

**T.C.
TRAKYA ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK VE TIBBİ BİLİŞİM
ANABİLİM DALI
YÜKSEK LİSANS PROGRAMI**

Tez Yöneticisi
Prof. Dr. Necdet SÜT

**VERİ MADENCİLİĞİNDE VERİ DÖNÜŞTÜRME
YÖNTEMLERİNİN SINIFLANDIRMA
ALGORİTMALARININ PERFORMANSLARINA OLAN
ETKİSİ**

(Yüksek Lisans Tezi)

Fatma Betül ÖRS

Referans no: 10322788

EDİRNE – 2020

**T.C.
TRAKYA ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK VE TIBBİ BİLİŞİM
ANABİLİM DALI
YÜKSEK LİSANS PROGRAMI**

Tez Yöneticisi
Prof. Dr. Necdet SÜT

**VERİ MADENCİLİĞİNDE VERİ DÖNÜŞTÜRME
YÖNTEMLERİNİN SINIFLANDIRMA
ALGORİTMALARININ PERFORMANSLARINA OLAN
ETKİSİ**

(Yüksek Lisans Tezi)

Fatma Betül ÖRS

Destekleyen kurum:

Tez no:

EDİRNE – 2020

T.C.
TRAKYA ÜNİVERSİTESİ
Sağlık Bilimleri Enstitü Müdürlüğü

O N A Y

Trakya Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı yüksek lisans programı çerçevesinde ve Necdet Süt danışmanlığında yüksek lisans öğrencisi Fatma Betül ÖRS tarafından tez başlığı “Veri Madenciliğinde Veri Dönüştürme Yöntemlerinin Sınıflandırma Algoritmalarının Performanslarına Olan Etkisi” olarak teslim edilen bu tezin tez savunma sınav 23/01/2020 tarihinde yapılarak aşağıdaki jüri üyeleri tarafından “**Yüksek Lisans Tezi**” olarak kabul edilmiştir.


İmza

Prof. Dr. Necdet SÜT

JÜRİ BAŞKANI (DANIŞMAN)


İmza

Doç. Dr. Hayriye Ertem VEHİD

JÜRİ ÜYESİ


İmza

Dr. Öğr. Üyesi Selçuk KORKMAZ

JÜRİ ÜYESİ

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylıyorum.

Prof. Dr. Tammam SİPAHİ

Enstitü Müdürü



TEŞEKKÜR

Tez çalışmam ve yüksek lisans eğitimim sürecinde bilimsel katkılarını ve hoşgörüsünü esirgemeyen Tıp Fakültesi Biyoistatistik Anabilim Dalı Başkanı ve tez danışmanım Sn. Prof. Dr. Necdet Süt'e sonsuz teşekkürlerimi sunarım. Bu süreçte yardımlarını ve desteklerini esirgemeyen Sn. Dr. Öğr. Üyesi Selçuk Korkmaz başta olmak üzere değerli dostum Merve'ye, arkadaşlarım Ecem ve Beyza'ya en içten teşekkürlerimi sunarım. Ayrıca her zaman yanımda olan ve bana desteğini sunan sevgili babama teşekkürü borç bilirim.

İÇİNDEKİLER

GİRİŞ VE AMAÇ	1
GENEL BİLGİLER	3
VERİ MADENCİLİĞİ	3
SINIFLANDIRMA	12
GEREÇ VE YÖNTEMLER	37
BULGULAR	41
TARTIŞMA.....	143
SONUÇ	148
ÖZET	150
SUMMARY	151
KAYNAKLAR.....	153
ŞEKİLLER LİSTESİ	164
ÖZGEÇMİŞ	167

SİMGE VE KISALTMALAR

DVM	: Destek vektör makineleri
NB	: Naif Bayes
EG	: Eşit genişlikte aralıklandırma
EF	: Eşit frekansta aralıklandırma
PKD	: Pozitif kestirim değeri
NKD	: Negatif kestirim değeri

GİRİŞ VE AMAÇ

Tıbbi uygulamalarda olumlu sonuçlar almak için doğru karar vermek büyük önem taşımaktadır. Sağlık alanında ve teknolojiye yaşanan gelişmeler sevindirici olmakla birlikte tıpta karşılaşılan bir takım belirsizlikler, klinisyenlerin karar vermesini güçleştirmektedir. 1975'ten itibaren hızla artan tanı ve tedavi seçenekleri ve paralelinde önemi artan maliyet yönleri, sağlık hizmeti sunucularına karar vermede zorluk yaratmaktadır (1). Ek olarak, kullanılan tanı testlerindeki kusurların ve tedavilerde etkili olabilecek değişkenlerin (klinisyenlerin uzmanlığı, hastaların demografik özellikleri, tıbbi geçmişi vb.) yarattığı belirsizlikler, doğru kararların verilmesini daha da zor hale getirmektedir (1-3). Tüm bu karmaşıklık karşısında klinisyenlerin bireysel tecrübelerine bağlı olmadan; objektif ve olası çelişkilerden uzak kararların verilebilmesi için bir çerçeve oluşturacak farklı yaklaşımlara ihtiyaç duyulmuştur (1,4). Birçok disiplini bünyesinde barındıran veri madenciliği, yeni klinik ilkeler geliştirmek ve edinilen bilgiler ile bilimsel açıdan kaliteli kararlar ortaya koyabilmek için kullanılan yardımcı araçlardan birisi olmuştur (1,5).

Veri madenciliği, araştırma yapılan alanda geçerli ve yararlı yeni bilgiler elde etmek için genellikle veri toplama, analize hazırlama, modelleme gibi adımlardan oluşan bir süreç olarak tanımlanmaktadır (5,6). Bu tezin temel konusu olan sınıflandırma, veri madenciliği tekniklerinden olup naif Bayes, karar ağaçları ve destek vektör makineleri başlıca kullanılan yöntemlerindedir.

Naif Bayes algoritması, Bayes teoremine dayanan sade bir sınıflandırma yöntemidir. Naif Bayes algoritmasının nadir sağlanan, güçlü varsayımları vardır. Buna rağmen birçok

problemleri veri karşısında iyi sonuçlar verebilmektedir. Sınıflandırma problemlerinde yeni bir gözlem için bağımlı değişkendirdeki sınıfların olasılıkları ile sınıf koşullu olasılıkların kullanılmasıyla sınıf tahmini yapılmaktadır (7). Karar ağaçları, veri setindeki değişkenler için belirli bir önem sırası göz önünde bulunduran hiyerarşik yapıda, gözetimli sınıflandırma yöntemidir (8,9). Burada değişken seçme yöntemleri ile değişken sırası belirlenir ve yukarıdan aşağıya doğru bir yol izlenerek yeni gözlemlerin sınıfları için karar verilir (10,11). ID3 karar ağacı algoritmasının geliştirilen versiyonu olan C5.0, sürekli veriyi modelleyebilmesi, çok boyutlu verilerde özellik seçimini otomatik olarak yapabilmesi, modelin önceden belirlenmesinin gerekli olmaması gibi avantajlara sahiptir (12). Çalışmada kullanılan diğer bir yöntem destek vektör makineleridir. Doğrusal olmayan ve çok boyutlu verilerde iyi performans göstermesi sebebiyle son yıllarda popülerliği artmıştır. Çok boyutlu uzayda bir sınıftaki gözlemler ile diğer sınıftaki gözlemler arasındaki alanı en büyükleleyen bir sınır çizerek sınıflandırma yapmaktadır (13).

Sınıflandırma yöntemlerinin avantajları ve dezavantajları vardır. Bu nedenle birbirlerine olan performans üstünlükleri, örneklem büyüklüğüne ve veri tipine bağlı olarak farklılık gösterebilmektedir. Daha sağlam ve dengeli sonuçlar elde edebilmek için algoritmaların dezavantajlı oldukları verilere, sınıflandırma öncesi bir takım yöntemler uygulanarak performans başarısında artış sağlanabilmektedir. Bu amaç doğrultusunda mevcut seçeneklerden birisi, veriler üzerinde dönüşüm uygulamaktır. Veri madenciliğinde, veri dönüşümü için minimum-maksimum normalizasyonu ve z-skor normalizasyonu sık kullanılmaktadır. Benzer amaçla, gruplandırma (binning) olarak da bilinen gözetimsiz diskritizasyon uygulamalarından eşit genişlikte ve eşit frekansta aralıklandırma yöntemleri kullanılan diğer bir yaklaşımdır (14).

Çalışmada, örneklem dağılımlarına (Normal, Ki-kare, F), gözlem sayılarına (100, 500, 1000, 10000) ve sınıf dağılım oranlarına (0,1, 0,2, 0,3, 0,4, 0,5) göre simülasyon ile üretilen farklı tipteki verilere, minimum-maksimum ve z-skor normalizasyonu ile eşit genişlikte ve eşit frekansta aralıklandırma diskritizasyon yöntemleri uygulanarak veri dönüşümünün, naif Bayes, C5.0 ve destek vektör makineleri üzerindeki etkilerinin incelenmesi amaçlandı.

GENEL BİLGİLER

VERİ MADENCİLİĞİ

Veri Madenciliği Tarihçesi

Bilgisayarların bellek kapasitesinde yaşanan iyileşmeler, diğer alanlar gibi sağlık kurum ve kuruluşlarının da daha fazla veri toplayıp saklamalarına katkı sağlamıştır. Giderek artan veri büyüklüğüyle birlikte verilerin işlenmesinde insanüstü bir performans ihtiyacı doğmuştur (15). 1950’lerde bilgisayarların keşfedilmesiyle oluşturulan veri madenciliği, 1980’lerin sonlarında bilgisayar gücü yardımıyla verilerin daha hızlı ve kolay işlenebilmesi amacıyla kullanılmaya başlanmış ve asıl popülerliğini 1990’larda kazanmıştır (5,15). Bünyesinde bulunan bazı algoritmaların ve yaklaşımların ortaya çıkışı çok daha eskiye dayanmakla birlikte “veri madenciliği” terimini ilk kez 1970 yılında Jorgenson ve ark. (16) yayımladıkları makalede kullanmıştır. Veri madenciliği günümüzde sağlık, ekonomi, finans, pazarlama gibi verilerinde kalıplar ve ilişkiler bulunan ve bu kalıpları bilgi avantajına dönüştürmek isteyen her alanda kullanılabilir.

Tıp’ta Uygulama Alanları

Tıbbi araştırmacılar, büyük veri tabanlı bilgi sisteminde depolanan verileri, verimli kullanma konusunda problem ile karşılaşmaktadır (17). Tıbbi bilgi sistemi veri tabanları, hasta kayıtları, hekim teşhisi ve hasta izleme bilgileri gibi pek çok veriyi içerir (18). Veri madenciliği, bu tür büyük verileri işleyerek sağlık sektörü araştırmacılarına farklı konularda

karar almalarında destek sağlamaktadır. Aşağıda çeşitli veri madenciliği uygulamaları sıralanmıştır (19):

1. Hastane kaynağının etkin yönetimi:

Sağlık kuruluşlarının, daha kaliteli bir sağlık hizmeti sunmak için hastane kaynaklarını etkili bir şekilde yönetmesi gerekmektedir. Hastane kaynaklarının yönetiminde, veri madenciliği oluşturduğu modeller ile yardımcı olmaktadır. Öncelikli hastaların belirlenmesi ve zamanında alınan etkili tedavi ile sonuçların iyileştirilmesi, tedavi maliyetlerinin düşürülmesi, internet aracılığıyla hastalara çeşitli hizmetler sunulması gibi hedeflere yönelik hizmet kalitesini artırıcı uygulamalar mevcuttur (20,21).

2. Hastane sıralamaları:

Hastanelerin sıralamasının belirlenmesinde gerekli analizleri gerçekleştirmek için farklı veri madenciliği yaklaşımlarından yararlanılmaktadır (22). Hastane sıralamaları, yüksek riskli hastalara öncelik verip tedavi edebilme yeteneklerine göre yapılmaktadır.

3. Daha iyi müşteri ilişkisi:

Veri madenciliği, sağlık kuruluşlarının hastalarla daha iyi ilişki kurabilmeleri için onların tercihlerini, ihtiyaçlarını ve davranışlarını anlamasına yardımcı olur. Müşteri Potansiyel Yönetimi A.Ş., tüketici sağlık hizmetlerine yönelik bir endeks geliştirmiştir.

4. Hastane enfeksiyon kontrolü:

Hastanelerdeki enfeksiyonu kontrol etmek için, veri madenciliği teknikleri ile uzmanlara yardımcı bir inceleme sistemi oluşturulmuştur (22).

5. Daha akıllı muayene teknikleri:

Veri Madenciliğinin kullanılmasıyla hekimler, mevcut tedaviler arasında hangi tekniğin daha etkili, aynı zamanda uygun maliyetli olduğunu öğrenebilir. Bunun yanında veri madenciliği, tedavilerin olası yan etkilerini belirlemede ve riskleri azaltmada yardımcı olur.

6. Geliştirilmiş hasta bakımı:

Dijital ortamda toplanan hastalara ait büyük veriler ile veri madenciliği, hastaların ihtiyaçlarını belirleyerek sağlık sistemi kalitesinin iyileşmesinde rol oynar (23).

7. Sigorta sahteciliğini azaltma:

Sağlık sigortaları, uygunsuz reçetelerin ve sahte işlemlerin saptanmasında veri madenciliği tekniklerini kullanmaktadır. Avustralya Sağlık Sigortası Komisyonu ve Texas Medicaid, hastane verilerini veri madenciliği teknikleri ile analiz ederek harcamalarda milyon dolarlık tasarruf sağlamıştır.

8. Yüksek riskli hastaları saptama:

Veri madenciliğinden elde edilen tahmin modeli ile sağlık hizmeti sunulan hasta, diğer hastalarla kıyaslanarak yüksek risk taşıyan hasta saptanır. Bu sistemde temel amaç, hastaya daha kaliteli sağlık hizmeti sunup aynı zamanda aldıkları hizmetlerin maliyetlerinde hastaların tasarruf etmelerini sağlamaktır (24).

9. Sağlık politikası planlaması:

Veri madenciliği, sağlık hizmetlerini iyileştirmek için etkili politikaların belirlenmesinde önemli bir role sahiptir. COREPLUS ve SAFS modelleri, hastanelerin sunduğu tıbbi bakım hizmetlerinin sonuçlarını ve tedavilerin maliyetini analiz etmek için veri madenciliği teknikleri ile geliştirilmiştir (25).

Veri Madenciliği Ve Bilgi Keşfi Süreci

Araştırmacılar, verilerdeki yapısal kalıpları bularak gizli kuralları ve ilişkileri keşfetmenin yanında ilgili alanda kullanılabilecek yeni bilgiler edinmeyi de istemektedir. Bu nedenle veri madenciliğinde amaç, büyük (genellikle) veriden kolay anlaşılabilir kurallar ile geçerli, kullanıma uygun yeni bilgiler keşfetmektir (26). Bilgi keşfi, verilerin nasıl depolanıp erişilebileceği, veri setlerini analiz etmek için algoritmaların nasıl kullanılacağı, insan ile makine arasındaki etkileşimin nasıl modelleneceği ve sonuçların nasıl sunulup yorumlanacağı dahil olmak üzere tüm bilgi çıkarma süreci ile ilgilidir (15,26). Veri madenciliği, genellikle “bilgilerin keşfedildiği süreç” olarak tanımlanmakla birlikte bilgi keşfi sürecinin bir parçası şeklinde de tanımlandığından bu sürecin işleyişini bilmek gerekmektedir (15,27).

Bilgi keşfi süreci:

Bilgi keşfi süreçlerinin daha iyi anlaşılması, projeyi planlarken ve uygularken takip edilecek bir yol haritası sunması için araştırmacıları ortak bir çerçevede buluşturacak standart bir model geliştirilmek istenmiştir (26). Bu amaç doğrultusunda 1990’lı yıllardan itibaren

gelişen bilgi keşfi süreci için birden fazla model önerilmiştir (28). Fayyad ve ark. (29-31)'nın (1996) dokuz aşamalı modeli akademik araştırmalara yönelik geliştirilirken altı aşamalı CRISP-DM (CRoss-Industry Standard Process for DM) modeli endüstri odaklı oluşturulmuştur (15,26). CRISP-DM, 1996'nın başlarında dört şirketten oluşan bir konsorsiyum (SPSS, NCR, Daimler Chrysler ve OHRA) tarafından geliştirilmiş ve resmi olarak 2000 yılında tanıtılmıştır (32,33). Cios ve ark. ise ilk olarak 2000 yılında CRISP-DM modelini akademik araştırma topluluğunun ihtiyaçlarına uyacak şekilde üzerinde değişiklik yaparak karma bir model önermiştir (34). 2005 yılında Cios ve Kurgan önceki modeli farklı alanlarda da kullanılabilir kılmak için adımların tanımlamalarında daha genel ve araştırma odaklı ifadeler ile değişiklik yaparak ikinci bir model ortaya sürmüştür (35).

Modelde yer alan adımlar aşağıdaki gibi tanımlanmıştır (35):

1. Problemleri alanı anlama:

İlk adımda, bir problemi tanımlamak ve proje hedeflerini belirlemek, terminolojiyi öğrenmek, kilit kişileri belirlemek ve probleme mevcut çözümler hakkında bilgi edinmek için alanda uzman kişilerle yakından çalışmak yer almaktadır. Problemin kısıtlamaları hakkında bilgi verilir. Proje hedefleri veri madenciliği ve bilgi keşfi hedeflerine çevrilmelidir ve potansiyel veri madenciliği araçlarının ilk seçimini içerebilir.

2. Verilerin anlaşılması:

Bu adım, veri formatına ve büyüklüğüne karar vermeyi ve örneklem verilerinin toplanmasını içerir. Mevcut bilgilere dayanarak değişkenler önemlerine göre sıralanabilir. Veriler, eksiksizlik, fazlalık, eksik değerlerin varlığı, değişken değerlerinin uygunluğu vb. açısından kontrol edilir.

3. Verilerin hazırlanması:

Bu adımda, hangi verilerin veri madenciliği yöntemleri için girdi olarak kullanılacağına karar verilir. Veri örnekleme, korelasyon ve önem testlerini uygulamayı ve verileri temizlemeyi (veri kayıtlarının eksiksizliğini kontrolü, gürültülü verilerin düzeltilmesi gibi) içerebilir. Temizlenen veriler daha sonra değişken seçimi ve çıkarma algoritmalarıyla boyutların azaltılması, diskritizasyon ile yeni değişkenlerin türetilmesi ve verilerin özetlenmesi (veri taneciklendirme) işlemlerinden geçirilebilir. Bunlar neticesinde kullanılması

planlanan veri madenciliği yöntemleri için belirli gereksinimleri karşılayan yeni veriler elde edilir.

4. Veri madenciliği:

Burada ön işlemden geçirilmiş verilerden yeni bilginin keşfedilmesi için çeşitli veri madenciliği yöntemleri kullanılır.

5. Keşfedilen bilginin değerlendirilmesi:

Değerlendirme, sonuçları anlamayı, keşfedilen bilginin yeni ve dikkat çekici olup olmadığını kontrol etmeyi, sonuçların etki alanı uzmanları tarafından yorumlanmasını ve keşfedilen bilginin etkisini kontrol etmeyi içerir. Yalnızca onaylanan modeller korunur ve sonuçları iyileştirmek için hangi alternatif önlemlerin alınabileceğini belirlemek amacıyla tüm süreç gözden geçirilir.

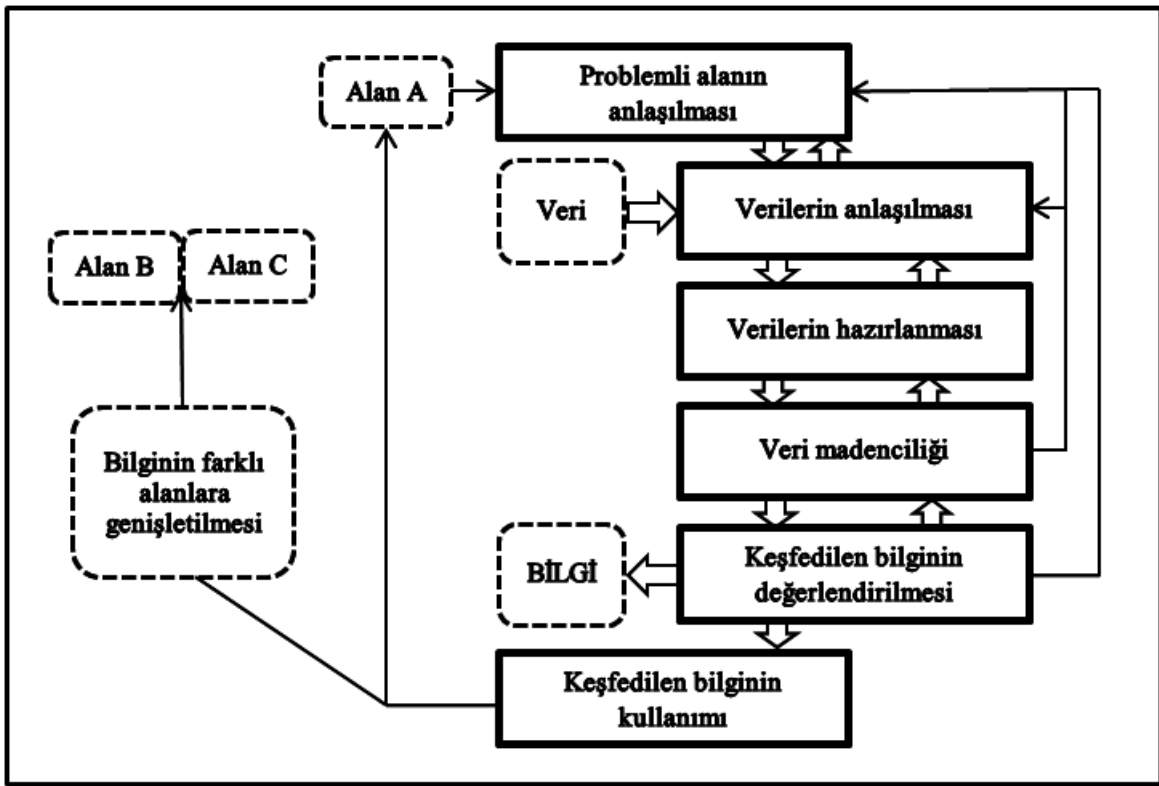
6. Keşfedilen bilginin kullanımı:

Bu son adım, keşfedilen bilginin nerede ve nasıl kullanılacağını planlamaktan ibarettir. Geçerli etki alanındaki uygulama alanı, diğer etki alanlarına genişletilebilir. Keşfedilen bilginin uygulanmasını izlemek için bir plan oluşturulur ve tüm proje belgelendirilir. Son olarak, keşfedilen bilgi dağıtılır.

Altı adımlı bu model, Şekil 1'de görselleştirilmiştir. Model, yinelemeli (iterative) ve etkileşimli (interactive) olduğundan bu süreçte gerektiğinde önceki adımlara geri dönüşler olabilmektedir (35). Modelde kullanılan bu mekanizma:

- Adım 2'den Adım 1'e; verileri daha iyi anlamak için alanda ek bilgiler gerekli olabilir.
- Adım 3'ten Adım 2'ye; belirli veri ön işleme algoritmaları (örneğin, veri dönüşümü veya diskritizasyon) seçilmeden önce veriler hakkında ek veya daha fazla özel bilgi gerekebilir.
- Adım 4'ten Adım 1'e; seçilen veri madenciliği yöntemleri tatmin edici sonuçlar üretmediğinde proje hedeflerinin değiştirilmesi gerekebilir.
- Adım 4'ten Adım 2'ye; verilerin yanlış yorumlandığı durumda veri madenciliği yönteminde hataya neden olabilmektedir. En yaygın senaryo, hangi yöntemin kullanılması konusunda verilerin yetersiz anlaşılması kaynaklı yaşanan belirsizliktir.

- Adım 4'ten Adım 3'e; veri hazırlama aşamasında bilinmeyen fakat sonra kullanılan yöntemin özel gereklilikleri nedeniyle tekrar verilerin düzenlenmesi gerekebilir.
- Adım 5'ten Adım 1'e; keşfedilen bilgi geçerli olmadığında tüm sürecin tekrarlanması gerekebilir. Alanın anlaşılmasında ya da yorumlanmasında, tasarımında, problem kısıtlamalarının, gereksinimlerinin ya da hedeflerinin anlaşılmasında yapılan yanlışlıklar bu problemin kaynağı olabilmektedir.
- Adım 5'ten Adım 4'e; keşfedilen bilgilerin yeni, ilginç veya yararlı olmadığı durumlarda, farklı yöntemler seçilerek 4. Adım tekrarlanabilir.



Şekil 1. Altı adımlı bilgi keşfi süreci

Verilerde Ön-Hazırlık İşlemleri

Veri madenciliği algoritmalarının başarısı verilerin kalitesiyle doğrudan ilişkili olduğundan ön-hazırlık işlemlerinin önemi büyüktür. Verilerin kalitesi ne kadar yüksek olursa algoritmalarından elde edilen sonuçların kalitesi de o derece yüksek olacaktır (14). Günümüzde ham veriler, artan büyüklükle beraber gürültülü veriler (noisy data), eksik değerler, hatalı veri girişi gibi farklı sorunlara daha açık hale gelmiştir (14,36). Bunlara ek olarak eldeki ham veriler, büyüklüklerine bağlı olarak süreci fazla uzatabilmektedir. Bu tür sorunlar nedeniyle

veri madenciliği algoritmalarını uygulamadan önce ham verileri birkaç açıdan ele almak gerekmektedir.

Veri temizleme (Data cleansing):

Öncelikle ham veriler; gürültülü veriler, aykırı değerler, eksik değerler, tekrarlanmış/kopyalanmış, hatalı kaydedilmiş veriler, süresi dolmuş veriler vb. yönlerden kontrol edilmelidir (36). Gürültülü ve aykırı değerleri saptamada veri özetlerinden veya grafiklerden yararlanılabilir (14). Gürültülü verilerin (verilerde rastgele hatalar sonucunda meydana gelen sapmaların) ve aykırı değerlerin varlığında düzgünleştirme (smoothing) teknikleri uygulanarak etkileri azaltılabilir veya yok edilebilir (14,36). Eksik değerler söz konusu olduğunda; ilgili gözlemlerin dikkate alınmaması, eksik değerlerin kestirimi, nicel değişkenler için ortalama veya medyan, nitel değişkenler için mod gibi değerlerle doldurulması uygulanan yaklaşımlardandır (36,37). Hatalı veri girişleri, doğru veriler ile düzeltilmeli, aynı verilerin varlığında ise tekrarlanan veriler silinmelidir (36).

Veri entegrasyonu:

Birden fazla kaynaktan toplanan verilerde aynı bilgiyi temsil eden bazı değişkenler, farklı isimlerde yazılmış veya farklı şekilde kodlanmış olabilir. Verilerin uyumlu biçimde birleştirilmesi için her bir değişkenin veri tiplerine, uygun değer aralıklarına, ilişkilerine yönelik bilgilere sahip olmak gerekmektedir. Verilerin uyum sorunu dışında, bir veya daha fazla değişken yoluyla elde edilebilen bir değişkenin varlığı, veri setinde “fazlalık” (redundancy) problemi yaratacaktır. Bu değişkenlerin saptanması için değişkenler arasındaki korelasyona bakmak gerekmektedir (14).

Veri indirgeme:

Birleştirme (aggregation), daha az sayıda değişken veya gözlemle daha “kararlı” veri elde etmek amacıyla, değişkenlerin veya gözlemlerin tek bir değişken veya gözlem altında birleştirilmesidir (36). Örnek olması açısından elimizde bir hastaneden elde edilen bölümlere göre farklı kan ürünlerinin haftalık transfüzyon uygulama sayısını gösteren bir veri olduğunu düşünelim. Birleştirme yoluyla haftaları aylara çevirebilir, -önemine göre- kan ürünlerini ayırmadan tek bir grupta toplayabiliriz. Böylece bölümlere göre aylık transfüzyon sayısını gösteren daha küçük bir veri elde etmiş oluruz.

Örnekleme, veri madenciliği yöntemlerini daha etkin kılmak için bilgi kayıplarına dikkat ederek verinin tamamını temsil etmek üzere daha küçük hacimde verinin çekilmesi işlemidir (36).

Özellik seçimi (feature selection), veri madenciliği algoritmasının performansında düşüşe yol açmadan modele anlamlı katkısı olmayan ve fazlalık yaratan değişkenleri veriden çıkartarak boyut azaltılmasında kullanılır (38). Değişken seçimi de denen bu yaklaşımda birçok yöntem kullanılmaktadır. Bu yöntemler özelliklerine göre filtreleme (filter), sarmal (wrapper), gömülü (embedded) yöntemler olmak üzere genel olarak üç grup altında toplanmıştır (39,40).

Veri dönüşümü:

Veri madenciliği uygulamalarında sonuçların daha verimli olması amacıyla veriler, uygulanması planlanan yöntemler için daha uygun verilere dönüştürülür. Veri dönüşümü için matematiksel dönüşümler (e^x , $\ln x$, x^n vb.), normalizasyon ve diskritizasyon gibi yöntemler kullanılmaktadır. Normalizasyon ile tüm sürekli değişkenler, aynı aralığa getirilerek ağırlıkları eşitlenir. Diğer bir yaklaşım olan diskritizasyon ile sürekli veriler, veri madenciliğinde nominal olarak analiz edilebilmektedir (13,35).

Veri Madenciliği Yöntem ve Teknikleri

Veriler arasındaki gizli kuralları ve ilişkileri bulmak için veri kümelerini araştırıp analiz etmede farklı yöntem ve teknikler kullanılmaktadır. Analizlerde kullanılan yöntem ve teknikler, kolay anlaşılabilirliği amacıyla özelliklerine göre gruplandırılır. Maimon ve ark. 2009 yılında yayımlanan çalışmalarında, veri madenciliği için oluşturdukları taksonomi Şekil 2'de gösterilmektedir (41).

Veri madenciliğinde yöntem ve teknikler arasındaki en temel ayırım, doğrulayıcı ve keşif odaklı olmalarına göre gerçekleşmektedir. Student t testi, ANOVA gibi klasik istatistiksel yöntemlerin dâhil olduğu doğrulayıcı yöntemler, bir hipotezin doğruluğunun test edilmesiyle ilgilidir ve veri madenciliği ile çok az ilişkilidir. Doğrulayıcı yöntemlerin aksine keşif odaklı yöntemler, veri madenciliğinin gizli yapıları, kuralları ve ilişkileri keşfetme amaçlarına yönelik yöntem ve tekniklerden oluştuğundan doğrudan veri madenciliğiyle ilişkisi bulunmaktadır (41).

Veri madenciliğinde kullanılan teknikler tanımlayıcı ve öngörücü özellikte olabilmektedir. Tanımlayıcı teknikler ile bir veri kümesinde gizli kalmış bilgiler ortaya çıkarılırken öngörücü teknikler ile mevcut bilgilere dayanarak nitel veya nicel yeni bilgiler tahmin edilebilmektedir (5). Keşif odaklılarda bu özelliklere göre ayrılan teknikler için tanımlayıcı yerine “gözetimsiz öğrenme” (unsupervised learning), öngörücü yerine “gözetimli öğrenme” (supervised learning) terimleri de kullanılmaktadır (41).



Şekil 2. Veri madenciliğinde yöntem ve tekniklerine ilişkin taksonomi

Veri madenciliğinde uygulanan tüm yöntemlerde bir girdiye (input) ihtiyaç vardır ve uygulama sonucunda bir çıktı (output) elde edilir. Çıktı; gizli kurallar, yapılar, sınıflandırmalar vb. ile ilgili üretilen yeni bir bilgiyi ifade etmektedir (41). Girdi ise veri madenciliği algoritmalarına öğretilmek üzere kullanılan değişken ve gözlemlerden oluşan veri setidir (42). Girdilerde gözetimli ve gözetimsiz öğrenme teknikleri, bağımlı değişkenin varlığına bağlı olarak birbirinden farklılık gösterir.

Gözetimsiz öğrenme teknikleri ile bağımlı değişkenin olmadığı verilerde gizli kalıplar belirlenir. Kümeleme (cluster), birliktelik (association) ve korelasyon (correlation) analizleri bu kategoride bulunan tekniklere örnek verilebilir (43). Gözetimli öğrenme tekniklerinde ise nitel veya nicel özellikteki bağımlı değişkenin yer aldığı veriler kullanılmaktadır. Bu tür

tekniklerde bağımsız değişkenler ile bağımlı değişken arasındaki ilişki modellenerek bağımlı değişken değeri bilinmeyen yeni gözlemler için tahminler geliştirilir (44).

Teknik ve yöntemler arasındaki sonraki ayırım, gözetimli öğrenme içinde bağımlı değişken türünden kaynaklanmaktadır. Bağımlı değişken değerleri nitel özellikte ise sınıflandırma, nicel özellikte ise kestirim (prediction) modelleri de denen regresyon modelleri kullanılmaktadır (12).

SINIFLANDIRMA

Sınıflandırma Öncesi Veri Dönüşümü

Veri madenciliğinde, veri dönüşümü uygulamaları veri problemlerini çözmeye sıkça tercih edilmektedir. Çalışmada, veri dönüşümü için normalizasyon ve gözetimsiz diskritizasyon yaklaşımları ele alındı.

Normalizasyon:

Verilerde bulunan sürekli değişkenler, birbirinden farklı aralıklarda dağılma eğilimi göstermektedir. Değişkenlerin bu dağılımı, ölçü birimlerinin farklılığından kaynaklanabilmektedir. Tıp'ta karşılaşılan bu değişkenlere örnek olarak hastalardan alınan ateş (C°) ve boy (m) ölçümleri verilebilir. Bazen de değişkenlerin, üre (mg/dL) ve kreatinin (mg/dL) gibi ölçü birimi aynı ancak referans aralıkları farklı olabilmektedir. Aralıklardaki farklılıklar, daha büyük aralıkta dağılan değişkenin, sonuçlar üzerinde daha büyük bir etkiye sahip olmasına yol açmaktadır. Bu nedenle, verilere dönüşüm uygulanarak aralık farklılığından kaynaklı etki arındırılmalıdır (37). Bu amaçla, normalizasyon yöntemleri kullanılmaktadır.

Normalizasyon, değişken değerlerini, daha küçük aralıklara ölçeklendirme işlemidir. Geniş aralıktaki değişkenlerin, küçük aralıktaki değişkenlere ağır basan etkisini standart hale getirebilme özelliği, normalizasyonu, özellikle mesafeye dayalı sınıflandırıcılar için yararlı kılmaktadır (14,37). İki farklı normalizasyon yöntemine yönelik açıklamalar aşağıda yer almaktadır (14):

a. Minimum-maksimum normalizasyonu:

Minimum-maksimum normalizasyonu, veride doğrusal bir dönüşüm uygulayarak minimum değer ve maksimum değeri, daha küçük yeni bir aralığın minimum ve maksimum

değerine dönüştürür. Minimum-maksimum normalizasyonu sonrasında, orijinal veri değerleri arasındaki ilişkiler korunur.

b. Z-Skor normalizasyonu:

Z-Skor normalizasyonu, değişken ortalamasını ve standart sapmasını kullanarak dönüşüm uygular. Bu yöntem, değişkenlerin gerçek minimum ve maksimum değeri bilinmediğinde veya aykırı değerlerin varlığında faydalıdır.

Diskritizasyon:

Veri madenciliği uygulamalarında, bazı algoritmalar sürekli verileri kendi içinde işlemekle birlikte, naif Bayes algoritmasındaki normal dağılım varsayımının aranması, karar ağaçlarının yavaşlaması gibi olumsuzluklar nedeniyle öğrenmenin verimi ve etkinliği azalmaktadır (45,46). Diskritizasyon uygulamaları, herhangi bir olasılık dağılım fonksiyonuna gerek olmadan, hızlı bir şekilde daha iyi sonuçlara ulaşmayı mümkün kılmaktadır (45,47).

Diskritizasyon, çok sayıdaki sürekli değişken değerlerini aralıklara bölüp birer etiket atayarak az sayıda kesikli değerlere dönüştürme sürecidir (14,48). Basit bir uygulama ile veriler indirgenerek anlaşılması ve yorumlanması kolay hale gelir. Elde edilen kesikli veriler ile daha yüksek sınıflandırma performanslarına ulaşıldığını gösteren çok sayıda çalışma mevcuttur. Setiono ve ark. (49) diskritizasyon yöntemlerinin, değişken seçimi için de kullanımının uygun olduğunu bildirmiştir. 1996 yılında yaptıkları çalışmada, diskritizasyon sonrası yeni değişkende tek bir grup oluşmuş ise bu değişkenin veriden çıkarılabileceğini söylemiştir.

Diskritizasyon uygulamalarında bir noktaya dikkat edilmesi gerekir. Sürekli değerlerin kategorik hale dönüşümü genel olarak bilgi kaybına yol açar. Bu nedenle başarılı bir diskritizasyon için uygulanan yöntemin, minimum seviyede bilgi kaybına neden olması beklenir (50).

Veri madenciliğinin aktif konularından biri olan diskritizasyon için birçok yöntem geliştirildi (46). Bu yöntemler; sınıf bilgisinin kullanımına (gözetimli (supervised) - gözetimsiz (unsupervised)), sürecin ilerleyiş yönüne (yukarıdan aşağıya (top-down) - aşağıdan yukarıya (bottom-up)), değişkenler arası ilişkinin dikkate alınıp alınmamasına (tek değişkenli (univariate) - çok değişkenli (multivariate)), değişken bilgilerinin tamamının veya

bir kısmının kullanımına (global (global) - lokal (local)) vb. durumlara bağılı olarak farklı kategoriler altında toplandı (14,51,52).

Çalışmada ele alınan eşit genişlikte aralıklandırma (equal-width interval) ve eşit frekansta aralıklandırma (equal-frequency interval), birer gözetimsiz, tek değişkenli ve global diskritizasyon yöntemidir (53,54).

Gözetimsiz diskritizasyon yöntemleri:

Sürekli değişken değerlerini, sınıf bilgisinden yararlanmadan kesikli hale dönüştüren yöntemlere gözetimsiz diskritizasyon yöntemleri denmektedir. Diskritizasyon sürecinde, veriler, küçükten büyüğe doğru sıralanır. Daha sonra diskritizasyon formülü ile aralık sınırları -kesme noktaları- belirlenir. Aralıklarda kalan sürekli değerler, ilgili aralığa atanan etiket değerine dönüştürülür (53).

Gözetimsiz diskritizasyon yöntemlerinden eşit genişlikte aralıklandırma ve eşit frekansta aralıklandırma, diğer yöntemlere göre daha eski olup onlar kadar sofistike değildir. (55).

a. Eşit genişlikte aralıklandırma diskritizasyonu (equal-width interval discretization):

Eşit genişlikte aralıklandırma (EG) diskritizasyonu, en basit yöntem olarak görülmektedir. Sürekli değişkenin dağılım aralığını, araştırmacı tarafından belirlenen k sayıda eşit genişlikte gruplara böler. Bölme sonucunda k-1 tane aralık oluşur (53).

Catlett (56) 1991 yılında gerçekleştirdiği çalışmada EG yönteminin aykırı değerlere karşı hassasiyetine dikkat çekmiştir. Bu yöntemin kısıtlılığı için belirlenen eşit genişlikteki aralıklarda, gözlem sayılarının dengesiz dağılımı söylenmektedir. Bazı aralıklarda daha fazla gözlem bulunurken bazılarında daha az hatta hiç gözlem bulunmayabilir (53).

b. Eşit frekansta aralıklandırma diskritizasyonu (equal-frequency interval discretization):

Eşit frekansta aralıklandırma (EF) diskritizasyonu, gözlem sayısını, k sayısına bölerek yaklaşık olarak eşit n/k sayıda gözlem içerecek şekilde değişkeni k sayıda gruba ayırır. Buradaki k, EG yönteminde olduğu gibi araştırmacı tarafından belirlenir. Bu yöntem, EG yönteminin kısıtlılığını aşarak gözlemleri aralıklara eşit sayıda dağıtmaya çalışır.

EF diskritizasyon yönteminin kısıtlılığı, ardışık olarak gelen aralıkların sınır değerleri aynı sürekli değere sahip olma ihtimalidir. Aynı değere sahip gözlemler farklı gruplara atanabilmektedir (57).

İki yaklaşımın genel özellikleri aşağıdaki gibidir:

Normalizasyon	Diskritizasyon
<ul style="list-style-type: none">➤ Farklı aralıklarda değişkenler, hatalı veriler, aykırı değerler gibi problemler, sınıflandırma için uygun hale getirilir.➤ Doğru sınıflandırma oranı artabilir.	<ul style="list-style-type: none">➤ Hatalı veri ve aykırı değer problemi giderilir ve sürekli veriler, bazı sınıflandırma algoritmaları için uygun hale getirilir.➤ Depolama alanında ve sürede tasarruf sağlanır.➤ Verilerin anlaşılıp yorumlanması kolaylaşır.➤ Doğru sınıflandırma oranı artabilir.

Sınıflandırma Tanımı ve Süreci

Sınıflandırma, araştırılan popülasyondaki bir gözlemi, bağımsız değişkenlerden elde edilen bilgilere göre nitel türdeki bağımlı değişkenin kategorilerinden birine yerleştirme işlemidir (8). Sınıflandırma uygulamalarında bağımlı değişkendeki kategorilerin her birine “sınıf” denmektedir. Tıbbi araştırmalarda bu sınıflara, üç farklı tedavi tipi, iki farklı hastalık teşhisi gibi örnekler verilebilir.

Sınıflandırma süreci, eğitim ve test olmak üzere temelde iki aşamadan oluşmaktadır. İlk aşamada, verilerin belirli bir kısmı bölünerek eğitim için kullanılmaktadır. Veri madenciliğinde algoritmalara girdi olarak verilen bu verilere “eğitim seti”, eğitim setlerinde bulunan bağımlı değişkenlere ise “sınıf değişkeni” denmektedir. Eğitim sonrasında sınıf değişkeni ile bağımsız değişkenler arasındaki ilişkiye göre model parametreleri belirlenir ve sınıflandırma için uygun bir kural oluşturulur. İkinci aşamada ise bu kurala göre yapılan sınıflandırmaların doğruluklarına bakılmaktadır. Bu aşamada, “test seti” denilen veri setinin geriye kalan kısmı kullanılarak sınıflandırma gerçekleştirilir ve çıktı olarak elde edilen sınıflar ile gözlemlerin gerçek sınıfları karşılaştırılır (12,58).

Veri setinin iki farklı bölüme ayrılması sınıflandırma yönteminin performansını değerlendirmek için önemlidir. Verinin tamamının eğitim için kullanılması durumunda, algoritma tüm bilgileri öğreneceği için veri setinden verilen bir gözlemin sınıfını otomatik

olarak söyleyecektir. Yeni bir gözlem için bakıldığında ise gerçek sınıfı bilinmediği için sınıf tahmininin doğruluğu hakkında yine emin olunamayacaktır. Her iki durumda da sınıflandırma modelinin başarısını değerlendirmek mümkün değildir (12). Bu nedenle verilerin, eğitim ve öğrenme sonucunda elde edilen modelin kontrolü için eğitim seti ve test seti olarak ikiye bölünmesi gerekli görülür.

Sınıflandırma sürecinde amaç, sınıfı bilinmeyen gözlemlerin sınıflarını belirlemeye yarayacak modelleri veya sınıflandırıcıları bulmaktır (8,12). Veri madenciliğinde sınıflandırma problemleri için naif Bayes, C5.0 ve destek vektör makineleri sık kullanılan sınıflandırıcılardır.

Naif Bayes

Naif Bayes (NB), Bayes teoreminden geliştirilen bir sınıflandırma yöntemidir (8). İngiliz matematikçi Thomas Bayes'in (1763) toplam olasılık formülünün tersini alarak elde ettiği formülü, Bayesci yaklaşımların temelini oluşturmuştur (8,59). Bayes formülünde koşullu olasılık kullanılmaktadır. Koşullu olasılık, örneklem uzayından ek koşullarla ayrılan alt bölümündeki olaylara ilişkin olasılıklardır (60). Koşullu olasılığın matematiksel ifadesi, iki olayın kesişim olasılıklarının marjinal olasılığa bölümü şeklindedir (59). Bayes formülünde bulunan $P(C_i)$ olasılıklarına önsel olasılık denmektedir ve sınıflandırma öncesi bilinen sınıfların olasılığıdır. $P(X_j/C_i)$ ise sonsal olasılıktır ve sınıf bilgisi koşulundaki olasılığı ifade etmektedir (60).

$$\underline{\text{Bayes Teoremi}} \quad P(C_i / X) = \frac{P(X_1 \cap X_2 \cap \dots \cap X_p / C_i) * P(C_i)}{P(X_1 \cap X_2 \cap \dots \cap X_p)}$$

⇓

$$\underline{\text{Naif Bayes}} \quad P(C_i / X_1, X_2, \dots, X_p) = \prod_{j=1}^p P(X_j / C_i) * P(C_i)$$

Veri madenciliğinde p bağımsız değişken sayısı arttıkça sınıf koşullu kesişim olasılıklarının tamamına ulaşmak zorlaşacağı için çözümü karmaşık bir hal almaktadır. Naif Bayes yönteminde işlemleri kolaylaştırmak amacıyla her bir sınıftaki değişkenlerin yani sınıf koşullu değişkenlerin birbirinden bağımsız olduğu varsayılır. Bu varsayım birçok kaynakta “koşullu bağımsızlık” olarak geçmektedir (7).

Koşullu bağımsızlık varsayımı, $p(A \cap B)=P(A)*P(B)$ olasılık kuralı ile Bayes formülünü, $P(X_j/C_i)$ sınıf koşullu olasılıkların ve $P(C_i)$ önsel olasılıkların çarpımı şeklinde basitleştirir (7). Paydada bulunan $P(X_1 \cap X_2 \cap \dots \cap X_p)$, sabit bir değer olduğundan sınıf belirlemede fark yaratmayacağı gerekçesiyle göz ardı edilir (61).

İki sınıflı bir veride sınıf tahmini için aşağıdaki olasılık bilgileri elimizde olsun.

Önsel olasılıklar:

$$P(\text{Riskli grup} = \text{Evet}) = 9/20=0,45$$

$$P(\text{Riskli grup} = \text{Hayır}) = 11/20=0,55$$

Sınıf koşullu olasılıklar:

$$P(\text{Yaş} = <30 / \text{Riskli grup} = \text{Evet}) = 2/9 = 0,222$$

$$P(\text{Yaş} = <30 / \text{Riskli grup} = \text{Hayır}) = 6/11 = 0,545$$

$$P(\text{Gelir düzeyi} = \text{orta} / \text{Riskli grup} = \text{Evet}) = 4/9 = 0,444$$

$$P(\text{Gelir düzeyi} = \text{orta} / \text{Riskli grup} = \text{Hayır}) = 7/11 = 0,636$$

$$P(\text{Cinsiyet} = \text{Kadın} / \text{Riskli grup} = \text{Evet}) = 6/9 = 0,667$$

$$P(\text{Cinsiyet} = \text{Kadın} / \text{Riskli grup} = \text{Hayır}) = 4/11 = 0,364$$

$$P(\text{Sigara} = \text{Evet} / \text{Riskli grup} = \text{Evet}) = 6/9 = 0,667$$

$$P(\text{Sigara} = \text{Evet} / \text{Riskli grup} = \text{Hayır}) = 4/11 = 0,364$$

Yukarıdaki olasılıklar kullanılarak, 30 yaşın altında, orta gelir düzeyinde, sigara tüketen kadın hastanın hastalık riski altında olma ve olmama olasılıkları ayrı ayrı hesaplanır:

$$P(X / \text{Riskli grup} = \text{Evet}) = 0,222 \times 0,444 \times 0,667 \times 0,667 = 0,044$$

$$P(X / \text{Riskli grup} = \text{Evet}) P(\text{Riskli grup} = \text{Evet}) = 0,0198$$

$$P(X / \text{Riskli grup} = \text{Hayır}) = 0,545 \times 0,636 \times 0,364 \times 0,364 = 0,046$$

$$P(X / \text{Riskli grup} = \text{Hayır}) P(\text{Riskli grup} = \text{Hayır}) = 0,025$$

Sınıfı bilinmeyen yeni bir gözlem için en uygun sınıfı belirlerken en yüksek olasılık dikkate alınır. NB formülüyle her bir sınıf için hesaplanan olasılıklar arasından en yüksek olasılık değerine sahip sınıf, yeni gözlemin atanacağı sınıf olur (12,59). Örnekte hastanın hastalık riski altında olmadığına karar verilir.

Naif Bayes yönteminde sınıflandırma yaparken dikkat edilmesi gereken özel bir durum vardır. Bu özel durumla karşılaşıldığında çözüm olarak Laplace kestirimi uygulanmaktadır.

Laplace Kestirimi:

Naif Bayes yönteminde sınıflandırma için kullanılan eğitim verilerinde, sınıf koşulu altındaki tüm değişken değerlerinin gözlenmediği, başka bir ifadeyle; belirli bir sınıfa ait gözlemi bulunmayan değişkenlerin varlığı durumunda hesaplanacak sınıf koşullu olasılık, sıfır olacaktır. Bayes formülü çarpımsal bir formül olmasından dolayı, bu özel durum tüm olasılıkların çarpımının sıfır olarak bulunmasına sebep olacaktır. Yani bir değişken değerinin gözlenmemesi, diğer değişken değerlerine bakılmaksızın o sınıfın kestirimini imkânsız hale getirecektir. Bu probleme çözüm olarak bir düzeltme uygulanır: Gözlemi bulunmayan değişkene ait değerlerin tüm koşullu olasılıklarının payına -genellikle- 1 sayısı eklenip payda sayısı da toplamda eklenen 1 sayısı kadar arttırılır ve işlem bu şekilde gerçekleştirilir. “Laplace kestirimi” denilen bu düzeltme ile bir sınıftaki tüm değerler en az bir kez gözlenmiş olur (7).

Tom Mitchel kitabında (1997) Laplace kestiriminin genel tanımlamasını;

$$\frac{n_c + mp}{n + m}$$

formülü ile yapmıştır (62). Burada p, 1/değişken değer sayısını (değişkendeki kategori sayısı), n, bir sınıftaki eğitilen örneklerin sayısını, n_c, ilgili sınıftaki değişken değerinin eğitilen örneklerin sayısını ve m ise eklenilecek toplam örnek sayısını ifade eder. Genellikle mp=1 olacak şekilde ekleme yapılmaktadır. Eğitim setindeki diğer değişkenlerle karşılaştırmak önemli ise daha büyük sayılar eklenebilir (63).

Veri setlerinde sürekli deęişkenlerin varlığı, bazı algoritmaların genel işleyişine uymamaktadır. Olasılıklara dayanan formül yapısı sebebiyle naif Bayes bu algoritmalarından birisidir. Bu nedenle sürekli deęişkenler için farklı çözümler geliştirilmeye çalışılmıştır.

Sürekli deęişkenler:

Sürekli deęişkenler ile karşılaşılan verilerde, normal dağılımın olasılık yoğunluk fonksiyonu (Gaussian) kullanılarak sınıflandırma işlemi gerçekleştirilir. Bu nedenle naif Bayes yönteminde sürekli deęişkenler için normal dağılım varsayımından söz edilir. Fakat verilerin normal dağılması kesin olarak gerekli görülen bir varsayım değildir. Farklı dağılımlarda da NB algoritması kullanıma uygun olabilmektedir. Deęişkenlerde normal dağılım dışında bilinen başka bir dağılım gözleniyorsa uygun olasılık yoğunluk fonksiyonu kullanılarak benzer şekilde sınıflar kestirilebilir. Bilinen herhangi bir dağılıma uymayan deęişkenler için çekirdek (Kernel) yoğunluk kestirimi kullanılmaktadır (63). “Elements of Statistical Learning” (64) kitabında kullanılan sınıf koşullu olasılık yoğunluk fonksiyonu formülü aşağıdaki gibi verilmiştir.

$$f_j(X) = \sum_{k=1}^p f_{jk}(X_k)$$

Alternatif olarak sürekli veriler, ön-işleme sürecinde diskritizasyon yöntemleri ile kesikli hale dönüştürülerek sınıflandırılabilir (63). Bu yöntemleri uygulayarak sınıflandırma performansında artış sağlamak mümkündür. Dougherty ve ark. (65) 1995 yılında yaptıkları çalışmada, NB'nin diskritizasyon uygulanan verileri kullanarak daha düşük sınıflandırma hatası elde ettiğine dair kanıtlar sundu. Bu çalışmada, diskritizasyon yönteminin, sürekli deęişkenlerde olasılık dağılımlarıyla ilgili herhangi bir varsayıma gerek duymadığı için etkili olduğunu öne sürdüler.

Naif Bayes'in genel özellikleri:

Naif Bayes algoritmasında, sınıf koşullu bağımsızlık varsayımının yanında deęişkenlerin eşit derecede önemliliği varsayımı da vardır. Bu varsayımlarının gerçek dünyada çok nadir sağlanması sebebiyle ismine “naif” denmiştir (7,59). Küçümseyici ismine rağmen NB, gerçek veri kümeleri üzerindeki uygulamalarda, özellikle deęişkenlerin

bağımsızlık problemini çözebilen özellik seçme prosedürlerinin bir kısmı ile birleştirildiğinde, karmaşık algoritmalarından daha iyi sonuçlar verebilmektedir. (63,64).

Güçlü varsayımlara rağmen NB, gürültülü verilerin ve anlamlı etkisi olmayan değişkenlerin varlığında sağlam bir sınıflandırıcı özelliğindedir (66,67).

Naif Bayes'in diğer bir özelliği de sınıflandırma sürecinde eğitim setindeki tüm bilgileri kullanmasıdır. Veri madenciliğinde kullanılan birçok algoritma düşük etkili değişkenleri görmezden gelirken NB, kestirilen olasılığa etkisi küçük bile olsa tüm değişkenleri işleme katar (7).

Naif Bayes algoritmasının kullanıldığı sınıflandırma uygulamalarında, bilgi güncellemesi için sıfırdan bir eğitim süreci gerektirmez. Bu nedenle NB, eğitim bilgilerinin sıklıkla güncellendiği sınıflandırma uygulamaları için uygun bir sınıflandırıcı olarak görülmektedir (68-70).

Tıpta naif Bayes:

Naif Bayes, Bayesci yaklaşıma dayandığından kullanımı birçok algoritmadan daha eskiye dayanmaktadır. Tıp'ta ilk kez 1961 yılında Warner ve ark. (71) tarafından bir çalışmada kullanılmıştır. Bu çalışma, kalp hastalığının tanısında yardımcı olasılıksal bir bilgisayar modeli geliştirmek amacıyla gerçekleştirilmiştir. İlerleyen yıllarda da özellikle tanı koyma araştırmalarında kullanılmaya devam etmiştir. 2016 yılında yayımlanan bir sistematik derleme çalışmasında, naif Bayes algoritmasının kolay uygulanabilirliği ve diğer sınıflandırıcılardan bir şekilde daha iyi sonuçlara ulaşılabilmesi nedenleriyle tıpta tanı belirlemede ve daha güvenilir kararlar vermede kullanıma uygun olduğu söylenmiştir (72).

Tanı koymak amacıyla NB, görüntü işleme çalışmalarında da kullanılmaktadır. Biyopsi örneklerinin mikroskopik incelemelerinde pratisyenlerin sübjektif kararları ve uzun süren tanı koyma süreci birer problem teşkil etmektedir. Adi ve ark. (73) belirttikleri bu problemler nedeniyle akciğer kanseri tanısı için 2017 yılında dijital görüntü işleme sistemi geliştirdi. Gri seviyeli eş oluşum matrisi (GLCM) ile özellik çıkarımı yaparak uyguladıkları naif Bayes algoritması ile biyopsi görüntü sınıflandırmasında %88.57 doğruluk oranı elde ettiler. Çalışmaları neticesinde biyopsi mikroskopik inceleme sürecinde dijital görüntü işleme tekniklerinin uygulanabilir olduğunu bildirdiler. Yang ve ark. (74) 2012 yılında yayımlanan makalelerinde %93,4 doğru sınıflandırma başarısına ulaştıkları naif Bayes sınıflandırmasına

dayanan elde taşınabilir ultrason ile memede lezyon belirleme sistemini önermiştir. Just ve ark. (75) 2017 yılında intihar düşüncesi olan kişilerde ölüm ve yaşamla ilgili kavramların (en ayırt edicileri; ölüm, zulüm, bela, kaygısız, iyi ve övgü), fMRI nöral göstergelerdeki değişimlerine göre intihar riskini değerlendirmek amacıyla yaptıkları araştırmada, intihar düşüncesi olan ve kontrol grupları için 17'şer kişi kullanmıştır. %91 oranında doğru sınıflandırma başarısı gösteren çalışmalarının biyolojik ve nörobilişsel bir temel oluşturduğunu söylediler.

Genetik ve ilaç geliştirme çalışmalarında NB ile başarılı sonuçlar elde edilmektedir. Jiang ve ark. (76) 2018 yılında senkron tümörlerde doku kaynaklı hepatobiliyer veya pankreatik kanser tanısının konmasında yardımcı olabilecek RNA dizilimi verileri kullanılarak naif Bayes algoritmasına dayanan bir yaklaşım geliştirdi. Oluşturulan model ile 10'lu çapraz geçerlilik yöntemiyle %95'ten yüksek başarıya ulaşılırken dış geçerlilik için kullanılan toplam 18 klinik kanser doku örneğinden (altısı negatif kontrol) 17 tanesini (%94,4) doğru olarak sınıflandırdıklarını bildirdiler. AK ve ark. (77) Leishmania donovani membran proteinlerinin ilaç hedefi veya aşı adayı olması açısından rolünün tahmin edilmesi amacıyla NB, EF, SVM, C4.5 sınıflandırıcılarını kullanarak yaptıkları çalışmada, 10'lu çapraz geçerlilik testi ile naif Bayes'in, %76.17 doğruluk oranıyla en yüksek sınıflandırma başarısı gösterdiğini bildirmiştir.

Türkiye'de naif Bayes uygulamaları:

Günakan ve ark. (78) 2019 yılında endometriyal kanserde lenf nodu tutulumunun kestirimi amacıyla naif Bayes algoritmasını kullanarak endometriyal kanser hastaları için tahmin modelleri oluşturdu. 2017 yılında Sayın ve ark. (79) tarafından karpal tünel sendromu üzerine 109 kişi ile gerçekleştirilen sınıflandırma çalışmasında, destek vektör makineleri, karar ağaçları, yapay sinir ağları ve naif Bayes yöntemleri kullanıldı ve en iyi performansı naif Bayes gösterdi.

Karar Ağaçları

Gözlemlerin sınıflandırılmasında kullanılan birçok algoritma, veri setindeki tüm değişkenleri aynı anda kullanmaktadır. Fakat gerçek hayatta her zaman değişkenlerin tamamına erişmek mümkün olmayabilir, hatta hepsini kullanmaya gerek bile olmayabilir. Bir doktor, tanı koymak için mümkün olan tüm testler yerine hastanın semptomlarını göz önünde bulundurarak belirli bazı testleri istemektedir. Öncelikli olarak istediği bu testlerin sonuçları,

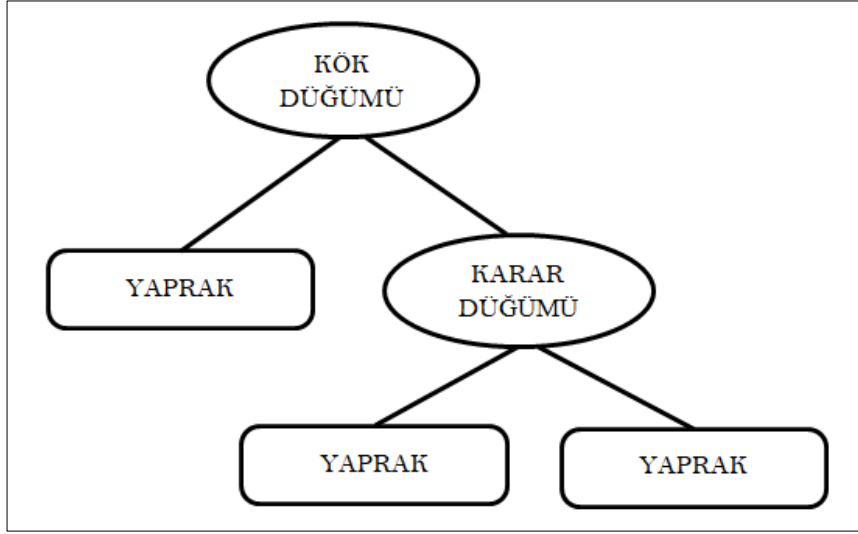
tanı koymada yeterli gelmediği takdirde daha sonra farklı testler de talep edebilmektedir (9). Gerçek hayatta karşılaşılan bu tip örneklerde olduğu gibi sınıf bilgisini kullanarak değişkenlerin önem sırasını göz önünde bulunduran, hiyerarşik yapıda gözetimli sınıflandırma yöntemleri geliştirilmiştir (8,9).

Tıbbi veri analizlerinde, veri madenciliği sonuçlarının insanlara anlaşılır bir şekilde iletebilmesi büyük önem taşımaktadır. Bu bakımdan, analiz sürecinde insan müdahalesine açık, tercihen düşük karmaşıklıkta fakat yüksek şeffaflıkta sonuçlar verebilen sembolik modellerin üretilmesi amacıyla sembolik algoritmalar kullanılmaktadır (80). Bir görsel sunarak yüksek doğruluk oranlarına ulaştırabilen ve “Eğer-İse” kuralına dayanan hiyerarşik yapıdaki yöntemlerden birisi olan karar ağaçları, bu amaç için sıkça kullanılmaktadır (9,80). Bazı araştırmacılar tarafından karar verme problemlerinde insan mantığına en yakın çalışan algoritma olduğu düşünülen bu yöntem, kolay anlaşılabilir ve kolay yorumlanabilen sonuçlar vermektedir (10,81). Nispeten hızlı olması, nicel ve nitel değişkenli karışık tipteki verilerde ve kayıp veri içeren veri setlerinde kolayca uygulanabilmesi, aykırı değerlere karşı dirençli olması gibi diğer başlıca özellikleri, karar ağaçlarını popüler kılmada etkili olmuştur (58,81). Bahsedilen bu tür özellikleri nedeniyle geliştirilmelerinden itibaren çok sık tercih edilmiş ve tıp, psikoloji, finansal analiz, moleküler biyoloji, astronomi, botanik ve bilgisayar bilimleri gibi çok farklı disiplinlerde kullanılmıştır (10,12,58).

Karar ağaçlarının temelini, araştırmacıların birbirinden habersiz olarak geliştirdikleri iki algoritma oluşturmuştur. 1970'lerin sonu, 1980'lerin başlarında makine öğrenme üzerine çalışan Quinlan (82), ID3 (Iterative Dichotomiser 3) olarak bilinen bir karar ağacı algoritması geliştirmiştir. İngiltere'de 1984 yılında, L. Breiman ve ark. (83) tarafından hem regresyon hem de sınıflandırma için geliştirilen algoritma, yazdıkları “Sınıflandırma ve Regresyon Ağaçları (Classification and Regression Trees - CART)” adlı kitapta tanıtıldı. Birbirine yakın zamanlarda geliştirilen ID3 ve CART algoritmalarında bağımsız olarak çalışılmış fakat verilerin eğitimi için benzer yaklaşımlar izlenmiştir (12,58). Yapı itibarıyla bir ağacı andıran bu yaklaşımlar, karar ağaçları olarak adlandırılmaktadır.

Karar ağaçlarının yapısı:

Bir karar ağacının yapısı, düğüm ve dallardan oluşmaktadır. Şekil 3'te gösterildiği gibi en üstteki başlangıç düğümüne “kök”, en sondaki bağımlı değişkene ait sınıfların yer aldığı düğüme ise “yaprak” denmektedir (11,12).



Şekil 3. Karar ağacı

Karar ağaçlarında yapraklar dışındaki tüm düğümler birer bağımsız değişkeni göstermektedir. Bağımsız değişken soruları (Örneğin; “Cinsiyeti nedir?” ve “Yaşı kaçtır?” gibi...) ile hangi sınıfa ulaşılabacağına karar verildiği için bu düğümlere, “karar düğümleri” de denmektedir (10,11). Her karar düğümünden çıkan dallar, değişkene ait tüm cevapları (Örneğin; “kadın - erkek”, “65 yaş ve altı - 65 yaş üstü” gibi...); farklı bir ifadeyle onların kategorilerini temsil etmektedir (12). İlgili değişkene ait kategorilerin sayısı kadar oluşan bu dallar, yaprak veya yeni bir karar düğümüne bağlanır. En son aşamada ulaşılan yapraklar, sınıflara ilişkin kararları ifade etmektedir (10).

Karar ağaçlarında sınıflandırma süreci:

Birçok karar ağacı algoritmasının temelinde “böl ve fethet (divide-and-conquer)” stratejisine dayanan Hunt’ın temel kavram öğrenme algoritması yatmaktadır (11). Burada, veri setinin tamamıyla kökten başlayıp sadece bir sınıftan olgular içeren altkümeler elde edilmesi amacıyla ardışık bir şekilde iki altkümeye bölünerek ilerleyen ve yaprakta son bulan yukarıdan aşağıya doğru bir yol (path) izlenir (10,11,58,61). Sınıflandırma sürecinde oluşturulan bu yollar, tüm sınıfları tanımlayan kurallar kümesini temsil etmektedir (11,84).

Sınıflandırma sürecinde izlenen yolun her bir adımında düğümler için en iyi bölünmeyi (split) sağlayan değişkenlere karar verilmelidir (10). Değişken kategorilerine göre gözlemlerin, bir sınıfta en yüksek oranda bulunmasına dikkat edilerek bilgi kazancı veya Gini indeksi gibi yöntemler ile bu değişkenleri belirlemek mümkündür (8,12).

Sınıflandırma süreci aşağıdaki gibi özetlenebilir:

1. Başlangıç düğümü olan kök düğümü için eğitim setini en anlamlı altkümelere bölen değişken seçilir (11). İkili bölünme gerçekleştiren algoritmalarda 2'den fazla kategorisi olan değişkenlerin kategorilerinde ikili olacak şekilde birleştirme yoluna gidilir. Sürekli değerlere sahip değişkenler için ise bir kesim noktası belirlenerek değişken değerleri iki grup altında toplanır. Her iki durumda da en yüksek sınıf oranlarına göre kategoriler seçilir (85).
2. Kök belirlendikten sonra veri seti alt kümelere bölünür ve işlem diğer karar düğümleri için de tekrarlanır (85).
3. Birinci ve ikinci adımlar süreç son bulana kadar devam eder. Tüm dallar bir yaprağa ulaştığı anda; altkümelerdeki gözlemlerin tamamı bir sınıfta toplandığında sınıflandırma süreci son bulur (9).

Değişken seçme kriterleri:

Bölme kriteri veya bölme kuralları da denilen değişken seçme kriterleri, her bir altkümedeki gözlemlerin safi (pure) bir sınıfa ait olduğu ideal duruma en yakın ayrımı sağlamaya yardımcı ölçülerdir (10,12). Bu ölçüler, her bir değişkenin sınıflar hakkında ne kadar bilgi taşıdıklarıyla ilgilenir (9). Aynı zamanda sürekli değişken değerlerinin dönüşümünde bölünme noktalarının belirlenmesi için de bu kriterler kullanılmaktadır (12).

En sık kullanılan ve yaygın olarak bilinen kriterler (8):

1. χ^2 kriteri, bağımsız değişkenlerin kalitatif veya kesikli düzende olduğu durumlarda kullanılabilir.
2. Gini indeksi, her tip bağımsız değişken için kullanılabilir.
3. Twoing kriteri, her tip bağımsız değişkende fakat bağımlı değişkendeki sınıf sayılarının 3'ten büyük olduğu durumda kullanılır; optimal bölünme için 3'ten fazla sınıf, iki sınıfa dönüştürülür.
4. Sıralı Twoing kriteri, sıralı düzendeki sınıf sayısı 3'ten büyük olduğunda bitişik sınıfların birleştirilmesiyle ikili sınıfa dönüştürülür.
5. Entropi veya bilgi kazancı, tüm bağımsız değişken tipleri için kullanılabilir.

Birden çok değişken seçme kriterinin varlığı, akla hangisinin en iyisi olduğu yönde sorular getirmektedir. Karar ağacının işlem süresinin, ağaç yüksekliğiyle katlanarak arttığı

yapılan çalışmalarda gösterilmiştir. Bu nedenle, daha sığ ağaçlar (örneğin, ikili bölme yerine çoklu ve daha dengeli bölmeler) üretebilme eğiliminde olan ölçümler seçilebilir. Bununla birlikte bazı araştırmalar, sığ ağaçlarda çok sayıda yaprağın oluştuğunu ve hata oranında artışın görüldüğünü bulmuştur. Yapılan karşılaştırma çalışmalarında, bir seçim ölçümünün diğer bir seçim ölçümüne göre belirgin bir üstünlüğü bulunamamıştır. Çoğu ölçüm oldukça iyi sonuçlar vermektedir (12).

C5.0 karar ağacı algoritması:

Quinlan, ID3 algoritmasını geliştirerek 1993 yılında C4.5 algoritmasını tanıttı (86). C4.5 ile ID3'ün aşağıdaki özellikleri iyileştirildi (87):

- a. Sürekli değişkenlerin kullanımı
- b. Kayıp verilerin kullanımı
- c. Farklı aralıktaki değişkenlerin kullanımı
- d. Karar ağaçlarını budama

C5.0 algoritması da C4.5'in geliştirilmiş versiyonudur. En dikkat çeken yeni özellikler için boosting ve winnowing teknikleri söylenebilir. Boosting teknikleri ile amaç, analizleri yinelemeli yaparak her yinelemeden sonra gözlem ağırlıklarını ayarlayıp tahminlerin güvenilirliğini artırmaktır. Sınıflandırmada hatalı sınıflandırılmış gözlemlere daha fazla ağırlık verilir, bu sayede genellikle daha iyi tahmin sonuçları elde edilir. Winnowing ise analiz için yararsız değişkenlerin elenmesi anlamına gelir (88). Bu özelliklere ek olarak gürültülü veri, aşırı öğrenme (over-fitting) ve hatalı budama problemlerine de çözümler getirilmiştir (89).

Entropi:

C5.0 algoritmasında kullanılan entropi, Claude Shannon (90) tarafından 1948 yılında bilgi teorisi üzerine yaptığı çalışmada tanıtıldı. Veri madenciliğinde, entropi terimiyle bilgi kazancı terimi birbiriyle ilişkilidir. Entropi, sınıflandırmada en uygun değişken sırasını belirlemek amacıyla en yüksek bilgi kazancını sağlayan değişkenlerin düğümlere seçilmesinde kullanılmaktadır (58). En yüksek bilgi kazancı sağlayan değişkenlerin seçilmesiyle sınıflandırmada gerekli görülen bilgi yani entropi minimum seviyeye indirilmiş olur (12). Eğer entropi değeri "0" olarak bulunursa gözlemlerin sınıflandırılması için daha

fazla bilgiye ihtiyaç duyulmadığı, tüm gözlemlerin aynı sınıfa ait olduğu anlaşılmaktadır. Entropi indeksine ait formül aşağıda verilmiştir (58):

$$I_{Entropi}(i) = - \sum_{j=1}^m f(i, j) \log_2(f(i, j))$$

$f(i, j)$, i düğümündeki altkümede bulunan m farklı sınıftan j sınıfına ait olguların oranını göstermektedir. Formülde bu oran, \log_2 tabanında yazılan haliyle çarpılmaktadır. Entropi bölme indeksin hesaplanmasında aşağıdaki formül kullanılmaktadır (58):

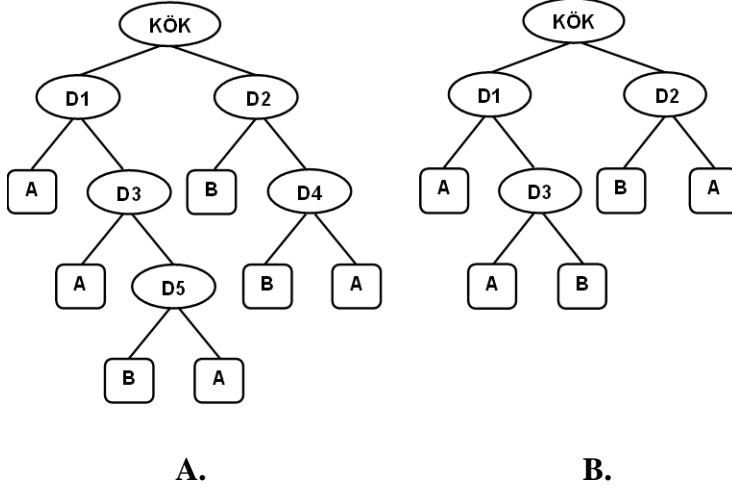
$$Entropi_{bölme} = \sum_{i=1}^p \frac{n_i}{n} I_{Entropi}(i)$$

En uygun bölme için hesaplanan entropi indeksleri arasında gözlenen en düşük değere sahip kategori ve kategorinin ait olduğu değişken seçilir (58). Entropi, sürekli değişkenlerin ikili hale dönüştürülmesinde de kullanılmaktadır (9).

Budama:

Bir karar ağacının büyük ve karmaşık yapıda olması, en büyük avantajlarından sade ve kolay anlaşılabilir olma özelliklerini yitirmesine neden olmaktadır. Ayrıca çok düğümlü karmaşık yapı, aşırı öğrenme (over-fitting) riskini doğurmaktadır (11). Özellikle ön işlemden geçirilmemiş ham verilerde gürültülü veya aykırı değerlerin varlığı bu soruna neden olmaktadır (58).

Aşırı öğrenme durumunda algoritma eğitilen veri setini aynen öğreneceğinden başka bir veri setinde kestirim amacıyla kullanmak sorun olacaktır (58). Sonuç olarak aşırı öğrenme, eğitim hatasını çok düşük kılmakla birlikte genel hata oranını (test hatası) yükselteceğinden istenmeyen bir durumdur (91). Bunun tam aksine az sayıdaki düğümden oluşan küçük bir ağaçta ise yetersiz öğrenme (under-fitting) sorunuyla karşılaşılabilir (83). Bu da hem eğitim hem de genel hata oranında yüksek sonuçlar elde edilmesine yol açmaktadır (91). Bu problemlere çözüm üretmek ve daha anlaşılabilir ağaçlar oluşturmak amacıyla ağaçlara budama yöntemleri uygulanmaktadır (11).



Şekil 4. A-Budama öncesi, B-Budama sonrası karar ağacı

Şekil 4'te bir karar ağacının budama öncesi ve budama sonrası gösterilmektedir. Budanan ağaçlarda eğitim setindeki doğruluk oranında ilk haline göre düşüş olabilirken genel doğruluk oranında artış gözlenebilmektedir (58).

Karar Ağaçlarının Genel Özellikleri:

Karar ağaçlarının genel özellikleri aşağıdaki gibi özetlenebilir:

1. Olasılık dağılımlarıyla ilgili herhangi bir varsayımı bulunmaması nedeniyle parametrik olmayan bir yöntem olarak görülmektedir (91).
2. Karar ağaçları, adım adım ilerleyerek her düğümde en iyi değişkene karar verir. Devamındaki düğümlerde en iyisi elde edilemese bile geri dönüşü olmadığı için açgözlü (greedy) bir yaklaşımdır (10).
3. Aynı veri setinden birden fazla karar ağacı oluşturulabilmesi nedeniyle buluşsal (heuristic) bir yaklaşımdır (61).
4. Büyük veri setlerinde bile kolayca hesaplanır ve hızlı bir şekilde ağaç oluşturulur (91).
5. Gürültülü verilere karşı sağlamdır (91). Sürekli değişkenler üzerinde yapılan matematiksel dönüşümlerden etkilenmez. Diğer sınıflandırıcılara göre ön-hazırlık işlemleri daha kolaydır (8).
6. Birbirleriyle yüksek ilişkisi bulunan değişkenlerin varlığında bu değişkenlerin biri kullanılmaktadır. Ancak çok sayıda ilişkili değişkenler bulunuyorsa karar ağaçlarını uygulamadan önce değişken seçme yöntemlerinin kullanımı daha doğru sonuçlara ulaştırmada yardımcı olabilmektedir (91).

Tıp'ta karar ağacı uygulamaları:

Karar ağaçları, tıbbi arařtırmalarda 20 yıldan fazla süredir sıkça kullanılmaktadır. 1997 yılında Cremilleux ve Robert (92), karar ağaçlarının tıptaki yerine yönelik genel bir çerçeve sundu. Bu konuda yapılan birkaç çalışmayla birlikte karar ağaçlarının evrimsel bir yaklaşım olduđu kararında uzlařıldı (93).

Kılavuz oluřturma ve klinik karar desteđi için geliřtirilen karar modellerinin, farklı bölgelerden kullanıcıların kolaylıkla eriřebilecekleri bađımsız yazılım eksikliđi nedeniyle kullanımı sınırlı olmaktadır. Bu eksikliđi gidermek için Sanders ve ark. (94) mevcut karar modelleri için "World Wide Web" tabanlı bir arayüz geliřtirdi. 1999 yılında yayımlanan bu çalışmada ani kardiyak ölümünü önleme stratejileri ile maliyet etkinliđini deđerlendiren bir karar modeli kullanılarak arayüzün işlevselliđi tanımlanır (93).

Adam ve ark. (95) 2002 yılında prostat kanserinin erken tespiti için daha iyi biyobelirteçler belirlemek amacıyla karar ağacı algoritması ile birleřtirilmiř bir protein biyoçip yüzeyi geliřtirilmiř lazer desorpsiyon / iyonizasyon kütle spektrometrisi yaklaşımı geliřtirdi. Prostat kanseri ile kanser dıřı (iyi huylu prostat hiperplazisi / sađlıklı) hastalardan alınan serumun geliřtirilmiř lazer desorpsiyonu / iyonizasyon kütle spektrometrisi protein profilleri sınıflandırmasında % 83 duyarlılık, % 97 özgülük, % 96 pozitif kestirim deđer ve %96 genel dođruluk oranı elde ettiklerini bildirdiler. Bu sınıflandırma sisteminin, prostat kanserinin erken tespiti için hassas ve yenilikçi bir yaklaşım sađlayacađını öne sürdüler.

Ekokardiyografi, renkli Doppler, BT ve MRG gibi yeni teknolojiler, oskültasyona göre kalp hastalıđı için daha dođru kanıtlar sađlamaktadır ancak bu yöntemlerin maliyeti yüksek, uygulanıřı karmařık ve büyük boyutundan dolayı kullanım yerleri sınırlıdır. Pavlopoulos ve ark. (96) kalp oskültasyonunun iç hastalıkları ve kardiyoloji eđitim programlarının çođunda deđerinin hafife alındıđı ve genç klinisyenlerin bu alanda yeterli eđitim almadıkları görüřündedir. Bu nedenle 2004 yılında klinisyenlerin kalp sesi teřhisini desteklemek amacıyla "temiz" Aort Stenozu ile "temiz" Mitral Regürjitasyon tanısı için karar ağaçlarına dayanan kural tabanlı bir yöntem geliřtirdi. Geliřtirilen bu yöntemin, karar destek sistemi için kullanıma uygun olduđu bildirildi.

Stukan ve ark. (97) kolorektal kanserden (mCRC) kaynaklanan yumurtalık metastazlarının klinik ve sonografik özelliklerini tanımlayarak mCRC'yi primer over kanserinden ayırmak amacıyla bir karar ağacı modeli geliřtirdi. 2019 yılında yaptıkları

çalışmada, sübjektif değerlendirme ve ADNEX modelinin değerlendirmesi ile karşılaştırdıkları bu model, mCRC ve primer OC ayırımında daha yüksek başarı gösterdi.

Türkiye’de karar ağacı uygulamaları:

Obstrüktif uyku apnesi sendromu, hem sağ hem de sol kalp ventrikülü etkileyen önemli bir hastalıktır. Polat ve ark. (98) (2008), C4.5 karar ağacı, yapay sinir ağı, yapay bağışıklık tanıma sistemi ve adaptif nöro-bulanık çıkarım sistemi dahil olmak üzere dört farklı algoritmayı karşılaştırmış ve obstrüktif uyku apne sendromu tanısında en iyi sınıflandırıcı sistemin C4.5 karar ağacı sınıflandırıcısı olduğunu göstermiştir.

Turan ve Şehirli (99) 2017 yılında komet görüntüleri kullanarak DNA hasarına ve derecesine karar vermek için dinamik zaman bükme yöntemi ve karar ağacı sınıflandırmasını birlikte uyguladıkları bir yaklaşım önerdi.

Yüz malformasyonları, dismorfik sendromlarda farklılık göstermesi sebebiyle bu sendromların erken teşhisi için önemli ve ayırt edici bilgiler içermektedir. Özdemir ve ark. (100) yüzdeki malformasyonları göz önünde bulundurarak farklı sendrom tiplerinin otomatik olarak tanınması amacıyla 2018 yılında karar ağacına dayanan bir sistem geliştirdi. Bir klinik uzman görüşüyle birlikte k-en yakın komşu ve yapay sinir ağı yöntemlerinden de sonuçlar aldıkları çalışmada, en yüksek başarıyı karar ağacının gösterdiğini bildirdiler.

Destek Vektör Makineleri

Destek vektör makineleri (DVM), ilk olarak 1963 yılında Vapnik ve Lerner (101) tarafından doğrusal sınıflandırma için önerildi. 1992 yılında ise Vapnik ve ark. (102) doğrusal olmayan sınıflandırmalar için algoritmada yaptıkları yeni düzenlemeleri içeren makaleyi yayımladı. Destek vektör makineleri algoritması, değişken sayısı kadar boyutlu uzayda tüm gözlemleri, değerlerine göre yerleştirir (103). Daha sonra gözlemler arasında sınıflarına göre birbirinden en iyi ayrımı sağlayan hiperdüzlemi bulur ve yeni gözlemler için bu hiperdüzleme göre sınıflandırma yapar (104).

Hiperdüzlem, p boyutlu uzayda p-1 boyutlu alt uzay düzlemidir. Matematiksel tanımlaması aşağıdaki denklem ile yapılmaktadır:

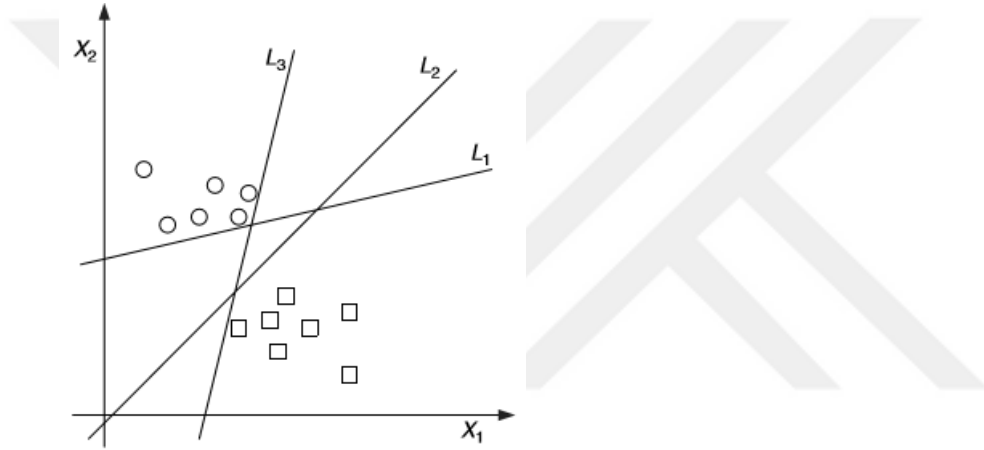
$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$$

Uzayda bulunan bir nokta için bu denklem uygulandığında;

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p < 0$$

$$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p > 0$$

sonuçları elde edilebilir. Sonuç negatif ise nokta bir sınıfta, aksi durumda yani sonuç pozitif ise diğer sınıftadır. Gözlemlerin, ikiye bölünen uzayda, hiperdüzlem denklem sonucunun işaretine göre hangi tarafta olduğu kolaylıkla bulunabilir (104).



Şekil 5. Farklı hiperdüzlemler tarafından noktaların ayrımı (105)

Mükemmel ayırım için hiperdüzlemler aşağı veya yukarıya doğru kaydırılarak veya döndürülerek birden fazla hiperdüzlem oluşturulabilir. Şekil 5'te 3 farklı hiperdüzlem örneğine yer verildi. Oluşturulan birden fazla hiperdüzlem arasından en iyi hiperdüzleme karar verilmelidir (104). Bu karar için en büyük ayrımı sağlayan maksimum marjin hiperdüzlemi (MMH) iyi bir seçim olacaktır (106). Marjin, eğitim setindeki her bir gözlemden hiperdüzleme olan dik mesafeler arasından en küçük uzaklıktır. MMH, gözlemlerin hiperdüzleme olan en küçük mesafesi yani marjini en büyük olan hiperdüzlemdir (104).

Destek vektörleri, her bir sınıftan MMH'ye en yakın olan noktalardır. Destek vektörleri, yüksek değişken sayısında bile bir sınıflandırma modeli için çok kompakt bir yol sağlar (106).

Doğrusal sınıflandırma:

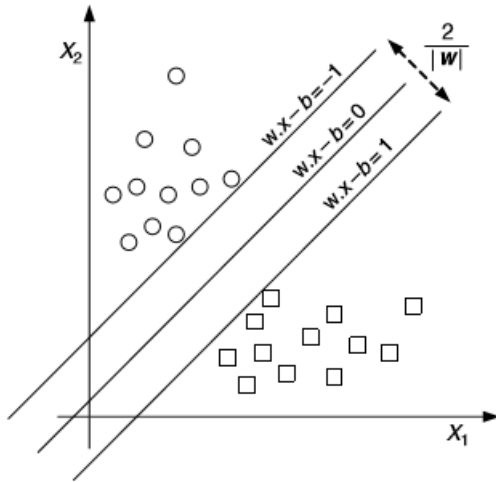
Destek vektör makinelerinde maksimum marjının, tanımlamalarında daha kolay anlaşılması için hiperdüzlemin doğrusal olduğu varsayılır (103,106). Sınıflar arası mümkün olan en homojen ayrımı sağlayan doğrusal hiperdüzlemi araştırma sürecinde p boyutlu uzay için aşağıdaki formül kullanılır:

$$\vec{w} \cdot \vec{x} - b = 0$$

Formülde bulunan \vec{w} , p ağırlık vektörü, b ise yandır. Maksimum marjini belirlemek için vektör makineleri ve birbirine paralel iki hiperdüzlemi bulmak gerekir. Şekil 6'da da gösterilen bu iki hiperdüzlem aşağıdaki formüller ile tanımlanmaktadır:

$$\vec{w} \cdot \vec{x} - b = 1$$

$$\vec{w} \cdot \vec{x} - b = -1$$



Şekil 6. Maksimum marjin hiperdüzlemi (105)

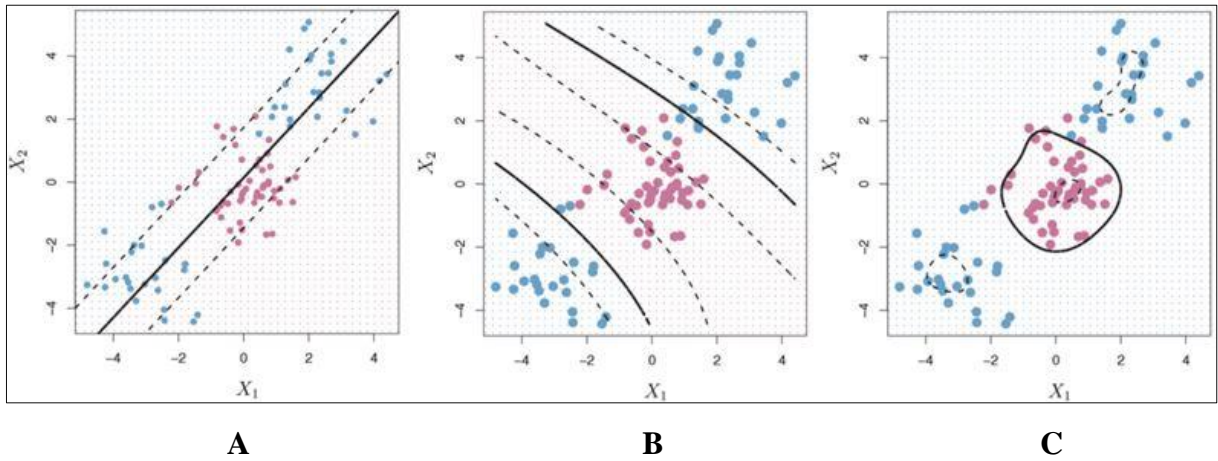
Eğitim verileri doğrusal olarak ayrılabilir özellikte ise bu hiperdüzlemler seçilerek mesafeleri maksimize edilmeye çalışılır. Geometri bilgisi ile hiperdüzlemler arasındaki mesafenin $2/\|\vec{w}\|$ olduğu bulunur. Formüldeki $\|\vec{w}\|$; orijinden \vec{w} vektörüne olan uzaklığı (Öklid form) ifade eder. Hiperdüzlemler arası mesafenin maksimum olabilmesi için $\|\vec{w}\|$ değerinin minimize edilmesi gerekir. Bu kuadratik programlama (QP) optimizasyon problemidir (105).

$$\begin{aligned} &\text{Minimize } \frac{1}{2} \|\bar{w}\|, \\ &\text{Kısıt } y_i(\bar{w} \cdot \bar{x}_i - b) \geq 1 - \xi_i, \\ &\xi_i \geq 0, \quad \sum_{i=1}^n \xi_i \leq C. \end{aligned}$$

Kısıtta yer alan y_i , işareti $[-1,1]$ olan gözlem sınıflarını temsil etmektedir (7). Kısıtta yer alan bir diğer sembol $\xi_i \dots \xi_n$, “gevşek değişkenler (slack variables)” olarak adlandırılır ve i . gözlemin marjine ve hiperdüzleme göre nerede konumlandığı bilgisini verir. Eğer $\xi_i = 0$ ise gözlem doğru taraftadır. Eğer $\xi_i > 0$ ise gözlem marjinin, $\xi_i > 1$ olduğunda ise hiperdüzlemin yanlış tarafındadır. Kısıtta yer alan bir diğer parametre olan C , bir ayarlama (tuning) parametresidir ve ξ_i 'lerin toplamı olarak idare edilebilir marjin ve hiperdüzlemin sınırlarının dışında olma şiddetini temsil eder. Uygulamalarda C sayısı azaldıkça, sınırlar daralarak marjin ihlali seyrekleşir ve böylece sınıflayıcının veri uyumunda artış gerçekleşir. Bu durumda yan miktarı düşük ancak varyans büyük olmaktadır. C sayısındaki artışta ise tersi durum gözlenmektedir (104).

Doğrusal olmayan sınıflandırma:

Gerçek dünyada karşılaşılan verilerde sınıflar, doğrusal hiperdüzlemin iki tarafına nadir olarak ayrılır. Şekil 7'de sol taraftaki grafikte görüldüğü gibi doğrusal hiperdüzlem, noktaları sınıflara göre ayırmada yetersiz kalmıştır.



Şekil 7. A-Linear Kernel Fonksiyonu, B-Polinom Kernel Fonksiyonu, C-Radyal Tabanlı Kernel Fonksiyonu (104).

Algoritma, doğrusal olarak ayrılamayan verilerde her gözlemi doğrusal olmayan bir çekirdek fonksiyonu kullanarak değiştirir (103). Bu çekirdekler ile eğitim verileri, doğrusal olarak ayrılabilceği daha yüksek boyutlu bir uzaya (çekirdek uzayı) dönüştürülür (107). Bu dönüşümlü değişken uzayında algoritma, maksimum marjin hiperdüzlemine uygun hale gelir (105).

Yaygın olarak kullanılan çekirdek fonksiyonları:

a. Lineer Kernel Fonksiyonu; $K(x_i, x_{i'}) = \sum_{j=1}^p x_{ij} \cdot x_{i'j}$

b. Polinom Kernel Fonksiyonu; $K(x_i, x_{i'}) = (1 + \sum_{j=1}^p x_{ij} \cdot x_{i'j})^d$

c. Radyal Tabanlı Kernel Fonksiyonu; $K(x_i, x_{i'}) = \exp(-\gamma \sum_{j=1}^p (x_{ij} - x_{i'j})^2)$

Kernel fonksiyonu seçimi büyük ölçüde verilere bağlıdır. Lineer kernel fonksiyonu, örneğin büyük seyrek verilerde (sparse data) sıkça kullanılırken polinom kernel fonksiyonu, yaygın olarak görüntü işlemede tercih edilir (107).

Destek vektör makinelerinin genel özellikleri:

1. DVM'nin başarısı, kernel seçimine ve parametrelerine bağlıdır. İyi bir model için kernel ve parametrelerin çeşitli kombinasyonlarının denenmesi gerekmektedir (106).
2. DVM, istatistiksel öğrenme teorisine dayanan bir yöntem olduğundan güvenilir bir teori temeli vardır (108).
3. Hem sınıflandırma hem de regresyon problemleri için kullanıma uygundur.
4. DVM, dışbükeylik özelliği olduğu için tek bir çözüm sunar. Yerel minimum çözümleri sunan sinir ağları ile karşılaştırıldığında mutlak (global) minimumu ile önemli bir avantaja sahiptir (105).
5. Hem eğitim hem de test sürecinde algoritma özellikle büyük verilerde hızı açısından dezavantaja sahiptir (105).

6. Büyük verilerde kernel fonksiyonları kullanıldığında, kernel matrisinin hesaplaması için büyük bellek ve hesaplama kaynakları gerekmektedir (107).

Tıp'ta destek vektör makineleri:

Brown ve ark. (109) 2000 yılında DNA mikrodizi hibridizasyon deneylerinden gen ekspresyon verilerini fonksiyonel olarak sınıflandırmak amacıyla destek vektör makineleri teorisine dayanan bir yöntem sundu. Çalışmalarında DVM'nin diğer sınıflandırıcılara göre genleri daha doğru sınıflandırdığını bildirdiler.

Protein-protein etkileşim bölgelerinin tanımlanması ve protein etkileşimlerinin özgüllüğüne ve gücüne katkıda bulunan spesifik amino asitlerin saptanması, akılcı ilaç tasarımı ve metabolik ve sinyal iletim ağlarının analizine kadar geniş uygulamalarda önemli bir problemdir. protein-protein etkileşim bölgeleri için kestirim yöntemlerinin gücünü artırmak amacıyla Sen ve ark. (110) 2004 yılında dört farklı yöntemi birleştirerek bir konsensüs metodolojisi geliştirdi. Bu birleştirme yaklaşımında, veri madenciliği yöntemlerinden destek vektör makineleri kullanılmış ve çalışma sonucunda diğer biyoinformatik problemlerinde, tahmin doğruluklarını iyileştirmek için benzer metodolojilerin geliştirilebileceği gösterildi.

Katalitik kalıntıların tanımlanması, protein fonksiyonu hakkında değerli bilgiler sağlayabilmektedir. Pugalenti ve ark. (111) dizi ve yapısal özellikler kullanılarak katalitik kalıntıların tanımlanması için 2008 yılında bir SVM yöntemi geliştirdi. Algoritma, 2096 katalitik kalıntıya uygulandı ve 254 katalitik tortu için test edildiğinde tamamını doğru şekilde tahmin edebilme başarısı gösterdi. Geliştirilen algoritmanın, protein yapılarından katalitik kalıntıların tanımlanmasını kolaylaştırmak için yararlı olduğu bildirildi.

Niu ve ark. (112) 2013 yılında aterosklerozun erken evresinde arteriyel pürüzlülüğü tanımlamak amacıyla karotis arter duvarının ultrason görüntülerinden çıkarılan doku özelliklerine dayanarak destek vektör makinesi sınıflandırıcısı ile bir yöntem geliştirdi. Çalışmada bu yöntem ile arteriyel yüzey pürüzlülüğünü tanımlamanın mümkün olduğu ve aterosklerozun erken saptanması ve de teşhisi için yararlı olduğu gösterildi.

Kalatsız ve ark. (113) tip II diyabet olan ve beyin hasarı olmayan hastaların beyin SPECT görüntülerinin sağlıklı deneklerden ayrımını araştırmak amacıyla 2013 yılında bir

görüntü işleme yöntemi geliştirdi. Bulgularında elde ettikleri yüksek sınıflandırma başarısı nedeniyle geliştirdikleri bu yöntemi, beyin SPECT uygulanan tip II diyabet hastalarında serebral kan akışını değerlendirmede doktorlara yardımcı bir sistem olarak önerdi.

Wei ve ark. (114) 2018 yılında herpes, dermatit ve sedef hastalıklarının teşhisi için destek vektör makinesine dayanan yeni bir yöntem geliştirdi. Başlangıçta cilt görüntülerini ön-işlemden geçirerek sınıflandırdıkları çalışmada, üç tip cilt hastalığının tespiti için elde ettikleri sonuçların yöntemin etkililiğini ve uygulanabilirliğini gösterdiğini söylediler.

Uddin ve ark. (115) 2019 yılında kronik obstrüktif akciğer hastalığında akut alevlenmelerin belirlenmesi, mortalite ve mali yükün azaltılması için büyük önem taşımaktadır. bu hastalık için farklı sınıflandırma modelleri geliştirmeyi ve aralarından en iyi modeli bulmak amacıyla karşılaştırma çalışması gerçekleştirdi. Elde edilen sonuçlardan DVM, en yüksek başarı performansı gösterdi.

Türkiye’de destek vektör makineleri uygulamaları:

Güler ve Koçer (116) 2005 yılında sağlıklı, nöropati ve miyopati deneklerinden elde edilen elektromiyogram (EMG) sinyallerinin sınıflandırılması amacıyla destek vektör makinesi (SVM) ve geri yayımlı sinir ağı algoritmalarını uyguladı. Kullandıkları iki algoritmanın performanslarını karşılaştırdıkları çalışmada, DVM daha iyi performans gösterdi.

Yılmaz ve ark. (117) sadece tek derivasyonlu bir elektrokardiyografi sinyalinden elde edilen veriler ile uyku evrelerinin ve obstrüktif apne dönemlerinin sınıflandırılma fizibilitesini araştırmak amacıyla 2010 yılında bir çalışma gerçekleştirdi. Çalışmalarında, kuadratik diskriminant analizi ile destek vektör makineleri benzer ancak her iki algoritma da k-en yakın komşu algoritmasından daha iyi sonuçlar vermiştir. Bu araştırmacılar, sadece tek derivasyonlu EKG gerektiren sınıflandırmanın, uyku evresi ve apne dönemi tanımlamaları için uygun olduğunu ve dolayısıyla evde kullanılabilecek basit bir otomatik sınıflandırma sistemi için yolu açabileceğini bildirdi.

Korkmaz ve ark. (118) ilaç keşfinin erken evresinde büyük bileşik koleksiyonlarında, aktif ve aktif olmayan molekülleri ayırt edebilen bir sınıflandırma yöntemi üzerine çalışma gerçekleştirdi. Farklı özellik seçme yaklaşımlarının (Pearson korelasyon katsayısı, özyinelemeli unsur eliminasyonu, sarıcı yöntemi ve altküme seçimi) karşılaştırıldığı bu

çalışmada, altküme seçimi ile en yüksek performans gösteren DVM'nin ilaç keşfinin erken evresinde sınıflandırma görevi için yararlı bir yöntem olabileceği bildirildi.



GEREÇ VE YÖNTEMLER

Veri ön-işleme sürecinde uygulanan normalleştirme ve z-skor yöntemleri ile gözetimsiz diskritizasyon yöntemlerinin, sınıflandırma algoritmalarının performansları üzerindeki etkileri, simülasyon ile üretilen veri setleri kullanılarak incelendi. Veri setleri, nicel türde 10 bağımsız değişken ve 2 kategoriden oluşan bir bağımlı değişken olmak üzere toplamda 11 değişkenli yapıda oluşturuldu.

Tablo 1. Örneklem dağılımı parametreleri

Değ.	Grup	Normal dağılım		Ki-kare dağılımı	F dağılımı	
		Ortalama	ss	SD1	SD1	SD2
1.	1	1,8	0,2	30	4	4
Değ.	0	2	0,2	32	5	5
2.	1	20	4	10	50	50
Değ.	0	22	5	8	10	10
3.	1	4	0,3	200	100	10
Değ.	0	4,2	0,2	195	100	50
4.	1	10	1	100	10	1
Değ.	0	9	0,8	95	5	15
5.	1	30	4	2	100	10
Değ.	0	28,5	3	4	100	5
6.	1	150	40	5	30	1
Değ.	0	180	50	7	10	5
7.	1	40	5	15	100	5
Değ.	0	41	4	17	100	50
8.	1	100	15	500	100	10
Değ.	0	95	20	475	10	10
9.	1	5	0,8	200	1	5
Değ.	0	4,5	0,5	210	10	5
10.	1	50	10	1000	10	5
Değ.	0	52	10	1050	10	3

Değ.: Değişken; **ss:** standart sapma; **SD1:** 1. serbestlik derecesi; **SD2:** 2. serbestlik derecesi.

Bağımsız değişken olarak üretilen nicel değişkenler, F dağılımı, ki-kare dağılımı ve normal dağılım olmak üzere üç farklı örneklem dağılımından üretildi. Örneklem dağılımları için kullanılan parametreler Tablo 1’de gösterilmektedir. Bağımlı değişkendeki kategoriler, “1” ve “0” olarak kodlandı. Bu kategoriler, 5 farklı dağılım oranlarında (sırasıyla oranlar: 0,1-0,9, 0,2-0,8, 0,3-0,7, 0,4-0,6, 0,5-0,5) oluşturuldu. Gözlem sayıları; 100, 500, 1000 ve 10000 olmak üzere dört farklı tipte ele alındı.

Çalışmada verilerin %80’i eğitim seti, geri kalan %20’si test seti olarak kullanıldı.

Verilerin üretimi ve çalışmada yapılan tüm analizler; R-Project (version 3.5.2) yazılımı ile entegre RStudio (Desktop 1.1.463) programı kullanılarak gerçekleştirildi. Verilerin üretimi ve simülasyondaki döngüler için “MASS (version 7.3-51.1)”, “copula (version 0.999-19.1)”, “extraDistr (version 1.8.11)”, “MiscTools (version 0.6-24)”, “foreach (version 1.4.7)”, “doParallel (version 1.0.15)” paketlerinden yararlandı.

Sınıflandırma öncesi uygulanan veri dönüştürme yöntemlerinin etkilerini incelemek amacıyla naif Bayes, C5.0 karar ağacı ve destek vektör makineleri algoritmaları ile çalışıldı. Naif Bayes (NB) algoritması için “naivebayes (version 0.9.6)” R paketi kullanıldı. Sürekli değişkenler üzerindeki NB uygulamalarında, normal dağılımdan üretilen veriler için “Gaussian”, diğer dağılımlar için “Kernel” olasılık yoğunluk fonksiyonları kullanıldı. C5.0 karar ağacı algoritması, “C50 (version 0.1.2)” paketi ile uygulandı. Destek vektör makineleri (DVM) için “caret (version 6.0-84)” ve “e1071 (version 1.7-2)” paketleri kullanıldı. DVM algoritmasında radyal tabanlı kernel fonksiyonu uygulandı. Radyal tabanlı fonksiyonun düzenleme (cost) parametresi için “1” ve “100”, gama (gamma) parametresi için “0,001” ve “0,1” değerleri belirlendi.

Veri dönüştürme yöntemlerinin algoritmalar üzerindeki etkilerini incelemek amacıyla ilk önce herhangi bir ön-işlem uygulamadan, ham veri ile sınıflandırma sonuçları alındı. Daha sonra üretilen verilere, minimum-maksimum normalizasyonu, z skor normalizasyonu ve gözetimsiz diskritizasyon yöntemlerinden eşit genişlikte aralıklandırma (EG) ve eşit frekansta aralıklandırma (EF) yöntemleri uygulanarak sınıflandırma analizi tekrarlandı. Veri dönüşümünde kullanılan yöntemlere ait formüller Tablo 2’de verildi.

Ham verideki değerler, minimum-maksimum normalizasyonu ve z-skor normalizasyonu uygulanarak sırasıyla; “0” ile “1”, ve “-3” ile “3” aralığına getirildi. Bu işlemler için R programında her iki yöntemin formüllerine göre yazılan fonksiyonlar

kullanıldı. EG ve EF diskritizasyon yöntemlerinin formülünde yer alan k değeri, çalışmada her iki yöntem için “10” olarak belirlendi. Veriler bu değere göre bölünerek kategorik hale dönüştürüldü. Gözetimsiz diskritizasyon yöntemleri için “arules (version 1.6-4)” paketi kullanıldı.

Tablo 2. Veri dönüştürme yöntemleri

Minimum-Maksimum normalizasyonu	$\frac{X_{\min} - x}{X_{\max} - X_{\min}}$	X_{\min} ; en küçük değer X_{\max} ; en büyük değer
Z-Skor normalizasyonu	$\frac{x - \bar{X}}{\sigma}$	\bar{X} ; değişken ortalaması σ ; standart sapma
Eşit genişlikte aralıklandırma diskritizasyonu (EG)	$a = \frac{X_{\max} - X_{\min}}{k}, \quad s = X_{\min} + (i \times a)$	k ; aralık sayısı a ; aralık genişliği i ; 1,2,...,k-1 s ; sınırlar
Eşit frekansta aralıklandırma diskritizasyonu (EF)	$f = \frac{n}{k}$	k ; aralık sayısı n ; gözlem sayısı f ; aralık frekansı

Simülasyon çalışmasındaki tüm uygulamalar 1000 kez tekrar edildi ve elde edilen bulgular ortalama ve standart sapma değerleri ile özetlendi.

Çalışma neticesinde, sınıflandırma algoritmalarının performanslarını değerlendirmek için “Genel Doğruluk Oranı (Accuracy)”, “Eğri Altında Kalan Alan (AUC)”, “Duyarlılık (Sensitivity)”, “Seçicilik (Specificity)”, “Pozitif Kestirim Değeri (Positive Predictive Value)”, “Negatif Kestirim Değeri (negative Predictive Value)” ölçüleri kullanıldı. Bağımlı değişkendeki sınıfların dengesiz dağılımlarında “Genel Doğruluk Oranı” yerine “Dengeli Doğruluk Oranı (Balanced Accuracy)” kriteri kullanıldı. Performans değerlendirmek amacıyla kullanılan bu kriterlere ilişkin tanımlama ve hesaplama formülleri Tablo 3’te verildi.

Tablo 3. Performans değerlendirme kriterleri

Doğ.	Gerçekte “Pozitif” ve “Negatif” olanların toplama göre oranı	$(GP+GN)/(GP+ GN+YP+YN)$
D. Doğ.	Duyarlılık ve seçicilik oranlarının ortalaması	$(GP/(GP+YN)+ GN/(GN+YP))/2$
Duy.	“Pozitif” olarak tahmin edilenlerin, gerçekte “Pozitif” olanlara göre oranı	$GP/(GP+YN)$
Seç.	“Negatif” olarak tahmin edilenlerin, gerçekte “Negatif” olanlara göre oranı	$GN/(GN+YP)$

Tablo 3. Devam Performans değerlendirme kriterleri

PKD	Gerçekte “Pozitif” olanların, “Pozitif” olarak tahmin edilenlere göre oranı	$GP/(GP+YP)$
NKD	“Negatif” olarak tahmin edilenlerin, gerçekte “Negatif” olanlara göre oranı	$GN/(GN+YN)$
AUC	Dikey ekseninde duyarlılığın, yatay ekseninde 1-seçiciliğin yer aldığı eğri altında kalan alan	

Doğ.: Genel doğruluk oranı; **D. Doğ.:** Dengeli Doğruluk oranı; **Duy.:** Duyarlılık oranı; **Seç.:** Seçicilik oranı; **PKD:** Pozitif kestirim değeri; **NKD:** Negatif kestirim değeri; **GP:** Gerçek pozitif; **GN:** Gerçek negatif; **YP:** Yanlış pozitif; **YN:** Yanlış negatif.

BULGULAR

NAİF BAYES ALGORİTMASI İLE SINIFLANDIRMA

Normal dağılımdan üretilen verilerin sınıflandırılması

Tablo 4. Normal Dağılım için NB ile elde edilen ortalama doğruluk oranları (%)

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	73,025±18,02	73,025±18,02	73,025±18,02	63,469±16,45	67,194±17,42
	0,2	81,744±12,10	81,744±12,10	81,744±12,10	70,688±12,56	74,338±13,13
	0,3	85,113±9,50	85,113±9,50	85,113±9,50	74,256±11,43	76,718±10,90
	0,4	86,500±7,79	86,500±7,79	86,500±7,79	77,008±9,23	78,554±9,57
	0,5	87,115±7,29	87,115±7,29	87,115±7,29	78,245±9,00	78,660±9,52
500	0,1	79,672±7,88	79,672±7,88	79,672±7,88	72,080±7,87	71,777±7,93
	0,2	84,858±5,21	84,858±5,21	84,858±5,21	79,821±5,49	80,621±5,60
	0,3	87,187±3,84	87,187±3,84	87,187±3,84	83,729±4,16	84,252±4,13
	0,4	88,534±3,22	88,534±3,22	88,534±3,22	85,694±3,55	85,853±3,53
	0,5	88,729±3,18	88,729±3,18	88,729±3,18	86,246±3,50	86,219±3,50
1000	0,1	80,513±5,49	80,513±5,49	80,513±5,49	75,500±5,57	74,242±5,79
	0,2	85,248±3,49	85,248±3,49	85,248±3,49	82,215±3,76	82,063±3,76
	0,3	87,448±2,90	87,448±2,90	87,448±2,90	85,111±3,09	85,213±3,00
	0,4	88,467±2,38	88,467±2,38	88,467±2,38	86,793±2,45	86,760±2,59
	0,5	88,824±2,21	88,824±2,21	88,824±2,21	87,239±2,37	87,120±2,39
10000	0,1	81,232±1,65	81,232±1,65	81,232±1,65	79,328±1,68	77,420±1,74
	0,2	85,564±1,15	85,564±1,15	85,564±1,15	84,150±1,18	83,778±1,18
	0,3	87,716±0,86	87,716±0,86	87,716±0,86	86,559±0,89	86,612±0,87
	0,4	88,759±0,74	88,759±0,74	88,759±0,74	87,735±0,76	87,930±0,76
	0,5	89,079±0,73	89,079±0,73	89,079±0,73	88,099±0,75	88,328±0,74

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Normal dağılımdan üretilen veriler ile elde edilen ortalama doğruluk oranları incelendiğinde, 100, 500, 1000 ve 10000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 4).

Çalışmada, 100 gözlem için elde edilen ortalama doğruluk oranı, sınıf dağılım oranı 0,1 olduğunda; ham veride $73,025 \pm 18,02$ olarak bulunurken eşit genişlikte aralıklandırma (EG) diskritizasyonu ile $63,469 \pm 16,45$, eşit frekansta aralıklandırma (EF) diskritizasyonu ile $67,194 \pm 17,42$ oranlarına geriledi. Sınıf dağılım oranı 0,2 olduğunda ortalama doğruluk oranlarının; ham veride $81,744 \pm 12,1$, EG diskritizasyonu ile $70,688 \pm 12,56$, EF diskritizasyonu ile $74,338 \pm 13,13$ olduğu görüldü. Bu oranlar sırasıyla; 0,3 sınıf dağılım oranı için $85,113 \pm 9,50$, $74,256 \pm 11,43$, $76,718 \pm 10,90$, 0,4 dağılım oranı için sırasıyla; $86,500 \pm 7,79$, $77,008 \pm 9,23$, $78,554 \pm 9,57$ olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda ortalama doğruluk oranlarının; ham veride $87,115 \pm 7,29$, EG diskritizasyonu ile $78,245 \pm 9,00$, EF diskritizasyonu ile $78,66 \pm 9,52$ olduğu görüldü.

Gözlem sayısı 500 olduğunda gerçekleştirilen sınıflandırma çalışmasında 0,1 sınıf dağılım oranı için elde edilen ortalama doğruluk oranları; ham veride $79,672 \pm 7,88$, EG ile $72,080 \pm 7,87$, EF ile $71,777 \pm 7,93$ olarak bulundu. Sınıf dağılımlarında denge oranı arttıkça doğru sınıflandırma oranlarında azalma gerçekleştiği görüldü. Sınıf dağılım oranı 0,5 olduğunda ham veri, EG ve EF diskritizasyonları ile oranlar sırasıyla; $88,729 \pm 3,18$, $86,246 \pm 3,5$, $86,219 \pm 3,5$ olarak belirlendi.

Gözlem sayısı 1000 olduğunda, sınıf dağılım oranı 0,1 iken elde edilen ortalama doğruluk oranlarının; ham veride $80,513 \pm 5,49$, EG diskritizasyonu ile $75,500 \pm 5,57$, EF diskritizasyonu ile $74,242 \pm 5,79$ olduğu görüldü. Sınıf dağılım oranı 0,5 olduğunda bu oranlar; ham veri ile $88,824 \pm 2,21$, EG diskritizasyonu ile $87,239 \pm 2,37$, EF diskritizasyonu ile $87,12 \pm 2,39$ olarak elde edildi.

Gözlem sayısı 10000 ve sınıf dağılım oranı 0,1 olduğunda elde edilen ortalama doğruluk oranları; ham veri ile $81,232 \pm 1,65$, EG diskritizasyonu ile $79,328 \pm 1,68$, EF diskritizasyonu ile $77,420 \pm 1,74$ iken bu oranlar 0,5 sınıf dağılımında sırasıyla; $89,079 \pm 0,73$, $88,099 \pm 0,75$, $88,328 \pm 0,74$ olarak bulundu.

Çalışmada, gözlem sayısı ve sınıfların denge dağılım oranı arttıkça doğruluk oranları artmaktadır. 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama doğruluk oranlarına Tablo 4’te yer verilmektedir.

Tablo 5. Normal Dağılım için NB ile elde edilen ortalama AUC değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,899±0,13	0,899±0,13	0,899±0,13	0,816±0,17	0,811±0,18
	0,2	0,927±0,08	0,927±0,08	0,927±0,08	0,840±0,12	0,847±0,12
	0,3	0,940±0,06	0,940±0,06	0,940±0,06	0,854±0,10	0,859±0,10
	0,4	0,944±0,05	0,944±0,05	0,944±0,05	0,864±0,08	0,870±0,08
	0,5	0,947±0,05	0,947±0,05	0,947±0,05	0,865±0,08	0,870±0,08
500	0,1	0,953±0,03	0,953±0,03	0,953±0,03	0,914±0,05	0,903±0,05
	0,2	0,956±0,03	0,956±0,03	0,956±0,03	0,931±0,03	0,925±0,03
	0,3	0,957±0,02	0,957±0,02	0,957±0,02	0,935±0,03	0,933±0,03
	0,4	0,958±0,02	0,958±0,02	0,958±0,02	0,939±0,02	0,937±0,02
	0,5	0,958±0,02	0,958±0,02	0,958±0,02	0,939±0,02	0,939±0,02
1000	0,1	0,956±0,02	0,956±0,02	0,956±0,02	0,935±0,03	0,924±0,03
	0,2	0,958±0,02	0,958±0,02	0,958±0,02	0,944±0,02	0,938±0,02
	0,3	0,957±0,02	0,957±0,02	0,957±0,02	0,944±0,02	0,941±0,02
	0,4	0,958±0,01	0,958±0,01	0,958±0,01	0,946±0,02	0,945±0,02
	0,5	0,958±0,01	0,958±0,01	0,958±0,01	0,947±0,01	0,946±0,01
10000	0,1	0,960±0,01	0,960±0,01	0,960±0,01	0,952±0,01	0,947±0,01
	0,2	0,959±0,01	0,959±0,01	0,959±0,01	0,953±0,01	0,950±0,01
	0,3	0,959±0,004	0,959±0,004	0,959±0,004	0,953±0,005	0,952±0,005
	0,4	0,959±0,004	0,959±0,004	0,959±0,004	0,953±0,004	0,953±0,004
	0,5	0,960±0,004	0,960±0,004	0,960±0,004	0,953±0,004	0,954±0,004

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Normal dağılımdan üretilen veriler ile elde edilen ortalama AUC değerleri incelendiğinde, 100, 500, 1000 ve 10000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 5).

Gözlem sayısı 100 ve sınıf dağılım oranı 0,1 olduğunda ortalama AUC değerleri; ham veride 0,899±0,13, EG diskritizasyonu ile 0,816±0,17, EF diskritizasyonu ile 0,811±0,18 olarak bulundu. Sınıf dağılım oranı 0,2 iken bu değerler sırasıyla; 0,927±0,08, 0,84±0,12, 0,847±0,12 olup sınıf dağılım oranı 0,3 iken 0,94±0,06, 0,854±0,1, 0,859±0,1 olarak bulundu. Sınıf dağılım oranı 0,4 iken bu değerler sırasıyla; 0,944±0,05, 0,864±0,08, 0,87±0,08 olarak belirlenirken sınıf dağılım oranı 0,5 olduğunda ortalama AUC değerlerinin sırasıyla; 0,947±0,05, 0,865±0,08, 0,87±0,08 olduğu bulgusuna ulaşıldı.

Gözlem sayısı 500 iken gerçekleştirilen sınıflandırma çalışmasında sınıf dağılım oranı 0,1 için ortalama AUC değerleri; ham veride $0,953\pm 0,03$, EG diskritizasyonu ile $0,914\pm 0,05$, EF diskritizasyonu ile $0,903\pm 0,05$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise sırasıyla; $0,958\pm 0,02$, $0,939\pm 0,02$, $0,939\pm 0,02$ değerlerine ulaşıldı. Gözlem sayısı 1000 olduğunda, sınıf dağılım oranı 0,1 iken; ham veride $0,956\pm 0,02$, EG diskritizasyonu ile $0,935\pm 0,03$, EF diskritizasyonu ile $0,924\pm 0,03$ olarak elde edilen ortalama AUC değerleri sınıf dağılım oranı 0,5 olduğunda sırasıyla; $0,958\pm 0,01$, $0,947\pm 0,01$, $0,946\pm 0,01$ olarak belirlendi. Son olarak gözlem sayısı 10000 ve sınıf dağılım oranı 0,1 iken ortalama AUC değerleri; ham veride $0,96\pm 0,01$, EG diskritizasyonu ile $0,952\pm 0,01$, EF diskritizasyonu ile $0,947\pm 0,01$ olarak bulunurken sınıf dağılım oranı 0,5 olduğunda bu değerler sırasıyla; $0,96\pm 0,004$, $0,953\pm 0,004$, $0,954\pm 0,004$ olarak saptandı.

Çalışmada, gözlem sayısı ve sınıfların denge dağılım oranı arttıkça AUC değerleri artmaktadır. 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama AUC değerlerine Tablo 5’te yer verilmektedir.

Tablo 6. Normal Dağılım için NB ile elde edilen ortalama duyarlılık oranları

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	$0,475\pm 0,36$	$0,475\pm 0,36$	$0,475\pm 0,36$	$0,295\pm 0,33$	$0,401\pm 0,34$
	0,2	$0,670\pm 0,24$	$0,670\pm 0,24$	$0,670\pm 0,24$	$0,465\pm 0,25$	$0,584\pm 0,25$
	0,3	$0,762\pm 0,18$	$0,762\pm 0,18$	$0,762\pm 0,18$	$0,570\pm 0,22$	$0,666\pm 0,20$
	0,4	$0,823\pm 0,14$	$0,823\pm 0,14$	$0,823\pm 0,14$	$0,679\pm 0,16$	$0,741\pm 0,16$
	0,5	$0,866\pm 0,11$	$0,866\pm 0,11$	$0,866\pm 0,11$	$0,754\pm 0,14$	$0,780\pm 0,14$
500	0,1	$0,606\pm 0,16$	$0,606\pm 0,16$	$0,606\pm 0,16$	$0,454\pm 0,16$	$0,457\pm 0,16$
	0,2	$0,725\pm 0,10$	$0,725\pm 0,10$	$0,725\pm 0,10$	$0,627\pm 0,11$	$0,655\pm 0,11$
	0,3	$0,793\pm 0,07$	$0,793\pm 0,07$	$0,793\pm 0,07$	$0,730\pm 0,08$	$0,753\pm 0,08$
	0,4	$0,843\pm 0,06$	$0,843\pm 0,06$	$0,843\pm 0,06$	$0,797\pm 0,06$	$0,812\pm 0,06$
	0,5	$0,879\pm 0,05$	$0,879\pm 0,05$	$0,879\pm 0,05$	$0,851\pm 0,05$	$0,858\pm 0,05$
1000	0,1	$0,622\pm 0,11$	$0,622\pm 0,11$	$0,622\pm 0,11$	$0,522\pm 0,11$	$0,503\pm 0,12$
	0,2	$0,732\pm 0,07$	$0,732\pm 0,07$	$0,732\pm 0,07$	$0,673\pm 0,07$	$0,679\pm 0,07$
	0,3	$0,796\pm 0,05$	$0,796\pm 0,05$	$0,796\pm 0,05$	$0,754\pm 0,06$	$0,765\pm 0,06$
	0,4	$0,841\pm 0,04$	$0,841\pm 0,04$	$0,841\pm 0,04$	$0,816\pm 0,04$	$0,823\pm 0,04$
	0,5	$0,879\pm 0,03$	$0,879\pm 0,03$	$0,879\pm 0,03$	$0,861\pm 0,04$	$0,865\pm 0,04$
10000	0,1	$0,636\pm 0,03$	$0,636\pm 0,03$	$0,636\pm 0,03$	$0,598\pm 0,03$	$0,563\pm 0,03$
	0,2	$0,738\pm 0,02$	$0,738\pm 0,02$	$0,738\pm 0,02$	$0,712\pm 0,02$	$0,709\pm 0,02$
	0,3	$0,801\pm 0,02$	$0,801\pm 0,02$	$0,801\pm 0,02$	$0,781\pm 0,02$	$0,786\pm 0,02$
	0,4	$0,845\pm 0,01$	$0,845\pm 0,01$	$0,845\pm 0,01$	$0,831\pm 0,01$	$0,837\pm 0,01$
	0,5	$0,881\pm 0,01$	$0,881\pm 0,01$	$0,881\pm 0,01$	$0,871\pm 0,01$	$0,875\pm 0,01$

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Normal dağılımdan üretilen veriler ile elde edilen ortalama duyarlılık oranları incelendiğinde, 100, 500, 1000 ve 10000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 6).

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama duyarlılık oranları sınıf dağılım oranı 0,1 olduğunda; ham veride $0,475\pm0,36$, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile $0,295\pm0,33$, eşit frekansta aralıklandırma (EF) diskritizasyonu ile $0,401\pm0,34$ olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama duyarlılık oranlarının; ham veride $0,67\pm0,24$, EG diskritizasyonu ile $0,465\pm0,25$, EF diskritizasyonu ile $0,584\pm0,25$ olduğu görüldü. Bu oranlar sırasıyla; sınıf dağılım oranı 0,3 için $0,762\pm0,18$, $0,57\pm0,22$, $0,666\pm0,2$, sınıf dağılım oranı 0,4 için $0,823\pm0,14$, $0,679\pm0,16$, $0,741\pm0,16$ olarak hesaplandı. Son olarak sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama duyarlılık oranları; ham veride $0,866\pm0,11$, EG diskritizasyonu ile $0,754\pm0,14$, EF diskritizasyonu ile $0,78\pm0,14$ olarak saptandı.

Gözlem sayısı 500 iken gerçekleştirilen sınıflandırma çalışmasında sınıf dağılım oranı 0,1 için elde edilen ortalama duyarlılık oranları; ham veride $0,606\pm0,16$, EG diskritizasyonu ile $0,454\pm0,16$, EF diskritizasyonu ile $0,457\pm0,16$ olarak bulundu. Sınıf dağılımında denge oranı arttıkça duyarlılığın arttığı görülen çalışmada sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama duyarlılık oranları; ham veride $0,879\pm0,05$, EG diskritizasyonu ile $0,851\pm0,05$, EF diskritizasyonu ile $0,858\pm0,05$ olarak elde edildi.

Gözlem sayısı 1000 ve sınıf dağılım oranı 0,1 iken elde edilen ortalama duyarlılık oranları; ham veride $0,622\pm0,11$, EG diskritizasyonu ile $0,522\pm0,11$, EF diskritizasyonu ile $0,503\pm0,12$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda elde edilen bu oranlar; ham veride $0,879\pm0,03$, EG diskritizasyonu ile $0,861\pm0,04$, EF diskritizasyonu ile $0,865\pm0,04$ olarak belirlendi.

Gözlem sayısı 10000 olduğunda ise yapılan sınıflandırma çalışmasında sınıf dağılım oranı 0,1 iken elde edilen ortalama duyarlılık oranları; ham veride $0,636\pm0,03$, EG diskritizasyonu ile $0,598\pm0,03$, EF diskritizasyonu ile $0,563\pm0,03$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ortalama duyarlılık oranları; ham veride $0,881\pm0,01$, EG diskritizasyonu ile $0,871\pm0,01$, EF diskritizasyonu ile $0,875\pm0,01$ olarak bulundu.

Çalışmada, gözlem sayısı ve sınıfların denge dağılım oranı arttıkça duyarlılık oranlarında artış gözlenmektedir. 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama duyarlılık oranlarına Tablo 6’da yer verilmektedir.

Tablo 7. Normal Dağılım için NB ile elde edilen ortalama seçicilik oranları

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,986±0,03	0,986±0,03	0,986±0,03	0,974±0,04	0,943±0,06
	0,2	0,965±0,05	0,965±0,05	0,965±0,05	0,949±0,05	0,903±0,08
	0,3	0,940±0,07	0,940±0,07	0,940±0,07	0,915±0,08	0,869±0,09
	0,4	0,907±0,09	0,907±0,09	0,907±0,09	0,861±0,10	0,830±0,11
	0,5	0,876±0,10	0,876±0,10	0,876±0,10	0,811±0,12	0,793±0,14
500	0,1	0,988±0,01	0,988±0,01	0,988±0,01	0,987±0,01	0,979±0,02
	0,2	0,972±0,02	0,972±0,02	0,972±0,02	0,970±0,02	0,958±0,02
	0,3	0,951±0,03	0,951±0,03	0,951±0,03	0,945±0,03	0,932±0,03
	0,4	0,928±0,03	0,928±0,03	0,928±0,03	0,917±0,04	0,905±0,04
	0,5	0,895±0,05	0,895±0,05	0,895±0,05	0,874±0,05	0,866±0,05
1000	0,1	0,988±0,01	0,988±0,01	0,988±0,01	0,988±0,01	0,982±0,01
	0,2	0,973±0,01	0,973±0,01	0,973±0,01	0,971±0,01	0,962±0,02
	0,3	0,953±0,02	0,953±0,02	0,953±0,02	0,948±0,02	0,939±0,02
	0,4	0,928±0,02	0,928±0,02	0,928±0,02	0,920±0,03	0,913±0,03
	0,5	0,897±0,03	0,897±0,03	0,897±0,03	0,884±0,03	0,877±0,03
10000	0,1	0,989±0,002	0,989±0,002	0,989±0,002	0,988±0,003	0,985±0,003
	0,2	0,973±0,004	0,973±0,004	0,973±0,004	0,971±0,004	0,967±0,005
	0,3	0,954±0,01	0,954±0,01	0,954±0,01	0,950±0,01	0,946±0,01
	0,4	0,930±0,01	0,930±0,01	0,930±0,01	0,924±0,01	0,922±0,01
	0,5	0,900±0,01	0,900±0,01	0,900±0,01	0,891±0,01	0,892±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Normal dağılımdan üretilen veriler ile elde edilen ortalama seçicilik oranları incelendiğinde, 100, 500, 1000 ve 10000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 7).

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama seçicilik oranları sınıf dağılım oranı 0,1 olduğunda; ham veride 0,986±0,03, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,974±0,04, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,943±0,06 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama seçicilik oranlarının; ham veride 0,965±0,05, EG diskritizasyonu ile 0,949±0,05, EF diskritizasyonu ile 0,903±0,08 olduğu görüldü. Bu oranlar sırasıyla; 0,3 sınıf dağılım oranı için 0,94±0,07, 0,915±0,08, 0,869±0,09, 0,4 sınıf dağılım oranı için 0,907±0,09, 0,861±0,1, 0,83±0,11 olarak bulundu.

Son olarak sınıf dağılım oranı 0,5 olduğunda elde edilen bu oranların; ham veride, $0,876\pm 0,1$, EG diskritizasyonu ile $0,811\pm 0,12$, EF diskritizasyonu ile $0,793\pm 0,14$ olduğu görüldü.

Gözlem sayısı 500 olduğunda, sınıf dağılım oranı 0,1 için elde edilen ortalama seçicilik oranları; ham veride $0,988\pm 0,01$, EG diskritizasyonu ile $0,987\pm 0,01$, EF diskritizasyonu ile $0,979\pm 0,02$ olarak bulunurken sınıf dağılım oranı 0,5 olduğunda bu oranlar sırasıyla; $0,895\pm 0,05$, $0,874\pm 0,05$, $0,866\pm 0,05$ olarak saptandı. Gözlem sayısı 1000 ve sınıf dağılım oranı 0,1 iken ortalama seçicilik oranları; ham veri ile $0,988\pm 0,01$, EG ile $0,988\pm 0,01$, EF ile $0,982\pm 0,01$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise bu oranların sırasıyla; $0,897\pm 0,03$, $0,884\pm 0,03$, $0,877\pm 0,03$ olduğu görüldü. Gözlem sayısı 10000 olduğunda, sınıf dağılım oranı 0,1 iken elde edilen ortalama seçicilik oranları; ham veri ile $0,989\pm 0,002$, EG ile $0,988\pm 0,003$, EF ile $0,985\pm 0,003$, sınıf dağılım oranı 0,5 olduğunda; ham veri ile $0,9\pm 0,01$, EG ile $0,891\pm 0,01$, EF ile $0,892\pm 0,01$ olarak belirlendi.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama seçicilik oranlarına Tablo 7’de yer verilmektedir.

Tablo 8. Normal Dağılım için NB ile elde edilen ortalama pozitif kestirim değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	$0,828\pm 0,30$	$0,828\pm 0,30$	$0,828\pm 0,30$	$0,590\pm 0,42$	$0,477\pm 0,37$
	0,2	$0,853\pm 0,19$	$0,853\pm 0,19$	$0,853\pm 0,19$	$0,728\pm 0,28$	$0,635\pm 0,24$
	0,3	$0,863\pm 0,14$	$0,863\pm 0,14$	$0,863\pm 0,14$	$0,761\pm 0,20$	$0,706\pm 0,17$
	0,4	$0,868\pm 0,11$	$0,868\pm 0,11$	$0,868\pm 0,11$	$0,782\pm 0,14$	$0,760\pm 0,13$
	0,5	$0,883\pm 0,09$	$0,883\pm 0,09$	$0,883\pm 0,09$	$0,809\pm 0,11$	$0,802\pm 0,11$
500	0,1	$0,858\pm 0,13$	$0,858\pm 0,13$	$0,858\pm 0,13$	$0,816\pm 0,16$	$0,725\pm 0,18$
	0,2	$0,872\pm 0,08$	$0,872\pm 0,08$	$0,872\pm 0,08$	$0,844\pm 0,09$	$0,800\pm 0,09$
	0,3	$0,876\pm 0,06$	$0,876\pm 0,06$	$0,876\pm 0,06$	$0,854\pm 0,07$	$0,830\pm 0,07$
	0,4	$0,888\pm 0,05$	$0,888\pm 0,05$	$0,888\pm 0,05$	$0,867\pm 0,05$	$0,853\pm 0,05$
	0,5	$0,895\pm 0,04$	$0,895\pm 0,04$	$0,895\pm 0,04$	$0,873\pm 0,04$	$0,867\pm 0,04$
1000	0,1	$0,860\pm 0,09$	$0,860\pm 0,09$	$0,860\pm 0,09$	$0,836\pm 0,10$	$0,766\pm 0,11$
	0,2	$0,874\pm 0,06$	$0,874\pm 0,06$	$0,874\pm 0,06$	$0,857\pm 0,06$	$0,822\pm 0,06$
	0,3	$0,880\pm 0,04$	$0,880\pm 0,04$	$0,880\pm 0,04$	$0,864\pm 0,05$	$0,846\pm 0,05$
	0,4	$0,887\pm 0,03$	$0,887\pm 0,03$	$0,887\pm 0,03$	$0,873\pm 0,04$	$0,864\pm 0,04$
	0,5	$0,896\pm 0,03$	$0,896\pm 0,03$	$0,896\pm 0,03$	$0,882\pm 0,03$	$0,877\pm 0,03$
10000	0,1	$0,865\pm 0,03$	$0,865\pm 0,03$	$0,865\pm 0,03$	$0,853\pm 0,03$	$0,811\pm 0,03$
	0,2	$0,873\pm 0,02$	$0,873\pm 0,02$	$0,873\pm 0,02$	$0,862\pm 0,02$	$0,842\pm 0,02$
	0,3	$0,881\pm 0,01$	$0,881\pm 0,01$	$0,881\pm 0,01$	$0,871\pm 0,01$	$0,863\pm 0,01$
	0,4	$0,889\pm 0,01$	$0,889\pm 0,01$	$0,889\pm 0,01$	$0,879\pm 0,01$	$0,877\pm 0,01$
	0,5	$0,898\pm 0,01$	$0,898\pm 0,01$	$0,898\pm 0,01$	$0,889\pm 0,01$	$0,890\pm 0,01$

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Normal dağılımdan üretilen veriler ile elde edilen ortalama pozitif kestirim değerleri incelendiğinde, 100, 500, 1000 ve 10000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 8).

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama pozitif kestirim değerleri (PKD) sınıf dağılım oranı 0,1 olduğunda; ham veride $0,828 \pm 0,3$, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile $0,59 \pm 0,42$, eşit frekansta aralıklandırma (EF) diskritizasyonu ile $0,477 \pm 0,37$ olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama PKD'nin; ham veride $0,853 \pm 0,19$, EG diskritizasyonu ile $0,728 \pm 0,28$, EF diskritizasyonu ile $0,635 \pm 0,24$ olduğu görüldü. Sınıf dağılım oranı 0,3 iken bu değerler; ham veri ile $0,863 \pm 0,14$, EG diskritizasyonu ile $0,761 \pm 0,2$, EF diskritizasyonu ile $0,706 \pm 0,17$ olarak hesaplandı. Sınıf dağılım oranı 0,4 iken elde edilen ortalama değerlerin; ham veride $0,868 \pm 0,11$, EG diskritizasyonu ile $0,782 \pm 0,14$, EF diskritizasyonu ile $0,76 \pm 0,13$ olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama PKD'nin; ham veride $0,883 \pm 0,09$, EG diskritizasyonu ile $0,809 \pm 0,11$, EF diskritizasyonu ile $0,802 \pm 0,11$ olduğu sonucuna ulaşıldı.

Gözlem sayısı 500 iken gerçekleştirilen sınıflandırma çalışmasında sınıf dağılım oranı 0,1 için elde edilen ortalama PKD; ham veride $0,858 \pm 0,13$, EG diskritizasyonu ile $0,816 \pm 0,16$, EF diskritizasyonu ile $0,725 \pm 0,18$ olarak bulundu. Sınıf dağılım oranı 0,2, 0,3 ve 0,4 olduğunda sırasıyla; ham veride $0,872 \pm 0,08$, $0,876 \pm 0,06$, $0,888 \pm 0,05$ olarak bulunurken EG diskritizasyonu ile $0,844 \pm 0,09$, $0,854 \pm 0,07$, $0,867 \pm 0,05$ ve EF diskritizasyonu ile $0,8 \pm 0,09$, $0,83 \pm 0,07$, $0,853 \pm 0,05$ ortalama değerlerine ulaşıldı. Son olarak sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama PKD; ham veride $0,895 \pm 0,04$, EG diskritizasyonu ile $0,873 \pm 0,04$, EF diskritizasyonu ile $0,867 \pm 0,04$ olarak bulundu.

Gözlem sayısı 1000 olduğunda gerçekleştirilen sınıflandırma çalışmasında sınıfların 0,1 dağılım oranı için elde edilen ortalama PKD; ham veride $0,86 \pm 0,09$, EG diskritizasyonu ile $0,836 \pm 0,1$, EF diskritizasyonu ile $0,766 \pm 0,11$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise elde edilen ortalama PKD; ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında $0,896 \pm 0,03$, EG diskritizasyonu ile $0,882 \pm 0,03$, EF diskritizasyonu ile $0,877 \pm 0,03$ olarak saptandı. Gözlem sayısı 10000 olduğunda yapılan sınıflandırma çalışmasında sınıf dağılım oranı 0,1 iken elde edilen ortalama PKD; ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında $0,865 \pm 0,03$, EG diskritizasyonu ile $0,853 \pm 0,03$, EF diskritizasyonu ile $0,811 \pm 0,03$ olarak bulundu. Sınıfların

0,5 dağılım oranı için ortalama PKD; ham veri ile minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında $0,898\pm 0,01$, EG diskritizasyonu ile $0,889\pm 0,01$, EF diskritizasyonu ile $0,89\pm 0,01$ olarak elde edildi.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama pozitif kestirim değerlerine Tablo 8’de yer verilmektedir.

Tablo 9. Normal Dağılım için NB ile elde edilen ortalama negatif kestirim değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,946±0,04	0,946±0,04	0,946±0,04	0,927±0,03	0,935±0,04
	0,2	0,924±0,05	0,924±0,05	0,924±0,05	0,879±0,05	0,899±0,06
	0,3	0,907±0,07	0,907±0,07	0,907±0,07	0,838±0,07	0,863±0,07
	0,4	0,891±0,08	0,891±0,08	0,891±0,08	0,808±0,08	0,835±0,09
	0,5	0,877±0,09	0,877±0,09	0,877±0,09	0,777±0,10	0,794±0,11
500	0,1	0,958±0,02	0,958±0,02	0,958±0,02	0,942±0,02	0,942±0,02
	0,2	0,935±0,02	0,935±0,02	0,935±0,02	0,913±0,02	0,918±0,02
	0,3	0,916±0,03	0,916±0,03	0,916±0,03	0,892±0,03	0,899±0,03
	0,4	0,900±0,03	0,900±0,03	0,900±0,03	0,873±0,03	0,880±0,03
	0,5	0,883±0,04	0,883±0,04	0,883±0,04	0,856±0,04	0,861±0,04
1000	0,1	0,959±0,01	0,959±0,01	0,959±0,01	0,949±0,01	0,947±0,01
	0,2	0,936±0,02	0,936±0,02	0,936±0,02	0,923±0,02	0,923±0,02
	0,3	0,917±0,02	0,917±0,02	0,917±0,02	0,900±0,02	0,904±0,02
	0,4	0,898±0,02	0,898±0,02	0,898±0,02	0,883±0,03	0,886±0,03
	0,5	0,882±0,03	0,882±0,03	0,882±0,03	0,865±0,03	0,868±0,03
10000	0,1	0,961±0,003	0,961±0,003	0,961±0,003	0,957±0,003	0,953±0,004
	0,2	0,937±0,01	0,937±0,01	0,937±0,01	0,931±0,01	0,930±0,01
	0,3	0,918±0,01	0,918±0,01	0,918±0,01	0,910±0,01	0,912±0,01
	0,4	0,900±0,01	0,900±0,01	0,900±0,01	0,891±0,01	0,895±0,01
	0,5	0,884±0,01	0,884±0,01	0,884±0,01	0,874±0,01	0,877±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Normal dağılımdan üretilen veriler ile elde edilen ortalama negatif kestirim değerleri incelendiğinde, 100, 500, 1000 ve 10000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 9).

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama negatif kestirim değerleri (NKD) sınıf dağılım oranı 0,1 olduğunda; ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında $0,946\pm 0,04$, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile $0,927\pm 0,03$, eşit frekansta aralıklandırma (EF) diskritizasyonu ile $0,935\pm 0,04$ olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda bu ortalama

değerlerin; ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında $0,924\pm 0,05$, EG diskritizasyonu ile $0,879\pm 0,05$, EF diskritizasyonu ile $0,899\pm 0,06$ olduğu görüldü. Sınıf dağılım oranı 0,3 iken ortalama NKD; ham veri ile minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında $0,907\pm 0,07$, EG diskritizasyonu ile $0,838\pm 0,07$, EF diskritizasyonu ile $0,863\pm 0,07$ olarak hesaplandı. Sınıf dağılım oranı 0,4 iken elde edilen ortalama NKD'nin; ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında $0,891\pm 0,08$, EG diskritizasyonu ile $0,808\pm 0,08$, EF diskritizasyonu ile $0,835\pm 0,09$ olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama NKD'nin; ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında $0,877\pm 0,09$, EG diskritizasyonu ile $0,777\pm 0,1$, EF diskritizasyonu ile $0,794\pm 0,11$ olduğu görüldü.

Gerçekleştirilen 500 gözlem sayılı sınıflandırma çalışmasında sınıf dağılım oranı 0,1 için elde edilen ortalama negatif kestirim değerleri; ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında $0,958\pm 0,02$, EG diskritizasyonu ile $0,942\pm 0,02$, EF diskritizasyonu ile $0,942\pm 0,02$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama NKD; ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında $0,883\pm 0,04$, EG diskritizasyonu ile $0,856\pm 0,04$, EF diskritizasyonu ile $0,861\pm 0,04$ olarak belirlendi.

Gözlem sayısının 1000 için gerçekleştirilen sınıflandırmada sınıf dağılım oranı 0,1 olduğunda elde edilen ortalama NKD; ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında $0,959\pm 0,01$, EG diskritizasyonu ile $0,949\pm 0,01$, EF diskritizasyonu ile $0,947\pm 0,01$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise bu değerler sırasıyla; $0,882\pm 0,03$, $0,865\pm 0,03$, $0,868\pm 0,03$ olarak elde edildi.

Gözlem sayısı 10000 olduğunda yapılan sınıflandırma çalışmasında sınıf dağılım oranı 0,1 iken elde edilen ortalama NKD; ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında $0,961\pm 0,003$, EG diskritizasyonu ile $0,957\pm 0,003$, EF diskritizasyonu ile $0,953\pm 0,004$ olarak bulundu. Sınıfların 0,5 tam dengeli dağılımında elde edilen ortalama NKD; ham veri ile minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında $0,884\pm 0,01$, EG diskritizasyonu ile $0,874\pm 0,01$, EF diskritizasyonu ile $0,877\pm 0,01$ olarak saptandı.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama negatif kestirim değerlerine Tablo 9’da yer verilmektedir.

Ki-kare dağılımından üretilen verilerin sınıflandırılması

Tablo 10. Ki-kare dağılımı için NB ile elde edilen ortalama doğruluk oranları (%)

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	61,144±15,45	61,144±15,45	61,144±15,45	56,758±13,45	61,603±16,01
	0,2	70,650±13,04	70,650±13,04	70,650±13,04	65,138±12,74	68,647±13,35
	0,3	75,605±10,82	75,608±10,82	75,605±10,82	70,257±11,19	72,080±11,28
	0,4	77,260±9,84	77,256±9,84	77,256±9,84	72,883±10,39	73,485±9,92
	0,5	77,990±8,97	77,990±8,97	77,990±8,97	72,865±10,07	72,810±9,86
500	0,1	67,572±7,69	67,572±7,69	67,573±7,69	62,794±7,09	65,639±7,78
	0,2	75,604±5,56	75,607±5,56	75,604±5,56	72,752±5,89	74,664±5,96
	0,3	80,080±4,45	80,081±4,45	80,081±4,45	78,010±4,70	78,704±4,69
	0,4	82,037±3,86	82,037±3,86	82,037±3,86	80,297±3,95	80,799±3,97
	0,5	82,718±3,64	82,720±3,64	82,719±3,64	81,171±3,92	81,564±3,70
1000	0,1	68,736±5,45	68,736±5,45	68,736±5,45	65,670±5,29	67,894±5,48
	0,2	76,855±3,86	76,855±3,86	76,855±3,86	74,758±4,02	76,285±3,94
	0,3	80,695±3,30	80,695±3,30	80,695±3,30	79,325±3,38	80,020±3,38
	0,4	82,773±2,82	82,771±2,82	82,771±2,82	81,670±2,85	82,175±2,85
	0,5	83,170±2,71	83,170±2,71	83,170±2,71	82,065±2,77	82,698±2,80
10000	0,1	71,448±1,78	71,448±1,78	71,448±1,78	69,093±1,76	71,295±1,73
	0,2	78,444±1,32	78,444±1,32	78,444±1,32	76,683±1,31	78,269±1,29
	0,3	81,921±1,03	81,921±1,03	81,921±1,03	80,483±1,03	81,671±1,05
	0,4	83,755±0,85	83,755±0,85	83,755±0,85	82,406±0,88	83,459±0,86
	0,5	84,291±0,82	84,291±0,82	84,291±0,82	82,958±0,86	84,020±0,84

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama doğruluk oranları sınıf dağılım oranı 0,1 olduğunda; ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında %61,144±15,45, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile %56,758±13,45, eşit frekansta aralıklandırma (EF) diskritizasyonu ile %61,603±16,01 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama doğruluk oranlarının sırasıyla; %70,650±13,04, %65,138±12,74, %68,647±13,35 oranlarında olduğu görüldü. Sınıf dağılım oranı 0,3 iken bu oranlar; ham veri ve z-skor normalizasyonu ile %75,605±10,82, minimum-maksimum normalizasyonu ile %75,608±10,82, EG ile %70,257±11,19, EF ile %72,080±11,28 olarak hesaplandı. Sınıf dağılım oranı 0,4 iken bu oranlar; ham veri ile %77,260±9,84 olarak elde edilirken minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında %77,256±9,84, EG ile %72,883±10,39, EF ile %73,485±9,92 olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda elde

edilen ortalama doğruluk oranlarının; ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında %77,99±8,97, EG ile %72,865±10,07, EF ile %72,81±9,86 olduğu sonucuna ulaşıldı.

Gözlem sayısı 500 olduğunda yapılan sınıflandırma çalışmasında, 0,1 sınıf dağılım oranı için elde edilen ortalama doğruluk oranları; ham veri ve minimum-maksimum normalizasyonu ile %67,572±7,69, z-skor normalizasyonu sonrasında %67,573±7,69, EG ile %62,794±7,09, EF ile %65,639±7,78 olarak bulundu. Bu değerlerin, sınıf dağılım oranı 0,5 olduğunda sırasıyla; %82,718±3,64, %82,720±3,64, %82,719±3,64, %81,171±3,92, %81,564±3,7 olduğu görüldü.

Gözlem sayısı 1000 ve sınıf dağılım oranı 0,1 iken elde edilen ortalama doğruluk oranları için; ham veri, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında %68,736±5,45, EG ile %65,67±5,29, EF ile %67,894±5,48 bulgularına ulaşılırken sınıf dağılım oranı 0,5 olduğunda ortalama doğruluk oranları sırasıyla; %83,17±2,71, %82,065±2,77, %82,698±2,8 olarak belirlendi.

Ortalama doğruluk oranları, 10000 gözlem sayısı için yapılan çalışmada, sınıf dağılım oranı 0,1 iken ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında %71,448±1,78, EG diskritizasyonu ile %69,093±1,76, EF diskritizasyonu ile %71,295±1,73 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise bu oranlar için sırasıyla; %84,291±0,82, %82,958±0,86, %84,02±0,84 bulgularına ulaşıldı.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama doğruluk oranlarına Tablo 10'da yer verilmektedir.

Tablo 11. Ki-kare dağılımı için NB ile elde edilen ortalama AUC değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,791±0,18	0,791±0,18	0,790±0,18	0,709±0,20	0,749±0,19
	0,2	0,836±0,12	0,836±0,12	0,836±0,12	0,771±0,14	0,785±0,13
	0,3	0,854±0,10	0,854±0,10	0,854±0,10	0,799±0,11	0,809±0,11
	0,4	0,855±0,09	0,855±0,09	0,855±0,09	0,807±0,10	0,810±0,10
	0,5	0,860±0,08	0,860±0,08	0,860±0,08	0,805±0,10	0,806±0,10
500	0,1	0,881±0,06	0,881±0,06	0,881±0,06	0,846±0,06	0,850±0,06
	0,2	0,896±0,04	0,896±0,04	0,896±0,04	0,877±0,04	0,882±0,04
	0,3	0,900±0,03	0,900±0,03	0,900±0,03	0,886±0,04	0,890±0,04
	0,4	0,902±0,03	0,902±0,03	0,902±0,03	0,889±0,03	0,894±0,03
	0,5	0,906±0,03	0,906±0,03	0,906±0,03	0,893±0,03	0,898±0,03

Tablo 11. Devam Ki-kare dağılımı için NB ile elde edilen ortalama AUC değerleri

1000	0,1	0,898±0,04	0,898±0,04	0,898±0,04	0,880±0,04	0,885±0,04
	0,2	0,905±0,03	0,905±0,03	0,905±0,03	0,894±0,03	0,899±0,03
	0,3	0,908±0,02	0,908±0,02	0,908±0,02	0,899±0,02	0,904±0,02
	0,4	0,911±0,02	0,911±0,02	0,911±0,02	0,903±0,02	0,908±0,02
	0,5	0,911±0,02	0,911±0,02	0,911±0,02	0,902±0,02	0,908±0,02
10000	0,1	0,919±0,01	0,919±0,01	0,919±0,01	0,908±0,01	0,915±0,01
	0,2	0,921±0,01	0,921±0,01	0,921±0,01	0,909±0,01	0,918±0,01
	0,3	0,921±0,01	0,921±0,01	0,921±0,01	0,910±0,01	0,919±0,01
	0,4	0,922±0,01	0,922±0,01	0,922±0,01	0,911±0,01	0,920±0,01
	0,5	0,922±0,01	0,922±0,01	0,922±0,01	0,911±0,01	0,920±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Ki-kare dağılımından üretilen veriler ile elde edilen ortalama AUC değerleri incelendiğinde, 500, 1000 ve 10000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 11).

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama AUC değerleri, sınıf dağılım oranı 0,1 olduğunda; ham veride ve minimum-maksimum normalizasyonu sonrasında 0,791±0,18, z-skor normalizasyonu sonrasında 0,790±0,18, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,709±0,2, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,749±0,19 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama AUC değerlerinin; ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında 0,836±0,12, EG diskritizasyonu ile 0,771±0,14, EF diskritizasyonu ile 0,785±0,13 olduğu görüldü. Bu değerler sırasıyla; sınıf dağılım oranı 0,3 olduğunda 0,854±0,1, 0,799±0,11, 0,809±0,11, sınıf dağılım oranı 0,4 olduğunda 0,855±0,09, 0,807±0,1, 0,81±0,1 değerlerine yükseldi. Son olarak sınıf dağılım oranı 0,5 olduğunda ise ortalama AUC değerleri için ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında 0,86±0,08, EG diskritizasyonu ile 0,805±0,1, EF diskritizasyonu ile 0,806±0,1 değerleri elde edildi.

Gözlem sayısı 500 olduğunda gerçekleştirilen çalışmada sınıf dağılım oranı 0,1 için elde edilen ortalama AUC değerleri; ham veride 0,881±0,06, EG diskritizasyonu ile 0,846±0,06, EF diskritizasyonu ile 0,85±0,06 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise elde edilen bu değerler sırasıyla; 0,906±0,03, 0,893±0,03, 0,898±0,03 olarak belirlendi.

Gözlem sayısı 1000 olduğunda ve sınıf dağılım oranı 0,1 iken ortalama AUC değerleri; ham veride $0,898\pm 0,04$, EG diskritizasyonu ile $0,88\pm 0,04$, EF diskritizasyonu ile $0,885\pm 0,04$ olarak elde edildi. Sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama AUC değerleri; ham veri ile $0,911\pm 0,02$, EG diskritizasyonu ile $0,902\pm 0,02$, EF diskritizasyonu ile $0,908\pm 0,02$ olarak bulundu.

Gözlem sayısı 10000 olduğunda yapılan sınıflandırma çalışmasında sınıf dağılım oranı 0,1 iken elde edilen ortalama AUC değerleri; ham veride $0,919\pm 0,01$, EG diskritizasyonu ile $0,908\pm 0,01$, EF diskritizasyonu ile $0,915\pm 0,01$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda sırasıyla; $0,922\pm 0,01$, $0,911\pm 0,01$, $0,92\pm 0,01$ değerleri saptandı.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama AUC değerlerine Tablo 11’de yer verilmektedir.

Tablo 12. Ki-kare dağılımı için NB ile elde edilen ortalama duyarlılık oranları

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	$0,248\pm 0,31$	$0,248\pm 0,31$	$0,248\pm 0,31$	$0,175\pm 0,27$	$0,295\pm 0,32$
	0,2	$0,474\pm 0,25$	$0,474\pm 0,25$	$0,474\pm 0,25$	$0,387\pm 0,24$	$0,489\pm 0,26$
	0,3	$0,613\pm 0,20$	$0,613\pm 0,20$	$0,613\pm 0,20$	$0,547\pm 0,21$	$0,599\pm 0,20$
	0,4	$0,701\pm 0,16$	$0,701\pm 0,16$	$0,701\pm 0,16$	$0,666\pm 0,17$	$0,676\pm 0,16$
	0,5	$0,783\pm 0,13$	$0,783\pm 0,13$	$0,783\pm 0,13$	$0,746\pm 0,14$	$0,725\pm 0,14$
500	0,1	$0,369\pm 0,15$	$0,369\pm 0,15$	$0,369\pm 0,15$	$0,275\pm 0,14$	$0,337\pm 0,16$
	0,2	$0,557\pm 0,11$	$0,557\pm 0,11$	$0,557\pm 0,11$	$0,504\pm 0,12$	$0,549\pm 0,12$
	0,3	$0,682\pm 0,08$	$0,682\pm 0,08$	$0,682\pm 0,08$	$0,651\pm 0,09$	$0,666\pm 0,09$
	0,4	$0,762\pm 0,07$	$0,762\pm 0,07$	$0,762\pm 0,07$	$0,743\pm 0,07$	$0,748\pm 0,07$
	0,5	$0,827\pm 0,05$	$0,827\pm 0,05$	$0,827\pm 0,05$	$0,819\pm 0,06$	$0,812\pm 0,06$
1000	0,1	$0,390\pm 0,11$	$0,390\pm 0,11$	$0,390\pm 0,11$	$0,331\pm 0,11$	$0,379\pm 0,11$
	0,2	$0,579\pm 0,08$	$0,579\pm 0,08$	$0,579\pm 0,08$	$0,542\pm 0,08$	$0,576\pm 0,08$
	0,3	$0,688\pm 0,06$	$0,688\pm 0,06$	$0,688\pm 0,06$	$0,670\pm 0,06$	$0,683\pm 0,06$
	0,4	$0,770\pm 0,05$	$0,770\pm 0,05$	$0,770\pm 0,05$	$0,761\pm 0,05$	$0,764\pm 0,05$
	0,5	$0,829\pm 0,04$	$0,829\pm 0,04$	$0,829\pm 0,04$	$0,822\pm 0,04$	$0,821\pm 0,04$
10000	0,1	$0,444\pm 0,04$	$0,444\pm 0,04$	$0,444\pm 0,04$	$0,398\pm 0,04$	$0,444\pm 0,03$
	0,2	$0,609\pm 0,03$	$0,609\pm 0,03$	$0,609\pm 0,03$	$0,579\pm 0,03$	$0,610\pm 0,03$
	0,3	$0,709\pm 0,02$	$0,709\pm 0,02$	$0,709\pm 0,02$	$0,691\pm 0,02$	$0,708\pm 0,02$
	0,4	$0,783\pm 0,01$	$0,783\pm 0,01$	$0,783\pm 0,01$	$0,771\pm 0,02$	$0,780\pm 0,01$
	0,5	$0,838\pm 0,01$	$0,838\pm 0,01$	$0,838\pm 0,01$	$0,831\pm 0,01$	$0,834\pm 0,01$

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Ki-kare dağılımından üretilen veriler ile elde edilen ortalama duyarlılık oranları incelendiğinde, 100, 500, 1000 ve 10000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf

dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 12).

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama duyarlılık oranları sınıf dağılım oranı 0,1 olduğunda; ham veride $0,248 \pm 0,31$, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile $0,175 \pm 0,27$, eşit frekansta aralıklandırma (EF) diskritizasyonu ile $0,295 \pm 0,32$ olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda; ham veride $0,474 \pm 0,25$, EG diskritizasyonu ile $0,387 \pm 0,24$, EF diskritizasyonu ile $0,489 \pm 0,26$ olduğu görüldü. Sınıf dağılım oranı 0,3 için bu oranlar sırasıyla; $0,613 \pm 0,2$, $0,547 \pm 0,21$, $0,599 \pm 0,2$ olarak saptandı. Sınıf dağılım oranı 0,4 için elde edilen bu oranlar sırasıyla; $0,701 \pm 0,16$, $0,666 \pm 0,17$, $0,676 \pm 0,16$ olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda ortalama duyarlılık oranlarının; ham veride $0,783 \pm 0,13$, EG diskritizasyonu ile $0,746 \pm 0,14$, EF diskritizasyonu ile $0,725 \pm 0,14$ olduğu bulgusuna ulaşıldı.

Gözlem sayısı 500 ve sınıf dağılım oranı 0,1 için elde edilen ortalama duyarlılık oranları; ham veride $0,369 \pm 0,15$, EG diskritizasyonu ile $0,275 \pm 0,14$, EF diskritizasyonu ile $0,337 \pm 0,16$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama duyarlılık oranları; ham veride $0,827 \pm 0,05$, EG diskritizasyonu ile $0,819 \pm 0,06$, EF diskritizasyonu ile $0,812 \pm 0,06$ olarak belirlendi.

Gözlem sayısı 1000 olduğunda gerçekleştirilen sınıflandırma çalışmasında sınıf dağılım oranı 0,1 iken elde edilen ortalama duyarlılık oranları; ham veride $0,39 \pm 0,11$, EG diskritizasyonu ile $0,331 \pm 0,11$, EF diskritizasyonu ile $0,379 \pm 0,11$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda elde edilen bu oranlar sırasıyla; $0,829 \pm 0,04$, $0,822 \pm 0,04$, $0,821 \pm 0,04$ olarak saptandı.

Gözlem sayısı 10000 için gerçekleştirilen çalışmalar incelendiğinde, sınıf dağılım oranı 0,1 iken elde edilen ortalama duyarlılık oranları; ham veride $0,444 \pm 0,04$, EG diskritizasyonu ile $0,398 \pm 0,04$, EF diskritizasyonu ile $0,444 \pm 0,03$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ortalama duyarlılık oranları; ham veride $0,838 \pm 0,01$, EG diskritizasyonu ile $0,831 \pm 0,01$, EF diskritizasyonu ile $0,834 \pm 0,01$ olarak bulundu.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama duyarlılık oranlarına Tablo 12'de yer verilmektedir.

Tablo 13. Ki-kare dağılımı için NB ile elde edilen ortalama seçicilik oranları

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,975±0,04	0,975±0,04	0,975±0,04	0,960±0,05	0,937±0,06
	0,2	0,939±0,06	0,939±0,06	0,939±0,06	0,916±0,08	0,884±0,08
	0,3	0,899±0,08	0,899±0,08	0,899±0,08	0,858±0,10	0,842±0,10
	0,4	0,844±0,11	0,844±0,11	0,844±0,11	0,792±0,13	0,794±0,12
	0,5	0,777±0,14	0,777±0,14	0,777±0,14	0,711±0,15	0,731±0,15
500	0,1	0,983±0,01	0,983±0,01	0,983±0,01	0,981±0,01	0,975±0,02
	0,2	0,955±0,02	0,955±0,02	0,955±0,02	0,951±0,02	0,945±0,03
	0,3	0,920±0,03	0,920±0,03	0,920±0,03	0,909±0,03	0,908±0,04
	0,4	0,879±0,04	0,879±0,04	0,879±0,04	0,863±0,05	0,868±0,05
	0,5	0,827±0,05	0,827±0,05	0,827±0,05	0,805±0,06	0,820±0,05
1000	0,1	0,985±0,01	0,985±0,01	0,985±0,01	0,983±0,01	0,979±0,01
	0,2	0,958±0,02	0,958±0,02	0,958±0,02	0,953±0,02	0,95±0,02
	0,3	0,926±0,02	0,926±0,02	0,926±0,02	0,916±0,02	0,918±0,02
	0,4	0,885±0,03	0,885±0,03	0,885±0,03	0,873±0,03	0,879±0,03
	0,5	0,834±0,04	0,834±0,04	0,834±0,04	0,819±0,04	0,833±0,04
10000	0,1	0,985±0,003	0,985±0,003	0,985±0,003	0,984±0,003	0,982±0,003
	0,2	0,960±0,01	0,960±0,01	0,960±0,01	0,954±0,01	0,955±0,01
	0,3	0,929±0,01	0,929±0,01	0,929±0,01	0,919±0,01	0,925±0,01
	0,4	0,893±0,01	0,893±0,01	0,893±0,01	0,877±0,01	0,889±0,01
	0,5	0,848±0,01	0,848±0,01	0,848±0,01	0,828±0,01	0,846±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Ki-kare dağılımından üretilen veriler ile elde edilen ortalama seçicilik oranları incelendiğinde, 100, 500, 1000 ve 10000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 13).

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama seçicilik oranları sınıf dağılım oranı 0,1 olduğunda; ham veride 0,975±0,04, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,96±0,05, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,937±0,06 olarak bulundu. Sınıf dağılım oranı 0,2 için ortalama seçicilik oranlarının; ham veride 0,939±0,06, EG diskritizasyonu ile 0,916±0,08, EF diskritizasyonu ile 0,884±0,08 olduğu görüldü. Sınıf dağılım oranı 0,3 iken bu oranlar; ham veri ile 0,899±0,08, EG ile 0,858±0,1, EF ile 0,842±0,1 olarak hesaplandı. Sınıf dağılım oranı 0,4 iken elde edilen bu oranlar sırasıyla; 0,844±0,11, 0,792±0,13, 0,794±0,12 olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda ise ham veride 0,777±0,14, EG diskritizasyonu ile 0,711±0,15, EF diskritizasyonu ile 0,731±0,15 oranları elde edildi.

Gözlem sayısı 500 iken gerçekleştirilen sınıflandırma çalışmasında sınıf dağılım oranı 0,1 için elde edilen ortalama seçicilik oranları; ham veride $0,983\pm 0,01$, EG diskritizasyonu ile $0,981\pm 0,01$, EF diskritizasyonu ile $0,975\pm 0,02$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ortalama seçicilik oranları; ham veride $0,827\pm 0,05$, EG diskritizasyonu ile $0,805\pm 0,06$, EF diskritizasyonu ile $0,82\pm 0,05$ olarak elde edildi. Gözlem sayısı 1000 ve sınıf dağılım oranı 0,1 iken elde edilen ortalama seçicilik oranları; ham veride $0,985\pm 0,01$, EG diskritizasyonu ile $0,983\pm 0,01$, EF diskritizasyonu ile $0,979\pm 0,01$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda bu oranlar sırasıyla; $0,834\pm 0,04$, $0,819\pm 0,04$, $0,833\pm 0,04$ olarak elde edildi. Gözlem sayısı 10000 ve sınıf dağılım oranı 0,1 olduğunda elde edilen ortalama seçicilik oranları; ham veri ile $0,985\pm 0,003$, EG diskritizasyonu ile $0,984\pm 0,003$, EF diskritizasyonu ile $0,982\pm 0,003$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise ham veri ile $0,848\pm 0,01$, EG diskritizasyonu ile $0,828\pm 0,01$, EF diskritizasyonu ile $0,846\pm 0,01$ oranlarına ulaşıldı.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama seçicilik oranlarına Tablo 13'te yer verilmektedir.

Tablo 14. Ki-kare dağılımı için NB ile elde edilen ortalama pozitif kestirim değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	$0,559\pm 0,44$	$0,559\pm 0,44$	$0,559\pm 0,44$	$0,344\pm 0,41$	$0,375\pm 0,37$
	0,2	$0,694\pm 0,28$	$0,694\pm 0,28$	$0,694\pm 0,28$	$0,569\pm 0,31$	$0,537\pm 0,25$
	0,3	$0,743\pm 0,19$	$0,743\pm 0,19$	$0,743\pm 0,19$	$0,642\pm 0,20$	$0,642\pm 0,19$
	0,4	$0,767\pm 0,15$	$0,767\pm 0,15$	$0,767\pm 0,15$	$0,697\pm 0,15$	$0,703\pm 0,14$
	0,5	$0,789\pm 0,11$	$0,789\pm 0,11$	$0,789\pm 0,11$	$0,731\pm 0,11$	$0,740\pm 0,12$
500	0,1	$0,721\pm 0,20$	$0,721\pm 0,20$	$0,721\pm 0,20$	$0,627\pm 0,25$	$0,614\pm 0,22$
	0,2	$0,765\pm 0,11$	$0,765\pm 0,11$	$0,765\pm 0,11$	$0,726\pm 0,12$	$0,720\pm 0,11$
	0,3	$0,789\pm 0,07$	$0,789\pm 0,07$	$0,789\pm 0,07$	$0,758\pm 0,07$	$0,761\pm 0,08$
	0,4	$0,811\pm 0,06$	$0,811\pm 0,06$	$0,811\pm 0,06$	$0,787\pm 0,06$	$0,794\pm 0,06$
	0,5	$0,829\pm 0,04$	$0,829\pm 0,04$	$0,829\pm 0,04$	$0,809\pm 0,05$	$0,820\pm 0,04$
1000	0,1	$0,749\pm 0,13$	$0,749\pm 0,13$	$0,749\pm 0,13$	$0,688\pm 0,15$	$0,676\pm 0,13$
	0,2	$0,780\pm 0,07$	$0,780\pm 0,07$	$0,780\pm 0,07$	$0,747\pm 0,07$	$0,746\pm 0,07$
	0,3	$0,801\pm 0,05$	$0,801\pm 0,05$	$0,801\pm 0,05$	$0,776\pm 0,05$	$0,783\pm 0,05$
	0,4	$0,819\pm 0,04$	$0,819\pm 0,04$	$0,819\pm 0,04$	$0,801\pm 0,04$	$0,810\pm 0,04$
	0,5	$0,834\pm 0,03$	$0,835\pm 0,03$	$0,835\pm 0,03$	$0,821\pm 0,03$	$0,832\pm 0,03$
10000	0,1	$0,770\pm 0,04$	$0,770\pm 0,04$	$0,770\pm 0,04$	$0,731\pm 0,04$	$0,732\pm 0,04$
	0,2	$0,793\pm 0,02$	$0,793\pm 0,02$	$0,793\pm 0,02$	$0,761\pm 0,02$	$0,774\pm 0,02$
	0,3	$0,812\pm 0,02$	$0,812\pm 0,02$	$0,812\pm 0,02$	$0,785\pm 0,02$	$0,802\pm 0,02$
	0,4	$0,829\pm 0,01$	$0,829\pm 0,01$	$0,829\pm 0,01$	$0,807\pm 0,01$	$0,825\pm 0,01$
	0,5	$0,846\pm 0,01$	$0,846\pm 0,01$	$0,846\pm 0,01$	$0,829\pm 0,01$	$0,844\pm 0,01$

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Ki-kare dağılımından üretilen veriler ile elde edilen ortalama pozitif kestirim değerleri incelendiğinde, 100, 500 ve 10000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 14).

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama pozitif kestirim değerleri (PKD) sınıf dağılım oranı 0,1 olduğunda; ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrası $0,559\pm 0,44$, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile $0,344\pm 0,41$, eşit frekansta aralıklandırma (EF) diskritizasyonu ile $0,375\pm 0,37$ olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama PKD'nin; ham veride $0,694\pm 0,28$, EG diskritizasyonu ile $0,569\pm 0,31$, EF diskritizasyonu ile $0,537\pm 0,25$ olduğu görüldü. Sınıfların 0,3 dağılım oranı için bu değerler sırasıyla; $0,743\pm 0,19$, $0,642\pm 0,2$, $0,642\pm 0,19$ olarak hesaplandı. Sınıf dağılım oranı 0,4 için elde edilen ortalama değerler; ham veride $0,767\pm 0,15$, EG diskritizasyonu ile $0,697\pm 0,15$, EF diskritizasyonu ile $0,703\pm 0,14$ olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda ortalama PKD'nin; ham veride $0,789\pm 0,11$, EG diskritizasyonu ile $0,731\pm 0,11$, EF diskritizasyonu ile $0,74\pm 0,12$ olduğu bulgusuna ulaşıldı.

Gözlem sayısı 500 iken gerçekleştirilen sınıflandırma çalışmasında sınıf dağılım oranı 0,1 için ortalama PKD; ham veride $0,721\pm 0,2$, EG diskritizasyonu ile $0,627\pm 0,25$, EF diskritizasyonu ile $0,614\pm 0,22$ olarak elde edildi. Sınıf dağılım oranı 0,5 olduğunda; ham veri ile $0,829\pm 0,04$, EG diskritizasyonu ile $0,809\pm 0,05$, EF diskritizasyonu ile $0,82\pm 0,04$ değerleri saptandı.

Gözlem sayısının 1000 ve sınıf dağılım oranının 0,1 olduğu çalışmada elde edilen ortalama PKD; ham veri ile $0,749\pm 0,13$, EG diskritizasyonu ile $0,688\pm 0,15$, EF diskritizasyonu ile $0,676\pm 0,13$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ham veride $0,834\pm 0,03$, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında; $0,834\pm 0,03$, EG diskritizasyonu ile $0,821\pm 0,03$, EF diskritizasyonu ile $0,832\pm 0,03$ olarak elde edildi.

Gözlem sayısı 10000 olduğunda yapılan sınıflandırma çalışmasında sınıf dağılım oranı 0,1 iken elde edilen ortalama PKD; ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında $0,77\pm 0,04$, EG diskritizasyonu ile $0,731\pm 0,04$, EF

diskritizasyonu ile $0,732\pm 0,04$ olarak saptandı. Sınıf dağılım oranı 0,5 olduğunda ise bu değerler sırasıyla; $0,846\pm 0,01$, $0,829\pm 0,01$, $0,844\pm 0,01$ olarak belirlendi.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama pozitif kestirim değerlerine Tablo 14’te yer verilmektedir.

Tablo 15. Ki-kare dağılımı için NB ile elde edilen ortalama negatif kestirim değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	$0,922\pm 0,03$	$0,922\pm 0,03$	$0,922\pm 0,03$	$0,913\pm 0,03$	$0,924\pm 0,03$
	0,2	$0,880\pm 0,05$	$0,880\pm 0,05$	$0,880\pm 0,05$	$0,859\pm 0,05$	$0,876\pm 0,06$
	0,3	$0,849\pm 0,07$	$0,849\pm 0,07$	$0,849\pm 0,07$	$0,820\pm 0,07$	$0,835\pm 0,07$
	0,4	$0,815\pm 0,09$	$0,815\pm 0,09$	$0,815\pm 0,09$	$0,786\pm 0,09$	$0,792\pm 0,09$
	0,5	$0,792\pm 0,11$	$0,792\pm 0,11$	$0,792\pm 0,11$	$0,747\pm 0,12$	$0,736\pm 0,11$
500	0,1	$0,934\pm 0,02$	$0,934\pm 0,02$	$0,934\pm 0,02$	$0,924\pm 0,01$	$0,930\pm 0,02$
	0,2	$0,897\pm 0,02$	$0,897\pm 0,02$	$0,897\pm 0,02$	$0,885\pm 0,02$	$0,894\pm 0,02$
	0,3	$0,872\pm 0,03$	$0,872\pm 0,03$	$0,872\pm 0,03$	$0,860\pm 0,03$	$0,865\pm 0,03$
	0,4	$0,848\pm 0,04$	$0,848\pm 0,04$	$0,848\pm 0,04$	$0,835\pm 0,04$	$0,839\pm 0,04$
	0,5	$0,829\pm 0,04$	$0,829\pm 0,04$	$0,829\pm 0,04$	$0,818\pm 0,05$	$0,815\pm 0,05$
1000	0,1	$0,936\pm 0,01$	$0,936\pm 0,01$	$0,936\pm 0,01$	$0,930\pm 0,01$	$0,934\pm 0,01$
	0,2	$0,901\pm 0,02$	$0,901\pm 0,02$	$0,901\pm 0,02$	$0,893\pm 0,02$	$0,900\pm 0,02$
	0,3	$0,874\pm 0,02$	$0,874\pm 0,02$	$0,874\pm 0,02$	$0,867\pm 0,02$	$0,871\pm 0,02$
	0,4	$0,853\pm 0,03$	$0,853\pm 0,03$	$0,853\pm 0,03$	$0,846\pm 0,03$	$0,849\pm 0,03$
	0,5	$0,831\pm 0,03$	$0,831\pm 0,03$	$0,831\pm 0,03$	$0,823\pm 0,03$	$0,824\pm 0,03$
10000	0,1	$0,941\pm 0,004$	$0,941\pm 0,004$	$0,941\pm 0,004$	$0,936\pm 0,003$	$0,941\pm 0,003$
	0,2	$0,908\pm 0,01$	$0,908\pm 0,01$	$0,908\pm 0,01$	$0,901\pm 0,01$	$0,907\pm 0,01$
	0,3	$0,882\pm 0,01$	$0,882\pm 0,01$	$0,882\pm 0,01$	$0,874\pm 0,01$	$0,881\pm 0,01$
	0,4	$0,860\pm 0,01$	$0,860\pm 0,01$	$0,860\pm 0,01$	$0,852\pm 0,01$	$0,858\pm 0,01$
	0,5	$0,840\pm 0,01$	$0,840\pm 0,01$	$0,840\pm 0,01$	$0,831\pm 0,01$	$0,836\pm 0,01$

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Ki-kare dağılımından üretilen veriler ile elde edilen ortalama negatif kestirim değerleri incelendiğinde, 100, 500, 1000 ve 10000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 15).

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama negatif kestirim değerleri (NKD) sınıf dağılım oranı 0,1 olduğunda; ham veride $0,922\pm 0,03$, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile $0,913\pm 0,03$, eşit frekansta aralıklandırma (EF) diskritizasyonu ile $0,924\pm 0,03$ olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda bu ortalama değerlerin; ham veride $0,88\pm 0,05$, EG diskritizasyonu ile $0,859\pm 0,05$, EF diskritizasyonu ile $0,876\pm 0,06$ olduğu görüldü. Sınıf dağılım oranı 0,3 için ortalama NKD; ham veri ile

0,849±0,07, EG diskritizasyonu ile 0,82±0,07, EF diskritizasyonu ile 0,835±0,07 olarak elde edildi. Sınıf dağılım oranı 0,4 iken ortalama NKD'nin; ham veride 0,815±0,09, EG diskritizasyonu ile 0,786±0,09, EF diskritizasyonu ile 0,792±0,09 olduğu görüldü. Son olarak sınıf dağılım oranı 0,5 olduğunda ortalama NKD; ham veri ile 0,792±0,11, EG diskritizasyonu ile 0,747±0,12, EF diskritizasyonu ile 0,736±0,11 olarak saptandı.

Gözlem sayısı 500 iken gerçekleştirilen sınıflandırma çalışmasında sınıf dağılım oranı 0,1 için elde edilen ortalama negatif kestirim değerleri; ham veride 0,934±0,02, EG diskritizasyonu ile 0,924±0,01, EF diskritizasyonu ile 0,93±0,02 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ortalama NKD; ham veride 0,829±0,04, EG diskritizasyonu ile 0,818±0,05, EF diskritizasyonu ile 0,815±0,05 olarak elde edildi.

Gözlem sayısı 1000 ve sınıf dağılım oranı 0,1 olduğu çalışmada ortalama NKD; ham veri ile 0,936±0,01, EG diskritizasyonu ile 0,93±0,01, EF diskritizasyonu ile 0,934±0,01 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda bu değerler; ham veri ile 0,831±0,03, EG diskritizasyonu ile 0,823±0,03, EF diskritizasyonu ile 0,824±0,03 olarak saptandı.

Gözlem sayısının 10000 ve sınıf dağılım oranı 0,1 olduğu sınıflandırma çalışmasından elde edilen ortalama NKD; ham veride 0,941±0,004, EG diskritizasyonu ile 0,936±0,003, EF diskritizasyonu ile 0,941±0,003 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ortalama NKD; ham veri ile 0,84±0,01, EG ile 0,831±0,01, EF ile 0,836±0,01 olarak elde edildi.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama negatif kestirim değerlerine Tablo 15'te yer verilmektedir.

F dağılımından üretilen verilerin sınıflandırılması

Tablo 16. F dağılımı için NB ile elde edilen ortalama doğruluk oranları (%)

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	63,764±16,87	63,231±16,69	63,542±16,86	56,269±12,82	71,958±17,92
	0,2	70,413±13,06	69,850±13,02	70,100±12,97	60,725±11,32	79,791±12,43
	0,3	74,785±10,56	74,019±10,66	74,242±10,67	66,250±11,20	83,461±9,73
	0,4	76,773±9,81	76,085±9,92	76,283±9,86	69,285±10,98	83,927±8,66
	0,5	78,790±9,09	78,055±9,13	78,240±9,09	71,250±10,92	84,145±7,91
500	0,1	73,487±8,12	72,908±8,04	73,088±8,07	59,263±6,49	77,543±8,19
	0,2	78,549±5,76	78,014±5,81	78,193±5,87	64,158±5,86	85,977±4,96
	0,3	82,139±4,86	81,800±4,94	81,931±4,92	67,609±5,31	89,064±3,77
	0,4	84,323±4,21	83,996±4,24	84,152±4,20	70,608±5,26	90,128±3,28
	0,5	86,137±3,60	85,871±3,68	85,968±3,66	72,114±5,09	90,393±2,97

Tablo 16. Devam F dağılımı için NB ile elde edilen ortalama doğruluk oranları (%)

1000	0,1	76,254±5,85	75,774±5,92	75,933±5,93	60,838±5,08	79,705±5,44
	0,2	81,643±4,34	81,376±4,38	81,461±4,33	64,183±4,44	87,153±3,33
	0,3	85,231±3,35	85,072±3,41	85,127±3,42	66,683±4,22	89,952±2,45
	0,4	87,199±2,81	87,019±2,86	87,114±2,86	69,322±4,28	90,840±2,20
	0,5	88,425±2,38	88,291±2,42	88,365±2,40	71,078±4,24	91,182±1,97
10000	0,1	82,917±2,51	82,858±2,51	82,894±2,50	59,511±3,52	82,827±1,67
	0,2	87,142±1,95	87,125±1,96	87,142±1,95	59,156±2,89	88,647±1,01
	0,3	89,295±1,54	89,297±1,55	89,303±1,55	59,288±2,64	90,735±0,77
	0,4	90,561±1,07	90,572±1,07	90,574±1,06	60,456±2,40	91,596±0,66
	0,5	90,752±0,90	90,779±0,90	90,771±0,90	62,186±1,71	91,862±0,62

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Sınıflandırmada 100 gözlemde elde edilen ortalama doğruluk oranları, sınıfların 0,1 dağılım oranı için ham veride %63,764±16,87 olarak bulunurken minimum-maksimum normalizasyonu sonrası %63,231±16,69, z-skor normalizasyonu sonrası %63,542±16,86, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile %56,269±12,82, eşit frekansta aralıklandırma (EF) diskritizasyonu ile %71,958±17,92 olarak saptandı. Sınıf dağılım oranı 0,2 olduğunda bu oranların; ham veride %70,413±13,06, minimum-maksimum normalizasyonu sonrası %69,850±13,02, z-skor normalizasyonu sonrası %70,100±12,97, EG diskritizasyonu ile %60,725±11,32, EF diskritizasyonu ile %79,791±12,43 olduğu görüldü. Sınıf dağılım oranı 0,3 iken bu oranlar; ham veri ile %74,785±10,56, olup minimum-maksimum normalizasyonu sonrasında %74,019±10,66, z-skor normalizasyonu sonrasında %74,242±10,67, EG diskritizasyonu ile %66,250±11,20, EF diskritizasyonu ile %83,461±9,73 olarak hesaplandı. Sınıf dağılım oranı 0,4 iken bu oranlar sırasıyla; %76,773±9,81, %76,085±9,92, %76,283±9,86, %69,285±10,98, %83,927±8,66 olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda ortalama doğruluk oranları; ham veride %78,79±9,09, minimum-maksimum normalizasyonu sonrasında %78,055±9,13, z-skor normalizasyonu sonrasında %78,24±9,09, EG diskritizasyonu ile %71,25±10,92, EF diskritizasyonu ile %84,145±7,91 olarak bulundu.

Gözlem sayısı 500 iken gerçekleştirilen sınıflandırma çalışmasında sınıf dağılım oranı 0,1 için elde edilen ortalama doğruluk oranları; ham veride %73,487±8,12, minimum-maksimum normalizasyonu sonrasında %72,908±8,04, z-skor normalizasyonu sonrasında %73,088±8,07, EG diskritizasyonu ile %59,263±6,49, EF diskritizasyonu ile %77,543±8,19 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise ortalama doğruluk oranları; ham veride %86,137±3,6, minimum-maksimum normalizasyonu sonrası %85,871±3,68, z-skor

normalizasyonu sonrası %85,968±3,66, EG diskritizasyonu ile %72,114±5,09, EF diskritizasyonu ile %90,393±2,97 olarak elde edildi. Gözlem sayısı 1000 ve sınıf dağılım oranı 0,1 iken elde edilen ortalama doğruluk oranları; ham veride %76,254±5,85, minimum-maksimum normalizasyonu sonrasında %75,774±5,92, z-skor normalizasyonu sonrasında %75,933±5,93, EG diskritizasyonu ile %60,838±5,08, EF diskritizasyonu ile %79,705±5,44 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise bu oranlar sırasıyla; %88,425±2,38, %88,291±2,42, %88,365±2,4, %71,078±4,24, e %91,182±1,97 olarak elde edildi. Gözlem sayısı 10000 olduğunda ise sınıf dağılım oranı 0,1 iken ham veri ile %82,917±2,51 olarak bulunan ortalama doğruluk oranları, minimum-maksimum normalizasyonu sonrasında %82,858±2,51, z-skor normalizasyonu sonrasında %82,894±2,50, EG diskritizasyonu ile %59,511±3,52, EF diskritizasyonu ile %82,827±1,67 şeklindedir. Sınıf dağılım oranı 0,5 olduğunda ise ortalama doğruluk oranları sırasıyla; %90,752±0,9, %90,779±0,9, %90,771±0,9, %62,186±1,71, %91,862±0,62 olarak bulundu.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama doğruluk oranlarına Tablo 16'da yer verilmektedir.

Tablo 17. F dağılımı için NB ile elde edilen ortalama AUC değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,802±0,19	0,794±0,20	0,797±0,20	0,750±0,20	0,851±0,15
	0,2	0,819±0,14	0,813±0,15	0,816±0,14	0,762±0,14	0,898±0,09
	0,3	0,831±0,12	0,822±0,12	0,826±0,12	0,782±0,12	0,922±0,07
	0,4	0,841±0,10	0,831±0,11	0,837±0,11	0,782±0,12	0,921±0,06
	0,5	0,847±0,10	0,838±0,10	0,841±0,10	0,79±0,110	0,924±0,06
500	0,1	0,884±0,08	0,878±0,08	0,881±0,08	0,784±0,09	0,936±0,04
	0,2	0,899±0,05	0,892±0,06	0,897±0,06	0,796±0,06	0,959±0,02
	0,3	0,906±0,04	0,903±0,04	0,905±0,04	0,803±0,05	0,967±0,02
	0,4	0,913±0,04	0,909±0,04	0,912±0,04	0,805±0,05	0,969±0,02
	0,5	0,919±0,03	0,916±0,03	0,917±0,03	0,807±0,05	0,970±0,01
1000	0,1	0,907±0,05	0,902±0,05	0,905±0,05	0,786±0,06	0,954±0,02
	0,2	0,922±0,03	0,920±0,03	0,921±0,03	0,791±0,05	0,967±0,01
	0,3	0,932±0,03	0,931±0,03	0,931±0,03	0,793±0,05	0,973±0,01
	0,4	0,935±0,02	0,934±0,02	0,934±0,02	0,794±0,04	0,974±0,01
	0,5	0,939±0,02	0,937±0,02	0,938±0,02	0,796±0,04	0,975±0,01
10000	0,1	0,961±0,01	0,960±0,01	0,961±0,01	0,723±0,04	0,970±0,005
	0,2	0,966±0,01	0,966±0,01	0,966±0,01	0,704±0,03	0,975±0,003
	0,3	0,967±0,005	0,968±0,005	0,968±0,005	0,695±0,03	0,977±0,003
	0,4	0,969±0,004	0,969±0,004	0,969±0,004	0,692±0,02	0,978±0,003
	0,5	0,969±0,004	0,969±0,004	0,969±0,004	0,692±0,02	0,978±0,003

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama AUC değerleri sınıf dağılım oranı 0,1 olduğunda; ham veride $0,802\pm 0,19$, minimum-maksimum normalizasyonu sonrasında $0,794\pm 0,2$, z-skor normalizasyonu sonrasında $0,797\pm 0,2$, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile $0,75\pm 0,2$, eşit frekansta aralıklandırma (EF) diskritizasyonu ile $0,851\pm 0,15$ olarak bulundu. Sınıf dağılım oranı 0,2 için ortalama AUC değerlerinin; ham veride $0,819\pm 0,14$, olarak bulurken minimum-maksimum normalizasyonu sonrası $0,813\pm 0,15$, z-skor normalizasyonu sonrası $0,816\pm 0,14$, EG diskritizasyonu ile $0,762\pm 0,14$, EF diskritizasyonu ile $0,898\pm 0,09$ olduğu görüldü. Sınıf dağılım oranı 0,3 için bu değerler sırasıyla; $0,831\pm 0,12$, $0,822\pm 0,12$, $0,826\pm 0,12$, $0,782\pm 0,12$, $0,922\pm 0,07$ olarak hesaplanırken sınıf dağılım oranı 0,4 için $0,841\pm 0,1$, $0,831\pm 0,11$, $0,837\pm 0,11$, $0,782\pm 0,12$, $0,921\pm 0,06$ olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda ise ortalama AUC değerlerinin sırasıyla; $0,847\pm 0,1$, $0,838\pm 0,1$, $0,841\pm 0,1$, $0,79\pm 0,11$, $0,924\pm 0,06$ değerlerine yükseldi.

Gözlem sayısı 500 ve sınıf dağılım oranı 0,1 için elde edilen ortalama AUC değerleri; ham veride $0,884\pm 0,08$, minimum-maksimum normalizasyonu sonrasında $0,878\pm 0,08$, z-skor normalizasyonu sonrasında $0,881\pm 0,08$, EG diskritizasyonu ile $0,784\pm 0,09$, EF diskritizasyonu ile $0,936\pm 0,04$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama AUC değerleri sırasıyla; $0,919\pm 0,03$, $0,916\pm 0,03$, $0,917\pm 0,03$, $0,807\pm 0,05$, $0,97\pm 0,01$ olarak belirlendi. Gözlem sayısı 1000 ve sınıf dağılım oranı 0,1 olduğunda ortalama AUC değerleri; ham veride $0,907\pm 0,05$, minimum-maksimum normalizasyonu sonrasında $0,902\pm 0,05$, z-skor normalizasyonu sonrasında $0,905\pm 0,05$, EG diskritizasyonu ile $0,786\pm 0,06$, EF diskritizasyonu ile $0,954\pm 0,02$ olarak elde edilirken sınıf dağılım oranı 0,5 olduğunda ortalama değerler sırasıyla; $0,939\pm 0,02$, $0,937\pm 0,02$, $0,938\pm 0,02$, $0,796\pm 0,04$, $0,975\pm 0,01$ olarak bulundu. Gözlem sayısı 10000 ve sınıf dağılım oranı 0,1 iken ortalama AUC değerleri; ham veride $0,961\pm 0,01$, minimum-maksimum normalizasyonu sonrasında $0,96\pm 0,01$, z-skor normalizasyonu sonrasında $0,961\pm 0,01$, EG diskritizasyonu ile $0,723\pm 0,04$, EF diskritizasyonu ile $0,97\pm 0,005$ olarak bulunurken sınıf dağılım oranı 0,5 olduğunda bu değerler sırasıyla; $0,969\pm 0,004$, $0,969\pm 0,004$, $0,969\pm 0,004$, $0,692\pm 0,02$, $0,978\pm 0,003$ olarak elde edildi.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama AUC değerlerine Tablo 17’de yer verilmektedir.

Tablo 18. F dağılımı için NB ile elde edilen ortalama duyarlılık oranları

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,319±0,33	0,311±0,33	0,316±0,33	0,143±0,26	0,497±0,35
	0,2	0,470±0,25	0,460±0,25	0,464±0,25	0,252±0,22	0,681±0,24
	0,3	0,572±0,21	0,558±0,21	0,561±0,21	0,405±0,21	0,772±0,18
	0,4	0,631±0,18	0,619±0,18	0,622±0,18	0,548±0,19	0,801±0,15
	0,5	0,701±0,16	0,688±0,16	0,690±0,16	0,710±0,16	0,838±0,12
500	0,1	0,496±0,16	0,485±0,16	0,488±0,16	0,195±0,13	0,573±0,16
	0,2	0,607±0,12	0,597±0,12	0,600±0,12	0,306±0,12	0,755±0,10
	0,3	0,693±0,10	0,687±0,10	0,689±0,10	0,405±0,11	0,832±0,07
	0,4	0,753±0,09	0,748±0,09	0,749±0,09	0,539±0,11	0,870±0,06
	0,5	0,815±0,07	0,811±0,07	0,812±0,07	0,714±0,09	0,901±0,04
1000	0,1	0,546±0,12	0,537±0,12	0,539±0,12	0,224±0,10	0,612±0,11
	0,2	0,665±0,09	0,660±0,09	0,661±0,09	0,302±0,09	0,776±0,07
	0,3	0,751±0,07	0,748±0,07	0,749±0,07	0,378±0,09	0,842±0,05
	0,4	0,810±0,06	0,807±0,06	0,808±0,06	0,504±0,10	0,878±0,04
	0,5	0,862±0,05	0,860±0,05	0,861±0,05	0,704±0,09	0,907±0,03
10000	0,1	0,670±0,05	0,668±0,05	0,669±0,05	0,194±0,07	0,673±0,03
	0,2	0,767±0,05	0,766±0,05	0,766±0,04	0,192±0,06	0,803±0,02
	0,3	0,827±0,04	0,826±0,04	0,827±0,04	0,206±0,07	0,853±0,01
	0,4	0,877±0,04	0,876±0,04	0,877±0,04	0,266±0,08	0,886±0,01
	0,5	0,913±0,03	0,913±0,03	0,913±0,03	0,646±0,22	0,912±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama duyarlılık oranları sınıf dağılım oranı 0,1 olduğunda; ham veride 0,319±0,33, minimum-maksimum normalizasyonu sonrası 0,311±0,33, z-skor normalizasyonu sonrası 0,316±0,33, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,143±0,26, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,497±0,35 şeklindedir. Sınıf dağılım oranı 0,2 olduğunda bu oranların; ham veride 0,47±0,25, minimum-maksimum normalizasyonu sonrası 0,46±0,25, z-skor normalizasyonu sonrası 0,464±0,25, EG diskritizasyonu ile 0,252±0,22, EF diskritizasyonu ile 0,681±0,24 olduğu görüldü. Sınıf dağılım oranı 0,3 için bu oranlar; ham veri ile 0,572±0,21, olup minimum-maksimum normalizasyonu sonrasında 0,558±0,21, z-skor normalizasyonu sonrasında 0,561±0,21, EG diskritizasyonu ile 0,405±0,21, EF diskritizasyonu ile 0,772±0,18 olarak saptandı. Sınıf dağılım oranı 0,4 için elde edilen bu oranlar sırasıyla; 0,631±0,18, 0,619±0,18, 0,622±0,18, 0,548±0,19, 0,801±0,15 olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama duyarlılık oranlarının; ham veride 0,701±0,16, minimum-maksimum normalizasyonu sonrasında 0,688±0,16, z-skor normalizasyonu sonrasında 0,69±0,16, EG diskritizasyonu ile 0,71±0,16, EF diskritizasyonu ile 0,838±0,12 olduğu sonucuna ulaşıldı.

Gözlem sayısı 500 ve sınıf dağılım oranı 0,1 olduğunda elde edilen ortalama duyarlılık oranları; ham veride $0,496\pm 0,16$, minimum-maksimum normalizasyonu ile $0,485\pm 0,16$, z-skor ile $0,488\pm 0,16$, EG diskritizasyonu ile $0,195\pm 0,13$, EF diskritizasyonu ile $0,573\pm 0,16$ olarak bulundu. Sınıf dağılım oranı artarken duyarlılık oranlarının da arttığı çalışmada 0,5 yani tam dengeli hale geldiğinde; ham veride $0,815\pm 0,07$, minimum-maksimum normalizasyonu ile $0,811\pm 0,07$, z-skor normalizasyonu sonrası $0,812\pm 0,07$, EG diskritizasyonu ile $0,714\pm 0,09$, EF diskritizasyonu ile $0,901\pm 0,04$ olarak belirlendi.

Gözlem sayısı 1000 olduğunda gerçekleştirilen sınıflandırma çalışmasında sınıf dağılım oranı 0,1 için ortalama duyarlılık oranları; ham veride $0,546\pm 0,12$, minimum-maksimum normalizasyonu sonrası $0,537\pm 0,12$, z-skor normalizasyonu sonrası $0,539\pm 0,12$, EG diskritizasyonu ile $0,224\pm 0,1$, EF diskritizasyonu ile $0,612\pm 0,11$ olarak bulundu. Sınıf dağılım oranı artarken duyarlılık oranlarının da arttığı çalışmada 0,5 yani tam dengeli hale geldiğinde; ham veride $0,862\pm 0,05$, minimum-maksimum normalizasyonu sonrasında $0,86\pm 0,05$, z-skor normalizasyonu sonrasında $0,861\pm 0,05$, EG diskritizasyonu ile $0,704\pm 0,09$, EF diskritizasyonu ile $0,907\pm 0,03$ değerlerine ulaşıldı.

Gözlem sayısı 10000 olduğunda yapılan çalışmada sınıf dağılım oranı 0,1 için edilen ortalama duyarlılık oranları; ham veride $0,67\pm 0,05$, minimum-maksimum normalizasyonu ile $0,668\pm 0,05$, z-skor ile $0,669\pm 0,05$, EG diskritizasyonu ile $0,194\pm 0,07$, EF diskritizasyonu ile $0,673\pm 0,03$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise bu oranlar sırasıyla; $0,913\pm 0,03$, $0,913\pm 0,03$, $0,913\pm 0,03$, $0,646\pm 0,22$, $0,912\pm 0,01$ olarak elde edildi.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama duyarlılık oranlarına Tablo 18’de yer verilmektedir.

Tablo 19. F dağılımı için NB ile elde edilen ortalama seçicilik oranları

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	$0,956\pm 0,05$	$0,954\pm 0,05$	$0,955\pm 0,05$	$0,982\pm 0,03$	$0,942\pm 0,06$
	0,2	$0,938\pm 0,06$	$0,937\pm 0,06$	$0,938\pm 0,06$	$0,962\pm 0,05$	$0,915\pm 0,07$
	0,3	$0,924\pm 0,07$	$0,922\pm 0,07$	$0,924\pm 0,07$	$0,920\pm 0,08$	$0,898\pm 0,08$
	0,4	$0,905\pm 0,09$	$0,902\pm 0,09$	$0,904\pm 0,09$	$0,837\pm 0,12$	$0,877\pm 0,10$
	0,5	$0,875\pm 0,11$	$0,873\pm 0,11$	$0,875\pm 0,11$	$0,715\pm 0,16$	$0,845\pm 0,12$
500	0,1	$0,974\pm 0,02$	$0,973\pm 0,02$	$0,974\pm 0,02$	$0,991\pm 0,01$	$0,978\pm 0,02$
	0,2	$0,964\pm 0,02$	$0,963\pm 0,02$	$0,964\pm 0,02$	$0,978\pm 0,02$	$0,964\pm 0,02$
	0,3	$0,950\pm 0,03$	$0,949\pm 0,03$	$0,950\pm 0,03$	$0,947\pm 0,03$	$0,949\pm 0,03$
	0,4	$0,934\pm 0,04$	$0,932\pm 0,04$	$0,934\pm 0,04$	$0,873\pm 0,06$	$0,933\pm 0,03$
	0,5	$0,908\pm 0,05$	$0,907\pm 0,05$	$0,908\pm 0,05$	$0,728\pm 0,08$	$0,907\pm 0,04$

Tablo 19. Devam F dağılımı için NB ile elde edilen ortalama seçicilik oranları

1000	0,1	0,979±0,01	0,979±0,01	0,979±0,01	0,992±0,01	0,982±0,01
	0,2	0,968±0,02	0,968±0,02	0,968±0,02	0,981±0,01	0,967±0,01
	0,3	0,953±0,02	0,953±0,02	0,954±0,02	0,956±0,02	0,957±0,02
	0,4	0,934±0,03	0,933±0,03	0,934±0,03	0,883±0,05	0,938±0,02
	0,5	0,906±0,04	0,905±0,04	0,906±0,04	0,718±0,08	0,917±0,03
10000	0,1	0,989±0,004	0,989±0,004	0,989±0,004	0,996±0,003	0,983±0,003
	0,2	0,976±0,01	0,976±0,01	0,976±0,01	0,991±0,01	0,970±0,005
	0,3	0,959±0,02	0,959±0,02	0,960±0,02	0,980±0,02	0,961±0,01
	0,4	0,935±0,02	0,935±0,02	0,935±0,02	0,943±0,05	0,946±0,01
	0,5	0,902±0,03	0,902±0,03	0,902±0,03	0,597±0,22	0,925±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama seçicilik oranları sınıf dağılım oranı 0,1 olduğunda; ham veride 0,956±0,05, minimum-maksimum normalizasyonu sonrası 0,954±0,05, z-skor normalizasyonu sonrası 0,955±0,05, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,982±0,03, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,942±0,06 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama seçicilik oranlarının; ham veride 0,938±0,06, olarak bulurken minimum-maksimum normalizasyonu sonrası 0,937±0,06, z-skor normalizasyonu sonrası 0,938±0,06, EG diskritizasyonu ile 0,962±0,05, EF diskritizasyonu ile 0,915±0,07 olduğu görüldü. Sınıf dağılım oranı 0,3 iken bu oranlar; ham veri ile 0,924±0,07 olup minimum-maksimum normalizasyonu sonrası 0,922±0,07, z-skor normalizasyonu sonrasında 0,924±0,07, EG diskritizasyonu ile 0,92±0,08, EF diskritizasyonu ile 0,898±0,08 olarak hesaplandı. Sınıf dağılım oranı 0,4 iken elde edilen bu oranlar sırasıyla; 0,905±0,09, 0,902±0,09, 0,904±0,09, 0,837±0,12, 0,877±0,1 olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda ise bu oranların sırasıyla; 0,875±0,11, 0,873±0,11, 0,875±0,11, 0,715±0,16, 0,845±0,12 olduğu görüldü.

Gözlem sayısı 500 ve sınıf dağılım oranı 0,1 için elde edilen ortalama seçicilik oranları; ham veride 0,974±0,02, minimum-maksimum normalizasyonu sonrası 0,973±0,02, z-skor normalizasyonu sonrası 0,974±0,02, EG diskritizasyonu ile 0,991±0,01, EF diskritizasyonu ile 0,978±0,02 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise elde edilen ortalama seçicilik oranlarının; ham veride 0,908±0,05, minimum-maksimum normalizasyonu sonrası 0,907±0,05, z-skor normalizasyonu sonrası 0,908±0,05, EG diskritizasyonu ile 0,728±0,08, EF diskritizasyonu ile 0,907±0,04 olduğu bulgusuna ulaşıldı.

Gözlem sayısı 1000 olduğunda, 0,1 sınıf dağılım oranı için elde edilen ortalama seçicilik oranları; ham veride 0,979±0,01, minimum-maksimum normalizasyonu sonrası

0,979±0,01, z-skor normalizasyonu sonrası 0,979±0,01, EG diskritizasyonu ile 0,992±0,01, EF diskritizasyonu ile 0,982±0,01 olarak bulundu. Bu oranlar, sınıf dağılım oranı 0,5 olduğunda ham veride 0,906±0,04, minimum-maksimum normalizasyonu sonrası 0,905±0,04, z-skor normalizasyonu sonrası 0,906±0,04, EG diskritizasyonu ile 0,718±0,08, EF diskritizasyonu ile 0,917±0,03 olarak elde edildi.

Gözlem sayısı 10000 olduğunda yapılan çalışmada sınıf dağılım oranı 0,1 iken elde edilen ortalama seçicilik oranları; ham veride 0,989±0,004, minimum-maksimum normalizasyonu ile 0,989±0,004, z-skor ile 0,989±0,004, EG diskritizasyonu ile 0,996±0,003, EF diskritizasyonu ile 0,983±0,003 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise ham veride 0,902±0,03, minimum-maksimum normalizasyonu ile 0,902±0,03, z-skor ile 0,902±0,03, EG diskritizasyonu ile 0,597±0,22, EF diskritizasyonu ile 0,925±0,01 oranlarına ulaşıldı.

Çalışmada tüm örneklem dağılımlarından üretilen verilerin sınıflandırılması sonucunda 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımları (0,2, 0,3, 0,4) elde edilen ortalama seçicilik oranlarına Tablo 19'da yer verilmektedir.

Tablo 20. F dağılımı için NB ile elde edilen ortalama pozitif kestirim değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,482±0,4	0,463±0,40	0,472±0,4	0,482±0,46	0,538±0,35
	0,2	0,691±0,28	0,684±0,29	0,690±0,28	0,650±0,38	0,698±0,21
	0,3	0,786±0,19	0,777±0,19	0,782±0,19	0,705±0,26	0,783±0,15
	0,4	0,829±0,15	0,823±0,15	0,825±0,15	0,708±0,18	0,826±0,12
	0,5	0,861±0,11	0,857±0,12	0,859±0,12	0,724±0,13	0,854±0,10
500	0,1	0,696±0,17	0,689±0,17	0,692±0,17	0,706±0,30	0,756±0,15
	0,2	0,816±0,10	0,809±0,10	0,813±0,10	0,777±0,16	0,846±0,08
	0,3	0,861±0,07	0,857±0,07	0,860±0,07	0,774±0,11	0,879±0,06
	0,4	0,887±0,05	0,884±0,05	0,887±0,05	0,745±0,08	0,898±0,05
	0,5	0,901±0,04	0,900±0,04	0,901±0,04	0,728±0,06	0,908±0,04
1000	0,1	0,753±0,11	0,746±0,11	0,751±0,11	0,777±0,19	0,795±0,10
	0,2	0,843±0,07	0,841±0,07	0,843±0,07	0,808±0,11	0,857±0,06
	0,3	0,876±0,05	0,875±0,05	0,876±0,05	0,793±0,09	0,894±0,04
	0,4	0,893±0,04	0,892±0,04	0,893±0,04	0,751±0,08	0,906±0,03
	0,5	0,903±0,03	0,903±0,03	0,903±0,03	0,718±0,05	0,917±0,03
10000	0,1	0,871±0,04	0,870±0,04	0,871±0,04	0,868±0,08	0,817±0,03
	0,2	0,892±0,03	0,893±0,03	0,893±0,03	0,863±0,08	0,870±0,02
	0,3	0,899±0,03	0,900±0,03	0,900±0,03	0,840±0,09	0,904±0,01
	0,4	0,901±0,03	0,902±0,03	0,902±0,03	0,796±0,10	0,916±0,01
	0,5	0,904±0,03	0,905±0,03	0,905±0,03	0,648±0,09	0,924±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama pozitif kestirim değerleri (PKD) sınıf dağılım oranı 0,1 olduğunda; ham veride $0,482\pm 0,4$, minimum-maksimum normalizasyonu sonrası $0,463\pm 0,4$, z-skor normalizasyonu sonrası $0,472\pm 0,4$, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile $0,482\pm 0,46$, eşit frekansta aralıklandırma (EF) diskritizasyonu ile $0,538\pm 0,35$ olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama PKD; ham veride $0,691\pm 0,28$ olarak bulunurken minimum-maksimum normalizasyonu sonrası $0,684\pm 0,29$, z-skor normalizasyonu sonrası $0,69\pm 0,28$, EG diskritizasyonu ile $0,65\pm 0,38$, EF diskritizasyonu ile $0,698\pm 0,21$ olduğu görüldü. Bu değerler sırasıyla 0,3 sınıf dağılım oranı için $0,786\pm 0,19$, $0,777\pm 0,19$, $0,782\pm 0,19$, $0,705\pm 0,26$, $0,783\pm 0,15$ olarak hesaplanırken 0,4 sınıf dağılım oranı için $0,829\pm 0,15$, $0,823\pm 0,15$, $0,825\pm 0,15$, $0,708\pm 0,18$, $0,826\pm 0,12$ olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda ise bu değerlerin sırasıyla; $0,861\pm 0,11$, $0,857\pm 0,12$, $0,859\pm 0,12$, $0,724\pm 0,13$, $0,854\pm 0,1$ olduğu görüldü.

Gözlem sayısı 500 ve sınıf dağılım oranı 0,1 için elde edilen ortalama PKD; ham veride $0,696\pm 0,17$, minimum-maksimum normalizasyonu ile $0,689\pm 0,17$, z-skor normalizasyonu ile $0,692\pm 0,17$, EG diskritizasyonu ile $0,706\pm 0,3$, EF diskritizasyonu ile $0,756\pm 0,15$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise bu değerlerin sırasıyla $0,901\pm 0,04$, $0,9\pm 0,04$, $0,901\pm 0,04$, $0,728\pm 0,06$, $0,908\pm 0,04$ değerlerine yükseldiği görüldü.

Gözlem sayısı 1000 ve sınıf dağılım oranı 0,1 iken elde edilen ortalama PKD; ham veride $0,753\pm 0,11$, minimum-maksimum normalizasyonu sonrasında $0,746\pm 0,11$, z-skor normalizasyonu sonrasında $0,751\pm 0,11$, EG diskritizasyonu ile $0,777\pm 0,19$, EF diskritizasyonu ile $0,795\pm 0,1$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda bu değerler sırasıyla; $0,903\pm 0,03$, $0,903\pm 0,03$, $0,903\pm 0,03$, $0,718\pm 0,05$, $0,917\pm 0,03$ olarak elde edildi.

Gözlem sayısının 10000 ve sınıf dağılım oranı 0,1 olduğu çalışmadan elde edilen ortalama PKD; ham veride $0,871\pm 0,04$, minimum-maksimum normalizasyonu sonrasında $0,87\pm 0,04$, z-skor normalizasyonu sonrasında $0,871\pm 0,04$ olarak bulunurken EG ile $0,868\pm 0,08$, EF ile $0,817\pm 0,03$ sonuçlarına ulaşıldı. Sınıf dağılım oranı 0,5 olduğunda ise bu oranlar sırasıyla; $0,904\pm 0,03$, $0,905\pm 0,03$, $0,905\pm 0,03$, $0,648\pm 0,09$, $0,924\pm 0,01$ olarak bulundu.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama pozitif kestirim değerlerine Tablo 20'de yer verilmektedir.

Tablo 21. F dağılımı için NB ile elde edilen ortalama negatif kestirim değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,928±0,03	0,927±0,03	0,927±0,03	0,912±0,03	0,945±0,04
	0,2	0,879±0,05	0,877±0,05	0,878±0,05	0,839±0,04	0,923±0,06
	0,3	0,839±0,07	0,834±0,07	0,836±0,07	0,787±0,06	0,906±0,07
	0,4	0,794±0,08	0,788±0,08	0,789±0,08	0,742±0,09	0,875±0,08
	0,5	0,757±0,10	0,748±0,10	0,749±0,10	0,723±0,12	0,850±0,1
500	0,1	0,946±0,02	0,945±0,02	0,945±0,02	0,917±0,01	0,954±0,02
	0,2	0,908±0,02	0,906±0,02	0,907±0,02	0,850±0,02	0,941±0,02
	0,3	0,880±0,03	0,878±0,03	0,878±0,03	0,789±0,03	0,930±0,03
	0,4	0,852±0,04	0,850±0,04	0,851±0,04	0,742±0,04	0,916±0,03
	0,5	0,834±0,05	0,831±0,05	0,832±0,05	0,723±0,06	0,903±0,04
1000	0,1	0,951±0,01	0,950±0,01	0,950±0,01	0,920±0,01	0,958±0,01
	0,2	0,921±0,02	0,920±0,02	0,920±0,02	0,849±0,02	0,946±0,02
	0,3	0,900±0,02	0,899±0,03	0,899±0,03	0,783±0,02	0,934±0,02
	0,4	0,882±0,03	0,880±0,03	0,881±0,03	0,730±0,04	0,921±0,02
	0,5	0,870±0,04	0,868±0,04	0,869±0,04	0,713±0,05	0,908±0,03
10000	0,1	0,964±0,01	0,964±0,01	0,964±0,01	0,918±0,01	0,964±0,004
	0,2	0,944±0,01	0,944±0,01	0,944±0,01	0,831±0,01	0,952±0,005
	0,3	0,929±0,02	0,929±0,02	0,929±0,02	0,743±0,01	0,939±0,01
	0,4	0,920±0,02	0,920±0,02	0,920±0,02	0,659±0,02	0,926±0,01
	0,5	0,914±0,02	0,914±0,02	0,914±0,02	0,664±0,08	0,913±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Sınıflandırma sonucunda 100 gözlem için elde edilen negatif kestirim değeri (NKD) için sınıf dağılım oranı 0,1 olduğunda; ham veride 0,928±0,03, minimum-maksimum normalizasyonu ile 0,927±0,03, z-skor ile 0,927±0,03, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,912±0,03, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,945±0,04 ortalama değerlerine ulaşıldı. Sınıf dağılım oranı 0,2 olduğunda ham veride 0,879±0,05 olarak bulunan bu ortalama değerlerin; minimum-maksimum normalizasyonu sonrası 0,877±0,05, z-skor normalizasyonu sonrası 0,878±0,05, EG diskritizasyonu ile 0,839±0,04, EF diskritizasyonu ile 0,923±0,06 olduğu görüldü. Sınıf dağılım oranı 0,3 için ortalama NKD sırasıyla; 0,839±0,07, 0,834±0,07, 0,836±0,07, 0,787±0,06, 0,906±0,07 olarak hesaplandı. Sınıf dağılım oranı 0,4 iken bu değerler sırasıyla; 0,794±0,08, 0,788±0,08, 0,789±0,08, 0,742±0,09, 0,875±0,08 olduğu gözlemlendi. Son olarak sınıf dağılım oranı 0,5 olduğunda ortalama NKD için sırasıyla; 0,757±0,1, 0,748±0,1, 0,749±0,1, 0,723±0,12, 0,85±0,1 değerlerine ulaşıldı.

Gözlem sayısı 500 ve sınıf dağılım oranı 0,1 için elde edilen ortalama negatif kestirim değerleri; ham veride 0,946±0,02, minimum-maksimum normalizasyonu sonrası 0,945±0,02, z-skor normalizasyonu sonrası 0,945±0,02, EG diskritizasyonu ile 0,917±0,01, EF

diskritizasyonu ile $0,954\pm 0,02$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ortalama NKD sırasıyla; $0,834\pm 0,05$, minimum-maksimum normalizasyonu sonrasında $0,831\pm 0,05$, z-skor normalizasyonu sonrasında $0,832\pm 0,05$, EG diskritizasyonu ile $0,723\pm 0,06$, EF diskritizasyonu ile $0,903\pm 0,04$ olarak elde edildi.

Gözlem sayısı 1000 olduğunda elde edilen negatif kestirim değerleri; sınıf dağılım oranı 0,1 için ham veride $0,951\pm 0,01$, minimum-maksimum normalizasyonu sonrası $0,95\pm 0,01$, z-skor normalizasyonu sonrası $0,95\pm 0,01$, EG diskritizasyonu ile $0,92\pm 0,01$, EF diskritizasyonu ile $0,958\pm 0,01$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda; ham veride $0,87\pm 0,04$, minimum-maksimum normalizasyonu sonrası $0,868\pm 0,04$, z-skor normalizasyonu sonrası $0,869\pm 0,04$, EG ile $0,713\pm 0,05$, EF ile $0,908\pm 0,03$ değerleri elde edildi.

Gözlem sayısının 10000 olduğu çalışmada, sınıf dağılım oranı 0,1 iken ortalama NKD; ham veride $0,964\pm 0,01$, minimum-maksimum normalizasyonu sonrasında $0,964\pm 0,01$, z-skor normalizasyonu sonrasında $0,964\pm 0,01$, EG diskritizasyonu ile $0,918\pm 0,01$, EF diskritizasyonu ile $0,964\pm 0,004$ olarak elde edildi. Sınıf dağılım oranı 0,5 olduğunda ise; ham veri ile $0,914\pm 0,02$, minimum-maksimum normalizasyonu ile $0,914\pm 0,02$, z-skor ile $0,914\pm 0,02$, EG ile $0,664\pm 0,08$, EF ile $0,913\pm 0,01$ bulgularına ulaşıldı.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama negatif kestirim değerlerine Tablo 21’de yer verilmektedir.

C5.0 KARAR AĞACI ALGORİTMASI İLE SINIFLANDIRMA

Normal dağılımdan üretilen verilerin sınıflandırılması

Tablo 22. Normal Dağılım için C5.0 ile elde edilen ortalama doğruluk oranları (%)

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	64,683±17,01	64,683±17,01	64,683±17,01	50,897±5,39	50,497±3,99
	0,2	71,263±13,20	71,263±13,20	71,263±13,20	53,063±8,10	53,403±8,68
	0,3	73,518±11,60	73,518±11,60	73,518±11,60	58,025±11,22	56,685±11,15
	0,4	74,896±10,40	74,896±10,40	74,896±10,40	64,967±11,15	65,179±11,38
	0,5	75,370±9,87	75,370±9,87	75,370±9,87	66,505±11,37	65,290±11,11
500	0,1	69,807±8,35	69,807±8,35	69,807±8,35	60,658±8,22	56,789±7,73
	0,2	75,953±6,04	75,951±6,05	75,951±6,05	70,316±6,87	70,131±6,86
	0,3	78,540±4,74	78,533±4,74	78,533±4,74	74,933±5,46	74,896±5,28
	0,4	79,661±4,38	79,661±4,38	79,663±4,39	76,898±4,60	76,642±4,79
	0,5	79,885±4,10	79,886±4,10	79,887±4,10	77,747±4,35	77,194±4,54

Tablo 22. Devam Normal Dağılım için C5.0 ile elde edilen ortalama doğruluk oranları (%)

1000	0,1	72,081±5,82	72,081±5,82	72,081±5,82	65,260±6,44	60,273±6,31
	0,2	77,341±4,23	77,341±4,23	77,341±4,23	73,689±4,52	72,122±4,80
	0,3	79,817±3,40	79,817±3,40	79,817±3,40	77,600±3,78	77,448±3,98
	0,4	80,870±2,97	80,870±2,97	80,870±2,97	79,582±3,13	79,214±3,24
	0,5	81,078±2,88	81,078±2,88	81,078±2,88	79,912±2,98	79,445±3,16
10000	0,1	76,253±2,01	76,251±2,02	76,251±2,02	73,999±2,00	71,875±2,15
	0,2	80,932±1,39	80,931±1,39	80,933±1,39	79,758±1,37	78,770±1,46
	0,3	83,121±1,08	83,121±1,08	83,121±1,08	82,490±1,08	81,907±1,03
	0,4	84,193±0,90	84,193±0,90	84,193±0,90	83,863±0,88	83,456±0,93
	0,5	84,562±0,82	84,562±0,82	84,562±0,82	84,343±0,79	84,091±0,87

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Normal dağılımdan üretilen veriler ile elde edilen ortalama doğruluk oranları incelendiğinde, 100 ve 1000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 22).

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama doğruluk oranları sınıf dağılım oranı 0,1 olduğunda; ham veride %64,683±17,01, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile %50,897±5,39, eşit frekansta aralıklandırma (EF) diskritizasyonu ile %50,497±3,99 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama doğruluk oranlarının; ham veride %71,263±13,20, EG diskritizasyonu ile %53,063±8,10, EF diskritizasyonu ile %53,403±8,68 olduğu görüldü. Sınıf dağılım oranı 0,3 iken bu oranlar sırasıyla; %73,518±11,60, %58,025±11,22, %56,685±11,15 olarak bulurken 0,4 sınıf dağılım oranında; %74,896±10,40, %64,967±11,15, %65,179±11,38 olarak elde edildi. Son olarak sınıf dağılım oranı 0,5 olduğunda ortalama doğruluk oranları için ham veride %75,37±9,87, EG diskritizasyonu ile %66,505±11,37, EF diskritizasyonu ile %65,29±11,11 olduğu bulgusuna ulaşıldı.

Gözlem sayısı 500 olduğunda, 0,1 sınıf dağılım oranı için elde edilen ortalama doğruluk oranları; ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında %69,807±8,35, EG diskritizasyonu ile %60,658±8,22, EF diskritizasyonu ile %56,789±7,73 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ham veride %79,885±4,1 olarak elde edilen ortalama doğruluk oranları; minimum-maksimum normalizasyonu ile %79,886±4,1, z-skor normalizasyonu ile %79,887±4,1, EG diskritizasyonu ile %77,747±4,35, EF diskritizasyonu ile %77,194±4,54 olarak belirlendi.

Gözlem sayısı 1000 olduğunda; 0,1 sınıf dağılım oranı için elde edilen ortalama doğruluk oranları; ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu %72,081±5,82, EG diskritizasyonu ile %65,260±6,44, EF diskritizasyonu ile %60,273±6,31 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda bu oranlar sırasıyla; %81,078±2,88, %79,912±2,98, %79,445±3,16 olarak elde edildi.

Gözlem sayısı 10000 olduğunda, 0,1 sınıf dağılım oranı için elde edilen ortalama doğruluk oranları; ham veride ve minimum-maksimum normalizasyonu sonrasında %76,253±2,01, z-skor normalizasyonu sonrasında %76,251±2,02, EG diskritizasyonu ile %73,999±2,00, EF diskritizasyonu ile %71,875±2,15 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise bu oranlar sırasıyla; %84,562±0,82, %84,562±0,82, %84,343±0,79, %84,091±0,87 olarak elde edildi.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama doğruluk oranlarına Tablo 22’de yer verilmektedir.

Tablo 23. Normal Dağılım için C5.0 ile elde edilen ortalama AUC değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,681±0,19	0,681±0,19	0,681±0,19	0,510±0,06	0,506±0,05
	0,2	0,748±0,15	0,748±0,15	0,748±0,15	0,534±0,09	0,538±0,09
	0,3	0,770±0,12	0,770±0,12	0,770±0,12	0,579±0,12	0,570±0,12
	0,4	0,783±0,11	0,783±0,11	0,783±0,11	0,648±0,12	0,645±0,12
	0,5	0,791±0,10	0,791±0,10	0,791±0,10	0,664±0,12	0,646±0,12
500	0,1	0,749±0,11	0,749±0,11	0,749±0,11	0,635±0,10	0,608±0,11
	0,2	0,812±0,07	0,812±0,07	0,812±0,07	0,738±0,08	0,731±0,08
	0,3	0,838±0,05	0,838±0,05	0,838±0,05	0,780±0,06	0,777±0,06
	0,4	0,850±0,05	0,850±0,05	0,850±0,05	0,799±0,05	0,794±0,05
	0,5	0,852±0,04	0,852±0,04	0,852±0,04	0,809±0,05	0,803±0,05
1000	0,1	0,777±0,08	0,777±0,08	0,777±0,08	0,704±0,08	0,659±0,08
	0,2	0,831±0,05	0,831±0,05	0,831±0,05	0,788±0,05	0,770±0,05
	0,3	0,851±0,04	0,851±0,04	0,851±0,04	0,817±0,04	0,812±0,04
	0,4	0,862±0,03	0,862±0,03	0,862±0,03	0,834±0,03	0,830±0,04
	0,5	0,865±0,03	0,865±0,03	0,865±0,03	0,837±0,03	0,834±0,03
10000	0,1	0,858±0,02	0,858±0,02	0,862±0,03	0,833±0,03	0,825±0,03
	0,2	0,883±0,02	0,883±0,02	0,883±0,02	0,871±0,02	0,870±0,02
	0,3	0,895±0,01	0,895±0,01	0,895±0,01	0,887±0,01	0,882±0,01
	0,4	0,901±0,01	0,901±0,01	0,901±0,01	0,894±0,01	0,891±0,01
	0,5	0,904±0,01	0,904±0,01	0,904±0,01	0,898±0,01	0,895±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Normal dağılımdan üretilen veriler ile elde edilen ortalama AUC değerleri incelendiğinde, 100, 500 ve 1000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 23).

Ortalama AUC değerleri 100 gözlem için incelendiğinde; sınıf dağılım oranı 0,1 iken ham veride $0,681\pm 0,19$ olarak bulunurken eşit genişlikte aralıklandırma (EG) diskritizasyonu ile $0,51\pm 0,06$, eşit frekansta aralıklandırma (EF) diskritizasyonu ile $0,506\pm 0,05$ değerlerine ulaşıldı. Sınıf dağılım oranı 0,2 olduğunda ortalama AUC değerleri sırasıyla; $0,748\pm 0,15$, $0,534\pm 0,09$, $0,538\pm 0,09$, sınıf dağılım oranı 0,3 olduğunda; $0,77\pm 0,12$, $0,579\pm 0,12$, $0,57\pm 0,12$ ve sınıf dağılım oranı 0,4 olduğunda; $0,783\pm 0,11$, $0,648\pm 0,12$, $0,645\pm 0,12$ değerleri bulundu. Dağılım oranı 0,5 olduğunda ise ham veri ile $0,791\pm 0,1$, EG diskritizasyonu ile $0,664\pm 0,12$ ve EF diskritizasyonu ile $0,646\pm 0,12$ değerleri elde edildi.

Gözlem sayısı 500 olduğunda, 0,1 sınıf dağılım oranı için ham veride $0,749\pm 0,11$, EG diskritizasyonu ile $0,635\pm 0,1$, EF diskritizasyonu ile $0,608\pm 0,11$ olarak elde edilen ortalama AUC değerleri, sınıf dağılım oranı 0,5 olduğunda sırasıyla; $0,852\pm 0,04$, $0,809\pm 0,05$, $0,803\pm 0,05$ olarak bulundu. Gözlem sayısı 1000 olduğunda, 0,1 sınıf dağılım oranı için ham veri ile $0,777\pm 0,08$, EG ile $0,704\pm 0,08$, EF ile $0,659\pm 0,08$ olarak bulunan ortalama AUC değerlerinin; sınıf dağılım oranı 0,5 olduğunda sırasıyla; $0,865\pm 0,03$, $0,837\pm 0,03$, $0,834\pm 0,03$ değerlerine yükseldiği görüldü. Gözlem sayısı 10000 olduğunda, 0,1 sınıf dağılım oranı için ortalama AUC değerleri; ham veri ve minimum-maksimum normalizasyonu ile $0,858\pm 0,02$, z-skor normalizasyonu ile $0,862\pm 0,03$, EG ile $0,833\pm 0,03$, EF ile $0,825\pm 0,03$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise sırasıyla $0,904\pm 0,01$, $0,898\pm 0,01$, $0,895\pm 0,01$ değerleri elde edildi.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama AUC değerlerine Tablo 23'te yer verilmektedir.

Tablo 24. Normal Dağılım için C5.0 ile elde edilen ortalama duyarlılık oranları

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	$0,349\pm 0,34$	$0,349\pm 0,34$	$0,349\pm 0,34$	$0,022\pm 0,11$	$0,012\pm 0,09$
	0,2	$0,528\pm 0,26$	$0,528\pm 0,26$	$0,528\pm 0,26$	$0,083\pm 0,18$	$0,087\pm 0,20$
	0,3	$0,613\pm 0,22$	$0,613\pm 0,22$	$0,613\pm 0,22$	$0,221\pm 0,27$	$0,189\pm 0,27$
	0,4	$0,690\pm 0,18$	$0,690\pm 0,18$	$0,690\pm 0,18$	$0,492\pm 0,22$	$0,504\pm 0,21$
	0,5	$0,745\pm 0,15$	$0,745\pm 0,15$	$0,745\pm 0,15$	$0,617\pm 0,19$	$0,597\pm 0,18$

Tablo 24. Devam Normal Dağılım için C5.0 ile elde edilen ortalama duyarlılık oranları

500	0,1	0,435±0,17	0,435±0,17	0,435±0,17	0,233±0,17	0,151±0,16
	0,2	0,594±0,12	0,594±0,12	0,594±0,12	0,467±0,14	0,469±0,15
	0,3	0,683±0,09	0,682±0,09	0,682±0,09	0,602±0,11	0,602±0,11
	0,4	0,743±0,08	0,743±0,08	0,743±0,08	0,690±0,09	0,690±0,10
	0,5	0,795±0,06	0,795±0,06	0,795±0,06	0,761±0,08	0,764±0,08
1000	0,1	0,477±0,12	0,477±0,12	0,477±0,12	0,326±0,13	0,224±0,13
	0,2	0,614±0,08	0,614±0,08	0,614±0,08	0,529±0,09	0,504±0,10
	0,3	0,699±0,07	0,699±0,07	0,699±0,07	0,647±0,08	0,647±0,08
	0,4	0,758±0,05	0,758±0,05	0,758±0,05	0,729±0,06	0,723±0,07
	0,5	0,803±0,05	0,803±0,05	0,803±0,05	0,789±0,05	0,785±0,06
10000	0,1	0,552±0,04	0,551±0,04	0,555±0,06	0,500±0,04	0,461±0,04
	0,2	0,669±0,03	0,669±0,03	0,669±0,03	0,636±0,03	0,624±0,03
	0,3	0,740±0,02	0,740±0,02	0,740±0,02	0,721±0,02	0,713±0,02
	0,4	0,792±0,02	0,792±0,02	0,792±0,02	0,783±0,02	0,779±0,02
	0,5	0,834±0,02	0,834±0,02	0,834±0,02	0,830±0,02	0,827±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Normal dağılımdan üretilen veriler ile elde edilen ortalama duyarlılık oranları incelendiğinde, 100 ve 1000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 24).

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama duyarlılık oranları sınıf dağılım oranı 0,1 olduğunda; ham veride 0,349±0,34, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,022±0,11, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,012±0,09 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama duyarlılık oranlarının; ham veride 0,528±0,26, EG diskritizasyonu ile 0,083±0,18, EF diskritizasyonu ile 0,087±0,2 olduğu görüldü. Sınıf dağılım oranı 0,3 iken bu oranlar; ham veri ile 0,613±0,22, EG diskritizasyonu ile 0,221±0,27, EF diskritizasyonu ile 0,189±0,27 olarak hesaplandı. Sınıf dağılım oranı 0,4 olduğunda; ham veride 0,69±0,18, EG diskritizasyonu ile 0,492±0,22, EF diskritizasyonu ile 0,504±0,21 değerleri bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda ortalama duyarlılık oranlarının; ham veride 0,745±0,15, EG diskritizasyonu ile 0,617±0,19, EF diskritizasyonu ile 0,597±0,18 olduğu görüldü.

Gözlem sayısı 500 olduğunda, sınıf dağılım oranı 0,1 için ham veri ile 0,435±0,17, EG diskritizasyonu ile 0,233±0,17, EF diskritizasyonu ile 0,151±0,16 olarak elde edilen ortalama

duyarlılık oranları, Sınıf dağılım oranı 0,5 olduğunda sırasıyla; 0,795±0,06, 0,761±0,08, 0,764±0,08 değerlerine yükseldiği görüldü.

Gözlem sayısı 1000 olduğunda, 0,1 sınıf dağılım oranı için ham veride 0,477±0,12, EG diskritizasyonu ile 0,326±0,13, EF diskritizasyonu ile 0,224±0,13 olarak elde edilen ortalama duyarlılık oranları, sınıf dağılım oranı 0,5 olduğunda sırasıyla; 0,803±0,05, 0,789±0,05, 0,785±0,06 olarak bulundu.

Gözlem sayısı 10000 olduğunda, 0,1 sınıf dağılım oranı için ortalama duyarlılık oranları; ham veride 0,552±0,04, minimum-maksimum normalizasyonu sonrasında 0,551±0,04, z-skor normalizasyonu sonrasında 0,555±0,06, EG diskritizasyonu ile 0,500±0,04, EF diskritizasyonu ile 0,461±0,04 olarak elde edildi. Sınıf dağılım oranı 0,5 olduğunda ortalama duyarlılık oranları; ham veri, minimum-maksimum ve z-skor normalizasyonu ile 0,834±0,02, EG ile 0,830±0,02, EF ile 0,827±0,01 olarak bulundu.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama duyarlılık oranlarına Tablo 24'te yer verilmektedir.

Tablo 25. Normal Dağılım için C5.0 ile elde edilen ortalama seçicilik oranları

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,945±0,06	0,945±0,06	0,945±0,06	0,996±0,02	0,998±0,01
	0,2	0,897±0,09	0,897±0,09	0,897±0,09	0,979±0,05	0,981±0,06
	0,3	0,857±0,11	0,857±0,11	0,857±0,11	0,940±0,10	0,945±0,10
	0,4	0,808±0,12	0,808±0,12	0,808±0,12	0,808±0,15	0,800±0,15
	0,5	0,762±0,15	0,762±0,15	0,762±0,15	0,713±0,18	0,709±0,18
500	0,1	0,961±0,02	0,961±0,02	0,961±0,02	0,980±0,02	0,984±0,02
	0,2	0,925±0,04	0,925±0,04	0,925±0,04	0,939±0,04	0,933±0,04
	0,3	0,888±0,05	0,888±0,05	0,888±0,05	0,896±0,05	0,896±0,05
	0,4	0,850±0,05	0,850±0,05	0,850±0,05	0,848±0,06	0,843±0,07
	0,5	0,803±0,07	0,803±0,07	0,803±0,07	0,794±0,08	0,780±0,08
1000	0,1	0,964±0,02	0,964±0,02	0,964±0,02	0,979±0,01	0,982±0,01
	0,2	0,933±0,02	0,933±0,02	0,933±0,02	0,945±0,02	0,938±0,03
	0,3	0,897±0,03	0,897±0,03	0,897±0,03	0,905±0,03	0,902±0,03
	0,4	0,860±0,04	0,860±0,04	0,860±0,04	0,863±0,04	0,861±0,04
	0,5	0,818±0,04	0,818±0,04	0,818±0,04	0,810±0,05	0,804±0,05
10000	0,1	0,975±0,006	0,975±0,006	0,975±0,01	0,980±0,004	0,978±0,005
	0,2	0,950±0,01	0,950±0,01	0,950±0,01	0,958±0,01	0,952±0,01
	0,3	0,923±0,01	0,923±0,01	0,923±0,01	0,929±0,01	0,924±0,01
	0,4	0,892±0,01	0,892±0,01	0,892±0,01	0,895±0,01	0,891±0,01
	0,5	0,857±0,01	0,857±0,01	0,857±0,01	0,857±0,01	0,854±0,02

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Normal dağılımdan üretilen veriler ile elde edilen ortalama seçicilik oranları incelendiğinde, 100, 500 ve 1000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 25).

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama seçicilik oranları, sınıf dağılım oranı 0,1 olduğunda; ham veride $0,945 \pm 0,06$, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile $0,996 \pm 0,02$, eşit frekansta aralıklandırma (EF) diskritizasyonu ile $0,998 \pm 0,01$ olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama seçicilik oranları; ham veride $0,897 \pm 0,09$ olarak bulunurken EF diskritizasyonu ile $0,981 \pm 0,06$ olarak elde edildi. Bu oranlar sırasıyla; 0,3 sınıf dağılım oranı için $0,857 \pm 0,11$, $0,94 \pm 0,1$, $0,945 \pm 0,1$, 0,4 dağılım oranı için $0,808 \pm 0,12$, $0,808 \pm 0,15$, $0,8 \pm 0,15$ olarak saptandı. Son olarak sınıf dağılım oranı 0,5 olduğunda sırasıyla; $0,762 \pm 0,15$, $0,713 \pm 0,18$, $0,709 \pm 0,18$ değerlerine ulaşıldı.

Gözlem sayısı 500 olduğunda, 0,1 sınıf dağılım oranı için elde edilen ortalama seçicilik oranları; ham veride $0,961 \pm 0,02$, EG diskritizasyonu ile $0,98 \pm 0,02$, EF diskritizasyonu ile $0,984 \pm 0,02$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise bu oranlar sırasıyla; $0,803 \pm 0,07$, $0,794 \pm 0,08$, $0,78 \pm 0,08$ olarak elde edildi. Gözlem sayısı 1000 olduğunda, sınıf dağılım oranı 0,1 için ham veride $0,964 \pm 0,02$, EG diskritizasyonu ile $0,979 \pm 0,01$, EF diskritizasyonu ile $0,982 \pm 0,01$ olarak elde edilen ortalama seçicilik oranları, sınıf dağılım oranı 0,5 olduğunda sırasıyla; $0,818 \pm 0,04$, $0,81 \pm 0,05$, $0,804 \pm 0,05$ değerlerine ulaşıldı. Gözlem sayısı 10000 olduğunda, 0,1 sınıf dağılım oranı için ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında $0,975 \pm 0,01$, EG ile $0,980 \pm 0,004$, EF ile $0,978 \pm 0,005$ olarak elde edilen ortalama seçicilik oranları, 0,5 sınıf dağılım oranı için sırasıyla; $0,857 \pm 0,01$, $0,857 \pm 0,01$, $0,854 \pm 0,022$ olarak belirlendi.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama seçicilik oranlarına Tablo 25’te yer verilmektedir.

Tablo 26. Normal Dağılım için C5.0 ile elde edilen ortalama pozitif kestirim değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	$0,471 \pm 0,38$	$0,471 \pm 0,38$	$0,471 \pm 0,38$	$0,417 \pm 0,44$	$0,365 \pm 0,35$
	0,2	$0,603 \pm 0,27$	$0,603 \pm 0,27$	$0,603 \pm 0,27$	$0,534 \pm 0,34$	$0,590 \pm 0,33$
	0,3	$0,678 \pm 0,20$	$0,678 \pm 0,20$	$0,678 \pm 0,20$	$0,658 \pm 0,27$	$0,641 \pm 0,25$
	0,4	$0,724 \pm 0,15$	$0,724 \pm 0,15$	$0,724 \pm 0,15$	$0,658 \pm 0,20$	$0,653 \pm 0,20$
	0,5	$0,773 \pm 0,12$	$0,773 \pm 0,12$	$0,773 \pm 0,12$	$0,702 \pm 0,15$	$0,694 \pm 0,15$

Tablo 26. Devam Normal Dağılım için C5.0 ile elde edilen ortalama pozitif kestirim değerleri

500	0,1	0,574±0,18	0,574±0,18	0,574±0,18	0,604±0,27	0,548±0,28
	0,2	0,676±0,11	0,676±0,11	0,676±0,11	0,677±0,14	0,654±0,13
	0,3	0,731±0,08	0,731±0,08	0,731±0,08	0,724±0,09	0,724±0,09
	0,4	0,772±0,06	0,772±0,06	0,772±0,06	0,758±0,07	0,753±0,07
	0,5	0,805±0,05	0,805±0,05	0,805±0,05	0,792±0,06	0,782±0,06
1000	0,1	0,611±0,13	0,611±0,13	0,611±0,13	0,652±0,16	0,605±0,20
	0,2	0,703±0,08	0,703±0,08	0,703±0,08	0,716±0,09	0,682±0,09
	0,3	0,748±0,05	0,748±0,05	0,748±0,05	0,75±0,06	0,743±0,06
	0,4	0,785±0,05	0,785±0,05	0,785±0,05	0,784±0,05	0,780±0,05
	0,5	0,817±0,04	0,817±0,04	0,817±0,04	0,808±0,04	0,803±0,04
10000	0,1	0,714±0,04	0,714±0,04	0,714±0,06	0,742±0,04	0,697±0,04
	0,2	0,771±0,03	0,771±0,03	0,771±0,03	0,790±0,02	0,765±0,03
	0,3	0,806±0,02	0,806±0,02	0,806±0,02	0,814±0,02	0,801±0,02
	0,4	0,831±0,02	0,831±0,02	0,831±0,02	0,833±0,02	0,827±0,02
	0,5	0,854±0,01	0,854±0,01	0,854±0,01	0,853±0,01	0,851±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Normal dağılımdan üretilen veriler ile elde edilen ortalama pozitif kestirim değerleri incelendiğinde, 100, 500 ve 1000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 26).

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama pozitif kestirim değerleri (PKD) sınıf dağılım oranı 0,1 olduğunda; ham veride 0,471±0,38, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,417±0,44, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,365±0,35 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama PKD'nin; ham veride 0,603±0,27 olarak bulurken EG diskritizasyonu ile 0,534±0,34, EF diskritizasyonu ile 0,59±0,33 değerlerine düştüğü görüldü. Sınıf dağılım oranı 0,3 iken bu değerler sırasıyla; 0,678±0,2, 0,658±0,27, 0,641±0,25 olarak hesaplandı. Sınıf dağılım oranı 0,4 olduğunda elde edilen ortalama değerlerin sırasıyla; 0,724±0,15, 0,658±0,2, 0,653±0,2 olduğu gözlemlendi. Son olarak sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama PKD için ham veriden 0,773±0,12, EG diskritizasyonu ile 0,702±0,15, EF diskritizasyonu ile 0,694±0,15 değerlerine ulaşıldı.

Gözlem sayısı 500 olduğunda, 0,1 sınıf dağılım oranı için ham veride 0,574±0,18, EG diskritizasyonu ile 0,604±0,27, EF diskritizasyonu ile 0,548±0,28 olarak bulunan ortalama

PKD, sınıf dağılım oranı 0,5 olduğunda sırasıyla; 0,805±0,05, 0,792±0,06, 0,782±0,06 değerlerine yükseldi.

Gözlem sayısı 1000 olduğunda, sınıf dağılım oranı 0,1 için ham veri ile 0,611±0,13, EG diskritizasyonu ile 0,652±0,16, EF diskritizasyonu ile 0,605±0,2 olarak elde edilen ortalama PKD, sınıf dağılım oranı 0,5 olduğunda sırasıyla; 0,817±0,04, 0,808±0,04, 0,803±0,04 olarak bulundu.

Gözlem sayısı 10000 olduğunda, 0,1 sınıf dağılım oranı için elde edilen ortalama PKD; ham veri ve minimum-maksimum normalizasyonu ile 0,714±0,04, z-skor yöntemi ile 0,714±0,06, EG diskritizasyonu ile 0,742±0,04, EF diskritizasyonu ile 0,697±0,04 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise elde edilen ortalama PKD; ham veri, minimum-maksimum normalizasyonu ve z-skor yöntemi ile 0,853±0,01, EG diskritizasyonu ile 0,853±0,01, EF diskritizasyonu ile 0,849±0,01 olarak elde edildi.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama pozitif kestirim değerlerine Tablo 26’da yer verilmektedir.

Tablo 27. Normal Dağılım için C5.0 ile elde edilen ortalama negatif kestirim değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,930±0,04	0,930±0,04	0,930±0,04	0,902±0,01	0,901±0,01
	0,2	0,887±0,06	0,887±0,06	0,887±0,06	0,812±0,03	0,814±0,04
	0,3	0,844±0,08	0,844±0,08	0,844±0,08	0,745±0,07	0,739±0,07
	0,4	0,805±0,10	0,805±0,10	0,805±0,10	0,713±0,09	0,715±0,09
	0,5	0,763±0,12	0,763±0,12	0,763±0,12	0,663±0,12	0,647±0,12
500	0,1	0,939±0,02	0,939±0,02	0,939±0,02	0,920±0,02	0,913±0,01
	0,2	0,902±0,03	0,902±0,03	0,902±0,03	0,877±0,03	0,877±0,03
	0,3	0,868±0,03	0,868±0,03	0,868±0,03	0,842±0,04	0,842±0,04
	0,4	0,834±0,04	0,834±0,04	0,834±0,04	0,807±0,04	0,806±0,05
	0,5	0,799±0,05	0,799±0,05	0,799±0,05	0,773±0,05	0,772±0,06
1000	0,1	0,943±0,01	0,943±0,01	0,943±0,01	0,929±0,01	0,919±0,01
	0,2	0,907±0,02	0,907±0,02	0,907±0,02	0,890±0,02	0,884±0,02
	0,3	0,875±0,02	0,875±0,02	0,875±0,02	0,858±0,03	0,857±0,03
	0,4	0,843±0,03	0,843±0,03	0,843±0,03	0,828±0,03	0,825±0,03
	0,5	0,808±0,04	0,808±0,04	0,808±0,04	0,795±0,04	0,792±0,04
10000	0,1	0,951±0,004	0,951±0,004	0,952±0,01	0,946±0,004	0,942±0,004
	0,2	0,920±0,01	0,920±0,01	0,920±0,01	0,913±0,01	0,910±0,01
	0,3	0,892±0,01	0,892±0,01	0,892±0,01	0,886±0,01	0,882±0,01
	0,4	0,866±0,01	0,866±0,01	0,866±0,01	0,861±0,01	0,858±0,01
	0,5	0,838±0,01	0,838±0,01	0,838±0,01	0,835±0,01	0,832±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Normal dağılımdan üretilen veriler ile elde edilen ortalama negatif kestirim değerleri incelendiğinde, 100, 500 ve 1000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 27).

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama negatif kestirim değerleri (NKD) sınıf dağılım oranı 0,1 olduğunda; ham veriden $0,93\pm 0,04$, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile $0,902\pm 0,01$, eşit frekansta aralıklandırma (EF) diskritizasyonu ile $0,901\pm 0,01$ olarak elde edildi. Sınıf dağılım oranı 0,2 olduğunda bu ortalama değerlerin sırasıyla; $0,887\pm 0,06$, $0,812\pm 0,03$, $0,814\pm 0,04$ olduğu görüldü. Sınıf dağılım oranı 0,3 iken ortalama NKD sırasıyla; $0,844\pm 0,08$, $0,745\pm 0,07$, $0,739\pm 0,07$, sınıf dağılım oranı 0,4 olduğunda $0,805\pm 0,1$, $0,713\pm 0,09$, $0,715\pm 0,09$ değerlerini almaktadır. Sınıf dağılım oranı 0,5 olduğunda ise ortalama NKD için sırasıyla; $0,763\pm 0,12$, $0,663\pm 0,12$, $0,647\pm 0,12$ değerlerine ulaşıldı.

Gözlem sayısının 500 olduğu sınıflandırma çalışmasında, 0,1 sınıf dağılım oranında ham veriden $0,939\pm 0,02$ olarak elde edilen ortalama negatif kestirim değerleri, EG diskritizasyonu ile $0,92\pm 0,02$, EF diskritizasyonu ile $0,913\pm 0,01$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise bu değerler sırasıyla; $0,799\pm 0,05$, $0,773\pm 0,05$, $0,772\pm 0,06$ değerlerine düştü.

Gözlem sayısı 1000 olduğunda, sınıf dağılım oranı 0,1 iken ortalama NKD; ham veride $0,943\pm 0,01$, EG ile $0,929\pm 0,01$, EF ile $0,919\pm 0,01$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise bu değerler sırasıyla; $0,808\pm 0,04$, $0,795\pm 0,04$, $0,792\pm 0,04$ değerlerine düştü.

Gözlem sayısı 10000 olduğunda, sınıf dağılım oranı 0,1 iken elde edilen ortalama NKD; ham veride $0,951\pm 0,004$, z-skor normalizasyonu ile $0,952\pm 0,01$, EG ile $0,946\pm 0,004$, EF ile $0,942\pm 0,004$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise ortalama NKD; ham veri, minimum-maksimum ve z-skor normalizasyonu ile $0,838\pm 0,01$, EG ile $0,835\pm 0,01$, EF ile $0,832\pm 0,01$ olarak elde edildi.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama negatif kestirim değerlerine Tablo 27'de yer verilmektedir.

Ki-kare dağılımından üretilen verilerin sınıflandırılması

Tablo 28. Ki-kare dağılımı için C5.0 ile elde edilen ortalama doğruluk oranları (%)

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	60,261±16,26	60,261±16,26	60,261±16,26	50,283±3,51	50,147±2,36
	0,2	65,822±13,53	65,822±13,53	65,822±13,53	51,522±6,17	51,453±5,98
	0,3	68,720±11,31	68,720±11,31	68,720±11,31	55,789±10,00	54,373±9,08
	0,4	70,156±10,63	70,156±10,63	70,156±10,63	64,090±11,37	62,392±11,59
	0,5	70,530±9,97	70,530±9,97	70,530±9,97	65,320±11,08	64,230±11,11
500	0,1	64,524±7,65	64,524±7,65	64,524±7,65	52,714±5,37	52,488±5,34
	0,2	70,301±5,89	70,301±5,89	70,301±5,89	63,554±6,46	63,603±6,55
	0,3	73,097±4,98	73,097±4,98	73,097±4,98	69,141±5,55	69,149±5,54
	0,4	74,699±4,59	74,699±4,59	74,699±4,59	71,820±4,97	71,540±4,73
	0,5	74,932±4,61	74,932±4,61	74,932±4,61	72,707±4,81	72,695±4,69
1000	0,1	65,878±5,90	65,878±5,90	65,878±5,90	55,666±5,81	55,337±6,12
	0,2	71,443±4,50	71,443±4,50	71,443±4,50	65,976±4,74	65,999±4,94
	0,3	74,315±3,63	74,314±3,63	74,314±3,63	71,247±3,85	71,526±4,16
	0,4	75,776±3,22	75,776±3,22	75,776±3,22	74,157±3,44	73,965±3,54
	0,5	76,165±3,06	76,165±3,06	76,165±3,06	74,779±3,10	75,010±3,29
10000	0,1	68,236±2,17	68,236±2,17	68,237±2,16	63,523±2,06	64,559±2,12
	0,2	74,355±1,56	74,357±1,56	74,357±1,56	71,708±1,50	72,391±1,57
	0,3	77,342±1,18	77,341±1,18	77,342±1,18	75,932±1,21	76,404±1,17
	0,4	78,904±1,03	78,906±1,03	78,906±1,03	78,229±0,96	78,419±1,01
	0,5	79,347±0,93	79,346±0,93	79,346±0,93	78,826±0,89	79,014±0,87

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Ki-kare dağılımından üretilen veriler ile elde edilen ortalama doğruluk oranları incelendiğinde, 100 ve 500 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 28).

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama doğruluk oranları sınıf dağılım oranı 0,1 olduğunda; ham veride %60,261±16,26, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile %50,283±3,51, eşit frekansta aralıklandırma (EF) diskritizasyonu ile %50,147±2,36 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama doğruluk oranlarının; ham veride %65,822±13,53, olarak bulunurken EG diskritizasyonu ile %51,522±6,17, EF diskritizasyonu ile %51,453±5,98 değerleri elde edildi. Bu oranları sırasıyla; 0,3 sınıf dağılımı için %68,720±11,31, %55,789±10,00, %54,373±9,08, 0,4 sınıf dağılımı için %70,156±10,63, %64,090±11,37, %62,392±11,59 olarak hesaplandı. Sınıf dağılım oranı 0,5 olduğunda ise bu oranların; ham veride %70,53±9,97, EG diskritizasyonu ile %65,32±11,08, EF diskritizasyonu ile %64,23±11,11 olduğu sonucuna ulaşıldı.

Gözlem sayısının 500 olduğu çalışmada, 0,1 sınıf dağılım oranı için elde edilen ortalama doğruluk oranları; ham veride %64,524±7,65, EG diskritizasyonu ile %52,714±5,37, EF diskritizasyonu ile %52,488±5,34 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ham veriden %74,932±4,61 olarak bulunan ortalama doğruluk oranları, EG diskritizasyonu ile %72,707±4,81, EF diskritizasyonu ile %72,695±4,69 oranlarına düştüğü görüldü.

Gözlem sayısı 1000 olduğunda gerçekleştirilen sınıflandırma çalışmasında sınıf dağılım oranı 0,1 iken elde edilen ortalama doğruluk oranları; ham veri, minimum-maksimum ve z-skor normalizasyonu ile %65,878±5,90, EG diskritizasyonu ile %55,666±5,81, EF diskritizasyonu ile %55,337±6,12 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ortalama doğruluk oranları sırasıyla; %76,165±3,06, %74,779±3,1, %75,01±3,29 olarak elde edildi.

Gözlem sayısı 10000 olduğunda yapılan sınıflandırma çalışmasında sınıf dağılım oranı 0,1 iken elde edilen ortalama doğruluk oranları; ham veride ve minimum-maksimum normalizasyonu sonrasında %68,236±2,17, z-skor normalizasyonu sonrasında %68,237±2,16, EG diskritizasyonu ile %63,523±2,06 ve EF diskritizasyonu ile %64,559±2,12 olarak elde edildi. Sınıf dağılım oranı 0,5 olduğunda ise ortalama doğruluk oranları sırasıyla; %79,347±0,93, %79,346±0,93, %78,826±0,89, %79,014±0,87 olarak bulundu.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama doğruluk oranlarına Tablo 28’de yer verilmektedir.

Tablo 29. Ki-kare dağılımı için C5.0 ile elde edilen ortalama AUC değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,644±0,20	0,644±0,20	0,644±0,20	0,505±0,04	0,502±0,03
	0,2	0,698±0,16	0,698±0,16	0,698±0,16	0,516±0,07	0,517±0,07
	0,3	0,731±0,13	0,731±0,13	0,731±0,13	0,560±0,11	0,548±0,10
	0,4	0,738±0,12	0,738±0,12	0,738±0,12	0,641±0,12	0,617±0,13
	0,5	0,744±0,11	0,744±0,11	0,744±0,11	0,647±0,12	0,632±0,12
500	0,1	0,703±0,11	0,703±0,11	0,703±0,11	0,546±0,08	0,544±0,08
	0,2	0,757±0,08	0,757±0,08	0,757±0,08	0,688±0,09	0,677±0,08
	0,3	0,783±0,06	0,783±0,06	0,783±0,06	0,726±0,07	0,727±0,06
	0,4	0,796±0,05	0,796±0,05	0,796±0,05	0,742±0,06	0,743±0,05
	0,5	0,801±0,05	0,801±0,05	0,801±0,05	0,746±0,05	0,751±0,05
1000	0,1	0,720±0,08	0,720±0,08	0,720±0,08	0,607±0,10	0,591±0,10
	0,2	0,773±0,05	0,773±0,05	0,773±0,05	0,721±0,06	0,723±0,06
	0,3	0,799±0,04	0,799±0,04	0,799±0,04	0,758±0,05	0,762±0,05
	0,4	0,811±0,04	0,811±0,04	0,811±0,04	0,776±0,04	0,780±0,04
	0,5	0,813±0,03	0,813±0,03	0,813±0,03	0,780±0,04	0,786±0,04

Tablo 29. Devam Ki-kare dağılımı için C5.0 ile elde edilen ortalama AUC değerleri

10000	0,1	0,806±0,03	0,806±0,03	0,806±0,03	0,762±0,04	0,770±0,03
	0,2	0,833±0,02	0,833±0,02	0,833±0,02	0,809±0,02	0,814±0,02
	0,3	0,844±0,01	0,844±0,01	0,844±0,01	0,827±0,01	0,832±0,01
	0,4	0,851±0,01	0,851±0,01	0,851±0,01	0,838±0,01	0,842±0,01
	0,5	0,853±0,01	0,853±0,01	0,853±0,01	0,840±0,01	0,845±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Ki-kare dağılımından üretilen veriler ile elde edilen ortalama AUC değerleri incelendiğinde, 100, 500 ve 1000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 29).

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama AUC değerleri, sınıf dağılım oranı 0,1 olduğunda; ham veride 0,644±0,2, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,505±0,04, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,502±0,03 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama AUC değerlerinin; ham veride 0,698±0,16 olarak bulunurken EG diskritizasyonu ile 0,516±0,07, EF diskritizasyonu ile 0,517±0,07 değerlerine yükseldiği görüldü. Sınıf dağılım oranı 0,3 için bu değerler; ham veri ile 0,731±0,13, EG ile 0,560±0,11, EF ile 0,548±0,1 olarak hesaplandı. Sınıf dağılım oranı 0,4 olduğunda ise bu değerler sırasıyla; 0,738±0,12, 0,641±0,12, 0,617±0,13 olarak gözlemlendi. Son olarak sınıf dağılım oranı 0,5 olduğunda ise ortalama AUC değerleri için sırasıyla; 0,744±0,11, 0,647±0,12, 0,632±0,12 bulgularına ulaşıldı.

Gözlem sayısı 500 olduğunda, sınıf dağılım oranı 0,1 için elde edilen ortalama AUC değerleri; ham veride 0,703±0,11, EG diskritizasyonu ile 0,546±0,08, EF diskritizasyonu ile 0,544±0,08 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise elde edilen ortalama AUC değerleri sırasıyla; 0,801±0,05, 0,746±0,05, 0,751±0,05 olarak elde edildi. Gözlem sayısı 1000 ve sınıf dağılım oranı 0,1 olduğunda ortalama AUC değerleri için ham veriden 0,720±0,08, EG diskritizasyonu ile 0,607±0,10, EF diskritizasyonu ile 0,591±0,10, bulgularına ulaşıldı. Sınıf dağılımı 0,5 oranına yükseldiğinde ortalama AUC değerleri de sırasıyla; 0,813±0,03, 0,78±0,04, 0,786±0,04 değerlerine yükseldi. Gözlem sayısı 10000 olduğunda, sınıf dağılım oranı 0,1 iken ortalama AUC değerleri; ham veri ile 0,806±0,03, EG ile 0,762±0,04, EF ile 0,770±0,03 olarak bulunurken 0,5 sınıf dağılım oranı olduğunda bu değerler sırasıyla; 0,853±0,01, 0,840±0,01, 0,845±0,01 değerlerine yükseldi.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama AUC değerlerine Tablo 29’da yer verilmektedir.

Tablo 30. Ki-kare dağılımı için C5.0 ile elde edilen ortalama duyarlılık oranları

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,264±0,33	0,264±0,33	0,264±0,33	0,008±0,07	0,004±0,05
	0,2	0,437±0,26	0,437±0,26	0,437±0,26	0,043±0,15	0,041±0,15
	0,3	0,546±0,21	0,546±0,21	0,546±0,21	0,183±0,26	0,143±0,25
	0,4	0,626±0,19	0,626±0,19	0,626±0,19	0,501±0,23	0,477±0,23
	0,5	0,704±0,15	0,704±0,15	0,704±0,15	0,635±0,19	0,620±0,19
500	0,1	0,336±0,16	0,336±0,16	0,336±0,16	0,064±0,12	0,059±0,12
	0,2	0,497±0,12	0,497±0,12	0,497±0,12	0,33±0,14	0,332±0,14
	0,3	0,602±0,10	0,602±0,10	0,602±0,10	0,498±0,12	0,505±0,12
	0,4	0,681±0,08	0,681±0,08	0,681±0,08	0,618±0,10	0,620±0,10
	0,5	0,748±0,07	0,748±0,07	0,748±0,07	0,717±0,09	0,716±0,09
1000	0,1	0,360±0,12	0,360±0,12	0,360±0,12	0,127±0,13	0,119±0,13
	0,2	0,514±0,09	0,514±0,09	0,514±0,09	0,377±0,1	0,379±0,11
	0,3	0,617±0,07	0,617±0,07	0,617±0,07	0,535±0,08	0,543±0,09
	0,4	0,694±0,06	0,694±0,06	0,694±0,06	0,655±0,07	0,651±0,08
	0,5	0,757±0,05	0,757±0,05	0,757±0,05	0,74±0,06	0,743±0,06
10000	0,1	0,399±0,04	0,399±0,04	0,399±0,04	0,293±0,05	0,318±0,05
	0,2	0,555±0,04	0,555±0,04	0,555±0,04	0,494±0,03	0,505±0,03
	0,3	0,655±0,03	0,655±0,03	0,655±0,03	0,617±0,03	0,627±0,03
	0,4	0,729±0,02	0,729±0,02	0,729±0,02	0,714±0,02	0,714±0,02
	0,5	0,791±0,02	0,791±0,02	0,791±0,02	0,788±0,02	0,785±0,02

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Ki-kare dağılımından üretilen veriler ile elde edilen ortalama duyarlılık oranları incelendiğinde, 100, 500 ve 1000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 30).

Sınıflandırma çalışmasında 100 gözlem için ortalama duyarlılık oranları, sınıf dağılım oranı 0,1 olduğunda; ham veride 0,264±0,33 olup eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,008±0,07, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,004±0,05 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama duyarlılık oranlarının; ham veride 0,437±0,26, EG diskritizasyonu ile 0,043±0,15, EF diskritizasyonu ile 0,041±0,15 olduğu görüldü. Sınıf dağılım oranı 0,3 iken bu oranlar sırasıyla; 0,546±0,21, 0,183±0,26, 0,143±0,25 olarak bulunurken 0,4 oranında; 0,626±0,19, 0,501±0,23, 0,477±0,23 olarak hesaplandı. Son olarak sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama duyarlılık oranlarının sırasıyla; 0,704±0,15, 0,635±0,19, 0,62±0,19 değerlerine yükseldiği bulundu.

Gözlem sayısı 500 iken gerçekleştirilen sınıflandırma çalışmasında, 0,1 sınıf dağılım oranı için elde edilen ortalama duyarlılık oranları; ham veri ile $0,336\pm 0,16$, EG ile $0,064\pm 0,12$, EF ile $0,059\pm 0,12$ olarak bulunurken sınıf dağılım oranı 0,5 olduğunda ortalama duyarlılık oranları sırasıyla; $0,748\pm 0,07$, $0,717\pm 0,09$, $0,716\pm 0,09$ olarak belirlendi.

Gözlem sayısı 1000 olduğunda gerçekleştirilen sınıflandırma çalışmasında sınıf dağılım oranı 0,1 iken elde edilen ortalama duyarlılık oranları; ham veride $0,360\pm 0,12$, EG diskritizasyonu ile $0,127\pm 0,13$, EF diskritizasyonu ile $0,119\pm 0,13$ olarak bulundu. Bu oranlar, 0,5 sınıf dağılım oranında sırasıyla; $0,757\pm 0,05$, $0,74\pm 0,06$, $0,743\pm 0,06$ değerlerine yükseldi.

Gözlem sayısı 10000 olduğunda yapılan sınıflandırma çalışmasında sınıf dağılım oranı 0,1 iken elde edilen ortalama duyarlılık oranları; ham veride, $0,399\pm 0,04$, EG ile $0,293\pm 0,054$, EF ile $0,318\pm 0,05$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ortalama duyarlılık oranları sırasıyla; $0,791\pm 0,02$, $0,788\pm 0,02$, $0,785\pm 0,02$ olarak elde edildi.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama duyarlılık oranlarına Tablo 30'da yer verilmektedir.

Tablo 31. Ki-kare dağılımı için C5.0 ile elde edilen ortalama seçicilik oranları

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	$0,941\pm 0,06$	$0,941\pm 0,06$	$0,941\pm 0,06$	$0,998\pm 0,02$	$0,999\pm 0,01$
	0,2	$0,880\pm 0,09$	$0,880\pm 0,09$	$0,880\pm 0,09$	$0,987\pm 0,05$	$0,988\pm 0,05$
	0,3	$0,829\pm 0,11$	$0,829\pm 0,11$	$0,829\pm 0,11$	$0,933\pm 0,11$	$0,944\pm 0,11$
	0,4	$0,777\pm 0,13$	$0,777\pm 0,13$	$0,777\pm 0,13$	$0,781\pm 0,15$	$0,771\pm 0,16$
	0,5	$0,707\pm 0,16$	$0,707\pm 0,16$	$0,707\pm 0,16$	$0,671\pm 0,18$	$0,665\pm 0,18$
500	0,1	$0,954\pm 0,03$	$0,954\pm 0,03$	$0,954\pm 0,03$	$0,990\pm 0,02$	$0,991\pm 0,02$
	0,2	$0,909\pm 0,04$	$0,909\pm 0,04$	$0,909\pm 0,04$	$0,941\pm 0,04$	$0,940\pm 0,04$
	0,3	$0,860\pm 0,05$	$0,860\pm 0,05$	$0,860\pm 0,05$	$0,885\pm 0,05$	$0,878\pm 0,06$
	0,4	$0,813\pm 0,06$	$0,813\pm 0,06$	$0,813\pm 0,06$	$0,819\pm 0,07$	$0,811\pm 0,07$
	0,5	$0,751\pm 0,07$	$0,751\pm 0,07$	$0,751\pm 0,07$	$0,737\pm 0,09$	$0,738\pm 0,09$
1000	0,1	$0,958\pm 0,02$	$0,958\pm 0,02$	$0,958\pm 0,02$	$0,986\pm 0,02$	$0,987\pm 0,02$
	0,2	$0,915\pm 0,03$	$0,915\pm 0,03$	$0,915\pm 0,03$	$0,943\pm 0,03$	$0,941\pm 0,03$
	0,3	$0,869\pm 0,04$	$0,869\pm 0,04$	$0,869\pm 0,04$	$0,890\pm 0,04$	$0,887\pm 0,04$
	0,4	$0,821\pm 0,04$	$0,821\pm 0,04$	$0,821\pm 0,04$	$0,828\pm 0,05$	$0,829\pm 0,05$
	0,5	$0,766\pm 0,05$	$0,766\pm 0,05$	$0,766\pm 0,05$	$0,755\pm 0,06$	$0,757\pm 0,06$
10000	0,1	$0,970\pm 0,01$	$0,970\pm 0,01$	$0,970\pm 0,01$	$0,980\pm 0,005$	$0,978\pm 0,01$
	0,2	$0,934\pm 0,01$	$0,934\pm 0,01$	$0,934\pm 0,01$	$0,944\pm 0,01$	$0,942\pm 0,01$
	0,3	$0,892\pm 0,01$	$0,892\pm 0,01$	$0,892\pm 0,01$	$0,901\pm 0,01$	$0,901\pm 0,01$
	0,4	$0,849\pm 0,02$	$0,849\pm 0,02$	$0,849\pm 0,02$	$0,851\pm 0,02$	$0,854\pm 0,01$
	0,5	$0,796\pm 0,02$	$0,796\pm 0,02$	$0,796\pm 0,02$	$0,789\pm 0,02$	$0,795\pm 0,02$

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Ki-kare dağılımından üretilen veriler ile elde edilen ortalama seçicilik oranları incelendiğinde, 100, 500 ve 1000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 31).

Sınıflandırma çalışmasında 100 gözlem için elde edilen ortalama seçicilik oranları, sınıf dağılım oranı 0,1 olduğunda; ham veride $0,941 \pm 0,06$, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile $0,998 \pm 0,02$, eşit frekansta aralıklandırma (EF) diskritizasyonu ile $0,999 \pm 0,01$ olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama seçicilik oranları ham veride $0,88 \pm 0,09$ olarak bulunurken EG diskritizasyonu ile $0,987 \pm 0,05$, EF diskritizasyonu ile $0,988 \pm 0,05$ değerlerine yükseldi. Sınıf dağılım oranı 0,3 iken ham veride $0,829 \pm 0,11$ olarak bulunan bu oranlara göre EG ve EF diskritizasyonu sonrasında sırasıyla elde edilen $0,933 \pm 0,11$, $0,944 \pm 0,11$ oranları ile artış sağlandı. Sınıf dağılım oranı 0,4 iken bu oranlar sırasıyla; $0,777 \pm 0,13$, $0,781 \pm 0,15$, $0,771 \pm 0,16$ olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda; ham veride $0,707 \pm 0,16$, EG diskritizasyonu ile $0,671 \pm 0,18$, EF diskritizasyonu ile $0,665 \pm 0,18$ ortalama seçicilik oranlarına ulaşıldı.

Gözlem sayısı 500 olduğunda, 0,1 sınıf dağılım oranı için ortalama seçicilik oranları; ham veride $0,954 \pm 0,03$, EG diskritizasyonu ile $0,99 \pm 0,02$, EF diskritizasyonu ile $0,991 \pm 0,02$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ortalama seçicilik oranları sırasıyla; $0,751 \pm 0,07$, $0,737 \pm 0,09$, $0,738 \pm 0,09$ olarak belirlendi. Gözlem sayısı 1000 olduğunda, sınıf dağılım oranı 0,1 iken ortalama seçicilik oranları; ham veride $0,958 \pm 0,02$, EG diskritizasyonu ile $0,986 \pm 0,02$, EF diskritizasyonu ile $0,987 \pm 0,02$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda bu oranlar sırasıyla; $0,766 \pm 0,05$, EG diskritizasyonu ile $0,755 \pm 0,06$, EF diskritizasyonu ile $0,757 \pm 0,06$ olarak elde edildi.

Gözlem sayısı 10000 olduğunda yapılan sınıflandırma çalışmasında, sınıf dağılım oranı 0,1 iken elde edilen ortalama seçicilik oranları; ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında $0,970 \pm 0,01$, EG diskritizasyonu ile $0,98 \pm 0,005$, EF diskritizasyonu ile $0,978 \pm 0,01$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ortalama seçicilik oranları; ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında $0,796 \pm 0,02$, EG diskritizasyonu ile $0,789 \pm 0,02$, EF diskritizasyonu ile $0,795 \pm 0,02$ olarak elde edildi.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama seçicilik oranlarına Tablo 31’de yer verilmektedir.

Tablo 32. Ki-kare dağılımı için C5.0 ile elde edilen ortalama pozitif kestirim değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,362±0,38	0,362±0,38	0,362±0,38	0,290±0,37	0,313±0,41
	0,2	0,509±0,28	0,509±0,28	0,509±0,28	0,499±0,32	0,516±0,32
	0,3	0,602±0,20	0,602±0,20	0,602±0,20	0,565±0,24	0,565±0,24
	0,4	0,669±0,15	0,669±0,15	0,669±0,15	0,625±0,18	0,599±0,19
	0,5	0,718±0,12	0,718±0,12	0,718±0,12	0,672±0,14	0,661±0,14
500	0,1	0,470±0,19	0,470±0,19	0,470±0,19	0,459±0,30	0,439±0,27
	0,2	0,588±0,11	0,588±0,11	0,588±0,11	0,599±0,16	0,604±0,16
	0,3	0,655±0,08	0,655±0,08	0,655±0,08	0,662±0,10	0,653±0,11
	0,4	0,713±0,07	0,713±0,07	0,713±0,07	0,701±0,07	0,695±0,08
	0,5	0,753±0,05	0,753±0,05	0,753±0,05	0,736±0,06	0,738±0,06
1000	0,1	0,500±0,14	0,500±0,14	0,500±0,14	0,524±0,21	0,520±0,2
	0,2	0,608±0,08	0,608±0,08	0,608±0,08	0,634±0,11	0,628±0,10
	0,3	0,673±0,06	0,673±0,06	0,673±0,06	0,682±0,07	0,680±0,07
	0,4	0,724±0,05	0,724±0,05	0,724±0,05	0,721±0,05	0,721±0,05
	0,5	0,766±0,04	0,766±0,04	0,766±0,04	0,754±0,04	0,757±0,04
10000	0,1	0,599±0,05	0,599±0,05	0,599±0,05	0,626±0,05	0,618±0,05
	0,2	0,677±0,03	0,677±0,03	0,677±0,03	0,688±0,03	0,687±0,03
	0,3	0,722±0,02	0,722±0,02	0,722±0,02	0,728±0,02	0,731±0,02
	0,4	0,763±0,02	0,763±0,02	0,763±0,02	0,762±0,02	0,765±0,02
	0,5	0,795±0,01	0,795±0,01	0,795±0,01	0,789±0,01	0,793±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Ki-kare dağılımından üretilen veriler ile elde edilen ortalama pozitif kestirim değerleri incelendiğinde, 100, 500 ve 1000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 32).

Sınıflandırma çalışmasında 100 gözlem için elde edilen ortalama pozitif kestirim değeri (PKD), sınıf dağılımı 0,1 oranındayken ham veri ile 0,362±0,38, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,29±0,37, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,313±0,41 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama PKD'nin; ham veri ile 0,509±0,28 olarak bulunurken EG ile 0,499±0,32 ve EF ile 0,516±0,32 olduğu görüldü. Bu değer, sınıf dağılım oranı 0,3 için ham veri ile 0,602±0,2 olup EG ve EF diskritizasyonu ile 0,565±0,24, olarak hesaplandı. Sınıf dağılım oranı 0,4 iken elde edilen ortalama pozitif kestirim değerlerinin; ham veride 0,669±0,15, EG ile 0,625±0,18, EF ile

0,599±0,19 olduğu gözlemlendi. Son olarak sınıf dağılım oranı 0,5 olduğunda, ortalama PKD sırasıyla; 0,718±0,12, 0,672±0,14, 0,661±0,14 olarak bulundu.

Gözlem sayısı 500 ve sınıf dağılım oranı 0,1 olduğunda ortalama PKD için ham veride 0,47±0,19, EG diskritizasyonu ile 0,459±0,3, EF diskritizasyonu ile 0,439±0,27 değerleri bulundu. Sınıf dağılımı tam dengeli (0,5) olduğunda ortalama PKD; ham veride 0,753±0,05, EG diskritizasyonu ile 0,736±0,06, EF diskritizasyonu ile 0,738±0,06 değerlerinde bulundu. Gözlem sayısı 1000 olduğunda, sınıf dağılım oranı 0,1 iken elde edilen ortalama PKD; ham veride 0,500±0,14 olarak bulunurken EG ve EF diskritizasyonu ile sırasıyla; 0,524±0,21, 0,520±0,2 değerine yükseldi. Sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama PKD; ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında 0,766±0,04, EG diskritizasyonu ile 0,754±0,04, EF diskritizasyonu ile 0,757±0,04 olarak bulundu. Gözlem sayısı 10000 olduğunda yapılan çalışmada, 0,1 sınıf dağılım oranı için ortalama PKD; ham veride ve minimum-maksimum normalizasyonu sonrasında 0,597±0,04, z-skor normalizasyonu sonrasında 0,599±0,05, EG diskritizasyonu ile 0,626±0,05, EF diskritizasyonu ile 0,618±0,05 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ortalama PKD; ham veri, minimum-maksimum normalizasyonu ve z-skor normalizasyonu ile 0,795±0,01, EG ile 0,789±0,01, EF ile 0,793±0,01 olarak elde edildi.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama pozitif kestirim değerlerine Tablo 32’de yer verilmektedir.

Tablo 33. Ki-kare dağılımı için C5.0 ile elde edilen ortalama negatif kestirim değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,921±0,03	0,921±0,03	0,921±0,03	0,901±0,01	0,900±0,005
	0,2	0,865±0,06	0,865±0,06	0,865±0,06	0,806±0,03	0,806±0,02
	0,3	0,815±0,07	0,815±0,07	0,815±0,07	0,734±0,06	0,727±0,06
	0,4	0,767±0,10	0,767±0,10	0,767±0,10	0,712±0,10	0,698±0,10
	0,5	0,716±0,12	0,716±0,12	0,716±0,12	0,663±0,13	0,651±0,13
500	0,1	0,929±0,02	0,929±0,02	0,929±0,02	0,905±0,01	0,905±0,01
	0,2	0,879±0,02	0,879±0,02	0,879±0,02	0,850±0,03	0,850±0,03
	0,3	0,836±0,03	0,836±0,03	0,836±0,03	0,806±0,04	0,807±0,04
	0,4	0,795±0,04	0,795±0,04	0,795±0,04	0,765±0,04	0,765±0,05
	0,5	0,752±0,05	0,752±0,05	0,752±0,05	0,727±0,06	0,728±0,06
1000	0,1	0,931±0,01	0,931±0,01	0,931±0,01	0,911±0,01	0,910±0,01
	0,2	0,883±0,02	0,883±0,02	0,883±0,02	0,859±0,02	0,859±0,02
	0,3	0,842±0,02	0,842±0,02	0,842±0,02	0,818±0,03	0,820±0,03
	0,4	0,802±0,03	0,802±0,03	0,802±0,03	0,784±0,03	0,783±0,03
	0,5	0,761±0,04	0,761±0,04	0,761±0,04	0,746±0,04	0,749±0,04

Tablo 33. Devam Ki-kare dağılımı için C5.0 ile elde edilen ortalama negatif kestirim değerleri

10000	0,1	0,936±0,004	0,936±0,004	0,936±0,004	0,926±0,004	0,928±0,004
	0,2	0,894±0,01	0,894±0,01	0,894±0,01	0,882±0,01	0,884±0,01
	0,3	0,858±0,01	0,858±0,01	0,858±0,01	0,846±0,01	0,849±0,01
	0,4	0,825±0,01	0,825±0,01	0,825±0,01	0,817±0,01	0,818±0,01
	0,5	0,792±0,01	0,792±0,01	0,792±0,01	0,788±0,01	0,788±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Ki-kare dağılımından üretilen veriler ile elde edilen ortalama negatif kestirim değerleri incelendiğinde, 100, 500 ve 1000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 33).

Sınıflandırma çalışmasında 100 gözlem için elde edilen ortalama negatif kestirim değeri (NKD), sınıf dağılım oranı 0,1 olduğunda; ham veride 0,921±0,03, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,901±0,01, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,9±0,005 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda bu ortalama değerlerin sırasıyla; 0,865±0,06, 0,806±0,03, 0,806±0,02 olduğu görüldü. Sınıf dağılım oranı 0,3 iken ortalama NKD; ham veri ile 0,815±0,07, EG 0,734±0,06, EF ile 0,727±0,06 olarak hesaplandı. Sınıf dağılım oranı 0,4 iken ortalama NKD için ham veri ile 0,767±0,1, EG ile 0,712±0,1, EF ile 0,698±0,1 değerleri elde edildi. Son olarak sınıf dağılım oranı 0,5 olduğunda bu değerler sırasıyla; 0,716±0,12, 0,663±0,13, 0,651±0,13 değerlerine düştü.

Gözlem sayısı 500 olduğunda yapılan çalışmada, sınıf dağılım oranı 0,1 için ortalama negatif kestirim değerleri; ham veri ile 0,929±0,02, EG diskritizasyonu ile 0,905±0,01, EF diskritizasyonu ile 0,905±0,01 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise bu değerler sırasıyla; 0,752±0,05, 0,727±0,06, 0,728±0,06 olarak belirlendi.

Gözlem sayısı 1000 olduğunda gerçekleştirilen sınıflandırma çalışmasında sınıf dağılım oranı 0,1 iken NKD için ham veri ile 0,931±0,01, EG ile 911±0,01, EF ile 0,910±0,01 değerleri bulundu. Sınıf dağılım oranı 0,5 olduğunda bu değerler sırasıyla; 0,761±0,04, 0,746±0,04, 0,749±0,04 olarak bulundu.

Gözlem sayısı 10000 olan sınıflandırma çalışmasında, sınıf dağılım oranı 0,1 iken ortalama NKD; ham veri ile 0,936±0,004, EG ile 0,926±0,004, EF ile 0,928±0,004 olarak

bulundu. Sınıf dağılım oranı 0,5 olduğunda ise bu değerler sırasıyla; 0,792±0,01, 0,788±0,01, 0,788±0,01 olarak bulundu.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama negatif kestirim değerlerine Tablo 33'te yer verilmektedir.

F dağılımından üretilen verilerin sınıflandırılması

Tablo 34. F dağılımı için C5.0 ile elde edilen ortalama doğruluk oranları (%)

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	66,908±17,63	66,908±17,63	66,908±17,63	52,928±8,99	51,364±6,97
	0,2	74,184±12,97	74,184±12,97	74,197±12,98	56,372±10,23	57,584±11,66
	0,3	78,008±10,93	78,008±10,93	78,004±10,92	60,007±10,29	64,342±12,87
	0,4	78,675±10,05	78,675±10,05	78,671±10,05	63,163±10,67	69,796±10,63
	0,5	78,600±9,26	78,600±9,26	78,590±9,26	66,025±10,84	71,040±10,71
500	0,1	75,998±8,33	75,998±8,33	75,995±8,33	57,706±6,95	62,276±9,08
	0,2	82,488±5,59	82,488±5,59	82,490±5,59	61,803±6,04	77,712±6,49
	0,3	84,810±4,31	84,810±4,31	84,811±4,31	65,289±5,79	83,031±4,75
	0,4	85,447±3,79	85,447±3,79	85,450±3,79	67,938±5,55	83,193±4,12
	0,5	85,874±3,70	85,875±3,70	85,879±3,70	69,460±5,00	83,863±4,28
1000	0,1	80,427±5,79	80,427±5,79	80,426±5,80	59,058±5,40	67,917±7,27
	0,2	84,473±3,88	84,469±3,87	84,472±3,87	62,741±4,94	81,372±4,65
	0,3	86,317±2,96	86,317±2,96	86,315±2,96	65,103±4,64	85,299±3,16
	0,4	86,954±2,64	86,954±2,64	86,943±2,64	67,730±4,57	85,517±2,85
	0,5	87,464±2,45	87,464±2,45	87,463±2,44	69,474±4,25	86,403±2,65
10000	0,1	84,932±1,76	84,932±1,76	84,933±1,76	59,264±3,69	78,863±2,14
	0,2	87,676±1,13	87,676±1,13	87,678±1,13	58,868±3,01	85,842±1,34
	0,3	89,158±0,92	89,158±0,92	89,152±0,93	58,946±2,81	88,489±0,89
	0,4	89,858±0,75	89,858±0,75	89,859±0,76	60,059±2,47	89,449±0,78
	0,5	90,065±0,69	90,065±0,69	90,077±0,69	61,995±1,60	89,841±0,70

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Sınıflandırma sonucunda 100 gözlem için elde edilen ortalama doğruluk oranları sınıf dağılım oranı 0,1 olduğunda; ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında %66,908±17,63, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile %52,928±8,99, eşit frekansta aralıklandırma (EF) diskritizasyonu ile %51,364±6,97 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama doğruluk oranlarının; ham veri ve minimum-maksimum normalizasyonu ile %74,184±12,97, z-skor normalizasyonu ile %74,197±12,98, EG diskritizasyonu ile %56,372±10,23, EF diskritizasyonu ile %57,584±11,66 olduğu görüldü. Sınıf dağılım oranı 0,3 iken ortalama doğruluk oranları; ham veri ve minimum-maksimum normalizasyonu ile %78,008±10,93 olup z-skor normalizasyonu

ile %78,004±10,92, EG diskritizasyonu ile %60,007±10,29, EF diskritizasyonu ile %64,342±12,87 olarak hesaplandı. Sınıf dağılım oranı 0,4 iken elde edilen ortalama doğruluk oranları için ham veri ve minimum-maksimum normalizasyonu ile %78,675±10,05, z-skor normalizasyonu ile %78,671±10,05, EG diskritizasyonu ile %63,163±10,67, EF diskritizasyonu ile %69,796±10,63 değerlerine ulaşıldı. Son olarak sınıf dağılım oranı 0,5 olduğunda ise bu oranlar sırasıyla; %78,6±9,26, %78,59±9,26, %66,025±10,84, %71,04±10,71'dir.

Gözlem sayısı 500 iken yapılan çalışmada, sınıf dağılım oranı 0,1 olduğunda elde edilen ortalama doğruluk oranları; ham veride ve minimum-maksimum normalizasyonu sonrasında %75,998±8,33, z-skor normalizasyonu sonrasında %75,995±8,33, EG diskritizasyonu ile %57,706±6,95, EF diskritizasyonu ile %62,276±9,08 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise bu oranların; ham veride minimum-maksimum normalizasyonu sonrasında %85,874±3,7, z-skor normalizasyonu sonrasında %85,875±3,7, %85,879±3,7, EG diskritizasyonu ile %69,46±5, EF diskritizasyonu ile %83,863±4,28 değerlerine yükseldiği belirlendi.

Gözlem sayısı 1000 olduğunda, 0,1 sınıf dağılım oranı için ortalama doğruluk oranları; ham veride ve minimum-maksimum normalizasyonu sonrasında %80,427±5,79, z-skor normalizasyonu sonrasında %80,426±5,80, EG diskritizasyonu ile %59,058±5,40, EF diskritizasyonu ile %67,917±7,27 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama doğruluk oranları sırasıyla; %87,464±2,45, %87,463±2,44, %69,474±4,25, %86,403±2,65 olarak elde edildi.

Gözlem sayısı 10000 olduğunda yapılan sınıflandırma çalışmasında sınıf dağılım oranı 0,1 iken elde edilen ortalama doğruluk oranları; ham veride %84,932±1,69, minimum-maksimum normalizasyonu sonrasında %84,932±1,69, z-skor normalizasyonu sonrasında %84,933±1,76, EG diskritizasyonu ile %59,264±3,69, EF diskritizasyonu ile %78,863±2,14 olarak bulundu. Sınıf dağılımı 0,5 oranına arttığında ise elde edilen ortalama doğruluk oranlarının, ham veride ve minimum-maksimum normalizasyonu sonrasında %90,065±0,69, z-skor normalizasyonu sonrasında %90,077±0,69, EG diskritizasyonu ile %61,995±1,60, EF diskritizasyonu ile %89,841±0,70 oranlarına yükselmiştir.

Çalışmadaki diğer sınıf dağılımlarında (0,2, 0,3, 0,4), 500, 1000 ve 10000 gözlem sayıları için elde edilen ortalama doğruluk oranlarına Tablo 34'te yer verildi.

Tablo 35. F dağılımı için C5.0 ile elde edilen ortalama AUC değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,694±0,19	0,694±0,19	0,694±0,19	0,541±0,11	0,515±0,07
	0,2	0,768±0,14	0,768±0,14	0,768±0,14	0,576±0,12	0,585±0,13
	0,3	0,813±0,11	0,813±0,11	0,813±0,11	0,619±0,12	0,664±0,14
	0,4	0,822±0,10	0,822±0,10	0,822±0,10	0,656±0,12	0,712±0,12
	0,5	0,834±0,09	0,834±0,09	0,834±0,09	0,684±0,12	0,716±0,11
500	0,1	0,808±0,10	0,808±0,10	0,808±0,10	0,592±0,08	0,674±0,11
	0,2	0,881±0,06	0,881±0,06	0,881±0,06	0,638±0,07	0,803±0,07
	0,3	0,906±0,04	0,906±0,04	0,906±0,04	0,682±0,07	0,863±0,06
	0,4	0,915±0,03	0,915±0,03	0,915±0,03	0,724±0,07	0,876±0,04
	0,5	0,921±0,03	0,921±0,03	0,921±0,03	0,752±0,06	0,886±0,05
1000	0,1	0,859±0,06	0,859±0,06	0,859±0,06	0,604±0,06	0,736±0,09
	0,2	0,907±0,04	0,907±0,04	0,907±0,04	0,643±0,06	0,855±0,05
	0,3	0,929±0,03	0,929±0,03	0,929±0,03	0,679±0,06	0,892±0,04
	0,4	0,934±0,02	0,934±0,02	0,934±0,02	0,724±0,06	0,908±0,03
	0,5	0,937±0,02	0,937±0,02	0,937±0,02	0,755±0,05	0,917±0,03
10000	0,1	0,902±0,04	0,902±0,04	0,902±0,04	0,597±0,04	0,895±0,02
	0,2	0,948±0,02	0,948±0,02	0,947±0,02	0,593±0,03	0,917±0,02
	0,3	0,959±0,01	0,959±0,01	0,959±0,01	0,600±0,04	0,951±0,01
	0,4	0,962±0,004	0,962±0,004	0,962±0,004	0,642±0,04	0,954±0,01
	0,5	0,962±0,004	0,962±0,004	0,962±0,004	0,679±0,02	0,954±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

F dağılımından üretilen veriler ile elde edilen ortalama AUC değerleri incelendiğinde, 100, 500 ve 1000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 35).

Sınıflandırma çalışmasında 100 gözlem için elde edilen ortalama AUC değerleri, sınıf dağılım oranı 0,1 olduğunda; ham veride 0,694±0,19, eşit genişlikte aralıklandırma (EG) diskritizasyonu ve eşit frekansta aralıklandırma (EF) diskritizasyonu sonrasında sırasıyla 0,541±0,11 ve 0,515±0,07 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama AUC değerleri; ham veride 0,768±0,14 olarak bulunurken EG diskritizasyonu ile 0,576±0,12, EF diskritizasyonu ile 0,585±0,13 olarak elde edildi. Bu değerler sırasıyla; 0,3 sınıf dağılım oranı için 0,813±0,11, 0,619±0,12, 0,664±0,14, 0,4 sınıf dağılım oranı için 0,822±0,1, 0,656±0,12, 0,712±0,12 olarak saptandı. Son olarak sınıf dağılım oranı 0,5 olduğunda sırasıyla ortalama AUC değerlerinin; 0,834±0,09, 0,684±0,12, 0,716±0,11 olduğu sonucuna ulaşıldı.

Gözlem sayısı 500 olan sınıflandırma çalışmasında, sınıf dağılım oranı 0,1 olduğunda ortalama AUC değerleri için ham veri ile 0,808±0,1, EG ile 0,592±0,08, EF ile 0,674±0,11

sonuçları elde edildi. Sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama AUC değerleri sırasıyla; 0,921±0,03, 0,752±0,06, 0,886±0,05 olarak bulundu.

Gözlem sayısı 1000 olduğunda, sınıf dağılım oranı 0,1 iken elde edilen ortalama AUC değerleri için ham veri ile 0,859±0,06, EG diskritizasyonu ile 0,604±0,06, EF diskritizasyonu ile 0,736±0,09 değerlerine ulaşıldı. Sınıf dağılım oranı 0,5 olduğunda ise elde edilen ortalama AUC değerlerinin sırasıyla; 0,937±0,02, 0,755±0,05, 0,917±0,03 değerlerine yükseldiği bulgusuna ulaşıldı.

Gözlem sayısı 10000 olduğunda, sınıf dağılım oranı 0,1 iken elde edilen ortalama AUC değerleri; ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında 0,902±0,04, EG diskritizasyonu ile 0,597±0,04, EF diskritizasyonu ile 0,895±0,02 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise bu değerler sırasıyla; 0,962±0,004, 0,679±0,02, 0,954±0,01 olarak elde edildi.

Çalışmadaki diğer sınıf dağılımlarında (0,2, 0,3, 0,4), 500, 1000 ve 10000 gözlem sayıları için elde edilen ortalama AUC değerlerine Tablo 35’te yer verildi.

Tablo 36. F dağılımı için C5.0 ile elde edilen ortalama duyarlılık oranları

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,380±0,35	0,380±0,35	0,380±0,35	0,067±0,18	0,032±0,15
	0,2	0,563±0,26	0,563±0,26	0,563±0,26	0,152±0,21	0,180±0,26
	0,3	0,670±0,21	0,670±0,21	0,670±0,21	0,267±0,23	0,381±0,30
	0,4	0,732±0,17	0,732±0,17	0,732±0,17	0,434±0,23	0,591±0,23
	0,5	0,784±0,14	0,784±0,14	0,783±0,14	0,637±0,20	0,714±0,22
500	0,1	0,549±0,17	0,549±0,17	0,548±0,17	0,161±0,14	0,266±0,19
	0,2	0,705±0,11	0,705±0,11	0,705±0,11	0,259±0,13	0,613±0,13
	0,3	0,775±0,08	0,775±0,08	0,775±0,08	0,358±0,14	0,725±0,10
	0,4	0,819±0,07	0,819±0,07	0,819±0,07	0,487±0,14	0,767±0,08
	0,5	0,851±0,06	0,851±0,06	0,851±0,06	0,678±0,12	0,820±0,07
1000	0,1	0,632±0,12	0,632±0,12	0,632±0,12	0,187±0,11	0,384±0,15
	0,2	0,735±0,08	0,735±0,08	0,735±0,08	0,273±0,11	0,685±0,10
	0,3	0,793±0,06	0,793±0,06	0,793±0,06	0,344±0,11	0,764±0,07
	0,4	0,829±0,05	0,829±0,05	0,829±0,05	0,473±0,14	0,800±0,06
	0,5	0,863±0,05	0,863±0,05	0,863±0,05	0,679±0,12	0,842±0,05
10000	0,1	0,712±0,04	0,712±0,04	0,712±0,04	0,192±0,07	0,601±0,04
	0,2	0,781±0,03	0,781±0,03	0,781±0,03	0,191±0,06	0,754±0,03
	0,3	0,830±0,02	0,830±0,02	0,830±0,02	0,202±0,07	0,818±0,02
	0,4	0,866±0,02	0,866±0,02	0,866±0,02	0,267±0,09	0,858±0,02
	0,5	0,890±0,02	0,890±0,02	0,890±0,02	0,598±0,24	0,888±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

F dağılımından üretilen veriler ile elde edilen ortalama duyarlılık oranları incelendiğinde, 1000 gözlem sayısında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 36).

Sınıflandırma çalışmasında 100 gözlem için elde edilen ortalama duyarlılık oranları sınıf dağılımı 0,1 olduğunda; ham veride minimum-maksimum ve z-skor normalizasyonu sonrasında $0,38\pm 0,35$ iken eşit genişlikte aralıklandırma (EG) diskritizasyonu ile $0,067\pm 0,18$, eşit frekansta aralıklandırma (EF) diskritizasyonu ile $0,032\pm 0,15$ olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama duyarlılık oranlarının sırasıyla; $0,563\pm 0,26$, $0,152\pm 0,21$, $0,18\pm 0,26$ olduğu görüldü. Bu oranlar 0,3 dağılım oranı için ham $0,67\pm 0,21$, $0,267\pm 0,23$, $0,381\pm 0,3$, 0,4 dağılım oranı için $0,732\pm 0,17$, $0,434\pm 0,23$, $0,591\pm 0,23$ olarak hesaplandı. Son olarak sınıf dağılım oranı 0,5 olduğunda ise ortalama duyarlılık oranlarının; ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında $0,784\pm 0,14$, EG ve EF diskritizasyonu sonrasında sırasıyla $0,637\pm 0,2$, $0,714\pm 0,22$ olduğu bulgusuna ulaşıldı.

Gözlem sayısının 500 olduğu sınıflandırma çalışmasında, 0,1 sınıf dağılım oranında elde edilen ortalama duyarlılık oranları için ham veride ve minimum-maksimum normalizasyonu sonrasında $0,549\pm 0,17$, z-skor normalizasyonu sonrasında $0,548\pm 0,17$, EG diskritizasyonu ile $0,161\pm 0,14$, EF diskritizasyonu ile $0,266\pm 0,19$ değerlerine ulaşıldı. Sınıf dağılım oranı 0,5 olduğunda ortalama duyarlılık oranları için; ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında $0,851\pm 0,06$, EG ve EF diskritizasyonu sonrasında ise sırasıyla $0,678\pm 0,12$, $0,82\pm 0,07$ değerleri elde edildi.

Gözlem sayısı 1000 olduğunda, sınıf dağılım oranı 0,1 iken elde edilen ortalama duyarlılık oranları; ham veri, minimum-maksimum ve z-skor normalizasyonu ile $0,632\pm 0,12$, EG diskritizasyonu ile $0,187\pm 0,11$, EF diskritizasyonu ile $0,384\pm 0,15$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda elde edilen bu oranlar sırasıyla; $0,863\pm 0,05$, $0,679\pm 0,12$, $0,842\pm 0,05$ olarak belirlendi.

Gözlem sayısı 10000 olduğunda, sınıf dağılım oranı 0,1 iken elde edilen ortalama duyarlılık oranları; ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında $0,712\pm 0,04$, EG ve EF diskritizasyonu sonrasında $0,192\pm 0,07$ ve $0,601\pm 0,04$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise ortalama duyarlılık oranları sırasıyla; $0,890\pm 0,02$, $0,598\pm 0,24$, $0,888\pm 0,01$ olarak elde edildi.

Çalışmadaki diğer sınıf dağılımlarında (0,2, 0,3, 0,4), 500, 1000 ve 10000 gözlem sayıları için elde edilen ortalama duyarlılık oranlarına Tablo 36’da yer verildi.

Tablo 37. F dağılımı için C5.0 algoritması ile elde edilen ortalama seçicilik oranları

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,959±0,06	0,959±0,06	0,959±0,06	0,992±0,03	0,995±0,02
	0,2	0,921±0,08	0,921±0,08	0,921±0,08	0,975±0,05	0,972±0,07
	0,3	0,890±0,10	0,890±0,10	0,890±0,10	0,934±0,10	0,906±0,12
	0,4	0,842±0,13	0,842±0,13	0,842±0,13	0,830±0,16	0,805±0,16
	0,5	0,789±0,14	0,789±0,14	0,789±0,14	0,683±0,20	0,707±0,20
500	0,1	0,971±0,02	0,971±0,02	0,972±0,02	0,993±0,01	0,979±0,02
	0,2	0,945±0,03	0,945±0,03	0,945±0,03	0,977±0,03	0,942±0,03
	0,3	0,921±0,04	0,921±0,04	0,921±0,04	0,948±0,05	0,936±0,04
	0,4	0,890±0,05	0,890±0,05	0,890±0,05	0,872±0,09	0,897±0,06
	0,5	0,867±0,06	0,867±0,06	0,867±0,06	0,711±0,12	0,857±0,08
1000	0,1	0,977±0,01	0,977±0,01	0,977±0,01	0,994±0,01	0,975±0,02
	0,2	0,954±0,02	0,954±0,02	0,954±0,02	0,982±0,02	0,942±0,02
	0,3	0,933±0,03	0,933±0,03	0,933±0,03	0,958±0,04	0,942±0,03
	0,4	0,910±0,04	0,910±0,04	0,910±0,04	0,882±0,08	0,911±0,04
	0,5	0,886±0,05	0,886±0,05	0,886±0,05	0,710±0,11	0,887±0,05
10000	0,1	0,987±0,004	0,987±0,004	0,987±0,004	0,996±0,004	0,979±0,005
	0,2	0,973±0,01	0,973±0,01	0,973±0,01	0,992±0,01	0,963±0,01
	0,3	0,953±0,01	0,953±0,01	0,952±0,01	0,979±0,03	0,951±0,01
	0,4	0,932±0,01	0,932±0,01	0,932±0,01	0,937±0,05	0,931±0,01
	0,5	0,911±0,02	0,911±0,02	0,911±0,02	0,642±0,24	0,908±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

F dağılımından üretilen veriler ile elde edilen ortalama seçicilik oranları incelendiğinde, 100 ve 1000 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 37).

Sınıflandırma çalışmasında 100 gözlem için elde edilen ortalama seçicilik oranı, sınıf dağılım oranı 0,1 olduğunda; ham veri ile 0,959±0,06 olup eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,992±0,03, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,995±0,02’dir. Sınıf dağılım oranı 0,2 olduğunda; ham veri ile 0,921±0,08, EG ile 0,975±0,05, EF ile 0,972±0,07 değerleri elde edildi. Sınıf dağılım oranı 0,3 için bu oranlar sırasıyla; 0,89±0,1, 0,934±0,1, 0,906±0,12 olarak bulunurken 0,4 sınıf dağılım oranında; 0,842±0,13, 0,83±0,16, 0,805±0,16 olarak elde edildi. Son olarak sınıf dağılım oranı 0,5 olduğunda ise bu oranların sırasıyla; 0,789±0,14, 0,683±0,2, 0,707±0,2 olduğu görüldü.

Gözlem sayısı 500 olan çalışma incelendiğinde, sınıf dağılım oranı 0,1 olduğunda ortalama seçicilik oranları için ham veri ve minimum-maksimum normalizasyonu ile $0,971\pm 0,02$ değerleri elde edilirken z-skor normalizasyonu sonrasında $0,972\pm 0,02$, EG diskritizasyonu ile $0,993\pm 0,01$, EF diskritizasyonu ile $0,979\pm 0,02$ değerlerine ulaşıldı. Sınıf dağılım oranı 0,5 olduğunda ise ham veri, minimum-maksimum ve z-skor normalizasyonu ile $0,867\pm 0,06$, EG ile $0,711\pm 0,12$, EF ile $0,857\pm 0,08$ oranları elde edildi.

Gözlem sayısı 1000 olduğunda, sınıf dağılım oranı 0,1 iken ortalama seçicilik oranları; ham veri, minimum-maksimum ve z-skor normalizasyonu ile $0,977\pm 0,01$, EG diskritizasyonu ile $0,994\pm 0,01$ ve EF diskritizasyonu ile $0,975\pm 0,02$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda bu oranlar sırasıyla; $0,886\pm 0,05$, $0,71\pm 0,11$, $0,887\pm 0,05$ olarak elde edildi.

Gözlem sayısı 10000 olduğunda, sınıf dağılım oranı 0,1 iken ortalama seçicilik oranları; ham veri ile $0,987\pm 0,004$, EG ile $0,996\pm 0,004$, EF ile $0,979\pm 0,005$, sınıf dağılım oranı 0,5 olduğunda ise sırasıyla; $0,911\pm 0,02$, $0,642\pm 0,24$, $0,908\pm 0,01$ değerleri bulundu.

Çalışmadaki diğer sınıf dağılımlarında (0,2, 0,3, 0,4), 500, 1000 ve 10000 gözlem sayıları için elde edilen ortalama seçicilik oranları Tablo 37’de yer verildi.

Tablo 38. F dağılımı için C5.0 ile elde edilen ortalama pozitif kestirim değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	$0,576\pm 0,39$	$0,576\pm 0,39$	$0,576\pm 0,39$	$0,535\pm 0,45$	$0,433\pm 0,35$
	0,2	$0,692\pm 0,26$	$0,692\pm 0,26$	$0,692\pm 0,26$	$0,669\pm 0,37$	$0,708\pm 0,31$
	0,3	$0,754\pm 0,19$	$0,754\pm 0,19$	$0,754\pm 0,19$	$0,706\pm 0,30$	$0,685\pm 0,23$
	0,4	$0,777\pm 0,15$	$0,777\pm 0,15$	$0,777\pm 0,15$	$0,684\pm 0,23$	$0,713\pm 0,19$
	0,5	$0,802\pm 0,12$	$0,802\pm 0,12$	$0,802\pm 0,12$	$0,691\pm 0,14$	$0,737\pm 0,14$
500	0,1	$0,707\pm 0,17$	$0,707\pm 0,17$	$0,707\pm 0,17$	$0,786\pm 0,29$	$0,615\pm 0,25$
	0,2	$0,774\pm 0,10$	$0,774\pm 0,10$	$0,774\pm 0,10$	$0,794\pm 0,19$	$0,736\pm 0,11$
	0,3	$0,814\pm 0,07$	$0,814\pm 0,07$	$0,814\pm 0,07$	$0,787\pm 0,15$	$0,840\pm 0,09$
	0,4	$0,837\pm 0,06$	$0,837\pm 0,06$	$0,837\pm 0,06$	$0,746\pm 0,12$	$0,840\pm 0,07$
	0,5	$0,868\pm 0,05$	$0,868\pm 0,05$	$0,868\pm 0,05$	$0,713\pm 0,07$	$0,858\pm 0,06$
1000	0,1	$0,762\pm 0,11$	$0,762\pm 0,11$	$0,763\pm 0,11$	$0,833\pm 0,21$	$0,649\pm 0,16$
	0,2	$0,806\pm 0,07$	$0,806\pm 0,07$	$0,806\pm 0,07$	$0,833\pm 0,14$	$0,755\pm 0,07$
	0,3	$0,840\pm 0,06$	$0,840\pm 0,06$	$0,840\pm 0,06$	$0,810\pm 0,12$	$0,854\pm 0,06$
	0,4	$0,862\pm 0,05$	$0,862\pm 0,05$	$0,862\pm 0,05$	$0,758\pm 0,10$	$0,860\pm 0,05$
	0,5	$0,886\pm 0,04$	$0,886\pm 0,04$	$0,886\pm 0,04$	$0,711\pm 0,07$	$0,884\pm 0,04$
10000	0,1	$0,858\pm 0,03$	$0,858\pm 0,03$	$0,858\pm 0,04$	$0,874\pm 0,11$	$0,762\pm 0,04$
	0,2	$0,879\pm 0,03$	$0,879\pm 0,03$	$0,880\pm 0,03$	$0,877\pm 0,10$	$0,836\pm 0,02$
	0,3	$0,884\pm 0,02$	$0,884\pm 0,02$	$0,883\pm 0,02$	$0,861\pm 0,12$	$0,877\pm 0,02$
	0,4	$0,895\pm 0,02$	$0,895\pm 0,02$	$0,895\pm 0,01$	$0,791\pm 0,12$	$0,893\pm 0,01$
	0,5	$0,910\pm 0,01$	$0,910\pm 0,01$	$0,910\pm 0,01$	$0,665\pm 0,09$	$0,907\pm 0,01$

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

F dağılımından üretilen veriler ile elde edilen ortalama pozitif kestirim değerleri incelendiğinde, 100, 500 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 38).

Sınıflandırma çalışmasında 100 gözlem için elde edilen ortalama pozitif kestirim değeri (PKD), sınıf dağılım oranı 0,1 olduğunda; ham veride $0,576\pm 0,39$, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile $0,535\pm 0,45$, eşit frekansta aralıklandırma (EF) diskritizasyonu ile $0,433\pm 0,35$ olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama PKD'nin; ham veride $0,692\pm 0,26$ olarak bulunurken EG diskritizasyonu ile $0,669\pm 0,37$, EF diskritizasyonu ile $0,708\pm 0,31$ değerleri elde edildi. Bu değerler sırasıyla; sınıf dağılım oranı 0,3 için $0,754\pm 0,19$, $0,706\pm 0,3$, $0,685\pm 0,23$, sınıf dağılım oranı 0,4 için $0,777\pm 0,15$, $0,684\pm 0,23$, $0,713\pm 0,19$ olarak hesaplandı. Son olarak sınıf dağılım oranı 0,5 olduğunda ise ortalama PKD için ham veri ile $0,802\pm 0,12$, EG diskritizasyonu ile $0,691\pm 0,14$, EF diskritizasyonu ile $0,737\pm 0,14$ bulgularına ulaşıldı.

Gözlem sayısı 500 olan çalışmada, 0,1 sınıf dağılım oranı için ortalama PKD; ham veri, minimum-maksimum ve z-skor normalizasyonu ile $0,707\pm 0,17$, EG diskritizasyonu ile $0,786\pm 0,29$, EF diskritizasyonu ile $0,615\pm 0,25$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise ortalama PKD sırasıyla; $0,868\pm 0,05$, $0,713\pm 0,07$, $0,858\pm 0,06$ olarak belirlendi.

Gözlem sayısı 1000 olduğunda, sınıf dağılım oranı 0,1 iken ortalama PKD; ham veri ve minimum-maksimum normalizasyonu ile $0,762\pm 0,11$, z-skor normalizasyonu ile $0,763\pm 0,11$, EG ile $0,833\pm 0,21$, EF ile $0,649\pm 0,16$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise; ham veri, minimum-maksimum ve z-skor normalizasyonu ile $0,886\pm 0,04$, EG ve EF ile sırasıyla $0,711\pm 0,07$, $0,884\pm 0,04$ olarak bulundu.

Gözlem sayısı 10000 olduğunda yapılan sınıflandırma çalışmasında sınıf dağılım oranı 0,1 iken elde edilen ortalama PKD için ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında $0,858\pm 0,04$, EG ile $0,874\pm 0,11$, EF ile $0,762\pm 0,04$ değerlerine ulaşıldı. Sınıf dağılım oranı 0,5 olduğunda ise ortalama PKD için sırasıyla; $0,910\pm 0,01$, $0,665\pm 0,09$, $0,907\pm 0,01$ değerleri elde edildi.

Çalışmadaki diğer sınıf dağılımlarında (0,2, 0,3, 0,4), 500, 1000 ve 10000 gözlem sayıları için elde edilen ortalama pozitif kestirim değerlerine Tablo 38'de yer verildi.

Tablo 39. F dağılımı için C5.0 ile elde edilen ortalama negatif kestirim değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,934±0,04	0,934±0,04	0,934±0,04	0,906±0,02	0,903±0,01
	0,2	0,897±0,06	0,897±0,06	0,897±0,06	0,823±0,04	0,829±0,05
	0,3	0,869±0,07	0,869±0,07	0,869±0,07	0,754±0,06	0,784±0,08
	0,4	0,833±0,09	0,833±0,09	0,833±0,09	0,697±0,09	0,764±0,11
	0,5	0,797±0,11	0,797±0,11	0,797±0,11	0,673±0,14	0,747±0,15
500	0,1	0,951±0,02	0,951±0,02	0,951±0,02	0,914±0,01	0,924±0,02
	0,2	0,928±0,03	0,928±0,03	0,928±0,03	0,841±0,02	0,908±0,03
	0,3	0,907±0,03	0,907±0,03	0,907±0,03	0,777±0,03	0,890±0,03
	0,4	0,883±0,04	0,883±0,04	0,883±0,04	0,723±0,05	0,855±0,04
	0,5	0,856±0,05	0,856±0,05	0,856±0,05	0,698±0,07	0,830±0,05
1000	0,1	0,960±0,01	0,960±0,01	0,960±0,01	0,917±0,01	0,935±0,01
	0,2	0,936±0,02	0,936±0,02	0,936±0,02	0,844±0,02	0,923±0,02
	0,3	0,914±0,02	0,914±0,02	0,914±0,02	0,775±0,03	0,904±0,02
	0,4	0,890±0,03	0,890±0,03	0,890±0,03	0,720±0,04	0,874±0,03
	0,5	0,869±0,04	0,869±0,04	0,868±0,04	0,697±0,06	0,851±0,04
10000	0,1	0,969±0,004	0,969±0,004	0,969±0,004	0,917±0,01	0,957±0,004
	0,2	0,947±0,01	0,947±0,01	0,947±0,01	0,831±0,01	0,940±0,01
	0,3	0,929±0,01	0,929±0,01	0,929±0,01	0,742±0,01	0,924±0,01
	0,4	0,912±0,01	0,912±0,01	0,913±0,01	0,658±0,02	0,908±0,01
	0,5	0,893±0,01	0,893±0,01	0,893±0,01	0,655±0,10	0,891±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

F dağılımından üretilen veriler ile elde edilen ortalama negatif kestirim değerleri incelendiğinde, 100 ve 500 gözlem sayılarında, 0,1, 0,2, 0,3, 0,4, 0,5 sınıf dağılım oranlarının her biri kendi içinde olmak üzere ham veride, minimum-maksimum normalizasyonu ve z-skor normalizasyonu sonrasında aynı değerleri aldığı görüldü (Tablo 39).

Sınıflandırma çalışmasında 100 gözlem için elde edilen ortalama negatif kestirim değerleri (NKD), sınıf dağılım oranı 0,1 olduğunda; ham veride 0,934±0,04 olarak bulunurken eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,906±0,02, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,903±0,01 sonuçlarına ulaşıldı. Bu değerler sırasıyla; sınıfların 0,2 dağılım oranı için 0,897±0,06, 0,823±0,04, 0,829±0,05, 0,3 dağılım oranı için 0,869±0,07, 0,754±0,06, 0,784±0,08, 0,4 dağılım oranı için 0,833±0,09, 0,697±0,09, 0,764±0,11 olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda ise ham veride 0,797±0,11 olup EG ile 0,673±0,14, EF ile 0,747±0,15 bulgularına ulaşıldı.

Gözlem sayısı 500 olan sınıflandırma çalışmasında, 0,1 sınıf dağılım oranı için ortalama negatif kestirim değerleri; ham veride 0,951±0,02, EG diskritizasyonu ile

0,914±0,01, EF diskritizasyonu ile 0,924±0,02 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise bu değerlerin sırasıyla; 0,856±0,05, 0,698±0,07, 0,83±0,05 olduğu saptandı.

Gözlem sayısı 1000 olduğunda, 0,1 sınıf dağılım oranı için elde edilen ortalama NKD; ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında 0,96±0,01, EG diskritizasyonu ile 0,917±0,01, EF diskritizasyonu ile 0,935±0,01 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise elde edilen değerler; ham veride ve minimum-maksimum normalizasyonu sonrasında 0,869±0,04, z-skor normalizasyonu sonrasında 0,868±0,04, EG diskritizasyonu ile 0,697±0,06, EF diskritizasyonu ile 0,851±0,04'tür.

Son olarak gözlem sayısının 10000 olduğu çalışmada, 0,1 sınıf dağılım oranı için elde edilen ortalama NKD; ham veride, minimum-maksimum ve z-skor normalizasyonu sonrasında 0,969±0,004, EG diskritizasyonu ile 0,917±0,01, EF diskritizasyonu ile 0,957±0,004 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise ortalama NKD sırasıyla; 0,893±0,01, 0,655±0,1, 0,891±0,01 olarak elde edildi.

Çalışmadaki diğer sınıf dağılımlarında (0,2, 0,3, 0,4), 500, 1000 ve 10000 gözlem sayıları için elde edilen ortalama negatif kestirim değerlerine Tablo 39'da yer verildi.

DESTEK VEKTÖR MAKİNELERİ İLE SINIFLANDIRMA

Normal Dağılımdan üretilen verilerin sınıflandırılması

Tablo 40. Normal dağılım için DVM ile elde edilen ortalama doğruluk oranları (%)

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	50,181±2,46	66,039±18,2	65,650±18,06	51,686±7,05	53,625±9,99
	0,2	50,656±3,77	78,763±13,1	76,963±12,96	58,184±11,35	62,150±12,69
	0,3	51,188±4,77	82,817±9,80	81,580±10,22	66,027±11,46	68,800±11,39
	0,4	51,863±5,94	84,269±8,07	83,817±8,65	71,485±9,86	73,117±10,36
	0,5	52,680±9,01	84,170±8,28	84,010±8,20	73,450±10,48	73,970±10,40
500	0,1	50,526±1,95	77,773±8,05	75,362±8,30	66,114±8,16	65,702±8,20
	0,2	52,689±4,71	83,943±5,13	82,439±5,22	76,679±5,95	75,853±5,80
	0,3	53,915±7,21	85,062±4,07	85,371±4,36	80,882±4,57	80,281±4,57
	0,4	54,793±9,07	85,948±3,43	86,620±3,71	83,045±4,08	82,495±4,01
	0,5	55,962±11,69	78,670±18,4	86,990±3,30	82,958±5,24	81,905±7,47
1000	0,1	52,058±3,35	79,627±5,66	77,914±5,93	72,478±6,24	70,071±6,07
	0,2	54,468±7,16	83,694±3,58	83,703±3,72	80,056±4,03	78,939±3,94
	0,3	55,160±9,42	85,262±2,93	86,325±2,97	83,283±3,29	82,452±3,13
	0,4	54,597±9,71	86,193±2,46	87,467±2,46	84,799±2,83	84,295±2,69
	0,5	55,934±11,7	80,721±15,0	87,457±2,44	85,174±2,58	84,587±2,56

Tablo 40. Devam Normal dağılım için DVM ile elde edilen ortalama doğruluk oranları (%)

10000	0,1	55,769±9,88	79,035±1,79	80,557±1,77	79,758±1,76	77,667±1,85
	0,2	53,971±9,73	83,319±1,16	85,106±1,15	84,287±1,16	83,802±1,20
	0,3	53,763±8,44	85,444±0,88	87,380±0,87	86,482±0,90	86,513±0,90
	0,4	55,185±6,39	86,425±0,84	88,481±0,76	87,309±0,78	87,745±0,78
	0,5	63,422±4,25	86,656±0,78	88,804±0,76	87,651±0,75	88,121±0,77

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Ortalama doğruluk oranları 100 gözlem için incelendiğinde, bağımlı değişkendeki sınıfların dağılım oranı 0,1 iken ham veride %50,181±2,46, minimum-maksimum normalizasyonu sonrası %66,039±18,2, z-skor normalizasyonu sonrası %65,65±18,06, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile %51,686±7,05, eşit frekansta aralıklandırma (EF) diskritizasyonu ile %53,625±9,99 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama doğruluk oranlarının; ham veride %50,656±3,77, olarak bulurken minimum-maksimum normalizasyonu sonrası %78,763±13,1, z-skor normalizasyonu sonrası %76,963±12,96, EG ile %58,184±11,35, EF ile %62,150±12,69 olduğu görüldü. Sınıf dağılım oranı 0,3 iken bu oranlar sırasıyla; %51,188±4,77, %82,817±9,80, %81,58±10,22, %66,027±11,46, %68,800±11,39 olarak hesaplandı. Sınıf dağılımının 0,4 olduğunda elde edilen ortalama doğruluk oranları sırasıyla; %51,863±5,94, %84,269±8,07, %83,817±8,65, %71,485±9,86, %73,117±10,36 olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda ise bu oranlar sırasıyla; %52,680±9,01, %84,170±8,28, %84,010±8,20, %73,450±10,48, %73,970±10,4 olarak elde edildi.

Gözlem sayısı 500 olan çalışmalar incelendiğinde, sınıf dağılım oranı 0,1 olduğunda ortalama doğruluk oranları için ham veride %50,526±1,95 olarak bulunurken minimum-maksimum normalizasyonu sonrası %77,773±8,05, z-skor normalizasyonu sonrası %75,362±8,30, EG ile %66,114±8,16, EF ile %65,702±8,20 olarak elde edildi. Sınıf dağılım oranı 0,5 olduğunda; ham veride %55,962±11,69, minimum-maksimum normalizasyonu sonrasında %78,670±18,4, z-skor normalizasyonu sonrasında %86,99±3,30, EG ile %82,958±5,24, EF ile %81,905±7,47 olarak ortalama doğruluk oranlarına ulaşıldı. Gözlem sayısı 1000 olduğunda gerçekleştirilen sınıflandırma çalışmaları incelendiğinde, 0,1 sınıf dağılım oranı için elde edilen ortalama doğruluk oranları; ham veride %52,058±3,35, minimum-maksimum normalizasyonu sonrası %79,627±5,66, z-skor normalizasyonu sonrası %77,914±5,93, EG ile %72,478±6,24, EF ile %70,071±6,07 olarak bulundu. Sınıf dağılım

oranı 0,5 olduğunda ise bu oranlar sırasıyla; %55,934±11,7, %80,721±15,0, %87,457±2,44, %85,174±2,58, %84,587±2,56 olarak bulundu. Gözlem sayısı 10000 olduğunda yapılan sınıflandırma çalışmasında sınıf dağılım oranı 0,1 iken ortalama doğruluk oranları; ham veride %55,769±9,88, minimum-maksimum normalizasyonu sonrası %79,035±1,79, z-skor normalizasyonu sonrası %80,557±1,77, EG ile %79,758±1,76, EF ile %77,667±1,85 olarak elde edildi. Sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama doğruluk oranları sırasıyla; %63,422±4,25, %86,656±0,78, %88,804±0,76, %87,651±0,75, %88,121±0,77 oranlarına yükseldi.

Çalışmadaki diğer sınıf dağılımlarında (0,2, 0,3, 0,4), 500, 1000 ve 10000 gözlem sayıları için elde edilen ortalama doğruluk oranlarına Tablo 40’ta yer verildi.

Tablo 41. Normal dağılım için DVM ile elde edilen ortalama AUC değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,523±0,25	0,887±0,15	0,891±0,16	0,683±0,25	0,741±0,22
	0,2	0,521±0,19	0,913±0,09	0,906±0,10	0,779±0,15	0,806±0,14
	0,3	0,509±0,17	0,920±0,07	0,912±0,08	0,805±0,12	0,817±0,11
	0,4	0,499±0,17	0,926±0,06	0,919±0,07	0,815±0,10	0,824±0,10
	0,5	0,486±0,17	0,925±0,07	0,918±0,07	0,812±0,11	0,824±0,10
500	0,1	0,613±0,17	0,931±0,05	0,927±0,05	0,884±0,06	0,877±0,06
	0,2	0,568±0,18	0,935±0,03	0,934±0,03	0,906±0,04	0,887±0,06
	0,3	0,530±0,17	0,935±0,03	0,936±0,03	0,910±0,03	0,905±0,03
	0,4	0,519±0,18	0,937±0,02	0,939±0,03	0,914±0,03	0,910±0,03
	0,5	0,549±0,19	0,795±0,33	0,940±0,02	0,910±0,06	0,896±0,11
1000	0,1	0,636±0,18	0,934±0,03	0,932±0,03	0,914±0,04	0,903±0,04
	0,2	0,594±0,17	0,936±0,02	0,940±0,02	0,927±0,02	0,919±0,03
	0,3	0,584±0,17	0,938±0,02	0,943±0,02	0,928±0,02	0,922±0,02
	0,4	0,564±0,16	0,939±0,02	0,945±0,02	0,930±0,02	0,925±0,02
	0,5	0,580±0,17	0,829±0,29	0,944±0,02	0,930±0,02	0,925±0,02
10000	0,1	0,700±0,12	0,938±0,01	0,938±0,01	0,950±0,01	0,944±0,01
	0,2	0,679±0,09	0,940±0,01	0,947±0,01	0,951±0,01	0,948±0,01
	0,3	0,675±0,07	0,940±0,01	0,951±0,01	0,951±0,01	0,951±0,01
	0,4	0,670±0,06	0,940±0,01	0,953±0,01	0,941±0,01	0,952±0,01
	0,5	0,672±0,05	0,941±0,01	0,954±0,004	0,942±0,01	0,952±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Ortalama AUC değerleri 100 gözlem için incelendiğinde; sınıf dağılım oranı 0,1 iken ham veride 0,523±0,25, minimum-maksimum normalizasyonu sonrası 0,887±0,15, z-skor normalizasyonu sonrası 0,891±0,16, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,683±0,25, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,741±0,22 değerlerine ulaşıldı. Sınıf dağılım oranı 0,2 olduğunda ortalama AUC değerleri ham veride 0,521±0,19

olarak bulunurken minimum-maksimum normalizasyonu sonrası $0,913\pm0,09$, z-skor normalizasyonu sonrası $0,906\pm0,10$, EG diskritizasyonu ile $0,779\pm0,15$, EF diskritizasyonu ile $0,806\pm0,14$ olduğu görüldü. Sınıf dağılım oranı 0,3 iken bu değerler sırasıyla; $0,509\pm0,17$, $0,920\pm0,07$, $0,912\pm0,08$, $0,805\pm0,12$, $0,817\pm0,11$ olarak saptandı. Sınıf dağılım oranı 0,4 için ortalama AUC değerlerinin; ham veride $0,499\pm0,17$, minimum-maksimum normalizasyonu sonrası $0,926\pm0,06$, z-skor normalizasyonu sonrası $0,919\pm0,07$, EG diskritizasyonu ile $0,815\pm0,10$, EF diskritizasyonu ile $0,824\pm0,10$ olduğu gözlemlendi. Son olarak sınıf dağılım oranı 0,5 olduğunda; ham veride $0,486\pm0,17$, minimum-maksimum normalizasyonu sonrasında $0,925\pm0,07$, z-skor normalizasyonu sonrasında $0,918\pm0,07$, EG diskritizasyonu ile $0,812\pm0,11$, EF diskritizasyonu ile $0,824\pm0,10$ ortalama AUC değerlerine ulaşıldı.

Gözlem sayısı 500 olan sınıflandırma çalışmalarında 0,1 sınıf dağılım oranı için elde edilen ortalama AUC değerleri; ham veride $0,613\pm0,17$, minimum-maksimum normalizasyonu sonrasında $0,931\pm0,05$, z-skor normalizasyonu sonrasında $0,927\pm0,05$, EG diskritizasyonu ile $0,884\pm0,06$, EF diskritizasyonu ile $0,877\pm0,06$ şeklindedir. Son olarak sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama AUC değerleri; ham veride $0,549\pm0,19$, minimum-maksimum normalizasyonu sonrasında $0,795\pm0,33$, z-skor normalizasyonu sonrasında $0,940\pm0,02$, EG diskritizasyonu ile $0,910\pm0,06$, EF diskritizasyonu ile $0,896\pm0,11$ olarak belirlendi.

Gözlem sayısı 1000 olduğunda yapılan çalışmalarda, 0,1 sınıf dağılım oranı için ortalama AUC değerleri; ham veride $0,636\pm0,18$, minimum-maksimum normalizasyonu $0,934\pm0,03$, z-skor normalizasyonu sonrası $0,932\pm0,03$, EG diskritizasyonu ile $0,914\pm0,04$, EF diskritizasyonu ile $0,903\pm0,04$ olarak elde edildi. Sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama AUC değerleri; ham veride $0,580\pm0,17$, minimum-maksimum normalizasyonu sonrası $0,829\pm0,29$, z-skor normalizasyonu sonrası $0,944\pm0,02$, EG diskritizasyonu ile $0,930\pm0,02$, EF diskritizasyonu ile $0,925\pm0,02$ olarak bulundu.

Gözlem sayısı 10000 olduğunda yapılan çalışmalar incelendiğinde, 0,1 sınıf dağılım oranı için elde edilen ortalama AUC değerleri; ham veri ile $0,700\pm0,12$, minimum-maksimum normalizasyonu ile $0,938\pm0,01$, z-skor normalizasyonu ile $0,938\pm0,01$, EG ile $0,950\pm0,01$, EF ile $0,944\pm0,01$ 'dir. Sınıf dağılım oranı 0,5 olduğunda ise bu değerler; ham veri ile $0,672\pm0,05$, minimum-maksimum normalizasyonu sonrası $0,941\pm0,01$, z-skor normalizasyonu sonrası $0,954\pm0,004$, EG ile $0,942\pm0,01$, EF ile $0,952\pm0,01$ olarak elde edildi.

Çalışmadaki diğer sınıf dağılımlarında (0,2, 0,3, 0,4), 500, 1000 ve 10000 gözlem sayıları için elde edilen ortalama AUC değerlerine Tablo 41’de yer verildi.

Tablo 42. Normal dağılım için DVM ile elde edilen ortalama duyarlılık oranları

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,005±0,05	0,331±0,37	0,323±0,36	0,038±0,14	0,080±0,21
	0,2	0,020±0,08	0,618±0,26	0,578±0,26	0,191±0,24	0,272±0,26
	0,3	0,041±0,12	0,742±0,19	0,698±0,20	0,395±0,24	0,454±0,23
	0,4	0,078±0,17	0,827±0,14	0,782±0,15	0,590±0,19	0,615±0,19
	0,5	0,546±0,40	0,823±0,13	0,830±0,13	0,727±0,15	0,730±0,15
500	0,1	0,012±0,04	0,573±0,16	0,521±0,17	0,336±0,17	0,328±0,17
	0,2	0,062±0,11	0,729±0,10	0,683±0,10	0,575±0,12	0,454±0,20
	0,3	0,097±0,17	0,773±0,08	0,766±0,08	0,694±0,09	0,680±0,09
	0,4	0,135±0,24	0,817±0,06	0,822±0,06	0,769±0,07	0,762±0,07
	0,5	0,595±0,31	0,783±0,25	0,866±0,05	0,823±0,08	0,813±0,11
1000	0,1	0,044±0,07	0,612±0,11	0,572±0,12	0,466±0,13	0,417±0,12
	0,2	0,099±0,16	0,717±0,07	0,706±0,07	0,639±0,08	0,618±0,08
	0,3	0,122±0,22	0,770±0,06	0,782±0,06	0,732±0,06	0,716±0,06
	0,4	0,123±0,25	0,819±0,05	0,832±0,04	0,797±0,05	0,788±0,05
	0,5	0,578±0,27	0,807±0,22	0,869±0,04	0,844±0,04	0,838±0,04
10000	0,1	0,119±0,20	0,596±0,04	0,623±0,04	0,609±0,04	0,570±0,04
	0,2	0,089±0,20	0,702±0,02	0,730±0,02	0,718±0,02	0,712±0,02
	0,3	0,100±0,18	0,769±0,02	0,796±0,02	0,784±0,02	0,786±0,02
	0,4	0,168±0,14	0,819±0,02	0,842±0,01	0,826±0,01	0,835±0,01
	0,5	0,517±0,07	0,859±0,01	0,880±0,01	0,871±0,01	0,872±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Ortalama duyarlılık oranları incelendiğinde, 100 gözlem için elde edilen ortalama duyarlılık oranları sınıf dağılım oranı 0,1 için ham veride 0,005±0,05, minimum-maksimum normalizasyonu sonrası 0,331±0,37, z-skor normalizasyonu sonrası 0,323±0,36 olarak bulunurken eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,038±0,14, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,08±0,21 olarak elde edildi. Sınıf dağılım oranı 0,2 olduğunda ortalama duyarlılık oranlarının; ham veride 0,020±0,08, minimum-maksimum normalizasyonu ile 0,618±0,26, z-skor ile 0,578±0,26, EG diskritizasyonu ile 0,191±0,24, EF diskritizasyonu ile 0,272±0,26 olduğu görüldü. Sınıf dağılım oranı 0,3 iken bu oranlar sırasıyla; 0,041±0,12, 0,742±0,19, 0,698±0,20, 0,395±0,24, 0,454±0,23 olarak hesaplandı. Sınıf dağılım oranı 0,4 için sırasıyla; 0,078±0,17, 0,827±0,14, 0,782±0,15, 0,59±0,19, 0,615±0,19 ortalama duyarlılık oranlarına ulaşıldı. Son olarak sınıf dağılım oranı 0,5 olduğunda ise bu değerlerin sırasıyla; 0,546±0,40, 0,823±0,13, 0,830±0,13, 0,727±0,15, 0,73±0,15 olduğu sonucuna ulaşıldı.

Gözlem sayısı 500 olduğunda gerçekleştirilen sınıflandırma çalışmasında 0,1 sınıf dağılım oranı için elde edilen ortalama duyarlılık oranları; ham veride $0,012\pm 0,04$, minimum-maksimum normalizasyonu sonrasında $0,573\pm 0,16$, z-skor normalizasyonu sonrasında $0,521\pm 0,17$, EG diskritizasyonu ile $0,336\pm 0,17$, EF diskritizasyonu ile $0,328\pm 0,17$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama duyarlılık oranları; ham veride $0,595\pm 0,31$, minimum-maksimum normalizasyonu sonrasında $0,783\pm 0,25$, z-skor normalizasyonu sonrasında $0,866\pm 0,05$, EG diskritizasyonu ile $0,823\pm 0,08$, EF diskritizasyonu ile $0,813\pm 0,11$ olarak belirlendi.

Gözlem sayısı 1000 olduğunda gerçekleştirilen çalışmalar incelendiğinde, sınıf dağılım oranı 0,1 iken elde edilen ortalama duyarlılık oranları; ham veride $0,044\pm 0,07$, minimum-maksimum normalizasyonu sonrasında $0,612\pm 0,11$, z-skor normalizasyonu sonrasında $0,572\pm 0,12$, EG diskritizasyonu ile $0,466\pm 0,13$, EF diskritizasyonu ile $0,417\pm 0,12$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda elde edilen bu oranlar; ham veride $0,578\pm 0,27$, minimum-maksimum normalizasyonu sonrasında $0,807\pm 0,22$, z-skor normalizasyonu sonrasında $0,869\pm 0,04$, EG diskritizasyonu ile $0,844\pm 0,04$, EF diskritizasyonu ile $0,838\pm 0,04$ şeklindedir.

Gözlem sayısı 10000 olduğunda yapılan sınıflandırma çalışmaları ele alındığında, sınıf dağılım oranı 0,1 iken elde edilen ortalama duyarlılık oranlarının; ham veride $0,119\pm 0,20$, minimum-maksimum normalizasyonu sonrasında $0,596\pm 0,04$, z-skor normalizasyonu sonrasında $0,623\pm 0,04$, EG diskritizasyonu ile $0,609\pm 0,04$, EF diskritizasyonu ile $0,57\pm 0,04$ olduğu görüldü. Sınıf dağılım oranı 0,5 olduğunda bu oranlar sırasıyla; $0,517\pm 0,07$, $0,859\pm 0,01$, $0,880\pm 0,01$, $0,871\pm 0,01$, $0,872\pm 0,01$ olarak bulundu.

Çalışmadaki diğer sınıf dağılımlarında (0,2, 0,3, 0,4), 500, 1000 ve 10000 gözlem sayıları için elde edilen ortalama duyarlılık oranlarına Tablo 42’de yer verildi.

Tablo 43. Normal dağılım için DVM ile elde edilen ortalama seçicilik oranları

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	$0,999\pm 0,01$	$0,990\pm 0,03$	$0,990\pm 0,03$	$0,996\pm 0,02$	$0,992\pm 0,03$
	0,2	$0,993\pm 0,03$	$0,957\pm 0,06$	$0,961\pm 0,05$	$0,973\pm 0,05$	$0,971\pm 0,05$
	0,3	$0,983\pm 0,05$	$0,914\pm 0,08$	$0,933\pm 0,07$	$0,926\pm 0,08$	$0,922\pm 0,08$
	0,4	$0,960\pm 0,09$	$0,858\pm 0,11$	$0,895\pm 0,09$	$0,840\pm 0,12$	$0,847\pm 0,12$
	0,5	$0,507\pm 0,41$	$0,860\pm 0,11$	$0,850\pm 0,12$	$0,742\pm 0,15$	$0,749\pm 0,15$

Tablo 43. Normal dağılım için DVM ile elde edilen ortalama seçicilik oranları

500	0,1	0,999±0,01	0,982±0,02	0,987±0,01	0,986±0,01	0,986±0,01
	0,2	0,991±0,02	0,950±0,03	0,966±0,02	0,958±0,03	0,965±0,02
	0,3	0,982±0,03	0,929±0,03	0,942±0,03	0,923±0,03	0,926±0,03
	0,4	0,961±0,07	0,902±0,04	0,910±0,04	0,892±0,04	0,888±0,04
	0,5	0,524±0,34	0,791±0,26	0,873±0,05	0,836±0,06	0,825±0,11
1000	0,1	0,997±0,01	0,980±0,01	0,986±0,01	0,984±0,01	0,984±0,01
	0,2	0,990±0,02	0,956±0,02	0,968±0,02	0,962±0,02	0,961±0,02
	0,3	0,981±0,03	0,935±0,02	0,945±0,02	0,933±0,02	0,933±0,02
	0,4	0,969±0,06	0,904±0,03	0,917±0,03	0,899±0,03	0,898±0,03
	0,5	0,541±0,31	0,807±0,24	0,880±0,03	0,860±0,04	0,854±0,04
10000	0,1	0,996±0,01	0,985±0,003	0,988±0,003	0,986±0,003	0,983±0,003
	0,2	0,990±0,01	0,964±0,01	0,972±0,01	0,967±0,01	0,964±0,01
	0,3	0,975±0,01	0,939±0,01	0,952±0,01	0,946±0,01	0,944±0,01
	0,4	0,935±0,01	0,909±0,01	0,927±0,01	0,920±0,01	0,920±0,01
	0,5	0,752±0,03	0,874±0,01	0,896±0,01	0,882±0,01	0,891±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Ortalama seçicilik oranları incelendiğinde, gözlem sayısı 100 ve sınıf dağılım oranı 0,1 olduğunda; ham veride 0,999±0,01, minimum-maksimum normalizasyonu sonrasında 0,990±0,03, z-skor normalizasyonu sonrasında 0,990±0,03, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,996±0,02, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,992±0,03 oranlarına ulaşıldı. Sınıf dağılım oranı 0,2 olduğunda ortalama seçicilik oranlarının; ham veride 0,993±0,03, minimum-maksimum normalizasyonu sonrasında 0,957±0,06, z-skor normalizasyonu sonrasında 0,961±0,05, EG diskritizasyonu ile 0,973±0,05, EF diskritizasyonu ile 0,971±0,05 olduğu görüldü. Sınıf dağılım oranı 0,3 iken bu oranlar; ham veri ile 0,983±0,05 olup minimum-maksimum ve z-skor normalizasyonu, EG diskritizasyonu ve EF diskritizasyonu sonrası bu oranlar sırasıyla; 0,914±0,08, 0,933±0,07, 0,926±0,08, 0,922±0,08 oranlarına düştü. Sınıf dağılım oranı 0,4 için elde edilen bu oranlar sırasıyla; 0,96±0,09, 0,858±0,11, 0,895±0,09, 0,84±0,12, 0,847±0,12 olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda ise; ham veride 0,507±0,41, minimum-maksimum normalizasyonu sonrasında 0,860±0,11, z-skor normalizasyonu sonrasında 0,850±0,12, EG diskritizasyonu ile 0,742±0,15, EF diskritizasyonu ile 0,749±0,15 oranları elde edildi.

Gözlem sayısı 500 iken gerçekleştirilen çalışmalar ele alındığında, 0,1 sınıf dağılım oranı için elde edilen ortalama seçicilik oranları; ham veride 0,999±0,01, minimum-maksimum normalizasyonu sonrasında 0,982±0,02, z-skor normalizasyonu sonrasında 0,987±0,01, EG diskritizasyonu ile 0,986±0,01, EF diskritizasyonu ile 0,986±0,01 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda bu oranlar sırasıyla; 0,524±0,34, 0,791±0,26,

0,873±0,05, EG 0,836±0,06, 0,825±0,11 olarak belirlendi. Gözlem sayısı 1000 olduğunda yapılan çalışmalar incelendiğinde, 0,1 sınıf dağılım oranı için elde edilen ortalama seçicilik oranları; ham veride 0,997±0,01 olup minimum-maksimum normalizasyonu, z-skor, EG diskritizasyonu ve EF diskritizasyonu sonrasında bu oran sırasıyla; 0,980±0,01, 0,986±0,01, 0,984±0,01, 0,984±0,01 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise ham veride 0,541±0,31, minimum-maksimum normalizasyonu sonrasında 0,807±0,24, z-skor normalizasyonu sonrasında 0,880±0,03, EG ile 0,86±0,04, EF ile 0,854±0,04 oranları elde edildi. Gözlem sayısı 10000 olduğunda yapılan çalışmada sınıf dağılım oranı 0,1 iken ham veride 0,996±0,01 elde edilen ortalama seçicilik oranı, minimum-maksimum normalizasyonu sonrası 0,985±0,003, z-skor normalizasyonu sonrası 0,988±0,003, EG ile 0,986±0, EF ile 0,983±0,003 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ortalama seçicilik oranları sırasıyla; 0,752±0,03, 0,874±0,01, 0,896±0,01, 0,882±0,01, 0,891±0,01 olarak elde edildi.

Çalışmadaki diğer sınıf dağılımlarında (0,2, 0,3, 0,4), 500, 1000 ve 10000 gözlem sayıları için elde edilen ortalama seçicilik oranlarına Tablo 43'te yer verildi.

Tablo 44. Normal dağılım için DVM ile elde edilen ortalama pozitif kestirim değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,293±0,43	0,837±0,30	0,816±0,33	0,519±0,46	0,560±0,44
	0,2	0,445±0,41	0,825±0,21	0,823±0,22	0,674±0,36	0,739±0,33
	0,3	0,514±0,39	0,811±0,15	0,841±0,15	0,730±0,25	0,746±0,24
	0,4	0,534±0,31	0,813±0,12	0,844±0,12	0,732±0,16	0,750±0,16
	0,5	0,531±0,22	0,865±0,10	0,857±0,10	0,750±0,12	0,754±0,12
500	0,1	0,483±0,44	0,798±0,15	0,826±0,15	0,757±0,22	0,753±0,23
	0,2	0,656±0,28	0,792±0,09	0,842±0,09	0,784±0,11	0,762±0,16
	0,3	0,563±0,33	0,827±0,07	0,853±0,07	0,800±0,07	0,802±0,07
	0,4	0,514±0,37	0,850±0,05	0,862±0,05	0,829±0,06	0,822±0,06
	0,5	0,556±0,15	0,812±0,17	0,874±0,04	0,835±0,07	0,824±0,10
1000	0,1	0,660±0,35	0,781±0,10	0,830±0,10	0,771±0,13	0,755±0,13
	0,2	0,653±0,28	0,808±0,06	0,849±0,06	0,811±0,07	0,802±0,07
	0,3	0,574±0,33	0,837±0,05	0,860±0,05	0,827±0,05	0,822±0,05
	0,4	0,560±0,36	0,853±0,04	0,871±0,04	0,842±0,04	0,839±0,04
	0,5	0,575±0,13	0,830±0,14	0,880±0,03	0,859±0,03	0,852±0,03
10000	0,1	0,407±0,38	0,811±0,03	0,851±0,03	0,827±0,03	0,791±0,03
	0,2	0,414±0,20	0,830±0,02	0,866±0,02	0,847±0,02	0,832±0,02
	0,3	0,516±0,11	0,845±0,01	0,877±0,01	0,862±0,01	0,858±0,01
	0,4	0,607±0,06	0,858±0,01	0,885±0,01	0,873±0,01	0,875±0,01
	0,5	0,674±0,04	0,872±0,01	0,895±0,01	0,881±0,01	0,889±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Ortalama pozitif kestirim deęerleri (PKD) 100 gözlem için incelendięinde, baęımlı deęiřkendeki 0,1 sınıf daęılım oranı için ham veride $0,293\pm 0,43$ olarak bulunurken minimum-maksimum normalizasyonu, z-skor, eřit genişlikte aralıklandırma (EG) diskritizasyonu ve eřit frekansta aralıklandırma (EF) diskritizasyonu sonrasında sırasıyla; $0,837\pm 0,30$, $0,816\pm 0,33$, $0,519\pm 0,46$, $0,56\pm 0,44$ deęerlerine ulařıldı. Sınıf daęılım oranı 0,2 olduęunda; ham veride $0,445\pm 0,41$, minimum-maksimum normalizasyonu sonrasında $0,825\pm 0,21$, z-skor normalizasyonu sonrasında $0,823\pm 0,22$, EG diskritizasyonu ile $0,674\pm 0,36$, EF diskritizasyonu ile $0,739\pm 0,33$ deęerleri elde edildi. Sınıf daęılım oranı 0,3 olduęunda bu deęerler sırasıyla; $0,514\pm 0,39$, $0,811\pm 0,15$, $0,841\pm 0,15$, $0,73\pm 0,25$, $0,746\pm 0,24$ 'tür. Sınıf daęılım oranı 0,4 iken elde edilen ortalama deęerlerin; ham veride $0,534\pm 0,31$, minimum-maksimum normalizasyonu sonrasında $0,813\pm 0,12$, z-skor normalizasyonu sonrasında $0,844\pm 0,12$, EG diskritizasyonu ile $0,732\pm 0,16$, EF diskritizasyonu ile $0,75\pm 0,16$ olduęu gözlemlendi. Son olarak sınıf daęılım oranı 0,5 olduęunda ise bu deęerler sırasıyla; $0,531\pm 0,22$, $0,865\pm 0,10$, $0,857\pm 0,10$, $0,75\pm 0,12$, $0,754\pm 0,12$ olarak elde edildi.

Gözlem sayısı 500 ve sınıf daęılım oranı 0,1 için ortalama PKD; ham veride $0,483\pm 0,44$, minimum-maksimum normalizasyonu sonrası $0,798\pm 0,15$, z-skor normalizasyonu sonrası $0,826\pm 0,15$, EG ve EF diskritizasyonu sonrası ise sırasıyla; $0,757\pm 0,22$, $0,753\pm 0,23$ olarak bulundu. Sınıf daęılım oranı 0,5 olduęunda ortalama PKD için sırasıyla; $0,556\pm 0,15$, $0,812\pm 0,17$, $0,874\pm 0,04$, $0,835\pm 0,07$, $0,824\pm 0,1$ deęerleri elde edildi.

Gözlem sayısı 1000 olduęunda geręekleřtirilen alıřmada 0,1 sınıf daęılım oranı için elde edilen ortalama PKD; ham veride $0,660\pm 0,35$ olup minimum-maksimum normalizasyonu, z-skor, EG diskritizasyonu, EF diskritizasyonu sonrasında sırasıyla $0,781\pm 0,1$, $0,830\pm 0,10$, $0,771\pm 0,13$, $0,755\pm 0,13$ olarak bulundu. Sınıf daęılım oranı 0,5 olduęunda ise bu deęerler sırasıyla; $0,575\pm 0,13$, $0,830\pm 0,14$, $0,880\pm 0,03$, $0,859\pm 0,03$, $0,852\pm 0,03$ olarak elde edildi.

Gözlem sayısı 10000 olduęunda, sınıf daęılım oranı 0,1 iken ortalama PKD; ham veride $0,407\pm 0,38$, minimum-maksimum normalizasyonu sonrasında $0,811\pm 0,03$, z-skor normalizasyonu sonrasında $0,851\pm 0,03$, EG diskritizasyonu ile $0,827\pm 0,03$, EF diskritizasyonu ile $0,791\pm 0,03$ olarak elde edildi. Sınıf daęılım oranı 0,5 olduęunda ise ham veri ile $0,674\pm 0,04$, minimum-maksimum normalizasyonu sonrasında $0,872\pm 0,01$, z-skor normalizasyonu sonrasında $0,895\pm 0,01$, EG diskritizasyonu ile $0,881\pm 0,01$, EF diskritizasyonu ile $0,889\pm 0,01$ deęerlerine ulařıldı.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama pozitif kestirim değerlerine Tablo 44’te yer verilmektedir.

Tablo 45. Normal dağılım için DVM ile elde edilen ortalama negatif kestirim değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,900±0,01	0,932±0,04	0,931±0,04	0,903±0,01	0,907±0,02
	0,2	0,802±0,01	0,913±0,06	0,904±0,05	0,830±0,04	0,845±0,05
	0,3	0,706±0,03	0,897±0,07	0,884±0,07	0,787±0,07	0,803±0,07
	0,4	0,613±0,04	0,889±0,08	0,868±0,08	0,763±0,09	0,776±0,09
	0,5	0,517±0,24	0,840±0,10	0,844±0,10	0,742±0,12	0,747±0,12
500	0,1	0,901±0,004	0,954±0,02	0,949±0,02	0,931±0,02	0,930±0,02
	0,2	0,809±0,02	0,934±0,02	0,925±0,02	0,901±0,03	0,878±0,04
	0,3	0,720±0,04	0,906±0,03	0,905±0,03	0,877±0,03	0,872±0,03
	0,4	0,634±0,07	0,882±0,03	0,886±0,04	0,854±0,04	0,850±0,04
	0,5	0,529±0,17	0,801±0,18	0,869±0,04	0,829±0,06	0,817±0,09
1000	0,1	0,904±0,01	0,958±0,01	0,954±0,01	0,943±0,01	0,938±0,01
	0,2	0,816±0,03	0,932±0,02	0,930±0,02	0,915±0,02	0,910±0,02
	0,3	0,727±0,05	0,905±0,02	0,910±0,02	0,891±0,02	0,885±0,02
	0,4	0,634±0,07	0,883±0,03	0,892±0,03	0,870±0,03	0,865±0,03
	0,5	0,544±0,15	0,823±0,14	0,872±0,03	0,847±0,03	0,842±0,03
10000	0,1	0,911±0,02	0,956±0,005	0,959±0,004	0,958±0,003	0,954±0,004
	0,2	0,815±0,04	0,928±0,01	0,935±0,01	0,932±0,01	0,931±0,01
	0,3	0,720±0,05	0,905±0,01	0,916±0,01	0,911±0,01	0,912±0,01
	0,4	0,631±0,05	0,883±0,01	0,898±0,01	0,888±0,01	0,893±0,01
	0,5	0,611±0,05	0,861±0,01	0,882±0,01	0,872±0,01	0,874±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Ortalama negatif kestirim değerleri (NKD) 100 gözlem için incelendiğinde, bağımlı değişkendeki sınıfların dağılım oranı 0,1 iken ham veri ile 0,900±0,01 değeri elde edilirken minimum-maksimum normalizasyonu sonrasında 0,932±0,04, z-skor normalizasyonu sonrasında 0,931±0,04, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,903±0,01, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,907±0,02 değerlerine ulaşıldı. Sınıf dağılım oranı 0,2 olduğunda bu ortalama değerlerin; ham veride 0,802±0,01, minimum-maksimum normalizasyonu sonrasında 0,913±0,06, z-skor normalizasyonu sonrasında 0,904±0,05, EG diskritizasyonu ile 0,830±0,04 ve EF diskritizasyonu ile 0,845±0,05 olduğu görüldü. Sınıf dağılım oranı 0,3 iken ortalama NKD sırasıyla; 0,706±0,03, 0,897±0,07, 0,884±0,07, 0,787±0,07, 0,803±0,07 olarak hesaplandı. Sınıf dağılım oranı 0,4 iken elde edilen ortalama NKD'nin; ham veride 0,613±0,04, minimum-maksimum normalizasyonu ile 0,889±0,08, z-skor ile 0,868±0,08, EG diskritizasyonu ile 0,763±0,09, EF diskritizasyonu ile 0,776±0,09 olduğu gözlemlendi. Son olarak sınıf dağılım oranı 0,5 olduğunda elde edilen

ortalama NKD'nin sırasıyla; 0,517±0,24, 0,840±0,10, 0,844±0,10, 0,742±0,12, 0,747±0,12 olduğu sonucuna varıldı.

Gözlem sayısı 500 iken gerçekleştirilen çalışmada 0,1 sınıf dağılım oranı için ham veride 0,901±0,004 elde edilen ortalama negatif kestirim değerleri; minimum-maksimum normalizasyonu sonrasında 0,954±0,02, z-skor normalizasyonu sonrasında 0,949±0,02, EG diskritizasyonu ile 0,931±0,02, EF diskritizasyonu ile 0,930±0,02 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama NKD; ham veride 0,529±0,17, minimum-maksimum normalizasyonu sonrasında 0,801±0,18, z-skor normalizasyonu sonrasında 0,869±0,04, EG ile 0,829±0,06, EF ile 0,817±0,09 olarak belirlendi.

Gözlem sayısı 1000 olduğunda gerçekleştirilen sınıflandırma çalışmasında sınıf dağılım oranı 0,1 iken elde edilen ortalama NKD; ham veride 0,904±0,01, minimum-maksimum normalizasyonu sonrasında 0,958±0,01, z-skor normalizasyonu sonrasında 0,954±0,01, EG diskritizasyonu ile 0,943±0,01, EF diskritizasyonu ile 0,938±0,01 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda bu değerler sırasıyla; 0,544±0,15, 0,823±0,14, 0,872±0,03, 0,847±0,03, 0,842±0,03 olarak elde edildi.

Gözlem sayısı 10000 olduğunda yapılan çalışmada sınıf dağılım oranı 0,1 iken elde edilen ortalama NKD; ham veride 0,911±0,02, minimum-maksimum normalizasyonu sonrasında 0,956±0,005, z-skor normalizasyonu sonrasında 0,959±0,004, EG diskritizasyonu ile 0,958±0,003, EF diskritizasyonu ile 0,954±0,004 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise elde edilen ortalama NKD sırasıyla; 0,611±0,05, 0,861±0,01, 0,882±0,01, 0,872±0,01, 0,874±0,01'dir.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama negatif kestirim değerlerine Tablo 45'te yer verilmektedir.

Ki-kare dağılımından üretilen verilerin sınıflandırılması

Tablo 46. Ki-kare dağılımı için DVM ile elde edilen ortalama doğruluk oranları (%)

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	51,461±6,46	57,964±14,04	58,133±13,98	50,350±3,44	50,319±3,52
	0,2	52,928±7,79	71,191±13,85	69,928±13,68	52,478±7,04	51,838±6,08
	0,3	55,326±9,91	78,913±10,47	76,293±10,97	56,745±9,57	56,679±9,48
	0,4	57,452±11,07	81,498±8,94	79,342±9,50	61,773±10,99	62,631±10,94
	0,5	58,595±12,68	81,105±9,18	80,010±9,27	65,395±11,63	66,100±11,22

Tablo 46. Devam Ki-kare dağılımı için DVM ile elde edilen ortalama doğruluk oranları (%)

500	0,1	51,964±4,06	69,500±8,70	67,456±8,89	54,121±5,19	52,731±4,48
	0,2	56,371±6,57	79,725±5,65	76,587±6,19	65,573±5,97	64,118±6,09
	0,3	60,402±9,38	82,616±4,24	80,761±4,69	72,697±5,08	71,628±5,02
	0,4	63,263±11,24	83,225±3,63	82,622±3,86	75,950±4,57	75,446±4,36
	0,5	67,559±12,26	72,740±20,52	83,539±3,66	82,427±3,91	81,925±4,36
1000	0,1	52,601±3,68	72,359±6,53	69,751±6,17	58,902±4,96	58,633±5,13
	0,2	58,532±6,59	80,464±3,69	77,649±4,02	70,354±4,36	70,048±4,20
	0,3	62,769±9,34	82,318±3,15	81,406±3,38	76,037±3,66	75,811±3,60
	0,4	66,622±11,27	83,710±2,77	83,496±2,78	79,009±3,10	78,748±3,12
	0,5	70,770±11,62	68,030±20,72	83,733±2,67	79,659±2,96	79,520±2,96
10000	0,1	55,635±3,16	74,857±1,80	72,102±2,03	68,945±2,05	71,387±2,10
	0,2	62,717±6,53	79,236±1,30	78,803±1,52	76,683±1,35	78,680±1,30
	0,3	66,337±10,39	82,387±1,05	82,173±1,08	80,432±1,10	82,070±1,01
	0,4	69,317±11,9	83,811±0,84	83,963±0,83	82,406±0,89	83,785±0,86
	0,5	69,673±12,84	83,734±1,98	84,474±0,84	82,964±0,87	84,242±0,82

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Ortalama doğruluk oranları 100 gözlem için incelendiğinde, sınıf dağılım oranı 0,1 olduğunda; ham veride %51,461±6,46, minimum-maksimum normalizasyonu sonrasında %57,964±14,04, z-skor normalizasyonu sonrasında %58,133±13,98, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile %50,350±3,44, eşit frekansta aralıklandırma (EF) diskritizasyonu ile %50,319±3,52 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama doğruluk oranlarının; ham veride %52,928±7,79, minimum-maksimum normalizasyonu sonrasında %71,191±13,85, z-skor normalizasyonu sonrasında %69,928±13,68, EG diskritizasyonu ile %52,478±7,04, EF diskritizasyonu ile %51,838±6,08 olduğu görüldü. Sınıf dağılım oranı 0,3 iken bu oranlar; ham veri ile %55,326±9,91 olup minimum-maksimum normalizasyonu, z-skor normalizasyonu, EG diskritizasyonu ve EF diskritizasyonu sonrasında sırasıyla; %78,913±10,47, %76,293±10,97, %56,745±9,57, %56,679±9,48 olarak hesaplandı. Sınıf dağılım oranı 0,4 olduğunda elde edilen ortalama doğruluk oranları sırasıyla; %57,452±11,07, %81,498±8,94, %79,342±9,50, %61,773±10,99, %62,631±10,94 olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda ise ortalama doğruluk oranları için ham veride %58,595±12,68, minimum-maksimum normalizasyonu sonrasında %81,105±9,18, z-skor normalizasyonu sonrasında %80,01±9,27, EG diskritizasyonu ile %65,395±11,63, EF diskritizasyonu ile %66,1±11,22 bulgularına ulaşıldı.

Gözlem sayısı 500 iken gerçekleştirilen sınıflandırma çalışmasında sınıf dağılım oranı 0,1 için elde edilen ortalama doğruluk oranları; ham veride %51,964±4,06, minimum-

maksimum normalizasyonu sonrasında %69,500±8,70, z-skor normalizasyonu sonrasında %67,456±8,89, EG diskritizasyonu ile %54,121±5,19, EF diskritizasyonu ile %52,731±4,48 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise bu oranlar; ham veride %67,559±12,26, minimum-maksimum normalizasyonu sonrasında %72,74±20,52, z-skor normalizasyonu sonrasında %83,539±3,66, EG ile %82,427±3,91, EF ile %81,925±4,36 olarak belirlendi.

Gözlem sayısı 1000 olduğunda yapılan çalışmada, sınıf dağılım oranı 0,1 iken elde edilen ortalama doğruluk oranı; ham veride %52,601±3,68 minimum-maksimum normalizasyonu sonrasında %72,359±6,53, z-skor normalizasyonu sonrasında %69,751±6,17, EG diskritizasyonu ile %58,902±4,96, EF diskritizasyonu ile %58,633±5,13 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ortalama doğruluk oranları; ham veride %70,77±11,62, minimum-maksimum normalizasyonu sonrasında %68,03±20,72, z-skor normalizasyonu sonrasında %83,733±2,67, EG ile %79,659±2,96, EF ile %79,52±2,96 olarak elde edildi.

Gözlem sayısı 10000 olduğunda yapılan çalışmada, 0,1 sınıf dağılım oranı için elde edilen ortalama doğruluk oranı; ham veride %55,635±3,16, minimum-maksimum normalizasyonu sonrasında %74,857±1,80, z-skor normalizasyonu sonrasında %72,102±2,03, EG diskritizasyonu ile %68,945±2,05, EF diskritizasyonu ile %71,387±2,10 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise ortalama doğruluk oranlarının; ham veride %69,673±12,84, minimum-maksimum normalizasyonu sonrasında %83,734±1,98, z-skor normalizasyonu sonrasında %84,474±0,84, EG diskritizasyonu ile %82,964±0,87, EF diskritizasyonu ile %84,242±0,82 olduğu saptandı.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama doğruluk oranlarına Tablo 46'da yer verildi.

Tablo 47. Ki-kare dağılımı için DVM ile elde edilen ortalama AUC değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,616±0,26	0,850±0,16	0,854±0,16	0,582±0,24	0,574±0,25
	0,2	0,632±0,19	0,888±0,10	0,876±0,11	0,659±0,19	0,661±0,18
	0,3	0,627±0,17	0,898±0,08	0,883±0,09	0,696±0,15	0,704±0,15
	0,4	0,609±0,17	0,901±0,07	0,883±0,08	0,706±0,15	0,716±0,14
	0,5	0,598±0,16	0,896±0,08	0,882±0,08	0,706±0,14	0,720±0,13
500	0,1	0,664±0,15	0,909±0,05	0,895±0,06	0,797±0,08	0,779±0,08
	0,2	0,660±0,17	0,916±0,03	0,906±0,04	0,831±0,05	0,823±0,05
	0,3	0,652±0,17	0,915±0,03	0,908±0,03	0,844±0,04	0,837±0,04
	0,4	0,656±0,18	0,916±0,03	0,910±0,03	0,848±0,04	0,845±0,04
	0,5	0,696±0,18	0,701±0,37	0,914±0,03	0,900±0,04	0,894±0,05

Tablo 47. Ki-kare dağılımı için DVM ile elde edilen ortalama AUC değerleri

1000	0,1	0,655±0,17	0,917±0,03	0,904±0,04	0,843±0,05	0,843±0,05
	0,2	0,681±0,17	0,916±0,02	0,911±0,03	0,862±0,03	0,863±0,03
	0,3	0,676±0,19	0,918±0,02	0,914±0,02	0,872±0,03	0,872±0,03
	0,4	0,689±0,19	0,920±0,02	0,917±0,02	0,878±0,03	0,877±0,03
	0,5	0,733±0,18	0,597±0,41	0,917±0,02	0,878±0,03	0,877±0,03
10000	0,1	0,716±0,19	0,921±0,01	0,911±0,03	0,903±0,01	0,913±0,01
	0,2	0,724±0,20	0,921±0,01	0,918±0,01	0,908±0,01	0,919±0,01
	0,3	0,692±0,24	0,921±0,01	0,920±0,01	0,909±0,01	0,920±0,01
	0,4	0,702±0,24	0,921±0,01	0,922±0,01	0,910±0,01	0,922±0,01
	0,5	0,697±0,24	0,921±0,01	0,922±0,01	0,911±0,01	0,922±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Gerçekleştirilen sınıflandırma çalışmasında 100 gözlem için elde edilen ortalama AUC değeri sınıf dağılım oranı 0,1 olduğunda; ham veride 0,616±0,26, minimum-maksimum normalizasyonu sonrasında 0,85±0,16, z-skor normalizasyonu sonrasında 0,854±0,16, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,582±0,24, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,574±0,25 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama AUC değeri; ham veride 0,632±0,19 olarak bulunurken minimum-maksimum normalizasyonu, z-skor, EG diskritizasyonu, EF diskritizasyonu sonrasında sırasıyla; 0,888±0,1, 0,876±0,11, 0,659±0,19, 0,661±0,18 değerlerine ulaşıldı. Sınıf dağılım oranı 0,3 sırasıyla bu değerler; 0,627±0,17, 0,898±0,08, 0,883±0,09, 0,696±0,15, 0,704±0,15 olarak elde edildi. Sınıf dağılım oranı 0,4 iken ortalama AUC değerlerinin; ham veride 0,609±0,17, minimum-maksimum normalizasyonu ile 0,901±0,07, z-skor ile 0,883±0,08, EG diskritizasyonu ile 0,706±0,15, EF diskritizasyonu ile 0,716±0,14 olduğu gözlemlendi. Son olarak sınıf dağılım oranı 0,5 olduğunda ise elde edilen ortalama AUC değerleri sırasıyla; 0,598±0,16, 0,896±0,08, 0,882±0,08, 0,706±0,14, 0,72±0,13'dir.

Gözlem sayısı 500 olduğunda yapılan çalışmada 0,1 sınıf dağılım oranı için ortalama AUC i; ham veride 0,664±0,15, minimum-maksimum normalizasyonu sonrasında 0,909±0,05, z-skor normalizasyonu sonrasında 0,895±0,06, EG diskritizasyonu ile 0,797±0,08, EF diskritizasyonu ile 0,779±0,08 olarak elde edildi. Sınıf dağılım oranı 0,5 olduğunda ortalama AUC değerleri sırasıyla; 0,696±0,18, 0,701±0,37, 0,914±0,03, 0,9±0,04, 0,894±0,05 olarak bulundu.

Gözlem sayısı 1000 olduğunda gerçekleştirilen sınıflandırma çalışmasında sınıf dağılım oranı 0,1 iken elde edilen ortalama AUC değerleri; ham veride %0,655±0,17, minimum-maksimum normalizasyonu sonrasında %0,917±0,03, z-skor normalizasyonu

sonrasında %0,904±0,04, EG diskritizasyonu ile %0,843±0,05, EF diskritizasyonu ile %0,843±0,05 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise ortalama değerler; ham veride 0,733±0,18, minimum-maksimum normalizasyonu sonrası 0,597±0,41, z-skor normalizasyonu sonrası 0,917±0,02, EG ile 0,878±0,03, EF ile 0,877±0,03 olarak elde edildi.

Gözlem sayısı 10000 olduğunda, sınıf dağılım oranı 0,1 iken elde edilen ortalama AUC değerleri; ham veride 0,716±0,19, minimum-maksimum normalizasyonu sonrasında 0,921±0,01, z-skor normalizasyonu sonrasında 0,911±0,03, EG diskritizasyonu ile 0,903±0,01, EF diskritizasyonu ile 0,913±0,01 olarak bulundu. Sınıfların tam dengeli (0,5) dağılımı için ortalama AUC değeri; ham veri ile 0,697±0,24 olarak bulunurken minimum-maksimum normalizasyonu, z-skor, EG diskritizasyonu, EF diskritizasyonu sonrasında sırasıyla; 0,921±0,01, 0,922±0,01, 0,911±0,01, 0,922±0,01 olduğu bulgusuna ulaşıldı.

Çalışmadaki diğer sınıf dağılımlarında (0,2, 0,3, 0,4), 500, 1000 ve 10000 gözlem sayıları için elde edilen ortalama AUC değerlerine Tablo 47’de yer verildi.

Tablo 48. Ki-kare dağılımı için DVM ile elde edilen ortalama duyarlılık oranları

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,036±0,14	0,168±0,29	0,172±0,29	0,010±0,07	0,009±0,07
	0,2	0,078±0,17	0,468±0,28	0,440±0,28	0,066±0,15	0,052±0,13
	0,3	0,145±0,22	0,679±0,21	0,609±0,21	0,188±0,21	0,195±0,21
	0,4	0,216±0,28	0,793±0,16	0,716±0,17	0,378±0,24	0,404±0,23
	0,5	0,627±0,40	0,821±0,13	0,803±0,13	0,646±0,17	0,661±0,17
500	0,1	0,046±0,09	0,408±0,18	0,365±0,18	0,090±0,11	0,060±0,09
	0,2	0,153±0,15	0,658±0,11	0,575±0,12	0,353±0,12	0,322±0,13
	0,3	0,264±0,23	0,758±0,08	0,692±0,09	0,547±0,10	0,526±0,10
	0,4	0,365±0,30	0,790±0,06	0,770±0,07	0,671±0,08	0,666±0,08
	0,5	0,680±0,30	0,731±0,30	0,835±0,06	0,822±0,06	0,820±0,06
1000	0,1	0,058±0,08	0,468±0,13	0,411±0,13	0,190±0,10	0,183±0,11
	0,2	0,201±0,15	0,674±0,07	0,597±0,08	0,454±0,09	0,446±0,09
	0,3	0,320±0,23	0,733±0,06	0,703±0,07	0,611±0,07	0,606±0,07
	0,4	0,444±0,30	0,800±0,05	0,786±0,05	0,718±0,05	0,715±0,06
	0,5	0,707±0,26	0,684±0,36	0,837±0,04	0,794±0,04	0,794±0,04
10000	0,1	0,121±0,07	0,522±0,04	0,458±0,04	0,397±0,04	0,447±0,04
	0,2	0,288±0,15	0,632±0,03	0,618±0,03	0,581±0,03	0,619±0,03
	0,3	0,391±0,25	0,730±0,03	0,716±0,02	0,691±0,02	0,717±0,02
	0,4	0,495±0,30	0,802±0,03	0,789±0,02	0,771±0,02	0,788±0,02
	0,5	0,739±0,25	0,849±0,03	0,846±0,01	0,831±0,01	0,841±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Sınıflandırma çalışmasında 100 gözlem için elde edilen ortalama duyarlılık oranları, sınıf dağılım oranı 0,1 olduğunda; ham veride $0,036\pm 0,14$, minimum-maksimum normalizasyonu sonrasında $0,168\pm 0,29$, z-skor normalizasyonu sonrasında $0,172\pm 0,29$, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile $0,01\pm 0,07$, eşit frekansta aralıklandırma (EF) diskritizasyonu ile $0,009\pm 0,07$ olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama duyarlılık oranlarının; ham veride $0,078\pm 0,17$, olarak bulunurken minimum-maksimum normalizasyonu sonrasında $0,468\pm 0,28$, z-skor normalizasyonu sonrasında $0,44\pm 0,28$, EG ile $0,066\pm 0,15$, EF ile $0,052\pm 0,13$ olduğu görüldü. Sınıf dağılım oranı 0,3 iken bu oranlar; ham veride $0,145\pm 0,22$ olup minimum-maksimum normalizasyonu sonrasında $0,679\pm 0,21$, z-skor normalizasyonu sonrasında $0,609\pm 0,21$, EG ile $0,188\pm 0,21$, EF ile $0,195\pm 0,21$ 'dir. Sınıf dağılım oranı 0,4 iken elde edilen bu oranlar; ham veride $0,216\pm 0,28$, minimum-maksimum normalizasyonu sonrasında $0,793\pm 0,16$, z-skor normalizasyonu sonrasında $0,716\pm 0,17$, EG ile $0,378\pm 0,24$, EF ile $0,404\pm 0,23$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise ortalama duyarlılık oranları; ham veride $0,627\pm 0,4$, minimum-maksimum normalizasyonu sonrasında $0,821\pm 0,13$, z-skor normalizasyonu sonrasında $0,803\pm 0,13$, EG ile $0,646\pm 0,17$, EF ile $0,661\pm 0,17$ değerlerine yükseldi.

Gözlem sayısı 500 olduğunda yapılan çalışmada, sınıf dağılım oranı 0,1 için elde edilen ortalama duyarlılık oranları; ham veride $0,046\pm 0,09$, minimum-maksimum normalizasyonu sonrasında $0,408\pm 0,18$, z-skor normalizasyonu sonrasında $0,365\pm 0,18$, EG ile $0,09\pm 0,11$, EF ile $0,06\pm 0,09$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise ortalama duyarlılık oranlarında artış gözlenerek; ham veride $0,68\pm 0,3$, minimum-maksimum normalizasyonu sonrasında $0,731\pm 0,3$, z-skor normalizasyonu sonrasında $0,835\pm 0,06$, EG ile $0,822\pm 0,06$, EF ile $0,82\pm 0,06$ olarak elde edildi. Gözlem sayısı 1000 ve sınıf dağılım oranı 0,1 iken elde edilen ortalama duyarlılık oranları; ham veride $0,058\pm 0,08$, minimum-maksimum normalizasyonu sonrasında $0,468\pm 0,13$, z-skor normalizasyonu sonrasında $0,411\pm 0,13$, EG ile $0,19\pm 0,1$, EF ile $0,183\pm 0,11$ olarak bulundu. Sınıf dağılımı 0,5 oranına yükseldiğinde bu oranlar sırasıyla; $0,707\pm 0,26$, $0,684\pm 0,36$, $0,837\pm 0,04$, $0,794\pm 0,04$, $0,794\pm 0,04$ olarak elde edildi. Gözlem sayısı 10000 olduğunda, sınıf dağılım oranı 0,1 iken elde edilen ortalama duyarlılık oranları; ham veride $0,121\pm 0,07$ olarak bulunurken minimum-maksimum normalizasyonu, z-skor, EG ve EF ile sırasıyla; $0,522\pm 0,04$, $0,458\pm 0,04$, $0,397\pm 0,04$, $0,447\pm 0,04$ değerlerine ulaşıldı. Sınıfların 0,5 dağılım oranı için ortalama duyarlılık oranı sırasıyla $0,739\pm 0,25$, $0,849\pm 0,03$, $0,846\pm 0,01$, $0,831\pm 0,01$, $0,841\pm 0,01$ olarak bulundu.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama duyarlılık oranlarına Tablo 48’de yer verildi.

Tablo 49. Ki-kare dağılımı için DVM ile elde edilen ortalama seçicilik oranları

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,993±0,02	0,991±0,03	0,990±0,03	0,998±0,01	0,997±0,02
	0,2	0,980±0,05	0,956±0,06	0,959±0,06	0,984±0,04	0,985±0,04
	0,3	0,961±0,07	0,899±0,09	0,917±0,08	0,947±0,08	0,938±0,08
	0,4	0,933±0,11	0,837±0,11	0,871±0,10	0,858±0,13	0,848±0,13
	0,5	0,544±0,41	0,801±0,14	0,797±0,14	0,662±0,17	0,661±0,17
500	0,1	0,993±0,01	0,982±0,02	0,984±0,01	0,993±0,01	0,995±0,01
	0,2	0,974±0,03	0,936±0,03	0,956±0,02	0,958±0,03	0,960±0,03
	0,3	0,944±0,06	0,894±0,04	0,923±0,03	0,907±0,04	0,907±0,04
	0,4	0,900±0,09	0,874±0,05	0,883±0,04	0,848±0,05	0,843±0,05
	0,5	0,672±0,31	0,724±0,30	0,836±0,05	0,826±0,05	0,818±0,06
1000	0,1	0,994±0,01	0,979±0,01	0,984±0,01	0,988±0,01	0,990±0,01
	0,2	0,969±0,03	0,935±0,02	0,956±0,02	0,953±0,02	0,955±0,02
	0,3	0,935±0,05	0,913±0,03	0,925±0,02	0,910±0,03	0,911±0,03
	0,4	0,888±0,08	0,874±0,04	0,884±0,03	0,862±0,03	0,860±0,03
	0,5	0,708±0,27	0,676±0,36	0,837±0,04	0,799±0,04	0,797±0,04
10000	0,1	0,992±0,01	0,975±0,004	0,984±0,003	0,982±0,003	0,981±0,003
	0,2	0,966±0,02	0,953±0,01	0,958±0,01	0,953±0,01	0,955±0,01
	0,3	0,936±0,04	0,917±0,01	0,928±0,01	0,918±0,01	0,924±0,01
	0,4	0,891±0,07	0,874±0,03	0,890±0,01	0,877±0,01	0,888±0,01
	0,5	0,654±0,31	0,826±0,06	0,844±0,01	0,829±0,01	0,844±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Ortalama seçicilik oranları 100 gözlem için incelendiğinde, 0,1 sınıf dağılım oranı için ham veride 0,993±0,02, minimum-maksimum normalizasyonu sonrasında 0,991±0,03, z-skor normalizasyonu sonrasında 0,99±0,03, eşit genişlikte aralıklandırma (EG) diskritizasyonu sonrasında 0,998±0,01, eşit frekansta aralıklandırma (EF) diskritizasyonu sonrasında 0,997±0,02 oranlarına ulaşıldı. Sınıf dağılım oranı 0,2 olduğunda ortalama seçicilik oranlarının; ham veride 0,98±0,05, olarak bulurken minimum-maksimum normalizasyonu sonrasında 0,956±0,06, z-skor normalizasyonu sonrasında 0,959±0,06, EG diskritizasyonu sonrasında 0,984±0,04, EF diskritizasyonu sonrasında 0,985±0,04 olduğu görüldü. Sınıf dağılım oranı 0,3 iken bu oranlar; ham veri ile 0,961±0,07 olup minimum-maksimum normalizasyonu, z-skor normalizasyonu, EG diskritizasyonu ve EF diskritizasyonu sonrasında sırasıyla; 0,899±0,09, 0,917±0,08, 0,947±0,08, 0,938±0,08 olarak hesaplandı. 0,4 sınıf dağılım oranı için bu oranlar; ham veride 0,933±0,11, minimum-maksimum normalizasyonu sonrasında 0,837±0,11, z-skor normalizasyonu sonrasında 0,871±0,1, EG diskritizasyonu

sonrasında $0,858\pm 0,13$, EF diskritizasyonu sonrasında $0,848\pm 0,13$ olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda elde edilen bu oranların; ham veride $0,544\pm 0,41$, minimum-maksimum normalizasyonu sonrasında $0,801\pm 0,14$, z-skor normalizasyonu sonrasında $0,797\pm 0,14$, EG diskritizasyonu ile $0,662\pm 0,17$, EF diskritizasyonu ile $0,661\pm 0,17$ olduğu görüldü.

Gözlem sayısı 500 olduğunda, sınıf dağılım oranı 0,1 için elde edilen ortalama seçicilik oranları; ham veride $0,993\pm 0,01$, minimum-maksimum normalizasyonu sonrasında $0,982\pm 0,02$, z-skor normalizasyonu sonrasında $0,984\pm 0,01$, EG diskritizasyonu ile $0,993\pm 0,01$, EF diskritizasyonu ile $0,995\pm 0,01$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise ortalama seçicilik oranlarında azalma olup ham veride $0,672\pm 0,31$, minimum-maksimum normalizasyonu sonrasında $0,724\pm 0,3$, z-skor normalizasyonu sonrasında $0,836\pm 0,05$, EG diskritizasyonu ile $0,826\pm 0,05$, EF diskritizasyonu ile $0,818\pm 0,06$ olarak belirlendi.

Gözlem sayısı 1000 olduğunda, sınıf dağılım oranı 0,1 iken elde edilen ortalama seçicilik oranları; ham veride $0,994\pm 0,01$, minimum-maksimum normalizasyonu sonrasında $0,979\pm 0,01$, z-skor normalizasyonu sonrasında $0,984\pm 0,01$, EG diskritizasyonu sonrasında $0,988\pm 0,01$, EF diskritizasyonu sonrasında $0,99\pm 0,01$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise ham veride $0,708\pm 0,27$, minimum-maksimum normalizasyonu sonrasında $0,676\pm 0,36$, z-skor normalizasyonu sonrasında $0,837\pm 0,04$, EG diskritizasyonu sonrasında $0,799\pm 0,04$, EF diskritizasyonu sonrasında $0,797\pm 0,04$ olarak elde edildi.

Gözlem sayısı 10000 olduğunda yapılan sınıflandırma çalışmasında sınıf dağılım oranı 0,1 iken elde edilen ortalama seçicilik oranları; ham veride $0,992\pm 0,01$, minimum-maksimum normalizasyonu sonrasında $0,975\pm 0,004$, z-skor normalizasyonu sonrasında $0,984\pm 0,003$, EG diskritizasyonu ile $0,982\pm 0,003$, EF diskritizasyonu ile $0,981\pm 0,003$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ortalama seçicilik oranları; ham veride $0,654\pm 0,31$, minimum-maksimum normalizasyonu sonrasında $0,826\pm 0,06$, z-skor normalizasyonu sonrasında $0,844\pm 0,01$, EG diskritizasyonu sonrasında $0,829\pm 0,01$, EF diskritizasyonu sonrasında $0,844\pm 0,01$ olarak elde edildi.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama seçicilik oranlarına Tablo 49'da yer verildi.

Tablo 50. Ki-kare dağılımı için DVM ile elde edilen ortalama pozitif kestirim değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,380±0,43	0,718±0,37	0,699±0,39	0,310±0,43	0,300±0,44
	0,2	0,498±0,40	0,771±0,26	0,772±0,27	0,534±0,41	0,479±0,43
	0,3	0,622±0,33	0,770±0,17	0,782±0,18	0,636±0,32	0,597±0,32
	0,4	0,706±0,20	0,782±0,13	0,802±0,14	0,661±0,23	0,662±0,22
	0,5	0,611±0,18	0,817±0,11	0,810±0,11	0,668±0,13	0,669±0,13
500	0,1	0,434±0,38	0,744±0,19	0,740±0,21	0,611±0,38	0,579±0,40
	0,2	0,606±0,20	0,730±0,10	0,775±0,11	0,692±0,15	0,684±0,17
	0,3	0,674±0,10	0,759±0,07	0,798±0,07	0,721±0,09	0,714±0,09
	0,4	0,712±0,07	0,810±0,06	0,817±0,06	0,749±0,07	0,742±0,06
	0,5	0,702±0,13	0,746±0,22	0,837±0,04	0,827±0,05	0,821±0,05
1000	0,1	0,513±0,32	0,729±0,13	0,753±0,13	0,667±0,23	0,689±0,23
	0,2	0,628±0,12	0,726±0,06	0,778±0,07	0,712±0,09	0,717±0,09
	0,3	0,682±0,07	0,786±0,05	0,803±0,05	0,747±0,06	0,746±0,06
	0,4	0,727±0,05	0,812±0,04	0,820±0,04	0,778±0,04	0,774±0,04
	0,5	0,721±0,14	0,711±0,22	0,839±0,03	0,799±0,04	0,797±0,04
10000	0,1	0,624±0,07	0,699±0,04	0,759±0,04	0,709±0,04	0,726±0,04
	0,2	0,682±0,03	0,770±0,02	0,789±0,02	0,755±0,02	0,774±0,02
	0,3	0,721±0,05	0,792±0,02	0,810±0,02	0,784±0,02	0,802±0,02
	0,4	0,744±0,08	0,811±0,03	0,827±0,01	0,807±0,01	0,824±0,01
	0,5	0,682±0,16	0,833±0,04	0,844±0,01	0,829±0,01	0,844±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Sınıflandırma çalışmasında 100 gözlem için elde edilen ortalama pozitif kestirim değeri (PKD) sınıf dağılım oranı 0,1 olduğunda; ham veride 0,38±0,43, minimum-maksimum normalizasyonu sonrasında 0,718±0,37, z-skor normalizasyonu sonrasında 0,699±0,39, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,31±0,43, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,3±0,44 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama PKD'nin; ham veride 0,498±0,4, olarak bulurken minimum-maksimum normalizasyonu sonrasında 0,771±0,26, z-skor normalizasyonu sonrasında 0,772±0,27, EG diskritizasyonu ile 0,534±0,41, EF diskritizasyonu ile 0,479±0,43 olduğu görüldü. Sınıf dağılım oranı 0,3 iken bu değerler; ham veri ile 0,622±0,33 olup minimum-maksimum normalizasyonu sonrasında 0,77±0,17, z-skor normalizasyonu sonrasında 0,782±0,18, EG diskritizasyonu ile 0,636±0,32, EF diskritizasyonu ile 0,597±0,32 olarak hesaplandı. Sınıf dağılım oranı 0,4 iken elde edilen ortalama değerlerin; ham veride 0,706±0,2, minimum-maksimum normalizasyonu sonrasında 0,782±0,13, z-skor normalizasyonu sonrasında 0,802±0,14, EG diskritizasyonu ile 0,661±0,23, EF diskritizasyonu ile 0,662±0,22 olduğu gözlemlendi. Son olarak sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama PKD'nin; ham veride 0,611±0,18, minimum-maksimum normalizasyonu sonrasında 0,817±0,11, z-skor normalizasyonu sonrasında

0,81±0,11, EG diskritizasyonu ile 0,668±0,13, EF diskritizasyonu ile 0,669±0,13 olduğu sonucuna varıldı.

Gözlem sayısı 500 iken gerçekleştirilen sınıflandırma çalışmasında sınıf dağılım oranı 0,1 için elde edilen ortalama PKD; ham veride 0,434±0,38, minimum-maksimum normalizasyonu sonrasında 0,744±0,19, z-skor normalizasyonu sonrasında 0,74±0,21, EG diskritizasyonu ile 0,611±0,38, EF diskritizasyonu ile 0,579±0,4 olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda ortalama PKD; ham veride 0,702±0,13, minimum-maksimum normalizasyonu sonrasında 0,746±0,22, z-skor normalizasyonu sonrasında 0,837±0,04, EG ile 0,827±0,05, EF ile 0,821±0,05 olarak elde edildi.

Gözlem sayısı 1000 olduğunda, 0,1 sınıf dağılım oranı için ortalama PKD; ham veride 0,513±0,32, minimum-maksimum normalizasyonu sonrasında 0,729±0,13, z-skor normalizasyonu sonrasında 0,753±0,13, EG diskritizasyonu ile 0,667±0,23, EF diskritizasyonu ile 0,689±0,23 olarak elde edildi. Sınıf dağılım oranı 0,5 olduğunda ise ortalama PKD; ham veride 0,721±0,14, minimum-maksimum normalizasyonu sonrasında 0,711±0,22, z-skor normalizasyonu sonrasında 0,839±0,03, EG diskritizasyonu ile 0,799±0,04, EF diskritizasyonu ile 0,797±0,04 olarak bulundu.

Gözlem sayısı 10000 olduğunda, 0,1 sınıf dağılım oranı için ortalama PKD; ham veride 0,624±0,07, minimum-maksimum normalizasyonu sonrasında 0,699±0,04, z-skor normalizasyonu sonrasında 0,759±0,04, EG diskritizasyonu ile 0,709±0,04, EF diskritizasyonu ile 0,726±0,04 olarak bulundu. Bu sonuçlarda sınıf dağılım oranı 0,5 artış gözlenmekte olup ham veri ile 0,682±0,16, minimum-maksimum normalizasyonu sonrasında 0,833±0,04, z-skor normalizasyonu sonrasında 0,844±0,01, EG diskritizasyonu ile 0,829±0,01, EF diskritizasyonu ile 0,844±0,01 değerlerine ulaşıldı.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama pozitif kestirim değerlerine Tablo 50’de yer verildi.

Tablo 51. Ki-kare dağılımı için DVM ile elde edilen ortalama negatif kestirim değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,903±0,01	0,916±0,03	0,916±0,03	0,901±0,01	0,901±0,01
	0,2	0,811±0,03	0,882±0,06	0,876±0,06	0,809±0,03	0,807±0,02
	0,3	0,729±0,06	0,874±0,07	0,851±0,07	0,735±0,05	0,735±0,05
	0,4	0,654±0,08	0,868±0,09	0,829±0,09	0,684±0,09	0,690±0,09
	0,5	0,628±0,18	0,827±0,11	0,811±0,11	0,660±0,13	0,671±0,13

Tablo 51. Devam Ki-kare dağılımı için DVM ile elde edilen ortalama negatif kestirim değerleri

500	0,1	0,904±0,01	0,938±0,02	0,933±0,02	0,908±0,01	0,905±0,01
	0,2	0,823±0,02	0,917±0,03	0,901±0,03	0,856±0,02	0,851±0,02
	0,3	0,755±0,05	0,897±0,03	0,876±0,03	0,825±0,03	0,818±0,03
	0,4	0,697±0,08	0,864±0,04	0,853±0,04	0,796±0,04	0,793±0,04
	0,5	0,702±0,13	0,758±0,20	0,838±0,05	0,825±0,05	0,822±0,05
1000	0,1	0,905±0,01	0,943±0,01	0,938±0,01	0,917±0,01	0,916±0,01
	0,2	0,830±0,02	0,920±0,02	0,905±0,02	0,875±0,02	0,874±0,02
	0,3	0,768±0,05	0,889±0,02	0,880±0,02	0,846±0,02	0,844±0,02
	0,4	0,723±0,09	0,869±0,03	0,862±0,03	0,822±0,03	0,820±0,03
	0,5	0,720±0,14	0,723±0,21	0,838±0,03	0,796±0,03	0,795±0,03
10000	0,1	0,910±0,01	0,948±0,004	0,942±0,004	0,936±0,004	0,941±0,004
	0,2	0,846±0,02	0,912±0,01	0,909±0,01	0,901±0,01	0,909±0,01
	0,3	0,789±0,06	0,888±0,01	0,884±0,01	0,874±0,01	0,884±0,01
	0,4	0,745±0,09	0,869±0,01	0,864±0,01	0,852±0,01	0,863±0,01
	0,5	0,658±0,20	0,847±0,03	0,846±0,01	0,830±0,01	0,842±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Sınıflandırma çalışmasında 100 gözlem için ortalama negatif kestirim değerleri (NKD), sınıf dağılım oranı 0,1 olduğunda; ham veride 0,903±0,01, minimum-maksimum normalizasyonu sonrasında 0,916±0,03, z-skor normalizasyonu sonrasında 0,916±0,03, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,901±0,01, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,901±0,01 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda bu ortalama değerlerin sırasıyla; 0,811±0,03, 0,882±0,06, 0,876±0,06, 0,809±0,03, 0,807±0,02 olduğu görüldü. Sınıf dağılım oranı 0,3 iken ortalama NKD sırasıyla; 0,729±0,06, 0,874±0,07, 0,851±0,07, 0,735±0,05, 0,735±0,05 olarak hesaplandı. Sınıf dağılım oranı 0,4 için ortalama NKD'nin sırasıyla; 0,654±0,08, 0,868±0,09, 0,829±0,09, 0,684±0,09, 0,69±0,09 olduğu gözlemlendi. Son olarak sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama NKD'nin; ham veride 0,628±0,18, minimum-maksimum normalizasyonu sonrasında 0,827±0,11, z-skor normalizasyonu sonrasında 0,811±0,11, EG diskritizasyonu ile 0,66±0,13, EF diskritizasyonu ile 0,671±0,13 olduğu sonucuna varıldı.

Gözlem sayısı 500 olduğunda, 0,1 sınıf dağılım oranı için ortalama negatif kestirim değerleri; ham veride 0,904±0,01, minimum-maksimum normalizasyonu sonrasında 0,938±0,02, z-skor normalizasyonu sonrasında 0,933±0,02, EG diskritizasyonu ile 0,908±0,01, EF diskritizasyonu ile 0,905±0,01 olarak elde edildi. Sınıf dağılım oranı 0,5 olduğunda ortalama NKD sırasıyla; 0,702±0,13, 0,758±0,2, 0,838±0,05, 0,825±0,05, 0,822±0,05'dir.

Gözlem sayısı 1000 olduğunda, sınıf dağılım oranı 0,1 iken elde edilen ortalama NKD; ham veride $0,905 \pm 0,01$, minimum-maksimum normalizasyonu sonrasında $0,943 \pm 0,01$, z-skor normalizasyonu sonrasında $0,938 \pm 0,01$, EG diskritizasyonu ile $0,917 \pm 0,01$, EF diskritizasyonu ile $0,916 \pm 0,01$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda değerler sırasıyla; $0,72 \pm 0,14$, $0,723 \pm 0,21$, $0,838 \pm 0,03$, $0,796 \pm 0,03$, $0,795 \pm 0,03$ 'tür.

Gözlem sayısı 10000 olduğunda yapılan çalışmada, sınıf dağılım oranı 0,1 iken elde edilen ortalama NKD; ham veride $0,91 \pm 0,01$, minimum-maksimum normalizasyonu ile $0,948 \pm 0,004$, z-skor normalizasyonu ile $0,942 \pm 0,004$, EG ile $0,936 \pm 0,004$, EF ile $0,941 \pm 0,004$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda sırasıyla; $0,658 \pm 0,2$, $0,847 \pm 0,03$, $0,846 \pm 0,01$, $0,83 \pm 0,01$, $0,842 \pm 0,01$ değerlerine ulaşıldı.

Çalışmadaki diğer sınıf dağılımlarında (0,2, 0,3, 0,4), 500, 1000 ve 10000 gözlem sayıları için elde edilen ortalama negatif kestirim değerlerine Tablo 51'de yer verildi.

F dağılımından üretilen verilerin sınıflandırılması

Tablo 52. F dağılımı için DVM ile elde edilen ortalama doğruluk oranları (%)

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	57,100±13,13	55,733±12,16	57,083±13,10	51,153±6,43	55,128±12,17
	0,2	60,741±11,63	59,484±10,96	60,219±11,27	55,266±9,68	67,047±13,34
	0,3	63,713±10,87	62,111±10,58	63,518±10,78	60,355±11,43	73,125±11,37
	0,4	67,371±10,98	62,308±9,63	66,410±10,52	65,604±11,52	76,379±9,98
	0,5	69,745±11,06	53,865±15,03	68,470±11,90	68,955±11,89	77,190±9,39
500	0,1	61,391±7,95	60,480±7,66	62,508±7,90	56,666±6,46	68,877±8,23
	0,2	65,774±6,63	63,778±5,85	68,951±6,66	60,903±6,02	79,394±5,67
	0,3	71,074±5,77	66,265±5,06	75,191±6,12	65,162±5,86	83,813±4,45
	0,4	77,061±5,13	67,918±4,54	78,821±5,06	68,288±5,45	85,450±3,72
	0,5	77,778±4,26	63,360±7,92	80,189±4,61	68,880±6,39	85,021±7,07
1000	0,1	62,444±6,12	61,584±5,77	65,403±6,08	58,081±5,17	72,972±5,96
	0,2	69,063±4,77	64,451±3,99	74,353±4,98	61,978±4,48	82,373±3,92
	0,3	75,065±4,09	66,567±3,30	80,070±3,87	64,738±4,48	86,420±2,80
	0,4	81,271±3,30	68,078±3,20	82,349±3,02	67,476±4,36	87,679±2,43
	0,5	80,720±2,90	64,862±4,42	83,103±3,01	68,634±4,20	88,107±2,29
10000	0,1	73,810±1,90	62,086±1,86	83,059±1,31	59,223±3,56	81,722±1,78
	0,2	81,093±1,29	64,141±1,40	83,059±1,31	59,078±2,95	87,908±1,06
	0,3	86,964±0,91	66,608±1,10	85,898±1,01	59,206±2,63	90,256±0,78
	0,4	87,536±0,77	68,345±1,05	87,359±0,88	60,334±2,38	91,163±0,68
	0,5	87,200±0,77	65,475±1,31	87,875±0,83	61,958±1,71	91,405±0,67

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Sınıflandırma çalışmasında 100 gözlem için elde edilen ortalama doğruluk oranları, 0,1 sınıf dağılım oranı için ham veride %57,100±13,13, minimum-maksimum normalizasyonu sonrasında %55,733±12,16, z-skor normalizasyonu sonrasında %57,083±13,10, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile %51,153±6,43, eşit frekansta aralıklandırma (EF) diskritizasyonu ile %55,128±12,17 oranları elde edildi. Sınıf dağılım oranı 0,2 olduğunda ortalama doğruluk oranlarının; ham veride %60,741±11,63, minimum-maksimum normalizasyonu sonrasında %59,484±10,96, z-skor normalizasyonu sonrasında %60,219±11,27, EG diskritizasyonu ile %55,266±9,68, EF diskritizasyonu ile %67,047±13,34 olduğu görüldü. Sınıf dağılım oranı 0,3 iken bu oranlar; ham veri ile %63,713±10,87, minimum-maksimum normalizasyonu sonrasında %62,111±10,58, z-skor normalizasyonu sonrasında %63,518±10,78, EG diskritizasyonu ile %60,355±11,43, EF diskritizasyonu ile %73,125±11,37 olarak hesaplandı. Sınıf dağılım oranı 0,4 iken elde edilen ortalama doğruluk oranları; ham veride %67,371±10,98, minimum-maksimum normalizasyonu sonrasında %62,308±9,63, z-skor normalizasyonu sonrasında %66,410±10,52, EG diskritizasyonu ile %65,604±11,52, EF diskritizasyonu ile %76,379±9,98 olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama doğruluk oranlarının; ham veride %69,745±11,06, minimum-maksimum normalizasyonu sonrasında %53,865±15,03, z-skor normalizasyonu sonrasında %68,47±11,9, EG diskritizasyonu ile %68,955±11,89, EF diskritizasyonu ile %77,19±9,39 olduğu sonucuna ulaşıldı.

Gözlem sayısı 500 olduğunda, 0,1 sınıf dağılım oranı için elde edilen ortalama doğruluk oranları; ham veride %61,391±7,95, minimum-maksimum normalizasyonu sonrasında %60,48±7,66, z-skor normalizasyonu sonrasında %62,508±7,9, EG diskritizasyonu ile %56,666±6,46, EF diskritizasyonu ile %68,877±8,23 olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda ise elde edilen ortalama doğruluk oranları; ham veride %77,778±4,26, minimum-maksimum normalizasyonu sonrasında %63,36±7,92, z-skor normalizasyonu sonrasında %80,189±4,61, EG diskritizasyonu ile %68,88±6,39, EF diskritizasyonu ile %85,021±7,07 değerlerine düştüğü görüldü.

Gözlem sayısı 1000 olduğunda, sınıf dağılım oranı 0,1 iken elde edilen ortalama doğruluk oranları; ham veride %62,444±6,12, minimum-maksimum normalizasyonu sonrasında %61,584±5,77, z-skor normalizasyonu sonrasında %65,403±6,08, EG diskritizasyonu ile %58,081±5,17, EF diskritizasyonu ile %72,972±5,96 olarak bulundu. Sınıf

dağılım oranı 0,5 olduğunda elde edilen ortalama doğruluk oranları; ham veride %80,72±2,9 olarak bulunurken minimum-maksimum normalizasyonu, z-skor normalizasyonu, EG diskritizasyonu, EF diskritizasyonu sonrasında sırasıyla %64,862±4,42, %83,103±3,01, %68,634±4,2, %88,107±2,29 oranlarına ulaşıldı.

Gözlem sayısı 10000 olduğunda yapılan sınıflandırma çalışmasında 0,1 sınıf dağılım oranı için elde edilen ortalama doğruluk oranları; ham veride %73,810±1,90, minimum-maksimum normalizasyonu sonrasında %62,086±1,86, z-skor normalizasyonu sonrasında %83,059±1,31, EG diskritizasyonu ile %59,223±3,56, EF diskritizasyonu ile %81,722±1,78 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise elde edilen ortalama doğruluk oranları; ham veride %87,2±0,77, minimum-maksimum normalizasyonu sonrasında %65,475±1,31, z-skor normalizasyonu sonrasında %87,875±0,83, EG diskritizasyonu ile %61,958±1,71, EF diskritizasyonu ile %91,405±0,67 oranlarına düştüğü görüldü.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama doğruluk oranlarına Tablo 52’de yer verildi.

Tablo 53. F dağılımı için DVM ile elde edilen ortalama AUC değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,670±0,28	0,620±0,29	0,650±0,29	0,659±0,25	0,763±0,22
	0,2	0,740±0,17	0,664±0,2	0,709±0,19	0,727±0,17	0,837±0,13
	0,3	0,760±0,14	0,680±0,17	0,731±0,16	0,705±0,15	0,857±0,10
	0,4	0,767±0,12	0,668±0,15	0,739±0,14	0,744±0,14	0,857±0,09
	0,5	0,761±0,12	0,531±0,22	0,747±0,14	0,756±0,13	0,857±0,09
500	0,1	0,806±0,09	0,707±0,13	0,815±0,10	0,741±0,10	0,898±0,06
	0,2	0,833±0,06	0,601±0,27	0,664±0,20	0,759±0,07	0,926±0,03
	0,3	0,843±0,05	0,701±0,07	0,866±0,07	0,765±0,07	0,934±0,03
	0,4	0,849±0,04	0,698±0,06	0,877±0,05	0,767±0,06	0,938±0,02
	0,5	0,851±0,04	0,681±0,11	0,882±0,04	0,760±0,08	0,925±0,11
1000	0,1	0,839±0,06	0,714±0,10	0,857±0,07	0,722±0,08	0,925±0,03
	0,2	0,867±0,04	0,705±0,07	0,895±0,04	0,730±0,06	0,944±0,02
	0,3	0,879±0,03	0,696±0,05	0,908±0,03	0,743±0,06	0,952±0,02
	0,4	0,882±0,03	0,696±0,04	0,909±0,02	0,754±0,05	0,953±0,01
	0,5	0,882±0,03	0,706±0,05	0,909±0,02	0,757±0,05	0,954±0,01
10000	0,1	0,927±0,01	0,702±0,03	0,937±0,01	0,635±0,05	0,964±0,01
	0,2	0,937±0,01	0,685±0,02	0,944±0,01	0,625±0,04	0,971±0,004
	0,3	0,941±0,01	0,686±0,02	0,947±0,01	0,627±0,04	0,974±0,003
	0,4	0,942±0,01	0,693±0,01	0,949±0,01	0,637±0,04	0,975±0,003
	0,5	0,941±0,01	0,708±0,02	0,950±0,01	0,670±0,03	0,943±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Sınıflandırma çalışmasında 100 gözlem için elde edilen ortalama AUC değerleri sınıf dağılım oranı 0,1 olduğunda; ham veride $0,67\pm0,28$, minimum-maksimum normalizasyonu sonrasında $0,62\pm0,29$, z-skor normalizasyonu sonrasında $0,65\pm0,29$, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile $0,659\pm0,25$, eşit frekansta aralıklandırma (EF) diskritizasyonu ile $0,763\pm0,22$ 'dir. Sınıf dağılım oranı 0,2 olduğunda ortalama AUC değerlerinin; ham veride $0,74\pm0,17$, minimum-maksimum normalizasyonu sonrasında $0,664\pm0,2$, z-skor normalizasyonu sonrasında $0,709\pm0,19$, EG diskritizasyonu ile $0,727\pm0,17$, EF diskritizasyonu ile $0,837\pm0,13$ arttığı görüldü. Sınıf dağılım oranı 0,3 iken bu değerler; ham veri ile $0,76\pm0,14$, minimum-maksimum normalizasyonu sonrasında $0,68\pm0,17$, z-skor normalizasyonu sonrasında $0,731\pm0,16$, EG diskritizasyonu ile $0,705\pm0,15$, EF diskritizasyonu ile $0,857\pm0,1$ olarak hesaplandı. Sınıf dağılım oranı 0,4 iken elde edilen ortalama AUC değerlerinin; ham veride $0,767\pm0,12$, minimum-maksimum normalizasyonu sonrasında $0,668\pm0,15$, z-skor normalizasyonu sonrasında $0,739\pm0,14$, EG diskritizasyonu ile $0,744\pm0,14$, EF diskritizasyonu ile $0,857\pm0,09$ olduğu gözlemlendi. Son olarak sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama AUC değerlerinin; ham veride $0,761\pm0,12$, minimum-maksimum normalizasyonu sonrasında $0,531\pm0,22$, z-skor normalizasyonu sonrasında $0,747\pm0,14$, EG diskritizasyonu ile $0,756\pm0,13$, EF diskritizasyonu ile $0,857\pm0,09$ olduğu sonucuna varıldı.

Gözlem sayısı 500 iken gerçekleştirilen sınıflandırma çalışmasında sınıf dağılım oranı 0,1 için elde edilen ortalama AUC değerleri; ham veride $0,806\pm0,09$, minimum-maksimum normalizasyonu sonrasında $0,707\pm0,13$, z-skor normalizasyonu sonrasında $0,815\pm0,1$, EG diskritizasyonu ile $0,741\pm0,1$, EF diskritizasyonu ile $0,898\pm0,06$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama AUC değerleri; ham veride $0,851\pm0,04$ olarak bulunurken minimum-maksimum normalizasyonu, z-skor, EG diskritizasyonu, EF diskritizasyonu sonrasında sırasıyla; $0,681\pm0,11$, $0,882\pm0,04$, $0,76\pm0,08$, $0,925\pm0,11$ olarak belirlendi.

Gözlem sayısı 1000 olduğunda gerçekleştirilen sınıflandırma çalışmasında, sınıf dağılım oranı 0,1 iken elde edilen ortalama AUC değerleri; ham veride $0,839\pm0,06$, minimum-maksimum normalizasyonu sonrasında $0,714\pm0,1$, z-skor normalizasyonu sonrasında $0,857\pm0,07$, EG diskritizasyonu ile $0,722\pm0,08$, EF diskritizasyonu ile $0,925\pm0,03$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise ham veride $0,882\pm0,03$ olarak elde edilen ortalama AUC değeri, minimum-maksimum normalizasyonu, z-skor, EG

diskritizasyonu, EF diskritizasyonu sonrasında sırasıyla; 0,706±0,05, 0,909±0,02, 0,757±0,05, 0,954±0,01 değerlerine yükseldi.

Gözlem sayısı 10000 olduğunda, sınıf dağılım oranı 0,1 iken elde edilen ortalama AUC değerleri; ham veride 0,927±0,01, minimum-maksimum normalizasyonu sonrasında 0,702±0,03, z-skor normalizasyonu sonrasında 0,937±0,01, EG diskritizasyonu ile 0,635±0,05, EF diskritizasyonu ile 0,964±0,01 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda bu değerler; ham veride 0,941±0,01, minimum-maksimum normalizasyonu sonrasında 0,708±0,02, z-skor normalizasyonu sonrasında 0,95±0,01, EG diskritizasyonu ile 0,662±0,03, EF diskritizasyonu ile 0,975±0,003 olarak elde edildi.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama AUC değerlerine Tablo 53’de yer verildi.

Tablo 54. F dağılımı için DVM ile elde edilen ortalama duyarlılık oranları

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,148±0,26	0,117±0,25	0,144±0,26	0,027±0,13	0,110±0,25
	0,2	0,228±0,23	0,198±0,22	0,217±0,23	0,128±0,21	0,375±0,27
	0,3	0,316±0,22	0,266±0,22	0,312±0,22	0,247±0,21	0,539±0,22
	0,4	0,468±0,22	0,301±0,20	0,429±0,23	0,469±0,24	0,663±0,18
	0,5	0,721±0,18	0,540±0,27	0,672±0,18	0,688±0,17	0,762±0,15
500	0,1	0,233±0,16	0,212±0,15	0,256±0,16	0,139±0,13	0,394±0,17
	0,2	0,342±0,13	0,055±0,14	0,198±0,22	0,244±0,13	0,624±0,11
	0,3	0,497±0,12	0,346±0,10	0,566±0,13	0,372±0,13	0,738±0,09
	0,4	0,702±0,11	0,434±0,09	0,694±0,10	0,524±0,12	0,805±0,07
	0,5	0,826±0,06	0,568±0,14	0,799±0,07	0,692±0,10	0,843±0,10
1000	0,1	0,256±0,12	0,234±0,12	0,317±0,12	0,168±0,11	0,476±0,12
	0,2	0,417±0,10	0,296±0,08	0,518±0,10	0,262±0,10	0,683±0,08
	0,3	0,575±0,08	0,350±0,07	0,664±0,08	0,347±0,11	0,782±0,05
	0,4	0,777±0,06	0,435±0,06	0,754±0,05	0,487±0,12	0,834±0,04
	0,5	0,847±0,04	0,576±0,09	0,831±0,04	0,675±0,10	0,874±0,04
10000	0,1	0,488±0,04	0,244±0,04	0,546±0,04	0,188±0,07	0,651±0,04
	0,2	0,653±0,03	0,288±0,03	0,691±0,03	0,192±0,06	0,788±0,02
	0,3	0,806±0,02	0,351±0,02	0,771±0,02	0,207±0,07	0,845±0,02
	0,4	0,855±0,01	0,439±0,02	0,831±0,02	0,269±0,09	0,880±0,01
	0,5	0,892±0,01	0,581±0,02	0,879±0,01	0,627±0,23	0,907±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Sınıflandırma çalışmasında 100 gözlem için elde edilen ortalama duyarlılık oranları, sınıf dağılım oranı 0,1 olduğunda; ham veride 0,148±0,26, minimum-maksimum normalizasyonu sonrasında 0,117±0,25, z-skor normalizasyonu sonrasında 0,144±0,26, eşit

genişlikte aralıklandırma (EG) diskritizasyonu ile $0,027\pm 0,13$, eşit frekansta aralıklandırma (EF) diskritizasyonu ile $0,11\pm 0,25$ olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama duyarlılık oranlarının; ham veride $0,228\pm 0,23$ olarak bulunurken minimum-maksimum normalizasyonu sonrasında $0,198\pm 0,22$, z-skor normalizasyonu sonrasında $0,217\pm 0,23$, EG diskritizasyonu ile $0,128\pm 0,21$, EF diskritizasyonu ile $0,375\pm 0,27$ olduğu görüldü. Sınıf dağılım oranı 0,3 iken bu oranlar; ham veride $0,316\pm 0,22$ olup minimum-maksimum normalizasyonu sonrasında $0,266\pm 0,22$, z-skor normalizasyonu sonrasında $0,312\pm 0,22$, EG diskritizasyonu ile $0,247\pm 0,21$, EF diskritizasyonu ile $0,539\pm 0,22$ olarak hesaplandı. Sınıf dağılım oranı 0,4 için sırasıyla; $0,468\pm 0,22$, $0,301\pm 0,2$, $0,429\pm 0,23$, $0,469\pm 0,24$, $0,663\pm 0,18$ değerlerine ulaşıldı. Son olarak sınıf dağılım oranı 0,5 olduğunda ortalama duyarlılık oranının; ham veride $0,721\pm 0,18$, minimum-maksimum normalizasyonu sonrasında $0,54\pm 0,27$, z-skor normalizasyonu sonrasında $0,672\pm 0,18$, EG diskritizasyonu ile $0,688\pm 0,17$, EF diskritizasyonu ile $0,762\pm 0,15$ olduğu sonucuna ulaşıldı.

Gözlem sayısı 500 olan çalışmada, sınıf dağılım oranı 0,1 için elde edilen ortalama duyarlılık oranları; ham veride $0,233\pm 0,16$, minimum-maksimum normalizasyonu sonrasında $0,212\pm 0,15$, z-skor normalizasyonu sonrasında $0,256\pm 0,16$, EG diskritizasyonu ile $0,139\pm 0,13$, EF diskritizasyonu ile $0,394\pm 0,17$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise ortalama duyarlılık oranları; ham veride $0,826\pm 0,06$, minimum-maksimum normalizasyonu sonrasında $0,568\pm 0,14$, z-skor normalizasyonu sonrasında $0,799\pm 0,07$, EG diskritizasyonu ile $0,692\pm 0,1$, EF diskritizasyonu ile $0,843\pm 0,1$ olarak elde edildi.

Gözlem sayısı 1000 olduğunda, sınıf dağılım oranı 0,1 iken elde edilen ortalama duyarlılık oranları; ham veride $0,256\pm 0,12$, minimum-maksimum normalizasyonu sonrasında $0,234\pm 0,12$, z-skor normalizasyonu sonrasında $0,317\pm 0,12$, EG diskritizasyonu ile $0,168\pm 0,11$, EF diskritizasyonu ile $0,476\pm 0,12$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ortalama duyarlılık oranı ham veride $0,847\pm 0,04$ olarak bulunurken minimum-maksimum normalizasyonu, z-skor, EG diskritizasyonu, EF diskritizasyonu ile sırasıyla; $0,576\pm 0,09$, $0,831\pm 0,04$, $0,675\pm 0,1$, $0,874\pm 0,04$ duyarlılık oranlarına ulaşıldı.

Gözlem sayısı 10000 olduğunda yapılan çalışmada sınıf dağılım oranı 0,1 iken elde edilen ortalama duyarlılık oranları; ham veri ile $0,488\pm 0,04$, minimum-maksimum normalizasyonu ile $0,244\pm 0,04$, z-skor normalizasyonu ile $0,546\pm 0,04$, EG ile $0,188\pm 0,07$, EF ile $0,651\pm 0,04$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ortalama duyarlılık oranları sırasıyla; $0,892\pm 0,01$, $0,581\pm 0,02$, $0,879\pm 0,01$, $0,627\pm 0,23$, $0,907\pm 0,01$ oranlarına yükseldi.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama duyarlılık oranlarına Tablo 54’te yer verildi.

Tablo 55. F dağılımı için DVM ile elde edilen ortalama seçicilik oranları

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	0,994±0,02	0,998±0,01	0,998±0,01	0,996±0,02	0,993±0,02
	0,2	0,987±0,03	0,991±0,03	0,987±0,03	0,977±0,04	0,966±0,05
	0,3	0,958±0,07	0,976±0,05	0,958±0,07	0,959±0,07	0,924±0,08
	0,4	0,880±0,13	0,945±0,08	0,899±0,11	0,843±0,12	0,864±0,11
	0,5	0,674±0,18	0,538±0,31	0,697±0,17	0,691±0,17	0,782±0,14
500	0,1	0,994±0,01	0,998±0,01	0,994±0,01	0,994±0,01	0,983±0,02
	0,2	0,973±0,02	0,993±0,03	0,991±0,03	0,974±0,03	0,964±0,02
	0,3	0,925±0,04	0,979±0,02	0,937±0,03	0,932±0,05	0,939±0,03
	0,4	0,839±0,06	0,925±0,04	0,882±0,05	0,842±0,07	0,904±0,04
	0,5	0,730±0,07	0,699±0,17	0,805±0,07	0,686±0,11	0,857±0,10
1000	0,1	0,993±0,01	0,998±0,004	0,991±0,01	0,993±0,01	0,983±0,01
	0,2	0,964±0,02	0,993±0,01	0,969±0,02	0,978±0,02	0,965±0,02
	0,3	0,927±0,03	0,981±0,01	0,938±0,02	0,947±0,04	0,947±0,02
	0,4	0,848±0,04	0,926±0,03	0,893±0,03	0,863±0,07	0,919±0,03
	0,5	0,768±0,05	0,721±0,10	0,831±0,04	0,697±0,09	0,888±0,03
10000	0,1	0,988±0,003	0,998±0,001	0,989±0,003	0,996±0,003	0,983±0,003
	0,2	0,969±0,01	0,995±0,003	0,971±0,01	0,990±0,01	0,970±0,01
	0,3	0,933±0,01	0,981±0,01	0,946±0,01	0,977±0,02	0,960±0,01
	0,4	0,896±0,01	0,927±0,01	0,917±0,01	0,938±0,05	0,943±0,01
	0,5	0,852±0,01	0,728±0,03	0,879±0,01	0,612±0,23	0,921±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Sınıflandırma çalışmasında 100 gözlem için elde edilen ortalama seçicilik oranları sınıf dağılım oranı 0,1 olduğunda; ham veride 0,994±0,02, minimum-maksimum normalizasyonu sonrasında 0,998±0,01, z-skor normalizasyonu sonrasında 0,998±0,01, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,996±0,02, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,993±0,02 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ham veride 0,987±0,03 olarak elde edilen ortalama seçicilik oranının, minimum-maksimum normalizasyonu, z-skor, EG diskritizasyonu ve EF diskritizasyonu sonrasında sırasıyla; 0,991±0,03, 0,987±0,03, 0,977±0,04, 0,966±0,05 değerlerine ulaşıldı. Sınıf dağılım oranı 0,3 iken bu oranlar sırasıyla; 0,958±0,07, 0,976±0,05, 0,958±0,07, ile 0,959±0,07, 0,924±0,08 olarak hesaplandı. Sınıfların 0,4 oranında dağılımı için elde edilen bu oranlar sırasıyla; 0,88±0,13, 0,945±0,08, 0,899±0,11, 0,843±0,12, 0,864±0,11 olarak bulundu. Son olarak sınıf dağılım oranı 0,5 olduğunda; ham veride 0,674±0,18, minimum-maksimum normalizasyonu

sonrasında $0,538\pm 0,31$, z-skor normalizasyonu sonrasında $0,697\pm 0,17$, EG diskritizasyonu ile $0,691\pm 0,17$, EF diskritizasyonu ile $0,782\pm 0,14$ değerleri elde edildi.

Gözlem sayısı 500 olan sınıflandırma çalışmasında, sınıf dağılım oranı 0,1 için elde edilen ortalama seçicilik oranları; ham veride $0,994\pm 0,01$, minimum-maksimum normalizasyonu sonrasında $0,998\pm 0,01$, z-skor normalizasyonu sonrasında $0,994\pm 0,01$, EG diskritizasyonu ile $0,994\pm 0,01$, EF diskritizasyonu ile $0,983\pm 0,02$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise ortalama seçicilik oranları sırasıyla; $0,73\pm 0,07$, $0,699\pm 0,17$, $0,805\pm 0,07$, $0,686\pm 0,11$, olarak elde edildi.

Gözlem sayısı 1000 olduğunda gerçekleştirilen sınıflandırma çalışmasında 0,1 sınıf dağılım oranı için ham veride $0,993\pm 0,01$ olarak elde edilen ortalama seçicilik oranları, minimum-maksimum normalizasyonu, z-skor, EG diskritizasyonu ve EF diskritizasyonu sonrasında sırasıyla; $0,998\pm 0,01$, $0,991\pm 0,01$, $0,993\pm 0,01$, $0,983\pm 0,01$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise bu oranlar için sırasıyla; $0,768\pm 0,05$, $0,721\pm 0,1$, $0,831\pm 0,04$, $0,697\pm 0,09$, $0,888\pm 0,03$ bulgusuna ulaşıldı.

Gözlem sayısı 10000 olduğunda, sınıf dağılım oranı 0,1 iken elde edilen ortalama seçicilik oranları; ham veride $0,988\pm 0,003$, minimum-maksimum normalizasyonu sonrasında $0,998\pm 0,001$, z-skor normalizasyonu sonrasında $0,989\pm 0,003$, EG diskritizasyonu ile $0,996\pm 0,003$, EF diskritizasyonu ile $0,983\pm 0,003$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise ortalama seçicilik oranları sırasıyla; $0,852\pm 0,01$, $0,728\pm 0,03$, $0,879\pm 0,01$, ile $0,612\pm 0,23$, $0,921\pm 0,01$ olarak belirlendi.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama seçicilik oranlarına Tablo 55'te yer verildi.

Tablo 56. F dağılımı için DVM ile elde edilen ortalama pozitif kestirim değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	$0,743\pm 0,40$	$0,879\pm 0,29$	$0,894\pm 0,28$	$0,432\pm 0,48$	$0,643\pm 0,44$
	0,2	$0,829\pm 0,31$	$0,879\pm 0,26$	$0,844\pm 0,30$	$0,604\pm 0,39$	$0,771\pm 0,29$
	0,3	$0,799\pm 0,28$	$0,859\pm 0,25$	$0,811\pm 0,25$	$0,769\pm 0,30$	$0,784\pm 0,21$
	0,4	$0,757\pm 0,22$	$0,824\pm 0,22$	$0,778\pm 0,21$	$0,681\pm 0,20$	$0,785\pm 0,15$
	0,5	$0,705\pm 0,13$	$0,547\pm 0,22$	$0,701\pm 0,14$	$0,701\pm 0,14$	$0,790\pm 0,12$
500	0,1	$0,835\pm 0,25$	$0,940\pm 0,16$	$0,855\pm 0,21$	$0,757\pm 0,33$	$0,746\pm 0,20$
	0,2	$0,774\pm 0,17$	$0,715\pm 0,36$	$0,879\pm 0,26$	$0,742\pm 0,21$	$0,818\pm 0,10$
	0,3	$0,746\pm 0,11$	$0,884\pm 0,11$	$0,803\pm 0,09$	$0,720\pm 0,14$	$0,842\pm 0,07$
	0,4	$0,750\pm 0,07$	$0,802\pm 0,10$	$0,801\pm 0,07$	$0,700\pm 0,10$	$0,851\pm 0,05$
	0,5	$0,756\pm 0,05$	$0,668\pm 0,12$	$0,806\pm 0,05$	$0,693\pm 0,07$	$0,856\pm 0,09$

Tablo 56. F dağılımı için DVM ile elde edilen ortalama pozitif kestirim değerleri

1000	0,1	0,808±0,19	0,938±0,12	0,816±0,16	0,766±0,24	0,770±0,12
	0,2	0,752±0,10	0,923±0,08	0,813±0,08	0,776±0,15	0,831±0,06
	0,3	0,774±0,06	0,891±0,07	0,823±0,05	0,762±0,12	0,865±0,05
	0,4	0,775±0,04	0,801±0,07	0,826±0,04	0,721±0,09	0,875±0,04
	0,5	0,786±0,03	0,682±0,07	0,832±0,04	0,696±0,05	0,887±0,03
10000	0,1	0,821±0,04	0,935±0,04	0,843±0,03	0,858±0,09	0,810±0,03
	0,2	0,840±0,02	0,934±0,03	0,855±0,02	0,849±0,10	0,868±0,02
	0,3	0,838±0,02	0,891±0,03	0,861±0,02	0,829±0,10	0,902±0,01
	0,4	0,845±0,01	0,802±0,02	0,869±0,01	0,782±0,10	0,911±0,01
	0,5	0,858±0,01	0,682±0,02	0,879±0,01	0,653±0,09	0,920±0,01

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Sınıflandırma çalışmasında 100 gözlem için elde edilen ortalama pozitif kestirim değeri (PKD), sınıf dağılım oranı 0,1 olduğunda; ham veride 0,743±0,4, minimum-maksimum normalizasyonu sonrasında 0,879±0,29, z-skor normalizasyonu sonrasında 0,894±0,28, eşit genişlikte aralıklandırma (EG) diskritizasyonu ile 0,432±0,48, eşit frekansta aralıklandırma (EF) diskritizasyonu ile 0,643±0,44 olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ortalama PKD'nin sırasıyla; 0,829±0,31, 0,879±0,26, 0,844±0,3, 0,604±0,39, 0,771±0,29 olduğu görüldü. Sınıf dağılım oranı 0,3 iken bu değerler sırasıyla; 0,799±0,28, 0,859±0,25, 0,811±0,25, 0,769±0,3, 0,784±0,21 olarak hesaplandı. Sınıf dağılım oranı 0,4 iken ortalama değerlerin sırasıyla; ham veride 0,757±0,22, minimum-maksimum normalizasyonu sonrasında 0,824±0,22, z-skor normalizasyonu sonrasında 0,778±0,21, EG diskritizasyonu ile 0,681±0,2, EF diskritizasyonu ile 0,785±0,15 olduğu gözlemlendi. Son olarak sınıf dağılım oranı 0,5 olduğunda elde edilen ortalama PKD'nin; ham veride 0,705±0,13, minimum-maksimum normalizasyonu sonrasında 0,547±0,22, z-skor normalizasyonu sonrasında 0,701±0,14, EG diskritizasyonu ile 0,701±0,14, EF diskritizasyonu ile 0,79±0,12 olduğu sonucuna varıldı.

Gözlem sayısının 500 olduğu sınıflandırma çalışmasında, sınıf dağılım oranı 0,1 için elde edilen ortalama PKD; ham veride 0,835±0,25, minimum-maksimum normalizasyonu sonrasında 0,94±0,16, z-skor normalizasyonu sonrasında 0,855±0,21, EG diskritizasyonu ile 0,757±0,33, EF diskritizasyonu ile 0,746±0,2 olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise ortalama PKD; ham veride 0,756±0,05, minimum-maksimum normalizasyonu sonrasında 0,668±0,12, z-skor normalizasyonu sonrasında 0,806±0,05, EG diskritizasyonu ile 0,693±0,07, EF diskritizasyonu ile 0,856±0,09 olarak elde edildi.

Gözlem sayısı 1000 olduğunda gerçekleştirilen sınıflandırma çalışmasında 0,1 sınıf dağılım oranı için ortalama PKD; ham veride 0,808±0,19, minimum-maksimum

normalizasyonu sonrasında $0,938\pm0,12$, z-skor normalizasyonu sonrasında $0,816\pm0,16$, EG diskritizasyonu ile $0,766\pm0,24$, EF diskritizasyonu ile $0,77\pm0,12$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ham veride $0,786\pm0,03$ olarak elde edilen ortalama PKD için minimum-maksimum normalizasyonu, z-skor, EG diskritizasyonu ve EF diskritizasyonu sonrasında sırasıyla; $0,682\pm0,07$, $0,832\pm0,04$, $0,696\pm0,05$, $0,887\pm0,03$ değerlerine ulaşıldı.

Gözlem sayısı 10000 olduğunda yapılan sınıflandırma çalışmasında 0,1 sınıf dağılım oranı için ortalama PKD; ham veride $0,821\pm0,04$, minimum-maksimum normalizasyonu ile $0,935\pm0,04$, z-skor ile $0,843\pm0,03$, EG diskritizasyonu ile $0,858\pm0,09$, EF diskritizasyonu ile $0,81\pm0,03$ olarak elde edildi. Sınıf dağılım oranı 0,5 olduğunda ise edilen ortalama PKD için ham veri ile $0,858\pm0,01$, minimum-maksimum normalizasyonu ile $0,682\pm0,02$, z-skor ile $0,879\pm0,01$, EG diskritizasyonu ile $0,653\pm0,09$, EF diskritizasyonu ile $0,92\pm0,01$ değerlerine ulaşıldı.

Çalışmada 500, 1000 ve 10000 gözlem sayıları için diğer sınıf dağılımlarından (0,2, 0,3, 0,4) elde edilen ortalama pozitif kestirim değerlerine Tablo 56’da yer verildi.

Tablo 57. F dağılımı için DVM ile elde edilen ortalama negatif kestirim değerleri

n	Oran	Ham veri	Min-Maks N.	Z-Skor N.	EG	EF
100	0,1	$0,914\pm0,03$	$0,911\pm0,02$	$0,914\pm0,03$	$0,902\pm0,01$	$0,910\pm0,02$
	0,2	$0,838\pm0,04$	$0,834\pm0,04$	$0,837\pm0,04$	$0,819\pm0,04$	$0,864\pm0,05$
	0,3	$0,770\pm0,06$	$0,760\pm0,06$	$0,769\pm0,06$	$0,752\pm0,06$	$0,829\pm0,07$
	0,4	$0,722\pm0,09$	$0,676\pm0,07$	$0,712\pm0,08$	$0,716\pm0,09$	$0,802\pm0,09$
	0,5	$0,723\pm0,14$	$0,501\pm0,23$	$0,691\pm0,13$	$0,699\pm0,14$	$0,780\pm0,11$
500	0,1	$0,921\pm0,02$	$0,920\pm0,02$	$0,924\pm0,02$	$0,912\pm0,01$	$0,936\pm0,02$
	0,2	$0,856\pm0,03$	$0,809\pm0,02$	$0,834\pm0,04$	$0,838\pm0,02$	$0,912\pm0,02$
	0,3	$0,812\pm0,04$	$0,778\pm0,03$	$0,837\pm0,04$	$0,778\pm0,03$	$0,894\pm0,03$
	0,4	$0,813\pm0,05$	$0,711\pm0,03$	$0,815\pm0,05$	$0,729\pm0,05$	$0,876\pm0,04$
	0,5	$0,809\pm0,05$	$0,612\pm0,11$	$0,803\pm0,05$	$0,694\pm0,07$	$0,846\pm0,09$
1000	0,1	$0,923\pm0,01$	$0,922\pm0,01$	$0,929\pm0,01$	$0,915\pm0,01$	$0,944\pm0,01$
	0,2	$0,869\pm0,02$	$0,850\pm0,02$	$0,890\pm0,02$	$0,842\pm0,02$	$0,924\pm0,02$
	0,3	$0,837\pm0,03$	$0,779\pm0,02$	$0,868\pm0,03$	$0,773\pm0,03$	$0,910\pm0,02$
	0,4	$0,852\pm0,03$	$0,712\pm0,02$	$0,846\pm0,03$	$0,720\pm0,04$	$0,893\pm0,03$
	0,5	$0,835\pm0,04$	$0,631\pm0,05$	$0,832\pm0,04$	$0,688\pm0,05$	$0,877\pm0,03$
10000	0,1	$0,946\pm0,004$	$0,922\pm0,004$	$0,951\pm0,004$	$0,917\pm0,01$	$0,962\pm0,004$
	0,2	$0,918\pm0,01$	$0,848\pm0,01$	$0,926\pm0,01$	$0,831\pm0,01$	$0,948\pm0,01$
	0,3	$0,918\pm0,01$	$0,779\pm0,01$	$0,906\pm0,01$	$0,742\pm0,01$	$0,935\pm0,01$
	0,4	$0,903\pm0,01$	$0,713\pm0,01$	$0,890\pm0,01$	$0,659\pm0,02$	$0,922\pm0,01$
	0,5	$0,888\pm0,01$	$0,635\pm0,01$	$0,879\pm0,01$	$0,657\pm0,09$	$0,908\pm0,01$

Min-Maks N.: Minimum-maksimum normalizasyonu; **Z-Skor N.:** Z-Skor normalizasyonu; **EG:** Eşit genişlikte aralıklandırma diskritizasyonu; **EF:** Eşit frekansta aralıklandırma diskritizasyonu.

Sınıflandırma çalışmasında 100 gözlem için elde edilen ortalama negatif kestirim değeri (NKD) sınıf dağılım oranı 0,1 olduğunda; ham veride ve z-skor normalizasyonu sonrasında $0,914\pm 0,03$, minimum-maksimum normalizasyonu sonrasında $0,911\pm 0,02$, eşit genişlikte aralıklandırma (EG) ile $0,902\pm 0,01$, eşit frekansta aralıklandırma (EF) diskritizasyonu ile $0,91\pm 0,02$ olarak bulundu. Sınıf dağılım oranı 0,2 olduğunda ham veride $0,838\pm 0,04$ olan ortalama değer, minimum-maksimum normalizasyonu, z-skor, EG diskritizasyonu ve EF diskritizasyonu sonrasında sırasıyla; $0,834\pm 0,04$, $0,837\pm 0,04$, $0,819\pm 0,04$, $0,864\pm 0,05$ olduğu görüldü. Sınıfların 0,3 dağılım oranı için ortalama NKD; ham veri ile $0,77\pm 0,06$ olup minimum-maksimum normalizasyonu, z-skor, EG diskritizasyonu ve EF diskritizasyonu sonrasında sırasıyla; $0,76\pm 0,06$, $0,769\pm 0,06$, $0,752\pm 0,06$, $0,829\pm 0,07$ olarak elde edildi. Sınıf dağılım oranı 0,4 olduğunda ortalama NKD'nin; ham veride $0,722\pm 0,09$, minimum-maksimum normalizasyonu, z-skor, EG diskritizasyonu ve EF diskritizasyonu sonrasında sırasıyla; $0,676\pm 0,07$, $0,712\pm 0,08$, $0,716\pm 0,09$, $0,802\pm 0,09$ olduğu gözlemlendi. Son olarak sınıf dağılım oranı 0,5 olduğunda ortalama NKD için sırasıyla; $0,723\pm 0,14$, $0,501\pm 0,23$, $0,691\pm 0,13$, $0,699\pm 0,14$, $0,78\pm 0,11$ bulgularına ulaşıldı.

Gözlem sayısı 500 olan sınıflandırma çalışmasında, sınıf dağılım oranı 0,1 için elde edilen ortalama negatif kestirim değerleri; ham veride $0,921\pm 0,02$, minimum-maksimum normalizasyonu sonrasında $0,92\pm 0,02$, z-skor normalizasyonu sonrasında $0,924\pm 0,02$, EG diskritizasyonu ile $0,912\pm 0,01$, EF diskritizasyonu ile $0,936\pm 0,02$ olarak bulundu. Sınıf dağılım oranı 0,5 olduğunda ise bu değerler sırasıyla; $0,809\pm 0,05$, $0,612\pm 0,11$, $0,803\pm 0,05$, $0,694\pm 0,07$, $0,846\pm 0,09$ olarak belirlendi. Gözlem sayısının 1000 olduğu çalışmada, 0,1 sınıf dağılım oranı için elde edilen ortalama NKD; ham veride $0,923\pm 0,01$, minimum-maksimum normalizasyonu sonrasında $0,922\pm 0,01$, z-skor normalizasyonu sonrasında $0,929\pm 0,01$, EG diskritizasyonu ile $0,915\pm 0,01$, EF diskritizasyonu ile $0,944\pm 0,01$ 'dir. Sınıf dağılım oranı 0,5 olduğunda ortalama NKD için ham veride $0,835\pm 0,04$ değeri bulunurken, minimum-maksimum normalizasyonu, z-skor, EG diskritizasyonu ve EF diskritizasyonu sonrasında sırasıyla; $0,631\pm 0,05$, $0,832\pm 0,04$, $0,688\pm 0,05$, $0,877\pm 0,03$ değerleri elde edildi. Gözlem sayısı 10000 olduğunda yapılan sınıflandırma çalışmasında, 0,1 sınıf dağılım oranı için elde edilen ortalama NKD; ham veride $0,946\pm 0,004$, minimum-maksimum normalizasyonu sonrasında $0,922\pm 0,004$, z-skor normalizasyonu sonrasında $0,951\pm 0,004$, EG diskritizasyonu ile $0,917\pm 0,01$, EF diskritizasyonu ile $0,962\pm 0,004$ 'tür. Sınıf dağılım oranı 0,5 olduğunda ham veri ile $0,888\pm 0,01$ olarak elde edilen ortalama NKD, minimum-maksimum

normalizasyonu ile $0,635\pm 0,01$, z-skor ile $0,879\pm 0,01$, EG diskritizasyonu ile $0,657\pm 0,09$, EF diskritizasyonu ile $0,908\pm 0,01$ olarak bulundu (Tablo 57).

SINIFLANDIRMA ALGORİTMALARININ KARŞILAŞTIRILMASI

Sınıflandırma algoritmalarının ham veride veya veri dönüşümleri sonrasında ulaşılan en yüksek doğruluk oranı, AUC değeri, duyarlılık ve seçicilik oranı ile pozitif kestirim değeri ve negatif kestirim değeri alınarak algoritmalar birbirleriyle karşılaştırıldı.



Tablo 58. Çalışmada elde edilen en yüksek doğruluk oranları (%)

		Doğruluk (%)											
Dağılım	Oran	N=100			N=500			N=1000			N=10000		
		NB	C5.0	DVM	NB	C5.0	DVM	NB	C5.0	DVM	NB	C5.0	DVM
Normal	0,1	73,0 H,M,Z	64,7 H,M,Z	66,0 M	79,7 H,M,Z	69,8 H,M,Z	77,8 M	80,5 H,M,Z	72,1 H,M,Z	79,6 M	81,2 H,M,Z	76,3 H	80,5 Z
	0,2	81,7 H,M,Z	71,3 H,M,Z	78,8 M	84,9 H,M,Z	76,0 H	83,9 M	85,2 H,M,Z	77,3 H,M,Z	83,7 Z	85,6 H,M,Z	80,9 Z	85,1 Z
	0,3	85,1 H,M,Z	73,5 H,M,Z	82,8 M	87,2 H,M,Z	78,5 H	85,4 Z	87,4 H,M,Z	79,8 H,M,Z	86,3 Z	87,7 H,M,Z	83,1 H,M,Z	87,4 Z
	0,4	86,5 H,M,Z	74,9 H,M,Z	84,3 M	88,5 H,M,Z	79,7 Z	86,6 Z	88,5 H,M,Z	81,0 H,M,Z	87,5 Z	88,8 H,M,Z	84,2 H,M,Z	88,5 Z
	0,5	87,1 H,M,Z	75,4 H,M,Z	84,2 M	88,7 H,M,Z	79,9 Z	87,0 Z	88,8 H,M,Z	81,1 H,M,Z	87,5 Z	89,1 H,M,Z	84,6 H,M,Z	88,8 Z
Ki-kare	0,1	61,6 EF	60,3 H,M,Z	58,1 Z	67,6 Z	64,5 EF	69,5 M	68,7 H,M,Z	65,9 H,M,Z	72,4 M	71,4 H,M,Z	68,2 Z	74,8 M
	0,2	70,6 H,M,Z	65,8 H,M,Z	71,2 M	75,6 H,M,Z	70,3 H,M,Z	79,7 M	76,8 H,M,Z	71,4 H,M,Z	80,5 M	78,4 H,M,Z	74,4 M,Z	79,2 M
	0,3	75,6 M	68,7 H,M,Z	78,9 M	80,1 M,Z	73,1 H,M,Z	82,6 M	80,1 H,M,Z	74,3 H,M,Z	82,3 M	81,9 H,M,Z	77,3 H,Z	82,4 M
	0,4	78,7 H,M,Z	70,2 H,M,Z	81,5 M	82,0 H,M,Z	74,7 H,M,Z	83,2 M	82,8 H	75,8 H,M,Z	83,7 M	83,8 H,M,Z	78,9 M,Z	83,9 Z
	0,5	78,0 H,M,Z	70,5 H,M,Z	81,1 M	82,7 M	74,9 H,M,Z	83,5 Z	83,2 H,M,Z	76,2 H,M,Z	83,7 Z	84,3 H,M,Z	79,3 H	84,5 Z
F	0,1	72,0 EF	66,9 H,M,Z	57,1 Z	77,5 EF	76,0 H,M	68,9 EF	79,7 EF	80,4 H,M	73,0 EF	82,9 H	84,9 Z	83,1 Z
	0,2	79,8 EF	74,2 H,M	67,0 EF	86,0 EF	82,5 Z	79,4 EF	87,1 EF	84,7 H	82,4 EF	88,6 EF	87,7 Z	87,9 EF
	0,3	83,5 EF	78,0 H,M,Z	73,1 EF	89,1 EF	84,8 Z	83,8 EF	90,0 EF	86,3 H,M	86,4 EF	90,7 EF	89,2 H,M	90,3 EF
	0,4	83,9 EF	78,7 H,M	76,4 EF	90,1 EF	85,4 Z	85,4 EF	90,8 EF	87,0 H,M	87,7 EF	91,6 EF	89,9 Z	91,2 EF
	0,5	84,1 EF	78,6 H,M	77,2 EF	90,4 EF	85,9 Z	85,0 EF	91,2 EF	87,5 H,M	88,1 EF	91,9 EF	90,1 Z	91,4 EF

H: Ham veri; **M:** Minimum-Maksimum Normalizasyonu; **Z:** Z-Skor Normalizasyonu; **EG:** Eşit Genişlikte Aralıklandırma Diskritizasyonu; **EF:** Eşit Frekansta Aralıklandırma Diskritizasyonu.

Çalışma neticesinde üç algoritmada ulaşılan en yüksek doğruluk oranlarına ve nasıl elde edildiklerine ilişkin bulgulara Tablo 58’de yer verildi.

Normal dağılımdan üretilen verilerde veri dönüşümü ile destek vektör makinelerinde tüm kombinasyonlarda artış görülürken naif Bayes algoritmasında herhangi bir artış gerçekleşmedi. NB algoritmasında Gaussian olasılık yoğunluk fonksiyonu ile gözlem sayısı ve sınıf dağılım oranlarına göre değişkenlik göstermeksizin diğer algoritmalarından daha yüksek genel doğruluk oranlarına ulaşıldı. DVM’de normal dağılımdan üretilen veriler için gözlem sayısının artmasıyla veri dönüşümlerindeki değişim dikkat çekmektedir. Gözlem sayısı arttıkça z-skor normalizasyonunun dengeliden dengesiz sınıf dağılımına doğru gidildiğinde minimum maksimuma göre üstünlüğünde artış görülmektedir. C5.0 algoritmasıyla yalnızca 500 gözlemdeki 0,4 ve 0,5 sınıf dağılımlarında ve 10000 gözlemdeki 0,2 sınıf dağılımında z-skor normalizasyonu ile doğruluk oranlarında artış olmakla birlikte Tablo 22’de yer alan bulgulara göre bu artışlar %0,01 oranından düşüktür.

Ki-kare dağılımı için incelendiğinde DVM’nin tüm veri tipinde veri dönüşümleri ile genel doğruluk oranında artış sağlandı. 100 gözlem ve 0,1 sınıf dağılım oranında, kernel olasılık yoğunluk fonksiyonu ile NB, geri kalan tüm kombinasyonlarda DVM algoritması en yüksek doğruluk oranlarına ulaştırdı. DVM’de 100 gözlemden büyük verilerde ve tam dengeli sınıf dağılımında (0,5) z-skor normalizasyonu ile en yüksek doğruluk oranları elde edilirken sınıflar dengeli dağılmadığında minimum-maksimum normalizasyonunun üstünlüğü dikkat çekmektedir. 100 gözlemde ise DVM’de 0,1 sınıf dağılımı dışındaki tüm dağılımlarda minimum-maksimum normalizasyonu uygulanan veriler ile en iyi sonuçlar bulundu.

F dağılımından üretilen verilerde 1000 ve 10000 gözlemdeki 0,1 sınıf dağılım oranı için C5.0 algoritmasının sırasıyla; ham veri ve z-skor dönüşümü ile geri kalan tüm kombinasyonlarda ise NB algoritmasının EF diskritizasyonu ile en yüksek sonuçları verdiği görüldü. DVM’de EF dönüşümü, doğruluğun artmasında diğer dönüşüm yöntemlerine göre çok daha etkili olmuştur.

Tablo 59. Çalışmada elde edilen en yüksek AUC değerleri

		AUC değerleri											
Dağılım	Oran	N=100			N=500			N=1000			N=10000		
		NB	C5.0	DVM	NB	C5.0	DVM	NB	C5.0	DVM	NB	C5.0	DVM
Normal	0,1	0,899 H,M,Z	0,681 H,M,Z	0,891 Z	0,953 H,M,Z	0,749 H,M,Z	0,931 M	0,956 H,M,Z	0,777 H,M,Z	0,934 M	0,960 H,M,Z	0,862 Z	0,950 EG
	0,2	0,927 H,M,Z	0,748 H,M,Z	0,913 H,M,Z	0,956 H,M,Z	0,812 H,M,Z	0,935 M	0,958 H,M,Z	0,831 H,M,Z	0,940 Z	0,959 H,M,Z	0,883 H,M,Z	0,951 EG
	0,3	0,940 H,M,Z	0,770 H,M,Z	0,920 H,M,Z	0,957 H,M,Z	0,838 H,M,Z	0,936 Z	0,957 H,M,Z	0,851 H,M,Z	0,943 Z	0,959 H,M,Z	0,895 H,M,Z	0,951 Z
	0,4	0,944 H,M,Z	0,783 H,M,Z	0,926 H,M,Z	0,958 H,M,Z	0,850 H,M,Z	0,939 Z	0,958 H,M,Z	0,862 H,M,Z	0,945 Z	0,959 H,M,Z	0,901 H,M,Z	0,953 Z
	0,5	0,947 H,M,Z	0,791 H,M,Z	0,925 H,M,Z	0,958 H,M,Z	0,852 H,M,Z	0,940 Z	0,958 H,M,Z	0,865 H,M,Z	0,944 Z	0,960 H,M,Z	0,904 H,M,Z	0,954 Z
Ki-kare	0,1	0,791 H,M	0,644 H,M,Z	0,854 Z	0,881 H,M,Z	0,723 H,M,Z	0,909 M	0,898 H,M,Z	0,720 H,M,Z	0,917 M	0,919 H,M,Z	0,806 H,M,Z	0,921 M
	0,2	0,836 H,M,Z	0,698 H,M,Z	0,888 M	0,896 H,M,Z	0,757 H,M,Z	0,916 M	0,905 H,M,Z	0,773 H,M,Z	0,916 M	0,921 H,M,Z	0,833 H,M,Z	0,921 M
	0,3	0,854 H,M,Z	0,731 H,M,Z	0,898 M	0,900 H,M,Z	0,783 H,M,Z	0,915 M	0,908 H,M,Z	0,799 H,M,Z	0,918 M	0,921 H,M,Z	0,844 H,M,Z	0,921 M
	0,4	0,855 H,M,Z	0,738 H,M,Z	0,901 M	0,902 H,M,Z	0,796 H,M,Z	0,916 M	0,911 H,M,Z	0,811 H,M,Z	0,920 M	0,922 H,M,Z	0,851 H,M,Z	0,922 Z,EF
	0,5	0,860 H,M,Z	0,744 H,M,Z	0,896 M	0,906 H,M,Z	0,801 H,M,Z	0,914 Z	0,911 H,M,Z	0,813 H,M,Z	0,917 Z	0,922 H,M,Z	0,853 H,M,Z	0,922 Z,EF
F	0,1	0,851 EF	0,694 H,M,Z	0,763 EF	0,936 EF	0,808 H,M,Z	0,898 EF	0,954 EF	0,859 H,M,Z	0,925 EF	0,970 EF	0,902 H,M,Z	0,964 EF
	0,2	0,898 EF	0,768 H,M,Z	0,837 EF	0,959 EF	0,881 H,M,Z	0,926 EF	0,967 EF	0,907 H,M,Z	0,944 EF	0,975 EF	0,948 H,M	0,971 EF
	0,3	0,922 EF	0,813 H,M,Z	0,857 EF	0,967 EF	0,906 H,M,Z	0,934 EF	0,973 EF	0,929 H,M,Z	0,952 EF	0,977 EF	0,959 H,M,Z	0,974 EF
	0,4	0,921 EF	0,822 H,M,Z	0,857 EF	0,969 EF	0,915 H,M,Z	0,938 EF	0,974 EF	0,934 H,M,Z	0,953 EF	0,978 EF	0,962 H,M,Z	0,975 EF
	0,5	0,924 EF	0,834 H,M,Z	0,857 EF	0,970 EF	0,921 H,M,Z	0,925 EF	0,975 EF	0,937 H,M,Z	0,954 EF	0,978 EF	0,962 H,M,Z	0,975 EF

H: Ham veri; **M:** Minimum-Maksimum Normalizasyonu; **Z:** Z-Skor Normalizasyonu; **EG:** Eşit Genişlikte Aralıklandırma Diskritizasyonu; **EF:** Eşit Frekansta Aralıklandırma Diskritizasyonu.

Çalışmada, naif Bayes, C5.0 ve destek vektör makinaları algoritmaları ile tüm simülasyon senaryolarından elde edilen en yüksek AUC değerlerine ve nasıl elde edildiklerine ilişkin bulgulara Tablo 59’da yer verildi.

Normal dağılımdan üretilen verilerde, Naif Bayes (NB) ile C5.0 algoritmalarında veri dönüşümü sonrası ham veriden elde edilen sonuçlara göre bir artış gerçekleşmedi. NB ile gözlem sayısı ve sınıf dağılım oranlarına göre değişkenlik göstermeksizin diğer algoritmalarından daha yüksek AUC değerine ulaşıldığı görüldü. DVM algoritmasında 100 gözlemden büyük verilerde en yüksek AUC değerlerine en sık z-skor normalizasyonu ile ulaşıldığı dikkat çekmektedir.

Ki-kare dağılımından üretilen verilerde, naif Bayes’te veri dönüşümü sonrası herhangi bir artış gerçekleşmedi. Tüm gözlem sayılarında ve sınıf dağılım oranlarında DVM algoritması ile en yüksek AUC değerlerine ulaşıldı. Fakat 10000 gözlemden 0,2-0,3-0,4-0,5 sınıf dağılımlarında NB ile DVM aynı yüksek AUC değerlerini verdi. DVM ile elde edilen en yüksek değerlerde sınıfların tam dengeli dağılımında z-skor normalizasyonu, dengesiz dağılımlarda ise minimum-maksimum normalizasyonu üstünlüğü dikkat çekmektedir.

F dağılımından üretilen verilerde, NB ve DVM EF diskritizasyonu sonrasında en yüksek AUC değerlerini verdiği görüldü. Gözlem sayısına veya sınıf dağılımına bağlı değişkenlik göstermeksizin naif Bayes algoritmasıyla en yüksek değerlere ulaşıldı.

C5.0 algoritmasında tüm dağılımlardan üretilen veriler üzerinde yapılan dönüşümler sonrasında AUC değerlerinde genellikle artış sağlanmadı.

Tablo 60. Çalışmada elde edilen en yüksek duyarlılık oranları

Duyarlılık oranları													
Dağılım	Oran	N=100			N=500			N=1000			N=10000		
		NB	C5.0	DVM	NB	C5.0	DVM	NB	C5.0	DVM	NB	C5.0	DVM
Normal	0,1	0,475 H,M,Z	0,349 H,M,Z	0,331 M	0,606 H,M,Z	0,435 H,M,Z	0,573 M	0,622 H,M,Z	0,477 H,M,Z	0,612 M	0,636 H,M,Z	0,555 H,M,Z	0,623 Z
	0,2	0,670 H,M,Z	0,528 H,M,Z	0,618 M	0,725 H,M,Z	0,594 H,M,Z	0,729 M	0,732 H,M,Z	0,614 H,M,Z	0,717 M	0,738 H,M,Z	0,669 H,M,Z	0,730 Z
	0,3	0,762 H,M,Z	0,613 H,M,Z	0,742 M	0,793 H,M,Z	0,683 H	0,773 M	0,796 H,M,Z	0,699 H,M,Z	0,782 Z	0,801 H,M,Z	0,740 H,M,Z	0,796 Z
	0,4	0,823 H,M,Z	0,690 H,M,Z	0,827 M	0,843 H,M,Z	0,743 H,M,Z	0,822 Z	0,841 H,M,Z	0,758 H,M,Z	0,832 Z	0,845 H,M,Z	0,792 H,M,Z	0,842 Z
	0,5	0,866 H,M,Z	0,745 H,M,Z	0,830 Z	0,879 H,M,Z	0,795 H,M,Z	0,866 Z	0,879 H,M,Z	0,803 H,M,Z	0,869 Z	0,881 H,M,Z	0,834 H,M,Z	0,880 Z
Ki-kare	0,1	0,248 H,M,Z	0,264 H,M,Z	0,172 Z	0,369 H,M,Z	0,336 H,M,Z	0,408 M	0,390 H,M,Z	0,360 H,M,Z	0,468 M	0,444 H,M,Z,EF	0,399 H,M,Z	0,522 M
	0,2	0,474 H,M,Z	0,437 H,M,Z	0,468 M	0,557 H,M,Z	0,497 H,M,Z	0,658 M	0,579 H,M,Z	0,514 H,M,Z	0,674 M	0,610 EF	0,555 H,M,Z	0,632 M
	0,3	0,613 H,M,Z	0,546 H,M,Z	0,679 M	0,682 H,M,Z	0,602 H,M,Z	0,758 M	0,688 H,M,Z	0,617 H,M,Z	0,733 M	0,709 H,M,Z	0,655 H,M,Z	0,730 M
	0,4	0,701 H,M,Z	0,626 H,M,Z	0,793 M	0,762 H,M,Z	0,681 H,M,Z	0,790 M	0,770 H,M,Z	0,694 H,M,Z	0,800 M	0,783 H,M,Z	0,729 H,M,Z	0,802 M
	0,5	0,783 H,M,Z	0,704 H,M,Z	0,821 M	0,827 H,M,Z	0,748 H,M,Z	0,835 Z	0,829 H,M,Z	0,757 H,M,Z	0,837 Z	0,838 H,M,Z	0,791 H,M,Z	0,849 M
F	0,1	0,497 EF	0,380 H,M,Z	0,148 H	0,573 EF	0,549 H,M	0,394 EF	0,612 EF	0,632 H,M,Z	0,476 EF	0,673 EF	0,712 H,M,Z	0,651 EF
	0,2	0,681 EF	0,563 H,M,Z	0,375 EF	0,755 EF	0,705 H,M,Z	0,624 EF	0,776 EF	0,735 H,M,Z	0,683 EF	0,803 EF	0,781 H,M,Z	0,788 EF
	0,3	0,772 EF	0,670 H,M,Z	0,539 EF	0,832 EF	0,775 H,M,Z	0,738 EF	0,842 EF	0,793 H,M,Z	0,782 EF	0,853 EF	0,830 H,M,Z	0,845 EF
	0,4	0,801 EF	0,732 H,M,Z	0,663 EF	0,870 EF	0,819 H,M,Z	0,805 EF	0,878 EF	0,829 H,M,Z	0,834 EF	0,886 EF	0,866 H,M,Z	0,880 EF
	0,5	0,838 EF	0,784 H,M	0,762 EF	0,901 EF	0,851 H,M,Z	0,843 EF	0,907 EF	0,863 H,M,Z	0,874 EF	0,913 Z	0,890 H,M,Z	0,907 EF

H: Ham veri; **M:** Minimum-Maksimum Normalizasyonu; **Z:** Z-Skor Normalizasyonu; **EG:** Eşit Genişlikte Aralıklandırma Diskritizasyonu; **EF:** Eşit Frekansta Aralıklandırma Diskritizasyonu.

Çalışmada, naif Bayes, C5.0 ve destek vektör makinaları algoritmaları ile elde edilen en yüksek duyarlılık oranlarına ve nasıl elde edildiklerine ilişkin bulgulara Tablo 60'da yer verildi.

Normal dağılımdan üretilen verilerde, Naif Bayes (NB) ile C5.0 algoritmalarında veri dönüşümü sonrası ham veriden elde edilen duyarlılık oranlarına göre bir artış gerçekleşmedi. 100 gözlem ve 0,4 sınıf dağılım oranı dışındaki tüm veri tiplerinde ham veriye uygulanan Gaussian olasılık yoğunluk fonksiyonu ile NB en yüksek duyarlılık oranlarına sahiptir. DVM bulgularına bakıldığında, gözlem sayısı arttıkça ve sınıflarda denge oranı azaldıkça z-skor normalizasyonunun minimum maksimuma göre üstünlüğü arttığı görülmektedir.

Ki-kare dağılımından üretilen verilerde, veri dönüşümü sonrası C5.0 algoritmasında duyarlılık oranlarında herhangi bir artış gerçekleşmedi. 100 gözlem ve 0,1-0,2 sınıf dağılım oranlarında DVM algoritması ile en yüksek duyarlılık oranlarına ulaşıldı. DVM ile elde edilen en yüksek değerlerde minimum-maksimum normalizasyonunun üstünlüğü dikkat çekmektedir.

F dağılımından üretilen verilerde, veri dönüşümü sonrası C5.0 algoritmasında duyarlılık oranlarında bir artış gerçekleşmedi. NB ve DVM her biri kendi içinde olmak üzere en yüksek duyarlılık oranlarını en sık EF diskritizasyonu sonrasında verdiği görüldü. Gözlem sayısına veya sınıf dağılımına bağlı olmaksızın naif Bayes algoritmasıyla duyarlılık oranlarında en iyi sonuçlara ulaşıldı.

Tablo 61. Çalışmada elde edilen en yüksek seçicilik oranları

Seçicilik oranları													
Dağılım	Oran	N=100			N=500			N=1000			N=10000		
		NB	C5.0	DVM	NB	C5.0	DVM	NB	C5.0	DVM	NB	C5.0	DVM
Normal	0,1	0,986 H,M,Z	0,998 EF	0,999 H	0,988 H,M,Z	0,984 EF	0,999 H	0,988 H,M,Z,EG	0,982 EF	0,997 H	0,989 H,M,Z	0,980 EG	0,996 H
	0,2	0,965 H,M,Z	0,981 EF	0,993 H	0,972 H,M,Z	0,939 EG	0,966 Z	0,973 H,M,Z	0,945 EG	0,990 H	0,973 H,M,Z	0,958 EG	0,990 H
	0,3	0,940 H,M,Z	0,945 EF	0,983 H	0,951 H,M,Z	0,896 EG,EF	0,982 H	0,953 H,M,Z	0,905 EG	0,981 H	0,954 H,M,Z	0,929 EG	0,975 H
	0,4	0,907 H,M,Z	0,808 H,M,Z,EG	0,960 H	0,928 H,M,Z	0,850 H,M,Z	0,961 H	0,928 H,M,Z	0,863 EG	0,969 H	0,930 H,M,Z	0,895 EG	0,935 H
	0,5	0,876 H,M,Z	0,762 H,M,Z	0,860 M	0,895 H,M,Z	0,803 H,M,Z	0,873 Z	0,897 H,M,Z	0,818 H,M,Z	0,880 Z	0,900 H,M,Z	0,857 H,M,Z,EG	0,896 Z
Ki-kare	0,1	0,975 H,M,Z	0,999 EF	0,990 Z	0,983 H,M,Z	0,991 EF	0,995 EF	0,985 H,M,Z	0,987 EF	0,994 H	0,985 H,M,Z	0,980 EG	0,992 H
	0,2	0,939 H,M,Z	0,988 EF	0,985 EF	0,955 H,M,Z	0,941 EG	0,974 H	0,958 H,M,Z	0,943 EG	0,969 H	0,960 H,M,Z	0,944 EG	0,966 H
	0,3	0,899 H,M,Z	0,944 EF	0,947 EG	0,920 H,M,Z	0,885 EG	0,944 H	0,926 H,M,Z	0,890 EG	0,935 H	0,929 H,M,Z	0,901 EG	0,936 H
	0,4	0,844 H,M,Z	0,781 EG	0,858 EG	0,879 H,M,Z	0,819 EG	0,900 H	0,885 H,M,Z	0,829 EF	0,888 H	0,893 H,M,Z	0,854 EF	0,891 H
	0,5	0,777 H,M,Z	0,707 H,M,Z	0,801 M	0,827 H,M,Z	0,751 H,M,Z	0,836 Z	0,834 H,M,Z	0,766 H,M,Z	0,837 Z	0,848 H,M,Z	0,796 H,M,Z	0,844 Z,EF
F	0,1	0,982 EG	0,995 EF	0,998 M,Z	0,991 EG	0,993 EG	0,998 M	0,992 EG	0,994 EG	0,998 M	0,996 EG	0,996 EG	0,998 M
	0,2	0,962 EG	0,975 EG	0,991 M	0,978 EG	0,977 EG	0,993 M	0,981 EG	0,982 EG	0,993 M	0,991 EG	0,992 EG	0,995 M
	0,3	0,924 H,Z	0,934 EG	0,976 M	0,950 H,Z	0,948 EG	0,979 M	0,957 EF	0,958 EG	0,981 M	0,980 EG	0,979 EG	0,981 M
	0,4	0,905 H	0,842 H,M,Z	0,945 M	0,934 H,Z	0,897 EF	0,925 M	0,938 EF	0,911 EF	0,926 M	0,946 EF	0,937 EG	0,943 EF
	0,5	0,875 H,Z	0,789 H,M,Z	0,782 EF	0,908 H,Z	0,867 H,M,Z	0,857 EF	0,917 EF	0,887 EF	0,888 EF	0,925 EF	0,911 H,M,Z	0,921 EF

H: Ham veri; **M:** Minimum-Maksimum Normalizasyonu; **Z:** Z-Skor Normalizasyonu; **EG:** Eşit Genişlikte Aralıklandırma Diskritizasyonu; **EF:** Eşit Frekansta Aralıklandırma Diskritizasyonu.

Çalışmada, naif Bayes, C5.0 ve destek vektör makinaları algoritmaları ile elde edilen en yüksek seçicilik oranlarına ve nasıl elde edildiklerine ilişkin bulgulara Tablo 61’de yer verildi.

Normal dağılımdan üretilen verilerde, NB algoritmalarında veri dönüşümü sonrası ham veriden elde edilen seçicilik oranlarına göre bir artış gerçekleşmedi. Sınıfların tam dengeli (0,5) dağıldığı verilerde naif Bayes (Gaussian), tüm dengesiz dağılımlarda ise DVM ile en yüksek duyarlılık oranlarına ulaşıldı. Her iki algoritma da en sık ham veri ile en yüksek değerlerini elde etti. C5.0 algoritmasında 100 gözlemden büyük, özellikle dengesiz sınıf dağılımlarında EG diskritizasyonunun, seçicilik oranını artırıcı yöndeki etkisi dikkat çekmektedir.

Ki-kare dağılımından üretilen verilerde, veri dönüşümü sonrası naif Bayes’te herhangi bir artış gerçekleşmedi. 100 gözlemde 0,1-0,2 sınıf dağılımlarında C5.0 (EF), 10000 gözlemde 0,4-0,5 sınıf dağılımlarında NB (Gaussian) en yüksek seçicilik oranlarına ulaştırırken geri kalan tüm veri tiplerinde DVM ile en yüksek seçicilik oranları elde edildi. DVM’de 100 gözlemden büyük ve sınıf dağılımları dengesiz yapıdaki verilerde en sık ham veri ile tam dengeli (0,5) dağılımda ise z-skor normalizasyonu ile en yüksek değerler elde edildi. C5.0 algoritmasında 100 gözlemde dengesiz dağılımlarda EF diskritizasyonunun üstünlüğü dikkat çekerken daha büyük gözlemlerde EG diskritizasyonu ile daha sık yüksek seçicilik oranlarına ulaşıldı.

F dağılımından üretilen verilerde, tüm gözlem sayıları için tam dengeli dağılımlarda NB’nin en yüksek seçicilik başarısı gösterdiği görüldü. 100 gözlemde 0,4 sınıf dağılımında yine NB en yüksek başarıyı gösterirken geri kalan tüm dengesiz sınıf dağılımlarında DVM ile en yüksek seçicilik oranlarına ulaşıldı. DVM kendi içinde tam dengeli dağılımda en yüksek oranına EF ile dengesiz dağılımlarda ise minimum-maksimum normalizasyonu ile ulaştı. EG diskritizasyonu özellikle dengesiz sınıf dağılımlarında seçicilik oranını artırıcı yönde en fazla C5.0 algoritması üzerinde etkili olduğu görülmektedir.

Tablo 62. Çalışmada elde edilen en yüksek pozitif kestirim değerleri

Pozitif Kestirim Değerleri													
Dağılım	Oran	N=100			N=500			N=1000			N=10000		
		NB	C5.0	DVM	NB	C5.0	DVM	NB	C5.0	DVM	NB	C5.0	DVM
Normal	0,1	0,828 H,M,Z	0,471 H,M,Z	0,837 M	0,858 H,M,Z	0,604 EG	0,826 Z	0,860 H,M,Z	0,652 EG	0,830 Z	0,865 H,M,Z	0,742 EG	0,851 Z
	0,2	0,853 H,M,Z	0,603 H,M,Z	0,825 M	0,872 H,M,Z	0,677 EG	0,842 Z	0,874 H,M,Z	0,716 EG	0,849 Z	0,873 H,M,Z	0,790 EG	0,866 Z
	0,3	0,863 H,M,Z	0,678 H,M,Z	0,841 Z	0,876 H,M,Z	0,731 H,M,Z	0,853 Z	0,880 H,M,Z	0,750 EG	0,860 Z	0,881 H,M,Z	0,814 EG	0,877 Z
	0,4	0,868 H,M,Z	0,724 H,M,Z	0,844 Z	0,888 H,M,Z	0,772 H,M,Z	0,862 Z	0,887 H,M,Z	0,785 H,M,Z	0,871 Z	0,889 H,M,Z	0,833 EG	0,885 Z
	0,5	0,883 H,M,Z	0,773 H,M,Z	0,865 M	0,895 H,M,Z	0,805 H,M,Z	0,874 Z	0,89 H,M,Z	0,817 H,M,Z	0,880 Z	0,898 H,M,Z	0,854 H,M,Z	0,895 Z
Ki-kare	0,1	0,559 H,M,Z	0,362 H,M,Z	0,718 M	0,721 H,M,Z	0,470 H,M,Z	0,744 M	0,749 H,M,Z	0,524 EG	0,753 Z	0,770 H,M,Z	0,626 EG	0,759 Z
	0,2	0,694 H,M,Z	0,516 H,M,Z	0,772 Z	0,765 H,M,Z	0,604 EF	0,775 Z	0,780 H,M,Z	0,634 EG	0,778 Z	0,793 H,M,Z	0,688 EG	0,789 Z
	0,3	0,743 H,M,Z	0,602 H,M,Z	0,782 Z	0,789 H,M,Z	0,662 EG	0,798 Z	0,801 H,M,Z	0,682 EG	0,803 Z	0,812 H,M,Z	0,731 EF	0,810 Z
	0,4	0,767 H,M,Z	0,669 H,M,Z	0,802 Z	0,811 H,M,Z	0,713 M	0,817 Z	0,819 H,M,Z	0,724 H,M,Z	0,820 Z	0,829 H,M,Z	0,765 EF	0,827 Z
	0,5	0,789 H,M,Z	0,718 H,M,Z	0,817 M	0,829 H,M,Z	0,753 M	0,837 Z	0,834 M,Z	0,766 H,M,Z	0,839 Z	0,846 H,M,Z	0,793 H,M,Z	0,844 Z
F	0,1	0,538 EF	0,576 H,M,Z	0,894 Z	0,756 EF	0,786 EG	0,940 M	0,795 EF	0,833 Z	0,938 M	0,871 H,Z	0,874 EG	0,935 M
	0,2	0,698 EF	0,708 EF	0,879 M	0,846 EF	0,794 EG	0,879 Z	0,857 EF	0,833 H,M,Z	0,923 M	0,893 M,Z	0,880 Z	0,934 M
	0,3	0,786 H	0,754 H,M,Z	0,859 M	0,879 EF	0,840 EF	0,884 M	0,894 EF	0,840 H,M,Z	0,891 M	0,904 EF	0,883 H,M	0,902 EF
	0,4	0,829 H	0,777 H,M,Z	0,824 M	0,898 EF	0,840 EF	0,851 EF	0,906 EF	0,862 H,M,Z	0,875 EF	0,916 EF	0,895 H,M,Z	0,911 EF
	0,5	0,861 H	0,802 H,M,Z	0,790 EF	0,908 EF	0,868 H,M,Z	0,856 EF	0,917 EF	0,886 H,M,Z	0,887 EF	0,924 EF	0,910 H,M,Z	0,920 EF

H: Ham veri; **M:** Minimum-Maksimum Normalizasyonu; **Z:** Z-Skor Normalizasyonu; **EG:** Eşit Genişlikte Aralıklandırma Diskritizasyonu; **EF:** Eşit Frekansta Aralıklandırma Diskritizasyonu.

Çalışmada, naif Bayes, C5.0 ve destek vektör makinaları algoritmaları ile tüm simülasyon senaryolarından elde edilen en yüksek pozitif kestirim değerlerine (PKD) ve nasıl elde edildiklerine ilişkin bulgulara Tablo 62’de yer verildi.

Normal dağılımdan üretilen verilerde, NB algoritmasında veri dönüşümü sonrası ham veriden elde edilen PKD’ye göre bir artış gerçekleşmedi. 100 gözlemde ve 0,1 sınıf dağılımında DVM (M) ile geri kalan tüm veri tiplerinde NB (Gaussian) ile en yüksek PKD elde edildi. C5.0 algoritmasında 100 gözlemden büyük, özellikle dengesiz sınıf dağılımlarında EG diskritizasyonunun etkisi dikkat çekmektedir.

Ki-kare dağılımından üretilen verilerde, veri dönüşümü naif Bayes’te herhangi bir artış gerçekleştirmedi. 1000 gözlem ve 0,2 sınıf dağılımı ve 10000 gözlem dışında kalan tüm veri tiplerinde PKD için DVM en yüksek başarıyı gösterdi. 10000 gözlemde en yüksek PKD’ye NB (Gaussian) ile ulaşıldı. DVM’de en yüksek pozitif kestirim değerleri en sık z-skor dönüşümü ile elde edildi. C5.0 algoritması kendi içinde, 100 gözlemde ham veri ile en yüksek seçicilik oranlarına ulaşırken dengesiz sınıf dağılımları için özellikle 1000 gözlemde EG’nin diğer dönüşümlere veya ham veriye göre üstünlüğü göze çarpmaktadır.

F dağılımından üretilen verilerde, 100 ile 500 gözlem sayıları için 0,4-0,5 sınıf dağılımlarında ve 1000 ile 10000 gözlem sayıları için 0,3-0,4-0,5 sınıf dağılımlarında NB’nin en yüksek PKD’ye ulaştığı görüldü. Geri kalan tüm dengesiz sınıf dağılımlarında DVM en yüksek başarıyı gösterdi. En yüksek PKD’yi NB’de 100 gözlemde ham veri, diğer gözlemlerde EF ile ulaşılırken DVM’de minimum maksimum normalizasyonunun başarı sıklığı dikkat çekmektedir. C5.0 algoritmasında kendi içinde PKD başarısı en sık ham veri ile elde edildiği görülmektedir.

Tablo 63. Çalışmada elde edilen en yüksek negatif kestirim değerleri

Negatif Kestirim Değerleri													
Dağılım	Oran	N=100			N=500			N=1000			N=10000		
		NB	C5.0	DVM	NB	C5.0	DVM	NB	C5.0	DVM	NB	C5.0	DVM
Normal	0,1	0,946 H,M,Z	0,930 H,M,Z	0,932 M	0,958 H,M,Z	0,939 H,M,Z	0,954 M	0,959 H,M,Z	0,943 H,M,Z	0,958 M	0,961 H,M,Z	0,952 H,M,Z	0,959 Z
	0,2	0,924 H,M,Z	0,887 H,M,Z	0,913 M	0,935 H,M,Z	0,902 H,M,Z	0,934 M	0,936 H,M,Z	0,907 H,M,Z	0,932 M	0,937 H,M,Z	0,920 H,M,Z	0,935 Z
	0,3	0,907 H,M,Z	0,844 H,M,Z	0,897 M	0,916 H,M,Z	0,868 H,M,Z	0,906 M	0,917 H,M,Z	0,875 H,M,Z	0,910 Z	0,918 H,M,Z	0,892 H,M,Z	0,916 Z
	0,4	0,891 H,M,Z	0,805 H,M,Z	0,889 M	0,900 H,M,Z	0,834 H,M,Z	0,886 Z	0,898 H,M,Z	0,843 H,M,Z	0,892 Z	0,900 H,M,Z	0,866 H,M,Z	0,898 Z
	0,5	0,877 H,M,Z	0,763 H,M,Z	0,844 Z	0,883 H,M,Z	0,799 H,M,Z	0,869 Z	0,882 H,M,Z	0,808 H,M,Z	0,872 Z	0,884 H,M,Z	0,838 H,M,Z	0,882 Z
Ki-kare	0,1	0,924 EF	0,921 H,M,Z	0,916 M,Z	0,934 H,M,Z	0,929 H,M,Z	0,938 M	0,936 H,M,Z	0,931 H,M,Z	0,943 M	0,941 H,M,Z,EF	0,936 H,M,Z	0,948 M
	0,2	0,880 H,M,Z	0,865 H,M,Z	0,882 M	0,897 H,M,Z	0,879 H,M,Z	0,917 M	0,901 H,M,Z	0,883 H,M,Z	0,920 M	0,908 H,M,Z	0,894 H,M,Z	0,912 M
	0,3	0,849 H,M,Z	0,815 H,M,Z	0,874 M	0,872 H,M,Z	0,836 H,M,Z	0,897 M	0,874 H,M,Z	0,842 H,M,Z	0,889 M	0,882 H,M,Z	0,858 H,M,Z	0,888 M
	0,4	0,815 H,M,Z	0,767 H,M,Z	0,868 M	0,848 H,M,Z	0,795 H,M,Z	0,864 M	0,853 H,M,Z	0,802 H,M,Z	0,869 M	0,860 H,M,Z	0,825 H,M,Z	0,869 M
	0,5	0,792 H,M,Z	0,716 H,M,Z	0,827 M	0,829 H,M,Z	0,752 H,M,Z	0,838 Z	0,831 H,M,Z	0,761 H,M,Z	0,838 Z	0,840 H,M,Z	0,792 H,M,Z	0,847 M
F	0,1	0,945 EF	0,934 H,M,Z	0,914 H,Z	0,954 EF	0,951 H,M,Z	0,936 EF	0,958 EF	0,960 H,M,Z	0,944 EF	0,964 H,M,Z,EF	0,969 H,M,Z	0,962 EF
	0,2	0,923 EF	0,897 H,M,Z	0,864 EF	0,941 EF	0,928 H,M,Z	0,912 EF	0,946 EF	0,936 H,M,Z	0,924 EF	0,952 EF	0,947 H,M,Z	0,948 EF
	0,3	0,906 EF	0,869 H,M,Z	0,829 EF	0,930 EF	0,907 H,M,Z	0,894 EF	0,934 EF	0,914 H,M,Z	0,910 EF	0,939 EF	0,929 H,M,Z	0,935 EF
	0,4	0,875 EF	0,833 H,M,Z	0,802 EF	0,916 EF	0,883 H,M,Z	0,876 EF	0,921 EF	0,890 H,M,Z	0,893 EF	0,926 EF	0,913 Z	0,922 EF
	0,5	0,850 EF	0,797 H,M,Z	0,780 EF	0,903 EF	0,856 H,M,Z	0,846 EF	0,908 EF	0,869 H,M	0,877 EF	0,913 H,M,Z	0,893 H,M,Z	0,908 EF

H: Ham veri; **M:** Minimum-Maksimum Normalizasyonu; **Z:** Z-Skor Normalizasyonu; **EG:** Eşit Genişlikte Aralıklandırma Diskritizasyonu; **EF:** Eşit Frekansta Aralıklandırma Diskritizasyonu.

Çalışmada, naif Bayes, C5.0 ve destek vektör makinaları algoritmaları ile elde edilen en yüksek negatif kestirim değerlerine (NKD) ve nasıl elde edildiklerine ilişkin bulgulara Tablo 63'te yer verildi.

Normal dağılımdan üretilen verilerde, NB ve C5.0 algoritmalarında veri dönüşümü sonrası ham veriden elde edilen NKD'ye göre bir artış gerçekleşmedi. Tüm veri tiplerinde en yüksek NKD NB (Gaussian) ile elde edildi. DVM bulgularına bakıldığında, gözlem sayısı arttıkça ve sınıflarda denge oranı azaldıkça z-skor normalizasyonunun minimum maksimuma göre üstünlüğü arttığı görülmektedir.

Ki-kare dağılımından üretilen verilerde, C5.0 algoritmasında veri dönüşümü sonrası ham veriden elde edilen NKD'ye göre bir artış gerçekleşmedi. 100 gözlemde ve 0,1 sınıf dağılımı dışındaki tüm veri tiplerinde DVM ile en yüksek NKD elde edildi. DVM'de en yüksek değerlere ulaşımında minimum-maksimum normalizasyon etkisinin yüksek olduğu görüldü. NB'de en sık olarak ham veri ile yüksek değerlere ulaşıldı.

F dağılımından üretilen verilerde, C5.0 algoritmasında veri dönüşümünün NKD'yi artırıcı yönde bir etkisi olmadı. 1000 ile 10000 gözlem sayıları için 0,1 sınıf dağılımında en yüksek NKD'ye C5.0 ile geri kalan tüm veri tiplerinde ise NB (EF) ile ulaşıldığı görüldü. DVM ve NB'de EF diskritizasyonu etkisi dikkat çekmektedir.

TARTIŞMA

Tez çalışmasında, naif Bayes, C5.0 ve destek vektör makinesi sınıflandırma algoritmaları kullanılarak farklı örneklem dağılımlarından simülasyon ile farklı gözlem sayılarında ve sınıfların farklı denge dağılım oranlarında üretilen verilerin ham hali ile birlikte dört farklı dönüşüm yöntemi ile elde edilen veriler üzerinden sonuçlar alındı. Üç algoritmanın, ham veri ile birlikte veri dönüşümleri sonrası performansları altı farklı performans değerlendirme kriteri ile karşılaştırıldı.

Yapılan çalışmada genel performans açısından naif Bayes, C5.0 ve destek vektör makineleri algoritmaları karşılaştırıldığında normal dağılım ve F dağılımından üretilen verilerde naif Bayes'in, ki-kare dağılımından üretilen verilerde ise destek vektör makinelerinin daha başarılı sonuçlar verdiği görüldü. Üç örneklem dağılımından üretilen ve herhangi bir ön-işlemden geçirilmeyen verilerden alınan sonuçlar ile normalizasyon sonrasında alınan sonuçlar karşılaştırıldığında, normalizasyon yöntemlerinin naif Bayes algoritmasının sınıflandırma performansı üzerinde genellikle artış sağlamadığı görüldü. C5.0 karar ağacı algoritmasında normalizasyon genellikle ham veriden alınan sonuçlara göre farklılık yaratmadı. DVM'de ise normalizasyon dönüşümleri ile seçicilik oranı dışındaki tüm performans ölçülerinde genellikle artış sağlandı.

Kumar ve ark. (119) Alzheimer verisi üzerinde minimum-maksimum normalizasyonunun doğrusal diskriminant analizi, k-en yakın komşu, naif Bayes, destek vektör makineleri, karar ağaçları, rastgele orman ve çok katmanlı algılayıcı algoritmalarının performanslarındaki rolünü araştırdı. Çalışmalarında algoritmaların genel performans

karşılaştırmasında naif Bayes daha iyi sonuçlar verdi. Minimum-maksimum normalizasyonunun karar ağaçları ve naif Bayes algoritmalarının performansını fazla etkilemediğini, DVM algoritmasının ise performansında önemli iyileşme yarattığını bildirdi.

Liu ve Li, (120) yaptıkları çalışmada minimum-maksimum normalizasyon ile maksimum normalizasyonun DVM algoritmasının performansındaki etkilerini karşılaştırdı. Çalışmalarında minimum-maksimum normalizasyonunun maksimum normalizasyonuna göre daha başarılı oldu. Minimum-maksimum normalizasyonunun DVM sınıflandırmasında hem zaman hem de performans açısından iyileşmeler sağladığını bildirdi.

Siledar ve Chaudhary (121), karar ağacında (C4.5) ve naif Bayes'te z-skor normalizasyonunun UCI veri ambarından edindikleri 30000 gözlemlili 24 değişkenli "Credit" verisi üzerinde etkisini araştırdı. Z-skor normalizasyonu sonrasında karar ağacında (C4.5) bir değişme olmadığını ancak NB'de artış olduğunu bildirdi. Çalışmamızda NB'de z-skor normalizasyonu sonrasında artış gerçekleşmemiştir. Bu farklılığın nedeni olarak normal dağılmayan veriler için "Gaussian" yerine "Kernel" olasılık yoğunluk fonksiyonunun tercih etmemizden kaynaklandığı düşünülmektedir.

Tez çalışmasında, DVM algoritmasında normal dağılımdan üretilen verilerde özellikle gözlem sayılarının artmasıyla z-skor dönüşümünün performansı artırmada daha etkili olduğu görüldü. Ki-kare dağılımından üretilen verilerde tam dengeli sınıf dağılımında z-skor normalizasyonu, sınıfların dengesiz dağıldığı verilerde ise minimum-maksimum normalizasyonu daha yüksek sonuçlar verdi. F dağılımından üretilen verilerde ise minimum-maksimum normalizasyonu performanslarda düşüşe sebep olurken 100 gözlemden büyük verilerde z-skor normalizasyonu ile daha başarılı sonuçlar elde edildi.

Zuriani ve Yuhanis (122), 5 farklı yerden topladıkları her biri 104 gözlemden oluşan verileri kullanarak normalizasyon yöntemlerini karşılaştırdı. En küçük kareler destek vektör makinelerinde minimum-maksimum normalizasyon ile z-skora normalizasyonuna göre daha iyi sonuçlar elde ettiler.

Çalışmamızda seçicilik oranları dönüşüm sonuçları açısından DVM'de diğer performans ölçülerinden farklılık gösterdi. DVM algoritmasında diğer performans kriterlerinde normalizasyon ile genellikle artış olurken seçicilik oranlarında, yalnızca tam dengeli sınıf dağılımlarında artış gerçekleşti. Dengesiz sınıf dağılımlarında ise normal dağılan verilerde genellikle artış gerçekleşmezken ki-kare dağılımında 100 gözlemlili verilerde artış

sağlandı. F dağılımında ise dengesiz sınıf dağılımlarında gözlem sayılarına göre değişmeksizin minimum-maksimum normalizasyonu ile daha yüksek seçicilik oranlarına ulaşıldı. C5.0 ve NB’de diğer performans kriterlerinde olduğu gibi artış gerçekleşmedi.

Suma ve ark. (123) UCI veri ambarından edindikleri “Breast” verisi üzerinde normalizasyon yöntemlerini doğrusal diskriminant analizi, destek vektör makineleri, naif Bayes, yapay sinir ağları ve karar ağacı algoritmalarında genel doğruluk oranı, duyarlılık ve seçicilik oranları açısından karşılaştırdı. Çalışmalarında en iyi performansı naif Bayes ile elde ettiler. NB’de minimum-maksimum normalizasyonu ile performans değerlerinde artış gerçekleşmezken z-skor normalizasyonu ile genel doğruluk oranlarında artış sağlandığı gösterildi. DVM için her iki yöntemin doğruluk oranlarını eşit olarak arttırdığını, duyarlılık ve seçicilik oranlarını ise minimum-maksimum normalizasyonunun arttırdığını gösterdiler. Suma ve ark.’nın çalışmasıyla çalışmamızdaki NB bulguları karşılaştırıldığında, bu farklılığın normal dağılmayan veriler için “Gaussian” yerine “Kernel” olasılık yoğunluk fonksiyonunun tercih etmemizden kaynaklandığı düşünülmektedir.

Çalışmamızda gözetimsiz diskritizasyon yöntemleri ile normal dağılım ve ki-kare dağılımından üretilen verilerde NB’nin performansında düşüş gerçekleştiği görüldü. F dağılımından üretilen ve gözetimsiz diskritizasyon yöntemlerinden eşit frekansta aralıklandırma (EF) yöntemi ile kategorik hale dönüştürülen verilerin kullanıldığı NB sınıflandırmasında ham veriden elde edilen sonuçlara göre daha başarılı sonuçlar alındı.

Yang ve ark. (124) UCI veri ambarından edindikleri 35 veri seti üzerinde farklı diskritizasyon yöntemlerini naif Bayes algoritmasının sınıflandırma sonucundaki hata oranlarını karşılaştırdı. Karşılaştırmalarında EF’nin EG’ye göre NB’de daha düşük hata oranları sağladığı görüldü.

Boulle (125), UCI veri ambarından edindiği 15 veri setini kullanarak EG ve EF yöntemleri için optimal bir k bölme sayısı belirlemek amacıyla bir çalışma gerçekleştirdi. Çalışmada EG ve EF için hesapladıkları optimal bölme sayısı ile “10” bölme sayısını, naif Bayes algoritmasının genel doğruluk oranları üzerinden karşılaştırdı. EG ve EF diskritizasyon yöntemleri “10” bölme sayısı için birbiriyle karşılaştırılmasında EF daha yüksek performans gösterdi.

Maslove ve ark. (126) laboratuvar ve fizyolojik verileri kullanarak naif Bayes ve karar ağaçlarında diskritizasyon yöntemlerini karşılaştırdı. Çalışmalarında 10 bölme sayısı için EG

ve EF dönüşümü sonrası karar ağaçlarında ham veriden daha kötü sonuçlar elde edildi. Naif Bayes'te ise laboratuvar verisinde EG performansta düşüşe neden olurken EF ile artış olduğu gösterildi. Fizyolojik veride ise bağımlı değişkendeki sınıfların klinik olarak belirlendiğinde EG EF'ye göre sınıfların kümeleme analizi uygulanarak belirlendiğinde ise EF EG'ye göre NB performansının artmasında daha etkili oldu.

Tez çalışmasında gözetimsiz diskritizasyon yöntemlerinin C5.0 algoritmasının doğruluk oranı, AUC değeri, duyarlılık oranı ve negatif kestirim değeri performanslarında düşüşe neden olduğu görüldü. Sınıflandırma performansları seçicilik oranı açısından ele alındığında, sınıfların dengesiz dağıldığı verilerde özellikle gözlem sayısının artmasıyla EG'nin daha iyi sonuçlar verdiği görüldü. Pozitif kestirim değeri açısından veri dönüşümlerinin performansa olan etkisi incelendiğinde, normal ve ki-kare dağılımından üretilen, sınıfların dengesiz dağıldığı, 1000 ve üstü gözlem sayılı verilerde EG'nin üstünlüğü dikkat çekmektedir.

Jiang ve ark. (127) UCI veri ambarından edindiği verileri kullanarak farklı sınıflandırıcılarda diskritizasyon öncesi sonuçlar ile farklı diskritizasyon yöntemlerinin uygulanması sonrasında elde edilen genel doğruluk oranlarını karşılaştırdı. Bu çalışmada, diskritizasyon yöntemlerinden EG ile karar ağaçlarında (C4.5) diskritizasyon öncesine göre genel doğruluk oranlarında düşme gerçekleşirken NB'de artış gerçekleştiği gösterildi. Çalışmamızda C5.0 karar ağacı algoritmasında ve NB'de EG ile genel doğruluk oranları düşmüştür. NB bulgularındaki çelişkinin nedeni olarak normal dağılmayan veriler için "Gaussian" yerine "Kernel" olasılık yoğunluk fonksiyonunun kullanılması düşünülmektedir.

Çalışmamızda gözetimsiz diskritizasyon yöntemleri ile DVM performansında artış sağlandı. Normal dağılımdan ve F dağılımından üretilen verilerde ise DVM performansında EF diskritizasyonu EG'ye göre daha etkili oldu. Ki-kare dağılımından üretilen verilerin diskritizasyonunda genellikle EG ile daha yüksek performanslar elde edildi. Diskritizasyon sonrası ulaşılan performans değerlerine göre normal dağılımdan ve F dağılımından üretilen verilerde naif Bayes, ki-kareden üretilen verilerde ise DVM daha yüksek başarı gösterdi.

Jung ve ark. (128) Ekson-dizi verileri üzerinde eğitilen ve RNA-seq verileri üzerinde test edilen modeller için, farklı özellik seçme yöntemlerinde, gözetimsiz diskritizasyon yöntemlerini; rastgele orman, destek vektör makineleri, naif Bayes ve PAM (Prediction Analysis of Microarrays) algoritmalarının performansları üzerinden değerlendirdi. Jung ve

ark. farklı sınıflandırıcılarda ve farklı özellik seçme yöntemlerinde gözetimsiz diskritizasyon yöntemleri arasında özellikle EF'nin daha iyi performans gösterdiğini bildirdi. Çalışmalarında en iyi başarıyı DVM algoritması gösterdi.

Yapılan literatür taramasında normalizasyon ve diskritizasyon yöntemlerini birlikte inceleyen çalışmaya rastlanmadı. Çalışmamızda normalizasyon ve diskritizasyon yöntemleri bir arada incelendiğinde, dönüşüm yöntemlerinin birbirlerine olan üstünlükleri naif Bayes, C5.0 ve destek vektör makineleri sınıflandırma algoritmalarında farklılık göstermektedir. C5.0 algoritmasında her iki yöntem ile genellikle performansta artış sağlanamadı. Naif Bayes'te F dağılımından üretilen verilerde EF diskritizasyon yöntemi ile daha iyi sonuçlar elde edildi. DVM'de verilerin dağılımlarına göre normalizasyon ve diskritizasyon yöntemlerinin birbirlerine olan üstünlüklerinde değişim gözlemlendi. DVM'de normal dağılımdan üretilen verilerin büyük çoğunluğunda z-skor, ki-kare dağılımından üretilen verilerin büyük çoğunluğunda ise minimum-maksimum normalizasyonu en iyi sonuçlara ulaşmada yardımcı oldu. F dağılımından üretilen verilerde ise EF diskritizasyonu DVM'nin performansını artırmada daha etkili oldu.

SONUÇ

Tez çalışmasında, sınıflandırma öncesi uygulanan normalizasyon ve gözetimsiz diskritizasyon yaklaşımlarının naif Bayes, C5.0 ve destek vektör makinesi algoritmaları üzerindeki etkilerini incelemek amacıyla simülasyon çalışması gerçekleştirildi. Çalışmada normal dağılım, ki-kare dağılımı ve F dağılımından simülasyon ile dört farklı gözlem sayısına ve sınıfların beş farklı denge dağılım oranlarına göre üretilen verilerin minimum-maksimum ve z-skor normalizasyonu ile eşit genişlikte (EG) ve eşit frekansta (EF) aralıklandırma diskritizasyonu dönüşümleri ele alındı.

Çalışmamızda sınıflandırma öncesi uygulanan normalizasyon ve diskritizasyon yöntemlerinden en çok etkilenen algoritma destek vektör makineleri olurken C5.0 karar ağacı algoritması en az etkilenen algoritma olarak belirlendi. C5.0 algoritması verilerin dağılımı, gözlem sayısı ve sınıf dağılım oranı açılarından değişkenlik gösteren veriler karşısında veri dönüşümleri açısından daha sağlam (robust) sonuçlar verdi.

Destek vektör makinelerinin performansında hem normalizasyon hem de diskritizasyon yöntemlerinin etkili oldukları ve daha iyi sonuçlar vermesinde katkı sağladıkları görüldü. Normalizasyon yöntemlerinin birbirine olan üstünlükleri, verilerin dağılımına, gözlem sayılarına ve sınıfların dengeli dağılıp dağılmamasına bağlı olarak değişkenlik gösterdi. Normal dağılım verilerde gözlem sayılarının artmasıyla birlikte performansı artırmada z-skor dönüşümünün etkisinin de arttığı görüldü. Ki-kare dağılımından üretilen verilerde tam dengeli sınıf dağılımında z-skor normalizasyonu, dengesiz sınıf dağılımlarında ise minimum-maksimum normalizasyonu daha iyi sonuçlar verdi. F

dağılımından üretilen verilerde ise minimum-maksimum normalizasyonuna göre z-skor normalizasyonu ile dönüştürülen verilerde daha başarılı sonuçlar elde edildi. Diskritizasyon ile yapılan dönüşümlerde verilerin dağılım şekillerine göre iki yöntemin birbirine olan üstünlüğünde değişme gerçekleşti. Normal dağılım ve F dağılımından üretilen veriler için EF diskritizasyonu EG'ye göre DVM performansında daha etkili oldu. Ki-kare dağılımından üretilen verilerin diskritizasyonunda genellikle EG ile daha yüksek performanslar elde edildi. DVM ile elde edilen sınıflandırma başarılarına göre her iki yaklaşım arasından normal ve ki-kare dağılımında normalizasyon yöntemleri, F dağılımında ise EF gözetimsiz diskritizasyon yöntemi daha etkilidir.

Naif Bayes algoritmasında genellikle normalizasyon yöntemleri ile performans değerlerinde artış gerçekleşmedi. Gözetimsiz diskritizasyon yöntemlerinin etkisi ise F dağılımından üretilen veriler ile yapılan sınıflandırmada net olarak görüldü. EF yöntemi ile dönüştürülen verilerden, herhangi bir ön-işlemden geçirilmeyen ham verilere göre daha iyi sonuçlar alındı.

Çalışmada kullanılan sınıflandırma algoritmalarının birbirlerine olan üstünlükleri incelendiğinde, normal dağılım ve F dağılımından üretilen verilerde sırasıyla; ham veri ve EF ile NB, ki-kare dağılımından üretilen verilerde normalizasyon yöntemleri ile DVM'nin daha başarılı sonuçlar verdiği görüldü.

ÖZET

Tez çalışmasında, sınıflandırma öncesi uygulanan normalizasyon ve gözetimsiz diskritizasyon yaklaşımlarının naif Bayes (NB), C5.0 ve destek vektör makineleri (DVM) algoritmaları üzerindeki etkilerini incelemek amacıyla simülasyon çalışması gerçekleştirildi.

Normalizasyon ve diskritizasyon yöntemlerinin üç algoritma üzerindeki etkilerinde farklılıklar gözlemlendi. Normalizasyon yöntemleri, C5.0 karar ağacı algoritması ve NB algoritmasında performansı artırmada genellikle etkisiz kaldı. DVM algoritmasında ise normalizasyon yöntemleri ile performans ölçülerinde artış sağlandı. DVM performanslarını artırmada en etkili normalizasyon yöntemi araştırıldığında, cevabın verilerin dağılımına, gözlem sayısına ve sınıfların dağılım oranlarına bağlı olarak değiştiği görüldü.

Gözetimsiz diskritizasyon yöntemleri, C5.0 algoritmasının performansında genellikle artış sağlamazken NB ve DVM ile daha iyi sonuçlar almada yardımcı oldu. Diskritizasyon yöntemleri, NB’de yalnızca F dağılımından üretilen verilerde etkili olurken DVM’de tüm örneklem dağılımlarında etkili oldu.

Çalışmada, C5.0 algoritması veri dönüşümlerinden en az etkilenirken DVM en çok etkilenen algoritma oldu. Genel performanslar açısından algoritmaları karşılaştırdığımızda, normal ve F dağılımından üretilen verilerin sınıflandırılmasında NB, ki-kare dağılımından üretilen verilerin sınıflandırılmasında ise DVM daha iyi performans gösterdi.

Anahtar kelimeler: Veri madenciliği, Sınıflandırma, Normalleştirme, Diskritizasyon.

**THE IMPACT OF DATA TRANSFORMING METHODS ON
PERFORMANCES OF CLASSIFICATION ALGORITHMS IN DATA
MINING
SUMMARY**

In this thesis, a simulation study was performed to investigate the effects of normalization and unsupervised discretization methods on naive Bayes (NB), C5.0 and support vector machine (SVM) algorithms.

The effects of normalization and discretization methods on the three algorithms were found to be change. Normalization methods were generally ineffective in improving the performance of the C5.0 decision tree algorithm and the NB algorithm. Performance measures of the SVM algorithm were increased with normalization methods. When the most effective normalization method was investigated, it was observed that the response varies depending on the distribution of data, the number of observations and the distribution rates of the classes.

Unsupervised discretization methods have generally not improved performance of the C5.0 algorithm, but have helped to achieve better results with NB and SVM. Unsupervised discretization methods increased NB performance only in classification of the datas produced from the F distribution, whereas SVM performance increased for datas produced from all sampling distributions.

In the study, the C5.0 algorithm was least affected by data transformations, while SVM was the most affected algorithm. According to the overall performance of the algorithms, NB showed higher performance in classification of datas produced from normal and F distributions, whereas SVM performed better in classification of datas generated from chi-square distribution than the other methods.

Keywords: Data mining, Classification, Normalization, Discretization.



KAYNAKLAR

1. Elstein AS. On the origins and development of evidence-based medicine and medical decision making. *Inflammation Research* 2004;53(2),184-9.
2. Sox HC, Higgins MC, Owens DK. *Medical decision making*. Chichester: John Wiley & Sons, 2013:1-6.
3. West AF, West RR. Clinical decision-making: coping with uncertainty. *Postgrad Med J* 2002;78:319-321.
4. Bae JM. The clinical decision analysis using decision tree. *Epidemiology and health* 2014;36.
5. Tufféry S (Translate: Riesco R.). *Data mining and statistics for decision making*. Chichester: John Wiley & Sons, 2011:1-24.
6. Torgo L. *Data mining with R: learning with case studies*. Boca Raton: Taylor & Francis Group, 2017:44.
7. Lantz B. *Machine learning with R*. Birmingham: Packt Publishing Ltd, 2015:89-124.
8. Tufféry S (Translate: Riesco R.). *Data mining and statistics for decision making*. Chichester: John Wiley & Sons; 2011:301-553.
9. Kubat M. *An introduction to machine learning*. Springer, 2015:113-133.
10. Williams G. *Data mining with Rattle and R: The art of excavating data for knowledge discovery*. New York: Springer Science & Business Media, 2011:205-244.

11. Cios KJ, Swiniarski RW, Pedrycz W. Data mining: a knowledge discovery approach. New York: Springer, 2007:381-416.
12. Jiawei H, Micheline K, Data Mining: Concepts and Techniques, San Francisco: Morgan Kaufmann Publishers, 2006:285-378.
13. Williams G. Data mining with Rattle and R: The art of excavating data for knowledge discovery. New York: Springer Science & Business Media, 2011:293-304.
14. Jiawei H, Micheline K, Data Mining: Concepts and Techniques, San Francisco: Morgan Kaufmann Publishers, 2006:47-104.
15. Fayyad U, Piatetsky-Shapiro G and Smyth P. From data mining to knowledge discovery in databases. AI magazine 2016;17(3):37-54.
16. Jorgenson, D. Hunter, J. Nadiri, M. The predictive performance of econometric models of quarterly investment behavior. Econometrica 1970;38(2):213-224.
17. Elmaghraby AS, Kantardzic MM, Wachowiak MP. Data mining from multimedia patient records. In: Triantaphyllou E, Felici G (Eds). Data Mining and Knowledge Discovery Approaches Based on Rule Induction Techniques. Boston: Springer; 2006. p.551-595.
18. Lord W, Wiggins D. Medical decision support systems advances in health care technology care shaping the future of medical. In: Spekowius G, Wendler T (Eds.) Springer; 2006. p.403-419.
19. Kalemci Ö. Veri madenciliği yöntemi ile prostat kanseri için erken uyarı protokollerinin geliştirilmesi (tez). İstanbul: İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü; 2018.
20. Koh HC, Tan G. Data Mining Application in Healthcare. Journal of Healthcare Information Management 2005;19(2):64-72.
21. Schuerenberg BK. An information excavation. Health Data Management 2003;11(6):80-82.
22. Obenshain MK. Application of data mining techniques to healthcare data. Infection Control & Hospital Epidemiology 2004;25(8):690-5.
23. Milley A. Healthcare and data mining. Health Management Technology 2000;21(8):44-7.

24. Ridinger M. American Healthways uses SAS to improve patient care. *DM Review* 2002;12(139).
25. Tomar D, Agarwal S. A survey on data mining approaches for healthcare. *Int. J. Bio-Science Bio-Technology* 2013;5(5):241–266.
26. Cios KJ, Swiniarski RW, Pedrycz W. *Data mining: a knowledge discovery approach*. New York: Springer, 2007:3-24.
27. Jiawei H, Micheline K, *Data Mining: Concepts and Techniques*, San Francisco: Morgan Kaufmann Publishers, 2006:1-45.
28. Kurgan LA, Musilek P. A survey of knowledge discovery and data mining process models. *The Knowledge Engineering Review* 2006;21(1):1-24.
29. Fayyad U, Piatetsky-Shapiro G, Smyth P. The KDD process for extracting useful knowledge from volumes of data. *Communications of the ACM* 1996;39(11):27-34.
30. Fayyad UM, Piatetsky-Shapiro G, Smyth P. Knowledge discovery and data mining: towards a unifying framework. In: Simoudis E, Han J, Fayyad U (Eds) *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining: 1996 August; Oregon, Portland*. AAAI Press; 1996, p.82-8.
31. Fayyad U, Piatetsky-Shapiro G, Smyth P. From data mining to knowledge discovery in databases. *AI magazine* 1996;17(3):37.
32. Shearer C. The CRISP-DM model: the new blueprint for data mining. *Journal of data warehousing* 2000;5(4):13-22.
33. Wirth R, Hipp J. CRISP-DM: Towards a standard process model for data mining. In: *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*. Citeseer; 2000, p.29-39.
34. Cios KJ, Teresinska A, Konieczna S, Potocka J, Sharma, S. A knowledge discovery approach to diagnosing myocardial perfusion. *IEEE Engineering in Medicine and Biology Magazine* 2000;19(4):17-25.
35. Cios K, Kurgan L. Trends in data mining and knowledge discovery. In: Pal N, Jain L (Eds). *Advanced Techniques in Knowledge Discovery and Data Mining*. Springer; 2005. p.1–26.
36. Gorunescu F. *Data Mining: Concepts, models and techniques*. Berlin: Springer Science & Business Media, 2011:1-42.

37. Larose DT, Chantal DL. Discovering knowledge in data: an introduction to data mining. New Jersey: John Wiley & Sons, 2014:16-50.
38. John GH, Kohavi R, Pflieger K. Irrelevant features and the subset selection problem. In: Cohen WW, Hirsh H (Eds) Machine Learning Proceedings 1994. San Francisco: Morgan Kaufmann; 1994. p.121-9.
39. Saeys Y, Inza I, Larranaga P. A review of feature selection techniques in bioinformatics, *Bioinformatics* 2007;23(19):2507-17.
40. Budak H. Özellik seçim yöntemleri ve yeni bir yaklaşım. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi* 2018;22:21-31.
41. Maimon O, Rokach L. Introduction to knowledge discovery and data mining. In: Maimon O, Rokach L (Eds). *Data mining and knowledge discovery handbook*. Boston: Springer; 2009. p.1-15.
42. Grąbczewski K. *Meta-learning in decision tree induction*. New York: Springer International Publishing, 2014:1-9.
43. Williams G. *Data mining with Rattle and R: The art of excavating data for knowledge discovery*. New York: Springer Science & Business Media, 2011:171-6.
44. Cios KJ, Swiniarski RW, Pedrycz W. *Data mining: a knowledge discovery approach*. New York: Springer, 2007:25-91.
45. Witten IH, Frank E. *Data mining: practical machine learning tools and techniques*. San Francisco: Elsevier, 2005:285-343.
46. Yang Y, Webb GI, Wu X. Discretization methods. In: Maimon O, Rokach L (Eds). *Data mining and knowledge discovery handbook second edition*, Boston: Springer; 2010. p.101-116.
47. Yang Y, WEBB GI. On why discretization works for naive-bayes classifiers. In: Gedeon TD, Fung LCC (Eds) *Australasian Joint Conference on Artificial Intelligence*. Boston: Springer; 2003, p. 440-452.
48. Olson DL, Delen D. *Advanced data mining techniques*. Berlin: Springer Science & Business Media, 2008:87-110.
49. Liu H, Setiono R. A probabilistic approach to feature selection-a filter solution. In *ICML*. San Francisco: Morgan Kaufmann Publishers Inc: 1996, July. 1996, p.319-327.

50. Jin R, Breitbart Y, Muoh C. Data discretization unification. *Knowledge and Information Systems* 2009;19(1):1-29.
51. BAY SD. Multivariate discretization for set mining. *Knowledge and Information Systems* 2001;3(4):491-512.
52. Chmielewski MR, Grzymala-Busse JW. Global discretization of continuous attributes as preprocessing for machine learning. *International journal of approximate reasoning* 1996;15(4):319-331.
53. Dash R, Paramguru RL, Dash R. Comparative analysis of supervised and unsupervised discretization techniques. *International Journal of Advances in Science and Technology* 2001;2(3):29-37.
54. Garcia S, Luengo J, Sáez JA, Lopez V, Herrera F. A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning. *IEEE Transactions on Knowledge and Data Engineering* 2012;25(4):734-50.
55. Liu H, Hussain F, Tan CL, Dash M. Discretization: An enabling technique. *Data mining and knowledge discovery* 2002;6(4):393-423.
56. Catlett J. Megainduction: machine learning on very large databases (Dissertation). Sydney: University of Sydney Basser Department of Computer Science; 1991.
57. Kotsiantis S, Kanellopoulos D. Discretization techniques: A recent survey. *GESTS International Transactions on Computer Science and Engineering* 2006;32(1):47-58.
58. Gorunescu F. *Data Mining: Concepts, models and techniques*. Berlin: Springer Science & Business Media, 2011:159-184.
59. Gorunescu F. *Data Mining: Concepts, models and techniques*. Berlin: Springer Science & Business Media, 2011:186-191.
60. İnal C, Günay S. *Olasılık ve matematiksel istatistik*. Ankara: Hacettepe Üniversitesi; 2010.
61. Zafarani R, Mohammad AA, Huan L. *Social media mining: an introduction*. Cambridge University Press, 2014:135-171.
62. Mitchell T. *Machine Learning*. McGraw Hill, 1997:179-180.
63. Witten IH, Frank E. *Data mining: practical machine learning tools and techniques*. San Francisco: Elsevier, 2005:83-142.

64. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning. New York: Springer series in statistics, 2001:191-218.
65. Dougherty J, Kohavi R, Sahami M. Supervised and unsupervised discretization of continuous features. In: Frieditis A, Russel SJ (Eds) Proceedings of the 12th international conference on machine learning; 1995, p.194–202.
66. Stephens CR, Huerta HF, Linares AR. When is the Naive Bayes approximation not so naive?. *Machine Learning* 2018;107(2):397-441.
67. Hsu CN, Huang HJ, Wong TT. Why discretization works for naive Bayesian classifiers. In: Langley P (Ed) Proceedings of the Seventeenth International Conference on Machine Learning; 2000, p.309–406.
68. Roy N, McCallum A. Toward optimal active learning through sampling estimation of error reduction. In: Brodley CE, Danyluk AP (Eds) Proceedings of the Eighteenth International Conference on Machine Learning; 2001, p.441–8.
69. Zaffalon M, Hutter, M. Robust feature selection using distributions of mutual information. Proceedings of the Fourteenth International Conference on Uncertainty in Artificial Intelligence; 2002.
70. Yang Y, Webb GI. Discretization for naive-Bayes learning: Managing discretization bias and variance. Monash University, School of Computer Science and Software Engineering; 2003. Report No:131.
71. Warner HR, Toronto AF, Veasey LG, Stephenson R. A mathematical approach to medical diagnosis application to congenital heart disease. *JAMA* 1961;177(3):177–183.
72. Langarizadeh M, Moghbeli F. Applying Naive Bayesian Networks to Disease Prediction: a Systematic Review. *Acta Inform Med* 2016 Oct;24(5):364–369.
73. Adi K, Widodo CE, Widodo AP, Gernowo R, Pamungkas A, Syifa RA. Naive Bayes algorithm for lung cancer diagnosis using image processing techniques. *Adv Sci Lett* 2017;23(3):2296-8.
74. Yang MC, Huang CS, Chen JH, Chang RF. Whole breast lesion detection using naive bayes classifier for portable ultrasound. *Ultrasound Med Biol* 2012;38(11):1870-80.
75. Just MA, Pan L, Cherkassky VL, McMakin L, Cha C, Nock MK et al. Machine learning of neural representations of suicide and emotion concepts identifies suicidal youth. *Nature human behaviour* 2017;1(12):911.

76. Jiang W, Shen Y, Ding Y, Ye C, Zheng Y, Zhao P, et al. A naive Bayes algorithm for tissue origin diagnosis (TOD-Bayes) of synchronous multifocal tumors in the hepatobiliary and pancreatic system. *Int J Cancer* 2018;142(2):357-368.
77. Sinha AK, Singh P, Prakash A, Pal D, Dube A, Kumar A. Putative Drug and Vaccine Target Identification in *Leishmania donovani* Membrane Proteins Using Naïve Bayes Probabilistic Classifier. *IEEE/ACM Trans Comput Biol Bioinform* 2017;14(1):204-211.
78. Günakan E, Atan S, Haberal AN, Küçükyıldız İA, Gökçe E, Ayhan A. A novel prediction method for lymph node involvement in endometrial cancer: machine learning. *International Journal of Gynecologic Cancer* 2019;29(2):320-4.
79. Sayin R, Keskin S, Hamamci M. Evaluation of several classification methods in carpal tunnel syndrome. *J Pak Med Assoc* 2017;67:1654-7.
80. Lavrac N, Zupan B. *Data Mining in Medicine. Data mining and knowledge discovery handbook.* Boston: Springer; 2009. p.1111-1131.
81. James G, Witten D, Hastie T, Tibshirani R. *An introduction to statistical learning with applications in R.* New York: Springer-Verlag, 2013:303-335.
82. Quinlan JR. Induction of decision trees. *Machine learning* 1986;1(1):81-106.
83. Breiman L., Friedman J.H., Olshen R.A., Stone C.J. *Classification Regression Trees,* Belmont: Wadsworth International Group, 1984.
84. Rokach L, Oded ZM. *Data mining with decision trees: theory and applications.* Vol. 69. World scientific, 2008.
85. Kabacoff Robert. *R in action: data analysis and graphics with R.* New York: Manning Publications 2011.
86. Quinlan JR. *C4. 5: Programs for machine learning.* Burlington: Morgan Kaufmann, 1993.
87. Hssina B, Merbouha A, Ezzikouri H, Erritali M. A comparative study of decision tree ID3 and C4. 5. *International Journal of Advanced Computer Science and Applications* 2014;4(2).
88. Mayor Eric. *Learning predictive analytics with R.* Birmingham: Packt Publishing Ltd, 2015:193-213.

89. Pandya R, Pandya J. C5. 0 algorithm to improved decision tree with feature selection and reduced error pruning. *International Journal of Computer Applications* 2015;117(16):18-21.
90. Shannon CE. A mathematical theory of communication. *Bell system technical journal* 1948;27(3):379-423.
91. Tan PN, Steinbach M, Karpatne A, Kumar V. *Introduction to Data Mining*. New York: US: Pearson, 2006:145-205.
92. Cremilleux B, Robert C, A theoretical framework for decision trees in uncertain domains: Application to medical data sets, *Lecture Notes In Artificial Intelligence* vol.1211, Berlin: Springer-Verlag; 1997. p.145-156.
93. Podgorelec V, Kokol P, Stiglic B, Rozman I. Decision trees: an overview and their use in medicine. *Journal of medical systems* 2002;26(5):445-463.
94. Sanders GD, Hagerty CG, Sonnenberg FA, Hlatky MA, Owens DK. Distributed decision support using a web-based interface: prevention of sudden cardiac death. *Medical Decision Making* 2000;19(2):157-166.
95. Adam BL, Qu Y, Davis JW, Ward MD, Clements MA, Cazares LH, et al. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. *Cancer research* 2002;62(13):3609-3614.
96. Pavlopoulos SA, Stasis AC, Loukis EN. A decision tree-based method for the differential diagnosis of Aortic Stenosis from Mitral Regurgitation using heart sounds. *Biomedical engineering online* 2004;3(1):21.
97. Stukan M, Alcazar JL, Gębicki J, Epstein E, Liro M, Sufliarska A, et al. Ultrasound and clinical preoperative characteristics for discrimination between ovarian metastatic colorectal cancer and primary ovarian cancer: A case-control study. *Diagnostics* 2019;9(4):210.
98. Polat K, Yosunkaya Ş, Güneş S. Comparison of different classifier algorithms on the automated detection of obstructive sleep apnea syndrome. *Journal of Medical Systems* 2008;32(3):243-250.
99. Turan MK, Sehirli E. A novel method to identify and grade DNA damage on comet images. *Computer methods and programs in biomedicine* 2017;147:19-27.

100. Özdemir ME, Telatar Z, Eroğul O, Tunca Y. Classifying dysmorphic syndromes by using artificial neural network based hierarchical decision tree. *Australasian physical & engineering sciences in medicine* 2018;41(2):451-461.
101. Vapnik VN, Lerner AY. Recognition of patterns with help of generalized portraits. *Avtomat. i Telemekh*, 1963;24(6):774-780.
102. Boser, B., Guyon, I., Vapnik, V.: A training algorithm for optimal margin classifiers. In: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 1992 July; Pittsburgh, 1992, p.144-152.
103. Bell J. *Machine learning: hands-on for developers and technical professionals*. John Wiley & Sons, 2014:139-160.
104. James G, Witten D, Hastie T, Tibshirani R. *An introduction to statistical learning with applications in R*. New York: Springer-Verlag, 2013:337-354.
105. Olson DL, Delen D. *Advanced data mining techniques*. Berlin: Springer Science & Business Media, 2008:111-123.
106. Lantz B. *Machine learning with R*. Birmingham: Packt Publishing Ltd, 2015:239-248.
107. Awad M, Khanna Rahul. *Efficient learning machines: theories, concepts, and applications for engineers and system designers*. Apress, 2015.
108. Shmilovici A. Support vector machines. In: Maimon O, Rokach L (Eds). *Data mining and knowledge discovery handbook*. Springer, Boston, MA, 2009. p. 231-247.
109. Brown MP, Grundy WN, Lin D, Cristianini N, Sugnet CW, Furey TS, et al. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proceedings of the National Academy of Sciences* 2000;97(1):262-7.
110. Sen TZ, Kloczkowski A, Jernigan RL, Yan C, Honavar V, Ho KM, et al. Predicting binding sites of hydrolase-inhibitor complexes by combining several methods. *BMC bioinformatics* 2004;5(1):205.
111. Pugalenti G, Kumar KK, Suganthan PN, Gangal R. Identification of catalytic residues from protein structure using support vector machine with sequence and structural features. *Biochemical and biophysical research communications* 2008;367(3):630-4.
112. Niu L, Qian M, Yang W, Meng L, Xiao Y, Wong KK, et al. Surface roughness detection of arteries via texture analysis of ultrasound images for early diagnosis of atherosclerosis. *PloS one* 2013;8(10).

113. Kalatzis I, Pappas D, Piliouras N, Cavouras D. Support vector machines based analysis of brain SPECT images for determining cerebral abnormalities in asymptomatic diabetic patients. *Medical informatics and the Internet in medicine* 2003;28(3):221-230.
114. Wei LS, Gan Q, Ji T. Skin disease recognition method based on image color and texture features. *Computational and mathematical methods in medicine* 2018.
115. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making* 2019;19(1):1-16.
116. Güler NF, Koçer S. Use of support vector machines and neural network in diagnosis of neuromuscular disorders. *Journal of medical systems* 2005;29(3):271-284.
117. Yılmaz B, Asyalı M H, Arıkan E, Yetkin S, Özgen F. Sleep stage and obstructive apneic epoch classification using single-lead ECG. *Biomedical engineering online* 2010;9(1):39.
118. Korkmaz S, Zararsiz G, Goksuluk D. Drug/nondrug classification using support vector machines with various feature selection strategies. *Computer methods and programs in biomedicine* 2014;117(2):51-60.
119. Kumar N, Manhas J, Sharma V. A comparative analysis to visualize the behavior of different machine learning algorithms for normalized and un-normalized data in predicting alzheimer's disease. *Journal of Computational and Theoretical Nanoscience*, 2019;16(9):3840-8.
120. Liu Z, Li W. A method of SVM with normalization in intrusion detection. *Procedia Environmental Sciences* 2011;11:256-262.
121. Siledar SB, Chaudhary SR. Comparative analysis of naive bayes classifier and decision tree C4. 5 on credit payment data set. *IJRET* 2017.
122. Zuriani M, Yuhanis Y. A comparison of normalization techniques in predicting dengue outbreak. In: *International Conference on Business and Economics Research*. 2011.
123. Suma VR, Renjith S, Ashok S, Judy MV. Analytical study of selected classification algorithms for clinical dataset. *Indian Journal of Science and Technology* 2016;9(11):1-9.

124. Boulle M. Optimal bin number for equal frequency discretizations in supervised learning. *Intelligent Data Analysis* 2005;9(2):175-188.
125. Maslove DM, Podchiyska T, Lowe HJ. Discretization of continuous features in clinical datasets. *Journal of the American Medical Informatics Association* 2012;20(3):544-553.
126. Jiang SY, Li X, Zheng Q, Wang LX. Approximate equal frequency discretization method. In: *WRI Global Congress on Intelligent Systems, 2009. IEEE; 2009, p. 514-8.*
127. Jung S, Bi Y, Davuluri RV. Evaluation of data discretization methods to derive platform independent isoform expression signatures for multi-class tumor subtyping. *BMC genomics* 2015;16(11):1-1.



ŞEKİLLER LİSTESİ

ŞEKİLLER

Şekil 1. Altı adımlı bilgi keşfi süreci	8
Şekil 2. Veri madenciliğinde yöntem ve tekniklerine ilişkin taksonomi.....	11
Şekil 3. Karar ağacı	23
Şekil 4. A-Budama öncesi, B-Budama sonrası karar ağacı	27
Şekil 5. Farklı hiperdüzlemler tarafından noktaların ayrımı	30
Şekil 6. Maksimum marjın hiperdüzlemi	31
Şekil 7. A-Linear Kernel Fonksiyonu, B-Polinom Kernel Fonksiyonu, C-Radyal Tabanlı Kernel Fonksiyonu.	32

TABLolar

Tablo 1. Örneklem dağılımı parametreleri	37
Tablo 2. Veri dönüştürme yöntemleri	39
Tablo 3. Performans değerlendirme kriterleri	39
Tablo 4. Normal Dağılım için NB ile elde edilen ortalama doğruluk oranları (%).....	41
Tablo 5. Normal Dağılım için NB ile elde edilen ortalama AUC değerleri.....	43
Tablo 6. Normal Dağılım için NB ile elde edilen ortalama duyarlılık oranları	44
Tablo 7. Normal Dağılım için NB ile elde edilen ortalama seçicilik oranları.....	46
Tablo 8. Normal Dağılım için NB ile elde edilen ortalama pozitif kestirim değerleri.....	47
Tablo 9. Normal Dağılım için NB ile elde edilen ortalama negatif kestirim değerleri	49
Tablo 10. Ki-kare dağılımı için NB ile elde edilen ortalama doğruluk oranları (%)	51

Tablo 11. Ki-kare dağılımı için NB ile elde edilen ortalama AUC değerleri.....	52
Tablo 12. Ki-kare dağılımı için NB ile elde edilen ortalama duyarlılık oranları	54
Tablo 13. Ki-kare dağılımı için NB ile elde edilen ortalama seçicilik oranları.....	56
Tablo 14. Ki-kare dağılımı için NB ile elde edilen ortalama pozitif kestirim değerleri.....	57
Tablo 15. Ki-kare dağılımı için NB ile elde edilen ortalama negatif kestirim değerleri	59
Tablo 16. F dağılımı için NB ile elde edilen ortalama doğruluk oranları (%)	60
Tablo 17. F dağılımı için NB ile elde edilen ortalama AUC değerleri.....	62
Tablo 18. F dağılımı için NB ile elde edilen ortalama duyarlılık oranları	64
Tablo 19. F dağılımı için NB ile elde edilen ortalama seçicilik oranları.....	65
Tablo 20. F dağılımı için NB ile elde edilen ortalama pozitif kestirim değerleri.....	67
Tablo 21. F dağılımı için NB ile elde edilen ortalama negatif kestirim değerleri	69
Tablo 22. Normal Dağılım için C5.0 ile elde edilen ortalama doğruluk oranları (%).....	70
Tablo 23. Normal Dağılım için C5.0 ile elde edilen ortalama AUC değerleri.....	72
Tablo 24. Normal Dağılım için C5.0 ile elde edilen ortalama duyarlılık oranları	73
Tablo 25. Normal Dağılım için C5.0 ile elde edilen ortalama seçicilik oranları.....	75
Tablo 26. Normal Dağılım için C5.0 ile elde edilen ortalama pozitif kestirim değerleri.....	76
Tablo 27. Normal Dağılım için C5.0 ile elde edilen ortalama negatif kestirim değerleri	78
Tablo 28. Ki-kare dağılımı için C5.0 ile elde edilen ortalama doğruluk oranları (%)	80
Tablo 29. Ki-kare dağılımı için C5.0 ile elde edilen ortalama AUC değerleri.....	81
Tablo 30. Ki-kare dağılımı için C5.0 ile elde edilen ortalama duyarlılık oranları	83
Tablo 31. Ki-kare dağılımı için C5.0 ile elde edilen ortalama seçicilik oranları	84
Tablo 32. Ki-kare dağılımı için C5.0 ile elde edilen ortalama pozitif kestirim değerleri	86
Tablo 33. Ki-kare dağılımı için C5.0 ile elde edilen ortalama negatif kestirim değerleri.....	87
Tablo 34. F dağılımı için C5.0 ile elde edilen ortalama doğruluk oranları (%)	89
Tablo 35. F dağılımı için C5.0 ile elde edilen ortalama AUC değerleri.....	91
Tablo 36. F dağılımı için C5.0 algoritması ile elde edilen ortalama duyarlılık oranları	92
Tablo 37. F dağılımı için C5.0 algoritması ile elde edilen ortalama seçicilik oranları	94
Tablo 38. F dağılımı için C5.0 ile elde edilen ortalama pozitif kestirim değerleri	95
Tablo 39. F dağılımı için C5.0 ile elde edilen ortalama negatif kestirim değerleri.....	97
Tablo 40. Normal Dağılım için DVM ile elde edilen ortalama doğruluk oranları (%).....	98
Tablo 41. Normal Dağılım için DVM ile elde edilen ortalama AUC değerleri	100
Tablo 42. Normal Dağılım için DVM ile elde edilen ortalama duyarlılık oranları.....	102
Tablo 43. Normal Dağılım için DVM ile elde edilen ortalama seçicilik oranları	103

Tablo 44. Normal Dağılım için DVM ile elde edilen ortalama pozitif kestirim değerleri	105
Tablo 45. Normal Dağılım için DVM ile elde edilen ortalama negatif kestirim değerleri	107
Tablo 46. Ki-kare dağılımı için DVM ile elde edilen ortalama doğruluk oranları (%).....	108
Tablo 47. Ki-kare dağılımı için DVM ile elde edilen ortalama AUC değerleri.....	110
Tablo 48. Ki-kare dağılımı için DVM ile elde edilen ortalama duyarlılık oranları.....	112
Tablo 49. Ki-kare dağılımı için DVM ile elde edilen ortalama seçicilik oranları.....	114
Tablo 50. Ki-kare dağılımı için DVM ile elde edilen ortalama pozitif kestirim değerleri.....	116
Tablo 51. Ki-kare dağılımı için DVM ile elde edilen ortalama negatif kestirim değerleri	117
Tablo 52. F dağılımı için DVM ile elde edilen ortalama doğruluk oranları (%).....	119
Tablo 53. F dağılımı için DVM ile elde edilen ortalama AUC değerleri.....	121
Tablo 54. F dağılımı için DVM ile elde edilen ortalama duyarlılık oranları.....	123
Tablo 55. F dağılımı için DVM ile elde edilen ortalama seçicilik oranları.....	125
Tablo 56. F dağılımı için DVM ile elde edilen ortalama pozitif kestirim değerleri.....	126
Tablo 57. F dağılımı için DVM ile elde edilen ortalama negatif kestirim değerleri	128
Tablo 58. Çalışmada elde edilen en yüksek doğruluk oranları (%).....	131
Tablo 59. Çalışmada elde edilen en yüksek AUC değerleri.....	133
Tablo 60. Çalışmada elde edilen en yüksek duyarlılık oranları	135
Tablo 61. Çalışmada elde edilen en yüksek seçicilik oranları.....	137
Tablo 62. Çalışmada elde edilen en yüksek pozitif kestirim değerleri.....	139
Tablo 63. Çalışmada elde edilen en yüksek negatif kestirim değerleri	141

ÖZGEÇMİŞ

Ad Soyad: Fatma Betül ÖRS

Doğum tarihi ve yeri : 23.09.1988 - EDİRNE

İletişim adresi: f.betul.oers@gmail.com

Eğitim

Üniversite : Mimar Sinan Güzel Sanatlar Üniversitesi
İstatistik 2011-2016

Yüksek Lisans : Trakya Üniversitesi
Biyostatistik ve Tıbbi Bilişim 2017-