



T.C.

**BURDUR MEHMET AKİF ERSOY ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI**

**ÇOCUKLAR İÇİN OTİZM SPEKTRUM BOZUKLUĞUNUN VERİ
MADENCİLİĞİ YÖNTEMLERİ İLE ANALİZİ**

Sümeyye ÇELİK

YÜKSEK LİSANS TEZİ

DANIŞMAN

Dr. Öğr. Üyesi Melike ŞİŞECİ ÇEŞMELİ

BURDUR – 2020



T.C.

**BURDUR MEHMET AKİF ERSOY ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ
YÖNETİM BİLİŞİM SİSTEMLERİ ANABİLİM DALI**

**ÇOCUKLAR İÇİN OTİZM SPEKTRUM BOZUKLUĞUNUN VERİ
MADENCİLİĞİ YÖNTEMLERİ İLE ANALİZİ**

Sümeyye ÇELİK

YÜKSEK LİSANS TEZİ

DANIŞMAN

Dr. Öğr. Üyesi Melike ŞİŞECİ ÇEŞMELİ

BURDUR – 2020

T.C.
BURDUR MEHMET AKİF ERSOY ÜNİVERSİTESİ
SOSYAL BİLİMLER ENSTİTÜSÜ

ETİK BEYAN

Burdur Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğine göre hazırlamış olduğum “ÇOCUKLAR İÇİN OTİZM SPEKTRUM BOZUKLUĞUNUN VERİ MADENCİLİĞİ YÖNTEMLERİ İLE ANALİZİ” adlı tezin hazırlanması sürecinde akademik etik ilkeleri ihlal etmediğimi taahhüt eder, tezimin kağıt ve elektronik kopyalarının Burdur Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım.

Sosyal Bilimler Enstitüsü Lisansüstü Eğitim-Öğretim Yönetmeliğinin ilgili maddeleri uyarınca gereğinin yapılmasını arz ederim.

- Tezimin tamamı her yerden erişime açılabilir.
- Tezim sadece Burdur Mehmet Akif Ersoy Üniversitesi yerleşkelerinde erişime açılabilir.
- Tezimin 3 yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin tamamı her yerden erişime açılabilir.

Sümeyye ÇELİK
18.06.2020

ÖNSÖZ

Tezimin konusunun belirlenmesinde ve ilerletilmesinde her türlü desteęi saęlayan, yardımlarını esirgemeyen, pozitif yönlendirmeleriyle ve güler yüzlü davranışlarıyla beni sürekli destekleyen değerli danışmanım Sayın Dr. Öğr. Üyesi Melike ŞİŞECİ ÇEŞMELİ'ye, bugünlere gelmemde pay sahibi olan eğitimin yanan meşaleleri öğretmenlerime ve akademisyenlerime, her anımda koşulsuz şartsız yanımda olan ve asla yardımlarını esirgemeyen değerli aileme;

Teşekkürü borç bilirim.



ÖZET

Çağımızda, teknoloji alanında yaşanan gelişmelere paralel olarak bilişim sistemlerinin kullanımında artış yaşanmaktadır. Bu gelişmeler ile birlikte kullanılan sistemlerin verimliliği artmakta ve maliyetleri de azaltmaktadır. Azalan maliyetler ise daha fazla verinin depolanmasına imkân tanımaktadır. Veri madenciliği teknikleri, veriler arasında gizli kalmış olan örüntüleri ortaya çıkarmayı amaçlaması sebebiyle, depolanan verilerden yararlanmak için oldukça rağbet görmektedir. Bu kapsamda veri madenciliği, tıp gibi birçok alanda yaygın bir biçimde kullanılmaktadır.

Teşhis ve tedavisi oldukça zor ve uzun bir süreçten oluşan Otizm spektrum bozukluğu doğuştan gelen ya da yaşamın ilk yıllarında ortaya çıkan karmaşık bir nöro-gelişimsel bozukluktur. Kullanılan ölçeklendirmeler için son gelişmeler sürekli olarak takip edilmeli, güncel veriler ile çalışılmalı ve çalışmalara güncel sistemler dâhil edilmelidir. Bu çalışmada gerçek ve güncel Otizm spektrum bozukluğu verilerinin çocuklar için olan alt kümesi kullanılmıştır.

Çalışma da veri seti üzerinde eksik verilerin tamamlanması, kategorik verilerin sayısallaştırılması, normalizasyon gibi veri ön işleme aşamaları gerçekleştirilmiştir. Veriler yapay sinir ağları ve dilsel kuvvetli sinir-bulanık sınıflayıcı ile sınıflandırılmış, k-means ve x-means ile kümelenmiştir. Her bir yöntemin sonuçları değerlendirilmiş ve performanslar karşılaştırılmıştır.

Anahtar kelimeler: Tıp alanında veri madenciliği, Hastalık teşhisi, Otizm spektrum bozukluğu, Sınıflandırma, Kümeleme

ABSTRACT

In our age, the use of information systems has been increasing in parallel with the developments in the field of technology. With these developments, the efficiency of the systems used increases and the costs decrease. Decreasing costs allow more data to be stored. Data mining techniques are widely sought to exploit the stored data as they aim to reveal patterns that have been hidden between data. In this context, data mining is widely used in many areas such as medicine.

Autism spectrum disorder is a complex neuro-developmental disorder that occurs during the first years of life. For the scales used, the latest developments should be continuously followed, up-to-date data should be studied and the current systems should be included in the studies. In this study, subset of actual and current Autism spectrum disorder data for children was used.

In the study, data pre-processing stages such as completion of missing data, digitization of categorical data and normalization were performed on the data set. Data were classified by artificial neural networks and neuro-fuzzy classifier with linguistic hedges, clustered with k-means and x-means. The results of each method were evaluated and the performances were compared.

Key words: Data mining in medicine, Diagnosis of disease, Autism spectrum disorder, Classification, Clustering

İÇİNDEKİLER TABLOSU

1. GİRİŞ	1
2. VERİ MADENCİLİĞİ	6
2.1 Veri Madenciliğinin Tıp Alanında Kullanımı	7
3. OTİZM SPEKTRUM BOZUKLUĞU	10
3.1 Otizm Spektrum Bozukluğu İle İlgili Yapılan Veri Madenciliği Çalışmaları	11
3.2 Otizm Spektrum Bozukluğu Konusunda Çocuklardan Toplanan Veri Seti İle Yapılan Çalışmalar.....	13
4. YÖNTEM.....	16
4.1. Sınıflandırma Yöntemleri	16
4.1.1. Yapay Sinir Ağları.....	16
4.1.2. Dilsel Kuvvetli Sinir-Bulanık Sınıflayıcı.....	18
4.2. Kümeleme Yöntemleri	22
4.2.1. K-Means	22
4.1.2. X-Means	24
4.3. Ön İşleme Teknikleri.....	29
4.4. Performans Ölçütleri.....	31
5. UYGULAMA.....	33
5.1. Otizm Spektrum Bozukluğu Veri Seti	33
5.2. Bulgular ve Değerlendirme	36
5.2.1. Yapay Sinir Ağı İle Elde Edilen Sonuçlar	37
5.2.2. Dilsel Kuvvetli Sinir-Bulanık Sınıflayıcı İle Elde Edilen Sonuçlar	43
5.2.3. K-Means İle Elde Edilen Sonuçlar	47
5.2.4. X-Means İle Elde Edilen Sonuçlar	49
5.3. Farklı Yöntemler ile Analiz Edilen Çocuklar İçin Otizm Spektrum Bozukluğu Verilerinin Tahminleme Başarılarının Karşılaştırılması	50

6. SONUÇ ve ÖNERİLER.....54



TABLolar DİZİNİ

Tablo 1: Thabtah'ın 2018 yılındaki aynı veri seti ile elde ettiği sonuçlar	13
Tablo 2: Örnek bir hata matrisi	32
Tablo 3: Veri setindeki özniteliklerin detayları.....	34
Tablo 4: Yapay Sinir Ağının performansı	37
Tablo 5: Kullanılan yöntemlerin performans değerleri	50
Tablo 6: Kullanılan yöntemlerin eğitim veri setleri için Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri	51
Tablo 7: Kullanılan yöntemlerin test veri setleri için Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri.....	52
Tablo 8: Çalışmanın performans değerleri ve aynı veri setini kullanan diğer çalışmalardaki performans değerleri	52

ŞEKİLLER DİZİNİ

Şekil 1: Örnek bir nöron yapısı	17
Şekil 2: Örnek bir yapay sinir ağı modeli	18
Şekil 3: A dilsel terimine uygulanan dilsel kuvvetlerin değişik p değerleri için yorumları	20
Şekil 4: İki-girişli, bir-çıkışlı bir bulanık kuralın normal ve dilsel kuvvetli halinin Mamdani çıkarım sistemindeki sonuçları.....	21
Şekil 5: Dilsel kuvvetli sinir-bulanık sınıflayıcı'nın çalışma yapısı	22
Şekil 6: Üç merkez içeren bir K-means yapısı.....	26
Şekil 7: Bölünme için rastgele seçilen bir vektör görüntüsü	26
Şekil 8: K-means'in üç ayrı bölgede (k= 2 ile) çalışmasının ilk basamağı	27
Şekil 9: K-means'lerin (k= 2 için) tamamlanması ile oluşan kümeler	27
Şekil 10: Sonuçta elde edilen kümeler.....	28
Şekil 11: Başarı oranının en yüksek olarak görüldüğü YSA modeli.....	38
Şekil 12: YSA'nın eğitildikten sonra eğitim verileri ile gerçekleşen uyumu	39
Şekil 13: YSA'nın eğitim aşamasındaki MSE değerine bakılarak oluşturulan performans grafiği	40
Şekil 14: Eğitilen ağın eğitim verisi için çizdirilen ROC eğrisi.....	41
Şekil 15: Eğitim verisi için çizdirilen hata matrisi	41
Şekil 16: Eğitilen ağın test verisi için çizdirilen ROC eğrisi.....	42
Şekil 17: Eğitilen ağın test verisi için çizdirilen hata matrisi	42
Şekil 18: Özniteliklerin DKSBS tarafından belirlenen önem düzeyleri.....	44
Şekil 19: Dilsel kuvvetli sinir-bulanık sınıflayıcı'a ait bulanık kuralların modeli	45
Şekil 20: Dilsel kuvvetli sinir-bulanık sınıflayıcı'nın eğitim performansı	46
Şekil 21: Eğitim veri için çizdirilen hata matrisi.....	46
Şekil 22: Test verisi için çizdirilen hata matrisi	47

Şekil 23: Gerçek sınıflar ve K-means ile elde edilen sınıflar	48
Şekil 24: K-means algoritması ile elde edilen sonuçlar için çizdirilen hata matrisi.....	48
Şekil 25: Gerçek sınıflar ve X-means ile elde edilen sınıflar	49
Şekil 26: X-means algoritması ile elde edilen sonuçlar için çizdirilen hata matrisi	50



KISALTMALAR

ADOS	:Autism Diagnostic Observation Schedule
AUC	:Area under the ROC Curve
BIC	:Bayesian Information Criterion
DKSBS	:Dilsel Kuvvetli Sinir-Bulanık Sınıflayıcı
DSM-V	:Diagnostic and Statistical Manual of Mental Disorders
EPR	:Electronic Patient Record
FN	:False Negatif
FP	:False Pozitif
MSE	:Mean Squared Error
OSB	:Otizm Spektrum Bozukluğu
Öğr.	:Öğretim
RMSE	:Root Mean Squared Error
ROC	:Receiver Operating Characteristic
SNP	:Single Nucleotide Polymorphisms
TN	:True Negatif
TP	:True Pozitif
YSA	:Yapay Sinir Ağları

1. GİRİŞ

Teknolojilerin gelişmesi ile birlikte her geçen gün veri üretimi katlanarak artmaktadır (Seçgin, 2016). Yapılan alışverişler, ziyaret edilen siteler, kullanılan ilaçlar, hastalık tanıları kayıt altına alınan birçok veriden sadece bazılarıdır. Gelişen teknoloji, veri üretimini arttırmanın yanında artan verileri kaydetme ve kullanma şekillerini de değiştirmiştir. Artık klasik dosyalama sistemleri yerini daha profesyonel ve çok yönlü kullanıma olanak tanıyan veri tabanlarına bırakmaktadır. Böylece verilerin daha sistematik olarak saklanabilmesine imkân tanınmaktadır.

Üretilen yüksek miktardaki verilerin aralarındaki örüntüleri tespit etmek ancak bilişim teknolojilerin kullanılması ile mümkün olabilmektedir (Levent, 2016) çünkü matematiksel hesaplamaları manuel olarak yapmak işlem sürelerinin çok uzamasının yanında, hata oranlarının da artmasına sebep olmaktadır. Bu açıdan bakıldığında, çok fazla verinin toplanmış olması tek başına bir anlam ifade etmemektedir. Fayda yaratabilmek için veri yığınları içerisinde çıkarımlarda bulunmak gerekmektedir. Çıkarımların bilgisayarlar ve diğer teknolojik araçlar ile yapılması, daha kısa sürede daha güvenilir kararlar alınmasına yardımcı olmaktadır (Türkoğlu, 2016). Bu sebepten dolayı bilişim sistemlerinin her alana entegre edilmesi ve kullanılması önemli bir ihtiyaçtır. Bu ihtiyaç devamında veri madenciliği kavramını ortaya çıkarmıştır.

Veri madenciliği, bilgiyi keşfetme yolculuğudur (Hong ve Wu, 2011). Veri madenciliği ile var olan ama daha keşfedilmemiş örüntüler aranmaktadır. Yani veri madenciliği; ham halde iken sadece yığın şeklinde duran ve bir anlam ifade etmeyen veriler arasından, amaca uygun olarak anlamlı ifadeler oluşturulmasına olanak tanıyan yöntemler bütünüdür. Veri madenciliğinin en temel amacına bakıldığı zaman cevher madenciliği ile aynı olduğu görülmektedir (Diler, 2016). Her iki madencilikte de var olan yığınlar arasından değerler üretilmeye çalışılmaktadır. Doğada birçok maden vardır ama kullanılmadığı sürece hiçbir anlamı yoktur. Fayda yaratmak için madenlere ulaşarak üzerlerinde işlemler yapmak gerekmektedir. Veri madenciliğinde de durum tıpkı böyledir. Zaten var olan örüntü ve ilişkiler keşfedilene kadar hiçbir anlam ifade etmezken, işlenerek ortaya çıkarıldıklarında çok değerli hale gelmektedir (Akdemir, 2016).

Veri madenciliği sađlık, biyoloji, telekomunikasyon, finans, pazarlama, sigortacılık, astronomi, tıp, genetik, kimya, cođrafi bilgi sistemleri, yüzey analizi, uzay bilimleri, atmosfer bilimleri, sosyal bilimler, davranış bilimleri, metin madenciliđi, internet madenciliđi gibi birçok alanda kullanılmaktadır (Talan, 2016). Veri madenciliđinde kullanılan birçok yöntem vardır. sınıflandırma, kümeleme ve birliktelik analizleri bu yöntemlerden en temel olanlarıdır (Gharib vd., 2010). Bu yöntemler şimdiye kadar geliştirilmiş olmakla beraber sürekli olarak geliştirilmeye devam edilmektedir (Yıldırım, 2016). Amaca göre bu yöntemlerden biri ya da birkaçı kullanılabilir (Koçak, 2016). Eğer çözülmek istenen problem için var olan yöntemler yeterli olmaz ise, o zaman yeni yöntemler geliştirilmektedir (Koyuncugil ve Özgülbaş, 2009).

Teknolojik gelişmeler her alanda olduđu gibi tıp alanında da oldukça kullanılmaktadır (Moskov, 2016). Günümüzde gelişen bu teknoloji ile birlikte hastanelerde kullanılan cihazlar ve uygulanan tedavi yöntemleri de geliştirilmektedir. Donanımsal ve yazılımsal olarak sürekli gelişen bu yöntemler sayesinde teşhis ve tedavilerin daha profesyonel şekilde uygulanmasına olanak tanınmaktadır (Tsumoto vd., 2017). Böylece insan hayatının söz konusu olduđu önemli süreçlerde, daha kısa zamanda ve daha etkili kararlar alınabilmektedir.

Tıp alanında söz konusu insan sađlığı olduđu için Ar-Ge çalışmalarının olabildiğince fazla olması ve desteklenmesi önemlidir. Çünkü insan hayatı maddi varlıklarla ölçülemeyecek kadar değerlidir (Reddy vd., 2012). Birçok tedavide erken teşhis büyük bir öneme sahiptir. Tedaviye geç kalındığı zaman hastalık daha fazla ilerleyebilmekte ve tedavisi daha zor olabilmektedir. Hatta bazı durumlarda geri dönüşümü mümkün olmayacak kayıplara sebep olabilmektedir. Bu gibi sebeplerden dolayı teşhis ve tedavi noktasında doktora yardımcı olması için kullanılan veri madenciliđi yöntemlerine literatürde sıkça rastlanmaktadır. Yapılan birçok çalışmada umut vadeden sonuçlar elde edilmiştir (Akdemir, 2016).

Tıp alanında yapılacak olan teknolojik çalışmalar birçok farklı disiplinin bir araya gelmesi ile oluşturulmaktadır. Çünkü hastalık ve teşhis noktasında uzman olan bir ya da birkaç hekim olmadan hastadan gerekli bilgiler alınamayacak ya da muayene sonucunda elde edilen bulgular değerlendirilemeyecektir. Hastalık ile ilgili olarak elde edilen bulguların değerlendirmesini alanında uzman olmayan kişilerin yapması mümkün

değildir. Her bir hastalık ayrı uzmanlık gerektirir ve hastalığın kişide olup olmadığının tespit edilmesi için spesifik detayların uzman tarafından biliniyor olması gereklidir. Aksi takdirde teşhis yanlış konulabilmekte ve tedavi aksayabilmektedir (Budak ve Bozkurt, 2013). Bu durum hastanın daha ciddi sorunlar yaşamasına sebep olabilmektedir. Sonuç olarak teşhis konusunda bilişim teknolojileri kullanılarak yapılacak bir çalışmada, bilgilerin alanında uzman kişilerden alınması gereklidir. Kullanılan parametreler doğruluğu uzmanlar tarafından kabul görmüş hali ile toplanmış olmalıdır ve analizler daha sonra gerçekleştirilmelidir. Bu analizleri gerçekleştirmek için de bilişim sistemlerini kullanabilen, uzmanından aldığı verileri işleyerek anlamlı bilgilere dönüştürebilecek kişilere ihtiyaç duyulmaktadır. Sonuç olarak başarılı bir çalışma yürütmek için farklı disiplinlerin bir arada çalışması oldukça önemlidir (Nilashi, 2017).

Teşhis konusunda son kararı verecek olan ve sonrasında tedaviyi yürütecek olan alanında uzman olan hekimlerdir. Bilişim teknolojileri ve veri madenciliği yöntemleri ile sadece hekimlere bu kararı vermelerinde yardımcı olunmaya çalışılmaktadır. Hekimlerde insandır ve duyguları vardır. Çok iyi bir eğitime ve tecrübeye sahip olsalar bile çok yoğun çalışma ortamlarında kaldıklarında bazen gözden kaçırdıkları noktalar olabilmektedir. Bu durum verecekleri kararları da etkileyebilmektedir (Çetinkaya Bozkurt vd., 2016). İşte bu noktada hekimlere destek amaçlı olarak veri madenciliği yöntemleri literatürde sıkça kullanılmaktadır (Reddy vd., 2012). Böylece çok fazla iş yüküne sahip olan hekimlerin üzerindeki iş yükünün hafifletilmesine olanak tanınarak daha verimli çalışmasına yardımcı olunmaya çalışılmaktadır (Altıntaş, 2010).

Veri madenciliği yapılırken makine öğrenmesinden faydalanılmaktadır. Böylece var olan örüntüleri öğrenen sistemler bazı çıkarsamalarda bulunabilmektedir (Albayrak ve Koltan Yılmaz, 2009). Bu şekilde yapılan hesaplamalar ile insan hatalarından dolayı ortaya çıkacak problemlerin önüne geçilebilmektedir.

Makinalar zeki sistemlerdir ve yapılan kodlamaları unutmaları gibi bir durum söz konusu değildir. Bozulmadıkları sürece aynı bilgileri kullanarak sürekli olarak işlem yapabilmekte, yeterli donanımlar sağlandığı takdirde aralıksız çalışmaya devam etmektedirler. Bu sebeple hesaplamaya dayalı işlemlerde insanların yorgunluklarından dolayı ortaya çıkabilecek hataların önüne geçilebilecektir. Bunların yanı sıra

gerçekleştirilecek olan işlemler çok daha kısa sürede halledilecek ve zamandan tasarruf sağlanacaktır. Zamandan tasarruf sağlanması ve işlemlerin kısa sürede doğru bir şekilde sonuçlandırılması için veri madenciliği yöntemleri oldukça sık kullanılmaktadır (Sağiroğlu, 2016).

Veri madenciliği yöntemleri uygulanırken en çok sıkıntının yaşandığı aşama verilerin toplanması aşamasıdır (Akal, 2016). Güvenlik dolayısı ile çoğu bilgi paylaşılmamakta ya da sınırlı olarak paylaşılmaktadır. Bu da yapılacak olan çalışmaların aksamasına sebep olmaktadır. Özellikle tıp alanındaki verilerin kişisel bilgiler olduğu düşünüldüğünde, hastaların sağlık bilgilerini elde etmek oldukça zor ve zahmetlidir. Bu sebepten dolayı sağlık verileri ile çalışılacak ise gerekli izinlerin alınması ve ayarlamaları yapılması ya da bazı kurum veya kuruluşların çalışmalarda kullanılması için yayınlamış oldukları herkese açık verilerden yararlanılması söz konusudur (Reddy vd., 2012).

Otizm, Türkiye de ve dünyada oldukça fazla görülen bir rahatsızlıktır (Kılınç ve Söğüt, 2018). Otizm spektrum bozukluğu, doğuştan olabilen ya da doğumdan sonra bir kaç yıl içerisinde ortaya çıkabilen karmaşık bir nöro gelişimsel bozukluktur (Torun ve İşeri, 2018). Özellikle çocukluk döneminde görülmektedir (Uğur vd., 2018). Beynin yapısını ya da işleyişini etkileyen bazı sinir sistemi sorunlarından kaynaklanmaktadır (Torun ve İşeri, 2018). Otizm spektrum bozukluğu gözlemlenen hastalarda sosyal ve bilişsel olarak gelişmelerin normalden daha yavaş olduğu görülmektedir (Uğur ve Göker, 2018). Örneğin çevrelerinden gelen uyarılara karşı tepkileri yeterli düzeyde değildir (Gülsöz ve Çıkılı, 2018). Birçok hastalıkta olduğu gibi otizmde de erken tedavi ve eğitim çok önemlidir. Bunun içinde erken teşhis önem kazanmaktadır. Çünkü çocukların hayattan uzak kalmadan, topluma kazandırılmaları gerekmektedir (Akman, 2018). Literatürde otizm spektrum bozukluğu konusunda birçok çalışma yer almaktadır. Bu çalışmaya da Thabtah (2017) tarafından gerçekleştirilen bilimsel çalışma yön vermiştir. Yazar çalışmasında otizm bozukluğu ile ilgili veri madenciliği çalışmalarını değerlendirmiştir. Yapılan araştırma ile son dönemde otizm verileri ile gerçekleştirilen veri madenciliği çalışmalarında verilerin güncelliğini yitirdiği ve yeni ölçeklendirmelerle toplanan verilere ihtiyaç duyuldu vurgusu yapılmıştır (Thabtah,2017). Devamında ise Thabtah ve arkadaşları (2018) tarafından gerçekleştirilen bilimsel çalışma ile güncel olarak oluşturulan ölçekten faydalanılarak,

ihtiyaç duyulan yeni veriler toplanmış ve sonuçlar yayınlanmıştır. Tarama testi ASDTests isimli bir mobil uygulama ortamında gerçekleştirilmiştir. Uygulamanın içinde yaş gruplarına göre ayrı kategoriler mevcuttur. Her kategoride on soru bulunmaktadır ve her biri kullanıcıların uygun cevabı dikkatlice seçmelerini sağlamak için bir resimle ilişkilendirilmiştir. Katılımcılara, bilgilerinin isimsiz tutulacağı ve sadece araştırma amacıyla paylaşılacağı bildirilmiştir. Değerlendirme tamamlanmadan önce, katılımcılardan araştırmanın amacını, gizlilik politikasını ve verilerin kullanımını açıklayan bir bilgilendirme yapılmıştır. Kullanılan veri setinde girdi olarak kullanılan öznelikler arasında otizm spektrum bozukluğu'nu ölçümleyen on sorunun yanında demografik özellikleri içeren bilgilerde mevcuttur. Çıktı parametresi ise teste giren bireyin otizm spektrum bozukluğu olup olmadığını belirtmek için kullanılan evet ya da hayır anlamı taşıyan sınıf değişkenidir. (Thabtah vd., 2018).

Otizm spektrum bozukluğu için Thabtah ve arkadaşların oluşturduğu güncel verilerin çocuklar için olan alt kümesi kullanılarak yapılan çalışmalar incelendiği zaman logistic regression, naive bayes gibi yöntemlerin kullanılması ile gerçekleştirilen sınıflandırma işlemleri olmakla birlikte yapay sinir ağları ve dilsel kuvvetli sinir-bulanık sınıflayıcı ile yapılan bir sınıflandırmaya ve herhangi bir yöntem ile kümelemeye rastlanmamıştır. Bu sebepten dolayı bu çalışmada öncelikle eksik verilerin tamamlanması, normalizasyon işleminin gerçekleştirilmesi gibi veri ön işleme yöntemleri uygulanmış devamında ise bu konuda literatürde eksikliği görülen kümeleme işleminin gerçekleştirilmesi ve sınıflandırma işlemi için denenmediği fark edilen iki yöntemin veri seti üzerinde uygulanması sağlanmıştır. Literatüre katkısı olması açısından çalışmada denenilen yöntemlerin tahminliğinin doğruluğu Doğruluk, Duyarlılık, Belirleyicilik, F-Ölçütü gibi birçok parametre ile verilmiş ve yöntemlerin başarı oranları karşılaştırılmıştır. Kümeleme işlemi için k-means ve x-means yöntemleri, sınıflandırma işlemi için ise yapay sinir ağları ve dilsel kuvvetli sinir-bulanık sınıflayıcı yöntemleri kullanılmıştır. Sonuçlara bakıldığı zaman tahminleme doğruluğu bakımından otizm spektrum bozukluğu'nun çocuklar için olan verilerinde, sınıflandırma yöntemlerinin kümeleme yöntemlerinden daha başarılı olduğu görülmektedir. Özellikle dilsel kuvvetli sinir-bulanık sınıflayıcı yönteminin tüm verileri doğru sınıflandırarak literatürdeki birçok yöntemden daha yüksek başarı oranına sahip olduğu için tercih edilebilecek yöntemler arasında en iyi yöntemlerden bir tanesi olabileceği sonucuna varılmaktadır.

2. VERİ MADENCİLİĞİ

Veri madenciliğinin literatürde birçok farklı tanımının olmasının yanında, en temel anlamına bakıldığında elde edilen verilerin veri madenciliği yöntemleri ile analiz edilerek anlamlandırılmaya çalışıldığı süreçtir. Veri madenciliği adından da anlaşıldığı üzere bir madenciliktir. Madencilik işinin çeşitli yöntemler yardımı ile verilere uygulanmasını sağlamaktadır (Akdemir, 2016). Böylece veriler üzerine uygulanan madencilik ile bilgi keşifleri gerçekleştirilmektedir. Yani var olan bilgiler, toplu halde duran ve ayrıştırılmamış olan veri yığınları içerisinde bulunarak keşfedilmektedir (Hong ve Wu, 2011).

Literatürde yapılan birkaç tanım şöyledir: Veri madenciliği, depolanmış veriden istatistiksel ve matematiksel teknikler yardımı ile anlamlı ve yeni ilişkiler keşfetme sürecidir (Akpınar, 2000). Veri madenciliği, depolanmış yüksek miktardaki veri içerisinde anlamlı kuralların çıkarılması işidir (Baykal, 2006). Veri madenciliği, önceden bilinmeyen anlamlı bilgilere ulaşılması ve bu bilgilerin kullanılması yöntemidir (Ayık vd., 2007). Çeşitli yöntemlerle toplanan veriler üzerine bir takım analizlerin yapılması ve çıkan sonuçların değerlendirilmesi veri madenciliği olarak adlandırılmaktadır (Ferreira vd., 2012). Veri Madenciliği; bilişim sistemleri, veri tabanları gibi teknolojik araçların kullanılması ile sistemli bir şekilde saklanan veriler üzerinde matematiksel algoritmalar ve istatistik tekniklerinin uygulanması sonucunda analizlerin gerçekleştirildiği ve verilerden anlamların çıkartıldığı bir süreçtir. Böylece süreçlerin daha doğru ve daha hızlı yönetilmesine imkan tanınmaktadır (Gharib vd., 2010).

Veri ham hali ile değersizdir ancak, bilişim sistemleriyle işlenerek bilgiye dönüştürülür ise anlam kazanmaktadır. Bu yüzden böyle bir bilgi keşfi sürecine ihtiyaç vardır. Veri madenciliği gerçekleştirilir iken belirli bir amaca yönelik gerçekleştirilmesi gereklidir. Eldeki verinin analiz edilerek yorumlanmasının yanı sıra ileriye dönük tahmin sistemlerinde de veri madenciliği yöntemlerine ihtiyaç duyulmaktadır. Teknolojik gelişmelerin bu kadar hızlı ilerlemesi ve bilgiye her yerden ulaşımın bu kadar kolaylaşması göz önünde bulundurulduğunda, veri madenciliği yöntemlerinin de hayatın içinde bulunması oldukça doğal karşılanmaktadır. Bu bağlamda gün geçtikçe analiz

edilmesi gereken verilerin boyutu büyümekte ve veri madenciliği yöntemlerinin önemi artmaktadır (Akal, 2016).

Veri madenciliği, verinin olduğu her alanda kullanılabilme imkânına sahip olduğu için kullanım alanı açısından çok büyük bir yelpazeye sahiptir. Özellikle verinin bilgiye dönüştürülmesinin öneminin artması ile birlikte, veri madenciliği yöntemlerinin kullanım alanı gittikçe artmaktadır. Pazarlama, Perakende ve Lojistik, Tıp, Devlet ve Savunma, Turizm, Eğlence, Üretim, Telekomünikasyon, Banka ve Sigortacılık, İnternet, Astronomi, Eğitim bu alanlardan bazılarıdır.

2.1 Veri Madenciliğinin Tıp Alanında Kullanımı

Tıp alanında hastalık teşhisi yapılırken dikkate alınması gereken birçok kıstas vardır. Hastalığın ciddi olabileceği ve alınacak yanlış bir karar ile çok daha büyük problemlere yol açılabileceği göz önünde bulundurulduğunda teşhis konulurken hızlı ve doğru karar verilmesi oldukça önemlidir. Bu noktada veri madenciliği yöntemlerinden faydalanmak gereklidir (Reddy vd., 2012).

Tıp alanında, veri madenciliği hastalıklı dokuların tespit edilmesi ve tanı konulmasında ya da tedavi edilmesinde destek olacak sistemlerin tasarlanmasında kullanılabilir (Brossette vd., 1998).

Bu konuda literatür de birçok araştırma yapılmıştır. Örneğin elektronik hasta kayıtları (EPR), hastaların tüm süreçlerinin kaydedildiği metinlerdir. EPR' ler madencilik yapılarak karar verme sürecine yardımcı olabilecek çok önemli kaynaklardır. EPR' lerde hastaneye gelen her hastanın şikâyetleri, fiziksel bulguları, laboratuvar sonuçları, radyografik çalışmaların bir kaydı, ilaçlarında olan değişikliklerin bir listesi ve takip bakımı için öneriler şeklinde bilgiler tutulmaktadır. Tsumoto'nun çalışmasında, hastanın kabulüne yönelik kompakt açıklamaları içeren bu metinler üzerine odaklanılmış ve bu bilgiler ışığında metin madenciliği yapılmıştır. Sonuçları değerlendirilmiş ve performanslar karşılaştırılmıştır (Tsumoto vd., 2017).

Gen araştırmaları sürekli gelişmektedir. Böylece üretilen bilgede her geçen gün artmaktadır. Önemli olan bu bilgileri analiz etmek ve kullanmaktır. Her hastalık için farklı çalışmalar yapılabilmektedir. Demircioğlu ve Bilge'nin çalışmalarında kanser hastalığı ile ilgili bir uygulama yapılmıştır. Temelde sınıflandırma yapılmakla birlikte

sınıflandırma doğruluğunu arttırmak için öznitelik seçimi gerçekleştirilmiştir (Demircioğlu ve Bilge 2015).

Estibüler aparatı insan işlevlerini yerine getirmeye yardımcı bir alettir ve yapılan testler sonucunda birçok veri üretilmektedir. Mirarchi ve arkadaşlarının çalışmasında üretilen bu veriler analiz edilmiştir. Gerekli veriler Üniversite Hastanesi veri tabanından yüklenen elektronik hasta kayıtları üzerinden elde edilmiştir. Yapılan araştırma sonucunda hastalığın ilişkili olduğu durumlar ortaya konulmuştur (Mirarchi vd., 2016).

Coulter ve arkadaşlarının çalışmasında kalp kasının bozukluğu konusunda bir çalışma yapılmıştır. Kalp kasının durumu konusunda veri madenciliği yöntemleri kullanılarak analizler gerçekleştirilmiştir. Yapılan çalışma sonucunda bazı belirtiler ile bazı ilaçlar arasında ilişki olduğunu ortaya konulmuştur (Coulter vd., 2007).

Göreke ve arkadaşlarının çalışmasında meme kanseri ile ilgili bir araştırma gerçekleştirilmiştir. Çalışmada bilişim sistemlerinin gücünden faydalanılmıştır. Böylece teknolojiyi kullanarak hastalığı teşhis etmeye yardımcı olmak hedeflenmiştir. Hastalığın doğru teşhisi için öncelikle öznitelikler belirlenmiştir ve yüksek başarı oranı elde edilmiştir (Göreke vd., 2014),

Diyabet dünya çapında oldukça fazla görülen bir hastalıktır. Bu sebepten dolayı İleri tıbbi ekipman olmadan, elde edilen tıbbi bilgileri kullanarak diyabet hastalığına tanı koyabilmek önemlidir. Böylece hastalığı veya hastalık riskini erken aşamada keşfetmek isteyen insanlara yardımcı olmak mümkündür. Bu öneminden dolayı Nilashi ve arkadaşlarının çalışmasında makine öğrenme tekniklerini kullanarak bir sistem geliştirilmiştir ve diyabet hastalığını sınıflandırılmıştır. Yöntem olarak kümeleme ve sınıflandırma yaklaşımları kullanılmıştır. Pima Hint Diyabet veri setindeki deneysel sonuçlar, önerilen yöntemin, doğruluğunu önemli ölçüde göstermektedir. Sonuç olarak hibrit akıllı sistemlerin, sağlık hizmeti verenlere bir karar destek sistemi olarak yardımcı olabileceği vurgulanmıştır (Nilashi vd., 2017).

Sensörler aracılığıyla günlük yaşam aktivitelerini tanımak mümkündür. Bu sensörlerden elde edilen verilerin, çeşitli makine öğrenimi veya bilgi odaklı teknikler kullanılarak sınıflandırılması ve etkinliklerin tanınması mümkündür. Böylece rutin olarak yapılan eylemler oluşturulabilmekte ve müdahale edilmesi gereken koşulların öngörülmesine imkân tanınmış olabilmektedir. Ihianle ve arkadaşlarının çalışmasında sensörlerden

alınan veriler veri madenciliği yöntemleri ile işlenmiş ve sonuçlar paylaşılmıştır (Ihianle vd., 2016).

Tıbbi bilginin de yardımıyla, veri madenciliği teknikleri iyi uygulandığında, hastanın aynı sorunu yaşayıp yaşamadığının tahmin edilebilmesi mümkündür. Peixoto ve arkadaşlarının çalışmasında sağlık çalışanlarına müdahaleleri öngörmede yardımcı olmak hedeflenmiştir. Bu amaçla yapılan çalışmada, modeller geliştirilerek kalp hastalığı veri sınıfları oluşturulmuştur (Peixoto vd., 2017).

Brossette ve arkadaşları (1998) tarafından yapılan bir çalışmada, hastanelerde çıkabilecek salgınların önceden tespit edilebilmesi için veri madenciliği yöntemlerini kullanan bir sistem önerilmiştir. Hastanenin önceki dönem kayıtlarında bulunan P. Aeruginosa hastalığının verileri üzerinde birliktelik kuralları ile yapılan çalışmadan elde edilen bilgilerle sistemin faydalı olabileceği yönleri ortaya konulmuştur (Brosette vd., 1998).

3. OTİZM SPEKTRUM BOZUKLUĞU

Otizm spektrum bozukluğu (OSB) dünya çapında oldukça hızlı büyüyen gelişimsel sakatlık tanısından biridir (Thabtah, 2017). OSB, doğuştan gelen ya da yaşamın ilk yıllarında ortaya çıkan karmaşık bir nöro-gelişimsel bozukluktur (Thabtah, 2018). Sosyal ilişkilerde, iletişim kurmada, öğrenmede ve konuşma yeteneklerinde sıkıntılar yaşanmasına sebep olmaktadır (De Campos Souza ve Guimaraes, 2018). OSB’de ilgi duyulan şeyler sınırlıdır ve bunun yanında kendini yineleyen davranışlarda ve hareketlerde bulunmaktadır (Akyol ve Karaci, 2018).

Genel olarak kriterler şöyledir:

1)Toplumsal İletişim ve Etkileşimde Güçlükler

- İlişki kurma ve sürdürmede zorlanma
- Göz kontağı kuramama
- Duyguları ifade edememe
- Etkileşim başlatma ve sürdürmede zorlanma

2) Sınırlı-Yineleyici Davranış Örüntüler (Tekrarlayıcı Davranışlar)

- Basmakalıp ve tekrarlayıcı motor hareketler
- Aynılıkta ısrar, rutine sıkı bağlılık
- Sınırlı ve yoğun ilgi alanı
- Duyusal az veya çok uyarılma

(Özeren, 2013).

OSB’nin teşhis ve tedavisi oldukça uzundur. Aynı zamanda teşhisin erken konulması ve tedaviye başlanması çok önemlidir (Sabuncuoğlu vd., 2015). Bu hastalığın erken fark edilmesi ileride yaşanacak sağlık sorunlarını önemli ölçüde azaltabilmekte ve çocuğun genel zihinsel sağlığını iyileştirici yöntemlerin uygulanmasına imkân tanımaktadır (van den Bekerom, 2017). Teşhis için en güncel olarak Amerikan Psikiyatri Birliği'nin yayımladığı kılavuz (DSM-V) kullanılır (Olçay Gül ve Tekin İftar, 2012).

Otizmin önemli bir genetik bileşeni olmasına rağmen, öncelikle davranışsal özelliklerle teşhis edilmektedir (Demirhan, 2018). Aileden bilgiler alınmakta, dikkati ölçen testler

uygulanmakta ve böylece davranışların uzmanlar tarafından yorumlanması sağlanmaktadır (Özkaya, 2013). Tanı koyabilecek kişiler, yalnızca konunun uzmanı olan hekimlerdir. Otizmlili çocukların dış görünüşleri diğer çocuklarla aynıdır ama davranışları farklılık göstermektedir. Tanı konulurken, çocuğun uzmanlar tarafından gözlenmesi, anne ve babalara çocuğun gelişimi ile ilgili sorular sorulması ve gelişim ile ilgili testlerinin yapılması gerekmektedir. Otizmin tanısı 1 yaşından itibaren konabilmektedir. Eğitimin olabildiğince erken başlaması çocuğun gelişimini olumlu yönde etkileyeceği için otizm tanısının erken konulması oldukça (Sabuncuoğlu vd., 2015).

3.1 Otizm Spektrum Bozukluğu İle İlgili Yapılan Veri Madenciliği Çalışmaları

OSB genellikle kişiliği ve iletişimi etkileyen bir rahatsızlıktır. Literatür incelendiği zaman OSB konusunda veri madenciliği yöntemlerinin de entegre edildiği bir çok çalışma olduğu gözlemlenmektedir. Örneğin Jiao ve arkadaşlarının çalışmasında hastalık teşhisi için kullanılan yöntemler karşılaştırılmıştır. Bu yöntemler bölgesel kalınlık ve bölgesel hacimdir. Çalışma 22 hasta ile gerçekleştirilmiş ve bulgular destek vektör makineleri, çok katmanlı algılayıcılar, karar ağaçları gibi makine öğrenme teknikleri ile analiz edilmiştir. Performanslar analiz edilip karşılaştırıldıktan sonra kalınlık tabanlı modelin hacme dayalı modele göre daha doğru sonuçlar verdiği ortaya konulmuştur (Jiao vd., 2010).

Jiao ve arkadaşlarının çalışmasında öncelikle Otizm spektrum bozukluğunun nöro gelişimsel bir bozukluk olduğu vurgulanmıştır. Amaç olarak beyindeki bölgesel kalınlık ve hacimlerin ölçümüne dayalı bir ayırt edicilik olabilir mi sorusunun cevabı aranmış ve bir model oluşturulmuştur. Oluşturulan bu model tek nükleotid polimorfizmleri (SNP) modeli ile karşılaştırılmıştır. Deney grubu olarak 66 beyin yapısı incelenerek beyindeki bölgesel kalınlık ve hacimler ölçülmüştür. Model oluşturulurken karar ağaçları, lojistik regresyon, destek vektör makineleri gibi makine öğrenme teknikleri kullanılmıştır. Sonucunda ise beyindeki bölgesel olarak kalınlık ve hacme dayalı yöntemin istenilen sonucu veremediği paylaşılmıştır. SNP tabanlı modelin daha doğru sonuçlar verdiği dile getirilmiştir.

Moradi ve arkadaşları (2017) temelde verilerin yorumlanması noktasında karşılaşılan zorluklarla başa çıkmak için makine öğrenimini önermişlerdir. Bu düşüncüyü doğrulamak için OSB'nin şiddetini tahmin etmek için makine öğrenimi yöntemlerini

kullanarak bir uygulama gerçekleştirmişlerdir. Uygulamada OSB olan 156 bireyden kortikal kalınlık ölçümlerine göre OSB'nin şiddetini tahmin etmeye çalışmışlardır ve gelecekteki çalışmalar için umut vadeden sonuçlara ulaştıklarını vurgulamışlardır.

Latkowski ve Osowski (2015) Yaptıkları çalışmalarında gen mikrodizileri kullanarak otizm olgusunu tanımak için bir araç sağlamayı amaçlamışlardır. Bunun için asıl önemli olan, otizmle sıkı sıkıya bağlı olan en önemli genleri keşfetmektir. Bu çalışmada da bir sınıflandırıcı grubu için en temsili girdi niteliklerini seçmek için farklı gen seçimi yöntemleri uygulanmıştır. Genetik algoritma ve destek vektör makinaları sınıflandırıcısı ile birleştirilmiş olan seçim sonuçlarına bakılmış ve otizm tanıma oranının artırıldığı görülmüştür. Sonuç olarak otizmin erken tanısı, çocukların tedavisi için son derece önemlidir. Erken tanı iyileşme ve normal sosyal iletişime geri dönüş olasılığını arttırdığı için erken tanı oldukça önemlidir.

Lu ve arkadaşları (2017) tarafından yaptıkları çalışmada otizm verileri kullanılmıştır. Öznitelik seçimi için birkaç yöntem uygulanmış ve sonuçları nihai sonuca entegre edilmiştir. Sonrasında oluşan gen kümesi sınıflandırmada kullanılarak otizm tanısına yardımcı olunmaya çalışılmıştır.

Wall ve arkadaşları (2012), Otizm Teşhis Gözlem Programı (ADOS) OSB'nin davranışsal değerlendirilmesi için en yaygın kullanılan araçlardan biri olduğunu belirtmişlerdir. Her biri kendi dil ve gelişim düzeyine göre belirli bir grup birey için hazırlanmış dört modülden oluşmaktadır. Ortalama olarak, bir modülün oluşturulması 30 ila 60 dakika sürer. Bu da teşhis için zamanın çok uzamasına yol açar. Bu yüzden Otizm sınıflandırması yapılırken veri madenciliği yöntemlerini kullanmak sürecin hızlanmasında fayda sağlamaktadır. Bu çalışmada da 612 birey için ADOS 1. Modülü 'nün tüm skorlarını incelemek için bir dizi makine öğrenme algoritması kullanılmıştır. Analizler sonucunda, ADOS'un 1. Modülünde yer alan 29 maddeden 8'inin, otizmi % 100 doğrulukla sınıflandırmaya yeterli olduğu görülmüştür.

Demirhan'ın çalışmasında, Thabtah'ın çalışması ile oluşturulan güncel veri setinin ergenler için olan alt kümesi kullanarak analizler gerçekleştirilmiştir. Yöntem olarak destek vektör makineleri, rastgele orman ve k-en yakın komşu algoritmalarını kullanmıştır. Bu üç yöntemde de 10-kat çapraz doğrulama kullanılarak yöntemlerin performansları değerlendirilmiştir. Sınıflandırmaların doğruluk oranlarına bakıldığı

zaman destek vektör makineleri yönteminin %95, k-en yakın komşu yönteminin %89 ve rastgele orman yönteminin %100 oranında başarıya eriştiği görülmüştür. Bu çalışmada OSB ergen tarama verilerini kullanarak yapılan sınıflandırmada en başarılı olduğu görülen yöntem rastgele orman yöntemi olmuştur. (Demirhan, 2018).

3.2 Otizm Spektrum Bozukluğu Konusunda Çocuklardan Toplanan Veri Seti İle Yapılan Çalışmalar

Literatür incelendiği zaman OSB konusunda veri madenciliği veya yapay zekâ tekniklerini içeren, OSB veri setinin farklı altkümelerini kullanan çok fazla çalışma olduğu görülmektedir. Bu sebepten dolayı literatürü sınırlandırabilmek için OSB veri setinin çocuklar için olan alt kümesini kullanan çalışmalara yer verilmiştir. Bu çalışmaların temelini topladığı verileri paylaşarak Thabtah (2018) oluşturmuştur.

Thabtah (2018) tarafından, detaylı olarak literatür değerlendirilmiş, daha önceki çalışmalara yer verilmiştir. Verilerin nasıl toplandığı anlatılmış ve uygulama aracı olarak kullanılan ASDTests tanıtılmıştır. Veriler arasındaki ilişki incelenerek grafikler oluşturulmuştur. ASD için oluşturulan veri seti için üç alt küme olarak belirlenen çocuk, ergen, yetişkin grupları için Logistic Regression ve Naïve Bayes algoritmaları ile analizler gerçekleştirilmiştir. Elde edilen sonuçlar Tablo 1’de gösterilmiştir.

Tablo 1: Thabtah’ın 2018 yılındaki aynı veri seti ile elde ettiği sonuçlar

Veri setinin alt kümeleri ve kullanılan yöntemler	Doğruluk
Ergen	
Logistic Regression	94.23
Naïve Bayes	91.34
Yetişkin	
Logistic Regression	99.85
Naïve Bayes	95.73
Çocuk	
Logistic Regression	97.94
Naïve Bayes	92.80

Tablo 1’de ifade edilen şekli ile çalışmada kullanılan yöntemlerin başarı oranına bakıldığı zaman veri setinin üç alt kümesi için de Logistic Regression yönteminin sınıflandırma başarısının daha iyi olduğu görülmüştür. En yüksek başarı oranı ise veri

setinin yetişkin alt kümesinin Logistic Regression ile sınıflandırıldığı durumda elde edilmiştir (Thabtah, 2018).

Akyol ve Karaci (2018), Thabtah'ın çalışması ile oluşturulan güncel veri setinin çocuklar için olan alt kümesini kullanarak öznitelik seçimi gerçekleştirilmiş ve ardından sınıflandırılma yapılmıştır. Çalışmada öncelikle ön işleme aşamaları gerçekleştirilmiş ve boş değerler içeren örnekler veri setinden atılmıştır. Böylece toplamda 292 örnekten oluşan veri seti 247 örneğe indirilmiştir. Veri setindeki kategorik veri türüne sahip öznitelikler sayısal değerlere çevrilmiştir. Bu ön işleme aşamasından sonra sınıflandırma gerçekleştirilmiştir. Sınıflandırma için Lojistik Regresyon ve Bulanık Kurallı Lojistik Regresyon kombinasyonu kullanılmıştır. Kullanılan yöntemlerin başarı oranına bakıldığı zaman Lojistik Regresyon %92.00, Bulanık Kurallı Lojistik Regresyon kombinasyonu %97.33 oranında doğruluğa sahiptir. Sonuç olarak bu çalışma için lojistik regresyonun bulanık kurallar ile birlikte çalıştırıldığında daha doğru bir tahminleme yaptığı gösterilmektedir (Akyol ve Karaci, 2018).

De Campos Souza ve Guimaraes (2018), Thabtah'ın çalışması ile oluşturulan güncel veri setinin çocuklar için olan alt kümesini kullanarak bir sınıflandırma gerçekleştirmiştir. Sınıflandırma işlemi gerçekleştirilmeden önce veriler ön işleme aşamasından geçirilmiş ve kategorik olan veri türleri sayısal ifadelerle dönüştürülmüştür. Yöntem olarak Bulanık Sinir Ağı Mimarisi kullanılmıştır ve bulanık kümelerin sayısının performansa etkisi değerlendirilmiştir. Modelin başarı oranını test etmek için veriler %70 eğitim %30 test verisi olarak ayrılmıştır. Sonuçlara bakıldığı zaman en düşük performansın bulanık küme sayısının 3 olarak kullanıldığı test verisinde %59 olarak, en yüksek performansın da bulanık küme sayısının 9 olarak kullanıldığı test ve eğitim setinde %100 olarak görülmüştür (De Campos Souza ve Guimaraes, 2018).

Hutchinson ve arkadaşları (2019), Thabtah'ın çalışması ile oluşturulan güncel veri setinin çocuklar için olan alt kümesini kullanarak sınıflandırma yapmışlardır. Sınıflandırmayı gerçekleştirmek için yöntem olarak Weka hazır paket programı kullanarak J48 Decision Tree ve Naive Bayes algoritmasını kullanmışlardır. İki yöntem ile elde edilen sonuçlar accuracy, sensitivity ve specificity değerleri olarak verilmiş ayrıca iki yöntemin başarıları karşılaştırılmıştır. Çalışmanın sonuçlarına bakıldığı

zaman; J48 Decision Tree için accuracy değerinin %92.8, sensitivity değerinin 0.901 ve specificity değerinin 0.934 olduğu Naive Bayes için ise accuracy değerinin %96.2, sensitivity değerinin 0.972 ve specificity değerinin 0.954 olduğu görülmektedir. Bu sonuçlar kapsamında kullanılan veri setini Naive Bayes algoritmasının J48 Decision Tree algoritmasından daha doğru sınıflandırdığı görülmektedir (Hutchinson vd, 2019).



4. YÖNTEM

4.1. Sınıflandırma Yöntemleri

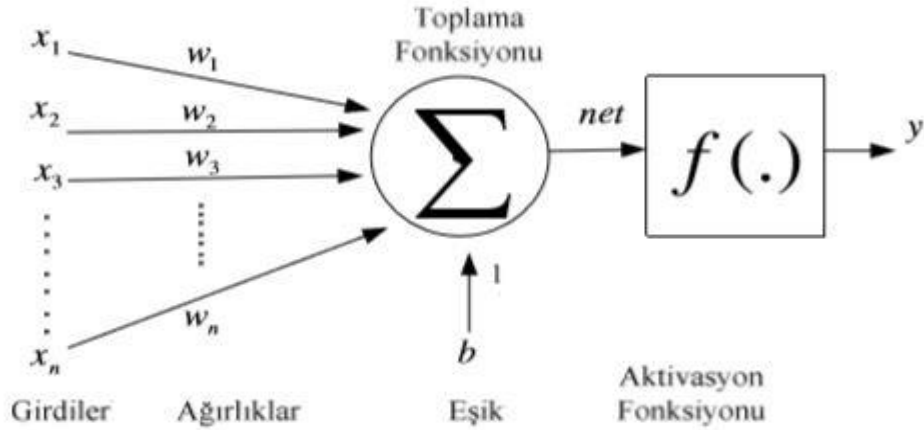
4.1.1. Yapay Sinir Ağları

Günümüzde teknolojik gelişmelerle birlikte bilgisayarlı sistemler sıklıkla kullanılmaktadır (Chine vd., 2016). Manuel olarak çok uzun zaman alacak hesaplamalar zeki araçlarla hızlı ve doğru bir şekilde yapılmaktadır (Carleo ve Troyer, 2017). Bu gibi gelişmeler, bilim insanlarına yeni fikirler vermiştir (Peurifoy vd., 2018). Bu fikirlerden bir tanesi de zeki sistemlere karar verdirme ve insan gibi davranma yetisini kazandırabilmek olmuştur. Bu düşünce, insanın olayları algılama ve karar verme sürecini inceleyerek başlamıştır (Küçükkocaoğlu vd. 2005). Böylece insan beyninin çalışma şekli taklit edilerek yapay bir karar verme mekanizması oluşturulmaya çalışılmıştır (Fırat ve Güngör, 2004).

Yapay sinir ağları (YSA), insan beyninin çalışma şekli ilham alınarak oluşturulmuş bir algoritmadır (Deb vd., 2018). Temelinde nörofizyologların ve psikologların sinir ağlarının nasıl çalıştığı konusunda ortaya koyduğu biyolojik bulgular kullanılmaktadır (Küçükkocaoğlu vd., 2005). Bu biyolojik bulgular yapısal ve işlevsel olarak sistemleştirilerek matematiksel bir model oluşturulmaya çalışılmıştır. Oluşturulan bu modele sinir ağları modeli denilmiştir (Akkaya vd., 2009). Diğer bir ifade ile YSA insan beyninin çalışma sistemini taklit ederek (Mehmet ve Akpınar, 2018), bilgisayarlı sistemlere karar verdirme becerisini kazandırmayı amaçlamaktadır (Kalantar vd., 2018). YSA, İnsan beyninin ezberleme, öğrenmeye, anlam çıkarma, bilgiler arasındaki ilişkileri ortaya koyma gibi birçok işlemini yapabilme amacıyla tasarlanmıştır. (Özsoy ve Fırat, 2004) (Akyılmaz ve Ayan, 2006). YSA ortadaki durumdan bir öğrenme gerçekleştirebildiği için yeni durumlara adapte olabilme yeteneğine sahiptir.

YSA'nın çalışma şekli, temelde nöron içerisinde gerçekleşmektedir. Örnek bir nöron yapısı Şekil 1'de gösterilmiştir.

Şekil 1: Örnek bir nöron yapısı

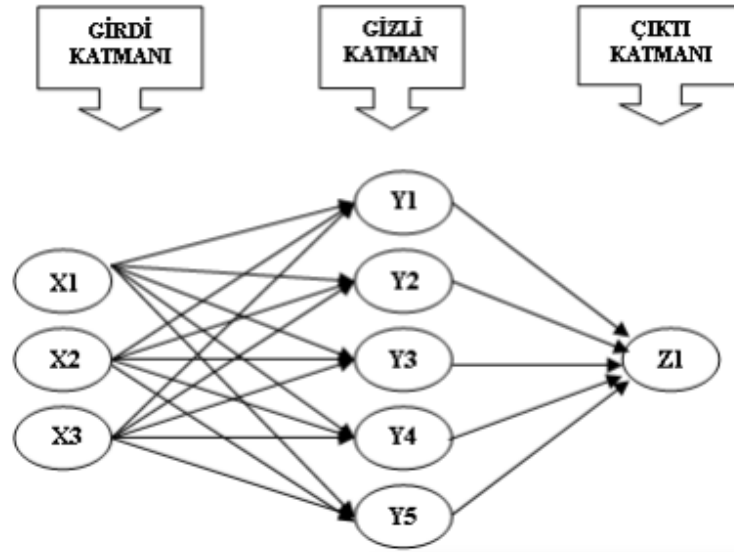


(Kaynar ve Taştan, 2009).

YSA'nın çalışma şekli, temelde nöron içerisinde gerçekleşmektedir. Yapay sinir hücresine işlem elemanı ya da yapay sinir hücresi de denilmektedir (Küçükkocaoğlu vd. 2007). Yapay nöron 5 ana bileşenden oluşmaktadır. Bu bileşenler girdiler, ağırlıklar, toplama (birleştirme) fonksiyonu, aktivasyon fonksiyonu ve çıkış şeklindedir (Akkaya vd., 2009). Bir nörona, nöron dışından gelen her bilgi girdidir. Ağırlık hücre dışından gelen girdinin etki düzeyini ifade etmektedir. Toplam fonksiyonu, her bir girdinin ağırlık değeri ile çarpıldıktan sonra toplanması ile oluşturulmaktadır. Son olarak toplam fonksiyonundan elde edilen değer bir transfer fonksiyonundan geçirilmesiyle çıktı oluşturulmaktadır (Çam vd., 2018).

Kullanılan faktörlere göre bir veya birden fazla girdi kullanılabilir. Çıktı ise istenilen sonuç kadardır (Villarrubia vd., 2018). Nöronların girdi ve çıktılarının birbiri ile ilişkilendirilmesi ile YSA oluşturulmaktadır. YSA içerisinde aynı doğrultuda bir araya gelen nöronlara katman denilmektedir (Calp, 2019). Bir YSA da temelde 3 tane katman bulunmaktadır (Cai ve Liu, 2018). Bu katmanlar girdi katmanı, gizli katman ve çıktı katmanı şeklindedir (Çam vd., 2018). Örnek bir YSA modeli Şekil 2'de gösterilmiştir.

Şekil 2: Örnek bir yapay sinir ağı modeli



(Akkaya vd., 2009).

Girdi katmanı, ilk katmandır ve dışarıdan gelen girdilerden meydana gelmektedir. Çıktı katmanı son katmandır ve elde edilen son çıktılar bu katmanda tutulmaktadır (Fırat ve Güngör, 2004). Girdi ve çıktının dışında kalan tüm katmanlara gizli katman denilmektedir. Gizli katmanlar, girdi katmanından alınan veriler üzerinde bir takım hesaplamalar yapılarak sonuçların çıktı katmanına iletiği ara katmanlardır (Aktaş, 2003). Temel hesaplamalar gizli katmanlardaki nöronlar aracılığı ile gerçekleştirildiği için gizli katman tasarımı oldukça önemlidir. Bu yüzden ağın performansını arttırmak için gizli katmanda ki nöron sayılarının doğru seçilmesi, oldukça önemlidir (Akkaya vd., 2009).

4.1.2. Dilsel Kuvvetli Sinir-Bulanık Sınıflayıcı

Dilsel kuvvetli sinir-bulanık sınıflayıcı (DKSBS), verileri sınıflandırmak için kullanılan yöntemlerden bir tanesidir. En temel özelliği sınıflandırma yapmadan önce özniteliklerin önem derecelerinin belirlenmesidir (Pençe vd, 2013). Özniteliklerin önem dereceleri belirlenirken bulanık çıkarsama yapılmaktadır. Böylece sınıflandırma sırasında, önem derecesi yüksek olan öznitelikler seçilirken, önem derecesi düşük olan öznitelikler devre dışı bırakılmış olmaktadır (Çetişli, 2006). Bulanık çıkarsama ile özniteliklerin önem dereceleri belirlendikten sonra sınıflandırma için YSA kullanılmaktadır. YSA eğitilirken, bulanık çıkarsama ile seçilen öznitelikler

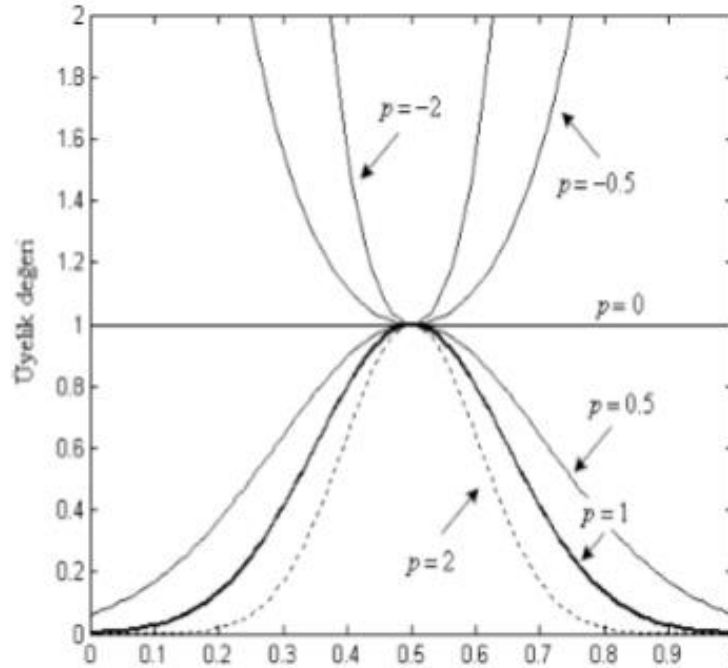
kullanıldığı için başarı oranı oldukça yüksektir (Çelik vd., 2018). Aynı zamanda kümeleri birbirinden ayırtırmayan öznitelikler kullanılmadığı için bellekten de tasarruf sağlanmaktadır. Bu durum hızın artmasının yanında, maliyetlerin de azalmasına yardımcı olmaktadır (Çetişli, 2006).

Sınıflandırma yapılırken bazı öznitelikler çıkarıldığında sınıflandırma başarısı düşmüyor ise o özniteliklerin sınıflandırmada kullanılmasına gerek yoktur (Huang ve Chow, 2005). Çünkü gerçek hayat problemlerinde de çözüme gidilirken oluşturulan alternatiflerin etki düzeyi farklı olabilmektedir. Önemli olan problemi çözüme götürecek olan daha hızlı ve daha doğru alternatifleri bulmaktır (Liu vd., 2005). Bu açıdan bakıldığı zaman dilsel kuvvetli sinir-bulanık sınıflayıcıda da gerçek hayat problemlerinin çözümüne yakın bir çözüm yolu izlenmektedir. Veri madenciliği açısından izlenen yol ise iki aşamada gerçekleştirilmektedir. Öncelikle ön işleme aşaması olarak sayılan özniteliklerin önem derecesinin belirlenmesi işlemi gerçekleştirilmektedir. Daha sonra belirlenen bu önem düzeylerine göre yapay sinir ağı eğitilerek sınıflandırma işlemi gerçekleştirilmektedir (Çelik vd., 2018). Böylece dilsel kuvvetli sinir-bulanık sınıflayıcı'yı diğer sınıflandırma yöntemlerinden ayıran en önemli özelliği bulanık çıkarsama ile öznitelik tespiti gerçekleştirmesidir (Fırat, 2008).

Dilsel kuvvetli sinir-bulanık sınıflayıcı, özniteliklerin önem düzeyini hesaplarken bulanık kuralları kullanmaktadır. Öznitelikler, bulanık kurallar ile aşamalı olarak belirlenmektedir (Çağlar vd., 2010). Bulanıklık, sistemi evet hayır gibi kesin cevapların kısıtlarından kurtarmaktadır (Koçyiğit ve Korürek, 2010). Yani bir özneliğin sonuç üzerinde etkisinin olup olmadığı var ya da yok şeklinde kesin olarak nitelendirilmek yerine, "az var", "var" yada "çok var" gibi aşamalı olarak nitelendirilmektedir. Böylece bulanık çıkarsamaların insan çıkarsamalarına benzerliği arttırılmış olmaktadır (Cetişli ve Kalkan, 2011).

A dilsel terimine uygulanan dilsel kuvvetlerin değişik p değerleri için yorumları Şekil 3'de gösterilmiştir.

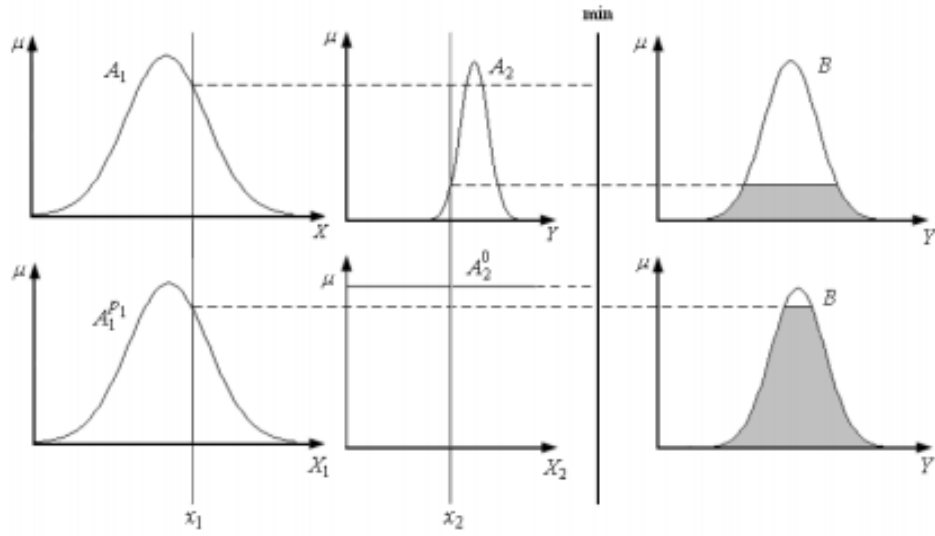
Şekil 3: A dilsel terimine uygulanan dilsel kuvvetlerin değişik p değerleri için yorumları



A dilsel terimine uygulanan, dilsel kuvvetlerin farklı p değerleri (- 2, -0.5, 0, 0.5, 1, 2) için kuvvetlendirilmiş A dilsel terimleri Şekil 3'de verilmiştir. Şekil 3'de görüldüğü gibi, p 'nin sıfırdan küçük olması durumunda ($p < 0$) üyelik fonksiyonunun yönü değişmektedir. Bunun sonucu olarak üyelik değeri 1'den büyük olmaktadır. Hesaplanan üyelik değerinin $[0 + 1]$ aralığının dışına çıkması çok istenilen bir durum değildir. Çünkü aralık $[0 + 1]$ 'in dışına çıkar ise kümeler bulanıklık özelliğini kaybetmektedir. Bu sebepten dolayı $p=0$ veya $p>0$ olmalıdır. Özniteliğin önem derecesi P değerine bakılarak oluşturulmaktadır. P değeri 1'e yaklaştıkça özniteliğin önem derecesi artmakta, 0'a yaklaştıkça azalmaktadır. P değeri 0 olarak hesaplanmış ise o zaman o özniteliğin, sınıflandırma üzerinde bir etkisinin olmadığı ortaya konulmuş olmaktadır (Cetişli ve Kalkan, 2011).

İki-girişli, bir-çıkışlı bir bulanık kuralın normal ve dilsel kuvvetli halinin Mamdani çıkarım sistemindeki sonuçları Şekil 4'de gösterilmiştir.

Şekil 4: İki-girişli, bir-çıkışlı bir bulanık kuralın normal ve dilsel kuvvetli halinin Mamdani çıkarım sistemindeki sonuçları.

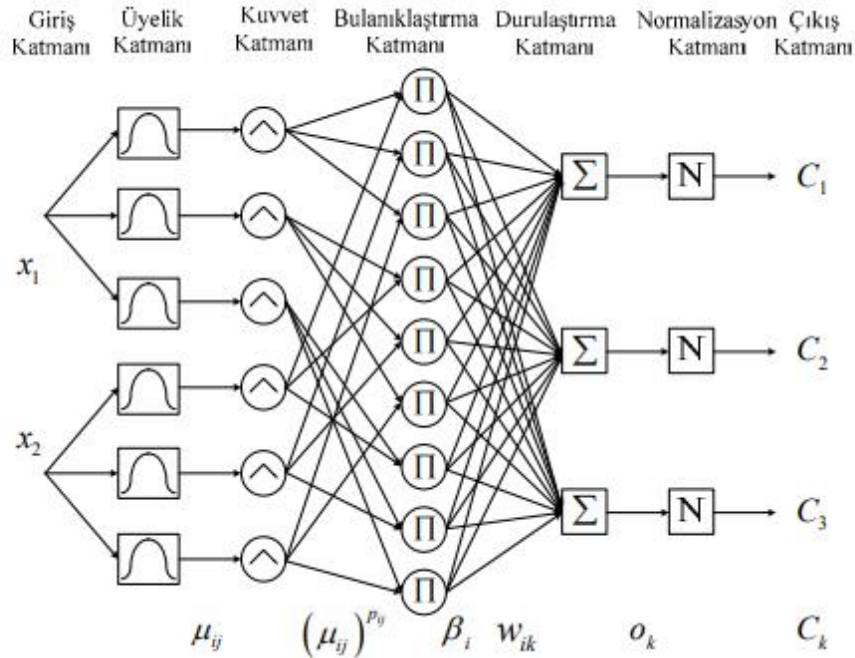


(Çetişli, 2006).

Şekil 4'de görüldüğü gibi, x_1 ve x_2 giriş değişkenleri için, B kümesinin yüksek bir derecede seçilmesi gerekiyor ise, A_2 bulanık kümesi buna engeldir. Yapılan seçimin doğruluğunun artırılması için yalnızca A_1 kullanılmalı ve A_2 devre dışı bırakılmalıdır. A_2 bulanık kümesini devreden çıkarmak için dilsel kuvvet olarak 0 kullanılmaktadır. Dolayısıyla X_2 uzayı da kuraldan çıkarılmış olmaktadır. Şekil 4'ün altında verilen çıkarsamada, X_2 uzayı çıkarsama için kullanılmadığı için B kümesinin seçilme derecesi artmaktadır (Çetişli, 2006).

Dilsel kuvvetli sinir-bulanık sınıflayıcının çalışma yapısı Şekil 5'de görüldüğü sıra ile gerçekleşmektedir.

Şekil 5: Dilsel kuvvetli sinir-bulanık sınıflayıcı'nın çalışma yapısı



(Cetişli ve Kalkan, 2011).

DKSBS'da takip edilen aşamalar; giriş katmanı, üyelik katmanı, kuvvet katmanı, bulanıklaştırma katmanı, durulaştırma katmanı, normalizasyon katmanı ve en son çıkış katmanı şeklindedir (Cetişli ve Kalkan, 2011).

4.2. Kümeleme Yöntemleri

4.2.1. K-Means

Kümeleme yöntemi, bir dizi elemanı benzer gruplarda toplama işlemidir. Kümeleme yöntemi veri setlerini gruplara ayırmak ve verilerdeki önemli ve gizli kalmış bilgileri açığa çıkarmak için en çok kullanılan yöntemlerden bir tanesidir (Çınaroğlu ve Bulut 2018). Kümeleme yönteminin, diğer bir veri madenciliği yöntemi olan sınıflandırmadan farkı şudur; sınıflandırmada sınıflar önceden bellidir ama kümelemede sınıflar önceden belli değildir (Doğan vd., 2018). K-means en eski kümeleme algoritmalarından bir tanesidir (MacQueen, 1967) ve hiyerarşik olmayan kümeleme yöntemleri arasında yer almaktadır (Yavuz vd., 2011). 1967 yılında J.B. MacQueen tarafından geliştirilen algoritma (Işık ve Çamurcu, 2007), yıllardır birçok alanda kullanılmaya devam

etmektedir (Bradley vd., 1998). K-means'ın tercih edilmesinin en temel sebebi yapısının sadeliği yanında model başarı oranının oldukça yüksek olmasıdır. K-means, kümeleri eşit büyüklükte ve homojen olarak oluşturma eğilimindedir (Işık ve Çamurcu, 2011) ve her elemanı bir kümeye atamaktadır. Böylece her eleman bir kümeye ait olmakta ve kümeler birbirinden kesin olarak ayrılmaktadır. Kümelemenin kullandığı öğrenme tekniği gözetimsiz öğrenmedir (Sariman, 2011). Yani kümeleme yapılırken önceden belirlenmiş kıstaslar yoktur. Kümeler, kümeyi temsil eden merkez noktaların belirlenmesi fikrine dayalı olarak oluşturulmaktadır (Güner vd., 2018). Kaç kümenin olacağı k parametresi ile ifade edilmektedir (Khaled vd., 1998). K parametresi aynı zamanda kaç tane küme merkezinin olacağı bilgisini taşıdığı için, K-means algoritması kullanılırken k parametresinin önceden girilmesi gerekmektedir (Bilgin ve Çamurcu, 2005). Kümeler belirlenirken; grup elemanların grup içi benzerliklerin en fazla ve gruplar arası benzerliklerin en az olması hedeflenmektedir (Demiralay ve Çamurcu 2005). Benzer olan kümeler oluşturulurken veriler arasındaki uzaklıklara bakılmaktadır. Bunun için eldeki veri kaç kümeye ayrılmak isteniyor ise o kadar küme merkezi belirlenmektedir (k parametresi) ve her elemanın bu belirlenen küme merkezlerine olan uzaklıkları tek tek hesaplanmaktadır. Hesaplamama sonucu her bir veri, kendine en yakın olan küme merkezine dâhil edilmektedir. Oluşan yeni kümelerin, küme merkezleri yeniden hesaplanmaktadır ve her bir elemanın yeni küme merkezlerine olan uzaklıkları da yeniden hesaplanmaktadır. Bu işlem hiçbir elemanın dâhil olduğu küme değişmeyinceye kadar devam etmektedir. En sonunda hiçbir elemanın dahil olduğu küme merkezi değişmiyor ise o zaman hesaplama durdurulmaktadır ve her eleman en son bulunduğu kümeye atanarak algoritma bitirilmektedir (Tekin vd., 2011).

Algoritmada uzaklık hesaplamalarından sonra birbirine en yakın olan elemanlar en çok benzerlik gösteren elemanlardır ve aralarında benzer oldukları tespit edilen elemanlar bir kümede toplanmaktadır (Çalışkan ve Soğukpınar, 2008).

K-means'ın en büyük eksiği verilerin kaç kümeye ayrılacağını, yani k parametresinin kaç olacağını tespit edememesidir. Bunun için, eğer verilerin kaç kümeden oluştuğu biliniyorsa k parametresi olarak bilinen küme sayısı girilebilmekte ya da farklı k parametreleri girilerek en uygun olanı tespit edilebilmektedir.

K-means algoritmasında elemanların küme merkezlerine olan uzaklıkları hesaplanırken kullanılan uzaklık formüllerinden bazıları şunlardır;

Öklid Uzaklığı,

City-Block (Manhattan) Uzaklığı,

Chebychev Uzaklığı.

En yaygın uzaklık formülü öklid uzaklığıdır. Bu çalışmada da kullanılan uzaklık öklid uzaklığı olmuştur. Öklid uzaklığı iki nokta arasındaki doğrusal uzaklığı vermektedir. Nesnelere arasındaki öklid uzaklığı sıfıra ne kadar yakınsa, nesnelere birbirine benzerliği o kadar fazladır (Güner vd., 2018). Noktalar $A = \{ x_1, x_2, \dots, x_n \}$, $B = \{ y_1, y_2, \dots, y_n \}$ şeklinde düşünüldüğünde öklid uzaklığı Eşitlik 1'deki gibi hesaplanmaktadır (Taşcı ve Onan 2016).

$$\text{Öklid uzaklığı (A, B)} = \left(\sqrt{\sum_{i=1}^n (x_i - y_i)^2} \right) \quad (1)$$

Girilen k adet küme için K-means algoritmasının çalışma şekli şu şekildedir;

- 1) İlk olarak, k adet rasgele küme merkezi belirlenir.
- 2) Her elemanın, tüm küme merkezlerine olan uzaklıkları hesaplanır ve elemanlar en yakın olan kümeye atanır.
- 3) Elemanlar en yakın oldukları kümeye atandıktan sonra, kümelerin küme merkezleri yeniden hesaplanır.
- 4) Herhangi bir elemanın ait olduğu küme de değişiklik oluyorsa, 2. Adıma gidilir ve algoritma devam eder. Hiçbir elemanın ait olduğu küme değişiyorsa algoritma sonlanır (Doğan vd., 2018).

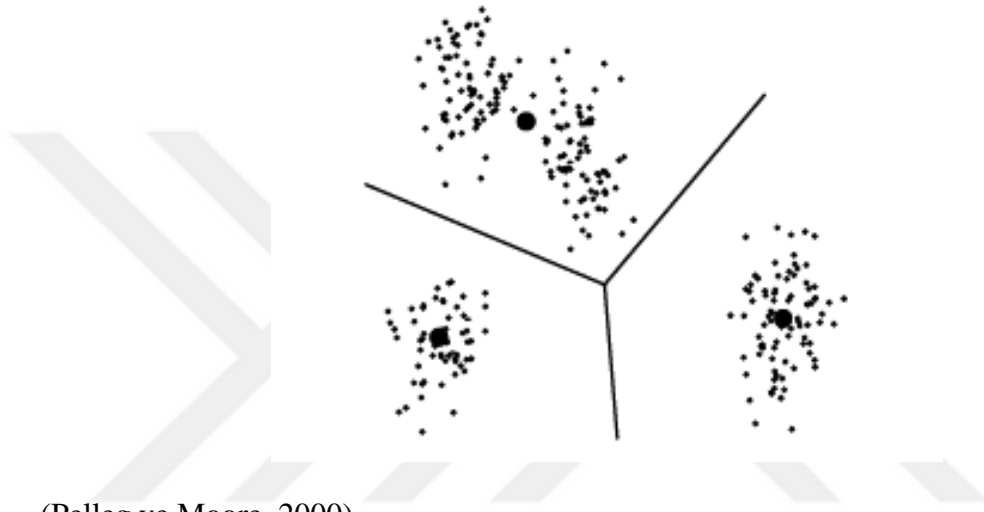
4.1.2. X-Means

K-means, kümeleme algoritmaları arasında oldukça popülerdir. K-means'in çok fazla tercih ediliyor olmasının sebebi kolay anlaşılabilir bir yapıya sahip olmasının yanında model başarımlarının oldukça yüksek olmasıdır. Diğer bir açıdan popüler olmasının yanında bazı eksikleri mevcuttur (Akçapınar vd., 2016). Bu eksiklerden en belirgin olanı; küme sayısını ifade eden k parametresinin kullanıcı tarafından sabit bir değer olarak sağlanması gerektiğidir. Küme sayısının sabit bir k değeri ile sınırlandırılması

küme sayısı için daha iyi olabilecek diğer alternatiflerin göz ardı edilmesi anlamına gelmektedir (Pelleg ve Moore, 2000). X-means'te veriler bir kd-tree'ye yerleştirilerek her aşamadaki istatistiksel veriler depolanmaktadır. Depolanan istatistiksel veriler içinde, belirli bir bölge için göz önünde bulundurulması gereken merkezlerin listesi de tutulmaktadır (Pelleg ve Moore, 2000). Böylece tüm alternatifler değerlendirilerek en iyi olanın seçilmesi imkânı sağlanmış olmaktadır (Muhr ve Granitzer, 2009). X-means algoritması, K-means algoritmasının geliştirilmiş hali olarak (Tsai ve Chiu, 2008) 2000 yılında Pelleg ve Moore tarafından geliştirilmiş bir kümeleme (bölütleme) algoritmasıdır (Pelleg ve Moore, 2000). Temelinde K-means algoritmasının çalışma şeklini kullanmasının yanı sıra eksik görülen yönünün tamamlanması amacıyla oluşturulmuştur (Kalogeratos ve Likas, 2012). X-means ile K-means'in eksikliği olarak görülen küme sayısının algoritma tarafından belirlenememesi sorununa çözüm sunulmaktadır (Akçapınar vd., 2016). X-means'te küme sayısı için sabit bir değer belirtilmek yerine makul olabilecek bir aralık belirtilmektedir. Böylece X-means belirtilen bu aralıktaki değerlerden (Steinley, 2006) en doğru oluşuna karar verdiği küme sayısını kendisi belirleyebilmektedir (Bholowalia ve Kumar, 2014). X-means yapısı içerisinde aşamalı olarak K-means algoritması çalıştırılmaktadır (Feng ve Hamerly, 2007). Alt kümelerin daha doğru oluşturulması için K-means her çalıştıktan sonra, hangi merkezlerin kendilerini bölmeleri gerektiği konusunda yerel kararlar oluşturularak ilerlenmektedir. Bu bölünme kararları bayesin bilgi kriteri (BIC) hesaplanarak verilmektedir (Jain, 2010; Hamerly ve Elkan, 2004). Her aşamada mevcut merkezlerin (ebeveyn) ve yeni oluşturulan yavru (çocuk) merkezlerin en iyi sonuçları blaclisting'e kaydedilmektedir. Kayıtlar tutulurken küme merkezlerinin yanında her k parametresi için model seçim kriterini puanlamak için kullanılan değerlerde tutulmaktadır (Pelleg ve Moore, 2000). Böylece merkez konumlarının dikkatlice analiz edilmesine imkan sağlanmaktadır (Tucker vd., 2010). Temelde, algoritma verilen aralığın alt sınırına eşit k ile başlamakta ve üst sınıra ulaşılan kadar aşamalı olarak yeni merkezler eklemeye devam etmektedir. Bu işlemler sırasında, en iyi puanı elde eden centroid seti kaydedilmektedir. Bir sonraki işlemde elde olan puan, sistemde kayıtlı olan puandan daha yüksek ise yeni puan listeye eklenmektedir. Böylece liste sürekli olarak güncellenmektedir. Güncelleme prosedürü özyinelemelidir ve bir bölge içindeki noktalar için olası centroidlerin listesini tutmaktadır. Görevi, söz konusu bölgelerin

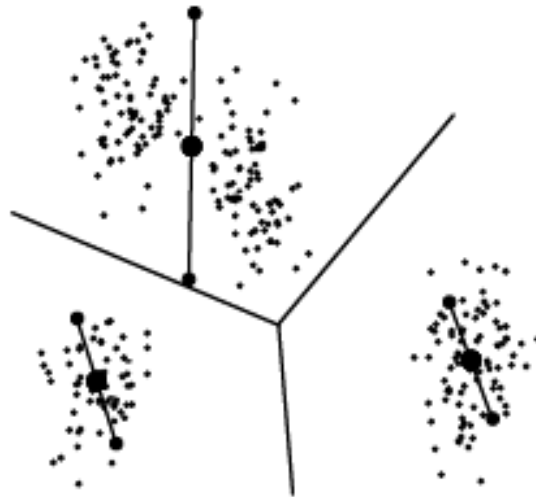
merkez noktalarını uygun değerler ile güncellemektir. İlk olarak, girilen k parametre listesinin en küçük değerindeki sayı kadar rasgele oluşturulan merkez noktalarının yazılması ile başlamaktadır (Pelleg ve Moore, 2000). Yeni değerler hesaplandıktan sonra daha iyiler bulundukça güncellenmektedir. En sonunda en yüksek değere sahip merkezler çıktı olarak kullanılmaktadır (Bholowalia ve Kumar, 2014). X-means algoritmasının çalışma yapısı aşamalı olarak Şekil 6-10 arasında görselleştirilmiştir.

Şekil 6: Üç merkez içeren bir K-means yapısı



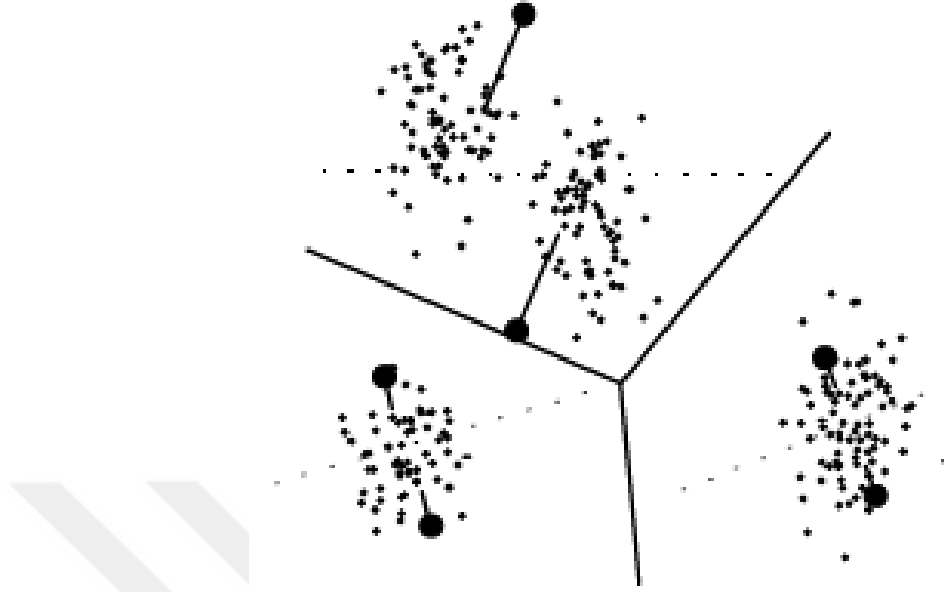
(Pelleg ve Moore, 2000).

Şekil 7: Bölünme için rastgele seçilen bir vektör görüntüsü



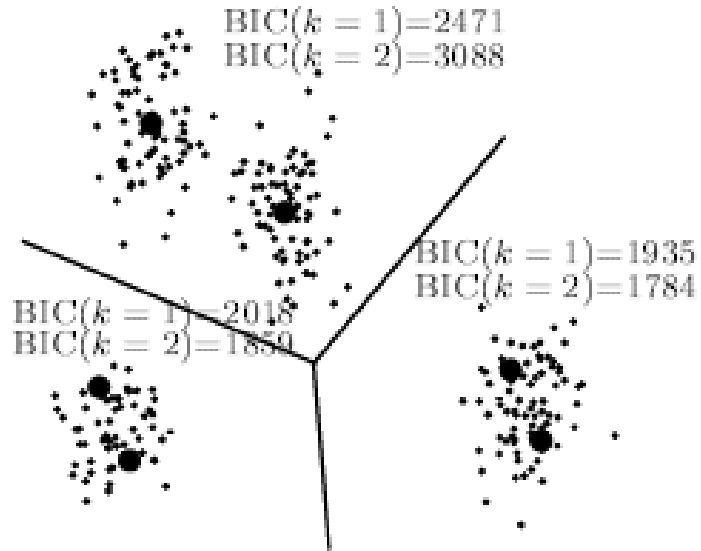
(Pelleg ve Moore, 2000).

Şekil 8: K-means'in üç ayrı bölgede ($k=2$ ile) çalışmasının ilk basamağı



(Pelleg ve Moore, 2000).

Şekil 9: K-means'lerin ($k=2$ için) tamamlanması ile oluşan kümeler



(Pelleg ve Moore, 2000).

Şekil 10: Sonuçta elde edilen kümeler



(Pelleg ve Moore, 2000).

Şekil 6-10 arasında X-means'in çalışma yapısı görsel olarak ifade edilmiştir. Şekiller ile ifade edilen çalışma aşamaları aşağıdaki gibidir.

Şekil 6'da merkez içeren bir K-means yapısı ve her bir merkezin sahip olduğu bölgenin sınırları gösterilmiştir. Yapı iyileştirme işlemi, her bir merkezin iki çocuğa bölünmesi ile başlamaktadır. İlk bölünme Şekil 7'de gösterilmiştir. Bölünmenin gerçekleşmesi için öncelikle merkez noktalardan geçecek şekilde ve uzaklıkları merkez noktaya eşit olacak şekilde rastgele seçilen bir vektör oluşturulmaktadır. Bu vektörün uç noktaları, merkez noktası olarak belirlenmektedir. Daha sonra her bölgede belirlenen bu merkez noktaları için yerel bir K-means ($k = 2$ ile) çalıştırılmaktadır. Her bölgede 2 olarak belirlenen K-means algoritmasının çalıştırıldığı durum Şekil 8'de gösterilmiştir. Şekil 9'da k 'nın 2 olarak belirlendiği K-means'lerin tamamlanması ile oluşan kümeler ve küme merkezleri gösterilmektedir. Yeni oluşan kümeler çocuk küme olarak adlandırılmaktadır. Önceki merkezlere ise ebeveyn denilmektedir. Bu noktada, tüm çocuk çiftlerinde ve ebeveynlerde BIC değeri hesaplanarak model seçim testi yapılmaktadır. BIC değeri hesaplanırken, bölünme gerçekleşmeden önceki ebeveynlerde küme merkezinin BIC değeri ve bölünme gerçekleştikten sonra oluşan çocuk küme merkezlerinin BIC değeri hesaplanmaktadır. Her hesaplamalardan sonra test sorusu "iki çocuğun burada gerçek

yapıyı modellediğine dair kanıt var mı, yoksa yapıyı ebeveyn daha iyi derecede mi modelliyor?" şeklindedir. Şekil 9 'da ifade edilen BIC değerleri bu sorunun cevabını vermektedir. Çünkü bölünme gerçekleşmeden önceki ve bölünme gerçekleştikten sonraki değerler bulunmaktadır. Testin sonucuna göre, ebeveyn veya yavruları öldürülmektedir. Son durum ise Şekil 10'da gösterilmiştir. Şekil 10'da ulaşılan, sonuç Şekil 9'da hesaplanan BIC değerlerinden sonra ebeveyn kümenin mi yoksa yavru kümelerin mi hayatta kalacağına belirlenmesinden sonra gerçekleşmektedir. Kalacak olan küme merkezleri belirlenirken ebeveyn kümenin (k 'nın 1 olduğu küme merkezi) BIC değeri, bölünme gerçekleştikten sonra oluşan yavru küme merkezlerinin (k 'nın 2 olduğu küme merkezi) BIC değerinden daha büyük ise yavru kümelerin küme merkezleri iptal edilmektedir. Durum tam tersi olur ise yani ebeveyn kümenin BIC değeri, yavru küme merkezlerinin BIC değerinden daha küçük ise ebeveyn kümelerin küme merkezleri iptal edilmektedir. Bu örnek için bakıldığı zaman üç bölgeden birinde bölünme gerçekleşirken, diğer iki bölgede bölünmenin gerçekleşmeyeceğine karar verildiği görülmektedir.

Bu şekildeki aşamaların hepsi k 'nın girilmiş olan üst sınırına gelinceye kadar çalıştırılmaktadır. Böylece kullanıcı tarafından girilen aralıktaki tüm k 'ların değerlendirilme şansı olmakta ve en uygun k değeri bulunabilmektedir.

4.3. Ön İşleme Teknikleri

Veri madenciliği yöntemleri ile analiz edilmek için toplanan veriler eksik olabilmektedir. Bu gibi durumlarda analizleri eksik veri ile gerçekleştirmek mümkündür. Aynı zaman da veri setinin kalitesini artırmak için eksik verilerin düzeltilmesi de tercih edilmektedir. Analizlerin eksiksiz veri üzerinde gerçekleştirilmesi kullanılan yöntemin başarı oranını arttırmada da oldukça önemli bir paya sahiptir. Eksik verinin düzeltilmesi veri ön işleme aşamasında yer almaktadır. Eksik veriyi düzeltmek için eksik verinin olduğu kayıtların çıkarılması (Sezgin ve Çelik, 2013), veya çeşitli yöntemlerle tamamlanması mümkündür (Aydın, 2007). Eksik olan verilerin tamamlanması için literatürde kabul görmüş birçok tamamlama yöntemi mevcuttur. Eksik değerler tamamlanırken dışardan genel bir sabit değer kullanılabilenkte, diğer verilerin özellikleri dikkate alınarak bir değer atanabilmekte (ortalama, mod ya da medyan gibi) (Aydın, 2007) ya da tahminlemeler sonucu bulunan değerler

kullanılabilmektedir (Regresyon analizi, Hot Deck, Naive Bayes İle Değer Atama, Son Gözlemi İleri Taşıma, Karar Ağaçlar, Beklenti Maksimizasyonu, Çoklu atama gibi) (Sezgin ve Çelik, 2013). Bu çalışmada eksik verilerin tamamlanması için diğer verilerin özellikleri dikkate alınarak bir değer ataması yapılmıştır. Veri seti içerisindeki her bir özneliğin frekansına bakılmış ve eksik olan veriler tamamlanırken frekansı en yüksek olan değer kullanılmıştır.

Günümüzde, veriler içinden anlam çıkarmak için veri madenciliği yöntemleri oldukça fazla kullanılmaktadır. Veri madenciliği ile bilginin keşfi süreci; veri seçimi ve toplanması, veri ön işleme, veri dönüştürme, veri madenciliği yöntemlerini uygulama ve en sonunda sonuçları değerlendirme şeklinde gerçekleşmektedir (Sebik ve Bülbül, 2018). Elde edilen verilerin kalitesi kullanılan yöntemin başarı oranını arttırmada çok büyük bir paya sahiptir (Oğuzlar, 2003). Verinin kalitesinin artırılması için kullanılan birçok veri ön işleme tekniği vardır (Mackinnon ve Glick, 1999). Veri setindeki eksik verilerin tamamlanması (Sezgin ve Çelik, 2013), gürültülü verilerin düzeltilmesi veya çıkarılması (Akgöbek ve Çakır, 2009), özneliklerin önem derecesinin belirlenmesi (Çetişli, 2006), bazı öznelikler üzerinde normalizasyon işleminin yapılması (Jayalakshmi ve Santhakumaran, 2011) kullanılan ön işleme tekniklerinden bazılarıdır. Bu çalışmada ön işleme tekniklerinden eksik verilerin tamamlanması ve normalizasyon kullanılmıştır. Literatürde kabul görmüş birçok normalizasyon çeşidi mevcuttur. Z-Score Normalizasyonu, Min-Max Normalizasyonu, Medyan Normalizasyonu, Sigmoid Normalizasyonu kabul gören normalizasyon yöntemlerinden bazılarıdır (Yavuz ve Deveci, 2012). Normalizasyon yöntemlerinin biri ya da birkaçı aynı anda kullanılabilmektedir. Bu çalışmada Min-Max normalizasyon yöntemi kullanılmıştır.

Min-Max Normalizasyon

Min-Max normalizasyon yöntemi ile veriler doğrusal olarak normalize edilmektedir. Min olarak ifade edilen değer minimum değer demektir ve veri içindeki en düşük sayıyı ifade etmektedir. Max olarak ifade edilen değer maksimum değer demektir ve veri içindeki en yüksek sayıyı ifade etmektedir. Min-Max ile normalize edilen veriler 0 ile 1 aralığına indirgenmiş olmaktadır (Yavuz ve Deveci, 2012). Veriler Min-Max yöntemi ile normalize edildikten sonra En düşük sayı 0, en yüksek sayı 1 değerini almaktadır. Aradaki değerler verilerin birbirine uzaklığı ile bağlantılı olarak büyüdükçe 1'e,

küçüldükçe 0'a yaklaşacak şekilde değerler almaktadır (Tunç ve Ülger, 2016). Verileri Min-Max yöntemi ile normalize etmek için (2) nolu formül kullanılmaktadır.

$$X' = \frac{X_i - X_{Min}}{X_{Max} - X_{Min}} \quad (2)$$

Bu eşitlikte;

X_i = Min-Max ile normalize edilecek veri

X_{Min} = Normalize edilecek veri seti içerisindeki en düşük sayı değeri

X_{Max} = Normalize edilecek veri seti içerisindeki en yüksek sayı değeri

X' = Hesaplama sonucunda normalize edilmiş olan veri

(Yavuz ve Deveci, 2012).

4.4. Performans Ölçütleri

Veri madenciliği, yapay zekâ gibi teknikler kullanılarak gerçekleştirilen tahminlemelerin ne kadar başarılı olduğunun veya başka bir ifade ile ne kadar güvenilir olduğun ortaya konulması için bir takım hesaplamalar yapılabilmektedir. Böylece elde edilen sonuçların doğruluğu ya da hatası ölçülebilir olmaktadır. Bu durum farklı yöntemlerin de karşılaştırılmasına imkân tanımaktadır. Yöntemlerin başarıları ölçülürken Doğruluk, Duyarlılık, Belirleyicilik, F-Ölçütü, Ortalama Hata (Mean Squared Error) ve Karekök Ortalama Hata (Root Mean Squared Error) gibi ölçütlerden faydalanılabilmektedir.

$$\text{Mean Squared Error (MSE): } \sum_{i=1}^N (Y_i - X_i)^2 / N \quad (3)$$

$$\text{Root Mean Squared Error (RMSE): } \sqrt{\sum_{i=1}^N (Y_i - X_i)^2 / N} \quad (4)$$

Eşitlik 3 ve 4 de kullanılan ifadelerin karşılıkları şöyledir;

Y_i = i .ci örnek için tahminlenen değer,

X_i = i .ci örnek için gerçek değer,

N = Örnek sayısı.

MSE ve RMSE kelime anlamlarına bakıldığı zaman ortalama hata ve karekök ortalama hata demektir. Hata değerlerinin düşük çıkması başarının da yüksek olduğunu ifade

etmektedir. Dolayısı ile MSE ve RMSE hesaplamalarında istenilen sonuç, değerlerin düşük çıkması şeklindedir.

Eğer sınıflandırılan ya da kümelenen veri setlerinde sonuç ikili (True False, Doğru Yanlış, Evet Hayır, Var Yok, 0 1 gibi) ise o zaman Hata Matrisine ve Doğruluk, Duyarlılık, Belirleyicilik, F-Ölçütü gibi değerlere de bakmak mümkündür. Hata Matrisi ve temelinde bu matrise dayanarak hesaplanan Doğruluk, Duyarlılık, Belirleyicilik, F-Ölçütü doğru sınıfa atanan ve yanlış sınıfa atılan örnek sayıları ile ilişkilidir. Tablo 2’de örnek bir hata matrisi verilmiştir.

Tablo 2: Örnek bir hata matrisi

Gerçek Sınıf	Model Tarafından Tahminlenen Sınıf		
		True	False
True	TP	FN	
False	FP	TN	

Tablo 2’deki satırlar verilerin gerçek sınıflarını, sütunlar ise verilerin model tarafından tahminlenen sınıflarını göstermektedir. Tablo içerisindeki ifadelerin anlamları ise şöyledir;

TP (True Pozitif): Gerçek sınıfı true, tahmin edilen sınıfı true olan örnek sayısı

FN (False Negatif): Gerçek sınıfı true, tahmin edilen sınıfı false olan örnek sayısı

FP (False Pozitif): Gerçek sınıfı false, tahmin edilen sınıfı true olan örnek sayısı

TN (True Negatif): Gerçek sınıfı false, tahmin edilen sınıfı false olan örnek sayısı

Dolayısı ile toplamda doğru sınıflandırılan örnek sayısı $TP + TN$, toplamda yanlış sınıflandırılan örnek sayısı ise $FP + FN$ ’dir.

TP, TN, FP, FN sayıları kullanılarak hesaplanan Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerlerinin de formülleri sırası ile Eşitlik 5, 6, 7, 8’de verilmiştir.

$$\text{Doğruluk} = \frac{TP + TN}{(TP + TN + FP + FN)} \quad (5)$$

$$\text{Duyarlılık} = \frac{TP}{(TP + FN)} \quad (6)$$

$$\text{Belirleyicilik} = \frac{TN}{(TN + FP)} \quad (7)$$

$$\text{F-Ölçütü} = \frac{2 * TP}{(2 * TP + FP + FN)} \quad (8)$$

5. UYGULAMA

5.1. Otizm Spektrum Bozukluğu Veri Seti

Bu çalışmada güncel ve gerçek OSB verilerinin çocuklar için olan alt kümesi kullanılmıştır. Verilerin ana kaynağı <http://fadifayez.com/autism-datasets/> internet sitesidir. Aynı zamanda Fadi Thabtah'ın 2017 yılında yaptığı bilimsel yayınla birlikte UCI machine learning repository sitesinde paylaşılmış ve kullanmak isteyenler için erişime açılmıştır. Kullanılan veri setinin adı Autistic Spectrum Disorder Screening Data for Children Data Set'tir. Veri seti, OSB'nin teşhisi için literatürde kabul gören son parametrelere göre hazırlanan soruların cevapları ile oluşturulmuştur. Cevaplar ASDTests uygulaması ile toplanmıştır. Uygulamanın içinde yaş gruplarına göre ayrı kategoriler mevcuttur. Her kategoride on soru bulunmaktadır ve her biri kullanıcıların uygun cevabı dikkatlice seçmelerini sağlamak için bir resimle ilişkilendirilmiştir. Katılımcılara, bilgilerinin isimsiz tutulacağı ve sadece araştırma amacıyla paylaşılacağı bildirilmiştir. Değerlendirme tamamlamadan önce, katılımcılardan araştırmanın amacını, gizlilik politikasını ve verilerin kullanımını açıklayan bir bilgilendirme yapılmıştır. Kullanılan veri setinde girdi olarak kullanılan öznitelikler arasında otizm spektrum bozukluğunu ölçümleyen on sorunun yanında demografik özellikleri içeren bilgilerde yer almaktadır. Toplamda 1100 kişiye uygulanan bu test, üç alt grup verilerini içermektedir. Gruplar çocuk, genç ve yetişkinler şeklindedir. Bu grupların oluşturulmasının sebebi ise OSB'nin teşhisi için sorulan bazı soruların gruplar arasında farklılık göstermesidir. Oluşturulan veri setinin çocuklar için olan alt kümesinde toplamda 292 örnek vardır ve bu çalışmada yalnızca çocuklar için olan küme kullanılmıştır. Veri seti sayısal ve kategorik veri türlerini içermektedir. Veri seti içerisindeki bazı örneklerde eksik veriler bulunmaktadır. Toplamda 21 öznitelikten oluşan veri setinde; 20 öznitelik bireylerin genel ve sağlık durumu ile ilgili bilgiler içerdiği için girdi olarak kullanılmış, 1 öznitelik ise bireyin otizmliliğini gösteren durumunu içerdiği için çıktı olarak kullanılmıştır. Girdi olarak kullanılan öznitelik parametrelerinden biri olan 'tarama yönteminin türü' bireyin yaş aralığının hangi grupta olduğunu ifade etmektedir. Yani otizm tarama testini gerçekleştiren bireyin çocuk, genç ya da yetişkin gruplarının hangisine dâhil olduğu bilgisini göstermektedir. Bu çalışmada sadece çocuklar için olan alt küme kullanılmıştır. Tüm veri seti çocuklara ait olduğu için grup bilgisine ihtiyaç duyulmamıştır. Bu sebepten dolayı veri seti

içerisinden bu öznelik çıkarılmıştır. Böylece girdi parametresinin sayısı 19' a düşürülmüştür. Özneliklerin türü ve soruların detayları Tablo 3'de detaylı olarak verilmiştir.

Tablo 3: Veri setindeki özneliklerin detayları

Öznelik numarası	Öznelik	Veri türü	Açıklama
1	Soru 1'in cevabı	Binary (0, 1)	Diğerleri duymadığında genellikle küçük sesler fark eder.
2	Soru 2'in cevabı	Binary (0, 1)	Genelde küçük detaylardan ziyade resmin tamamına odaklanır.
3	Soru 3'in cevabı	Binary (0, 1)	Bir sosyal grupta, birkaç farklı insanın konuşmasını kolayca takip edebilir.
4	Soru 4'in cevabı	Binary (0, 1)	Farklı etkinlikler arasında gidip gelmeyi kolay buluyor.
5	Soru 5'in cevabı	Binary (0, 1)	Adaylarla sohbe nasıl devam edeceğini bilmiyor.
6	Soru 6'in cevabı	Binary (0, 1)	Sosyal sohbetinde iyidir.
7	Soru 7'in cevabı	Binary (0, 1)	Bir hikaye okuduğunda, karakterin niyetlerini veya duygularını çözmekte zorlanır.
8	Soru 8'in cevabı	Binary (0, 1)	Okul öncesi eğitimindeyken, diğer çocuklarla oyun oynamayı sevmekten zevk alır.
9	Soru 9'in cevabı	Binary (0, 1)	Birinin ne düşündüğünü veya hissettiğini, sadece yüzlerine bakarak kolayca anlayabilir.
10	Soru 10'in cevabı	Binary (0, 1)	Yeni arkadaşlar edinmekte zorlanıyor.

11	Yaş	Number	Bireyin yıl olarak kaç yaşında olduğu bilgisi.
12	Cinsiyet	String	Bireyin erkek ya da kadın olduğu bilgisi.
13	Etnik köken	String	Bireyin hangi etnik kökene sahip olduğu bilgisi.
14	Sarılık ile doğmuş olmak	Boolean (evet-hayır)	Bireyin sarılık ile doğup doğmama bilgisi.
15	PDD'li aile üyesi	Boolean (evet-hayır)	Bireyin herhangi bir aile üyesinin PDD'sinin olup olmadığı bilgisi.
16	İkamet edilen ülke	String	Bireyin ikamet ettiği ülkenin bilgisi.
17	Daha önce tarama uygulamasının kullanılması	Boolean (evet-hayır)	Kullanıcının daha önce bir tarama uygulaması kullanıp kullanmadığı bilgisi.
18	Puanlama sonucu	Integer	Nihai puan, kullanılan tarama yönteminin puanlama algoritmasına dayanarak elde edilmiştir.
19	Testi kim tamamlıyor	String	Bireyin testini kimin gerçekleştirdiği (Veli, kendisi, bakıcı, sağlık personeli, klinisyen, vb.) bilgisi.
20	Otizmli olma durumu (Çıktı)	Boolean (evet-hayır)	Bireyin otizmli olma durumunun bilgisi

(Thabtah vd., 2018)

Veri setinde bulunduğu halde veri setinin hangi yaş kategorisine uygulandığı bilgisini taşıyan ve tüm bireylerin çocuk olduğunu ifade etmek için 1 ile doldurulmuş olan

“tarama yöntemi türü” özniteliği veri setinden çıkarılmıştır. Boolean türünde olan bireyin otizmlili olup olmama durumu ise sınıf etiketi olarak kullanılmıştır.

Çalışmada öncelikle öncelikle tüm öznitelikler sayısal değerlere çevrilmiştir. Bir sonraki aşamada veri seti üzerinde daha etkin analiz yapabilmek için eksik verilerin tamamlanması, verilerin 0 ile 1 arasında normalize edilmesi gibi veri ön işleme aşamaları gerçekleştirilmiştir. Daha sonra veriler yapay sinir ağları ve DKSBS ile sınıflandırılmış, k-means ve x-means ile kümelendirilmiştir. Tasarlanan ileri beslemeli YSA için kullanılan nöron sayılarının başarı oranına etkisi incelenmiş ve farklı yapıda düğümleri olan modeller test edilmiştir. Sınıflandırma yöntemlerinin model başarı oranını test etmek için veriler %70 eğitim, %30 test olacak şekilde ayrılmıştır. Sonuçlara bakıldığında zaman veriler hem yapay sinir ağları hem de DKSBS ile sınıflandırıldığında başarı oranını hem test hem de eğitim setinde %100 olduğu görülmüştür. Kümeleme yöntemleri ön öğreticisiz yöntemlerdir. Dolayısıyla kümelenen veriler eğitim ve test olarak ikiye ayrılmamıştır ve verilerin tamamı eğitim için kullanılmıştır. Kümelenen verilerin başarı oranı k-means de % 89,73, x-means de %88,0 şeklindedir.

5.2. Bulgular ve Değerlendirme

Literatür incelendiğinde otizm konusunda veri madenciliği yöntemleri kullanılarak gerçekleştirilen çok fazla çalışma olduğu görülmektedir. Thabtah (2017), çalışmasında bu konuya dikkat çekmiş, gerçekleştirilen çalışmalarda oldukça güzel sonuçların elde edilmiş olmasıyla birlikte kullanılan verilerin güncelliğini yitirdiğinden bahsetmiştir. Bu sebepten dolayı da güncel verilerle çalışılması gerektiğini vurgulamıştır. Devamında ise Thabtah (2018), yeni bir çalışma ile güncel ölçeğe uygun şekilde toplanan verileri paylaşmıştır. Bu çalışmada da Thabtah (2018), tarafından güncel ölçeğe uygun şekilde toplanan ve yapılan bilimsel yayın ile paylaşılan verilerin çocuklar için olan alt kümesi kullanılmıştır. Kullanılan veri seti 292 örnekten oluşmaktadır ve veri setinde toplamda 21 öznitelik mevcuttur. Veriler çok eski olmadığından bu veriler ile yapılan çalışmalarda sınırlıdır ve bu çalışma ile elde edilen bulgular diğer çalışmalarla karşılaştırılmıştır.

Veri setinin orijinal hali sayısal ve string ifade türünde değişkenler içermektedir. Bu çalışma kapsamında kullanılan normalizasyon işlemi ve sınıflandırma işlemi için

kullanılan yöntemler ise string ifadelerle çalıştırılmamaktadır. Bu sebepten dolayı öncelikle tüm veri seti sayısal değerlere dönüştürülmüştür. Bu dönüşüm yapılırken string ifadelerle etiketlenmiş olan kategorik veriler, sayısal ifadeler ile etiketlenecek şekilde değiştirilmiştir. Örneğin bayan ya da erkek olarak etiketlenen kategorik veriler 0 ve 1 ile değiştirilmiştir. Etnik köken, ikamet edilen ülke gibi değerler de aynı şekilde 1'den başlanarak sıralı olarak etiketlenmiştir.

Tüm veriler sayısal hale getirildikten sonra her bir özneliğin kendi arasında frekansı çıkarılmıştır. Veri seti içerisindeki boş değerler frekansı en yüksek olan değer ile doldurulmuştur. Devamında ise birçok normalizasyon yöntemi denemiş en yüksek başarı oranına sahip literatürde de oldukça fazla kullanılan Min-Max normalizasyon yöntemi ile elde edilen değerler oluşturulmuştur. Yöntemlerin başarılarının karşılaştırılması için veri setinin bir kısmı eğitim bir kısmı test verileri olarak ayrılmıştır. Bu oranın yüzde kaç olarak ayrılması gerektiği konusunda literatür incelendiğinde çok farklı oranların kullanılmış olduğu görülmekle birlikte, en yaygın oran %70 eğitim verisi, %30 test verisi olarak ayrıldığı görülmektedir. Genel yaklaşıma uyması açısından bu çalışmada da %70'e karşılık gelen ve rastgele olarak belirlenen 205 örnek eğitim, %30'a karşılık gelen ve rastgele olarak belirlenen 87 örnek ise test için kullanılmıştır. Veriler YSA ve DKSBS ile sınıflandırılmış, K-Means ve X-Means ile kümeleneştir.

5.2.1. Yapay Sinir Ağı İle Elde Edilen Sonuçlar

Tasarlanan ileri beslemeli YSA ile kullanılan nöron sayılarının başarı oranına etkisi incelenmiş ve farklı yapıda düğümleri olan modeller test edilmiştir.

YSA'nın performansı Tablo 4'de verilmiştir.

Tablo 4: Yapay Sinir Ağının performansı

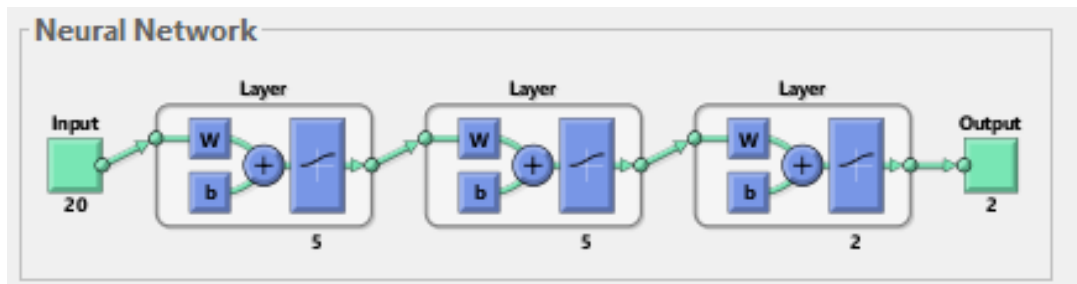
Giriş katmanında ki nöron sayısı	1.Gizli katmanda ki nöron sayısı	2.Gizli katmanda ki nöron sayısı	Çıkış katmanında ki nöron sayısı	Eğitim verisindeki doğruluk oranı	Test verisindeki doğruluk oranı
20	1	1	2	%91.21	%90.80
20	10	10	2	%100	%97.70

20	5	5	2	%100	%98.85
20	30	30	2	%100	%96.56
20	15	15	2	%100	%96.56
20	50	50	2	%100	%97.70
20	20	20	2	%100	%96.56

Tasarlanan modellerin başarı oranına bakıldığı zaman oldukça yüksek oranda başarıya ulaşıldığı görülmüştür. Farklı nöron sayıları ile tasarlanan modeller içerisinde en yüksek performans giriş katmanının da 20, iki gizli katmanda 5 ve çıkış katmanında 2 nöron kullanılarak tasarlanan modelde görülmüştür. Eğitim setinde %100 test setinde ise %98.85 oranında doğrulukla başarı elde edilmiştir. Bu durum eğitim setindeki tüm örneklerin doğru sınıflandırıldığı, test veri setinde ise yalnızca bir örneğin yanlış sınıflandırıldığı anlamına gelmektedir. Tablo 4’de görüldüğü gibi gizli katmandaki nöron sayısını arttırmak başarı oranının artması ile doğru orantılı değildir. Nöron sayısını arttırmak bazı veri setleri için modelin başarı oranını arttırmakla beraber, bazı veri setleri için modelin başarı oranını düşürmektedir. Dolayısı ile her zaman nöron sayısının çok fazla ya da çok az kullanılması ile alakalı sabit bir yaklaşımı benimsemek yerine birkaç alternatif belirlenerek en başarılı olan modelin tercih edilmesi başarı oranını arttırabilmektedir.

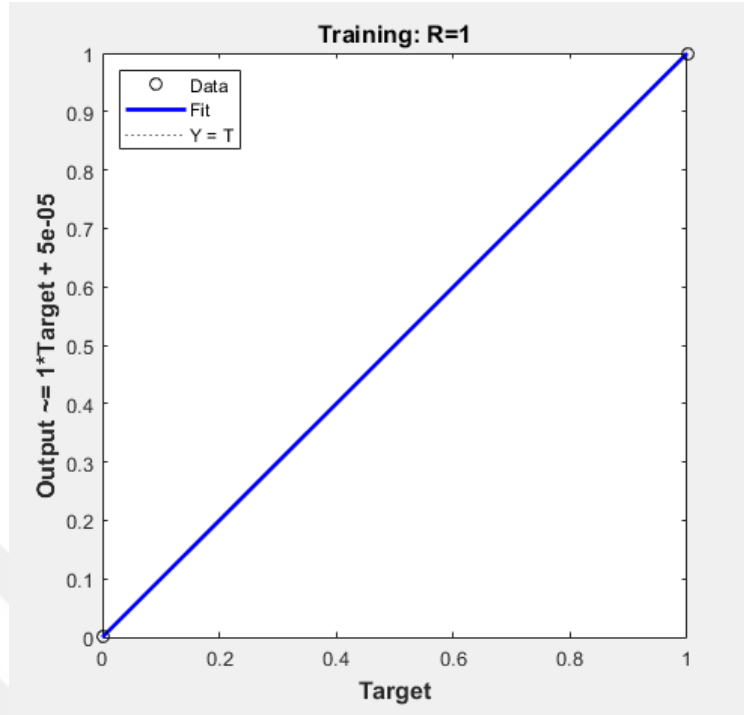
Başarı oranının en yüksek olarak görüldüğü YSA modeli Şekil 11’de gösterilmiştir.

Şekil 11: Başarı oranının en yüksek olarak görüldüğü YSA modeli



YSA'nın eğitildikten sonra eğitim verileri ile gerçekleşen uyumu Şekil 12’de gösterilmiştir.

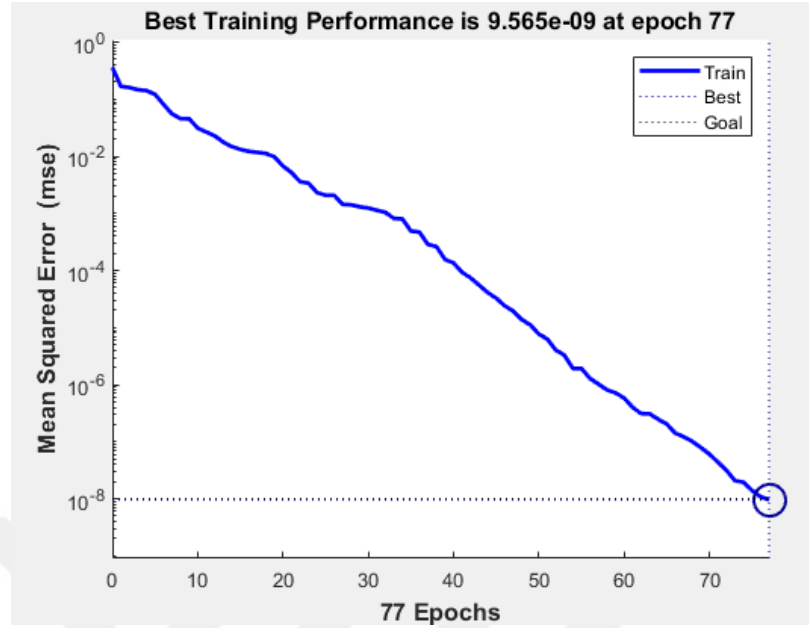
Şekil 12: YSA'nın eğitildikten sonra eğitim verileri ile gerçekleşen uyumu



Şekil 12'deki regresyon eğrisinde de görüldüğü gibi r değeri 1'dir. Yani eğitim eğrisi ile gerçek verilerin oluşturduğu eğri bire bir uymaktadır. Bu durum veriler ile uydurulan modelin tam anlamı ile örtüşüğünü ve eğitim verilerinin tamamının doğru sınıflandırıldığını göstermektedir.

Şekil 13'te geliştirilen YSA'nın eğitim aşamasındaki MSE değerine bakılarak oluşturulan performans grafiği verilmiştir.

Şekil 13: YSA'nın eğitim aşamasındaki MSE değerine bakılarak oluşturulan performans grafiği



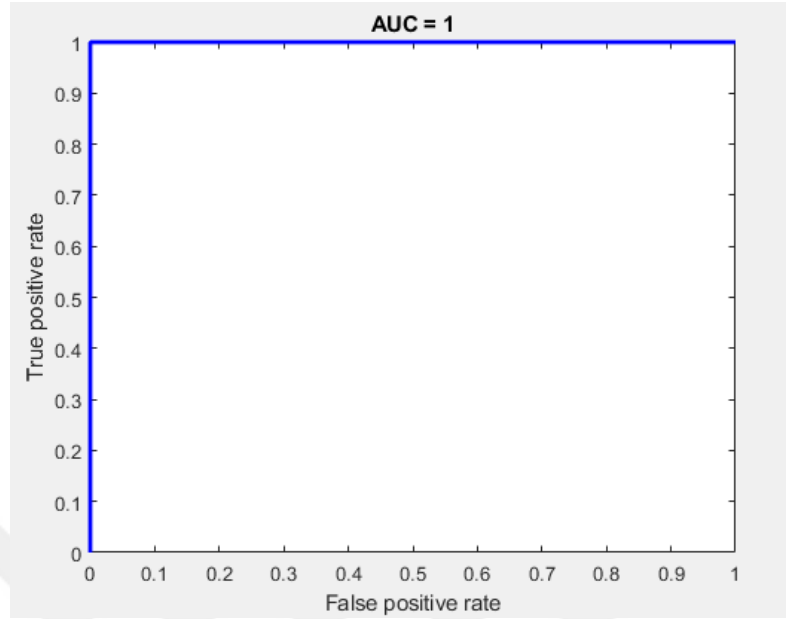
Şekil 13'de görüldüğü gibi ağın eğitimi sırasında hata oranı oldukça düşüktür. Bu durum ağın iyi eğitildiği anlamına gelmektedir.

Ağın eğitim verisi için ürettiği hata değerlerine detaylı olarak bakıldığında MSE değerinin $2.611e-08$, RMSE değerinin ise $1.616e-04$ olduğu görülmektedir.

Eğitilen ağın eğitim verileri için MSE ve RMSE değerlerinden de anlaşıldığı gibi ağın hata değerleri oldukça düşüktür.

Eğitilen ağın eğitim verisi için çizdirilen ROC eğrisi (Receiver Operating Characteristic) Şekil 14'de verilmiştir.

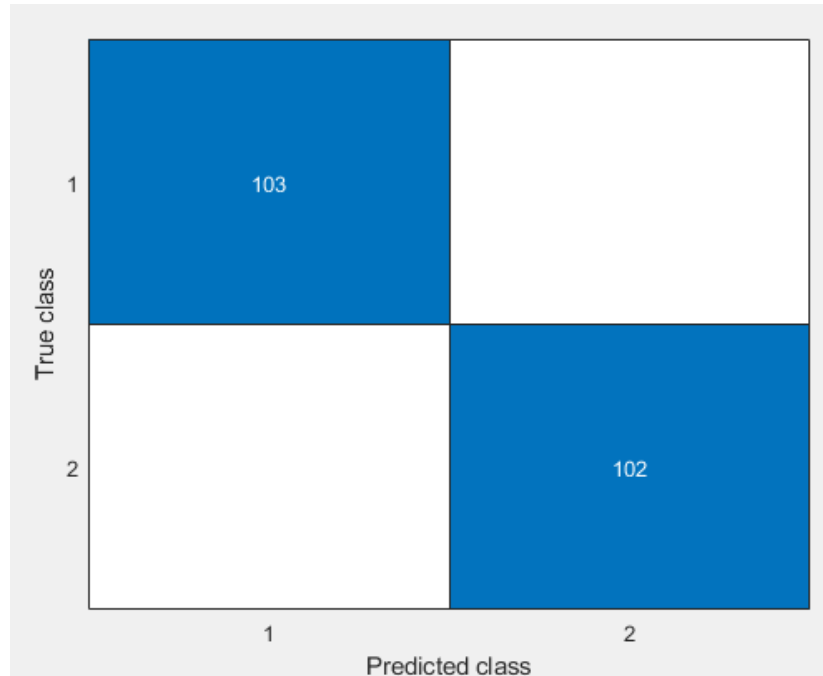
Şekil 14: Eğitilen ağın eğitim verisi için çizdirilen ROC eğrisi



Şekil 14’de görüldüğü gibi ROC eğrisi altında kalan alan (AUC) değeri 1’dir. Bu sonuç, eğitim verisinde yapılan sınıflandırmada tam bir başarı sağlandığını göstermektedir.

Eğitilen ağın eğitim verisi için çizdirilen hata matrisi Şekil 15’de verilmiştir.

Şekil 15: Eğitim verisi için çizdirilen hata matrisi

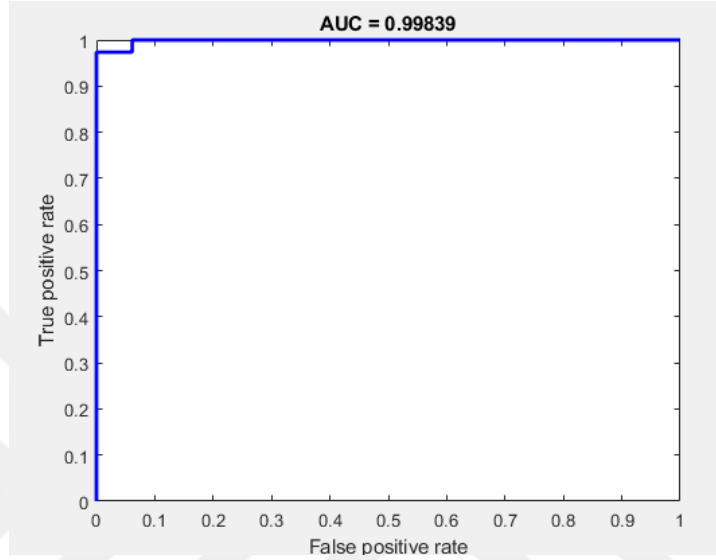


Şekil 15’de görüldüğü gibi eğitilen ağın eğitim veri seti için gerçekleştirdiği sınıflandırmanın tamamı doğrudur.

Eđitilen ađın, eđitim verileri iin gerekleřtirdiđi Dođruluk, Duyarlılık, Belirleyicilik ve F-Ölütü deđerleri sırası ile 1,1,1,1 řeklindeyir. Bu durum eđitim verilerinin tamamının dođru sınıflandırıldıđını gstermektedir.

Eđitilen ađın test verisi iin izdirilen ROC eđrisi řekil 16’da verilmiřtir.

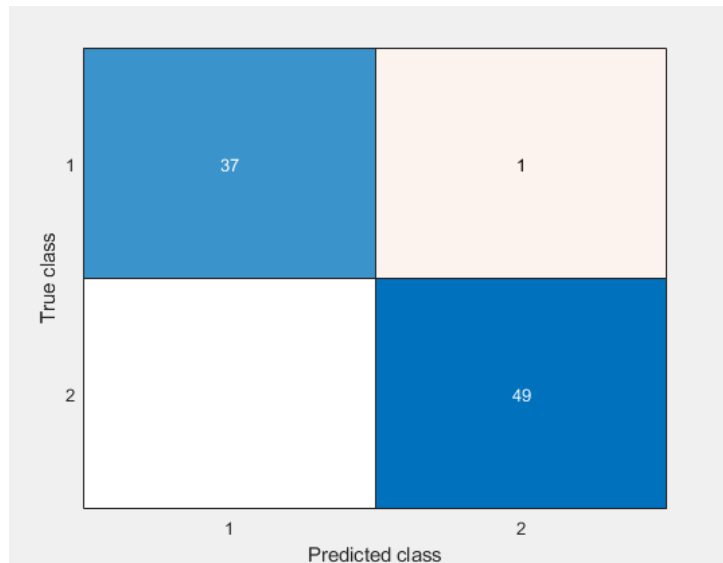
řekil 16: Eđitilen ađın test verisi iin izdirilen ROC eđrisi



řekil 16’da grldđ zere ROC eđrisi altında kalan alan (AUC) deđerleri 1’e ok yakın ıkmıřtır. Bu durum, test verisinde yapılan sınıflandırmada olduka yksek bir bařarı sađlandıđını gstermektedir.

Eđitilen ađın test verisi iin izdirilen hata matrisi řekil 17’de verilmiřtir.

řekil 17: Eđitilen ađın test verisi iin izdirilen hata matrisi



Şekil 17’de görüldüğü gibi eğitilen ağın test veri seti için gerçekleştirdiği sınıflandırmanın çok yüksek bir kısmı doğru tahminlenmiş, yalnızca bir örnek yanlış sınıflandırılmıştır. Bu durum test verisi için oldukça yüksek bir başarı anlamı taşımaktadır.

Ağın test verisi için ürettiği hata değerlerine detaylı olarak bakıldığında MSE değerinin $4.822e-04$, RMSE değerinin ise 0.022 olduğu görülmektedir.

Eğitilen ağın eğitim verileri için MSE ve RMSE değerlerinden de anlaşıldığı gibi ağın hata değerleri oldukça düşüktür.

Eğitilen ağın, test verileri için gerçekleştirdiği Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri sırası ile 0.989, 0.974, 1, 0.987 şeklindedir. Bu durum eğitilen ağın test verileri içinde oldukça başarılı sonuçlar verdiğini göstermektedir.

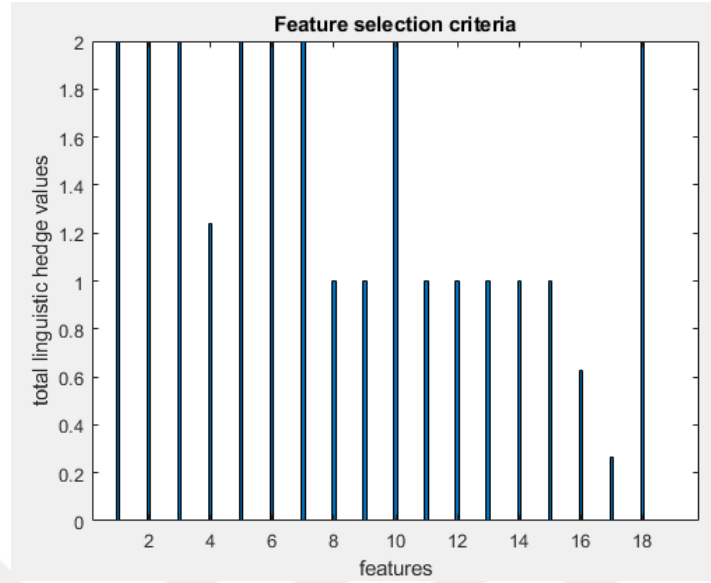
5.2.2. Dilsel Kuvvetli Sinir-Bulanık Sınıflayıcı İle Elde Edilen Sonuçlar

DKSBS ile öncelikle öznitelik seçimi gerçekleştirilmiş devamında ise sınıflandırma işlemi yapılmıştır. Yapılan sınıflandırmada veriler %70 eğitim, %30 test verisi şeklinde ayrılarak gerçekleştirilmiştir. Sınıflandırma sonuçlarına bakıldığı zaman eğitim ve test verilerinde %100 oranda başarı elde edildiği görülmüştür. Yöntemin eğitim verileri için hesaplanan hata değerlerine bakıldığı zaman MSE değerinin $3.099e-32$, RMSE değerinin ise $1.760e-16$ olduğu görülmektedir.

Gerçekleştirilen öznitelik seçimi, özniteliklerin önem düzeyine bakılarak belirlenmiştir.

Özniteliklerin DKSBS tarafından belirlenen önem düzeyleri Şekil 18’de verilmiştir.

Şekil 18: Özniteliklerin DKSBS tarafından belirlenen önem düzeyleri



Şekil 18’de numaralandırılmış özniteliklerin sırası Tablo 3’deki sıra ile aynı olacak şekilde 1’den 19’a kadar sıralı olarak ilerlemiştir. Dilsel kuvvetli sinir-bulanık sınıflayıcı’nın sınıflandırma için öznitelikler üzerinde belirlemiş olduğu önem derecelerine bakıldığı zaman beş farklı önem düzeyine sahip grupların oluştuğu görülmektedir. Bu grupları en büyük öneme sahip onları 5 olacak şekilde, 5 den 1’e kadar derecelendirilecek olursa; 5 önem düzeyine sahip öznitelikler 1,2,3,5,6,7,10,18 numaralı olanlar, 4 önem düzeyine sahip öznitelik 4 numaralı olan, 3 önem düzeyine sahip öznitelikler 8,9,11,12,13,14,15 numaralı olanlar, 2 önem düzeyine sahip öznitelik 16 numaralı olan ve 1 önem düzeyine sahip olan 17 numaralı olan özniteliktir. 19 numaralı özniteliğin ise önem derecesi 0 olarak belirlenmiştir.

Dilsel kuvvetli sinir-bulanık sınıflayıcı’nın ifade ettiği özniteliklerin katkı dereceleri Şekil 19’da görülmektedir.

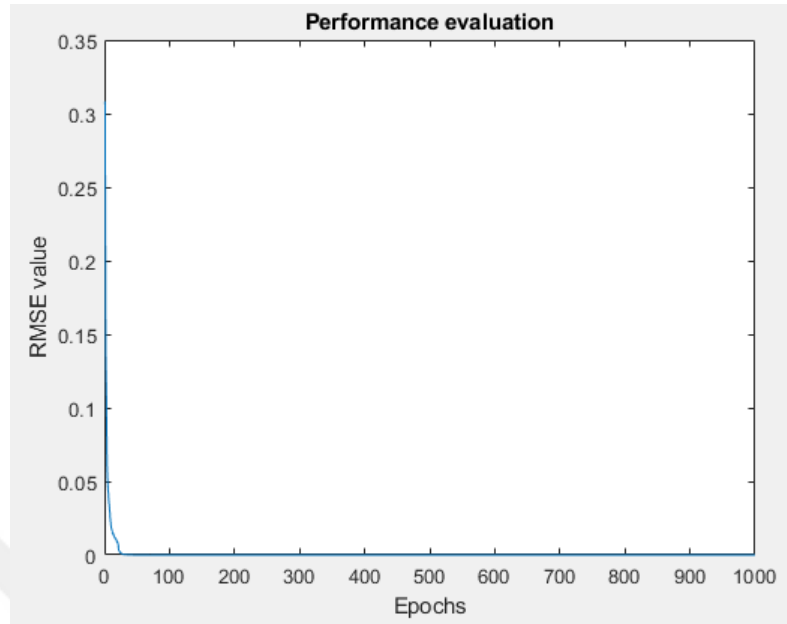
Şekil 19: Dilsel kuvvetli sınır-bulanık sınıflayıcı'ya ait bulanık kuralların modeli



Tablo 3’de yer alan 1,2,3,5,6,7,10 ve 18 numaralı öznitelikler olan çocuğun; diğerleri duymadığında genellikle küçük sesleri fark edip etmediği, genelde küçük detaylardan ziyade resmin tamamına odaklanıp odaklanmadığı, bir sosyal grupta, birkaç farklı insanın konuşmasını kolayca takip edebilip edemediği, adaylarla sohbe nasıl devam edeceğini bilip bilmediği, sosyal sohbetinde iyi olup olmadığı, bir hikaye okuduğunda, karakterin niyetlerini veya duygularını çözmekte zorlanıp zorlanmadığı, yeni arkadaşlar edinmeyi zor bulup bulmadığı ve kullanılan tarama yönteminin puanlama algoritmasına dayanarak elde edilmiş nihai puanın yer aldığı öznitelikler sınıflandırma üzerinde en çok etkiye sahip olan öznitelikler olarak belirlenmiştir. İkinci en etkili öznitelik olarak 4 numara ile ifade edilen; farklı etkinlikler arasında gidip gelmeyi kolay bulup bulmadığı durumu belirlenmiştir. Üçüncü en etkili öznitelikler 8,9,11,12,13,14,15 numaralarıyla ifade edilen; okul öncesi eğitimindeyken, diğer çocuklarla oyun oynamayı sevip sevmediği, birinin ne düşündüğünü veya hissettiğini, sadece yüzlerine bakarak kolayca anlayabilip anlayamadığı, bireyin yıl olarak kaç yaşında olduğu bilgisi, bireyin erkek ya da kadın olduğu bilgisi, bireyin hangi etnik kökene sahip olduğu bilgisi, bireyin sarılık ile doğup doğmama bilgisi ve bireyin herhangi bir aile üyesinin PDD’sinin olup olmadığı bilgisi’dir. Dördüncü en etkili öznitelik 16 numara ile ifade edilen; bireyin ikâmet ettiği ülkenin bilgisidir. Beşinci en etkili öznitelik 17 numara ile ifade edilen; kullanıcının daha önce bir tarama uygulaması kullanıp kullanmadığı bilgisidir.

Dilsel kuvvetli sınır-bulanık sınıflayıcı’nın eğitim performansı Şekil 20’de gösterilmiştir.

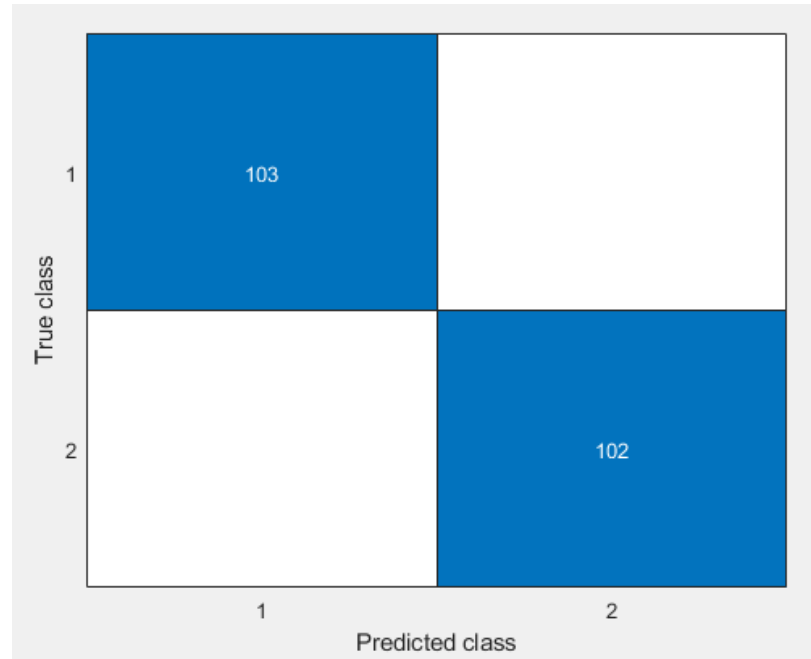
Şekil 20: Dilsel kuvvetli sinir-bulanık sınıflayıcı'nın eğitim performansı



Şeki 20'de görüldüğü gibi dilsel kuvvetli sinir-bulanık sınıflayıcı'nın hata değerleri oldukça düşüktür.

Dilsel kuvvetli sinir-bulanık sınıflayıcı'nın eğitildikten sonra eğitim verisi için çizdirilen hata matrisi Şekil 21'de verilmiştir.

Şekil 21: Eğitim veri için çizdirilen hata matrisi

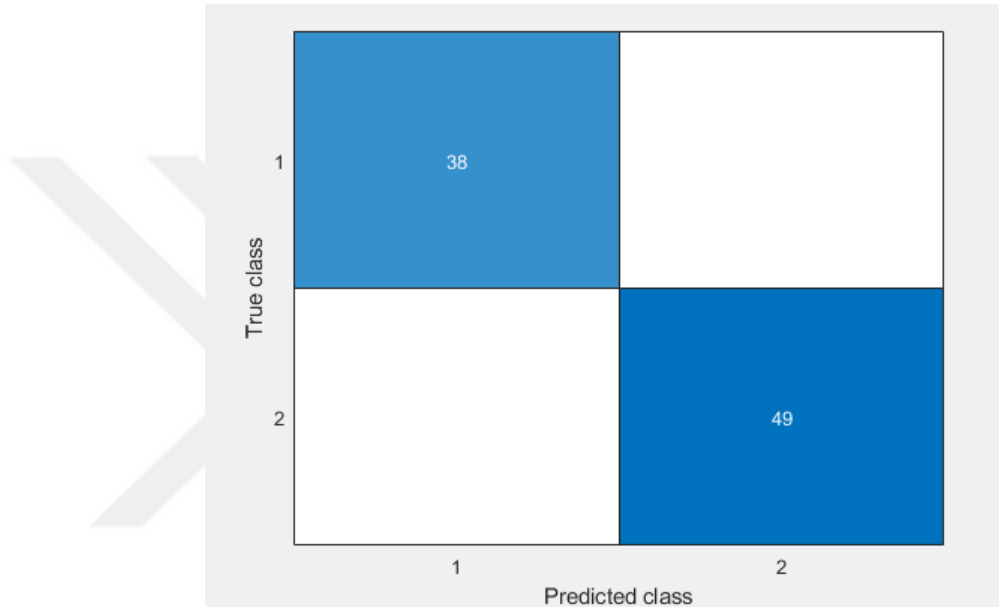


Şekil 21'de görüldüğü gibi eğitilen dilsel kuvvetli sinir-bulanık sınıflayıcı'nın eğitim veri seti için gerçekleştirdiği sınıflandırmanın tamamı doğrudur.

Eđitilen dilsel kuvvetli sinir-bulanık sınıflayıcı'nın, eğitim verileri için gerçekleştirdiđi Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri ise sırası ile 1,1,1,1 şeklindedir. Bu durum eğitim verilerinin tamamının doğru sınıflandırıldığını göstermektedir.

Dilsel kuvvetli sinir-bulanık sınıflayıcı'nın eğitimdikten sonra test verisi için çizdirilen hata matrisi Şekil 22'de verilmiştir.

Şekil 22: Test verisi için çizdirilen hata matrisi



Şekil 22'de görüldüğü gibi eğitimli dilsel kuvvetli sinir-bulanık sınıflayıcı'nın test veri seti için gerçekleştirdiđi sınıflandırmanın tamamı doğrudur.

Eđitilen dilsel kuvvetli sinir-bulanık sınıflayıcı'nın, test verileri için gerçekleştirdiđi Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri ise sırası ile 1,1,1,1 şeklindedir. Bu durum eğitim verilerinin tamamının doğru sınıflandırıldığını göstermektedir.

5.2.3. K-Means İle Elde Edilen Sonuçlar

Kümeleme algoritmaları ön öğreticisiz yöntemlerdir ve verilerin sınıfları önceden belli değildir. Bu sebepten dolayı veriler eğitim ve test verisi olarak ayrılmamış, tamamı eğitim için kullanılmıştır.

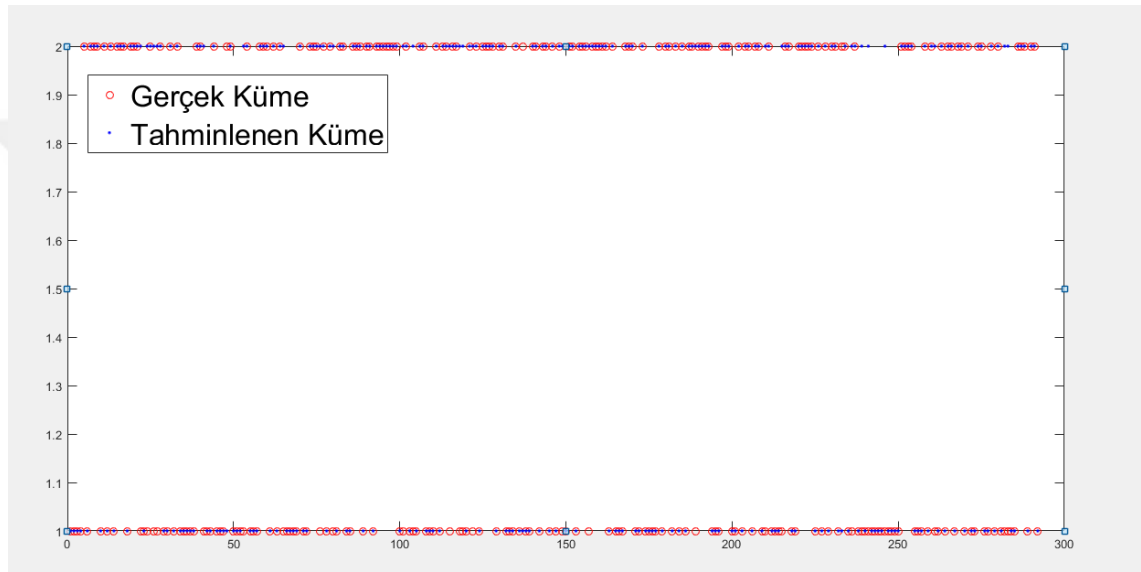
K-means algoritmasında k parametresi ile ifade edilen küme sayısı önceden belirlenmeli ve çalıştırılmadan önce algoritmaya verilmelidir. Bu çalışmadaki verilerin OSB olup

olmadığını gösterecek şekilde iki sınıfı vardır. Dolayısı ile kümeler oluşturulurken iki küme olması istendiğinden k parametresi iki olarak girilmiş ve veriler ikiye kümelendi.

K-means algoritması ile kümelenen verilerin sınıfları, gerçek küme sınıfları ile karşılaştırıldığında toplamda 292 veriden 262'si doğru sınıflandırılmış ve sınıflandırmanın doğruluk oranı % 89,73 olmuştur.

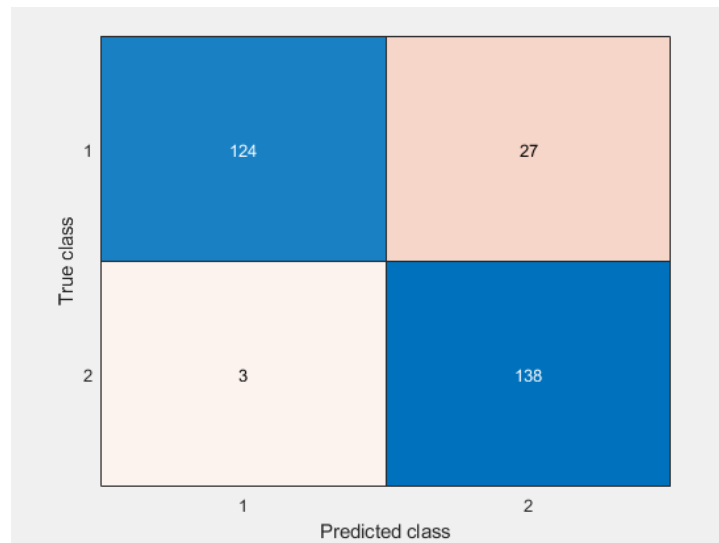
Gerçek sınıflar ve K-means ile elde edilen sınıflar Şekil 23'de verilmiştir.

Şekil 23: Gerçek sınıflar ve K-means ile elde edilen sınıflar



K-means algoritması ile elde edilen sonuçlar için çizdirilen hata matrisi Şekil 24'de verilmiştir.

Şekil 24: K-means algoritması ile elde edilen sonuçlar için çizdirilen hata matrisi



Şekil 24’de görüldüğü gibi K-means algoritması ile kümelenen verilerin sınıfları çoğunlukla doğru tahminlenmiş olup, toplamda yanlış tahminlenen örnek sayısı 30 olmuştur.

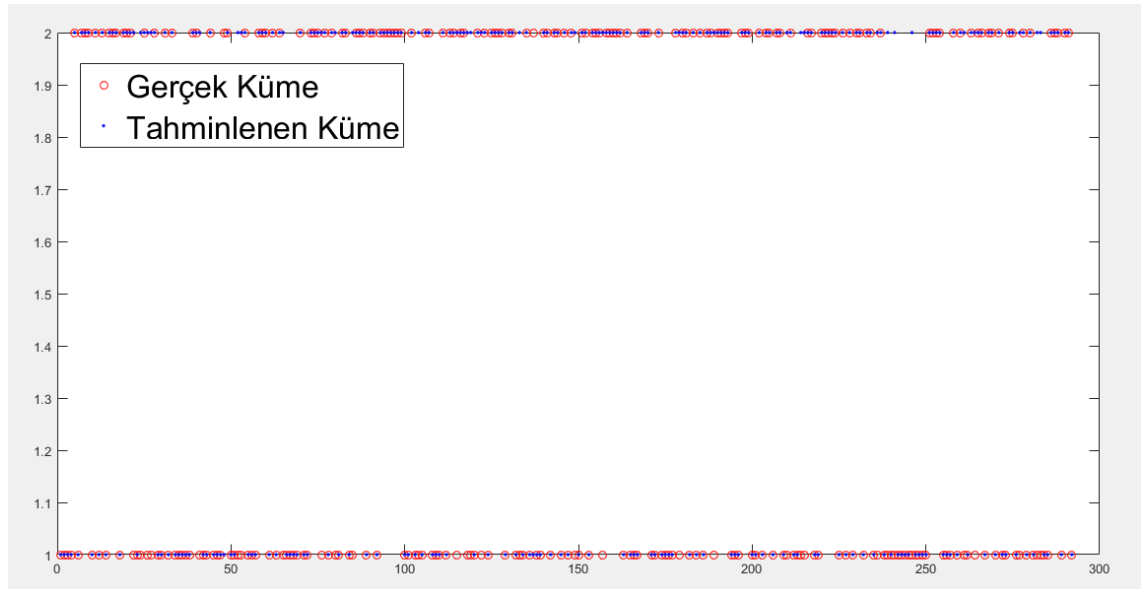
K-means algoritması ile kümelenen verilerin Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri sırası ile 0.897, 0.821, 0.979, 0.892 şeklindedir.

5.2.4. X-Means İle Elde Edilen Sonuçlar

X-means algoritması K-means algoritmasının geliştirilmesiyle oluşturulmuş daha güncel bir kümeleme yöntemidir. X-means, K-means’in çalışma yapısını kullanmasının yanında K-means’in geliştirilmiş halidir. Küme sayısının sabit bir değer olarak belirlenmesinin yerine bir aralık belirlenmesi ile çalışmaktadır. Böylece veri seti için en uygun olan küme sayısı algoritma tarafından belirlenebilmektedir. Küme sayısı belirlenirken, en iyi küme sayısını belirlemek için Bayesin bilgi kriteri kullanılmaktadır. Bu çalışmada, sınıf aralığı 2 ile 4 arasında girilerek çalıştırılan X-means algoritmasında küme sayısı algoritma tarafından 2 olarak belirlenmiştir. X-means algoritması ile ikiye kümelenen verilerin sınıfları, gerçek küme sınıfları ile karşılaştırıldığında toplamda 292 veriden 257’si doğru sınıflandırılmış ve sınıflandırmanın doğruluk oranı %**88,02** olmuştur.

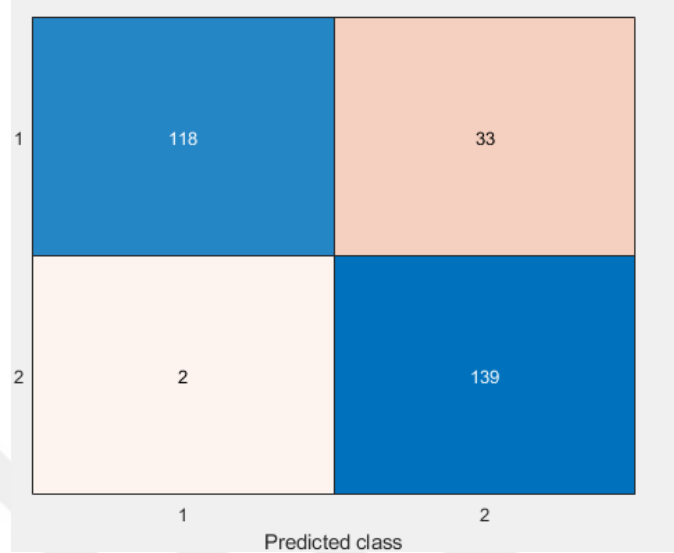
Gerçek sınıflar ve X-means ile elde edilen sınıflar Şekil 25’de verilmiştir.

Şekil 25: Gerçek sınıflar ve X-means ile elde edilen sınıflar



X-means algoritması ile elde edilen sonuçlar için çizdirilen hata matrisi Şekil 26’da verilmiştir.

Şekil 26: X-means algoritması ile elde edilen sonuçlar için çizdirilen hata matrisi



Şekil 26’da görüldüğü gibi X-means algoritması ile kümelenen verilerin sınıfları çoğunlukla doğru tahminlenmiş olup, toplamda yanlış tahminlenen örnek sayısı 35 olmuştur. X-means algoritması ile kümelenen verilerin Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri sırası ile 0.880, 0.781, 0.986, 0.871 şeklindedir.

5.3. Farklı Yöntemler ile Analiz Edilen Çocuklar İçin Otizm Spektrum Bozukluğu Verilerinin Tahminleme Başarılarının Karşılaştırılması

Bu çalışmada OSB veri setinin çocuklar için olan alt kümesi YSA ve DKSBS ile sınıflandırılmış, K-means ve X-means yöntemleri ile kümeleneştir. Kullanılan yöntemlerin performans değerleri Tablo 5’de verilmiştir.

Tablo 5: Kullanılan yöntemlerin performans değerleri

Yöntem	Eğitim setinde doğruluk oranı	Test setinde doğruluk oranı
Yapay sinir ağları	%100	%98.85
Dilsel kuvvetli sinir-bulanık sınıflayıcı	%100	%100
K-Means	%89,73	Tüm veriler eğitim için kullanılmıştır
X-Means	%88,02	Tüm veriler eğitim için kullanılmıştır

Tablo 5’de görüldüğü gibi YSA ile sınıflandırma başarı oranı oldukça yüksek olmasına rağmen en yüksek başarı oranı DKSBS ile elde edilmiştir. DKSBS ile yapılan sınıflandırma da eğitim ve test verisi için %100 oranında doğruluk gerçekleştirilmiştir. Bu durum dilsel kuvvetli sinir-bulanık sınıflayıcının veri setindeki tüm örnekleri doğru sınıflandırdığı anlamını taşımaktadır. Kümeleme yöntemlerinin başarı oranları incelendiği zaman genel anlamda sınıflandırma yöntemlerine göre daha az başarılı olduğu ve K-means yöntemi ile elde edilen sonuçların, X-means yöntemine göre daha başarılı olduğu görülmektedir.

Kullanılan Yöntemlerin Eğitim veri setleri için Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri Tablo 6’da verilmiştir.

Tablo 6: Kullanılan yöntemlerin eğitim veri setleri için Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri

Yöntem	Doğruluk	Duyarlılık	Belirleyicilik	F-Ölçütü
Yapay sinir ağları	1	1	1	1
Dilsel kuvvetli sinir-bulanık sınıflayıcı	1	1	1	1
K-Means	0.897	0.821	0.979	0.892
X-Means	0.880	0.781	0.986	0.871

Tablo 6’da sınıflandırma ve kümeleme yöntemlerinin eğitim verisi için hesaplanan Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri görülmektedir. Görüldüğü üzere sınıflandırma yöntemlerinde hesaplanan tüm değerler 1’dir. Kümeleme yöntemlerinde ise ortalama olarak K-means için hesaplanan değerler X-means’ten daha yüksektir.

Kullanılan Yöntemlerin test veri setleri için Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri Tablo 7’de verilmiştir.

Tablo 7: Kullanılan yöntemlerin test veri setleri için Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri

Yöntem	Doğruluk	Duyarlılık	Belirleyicilik	F-Ölçütü
Yapay sinir ağları	0.989	0.974	1	0.987
Dilsel kuvvetli sinir-bulanık sınıflayıcı	1	1	1	1
K-Means	Tüm veriler eğitim için kullanılmıştır	Tüm veriler eğitim için kullanılmıştır	Tüm veriler eğitim için kullanılmıştır	Tüm veriler eğitim için kullanılmıştır
X-Means	Tüm veriler eğitim için kullanılmıştır	Tüm veriler eğitim için kullanılmıştır	Tüm veriler eğitim için kullanılmıştır	Tüm veriler eğitim için kullanılmıştır

Tablo 7’de görüldüğü gibi test verisi için DKSBS ile yapılan sınıflandırmanın Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri 1 olarak hesaplanmıştır. Bu durum dilsel kuvvetli sinir-bulanık sınıflayıcının test verisinin tamamını doğru sınıflandırdığı anlamı taşımaktadır. Test verisi için YSA ile yapılan sınıflandırmanın Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri de 1’e oldukça yakın değerler çıkmıştır. Bu durum test verisi için dilsel kuvvetli sinir-bulanık sınıflayıcıdan daha başarısız olmakla birlikte kullanılabilir bir yöntem olduğunu ortaya koymaktadır. K-Means ve X-Means ile kümelenen verilerin tamamı eğitim için kullanılmış, dolayısı ile test verisi için hesaplanan Doğruluk, Duyarlılık, Belirleyicilik ve F-Ölçütü değerleri bulunmamaktadır. Çalışmanın performans değerleri ve aynı veri setini kullanan diğer çalışmalardaki performans değerleri Tablo 8’de verilmiştir.

Tablo 8: Çalışmanın performans değerleri ve aynı veri setini kullanan diğer çalışmalardaki performans değerleri

Çalışma	Yöntem	Performans
Thabtah, 2018	Logistic regression	%97.94
Thabtah, 2018	Naive bayes	%92.80
Akyol ve Karaci, 2018	Lojistik regresyon	%92.00
Akyol ve Karaci, 2018	Bulanık kuralı lojistik regresyon kombinasyonu	%97.33

De Campos Souza ve Guimaraes, 2018	Bulanık sinir ağı mimarisi	%100
Hutchinson vd., 2019	J48 decision tree	%92.80
Hutchinson vd., 2019	Naive bayes	%96.20
Bu çalışma	Yapay sinir ağları	%98.85
Bu çalışma	Dilsel kuvvetli sinir-bulanık sınıflayıcı	%100
Bu çalışma	K-Means	%89,73
Bu çalışma	X-Means	%88,02

Tablo 8’de görüldüğü gibi aynı veri seti farklı çalışmalarda farklı yöntemlerle sınıflandırılmıştır ama bu çalışma dışında bir kümeleme işlemine rastlanmamıştır. Yapılan çalışmalardaki sonuçlar incelendiği zaman genel olarak başarı oranının %90’ın üzerinde olduğu görülmekle birlikte %100 olan bir çalışma da bulunmaktadır. Bu çalışma ile elde edilen sonuçlar ile diğer çalışmalarda elde edilen sonuçlar karşılaştırıldığı zaman ise verilerin DKSBS ve YSA ile sınıflandırıldığında sınıflandırma başarısının logistic regression, naive bayes, bulanık kuralı lojistik regresyon kombinasyonu ve j48 decision tree ile sınıflandırılan çalışmalardaki sınıflandırma başarısından daha yüksek olduğu görülmektedir. Bulanık sinir ağı mimarisi ile gerçekleştirilen sınıflandırma ve bu çalışma kapsamın da DKSBS ile gerçekleştirilen sınıflandırmanın başarı oranları ise aynıdır. İki yöntem ile sınıflandırılan verilerdeki doğruluk aynı orandadır ve bu oran %100 şeklindedir. Bu durum iki yöntemde tüm örnekleri doğru sınıflandırdığı anlamına gelmektedir. Bu çalışma kapsamında gerçekleştirilen kümeleme işlemi sonuçları incelendiği zaman ise kümeleme yöntemlerinden K-means yönteminin X-means yöntemine göre daha başarılı olduğu ve iki kümeleme yönteminin de başarı oranının sınıflandırma yöntemlerine göre daha düşük olduğu sonucuna ulaşılmaktadır.

6. SONUÇ ve ÖNERİLER

Bu çalışma ile OSB konusunda veri madenciliği, yapay zeka gibi teknikleri kullanarak gerçekleştirilen çalışmalar incelenmiştir. Bu konuda oldukça başarılı sonuçlara ulaşan birçok çalışmanın olduğu görülmektedir. Bunun yanı sıra bilimsel bir yayın ile son dönemde gerçekleştirilen çalışmalarda güncelliğini yitiren veri setlerinin kullanıldığı eleştirisi dikkat çekmektedir. Yayının detayları incelendiği zaman güncel ve başarılı yöntemlerin kullanılarak teknolojik gelişmelerden faydalandığı fakat deneysel çalışmalarda kullanılan verilerin güncelliğini yitirdiği vurgusu yapıldığı görülmektedir. Çalışmanın devamında ise eksiklik olarak görülen güncel olmayan veriler, geliştirilen son ölçeğe dayalı olarak toplanmış ve bir takım işlemler gerçekleştirildikten sonra paylaşılmıştır. Veri seti çocuk, genç, yetişkin şeklinde gruplandırılmak üzere toplamda 1100 örnek içermektedir. Bu çalışmada ise aynı veri setinin 292 örnek içeren çocuklar için olan alt küme kullanılmıştır. Toplamda 20 girdiden oluşan fakat tüm örneklerde aynı olduğu için çıkarılan bir öznelikten sonra 19'a düşürülen girdiler sonucunda bireyin otizmlili olma durumunu ifade eden çıktının tahminlenmesi için veri setine veri madenciliği yöntemlerinden olan sınıflandırma ve kümeleme işlemleri uygulanmıştır. Sınıflandırma ve kümeleme işlemleri uygulanmadan önce öncelikle tamamen numeric değerlere dönüştürülen veri seti üzerinde özneliklerin frekansına bakılarak eksik veriler tamamlanmış ve normalizasyon işlemleri gerçekleştirilmiştir. Yöntemlerin çalıştırılması için hazır hale getirilen veri seti son olarak %70 eğitim, %30 test verisi olmak üzere ayrılmıştır. Sınıflandırma işlemi için veriler test ve eğitim verisi olacak şekilde ayrılarak sınıflandırılmış, elde edilen sonuçlar birçok parametre ile ifade edilmiştir. Sınıflandırma işlemi için kullanılan yöntemler YSA ve DKSBS olmuştur. Kümeleme yöntemleri önöğreticisiz yöntemler olduğundan dolayı veri seti eğitim ve test olarak ayrılmamış tamamı eğitim verisi olacak şekilde kümeleneştir. Kümeleme işlemi için K-means ve X-means yöntemleri kullanılmıştır. Elde edilen sonuçlar birçok parametre ile ifade edilmiş olmakla birlikte tüm yöntemlerde ortak paydada karşılaştırma yapılabilmesi için temel olarak Doğruluk, Duyarlılık, Belirleyicilik, F-Ölçütü kullanılmıştır.

Sınıflandırma sonuçlarına bakıldığı zaman eğitim veri seti için YSA ve DKSBS olmak üzere iki yönteminde başarı oranının %100 olduğu görülmüştür. Dolayısı ile iki yöntemin eğitim veri seti için hesaplanan Doğruluk, Duyarlılık, Belirleyicilik, F-Ölçütü

değerleri 1 olarak hesaplanmıştır. Bu durum iki yönteminde eğitim verilerinin tamamını doğru sınıflandırdığı anlamına gelmektedir. Test seti için elde edilen sonuçlara bakıldığı zaman başarı oranlarının birbirinden farklı olduğu görülmektedir. Dilsel kuvvetli sinir-bulanık sınıflayıcı'nın test verisi için gerçekleştirdiği başarı oranı eğitim verisindeki başarı oranı ile aynı olmuştur. Dolayısı ile test seti için hesaplanan Doğruluk, Duyarlılık, Belirleyicilik, F-Ölçütü değerleri 1 olarak hesaplanmıştır. Bu durum test verilerinin DKSBS tarafından tamamının doğru sınıflandırdığı anlamına gelmektedir. YSA'nın test verisi için gerçekleştirdiği başarı oranı ise %98.85 olmuştur. Test seti için hesaplanan Doğruluk, Duyarlılık, Belirleyicilik, F-Ölçütü değerleri sırası ile 0.989, 0.974, 1, 0.987 olarak hesaplanmıştır. Bu sonuçlar YSA ile yapılan sınıflandırmanın test verisinde 87 örnekten yalnızca 1 tanesini yanlış sınıflandırdığı geriye kalan 86 örneği doğru sınıflandırdığını anlamını taşımaktadır. Sınıflandırma yöntemlerinin başarılarına bakıldığı zaman iki yönteminde oldukça başarılı olduğu ve veri setini sınıflandırmak için kullanılabileceği görülmekle birlikte, tüm verileri doğru sınıflandırdığından dolayı dilsel kuvvetli sinir-bulanık sınıflayıcının daha başarılı olduğu söylenebilmektedir.

Kümeleme sonuçlarına bakıldığı zaman eğitim ve test verisi olarak ayrılmadığı tüm verilerin eğitim için kullanıldığı görülmektedir. Dolayısı ile K-means ve X-means yöntemi ile kümelenen verilerde yalnızca eğitim verisi için sonuçlar bulunmaktadır. K-means ile kümelenen verilerin başarı oranı %89,73 olmuştur ve verilerin hesaplanan Doğruluk, Duyarlılık, Belirleyicilik, F-Ölçütü değerleri sırası ile 0.897, 0.821, 0.979, 0.892 şeklindedir. Bu sonuçlar K-means ile kümelenen verilerin 262 tanesinin doğru, 30 tanesinin yanlış sınıflandırıldığı anlamına gelmektedir. X-means ile kümelenen verilerin başarı oranı ise %88,02 olmuştur ve verilerin hesaplanan Doğruluk, Duyarlılık, Belirleyicilik, F-Ölçütü değerleri sırası ile 0.880, 0.781, 0.986, 0.871 şeklindedir. Bu sonuçlar X-means ile kümelenen verilerin 257 tanesinin doğru, 35 tanesinin yanlış sınıflandırıldığı anlamına gelmektedir. Bu çalışma kapsamında elde edilen sonuçlara göre kümeleme yöntemlerinin başarılarına bakıldığı zaman sonuçların genel anlamda yüksek olduğu fakat sınıflandırma yöntemlerine göre daha düşük kaldığı görülmekle birlikte bu çalışmadaki veri seti için K-means yönteminin X-means yöntemine göre daha başarılı sonuçlar verdiği tespit edilmiştir.

Bu çalışmada elde edilen sonuçlar ile aynı veri seti ile yapılan diğer çalışmalardaki sonuçlar karşılaştırıldığı zaman Thabtah (2018), logistic regression ile yaptığı

sınıflandırmanın başarı oranı %97.94, naive bayes ile yaptığı sınıflandırmanın başarı oranı %92.80 bulmuştur. Akyol ve Karaci (2018), lojistik regresyon ile yaptıkları sınıflandırmanın başarı oranı %92.00, bulanık kurallı lojistik regresyon kombinasyonu ile yaptıkları sınıflandırmanın başarı oranı %97.33 olarak elde etmişlerdir. De Campos Souza ve Guimaraes (2018), bulanık sinir ağı mimarisi ile yaptıkları sınıflandırmanın başarı oranı %100 olarak tespit etmişlerdir. Hutchinson ve arkadaşları (2019), j48 decision tree ile yaptıkları sınıflandırmanın başarı oranını %92.80, naive bayes ile yaptıkları sınıflandırmanın başarı oranını %96.20 olarak bulmuşlardır. Bu durum genel olarak veri setinin sınıflandırmaya uygun bir veri seti olduğunu göstermektedir. Aynı zamanda bu çalışma ile %100 oranında başarıya ulaşan DKSBS yönteminin kullanılan veri seti için literatürdeki birçok yöntemden daha başarılı sonuçlar elde ettiğini de göstermektedir.

Bu çalışmanın literatüre olan katkıları ise şöyledir; öncelikle aynı veri seti ile yapılan çalışmalar incelendiği zaman logistic regression, naive bayes, bulanık kurallı lojistik regresyon kombinasyonu, bulanık sinir ağı mimarisi, j48 decision tree gibi yöntemlerin kullanılması ile yapılan sınıflandırma işlemleri olmakla birlikte çocuklar için olan alt kümede herhangi bir yöntem ile kümeleme işleminin gerçekleştirildiği bir çalışmaya rastlanmamış olmakla birlikte, YSA ve DKSBS ile yapılan bir sınıflandırmaya da rastlanmamıştır. Sonuçlara bakıldığı zaman tahminleme doğruluğu bakımından OSB'nin çocuklar için olan verilerinde, sınıflandırma yöntemlerinin kümeleme yöntemlerinden daha başarılı olduğu görülmektedir. Özellikle DKSBS yönteminin tüm verileri doğru sınıflandırarak literatürdeki birçok yöntemden daha yüksek başarı oranına sahip olduğu için tercih edilebilecek yöntemler arasında en iyi yöntemlerden bir tanesi olabileceği sonucuna varılmaktadır. Dolayısı ile bu çalışma ile kullanılan veri seti için farklı yöntemlerin denenmesine imkân tanınarak literatürdeki birçok çalışmadan daha başarılı sonuçlar elde ettiği görülen bir yöntem ile sınıflandırılma yapılmasına imkan tanımaktadır. Bunun yanı sıra elde edilen sonuçların değerlendirilmesi için kullanılan parametreler birçok çalışmadan daha detaylı olarak verilmiştir. Bu durum elde edilen sonuçların daha detaylı yorumlanmasına imkan tanımaktadır.

İleriki çalışmalarda ise veri setinin farklı alt kümeleri kullanılarak veya tüm alt kümeleri aynı anda kullanılarak bir takım analizler gerçekleştirilebilir. Yapılacak çalışmalarda karar destek sistemleri, uzman sistemler, görüntü işleme teknikleri gibi daha farklı

teknolojiler ile desteklenen daha geniş kapsamlı veri setlerine uygulanan çalışmalar gerçekleştirilebilir. Verilerin yeniden güncel olarak toplanması için alanında uzman hekim veya hekimler ile birlikte ortak olarak yeni çalışmalar yürütülebilir. Devamında ise elde edilen bulgular ile sürekli öğrenmeye devam eden ve normal hayatta da kullanılabilen somut ürünler ortaya konulabilir.



KAYNAKÇA

- Akal, Ş. (2016), “Sağlık Alanında Akıllı Teknolojilerin Kullanımı İçin Bir Uygulama Önerisi”, *Uluslararası Yönetim Bilişim Sistemleri Konferansı*, 3, (291-297).
- Akçapınar, G., Altun, A. ve Aşkar, P. (2016), “Çevrimiçi Öğrenme Ortamındaki Benzer Öğrenci Gruplarının Kümeleme Yöntemi İle Belirlenmesi”, *Eğitim Teknolojisi Kuram ve Uygulama*, C.6, S.2, (46-64). DOI: 10.17943/etku.91440.
- Akdemir, Ç. (2016), “Hilenin Veri Madenciliği İle Ortaya Çıkarılması ve Perakende Sektöründe Bir Uygulama”, Yayınlanmamış Doktora Tezi, İstanbul: Marmara Üniversitesi Sosyal Bilimler Enstitüsü.
- Akgöbek, Ö., ve Çakır, F. (2009), “Veri Madenciliğinde Bir Uzman Sistem Tasarımı”, *Akademik Bilişim Konferansı*, 9, (801-806).
- Akpınar, H. (2000), “Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği. *İÜ İşletme Fakültesi Dergisi*, C.29, S.1, (1-22).
- Akkaya, G. C., Demireli, E., Yakut, Ü. H., ve Yakut, H. (2009), “İşletmelerde Finansal Başarısızlık Tahminlemesi: Yapay Sinir Ağları Modeli İle İmkb Üzerine Bir Uygulama”, *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, C.10, S.2, (187-216).
- Akman, B. (2018), “Okul Öncesi Öğretmenlerinin Otizmli Çocukların Kaynaştırılması Hakkındaki Düşüncelerinin İncelenmesi”, *Ozel Eğitim Dergisi*, C.19, S.1, (105-128).
- Aktaş, R. (2003), “Mali Başarısızlığın Öngörülmesi: İstatistiksel Yöntemler ve Yapay Sinir Ağı Karşılaştırılması”, *Ankara Üniversitesi SBF Dergisi*, C.58, S4. (1-24).
- Akyılmaz, O., ve Ayan, T. (2010), “Esnek Hesaplama Yöntemlerinin Jeodezide Uygulamaları”, *İTÜ Dergisi*, C.5, S.1, (261-268).
- Akyol, K., and Karaci, A. (2018), “A Study on Autistic Spectrum Disorder For Children Based On Feature Selection And Fuzzy Rule”, *International Congress on Engineering and Life Science*, (804–807).
- Albayrak, A. S. ve Koltan Yılmaz, Ş. (2009), “Veri Madenciliği: Karar Ağacı Algoritmaları ve İMKB Verileri Üzerine Bir Uygulama”, *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, C.14, S.1, (31-52).

Altıntaş, Y. Y. (2010), “Veri Madenciliğinin Tıpta Kullanımı ve Bir Uygulama: Hemodiyaliz Hastaları İçin Risk Seviyelerine Göre Risk Faktörlerinin Etkileşimlerinin İncelemesi”, Yayınlanmamış Yüksek Lisans Tezi. Ankara: Gazi Üniversitesi Fen Bilimleri Enstitüsü.

Aydın, S. (2007), “Veri Madenciliği ve Anadolu Üniversitesi Uzaktan Eğitim Sisteminde Bir Uygulama”, Yayınlanmamış Doktora Tezi. Eskişehir: Anadolu Üniversitesi Sosyal Bilimler Enstitüsü.

Ayık, Y. Z., Özdemir, A., ve Yavuz, U. (2007), “Lise Türü ve Lise Mezuniyet Başarısının, Kazanılan Fakülte İle İlişkisinin Veri Madenciliği Tekniği İle Analizi”, *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, C.10, S.2, (441-454).

Baykal, A. (2006), “Veri Madenciliği Uygulama Alanları”, *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, S.7, (95-107).

Bholowalia, P., and Kumar, A. (2014), “EBK-Means: A Clustering Technique Based On Elbow Method And K-Means In WSN”, *International Journal of Computer Applications*, Vol.105, Issue.9, (17-24).

Bilgin, T. T., ve Çamurcu, Y. (2005), “Dbscan, Optics ve K-Means Kümeleme Algoritmalarının Uygulamalı Karşılaştırılması”, *Politeknik Dergisi*, C.8, S.2, (139-145).

Brossette, S. E., Sprague, A. P., Hardin, J. M., Waites, K. B., Jones, W. T., and Moser, S. A. (1998), “Association Rules And Data Mining İn Hospital İnfection Control And Public Health Surveillance”, *Journal of the American medical informatics association*, Vol.5, Issue.4, (373-381).

Bradley, P. S., Fayyad, U. M., and Reina, C. (1998), “Scaling Clustering Algorithms to Large Databases”, *KDD*, Vol. 98, (9-15).

Budak, E. Ç., ve Bozkurt, M. R. (2013), “Vertebra Lomber Disklerde Meydana Gelen Bozulmaların Manyetik Rezonans Görüntüleme (Mrg) ile Analizi”, *AJIT-e: Online Academic Journal of Information Technology*, C.4, S.11, (125-144).

Coulter, D. M., Bate, A., Meyboom, R. H., Lindquist, M., and Edwards, I. R. (2001), “Antipsychotic Drugs and Heart Muscle Disorder in International Pharmacovigilance: Data Mining Study”, *Bmj*, Vol.322, Issue.7296, (1207-1209).

- Cai, Z., and Liu, J. (2018), "Approximating Quantum Many-Body Wave Functions Using Artificial Neural Networks", *Physical Review B*, Vol.97, Issue.3.
- Calp, M. H. (2019), "İşletmeler İçin Personel Yemek Talep Miktarının Yapay Sinir Ağları Kullanılarak Tahmin Edilmesi" *Politeknik Dergisi*, C.22, S.3, (675-686).
- Carleo, G., and Troyer, M. (2017), "Solving The Quantum Many-Body Problem with Artificial Neural Networks", *Science*, Vol.355, Issue.6325, (602-606).
- Chine, W., Mellit, A., Lughi, V., Malek, A., Sulligoi, G., and Pavan, A. M. (2016), "A Novel Fault Diagnosis Technique for Photovoltaic Systems Based on Artificial Neural Networks", *Renewable Energy*, vol.90, (501-512).
- Çağlar, M. F., Çetişli, B., ve Toprak, İ. B. (2010), "Automatic Recognition of Parkinson's Disease From Sustained Phonation Tests Using ANN and Adaptive Neuro-Fuzzy Classifier", *Mühendislik Bilimleri ve Tasarım Dergisi*, C.1, S.2, (59-64).
- Çalışkan, S. K., ve Soğukpınar, İ. (2008), "Kxknn: K-Means ve K En Yakın Komşu Yöntemleri İle Ağlarda Nüfuz Tespiti", *EMO Yayınları*, (120-24).
- Çam, S., Sigeze, Ç. S., ve Ballı, E. (2018), "Türkiye'nin Enerji Verimliliğinin Yapay Sinir Ağı ve ARDL Yaklaşımı ile Analizi", *Ege Akademik Bakış Dergisi*, C.18, S.4, (661-670).
- Çetinkaya Bozkurt, Ö., Kalkan, A. ve Şişeci Çeşmeli, M. (2016), "Karar Destek Sistemlerinin İşletme Yönetimi Açısından Önemi: Mermer İşletmelerinde Bir Araştırma", *Uluslararası Yönetim Bilişim Sistemleri Konferansı*, (3), (153-162).
- Çetişli B. (2006), "Öznelik Seçiminde Dilsel Kuvvetli Sinir Bulanık Sınıflayıcı Kullanımı", *Eskişehir Osmangazi Üniversitesi Mühendislik ve Mimarlık Fakültesi Dergisi*, C.19, S.2, (109-130).
- Çınaroğlu, S., and Bulut, H. (2018), "New Initialization Approaches for The K-Means And Particle Swarm Optimization Based Clustering Algorithms", *Journal of the Faculty of Engineering and Architecture of Gazi University*, Vol.33, Issue.2, (413-423).
- Çelik, S., Bozkurt, Ö. Ç., ve Çeşmeli, M. Ş. (2018), "İnsan Omurgası Veri Setinin Sinir-Bulanık Sınıflayıcı İle Öznelik Tespiti ve Sınıflandırılması", *Yönetim Bilişim Sistemleri Dergisi*, C.4, S.1, (39-52).

Cetiřli, B., ve Kalkan, H. (2011), "Classification of Multispectral Satallite Images By Using Adaptive Neuro-Fuzzy Classifier with Linguistic Hedges", *Signal Processing and Communications Applications Conference*, 19, (50-53).

Deb, C., Lee, S. E., ve Santamouris, M. (2018), "Using Artificial Neural Networks to Assess HVAC Related Energy Saving in Retrofitted Office Buildings", *Solar Energy*, Vol.163, (32-44).

Demirciođlu, H. Z. ve Bilge, H. ř. (2015), "Yumurtalık Kanseri Veri Kumesindeki Gen İfadelerinin Veri Madenciliđi ile Analizi", *Marmara Fen Bilimleri Dergisi*, C.27, S.4, (125-134).

Demiralay, M., ve amurcu, A. Y. (2005), "Cure, Agnes ve K-Means Algoritmalarındaki Kumeleme Yeteneklerinin Karřılařtırılması", *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, C.4, S.8, (1-18).

Demirhan, A. (2018), "Makine Öğrenmesi Yöntemlerinin Otizm Spektrum Bozukluđu Olgularının Belirlenmesindeki Başarımı", *Mugla Journal of Science and Technology*, C.4, S.1, (79-84).

De Campos Souza, P. V., and Guimaraes, A. J. (2018), "Using Fuzzy Neural Networks for Improving the Prediction of Children with Autism Through Mobile Devices", *Symposium on Computers and Communications* (01086-01089).

Diler, S. (2016), "Veri Madenciliđi Süreçleri ve Karar Ađaçları Algoritmaları ile Bir Uygulama", Yayınlanmamıř Yüksek Lisans Tezi. Van: Yüzüncü Yıl Üniversitesi Fen Bilimleri Enstitüsü.

Dođan, B., Buldu, A., Demir, Ö., ve Erol, B. (2018), "Sigortacılık Sektöründe Müřteri İliřki Yönetimi İçin Kumeleme Analizi", *Karaelmas Fen ve Mühendislik Dergisi*, C.8, S.1, (11-18).

Feng, Y., ve Hamerly, G. (2007), "PG-Means: Learning The Number of Clusters in Data", In *Advances in neural information processing systems*, (393-400).

Ferreira, D., Oliveira, A., and Freitas, A. (2012), "Applying Data Mining Techniques to Improve Diagnosis in Neonatal Jaundice", *BMC Medical Informatics And Decision Making*, Vol.12, Issue.1, 143, (1-6).

- Fırat, M., ve Güngör, M. (2004), “Askı Madde Konsantrasyonu ve Miktarının Yapay Sinir Ağları ile Belirlenmesi”, *Teknik Dergi*, C.15, S.73, (3267-3282).
- Fırat, M., Yurdusev, M. A., ve Mermer, M. (2008), “Uyarlamalı Sinirsel Bulanık Mantık Yaklaşımı İle Aylık Su Tüketiminin Tahmini”, *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, C.23, S.2, (449 – 457).
- Gharib, T. F., Nassar, H., Taha, M., and Abraham, A. (2010), “An Efficient Algorithm For Incremental Mining of Temporal Association Rules”, *Data & Knowledge Engineering*, Vol.69, Issue.8, (800-815).
- Göreke, V., Uzunhisarcıklı, E., ve Güven, A. (2014), “Gri Seviyeli Eşoluşum Matrisleri Kullanılarak Sayısal Mamogram Görüntüsünden Doku Özneliklerinin Çıkarılması ve Yapay Sinir Ağı İle Kitle Tespiti”, *Tıp Teknolojileri Ulusal Kongresi-TıpTekno'14*, (95-98).
- Gülsöz, T., ve Çıkkılı, Y. (2018), “Otizm Spektrum Bozukluğu Olan Öğrencilere Soğuk İçecek Hazırlama ve Sunma Becerisinin Video Model ile Öğretimin Etkililiği”, *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, C.18, S.1, (210-229).
- Güner, Ö. Ü. S., Codal, Ö. Ü. K. S., Geçer, H. S., ve Coşkun, E. (2018), “Trafik Kaza Desenlerinin Tanımlanmasında K-Means Kümeleme Algoritmasının Kullanılması: Sakarya İli Uygulaması”, *İşletme Bilimi Dergisi*, C.6, S.3, (89-105).
- Hamerly, G., ve Elkan, C. (2004), “Learning The K in K-Means. In *Advances in Neural Information Processing Systems*, (281-288).
- Hong, T. P., and Wu, C. W. (2011), “Mining Rules From an Incomplete Dataset with a High Missing Rate”, *Expert Systems with Applications*, Vol.38, Issue.4, (3931-3936).
- Huang, D., and Chow, T. W. (2005), “Efficiently Searching The Important Input Variables Using Bayesian Discriminant”, *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol.52, Issue.4, (8785-793).
- Hutchinson, J., Schauer, I., and Seetan, R. (2019), “A Comparative Study Of Data Mining Techniques Used To Test Predictive Accuracy Of Autism Spectrum Disorder Screening Process”, In *34th Annual Conference of The Pennsylvania Association of Computer and Information Science Educators*, (70-75).

Ihianle, I. K., Naeem, U., and Tawil, A. R. (2016), "Recognition of Activities of Daily Living From Topic Model", *Procedia Computer Science*, Vol.98, (24-31).

Işık, M., ve Çamurcu, A. Y. (2007), "K-Means, K-Medoids ve Bulanık C-Means Algoritmalarının Uygulamalı Olarak Performanslarının Tespiti", *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi* C.6, S.13, (31-45).

Işık M., ve Çamurcu, A. Y. (2011), "Document Clustering Using K-Means and Hyperspherical Fuzzy C-Means Algorithms", *Marmara Fen Bilimleri Dergisi*, C.22, S.1, (1-18).

Jain, A. K. (2010), "Data Clustering: 50 Years Beyond K-Means", *Pattern Recognition Letters*, C.31, S.8, (651-666).

Jayalakshmi, T., ve Santhakumaran, A. (2011), "Statistical Normalization and Back Propagation for Classification", *International Journal of Computer Theory and Engineering*, C.3, S.1, (1793-8201).

Jiao, Y., Chen, R., Ke, X., Cheng, L., Chu, K., Lu, Z., and Herskovits, E. H. (2011), "Predictive Models for Subtypes of Autism Spectrum Disorder Based on Single-Nucleotide Polymorphisms and Magnetic Resonance Imaging. *Advances in Medical Sciences*, Vol.56, Issue.2, (334-342).

Jiao, Y., Chen, R., Ke, X., Chu, K., Lu, Z., and Herskovits, E. H. (2010), "Predictive Models of Autism Spectrum Disorder Based on Brain Regional Cortical Thickness. *Neuroimage*, Vol.50, Issue.2, (589-599).

Kalantar, B., Pradhan, B., Naghibi, S. A., Motevalli, A., and Mansor, S. (2018), "Assessment of The Effects of Training Data Selection on The Landslide Susceptibility Mapping: A Comparison Between Support Vector Machine (SVM), Logistic Regression (LR) and Artificial Neural Networks (ANN)", *Geomatics, Natural Hazards and Risk*, Vol.9, Issue.1, (49-69).

Kalogeratos, A., and Likas, A. (2012), "Dip-Means: an Incremental Clustering Method For Estimating The Number Of Clusters", *In Advances in Neural Information Processing Systems*, (2393-2401).

- Kaynar, O., ve Taştan, S. (2009). “Zaman Serisi analizinde Mlp Yapay Sinir Ağları ve Arıma Modelinin Karşılaştırılması” *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, S.33, 161-172.
- Khaled, A., Sanjay, R., and Vineet, S. (1998), “An Efficient K-Means Clustering Algorithm. *In IPPS: 11th International Parallel Processing Symposium*.
- Koçak, Y. (2016), “Veri Madenciliği Tekniklerini Kullanarak Sosyal Ağ Tabanlı Sınıflandırıcı Geliştirilmesi”, Yayımlanmamış Yüksek Lisans Tezi. Ankara: Tobb Ekonomi ve Teknoloji Üniversitesi Fen Bilimleri Enstitüsü.
- Küçükkocaoğlu, G., Benli Keskin, Y., ve Küçüksözen, C., (2005), “Finansal Bilgi Manipülasyonunun Tespitinde Yapay Sinir Ağı Modelinin Kullanımı”, *İMKB Dergisi*, C.9, S.36, (1-23).
- Koçyiğit, Y., ve Korürek, M. (2010), “EMG İşaretlerini Dalgacık Dönüşümü ve Bulanık Mantık Sınıflayıcı Kullanarak Sınıflama”, *İTÜDERGİSİ/d*, C.4, S.3, (25-31).
- Kılınç, G. E., ve Söğüt, M. Ü. (2018), “Mikrobiyotaya Güncel Bir Yaklaşım: Otizm ve Mikrobiyota”, *Turkiye Klinikleri Journal of Health Sciences*, C.3, S.1, (88-94).
- Koyuncuğil, A. S. ve Özgülbaş, N. (2009), “Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları. *Bilişim Teknolojileri Dergisi*, C.2, S.2, (21-32).
- Küçükkocaoğlu, G., Benli, Y. K., ve Küçüksözen, C. (2007), “Finansal Bilgi Manipülasyonunun Tespitinde Yapay Sinir Ağı Modelinin Kullanımı”, *İMKB Dergisi*, C.9, S.36, (1-30).
- Latkowski, T., and Osowski, S. (2015), “Computerized System For Recognition of Autism on The Basis of Gene Expression Microarray Data” *Computers in Biology and Medicine*, vol.56, (82-88).
- Levent, E. B. (2016), “Veri Madenciliği ve Havacılık Sektöründe Bir Uygulama”, Yayımlanmamış Yüksek Lisans Tezi. İstanbul: Yıldız Teknik Üniversitesi Sosyal Bilimler Enstitüsü.
- Liu, H., Dougherty, E. R., Dy, J. G., Torkkola, K., Tuv, E., Peng, H., ... and Zhao, Z. (2005), “Evolving Feature Selection”, *IEEE Intelligent systems*, C.20, S.6, (64-76).

- Lu, H., Chen, J., Yan, K., Jin, Q., Xue, Y., and Gao, Z. (2017), “A Hybrid Feature Selection Algorithm For Gene Expression Data Classification. *Neurocomputing*”, Vol.256, (56-62).
- Mackinnon, M. J., and Glick, N. (1999), “Applications: Data Mining and Knowledge Discovery in Databases—an Overview”, *Australian & New Zealand Journal of Statistics*, Vol.41, S.3, (255-275).
- MacQueen, J. (1967), “Some Methods for Classification and Analysis of Multivariate Observations”, *In Proceedings of The Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol.1, Issue.14, (281-297).
- Mehmet, D., ve Akpınar, E. (2018), “Sabit ve Hareketli Hava İle Isıtılan Güneş Kolektörlerinin Isıl Verimliliklerinin Karşılaştırılması ve Yapay Sinir Ağları İle Modellenmesi”, *Fırat Üniversitesi Mühendislik Bilimleri Dergisi*, C.30, S.1, (41-46).
- Mirarchi, D., Petrolo, C., Canino, G., Vizza, P., Cuomo, S., Chiarella, G., and Veltri, P. (2016), “Applying Mining Techniques to Analyze Vestibular Data”, *Procedia Computer Science*, Vol.98, (467-472).
- Moradi, E., Khundrakpam, B., Lewis, J. D., Evans, A. C., and Tohka, J. (2017), “Predicting Symptom Severity in Autism Spectrum Disorder Based on Cortical Thickness Measures in Agglomerative Data”, *NeuroImage*, Vol.144, (128-141).
- Moskov, D. (2016), “Knowledge Extraction From Published Papers in Literature For The Catalytic Methanol Production From Synthesis Gas Using Data Mining Tools”, Yayınlanmamış Yüksek Lisans Tezi. İstanbul: Boğaziçi Üniversitesi. Fen Bilimleri Enstitüsü.
- Muhr, M., and Granitzer, M. (2009), “Automatic Cluster Number Selection Using A Split and Merge K-Means Approach”, *In 20th International Workshop on Database and Expert Systems Application*, (363-367).
- Nilashi, M., Ibrahim, O., Dalvi, M., Ahmadi, H., and Shahmoradi, L. (2017), “Accuracy Improvement For Diabetes Disease Classification: A Case on A Public Medical Dataset. *Fuzzy Information and Engineering*, Vol.9, Issue.3, (345-357).
- Oğuzlar, A. (2003), “Veri Ön İşleme”, *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, S.21, (67-76).

Olçay Gül, S., ve Tekin İftar, E. (2012), “Otizm Spektrum Bozukluğu Tanısı Olan Bireyler İçin Sosyal Öykülerin Kullanımı”, *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Özel Eğitim Dergisi*, C.13, S.2, (1-24).

Özeren, S. G. (2013), “Otizm Spektrum Bozukluğu (Os) ve Hastalığa Kanıt Penceresinden Bakış”, *Acıbadem Üniversitesi Sağlık Bilimleri Dergisi*, C.4, S.2, (57-63).

Özkaya, B. T. (2013), “Yaygın Gelişimsel Bozukluklardan Otizm Spektrum Bozukluğuna Geçiş: Dsm-V'de Karşımıza Çıkacak Değişiklikler”, *Psikiyatride Güncel Yaklaşımlar*, C.5, S.2, (127-139).

Özsoy, İ., ve Fırat, M. (2004), “Kirişsiz Döşemeli Betonarme Bir Binada Oluşan Yatay Deplasmanın Yapay Sınır Ağları İle Tahmini”, *Deü Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, C.6, S.1, (51-63).

Peixoto, R., Ribeiro, L., Portela, F., Santos, M. F., and Rua, F. (2017), “Predicting Resurgery in Intensive Care-A Data Mining Approach”, *Procedia Computer Science*, Vol.113, (577-584).

Pelleg, D., and Moore, A. W. (2000), “X-Means: Extending K-Means with Efficient Estimation of The Number of Clusters”, *In Icml*, Vol.1, (727-734).

Pençe, İ., Çetişli, B. (2013), “El Yazı Karakterlerinin Kapalı Cebirsel Eğrilerle Modellenmesi ve Sınıflandırılması”, *Sigma*, C.5, (1-7).

Peurifoy, J., Shen, Y., Jing, L., Yang, Y., Cano-Renteria, F., DeLacy, B. G., ... and Soljačić, M. (2018), “Nanophotonic Particle Simulation and Inverse Design Using Artificial Neural Networks”, *Science Advances*, Vol.4, Issue.6, (1-7).

Reddy, S. K., Kodali, S. R., ve Gundabathina, J. L. (2012), “Classification of Vertebral Column Using Naïve Bayes Technique”, *International Journal Of Computer Applications*, Vol.58, Issue.7, (38-42).

Sabuncuoğlu, M., Cebeci, S., Rahbar, M. H., and Hessabi, M. (2015), “Autism Spectrum Disorder and Attention Deficit Hyperactivity Disorder: Knowledge and Attitude of Family Medicine Residents in Turkey”, *Turkish Journal of Family Medicine & Primary Care*, C.9, S.2, (46-53).

- Sađırođlu, . (2016) “Türkiye’nin İller Bazında Enerji İhtiyacının Veri Madenciliđi İle Analizi ve Lisanssız Yenilenebilir Enerji Yatırımları İin Karar Destek Önermesi”, Yayınlanmamıř Doktora Tezi. Adana: ukurova Üniversitesi Fen Bilimleri Enstitüsü.
- Sariman, G. (2011), “Veri Madenciliđinde Kümeleme Teknikleri Üzerine Bir alıřma: K-Means ve K-Medoids Kümeleme Algoritmalarının Karřılařtırılması”, *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, C.15, S.3, (192-202).
- Sebik, N. B., ve Bülbül, H. İ. (2018), “Veri Madenciliđi Modellerinin Akciđer Kanseri Veri Seti Üzerinde Başarılarının İncelenmesi”, *TÜBAV Bilim Dergisi*, C.11, S.3, (1-7).
- Segin, S. (2016), “Supporting Managerial Decisions Using Data Mining: Evidence From Telecommunication Industry in Turkey”, Yayınlanmamıř Doktora Tezi. İzmir: Dokuz Eylül Üniversitesi Fen Bilimleri Enstitüsü.
- Sezgin, E., ve elik, Y. (2013), “Veri Madenciliđinde Kayıp Veriler İin Kullanılan Yöntemlerin Karřılařtırılması”, *Akademik Biliřim Konferansı*, (23-25).
- Steinley, D. (2006), “K- Means Clustering: A Half- Century Synthesis”, *British Journal of Mathematical and Statistical Psychology*, Vol.59, Issue.1, (1-34).
- Talan, M. İ. (2016), “Veri Madenciliđi İle Karpal Tünel Sendromuna Yönelik Ön Tanı Destek ve Hasta Takip Sisteminin Geliřtirilmesi”, Yayınlanmamıř Yüksek Lisans Tezi. Ankara: Gazi Üniversitesi Biliřim Enstitüsü.
- Tařcı, E., ve Onan, A. (2016), “K-En Yakın Komřu Algoritması Parametrelerinin Sınıflandırma Performansı Üzerine Etkisinin İncelenmesi”, *Akademik Biliřim*.
- Tekin, R., Kaya, Y., ve Tađluk, M. E. (2011), “K-Means ve Ysa Temelli Hibrit Bir Model İle Epileptik Eeg İřaretlerinin Sınıflandırılması”, *Elektrik Elektronik Bilgisayar Sempozyumu*, (277-283).
- Thabtah, F. (2017), “Autism Spectrum Disorder Screening: Machine Learning Adaptation and Dsm-5 Fulfillment”, *In Proceedings of the 1st International Conference on Medical and Health Informatics*, (1-6).
- Thabtah, F. (2018), “An Accessible and Efficient Autism Screening Method For Behavioural Data and Predictive Analyses”, *Health Informatics Journal*, Vol.25, Issue.4, (1739-1755).

Thabtah, F., Kamalov, F., and Rajab, K. (2018), “A New Computational Intelligence Approach to Detect Autistic Features For Autism Screening”, *International Journal of Medical Informatics*, Vol.117, 112-124.

Torun, Y. T., ve İşeri, E. (2018), “Bebeklik Döneminde Önemli Bir Psikopatoloji: Otizm”, *Türkiye Klinikleri Journal of Child Psychiatry-Special Topics*, C.4, S.1, (48-55).

Tsai, C. Y., and Chiu, C. C. (2008), “Developing A Feature Weight Self-Adjustment Mechanism For A K-Means Clustering Algorithm”, *Computational Statistics & Data Analysis*, Vol.52, Issue.10, (4658-4672).

Tsumoto, S., Kimura, T., Iwata, H., and Hirano, S. (2017), “Mining Text For Disease Diagnosis”, *Procedia Computer Science*, Vol.122, (1133-1140).

Tucker, C. S., Kim, H. M., Barker, D. E., and Zhang, Y. (2010), “A Relief Attribute Weighting and X-Means Clustering Methodology For Top-Down Product Family Optimization”, *Engineering Optimization*, Vol.42, Issue.7, (593-616).

Tunç, A., ve Ülger, İ. (2016), “Veri Madenciliği Uygulamalarında Özellik Seçimi İçin Finansal Değerlere Binning ve Five Number Summary Metotları İle Normalizasyon İşleminin Uygulanması”, XVIII. *Akademik Bilişim Konferansı*, Adnan Menderes Üniversitesi, 30.

Türkoğlu, T. (2016), “Çoklu Ölçüt Oy Değerleri Üzerinden Veri Madenciliği”, Yayımlanmamış Yüksek Lisans Tezi. Eskişehir: Anadolu Üniversitesi Fen Bilimleri Enstitüsü.

Uci Machine Learning Repository, *Autistic Spectrum Disorder Screening Data for Children* *Data Set*, <https://archive.ics.uci.edu/ml/datasets/Autistic+Spectrum+Disorder+Screening+Data+for+Children++>, (03.02.2018).

Uğur, Ç., ve Göker, Z. (2018), “Bir Çocuk Psikiyatrisi Ünitesine Başvuran Olgularda Otizm Spektrum Bozukluğu: Demografik ve Klinik Bulgular”, *Türkiye Çocuk Hastalıkları Dergisi*, C.13, S.3, (177-183).

Uğur, Ç., Sertçelik, M., Üneri, Ö., Dinç, G. Ş., Sekmen, E., ve Solmaz, E. (2018), “Evaluation of Serum Urotensin-II Levels of Children with ADHD and Autism Spectrum Disorder”, *Anadolu Psikiyatri Dergisi*, C.19, S.1, (80-86).

- Van den Bekerom, B. (2017), "Using Machine Learning For Detection of Autism Spectrum Disorder", In *Proc. 20th Student Conf.* (1-7).
- Villarrubia, G., De Paz, J. F., Chamoso, P., and De la Prieta, F. (2018), "Artificial Neural Networks Used in Optimization Problems", *Neurocomputing*, Vol.272, (10-16).
- Yavuz, Ü., Ekim, U. and Köklü, M. (2011), "Üniversite Öğrencilerin Ortak Zorunlu Derslerdeki Başarılarının K-Means Algoritması İle İncelenmesi", *NWSA: Engineering Sciences*, C.6, S.1, (342-347).
- Wall, D. P., Kosmicki, J., Deluca, T. F., Harstad, E., and Fusaro, V. A. (2012), "Use of Machine Learning to Shorten Observation-Based Screening and Diagnosis of Autism. *Translational psychiatry*, Vol.2 Issue.4, (1-8).
- Yavuz, S., ve Deveci, M. (2012), "İstatiksel Normalizasyon Tekniklerinin Yapay Sinir Ağı Performansına Etkisi" *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, S.40, (167-187).
- Yıldırım, M. (2016), "İldeki Kurumlar Arası Çalışma Performansının Arttırılmasında Veri Madenciliği Tekniklerinin Kullanılması", Yayınlanmamış Yüksek Lisans Tezi. Elazığ: Fırat Üniversitesi Fen Bilimleri Enstitüsü.

ÖZGEÇMİŞ

Kişisel Bilgiler :

Adı ve Soyadı : Sümeyye ÇELİK

Doğum Yeri :Gümüşhane

Medeni Hali :Bekâr

Eğitim Durumu :

Lisans Öğrenimi : Burdur Mehmet Akif Ersoy Üniversitesi / Bucak Zeliha Tolunay Uygulamalı Teknoloji ve İşletmecilik Yüksek Okulu /Yönetim Bilişim Sistemleri Bölümü (2013-2017)

Yüksek Lisans Öğrenimi : Burdur Mehmet Akif Ersoy Üniversitesi / Sosyal Bilimler Enstitüsü / Yönetim Bilişim Sistemleri Tezli Yüksek Lisans (2017-)

Yabancı Dil(ler) ve Düzeyi : İngilizce (iyi seviye)

İş Denevimi : Adana Alparslan Türkeş Bilim ve Teknoloji Üniversitesi / Yönetim Bilişim Sistemleri Bölümü / Araştırma Görevlisi (2019-)

Bilimsel Yayınlar ve Çalışmalar: