

T.C.
YÜZÜNCÜ YIL ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

ÇOK DEĞİŞKENLİ REGRESYON AĞAÇLARI VE UYGULAMASI

Duygu KORKMAZ
BİYOİSTATİSTİK ANABİLİM DALI
(TIP PROGRAMI)
YÜKSEK LİSANS TEZİ

DANIŞMAN
Prof. Dr. Sıddık KESKİN

VAN-2018

T.C.
YÜZÜNCÜ YIL ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

ÇOK DEĞİŞKENLİ REGRESYON AĞAÇLARI VE UYGULAMASI

Duygu KORKMAZ
BİYOİSTATİSTİK ANABİLİM DALI
(TIP PROGRAMI)
YÜKSEK LİSANS TEZİ

DANIŞMAN
Prof. Dr. Sıddık KESKİN

VAN-2018

T.C.
YÜZÜNCÜ YIL ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

ÇOK DEĞİŞKENLİ REGRESYON AĞAÇLARI VE UYGULAMASI

Duygu KORKMAZ
BİYOİSTATİSTİK ANABİLİM DALI
(TIP PROGRAMI)
YÜKSEK LİSANS TEZİ

Prof. Dr. Sıddık KESKİN
(Jüri Başkanı)

Prof. Dr. Sinan ÇALIK
(Üye)

Prof. Dr. H. Eray ÇELİK
(Üye)

TEZ KABUL TARİHİ
03/08/2018

TEŐEKKÜR

Bu tez alıőmasında verdiđi destekten dolayı, her tŸrlŸ ilgi ve yardımını esirgemeyen, hem bilgi birikimi hem de akademiye verdiđi emekle bana her zaman gerek bir bilim insanı rneđi olan sađduyulu danıőmanım Sayın Prof. Dr. Sıddık KESKİN'e, deđerli katkılarından, tavsiyelerinden, ilgi ve alakasından dolayı Prof. Dr. Zeliha Nazan ALPARSLAN'a, ayrıca alıőmalarımda bana her zaman destek olan aileme ve deđerli dostlarıma en iten teőekkŸrlerimi sunarım.



İÇİNDEKİLER

KABUL VE ONAY	III
TEŞEKKÜR	IV
İÇİNDEKİLER.....	V
SİMGELER VE KISALTMALAR	VI
ŞEKİLLER LİSTESİ.....	VII
TABLolar LİSTESİ	VIII
1. GİRİŞ.....	1
2. GENEL BİLGİLER.....	3
2.1. Sınıflandırma ve Regresyon Ağaçları Yöntemi Literatür Özeti.....	3
2.2. Çok Değişkenli Regresyon Ağaçları Yöntemi Literatür Özeti.....	6
3. GEREÇ YÖNTEM.....	10
3.1. Gereç.....	10
3.2. Yöntem	10
3.2.1. Sınıflandırma ve regresyon ağacı	11
3.2.2. Tek değişkenli regresyon ağacı	12
3.2.3. Çok değişkenli regresyon ağacı.....	13
4. BULGULAR	16
5. TARTIŞMA ve SONUÇ	37
ÖZET	41
SUMMARY	42
KAYNAKLAR.....	43
ÖZGEÇMİŞ.....	46
EK 1. AĞAÇ OLUŞUMU İÇİN R KODLARI.....	47
EK 2. ETİK KURUL BELGESİ.....	49
EK 3. TEZ ORJİNALLİK RAPORU.....	50

SİMGELER VE KISALTMALAR

Σ	: Toplam sembolü
X_j	: Sürekli bir değişken
D	: Düğüm
t_{sol}	: Sol düğüm
$t_{sağ}$: Sağ düğüm
p_{sol}	: t_{sol} düğümüne giden gözlemlerin kesirleri
$P_{sağ}$: $t_{sağ}$ düğümüne giden gözlemlerin kesirleri
t_p	: Ebeveyn düğüm
S	: Aday bölünme
$i(t)$: t. düğümün heterojenliği
y	: t. noda ait X_n gözlemi için cevap değişkeninin değeri
$\bar{y}(t)$: t. düğümde bulunan tüm gözlemlerin ortalaması
$Y_{ij(k)}$: k test setindeki gözlem değeri
$\hat{y}_{j(k)}$: k test setindeki gözlem için değişken j'nin tahmini değeri
\bar{y}_j	: j değişkeni için genel ortalama
N	: Gözlem sayısı
P	: Yanıt matrisi Y'deki değişken sayısı
Di	: Sapma (Deviance)
ALB	: Albümin
ALT	: Alanin Aminotransferaz
AST	: Aspartat Aminotransferaz
CART	: Classification and Regression Trees
CVRE	: Cross Validated Relative Error
ÇDRA	: Çok Değişkenli Regresyon Ağacı
GLB	: Globülin
HB	: Hemoglobin
MRT	: Multivariate Regression Trees
PCR	: Polimeraze Chain Reaction
PLT	: Platelets
RE	: Relative Error
TSSB	: Travma Sonrası Stres Bozukluğu
WBC	: White Blood Cell

ŞEKİLLER LİSTESİ

- Şekil 1.** Tüm değişkenlerin modele dahil edilmesiyle oluşan çok değişkenli regresyon ağacı.16
- Şekil 2.** ALB değişkeni modelden çıkarıldığında oluşan çok değişkenli regresyon ağacı20
- Şekil 3.** ALB ve HB değişkenlerinin modelden çıkarılması ile oluşan çok değişkenli regresyon ağacı23
- Şekil 4.** ALB, HB ve GLB değişkenlerinin modelden çıkarılması ile oluşan çok değişkenli regresyon ağacı26
- Şekil 5.** ALB, HB, GLB ve ALT değişkenlerinin modelden çıkarılması ile oluşan çok değişkenli regresyon ağacı.....29
- Şekil 6.** ALB, HB, GLB, ALT ve PLT değişkenlerinin modelden çıkarılması ile oluşan çok değişkenli regresyon ağacı.....32
- Şekil 7.** Tüm değişkenlerin modele dahil edilmesi ile 1SE kuralına göre oluşan çok değişkenli regresyon ağacı35

TABLolar LİSTESİ

Tablo 1. Tüm deęişkenlerin modele dahil edilmesi ile oluşturulan tablo.....	19
Tablo 2. Tüm deęişkenlerin modele dahil edilmesi ile oluşan hata çizelgesi	20
Tablo 3. Tüm deęişkenlerin modele dahil edilmesi ile ağaç bölünmeleri ve tüm ağaca ait sapma (deviance) deęerleri %	20
Tablo4. ALB deęişkeni modelden çıkarıldığında oluşturulan tablo	22
Tablo 5. ALB deęişkeni modelden çıkarıldığında hesaplanan hata çizelgesi	23
Tablo 6. ALB deęişkeni modelden çıkarıldığında ağaç bölünmeleri ve tüm ağaca ait sapma (deviance) deęerleri %	23
Tablo 7. ALB ve HB deęişkenleri modelden çıkarıldığında oluşturulan tablo	25
Tablo 8. ALB ve HB deęişkenleri modelden çıkarıldığında hesaplanan hata çizelgesi.....	26
Tablo 9. ALB ve HB deęişkenleri modelden çıkarıldığında ağaç bölünmeleri ve tüm ağaca ait sapma (deviance) deęerleri %	26
Tablo 10. ALB, HB ve GLB deęişkenleri modelden çıkarıldığında oluşturulan tablo	28
Tablo 11. ALB, HB ve GLB deęişkenleri modelden çıkarıldığında hata çizelgesi	29
Tablo 12. ALB, HB ve GLB deęişkenleri modelden çıkarıldığında ağaç bölünmeleri ve tüm ağaca ait sapma (deviance) deęerleri %	29
Tablo 13. ALB, HB, GLB ve ALT deęişkenleri modelden çıkarıldığında oluşturulan tablo ...	31
Tablo 14. ALB, HB, GLB ve ALT deęişkenleri modelden çıkarıldığında hesaplanan hata çizelgesi	32
Tablo 15. ALB, HB, GLB ve ALT deęişkenleri modelden çıkarıldığında ağaç bölünmeleri ve tüm ağaca ait sapma (deviance) deęerleri %	32
Tablo 16. ALB, HB, GLB, ALT ve PLT deęişkenleri modelden çıkarıldığında oluşturulan tablo.....	34
Tablo 17. ALB, HB, GLB, ALT ve PLT deęişkenleri modelden çıkarıldığında hesaplanan hata çizelgesi	34
Tablo 18. ALB, HB, GLB, ALT ve PLT deęişkenleri modelden çıkarıldığında ağaç bölünmeleri ve tüm ağaca ait sapma (deviance) deęerleri %	34
Tablo 19. Tüm deęişkenlerin modele dahil edildiğinde 1SE kuralına göre oluşturulan tablo...36	
Tablo 20. Tüm deęişkenlerin modele dahil edildiğinde 1SE kuralına göre hesaplanan hata çizelgesi	36
Tablo 21. Tüm deęişkenlerin modele dahil edildiğinde 1SE kuralına göre oluşturulan ağaç bölünmeleri ve tüm ağaca ait sapma (deviance) %.....	36

1. GİRİŞ

Araştırma, herhangi bir bilinmeyen sorun ile ilişkili olarak planlı ve sistematik veri toplama ve bu verinin analizi ve yorumlanması yoluyla soruna güvenilir bir çözüm getirme sürecidir. Araştırmacının amacına ulaşmasına olanak sağlayan araştırma, bilginin ilerlemesindeki en önemli süreçtir (Singh, 2006).

Bilimsel araştırmalarda ise istatistik yöntemler önemli bir yere sahiptir. Özellikle, deneysel araştırmalarda kurulan hipotezin anlamlı olup olmadığını sınımanın yolu istatistik yöntemlere dayanmaktadır. Bilimsel araştırmalarda, ele alınan probleme ilişkin bütün popülasyonun incelenmesi, uzun zaman gerektirdiğinden ve/ya ekonomik olmadığından çoğu kere mümkün olmamaktadır. Bu nedenle popülasyondan alınan örnekler yardımıyla popülasyon parametrelerine ilişkin tahminler yapılmaktadır. Örnekten veya örneklemeden elde edilen sonuçların da doğru ve güvenilir olması istenir. Doğru ve güvenilir sonuçlar elde etmenin yolu temelde uygun istatistik yöntemlerin kullanılmasına dayanır.

Bilimsel araştırmalarda, istatistik analiz yöntemleri; genel olarak parametrik ve parametrik olmayan yöntemler şeklinde ayrılabilir gibi tek değişkenli ve çok değişkenli yöntemler olarak da ayrılmaktadır. Bu yöntemler, değişkenler arasındaki doğrusal ve doğrusal olmayan ilişkileri bulmayı amaçlamaktadır.

Değişkenler arasındaki ilişkiler gerek doğrusal yöntemlerle gerekse doğrusal olmayan yöntemlerle incelendiğinde, anlaşılmasının ve yorumlanmasının kolaylığı açısından çoğunlukla tek değişkenli yöntemler tercih edilmektedir. Diğer bir ifadeyle, değişkenler arasındaki ilişkiler ikili olarak ele alınmakta ve böylece diğer değişkenlerin etkisinden bağımsız olarak incelenmektedir. Oysa ki bu ilişkiler, doğada oldukça karmaşık yapıdadır. Birden fazla cevap değişkeni ve birden fazla açıklayıcı değişken arasındaki ilişkilerle, birçok alanda yaygın olarak karşılaşmaktadır. Sağlık bilimlerinde fizyolojik özellikler ile biyokimyasal özellikler arasındaki ilişkiler, ekonomide sağlık harcamaları ile eğlence harcamaları arasındaki ilişkiler, çevre bilimlerinde türler ile çevresel değişkenler arasındaki ilişkiler buna örnek verilebilir. Genel olarak düşünüldüğünde, bu değişkenler arasındaki ilişkiler bireysel ya da birlikte ele alınarak incelenebilir. Ancak, daha önceden de belirtildiği üzere, bu ilişkilerin ayrı ayrı bağımsız olarak ele alıp incelenmesiyle, birlikte tek modelde ele alınarak incelenmesi sonucunda, elde edilen sonuçların ve yorumlamaların farklı olacağı aşikardır. Değişkenlerin ayrı ayrı ele alınarak incelenmesinde; genelleştirilmiş doğrusal modeller, genelleştirilmiş

eklemeli modeller ve CART (Sınıflama ve Regresyon Ağacı) gibi yöntemler kullanılabilir. Ancak bu yöntemler, genel olarak doğrusal ilişkilere dayalı analiz yapmaktadır. Diğer yandan bu yöntemlerin birçoğu, veri setindeki dağılımın dengesiz olmasına karşı hassas olmakta ve kayıp gözlemler olması durumunda sonuçlar değişebilmektedir. Değişkenler arasındaki doğrusal olmayan ilişkilerin ve yüksek derecedeki etkileşimlerin incelenmesi istendiğinde bu yöntemler yetersiz kalmaktadır. Bunların yerine değişkenleri birlikte ele alarak inceleyen yöntemler geliştirilmiştir. Çok değişkenli istatistik analiz yöntemleri olarak adlandırılan bu yöntemler, genel olarak karmaşık veri setlerini analiz etmede giderek yaygın bir şekilde kullanılmaktadır.

Bu yöntemler, çok sayıda bağımsız ve/veya bağımlı değişken arasındaki ilişkileri analiz etmede, birkaç tek değişkenli analiz yapmak yerine, bir defada birlikte analiz etme imkanı verir (Anderson, 1958).

Sınıflandırma ve Regresyon Ağacı (CART, Classification and Regression Tree) yöntemi, parametrik olmayan bir yöntem olması ve tek değişkenli cevap değişkeni için sonuçların kolay yorumlanabilir olması nedeniyle yaygın olarak kullanılmaktadır. Ancak, birden fazla cevap değişkeni ve açıklayıcı değişken olması durumunda, CART'ın her cevap değişkeni için ayrı ayrı kullanılarak analiz edilmesinin yerine, bu cevap değişkenlerinin birlikte analiz edilmesine olanak sağlayan Çok Değişkenli Regresyon Ağacı (ÇDRA) yöntemi geliştirilmiştir. İlk olarak De'ath (2002) tarafından önerilen ve CART'ın genişletilmiş hali olan Çok değişkenli regresyon ağacı yönteminde, tek değişkenli regresyon ağacındaki tek bir cevap değişkeninin yerini çoklu cevap değişkenleri alır. Çok Değişkenli Regresyon Ağacı, varsayımların esnekliği, sonuçların doğru ve güvenilir olması ve ayrıca kolay yorumlanabilirliği ve interaksiyonları göstermesi bakımından diğer alternatif yöntemlere göre avantajlıdır. Çok değişkenli regresyon ağacı yöntemi; açıklama, tanımlama ve tahmin amaçlı kullanılabilir. Cevap değişkenlerinin sürekli olması dışında, açıklayıcı değişkenlere ilişkin varsayımlar hakkında esneklik.

Yapılan literatür incelemesi sonucunda, çok değişkenli regresyon ağacı yöntemi ile ilgili Türkçe literatürün yok denecek kadar az olduğu dikkat çekmiştir. Bu eksikliğin giderilmesi için tasarlanan bu çalışmada; yöntemin tanıtılması, temel kavramların verilmesi ve teorik yapısının, uygulama ile birlikte incelenmesi amaçlanmıştır.

2. GENEL BİLGİLER

2.1. Sınıflandırma ve Regresyon Ağaçları Yöntemi Literatür Özeti

Breiman ve ark. (1984), Sınıflandırma ve Regresyon Ağaçları yöntemine ilişkin tanımlar ve ayrıntılı teorik bilgilere değinerek farklı alanlarda uygulamalara yer vermişlerdir.

Reibnegger ve ark. (1991), klinik kimyada kronik karaciğer rahatsızlığı olan hastalardan edinilen veri setini kullanarak, sınıflandırma probleminde; Yapay Sinir Ağları, Doğrusal Diskriminant Analizi ve Sınıflandırma ve Regresyon Ağaçları yöntemlerini karşılaştırmışlar ve sonuç olarak, Yapay Sinir Ağları Yönteminin, verideki gizli yapının keşfinde daha başarılı olduğunu vurgulamışlardır.

Curran ve ark. (1993), yaptıkları çalışmada, malign gliomların tedavisinde hem tedavi öncesi hem de tedaviye bağlı değişkenlerle bir analiz yapılması gerektiğine dayanarak, Regresyon Ağacı Modeli oluşturmuşlar ve prognostik faktörlerin, tedavi sırasındaki küçük değişikliklerden daha etkili olduğunu vurgulamışlardır.

Falconer ve ark. (1994), 1987'den 1991'e kadar toplanan verilerde; Sınıflandırma Ağacı yaklaşımını kullanarak, inme (felç ya da beyin krizi) hastalarının rehabilitasyon sonucunu tahmin etmeye çalışmışlardır. Sonuçta sınıflandırmanın başarısının %88 olduğunu belirterek, yöntemin kullanılabilir olabileceğini vurgulamışlardır.

Hess ve ark. (1999), bilinmeyen primerkarsinom hastalarına ait klinik özelliklerin ve sağkalım sürelerinin farklılık gösterdiği gerekçesiyle, bu hastalığa ilişkin, yeni prognostik faktörleri tanımlamayı ve klinik değişkenler ile sağkalım üzerindeki etkiler arasındaki etkileşimleri araştırmayı amaçlamışlardır. Bu amaca uygun olarak 1000 hasta, 1 Ocak 1987'den 30 Kasım 1994'e kadar Texas M.D. Anderson Kanseri Merkezi'ne sevk edilerek bu hastalara ilişkin klinik veriler bir veri tabanına kaydedilmiştir ve kaydedilen bu veriler üzerinde uygulanan CART yönteminin, hastalığa ilişkin önemli prognostik değişkenleri saptadığı ve hastaları benzer klinik özelliklere göre sınıflandırdığı belirlenmiştir. Ayrıca CART yönteminin daha önce belirlenmemiş hasta alt kümelerini tanımlamadaki başarısından hareketle, gelecekteki klinik çalışmalarda homojen hasta popülasyonlarını belirlemede yararlı olabileceği ifade edilmiştir.

De'ath ve Fabricius (2000), ekolojik veri setlerinin çoğu zaman kayıp gözlem ve yüksek derecede interaksiyonlar içerdiğini ve karmaşık bir yapıda olduğunu vurgulayarak, mercan taksonomisi ve varlığını (bolluğunu) içeren bir veri seti üzerinde ağaç yaklaşımının başarılı bir performans sergilediğini ortaya koymuşlardır. Bu tür veri setlerini analiz etmede; Doğrusal modellerle, Regresyon Ağacı yöntemini karşılaştırarak ağaç yöntemlerinin üstünlüğünü vurgulamışlar ve Doğrusal Modellerin gizli yapıyı bulmada başarısız olduğuna dikkat çekmişlerdir.

Liaw ve Wiener (2002), çalışmalarında Denetimli Öğrenme konusu olan Sınıflandırma ve Regresyon Ağaçları Yöntemlerini iki ayrı veri seti üzerinde deneyerek, bu yöntemlerin bir koleksiyonu olan Random Forest yöntemini, Sınıflandırma ve Regresyon Ağaçları yöntemlerine ayrı ayrı uygulamışlardır. Yöntemlerin R paket programı üzerinde nasıl uygulanacağını ayrıntılı göstererek, Random Forest yöntemini ayrıca denetimsiz öğrenme tekniği olarak R paket programında bir veri seti üzerinde denemişlerdir.

Lemon ve ark. (2003), çalışmalarında, CART'a ilişkin sürece genel metodolojik bir bakış sunmayı amaçlayıp, yöntemin halk sağlığındaki önemine dikkat çekerek, bu yöntemin halk sağlığı literatüründeki yerine değinmişlerdir. Bölünme kriterleri, durma kriteri, budama süreci, sınıflandırma hatası gibi yönetime ilişkin önemli başlıklara yer verdikten sonra örnek üzerinde oluşturdukları modeli inceleyip sonuçlarını değerlendirmişlerdir.

Timofeev (2004), yaptığı çalışmada, Sınıflandırma ve Regresyon Ağaçları yöntemine ilişkin maksimum ağaç boyutunu seçmede Gini ve Twoing algoritmaları ile birlikte, ağaç için doğru boyutu hesaplamada çapraz geçerlik kavramına değinerek, birisi simülasyon verisi olmak üzere, yöntemi 3 ayrı veri setinde incelemiş ve yöntemin avantaj ve dezavantajlarından bahsetmiştir.

Fonarow ve ark. (2005), Akut Dekompanse Kalp Yetmezliği olan hastalarda, mortalite (ölüm) riskini tahmin etmek için Sınıflandırma ve Regresyon Ağacı yöntemini kullanmışlardır. Bunun için hastaneye yatış esnasında elde edilen yaşamsal belirti ve laboratuvar verileri kullanılarak, mortalite (ölüm) riski (düşük, orta ve riskli) üç kategoriye ayrılmıştır. Sonuç olarak, 39 değişkenden sadece 3'ünün mortalite (ölüm) riskini belirlemede önemli olduğu ve Sınıflandırma Ağacında yer aldığı belirtilmiştir.

Prasad (2006), Kanada İklim Merkezi küresel dolaşım modeline göre mevcut ve gelecekteki iklim senaryoları altında tahmini bitki örtüsü haritalaması için Regresyon Ağaçları,

Bagging Ağaçları, Random Forest ve MARS olmak üzere 4 yöntemi karşılaştırmıştır. Çalışma sonucunda, Regresyon Ağaçlarının klasik yöntemlere göre iyi bir performans sergilediği ve avantajlı olduğu belirtilmiştir. Bunun yanı sıra, Bagging Ağaç ve Random Forest yöntemleri, Mars ve Regresyon Ağaçlarından daha iyi bir haritalama yaptığından, topluluk yöntemlerinin tek bir regresyon ağacından daha etkili olduğu vurgulanmıştır. Çalışmada, Random Forest yönteminin, mevcut dağılım ve gelecekteki potansiyel habitatları haritalamada üstün öngörü kabiliyetine sahip olduğu, Bagging ve Regresyon Ağaçlarının ise kolay yorumlanabilir olması nedeniyle ön plana çıktığı vurgulanarak, bu üç yöntemin kombinasyonunun etkin bir şekilde kullanılabilir olduğuna dikkat çekilmiştir.

Moisen (2008), yaptığı çalışmada, Sınıflandırma ve Regresyon Ağacı yöntemine ilişkin fikir rehberliği yapmayı amaçlayarak, yöntemin kullanımında karşılaşılan sorunları, terminolojisini veson uzantılarını genel hatlarıyla ele almış ve konuyu basitçe açıklamaya çalışmıştır.

Strobl ve ark. (2009), yaptıkları çalışmada, tekrarlı bölünme yöntemlerinin birçok alanda parametrik olmayan yöntemler için yaygın olduğunu belirterek, tekrarlı bölünme temeline dayalı geliştirilen yöntemleri sunmuşlardır. Ayrıca yöntemi, çeşitli uygulama alanlarındadüşük ve yüksek boyuttaki veri setleri üzerinde deneyerek, yöntemin başarısını incelemişler ve kısıtlı yanlarına değinmişlerdir.

Loh (2011), yaptığı çalışmada, Sınıflandırma ve Regresyon Ağacı yöntemine ilişkin tanımlamaları yaptıktan sonra, yöntemin teorik alt yapısını ele alarak, iki örnek üzerinde yönteme ilişkin bazı yaygın algoritmalara değinmiş ve yöntemin avantaj ve dezavantajlarından bahsetmiştir.

Speybroeck (2012), çalışmasında Sınıflandırma ve Regresyon Ağaçları yöntemini ayrıntılı ele alarak, teorik bilgilerle birlikte yöntemin avantajlarına değinmiştir.

Zimmerman ve ark. (2016), grip tanısına uygun reçete yazımında, hızlı ve uygun maliyetli yöntemlere ihtiyaç duyulması ve ilgili problemlere çözüm yolu bulunması amacıyla, Sınıflandırma ve Regresyon Ağaçları (CART) yöntemini ele almışlardır. 2011-2012 yılları arasında öksürük veya ateşle akut solunum yolu rahatsızlığı tedavisi için 5 yaş üzeri hastalardan, ayaktan tedavi merkezlerinde, grip teşhisinde kullanılan PCR testi için, burun sürüntüsüile birlikte, demografik özellikleri, semptomları, kişisel özellikleri ve alınan diğer kayıtları içeren veri setine Sınıflandırma ve Regresyon Ağacı yöntemini uygulamışlardır. Sonuçta; ateş,

öksürük ve yorgunluk belirtileri değişkenlerinin modele katkı bakımından önemli değişkenler olduğu belirtilerek; duyarlılık, özgüllük, pozitif prediktif değer ve negatif prediktif değerlerinin sırasıyla; % 84, % 48, % 23 ve % 94 olduğu bildirilmiştir.

Suchetana ve ark. (2017), 2004-2008 dönemi için 106 belediye arıtma tesisinden elde edilen bir örneklemin aylık atık raporu verileri ve amonyak deşarj izin limitleri ile ABD atıksu arıtma tesisi uyumluluğunu etkileyen faktörleri değerlendirmek için bir Regresyon Ağacı tabanlı yaklaşım geliştirmişlerdir. Standart Regresyon analizlerinden farklı olarak, CART yönteminin, girdi değişkenleri arasında karmaşık üst düzey etkileşimleri yakalama ve bu etkileşimleri görsel olarak sunma konusunda yüksek bir kabiliyete sahip olduğunu belirtmişlerdir. Modelin, mevcut gelecekteki limitleri belirleyebilmede başarılı olduğu vurgulanarak, Regresyon Ağaçları yönteminin doğru sınıflama oranının %70 olduğu belirtilmiştir.

Ebel ve Hansen (2017), gereksiz antibiyotik reçetelemeyi azaltmak için, akut rinosinüzit ve akut bakteriyel rinosinüzit tanısı koymada bir klinik karar verme kuralı geliştirme amacıyla çalışma yapmışlardır. Akut rinosinüziti ve akut bakteriyel rinosinüziti, yüksek bir doğrulukla teşhis etmek için Sınıflandırma ve Regresyon Ağacı yöntemini kullanarak; belirtileri, semptomları ve CRP'yi bir araya getiren bir dizi klinik karar kuralı geliştirmişlerdir. Performans değerlendirme ölçüsü olan ROC Eğrisi altında kalan alan, 0.783 ile 0.827 arasında bulunmuş olup, yöntemin başarılı bir sınıflandırma gösterdiği vurgulanmıştır.

Ross ve Kearney (2017), bireysel psikolojik faktörlerin, kötü muameleyi takiben gençlerde travma sonrası stres bozukluğu (TSSB) semptomları riskini artırdığını, ancak bu çalışmanın yeni katkısının, belirli TSSB semptom kümeleri üzerindeki bu faktörler arasındaki etkileşimli ilişkilere odaklanması olduğunu belirtmişlerdir. Çalışmada, TTSB'nin yeniden yaşanması, önlenmesi ve aşırı duygusal belirtileri açısından en çok risk altındaki kötü muamele gören gençlerin alt gruplarını tanımlamak için Sınıflandırma ve Regresyon Ağacı (CART) yöntemi kullanılmıştır. Araştırmanın temel amacı, birlikte düşünülen belirgin bilişsel, etkili ve demografik değişkenlerin kötü muamele gören gençlerde, TSSB'nin kilit yönlerini nasıl öngördüğü konusuna daha fazla açıklık getirmektir.

2.2. Çok Değişkenli Regresyon Ağaçları Yöntemi Literatür Özeti

Çok Değişkenli Regresyon Ağaçları (MRT) ilk olarak De'ath (2002) tarafından tanıtılmıştır. Çok değişkenli regresyon ağacının, veri setinde kayıp gözlemler olması ile birlikte,

değişkenler arasında doğrusal olmayan ilişkiler ve üst düzey etkileşim içerebilen karmaşık ekolojik verileri analiz etmede kullanılabilen vurgulanmıştır. Çalışmada, öncelikle Tek Değişkenli Regresyon Ağacı yöntemine ilişkin bilgiler verildikten sonra, Çok Değişkenli Regresyon Ağaçları Yöntemi, Uzaklık Temelli ve Oto İlişkili olmak üzere iki farklı şekilde ele alınmıştır. Bir simülasyon verisi üzerinde Yöntem, Redundancy Analize Kanonik Uyum Analizi ile karşılaştırılmıştır. Sonuçta, Çok Değişkenli Regresyon Ağaçları Yönteminin, diğer yöntemlere göre avantajlı olduğu vurgulanmıştır.

Larsen ve Speckman (2004) yaptıkları çalışmada, jeoloji, toprak tipi ve arazi türü gibi çevresel değişkenler ile bitki türlerinin varlığı (bolluk veya çokluk, abundance) arasındaki ilişkileri belirlemeyi amaçlamışlardır. Cevap değişkeni olarak kullanılan tür gruplarını, en fazla etkileyen faktörleri belirlemek için Çok Değişkenli Regresyon Ağacı yöntemini kullanmışlardır. 12 türe özgü yapılan analiz sonrasında, Çok Değişkenli Regresyon Ağacı sonuçlarının yorumlanmasının kolay olduğu, yöntemin kullanılabilir olduğu ve iyi bir performans gösterdiği vurgulanmıştır.

Questier ve ark. (2005), Denetimli (supervised) ve Denetimsiz (unsupervised) yöntemlerden özellik seçimine (FutureSelection) dayalı Çok Değişkenli Regresyon Ağaçları (MRT) ile Sınıflandırma ve Regresyon Ağaçlarını (CART) tanımlayarak teorik bilgiler vermişlerdir. Orijinal değişkenlerin (x) sadece bağımsız değişken olduğunda değil, aynı zamanda cevap değişkeni olduğunda ($x = y$) Oto-ilişkili Çok Değişkenli Regresyon Ağaçlarını önermişlerdir. Sonuç olarak, dört farklı veri seti üzerinde bu yöntemleri göstererek yöntemlerin avantajlarından bahsetmişlerdir. Bu avantajları ise: parametrik olmayan yöntem olması nedeniyle varsayım gerektirmemesi, hızlı ve basit olması, kolay yorumlanabilir olması, eksik gözlem olması durumunda kullanılabilir olması, gürültülü açıklayıcı ve cevap değişkenlerine ve outlier (sapan) değerlere karşı sağlam (robust) olması ayrıca çapraz geçerlik model seçimine olanak sağlaması olarak özetlemişlerdir.

Smyth ve ark. (2006), pek çok veri setinde ortak olan bir yapının "gürültü değişkenleri" olduğunu ve gürültü değişkenlerinin, ilgili verileri içermediği gibi veri kümesinin altında yatan yapıyı da gizlediğini belirtmişlerdir. Veri kümelerindeki boyut arttıkça, gürültülü değişkenler de arttığından bu probleme bir çözüm olarak faktör analizörleri ile otomatik ilişkili Çok Değişkenli Regresyon Ağaçlarından oluşan karışık modelleri önermişlerdir. Temel Bileşen Skorları olan Çok Değişkenli Regresyon Ağaçları ve Faktör Skorlu Çok Değişkenli Regresyon Ağaçlarını iki veri seti üzerinde uygulayarak, ağaç yöntemlerin üstün kümeleme sonuçları

verdiğini göstermişlerdir. Bu iki veri seti kullanılarak elde edilen sonuçlara göre, ağaç yöntemlerinin, diğer incelenen yöntemler arasında en istikrarlı yöntemler olduğu vurgulanmıştır.

Hsiao ve Shih (2007), Tek değişkenli yöntemlerden elde edilen sonuçlar göz önüne alındığında, Çok Değişkenli Regresyon Ağaçlarında kullanılan bölme yönteminin, hala seçim yanlılığı olduğuna inanıldığını vurgulamışlardır. Bu probleme ilişkin bölen değişken olarak sıralı veya kategorik öngörü kullanabilen tarafsız bir seçim yöntemini geliştirmişler ve bir simülasyon çalışmasında sonuçları göstermişlerdir. Sonuç olarak, artıkların olasılık tablosu analizi ile, her düğümde sabit veya doğrusal modele uyabilen Çok Değişkenli Regresyon Ağaçları için tarafsız bir değişken seçme yöntemini önermişlerdir. Önerilen yöntemin, Guide'ın seçim yönteminin kısa bir uzantısı (Loh, 2002) olduğu ve Guide'ın etkileşim(interaksiyon) testinin, bu seçme yöntemini Çok Değişkenli Regresyon Ağacı Yöntemi için daha eksiksiz yapmak amacıyla, bu uzantıya eklenebileceğini vurgulamışlardır.

Ruppert ve ark. (2010), 2 farklı dönemde (1991 ve 1995 yılları) St. Lawrence Körfezi'ndeki topluluk bileşiminin gözlenen yapılarında tür zenginliğine çevresel değişkenlerin etkisini belirlemek amacıyla, Çok Değişkenli Regresyon Ağacı (MRT) yöntemini kullanmışlardır. Yöntem ile farklı okyanus iklimi dönemlerinde çevresel koşulların ve türlerin bolluğunun, toplum kompozisyonu dağılışını nasıl etkilediği ve çevresel değişkenlik ile toplum kompozisyonu modelleri arasındaki ilişkinin farklılıkları değerlendirmişlerdir. Sonuç olarak, St. Lawrence Körfezi'ndeki topluluk dinamiklerinin, tür çeşitliliği, tuzluluk ve sıcaklıkla yakından ilişkili olduğu vurgulanmıştır.

Haman ve ark. (2011), random forest yöntemiyle nötr ve uyarlanabilir genetik varyasyonunu coğrafik şekillerle analiz etmek ve ayrıca bu genetik bilgiyi tohum transfer önerilerinin gelişimi ve tohum alanlarının tarifi için coğrafik değişkenlerle birleştirmek için yeni bir yaklaşım tanımlamışlardır. Genetik veri setini ikiye bölmede çevresel veya coğrafik tahmin edici değişkenleri kullanarak, genetik varyasyonun analizinde Çok Değişkenli Regresyon Ağaçları yöntemini kullanmışlardır. Daha önce yapılan çalışmaların, Çok Değişkenli Regresyon Ağacı yönteminin karmaşık genetik bilgiyi bütünleştirerek ve görselleştirerek öznel kararları nasıl destekleyebileceğini gösterdiğinden hareketle, bu yöntemin riskten kaçınma ve büyüme optimizasyonu stratejileri için uygulanabilir olduğunu ve tohum alanlarının genetik analizi ve coğrafi haritalama işlemleri için uygun olabileceğini belirtmişlerdir.

Cannon (2012), otomatik, hiyerarşik ve kural tabanlı bir kümeleme algoritması olan Çok Değişkenli Regresyon Ağacı (MRT) yöntemini kullanarak küresel bir iklim sınıflaması modeli tanımlamıştır. Homojen sıcaklık ve yağış bölgelerini tanımlama yeteneği açısından sonuçlar K"oppenGeiger sistemiyle karşılaştırmıştır. Sonuçta, MRT algoritması; otomatik, hiyerarşik ve kural tabanlı olduğundan, algoritmanın iklim sınıfı sisteminin hızlı bir şekilde tanımlanmasına ve kolayca yorumlanmasına olanak sağladığı ifade edilmiştir. Ayrıca, sınıflandırma performansının genel analizinden elde edilen sonuçlara göre MRT yönteminin, özellikle yağış önlemleri için K"oppenGeiger sınıflandırmasından önemli derecede daha iyi performans gösterdiği belirtilmiştir.

Skipper (2014), yaptığı çalışmada, bir lise istatistik müfredatı kapsamında teknoloji kullanımının zamansal sıklığını incelemiştir ve bu kullanımda Çok Değişkenli Regresyon Ağaçları yöntemleri ile araştırma kümeleri üretmiştir. Çalışmada, Öklid uzaklığı ve Kullback Leibler farkının iki versiyonu incelenerek Öklid kümelerinin sapan gözlemlere karşı daha sağlam/dirençli ve daha az çapraz geçerlilik hatasına sahip olduğu vurgulanmıştır.

3. GEREÇ VE YÖNTEM

3.1. Gereç

Çalışmanın uygulama aşamasında; yöntemin uygulanabilirliğinin gösterilmesi ve elde edilecek sonuçlar için, biyolojik yorum yerine, yöntemle ilişkin istatistik yorumlamaların yapılması amaçlanmıştır. Bu nedenle çalışmada gerçek veriler yerine, hipotetik veriler kullanılmıştır. Hipotetik verilerin oluşturulmasında, Suvak ve ark., (2017) tarafından yapılan çalışmaya benzer bir senaryo dikkate alınmıştır. Bunun için Van, Bitlis ve Ağrı illerine ait 21 ilçeden 95 hasta alındığı varsayılmıştır. Bu hastaların; Yaş (yıl), Beyaz Küre [WBC, (K/mm³)], Hemoglobin [HB, (g/dL)], Alanin Aminotransferaz [(ALT, (g/dL)], Aspartat Aminotransferaz [AST (g/dL)], Trombosit [PLT, (K/mm³)], Albümin [ALB, (g/dL) ve Globülin [GLB, (g/dL)] özellikleri açıklayıcı değişken olarak alınmıştır. İlçelerde görülen Brucella, Hepatoselüler karsinom (Hepatocellular carcinoma, HCC) ve Siroz (Cirrhosis) hastalıklarının varlığı ise cevap değişkenleri olarak varsayılmıştır. Çalışmaya dahil edilen açıklayıcı değişkenler için; bu değişkenlerin gerçek değerlerine benzer ortalama ve standart sapmalar kullanılarak, bilgisayar ortamında hipotetik veriler oluşturulmuştur. Açıklayıcı değişkenlerden Trombosit (Platelet) değişkeni, 100'e bölünerek analize dahil edilmiştir.

3.2. Yöntem

3.2.1. Sınıflandırma ve regresyon ağacı

İlk defa Breiman ve ark. (1984), tarafından tanımlanan ve sınıflandırma veya modelleme amaçlı kullanılabilen sınıflandırma ve regresyon ağacı (CART) yönteminde, tek bir bağımlı değişkendeki varyasyonu açıklamak için bağımsız değişkenler setine dayalı ikili bölünme (binary partition) işlemi ya da prosedürü uygulanır.

CART yönteminde, cevap değişkeni olarak kategorik ya da sürekli değişken alınabilir. Cevap değişkeninin kategorik olması durumunda yöntem; sınıflandırma ağacı, sürekli olması durumunda ise regresyon ağacı olarak adlandırılır. Diğer yandan, cevap değişkeninin yalnızca bir değişken olması durumunda yöntem, tek değişkenli sınıflandırma ve regresyon ağacı olarak ifade edilirken, bir den fazla cevap değişkeni olması durumunda çok değişkenli regresyon ağacı olarak ifade edilir.

3.2.2. Tek deęişkenli regresyon ağacı

CART, veri setini karşılıklı olarak düęüm (node) adı verilen alt gruplara ayırır ve bu düęümlerde cevap deęişkeni bakımından birbirine oldukça yakın veya benzer nesnelere yer alır. Ağacın başlama noktası, kök veya ebeveyn düęüm olarak adlandırılan düęümdür. Bu düęüm, veri setindeki tüm nesnelere ya da bireyleri içerir. Daha sonra tekrarlı ikili bölünme süreci (binary splitting procedure) uygulanarak veri seti iki gruba ayrılır. İki gruba ayrılma ile oluşan yeni düęümleri birbirinden ayırmak üzere, sağ ve sol düęüm ifadeleri kullanılır. Bölünme ile oluşturulan bu yeni düęümlere (gruplara) alt grup ya da çocuk düęüm (child node) adı verilir. Bu süreç, özyinelemeli (recursive) olup, her bir alt grubun ebeveyn veya kök düęüm olarak değerlendirilmesi sürecine kadar tekrar edilir. Her bir bölünme, tek bir açıklayıcı deęişkene göre bir kesim noktası, yani rakamsal bir deęer bulunarak veya açıklayıcı deęişkenler kategorik deęişken ise bu deęişkenlerin bir veya birden fazla kategorisi ile ilişkili olarak yürütülür. Buna göre bölünmede aşağıdaki kurallar dikkate alınır.

- a) X_j , sürekli bir deęişken olmak üzere “ t ” bu deęişken için kesim noktası ve herhangi bir düęüm de “ D ” ile gösterildiğinde; $D_{sol} = \{i \in \mathbb{N} : x_{ij} \leq t\}$, $D_{saę} = \{i \in \mathbb{N} : x_{ij} > t\}$ olarak bölünme gerçekleştirilir (Larsen ve Speckman, 2004).
- b) X_j , ordinal yada sıralı deęişken ise bölünme “a” daki kurala göre gerçekleşir.
- c) X_j , L adet seviye içeren bir kategorik deęişken ise tüm 2^L alt kümeleri göz önünde bulundurulur. Sağ ve sol alt gruplara her bölünme için ters çevrilmiş alt setler ile eşdeęer bölünme yapılır. Böylece $2^L - 1$ bölünme dikkate alınır (Larsen ve Speckman, 2004).

Her bir düęüm için mümkün olan bütün bölünmeler değerlendirilir. Modelde yer alan tüm tahmin edici deęerler ve bunların olası eşik deęerleri test edilir. Sonuç olarak, en iyi bölünmeyi yapan dikkate alınarak bunda karar kılınır. En iyi bölünme deęeri, açıklayıcı deęişken bakımından, cevap deęişkeni için alt gruplardaki (çocuk düęümlerdeki) heterojenlięi ifade eden ve safsızlık (impurity) olarak adlandırılan düęümler içerisindeki varyasyonu minimize edecek bölünme deęeri olarak tanımlanır. Herhangi bir bölünme için uyum iyilięi, ebeveyn veya kök düęüm ile bundan oluşturulan alt gruplar arasındaki heterojenlik veya safsızlık deęerindeki azalma olarak tanımlanır.

$$\Delta i(S, t_p) = i_p(t_p) - p_{sol}i(t_{sol}) - p_{saę}i(t_{saę}) \quad (1)$$

Eşitlikte (1)' de s aday bölünme, p_{sol} ve $p_{sağ}$ sırasıyla; t_p ebeveyn düğümünden, t_{sol} ve $t_{sağ}$ alt gruplarına (çocuk düğümlerine) giden gözlemlerin kesirleridir (fraksiyonlarıdır). En iyi ayırıcı (splitter), $\Delta i(S, t_p)$ değerini maksimize eden ayırıcı ya da bölendir (Put ve ark., 2005).

Heterojenlik ölçüsü: Düğümlerin heterojenlik ölçüsü olarak farklı kriterler önerilmiştir. Regresyon ağaçları için cevap değişkenin ortalaması ile ilişkili olarak düğümlerin ortalamasına ait genel kareler toplamı, en yaygın heterojenlik ölçüsü olarak kullanılmaktadır.

$$i(t) = \sum_{x_n \in t} (y_n - \bar{y}(t))^2 \quad (2)$$

Eşitlikte $i(t)$, t . düğümün heterojenliği, y_n t . noda ait X_n gözlemi için cevap değişkeninin değeri ve $\bar{y}(t)$, t . düğümde bulunan tüm gözlemlerin ortalamasıdır.

Diğer bir kriter ise biraz daha sağlam ya da robust olarak adlandırılan düğümleri medyanın değerine bağlı mutlak sapmalardır (Absolute deviations) (Put ve ark., 2005).

Ağacın her bir düğümüne bir isim veya sınıf atanır. Regresyon ağacı için bu basitçe düğüm için ortalama olarak ifade edilir.

Ayırma ya da bölünme süreci, durdurma kriterine (stopping criterion) ulaşıncaya kadar devam eder. Yani bütün alt gruplar (çocuk düğümler) homojen olana kadar ya da düğüm içerisindeki nesne sayısı araştırmacı tarafından belirlenen sayıya ulaşıncaya kadar devam eder. Böylece ağaç maksimum ağaç veya eğitim veri setindeki en iyi oluncaya kadar devam eder. Ancak bu ağaç için genel olarak aşırı uyum (overfit) söz konusudur. Diğer bir ifadeyle bu maksimum ağaç, yeni örnekler için zayıf bir tahmin etme yeteneğine sahiptir. Eğitim veri seti tanımlanmış ise maksimum ağaç en iyi seçimdir. Tahmin amaçlı durumlarda optimum ağaç, maksimum ağaçtan budama prosedürüne dayalı olarak elde edilen daha küçük ağaç olur. Budama süreci maksimum ağacın terminal dallarındaki yani uç düğümlerindeki en kötü dalların kesilerek uzaklaştırılmasına dayanır. Genel olarak optimum ağaç, çapraz geçerlilik yöntemlerine ya da bağımsız testlere dayalı yöntemlere bağlı olarak belirlenir veya seçilir.

3.2.3. Çok Değişkenli Regresyon Ağacı (Multivariate Regression Trees, MRT)

Tek değişkenli regresyon ağaçlarında yalnızca bir cevap değişkeni dikkate alınırken, çok değişkenli regresyon ağaçlarında birden fazla cevap değişkeni eş zamanlı olarak değerlendirilir. Çok değişkenli regresyon ağacında, ikili ayrılmalar ya da bölünmeler için çok değişkenli karar verme kuralları yoktur. Ancak cevap değişkeninin birden fazla olduğu unutulmamalıdır. Çok değişkenli regresyon ağacının oluşturulması, tek değişkenli regresyon ağacının oluşturulmasına benzerlik göstermekte ve birçok parametresi de aynı şekilde tanımlanmaktadır. Ancak heterojenlik ölçüsünde değişikliklerin yanı sıra, alt grupların (çocuk düğümlerin) adlandırılmasında da modelde kullanılan cevap değişkenlerine göre bir miktar farklılıklar oluşmaktadır. CART ile karşılaştırıldığında, tekrarlı ikili ayrışma sürecinin aynı şekilde kullanıldığı söylenebilir. Ancak burada CART' tan farklı olarak Çok Değişkenli Regresyon Ağacı'nda, çok değişkenli heterojenlik ölçüsü, Karesele Öklid uzaklığına bağlı olarak hesaplar.

Geometrik olarak, Karesele Öklid uzaklığı basitçe, düğüm sentroidi etrafında yer alan nesnelere, diğer bir ifade ile düğümleri oluşturan nesnelere ya da gözlemlerin sentroid merkezine olan Karesele Öklid uzaklığı olarak ifade edilir. Her birisinden p adet değişkene ait değer elde edilmiş n adet gözlem içeren herhangi bir düğüm için heterojenlik;

$$impurity = \sum_{i=1}^n \sum_{j=1}^p (y_{ij} - \bar{y}_j)^2 \quad (3)$$

olarak ifade edilir (Questier ve ark., 2005).

Heterojenlik ölçüsü, düğümlerde yer alan gözlemlere ait cevap değişkeni bakımından genel kareler toplamıdır. Yani cevap değişkenine ait değerlerin, düğümlere ait çok değişkenli ortalamadan (düğüm sentroidinden) olan sapma kareler toplamıdır. Böylece CART ve MRT'nin her ikisinde de karar ağacı, her bir düğüm içerisindeki heterojenliğin minimize edilmesi, düğümler arasındaki varyasyonun ise maksimize edilmesi temeline dayalı olarak oluşturulur.

Oluşturulan bir ağacın verilere uygunluğunun belirlenmesinde önemli ölçülerden birisi göreceli yada rölatif hatadır (Relative Error, RE). RE, her bir düğümde, Grup İçi Kareler Toplamının, Genel Kareler Toplamına oranı olarak tanımlanır. Böylece, RE model sonrasında, ağaç tarafından açıklanamayan varyasyonun bir ölçüsüdür.

Ancak, grup içi kareler toplamını minimize etmek her zaman yeterli olmayabilir. Bunun için, De'ath (2002) yeniden örnekleme ve çapraz doğrulama yaparak ağacı budamayı (yani dalları çıkarmayı) önermektedir. Bunun için veriler rastgele olarak, tümü kabaca eşit büyüklükteki önceden tanımlanmış sayıda gruba atanır. Grupların her biri, sırasıyla, analizden çıkarılır ve kalan verilerden (“eğitim” verileri olarak adlandırılır) bir ağaç oluşturulur. Hariç tutulan gruptan (“test” grubu olarak adlandırılır) gözlemler, daha önce eğitim verilerinden türetilen ayırma kurallarını izleyerek ağacın terminaldüğümlerine ayrı ayrı atanır.

Bir yaprağın yada düğümün sentroidi ile test gözlemi arasında bir uzaklık hesaplanır. Bu uzaklık, çapraz geçerlilik rölatif hata (Cross validation relative error, CVRE) olarak adlandırılır ve aşağıdaki eşitlik ile hesaplanır.

$$CVRE = \sum_{k=1}^K \frac{\sum_{i=1}^n \sum_{j=1}^p (y_{ij(k)} - \hat{y}_{j(k)})^2}{\sum_{i=1}^n \sum_{j=1}^p (y_{ij} - \bar{y}_j)^2} \quad (4)$$

(4) no'lu eşitlikte $Y_{ij(k)}$, k test setindeki gözlem değeri, $\hat{y}_{j(k)}$ bu gözlem için değişken j'nin tahmini değeri (tahmini yaprağın merkezi), \bar{y}_j , j değişkeni için genel ortalama, n gözlem sayısı, p yanıt matrisi Y'deki değişken sayısıdır.

Uygun ağaç boyutu, toplam yaprakların (1, 2, 3, vb.) ardışık sayıları için CVRE'nin hesaplanmasından elde edilebilir. Çoğu veri kümesi için CVRE ilk önce azaltılmalıdır.

Sapma (deviance) istatistiği: İki olasılık modeli arasındaki uzaklığı ifade etmede kullanılır. Bu istatistik, çoğunlukla model karşılaştırmalarında kullanılmakla birlikte, standart regresyon analizi ve varyans analizinde de kullanılmaktadır. Yalnızca sabit terimi içeren modelle karşılaştırıldığında, ilgili modelin uyum iyiliğinin bir ölçüsü olarak kullanılır. Sınıflandırma ve regresyon ağaçları yönteminde sapma (deviance) ise n adet bireylerin atandığı sınıflarda kullanılır.

Ağaç, bireylerin sınıf üyeliği için yani ilgili düğüme atanması için olasılığın bir ölçüsü olarak değerlendirildiğinde, her bir düğüm için olasılık dağılımı bütün gruplarda ya da alt gruplarda hesaplanabilir. Böylece ağacın yaprakları rastgele örneklem olarak multinomiyal dağılım gibi düşünüldüğünde, ağacın sapması (deviance), bütün yapraklar üzerinden bu olasılıkların toplamı olarak

$$D_i = -2 \sum_k n_{ik} \log p_{ik} \quad (5)$$

eşitliği ile ifade edilir.

Benzer şekilde regresyon ağacında ise bu istatistik j. birey için kareler toplamı olarak

$$D_i = \sum_j (y_j - \mu_i)^2 \quad (6)$$

ile ifade edilir.

Ağaç oluşturma süreci hiyerarşik olasılık modellerine benzer olduğundan regresyon analizindeki ileri (forward) doğru değişken seçimi yöntemine oldukça benzerdir.

Regresyon ağacında en iyi modelin mümkün olan sayıda parametre içeren ve cevap değişkenindeki varyasyonu açıklayan model olduğu düşünüldüğünde, bu modele doymuş model adı verilir. Dolayısıyla sapma (deviance), basitçe aday model ile doymuş model arasındaki uyum bakımından farkı ifade eder. Dolayısı ile regresyon ağacında doymuş model birçok terminal düğümü içermekte ve böylece cevap değişkenine en iyi uyumu sağlamaktadır. Regresyon ağacında daha basit bir hesaplama ile sapma (deviance) istatistiği, düğüm içerisindeki artık kareler toplamının tüm düğümler üzerinden toplanması ile bulunur. Diğer bir ifade ile gözlenen değerler ile beklenen değerler arasındaki farklara ilişkin kareler toplamıdır. En küçük kareler regresyonunda ise hata kareler toplamıdır.

Sınıflandırma ağacında ise artıklara ait kareler toplamı uyum eksikliğinin (lack of fit) en iyi ölçüsü olarak uygun olmadığından bunun yerine alternatif sapma (deviance) ölçüsü olarak entropi ölçüsünü minimize etmeye dayalı Gini İndeksi kullanılır.

MRT 'de her bir cevap değişkeni için etiketler histogram ile gösterilir. Böylece verilen düğünde cevap değişkenlerinin dağılımı görülür.

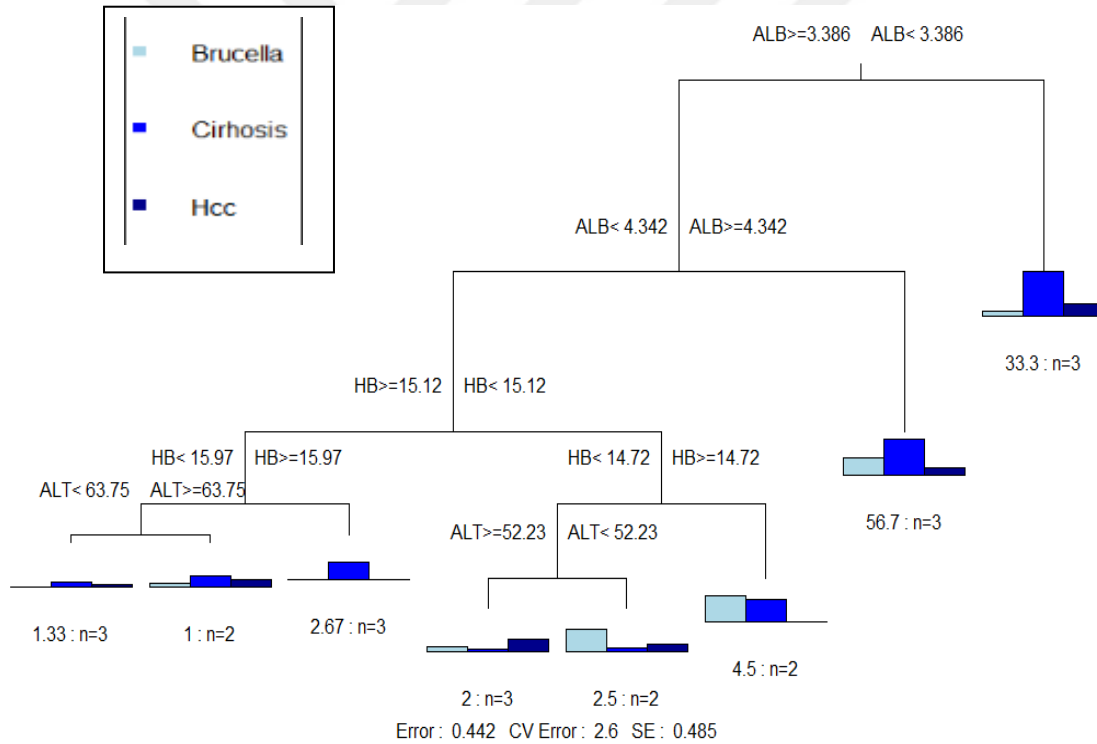
Bu çalışmada analizler R programlama dilinde yürütüldü. Ağacı inşa etmek için Mvpart paketi kullanılırken, sapma (deviance) istatistiğinin hesabı için Mvpartwrap paketi kullanıldı.

4. BULGULAR

Uygulama aşamasında ilk önce modele 8 adet açıklayıcı değişkenin [Yaş, Beyaz Küre (WBC), Hemoglobin (HB), Alanin Aminotransferaz (ALT), Aspartat Aspartat Aminotransferaz (AST), Trombosit (PLT), Albümin (ALB) ve Globülin (GLB)]tamamı dahil edilerek ağaç oluşturulmuştur. Sonraki adımlarda ise; modele en çok katkı sağlayan değişkenler sırası ile çıkartılarak, kalan değişkenler ile ağaç oluşturulmuştur.

Uygulama 1: Tüm değişkenlerin modele dahil edilmesi ile oluşan çok değişkenli regresyon ağacı

Tüm değişkenlerin modele dahil edilmesi ile oluşan çok değişkenli regresyon ağacı Şekil 1’de, bu ağaca ilişkin sonuçlar ise Tablo 1, 2 ve 3’ te verilmiştir.



Şekil 1. Tüm değişkenlerin modele dahil edilmesi ile oluşan çok değişkenli regresyon ağacı

Şekil 1’ de görüldüğü üzere; modele tüm değişkenler dahil edildiğinde, ilk bölünmenin ALB değişkenine göre olduğu görülür. Buna göre ALB için kesim değeri 3.386 olarak bulunmuş olup, bu değerden küçük olanlar sağ düğüme, büyük olanlar ise sol düğüme ayrılacak şekilde bölünme gerçekleşmiştir. Sol düğüm, ALB değişkenine göre iki alt gruba ayrılırken; sağ düğüm, 33.3 rölatif hata değerine sahip 3 ilçe ile terminal düğüm olmuştur. Sol düğümü iki alt düğüme ayıran ALB değişkeni için kesim değeri 4.342 olarak bulunmuştur.

ALB' ye göre ikiye ayrılan düğümlerden sol düğüm HB değişkenine göre yeniden iki alt düğüme ayrılırken, sağ düğüm 56.7 rölaf hataya sahip 3 ilçe ile terminal düğüm olmuştur.

ALB' ye göre ikiye ayrılan düğümlerden sol düğümde HB için kesim değeri 15.12 olarak bulunmuştur. HB değeri 15.12 den büyük olan soldaki düğüm, HB değişkenine göre yeniden iki alt düğüme bölünürken, sağdaki düğüm de yine HB değişkenine göre bölünmüştür. HB değeri 15.12 den büyük olan soldaki düğümü ikiye bölen HB değişkeni için kesim noktası 15.97 olarak gerçekleşirken; HB değeri 15.12 den küçük olan sağdaki düğümü ikiye bölen HB değişkeni için kesim noktası 14.72 olarak gerçekleşmiştir.

Solda yer alan HB değişkenine göre sağ ve sol düğüme ayrılan alt düğümlerden (alt gruplardan) soldaki düğüm, ALT değişkenine göre 2 alt düğüme ayrılırken; sağdaki düğüm 2.67 rölaf hata ve 3 ilçe ile terminal düğüm olmuştur. Sağda yer alan HB değişkenine göre sağ ve sol düğüme ayrılan alt düğümlerden (alt gruplardan) soldaki düğüm, ALT değişkenine göre 2 alt düğüme ayrılırken; sağdaki düğüm 4.5 rölaf hata ve 2 ilçe ile terminal düğüm olmuştur.

Solda yer alan ALT değişkeni için kesim değeri 63.75 olarak gerçekleşirken, sağda yer alan ALT değişkeni için kesim değeri 52.23 olarak gerçekleşmiştir. Solda yer alan ALT' ye göre bölünen alt gruplardan soldaki düğümde, 1.33 rölaf hata ile 3 ilçe yer alırken, sağdaki düğümde 1 rölaf hata ile 2 ilçe yer almıştır. Sağda yer alan ALT' ye göre bölünen alt gruplardan soldaki düğümde, 2 rölaf hata ile 3 ilçe yer alırken, sağdaki düğümde 2.5 rölaf hata ile 2 ilçe yer almıştır.

Modele dahil edilen 8 adet değişken ile oluşturulan ağaç için "Error" olarak verilen değer artıklara ait hatadır (Residual error). Diğer bir ifade ile ağaca ait genel rölaf hata 0.442 olarak bulunmuştur. Rölaf hatanın 1-R2 değerine eşit olduğu hatırlanacak olursa, modele ait R2değerinin % 55.8 olarak bulunduğu söylenebilir.

Düğümleri ayıran değişkenlerin yanısıra, diğer değişkenlerin de düğümde yer alan kesim değerleri, düğümlerdeki gözlem sayıları, ilgili düğüme düşen cevap değişkenlerine ait ortalamalar ve Toplanmış hata kareler ortalmaları Tablo 1' de verilmiştir.

Böylece ALB değişkenine göre ayrılan ikinci düğümde; sağ ve sol düğümleri ayırmada, GLB değişkeni için kesim değeri 4.30 olarak bulunurken, HB için 15.9, PLT ve AST için de sırasıyla 1580.6 ve 64.9 olarak bulunmuştur.

Bu düğümde yer alan 8 ilçe için cevap değişkenlerinden Brucellaya ait ortalama, 1.28 olarak bulunurken, Siroz ve Hcc' ye ait ortalamalar, sırasıyla 1.94 ve 0.72 olarak bulunmuştur. Bu düğüme ait toplanmış hata kareler ortalaması ise 8.68 olarak bulunmuştur. Tablo 2' verilen maliyet karmaşası (cost complexity) değeri, ağaç boyutunu kontrol altında tutmak için bir nevi ceza terimi gibi işlev görür ve bölünme sayısına bağlı olarak monotoniktir.

Tablo 2' verilen maliyet karmaşası (cost complexity) değeri, ağaç boyutunu kontrol altında tutmak için bir nevi ceza terimi gibi işlev görür ve bölünme sayısına bağlı olarak monotoniktir. Dolayısıyla bunun küçük değeri ağacın daha kompleks olacağını yani bölünme sayısının büyük olacağını gösterir. Tablo 2' de yer alan Çapraz Geçerlik Hatası (cross validation error), 10 kat çapraz geçerliğe dayalı olup, rölatif olarak yalnızca sabit terimi içeren modelden olan sapmayı belirtir. Çapraz Geçerlik Hatasının, rölatif hatadan büyük olması beklenen durumdur. Aynı veri seti için her ikisi de test ve fit edilen model sonuçları için iyimser bir teşhis ölçüsüdür. Tablo 2' de rölatif hata 0.80 ve 1.00 değerlerini alırken, Çapraz Geçerlik Hatası 1.10 ve 2.32, çapraz geçerlik standart hatası ise 0.32 ve 0.49 değerlerini almıştır.

Ağacın daha karmaşık olması ile birlikte rölatif hata düşer. Ancak, bu durum Çapraz Geçerlik Hatası için geçerli değildir.

Tablo 3' te tüm değişkenler modele dahil edildiğinde, cevap değişkenleri bakımından ağaç bölünmeleri için hesaplanan sapma değerleri verilmiştir. Tablo 3' te görüldüğü üzere; değerler, % 100 ile % 10.58 arasında değişim göstermiştir. Dolayısıyla, Siroz ve HCC için ALB'nin 4.342 den küçük olduğu bölünme ile oluşan düğümlerden Brucella için sapma değeri daha düşük bulunmuştur.

Buna göre adı geçen cevap değişkeni bakımından ilgili düğüme yapılacak atamalarda hatanın daha düşük olacağı diğer bir ifade ile doğruluk değerinin daha yüksek olacağı söylenebilir.

Tablo 1. Tüm deęişkenlerin modele dahil edilmesi ile oluşturulan tablo

	1. Düğüm	2. Düğüm	3. Düğüm	4. Düğüm	5. Düğüm	6. Düğüm
	ALB <3.386 ise R	ALB <4.342 ise L		HB <15.12 ise R		HB <15.97 ise L
	HB <14.5 ise L	GLB <4.30 ise L		PLT <1574.7 ise L		PLT <1579.5 ise L
	GLB <3.90 ise L	HB <15.9 ise L		AST <77.9 ise L		ALB <3.8 ise R
	ALT <52.8 ise R	PLT <1580.6 ise R		WBC <5577.7 ise R		GLB <3.79 ise R
	PLT <1750.9 ise L	AST <64.9 ise R		YAS <52.9 ise L		YAS <43.9 ise L
Gözlem Sayısı	21	18	3	15	3	8
Ortalama	1.19, 2.52, 0.86	1.28, 1.94, 0.72	0.67, 6, 1.67	1.07, 1.33, 0.67	2.33, 5, 1	0.12, 1.5, 0.37
Toplanmış Hata Kareler Ortalaması	11.19	8.68	11.1	4.1	18.9	1.34

Tablo 1. (Devam)

	7. Düğüm	8. Düğüm	9. Düğüm	10. Düğüm	11. Düğüm	12. Düğüm	13. Düğüm	14. Düğüm	15. Düğüm
	HB <14.72 ise L	ALT <63.75 ise L		ALT <52.23 ise R					
	WBC <5669 ise R	AST <71.6 ise L		PLT <1622 ise R					
	ALT <52.2 ise R	PLT <1536 ise L		GLB <4.04 ise L					
	PLT <1623 ise R	WBC <6203 iseL		AST <68.5 ise R					
	AST <77.8 ise L	HB <15.36 ise L		ALB <3.87 ise R					
Gözlem Sayısı	7	5	3	5	2	3	2	3	2
Ortalama	2.14, 1, 1	0.2, 1, 0.6	0, 2.33, 0	1.6, 0.4, 1.4	3.5, 3, 0	0, 0.67, 0.33	0.5, 1.5, 1	0.67, 0, 1.67	3, 0.5, 1
Toplanmış Hata Kareler Ortalaması	4.82	0.8	0.89	2.32	2.25	0.44	0.5	0.67	1.25

Tablo 2. Tüm değişkenlerin modele dahil edilmesi ile oluşan hatalar

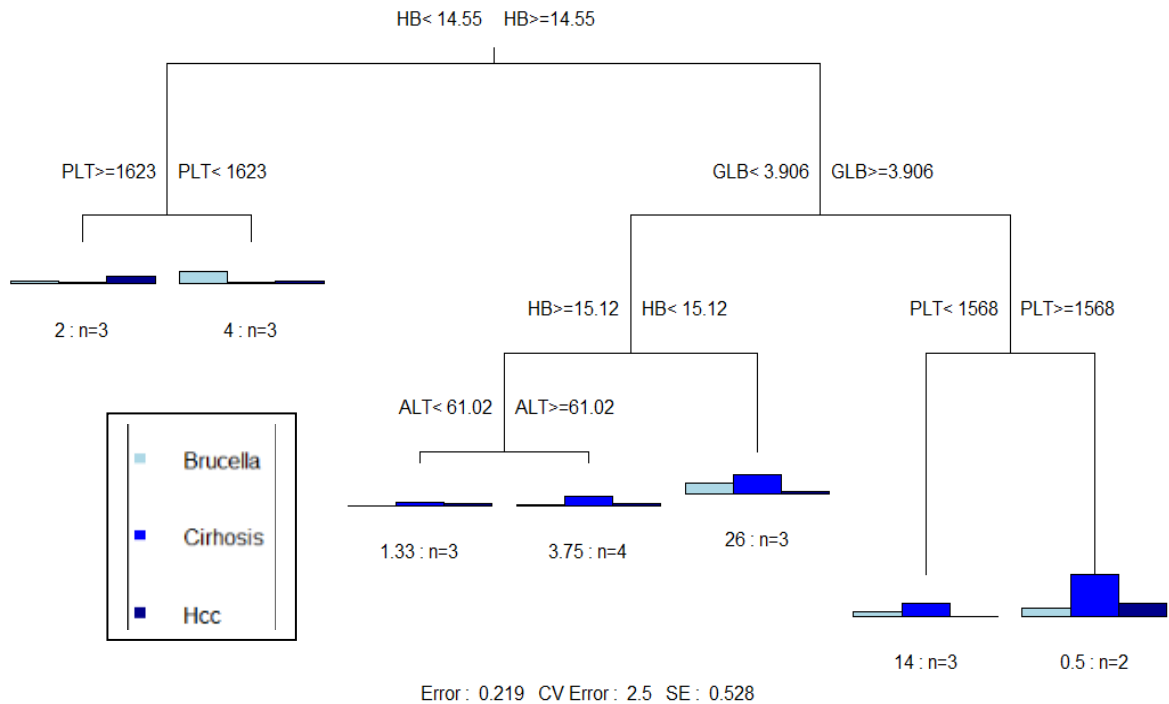
D	Maliyet Karmaşası (Cost complexity)	n bölünme	Rölatif Hata	Çapraz Geçerlik Hatası	Çapraz Geçerlik Standart Hatası
1	0.19	0	1.00	1.10	0.32
2	0.16	1	0.80	2.32	0.49

Tablo 3. Tüm değişkenlerin modele dahil edilmesi ile ağaç bölünmeleri ve tüm ağaca ait sapma (deviance) değerleri (%)

	ALB<3.386	ALB<4.342	HB<15.12	HB<15.97	ALT<63.75	HB<14.72	ALT<52.23
Brucella	-	10.58	88.71	-	18.00	29.28	92.02
Cirhosis	92.85	88.68	-	81.63	50.00	54.82	-
Hcc	-	-	-	16.53	32.00	15.89	-
Toplam Sapma (Deviance)	92.85	99.26	88.71	98.16	100.00	99.99	92.02

Uygulama 2: ALB değişkeninin modelden çıkarılması ile oluşan çok değişkenli regresyon ağacı

ALB değişkeninin modelden çıkarılması ile oluşan çok değişkenli regresyon ağacı Şekil 2’de, bu ağaca ilişkin sonuçlar ise Tablo 4, 5 ve 6’ da verilmiştir.



Şekil 2. ALB değişkeni modelden çıkarıldığında oluşan çok değişkenli regresyon ağacı

Şekil 2' de görüldüğü üzere; ALB değişkeni analizden çıkarıldığında, ilk bölünmenin HB değişkenine göre olduğu görülür. Buna göre HB değişkeni için kesim değeri 14.55 olarak bulunmuş olup, bu değerden küçük olanlar sol düğüme, büyük olanlar ise sağ düğüme ayrılacak şekilde bölünme gerçekleşmiştir. Sol düğüm, PLT değişkenine göre iki alt gruba ayrılırken; sağ düğüm, GLB değişkenine göre iki alt düğüme ayrılmıştır.

Sol düğümü iki alt düğüme ayıran PLT değişkeni için kesim değeri 1623 olarak bulunurken, sağ düğümü ayıran GLB değişkeni için ise kesim değeri 3.906 olarak bulunmuştur. PLT değeri 1623'ten büyük olan soldaki düğüm, 2 rölatif hata değerine sahip 3 ilçe ile terminal düğüm olurken, PLT değeri 1623 ten küçük olan sağdaki düğüm, 4 rölatif hata değerine sahip 3 ilçe ile terminal düğüm olmuştur.

GLB değişkeni için kesim noktası 3.906 olarak gerçekleşmiş ve buna göre ayrılan iki alt gruptan soldaki düğüm HB değişkenine göre yeniden iki alt düğüme bölünürken, sağdaki düğüm de PLT değişkenine göre yeniden iki alt düğüme bölünmüştür.

PLT için buradaki kesim değeri ise 1568 olarak bulunmuştur. PLT değeri 1568 den küçük olan soldaki düğüm, 14 rölatif hata değerine sahip 3 ilçe ile terminal düğüm olurken, sağdaki düğüm, 0.5 rölatif hata değerine sahip 2 ilçe ile terminal düğüm olmuştur.

HB için ise buradaki kesim değeri ise 15.12 olmuştur. HB değeri 15.12 den büyük olan soldaki düğüm, ALT değişkenine göre yeniden iki alt düğüme bölünürken, sağdaki düğüm, 26 rölatif hata değerine sahip 3 ilçe ile terminal düğüm olmuştur.

ALT değişkeni için kesim noktası 61.02 olarak gerçekleşmiş ve buna göre ayrılan iki alt gruptan soldakinde 1.33 rölatif hata ile 3 ilçe yer almış, sağdakinde ise 3.75 rölatif hata ile 4 ilçe yer almıştır.

Böylece toplam olarak ağaç 7 terminal düğümden oluşmuştur. Bu modele ait R2 değeri ise %78,1 olarak bulunmuştur.

Modele ait çapraz geçerlik rölatif hata 2.5 olurken; standart hata değeri 0.528 olarak hesaplanmıştır.

Tablo 4. ALB deęişkeni modelden çıkarıldığında oluşturulan tablo

	1. Düğüm	2. Düğüm	3. Düğüm	4. Düğüm	5. Düğüm	6. Düğüm
	HB <14.55 ise L	PLT <1623 ise R	GLB <3.906 ise L			HB <15.12 ise R
	GLB <3.91 ise L	WBC <5669.1 ise L	ALT <33.9 ise R			WBC <4643.7 ise R
	ALT <52.8 ise R	ALT <52.2 ise R	PLT <1568.3 ise L			PLT< 1599.8 ise L
	PLT <1759 ise L	YAS <51 ise L	WBC <6443.3 ise L			YAS <61.7 ise L
	WBC <8000 ise R	HB <14.28 ise R	HB <15.05 ise R			ALT <42.69 ise R
Gözlem Sayısı	21	6	15	3	3	10
Ortalama	1.19, 2.52, 0.86	1.67, 0.33, 1.17	1, 3.4, 0.7	0.67, 0.33, 1.67	2.67, 0.33, 0.67	0.8, 2.3, 0.5
Toplanmış Hata Kareler Ortalaması	11.2	2.25	11.9	0.67	1.33	6.22

Tablo 4. (Devam)

	7. Düğüm	8. Düğüm	9. Düğüm	10. Düğüm	11. Düğüm	12. Düğüm	13. Düğüm
	PLT <1568.3 ise L	ALT <61.02 ise L					
	YAS <49.8 ise R	PLT <1559.4 ise L					
	HB <16.3 ise R	YAS <43.9 ise L					
	ALT <59.2 ise R	WBC <5776.4 ise L					
	WBC <6667 ise L	HB <15.36 ise L					
Gözlem Sayısı	5	7	3	3	2	3	4
Ortalama	1.4, 5.6, 1.2	0.14, 1.4, 0.43	2.33, 4.33, 0.67	1, 3, 0	2, 9.5, 3	0, 0.67, 0.33	0.25, 2, 0.5
Toplanmış Hata Kareler Ortalaması	15.4	1.18	8.67	4.67	0.25	0.44	0.94

Tablo 5. ALB değışkeni modelden çıkarıldığında hesaplanan hata çizelgesi

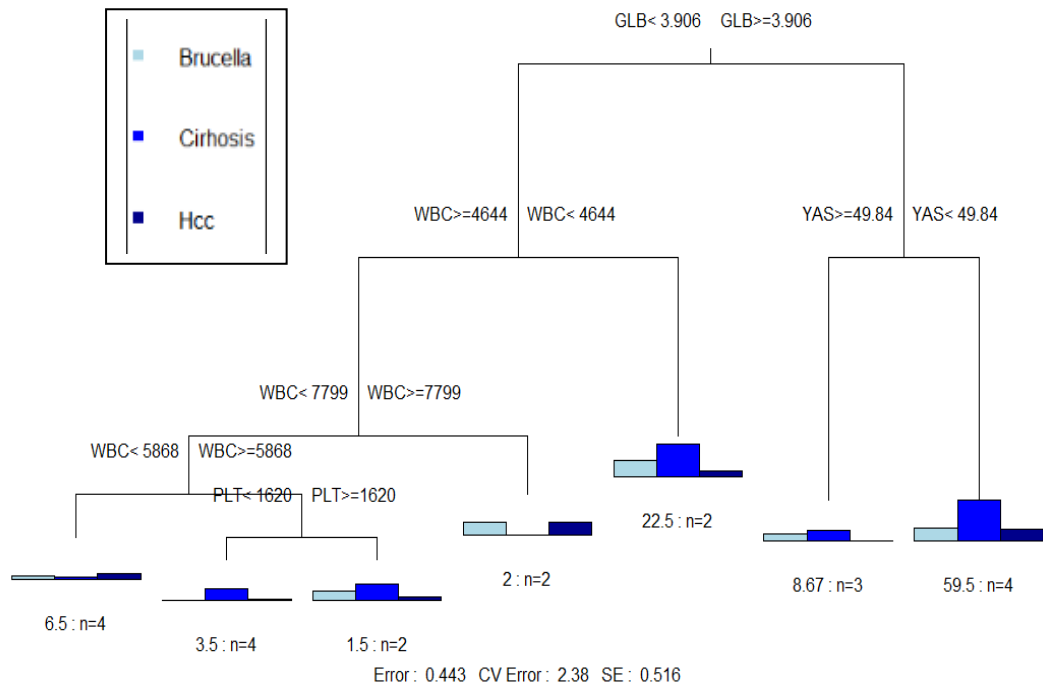
	Maliyet Karmaşası	n bölünme	Rölatif Hata	Çapraz Geçerlik Hatası	Çapraz Geçerlik Standart Hatası
1	0.21	0	1.00	1.11	0.32
2	0.12	3	0.38	2.28	0.50
3	0.03	4	0.27	2.42	0.53
4	0.01	5	0.23	2.46	0.53
5	0.00	6	0.22	2.48	0.52

Tablo 6. ALB değışkeni modelden çıkarıldığında ağaç bölümleri ve tüm ağaca ait sapma (deviance) değeri (%)

	HB<14.54	PLT<1623	GLB<3.906	HB<15.12	ALT<61.02	PLT<1568
Brucella	-	80.00	-	36.09	-	-
Cirrhosis	93.70	-	92.75	63.48	95.17	80.86
Hcc	-	20.00	-	-	-	17.23
Toplam Tür Sapması	93.70	100.00	92.75	99.57	95.17	98.09

Uygulama 3: ALB ve HB değışkenlerinin modelden çıkarılması ile oluşan çok değışkenli regresyon ağacı

ALB ve HB değışkenlerinin modelden çıkarılması ile oluşan çok değışkenli regresyon ağacı Şekil 3'te, bu ağaca ilişkin sonuçlar ise Tablo 7, 8 ve 9' da verilmiştir.



Şekil 3. ALB ve HB değışkenleri modelden çıkarıldığında oluşan çok değışkenli regresyon ağacı

Şekil 3' te görüldüğü üzere; ALB ve HB değişkenleri modelden çıkarıldığında, ilk bölünmenin GLB değişkenine göre olduğu görülür. Buna göre GLB için kesim değeri 3.906 olarak bulunmuş olup, bu değerden küçük olanlar sol düğüme, büyük olanlar ise sağ düğüme ayrılacak şekilde bölünme gerçekleşmiştir. Sol düğüm, WBC değişkenine göre iki alt gruba ayrılırken; sağ düğüm, YAS değişkenine göre iki alt düğüme ayrılmıştır.

Sol düğümü iki alt düğüme ayıran WBC değişkeni için kesim değeri 4644 olarak bulunurken, sağ düğümü ayıran YAS değişkeni için ise kesim değeri 49.84 olarak bulunmuştur. YAS değeri 49.84 ten büyük olan soldaki düğümden 8.67 rölatif hata ile 3 ilçe yer alırken, YAS değeri 49.84 den küçük olan sağdaki düğümden 59.5 rölatif hata ile 4 ilçe yer almıştır.

Sol düğümü iki alt düğüme ayıran WBC değeri 4644'den küçük olan sağdaki düğüm 22.5 rölatif hata ile 2 ilçeye sahip olurken, WBC değeri 4644'den büyük olan soldaki düğüm WBC değişkenine göre yine iki alt gruba ayrılmıştır. WBC için buradaki kesim değeri ise 7799 olmuştur. WBC değeri 7799 dan büyük olan sağdaki düğümden 2 rölatif hata ile 2 ilçe yer alırken, WBC değeri 7799 dan küçük olan soldaki düğüm WBC değişkenine göre yine iki alt düğüme bölünmüştür.

WBC için buradaki kesim değeri ise 5868 olmuştur. WBC değeri 5868 den küçük olan soldaki düğümden 6.5 rölatif hata ile 4 ilçe yer alırken, WBC değeri 5868den büyük olan sağdaki düğüm, PLT değişkenine göre yeniden iki alt düğüme ayrılmıştır. PLT değeri 1620 den küçük olan soldaki düğüm, 3.5 rölatif hata değerine sahip 4 ilçe ile terminal düğüm olurken, PLT değeri 1620 den büyük olan sağdaki düğüm, 1.5 rölatif hata değerine sahip 2 ilçe ile terminal düğüm olmuştur.

Böylece toplam olarak ağaç 7 terminal düğümden oluşmuştur. Bu modele ait R^2 değeri ise % 55.7 olarak bulunmuştur.

Modele ait çapraz geçerlik rölatif hata 2.38 olurken; standart hata değeri 0.516 olarak hesaplanmıştır.

Tablo 7. ALB ve HB deęişkenleri modelden çıkarıldığında oluşturulan tablo

	1. Düğüm	2. Düğüm	3. Düğüm	4. Düğüm	5. Düğüm	6. Düğüm
	GLB <3.906 ise L	WBC <4644 ise R	YAS <49.84 ise R	WBC <7799 ise L		
	ALT <52.8 ise R	YAS <60.4 ise L	WBC <8105.6 ise R	ALT <63.6 ise L		
	PLT <1750.9 ise L	ALT <31.17 ise R	ALT <69.2 ise R	PLT <1554.6 ise L		
	WBC <8000.7 ise R	AST <77.1 ise R	AST <63.07 ise L	GLB <3.51 ise L		
	YAS <60.4 ise R	GLB <3.50 ise L	PLT <1568.4 ise L	YAS <57.4 ise L		
Gözlem Sayısı	21	14	7	12	2	3
Ortalama	1.19, 2.52, 0.86	1, 1.64, 0.79	1.57, 4.29, 1	0.75, 1.08, 0.75	2.5, 5, 1	1, 1.67, 0
Toplanmış Hata Kareler Ortalaması	11.2	6.40	15.9	2.95	11.25	2.89

Tablo 7. (Devam)

	7. Düğüm	8. Düğüm	9. Düğüm	10. Düğüm	11. Düğüm	12. Düğüm	13. Düğüm
		WBC <5868 ise L			PLT <1620 ise L		
		PLT <1660.8 ise L			WBC <6144.5 ise L		
		ALT <86.06 ise R			AST <84.8 ise L		
		GLB <3.47 ise L			ALT <67.4 ise L		
		AST <71.72 ise L			GLB <3.73 ise R		
Gözlem Sayısı	4	10	2	4	6	4	2
Ortalama	2, 6.25, 1.75	0.5, 1.3, 0.5	2, 0, 2	0.5, 0.25, 0.75	0.5, 2, 0.33	0, 1.75, 0.25	1.5, 2.5, 0.5
Toplanmış Hata Kareler Ortalaması	14.88	2.31	1	1.63	1.48	0.88	0.75

Tablo 8. ALB ve HB değişkenleri modelden çıkarıldığında hesaplanan hata çizelgesi

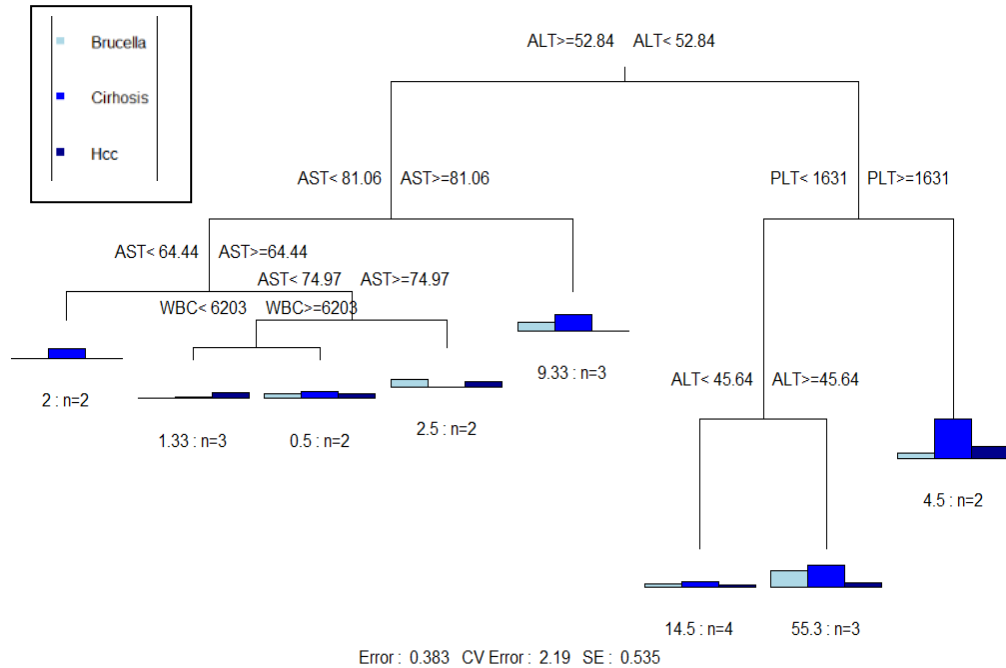
	Maliyet Karmaşası	n bölünme	Rölatif Hata	Çapraz Geçerlik Hatası	Çapraz Geçerlik Standart Hatası
1	0.16	0	1.00	1.11	0.32
2	0.14	2	0.67	2.11	0.49
3	0.04	3	0.54	2.30	0.52
4	0.03	4	0.49	2.32	0.52
5	0.02	5	0.46	2.38	0.52
6	0.00	6	0.44	2.37	0.51

Tablo 9. ALB ve HB değişkenleri modelden çıkarıldığında ağaç bölümleri ve tüm ağaca ait sapma (deviance) değerleri (%)

	GLB<3.906	WBC<4644	WBC<7799	WBC<5868	PLT<1620	YAS<49.84
Brucella	-	16.58	36.34	-	78.26	-
Cirhosis	94.94	83.07	27.30	94.64	19.56	83.80
Hcc	-	-	36.34	-	-	12.21
Toplam Tür Sapması	94.94	99.65	99.98	94.64	97.82	96.01

Uygulama 4: ALB, HB ve GLB değişkenlerinin modelden çıkarılması ile oluşan çok değişkenli regresyon ağacı

ALB, HB ve GLB değişkenlerinin modelden çıkarılması ile oluşan çok değişkenli regresyon ağacı Şekil 4’de, bu ağaca ilişkin sonuçlar ise Tablo 10, 11 ve 12’ de verilmiştir.



Şekil 4. ALB, HB ve GLB değişkenleri modelden çıkarıldığında oluşan çok değişkenli regresyon ağacı

Şekil 4' te görüldüğü üzere; ALB, HB ve GLB değişkenleri modelden çıkarıldığında, ilk bölünmenin ALT değişkenine göre olduğu görülür. Buna göre ALT için kesim değeri 52.84 olarak bulunmuş olup, bu değerden büyük olanlar sol düğüme, küçük olanlar ise sağ düğüme ayrılacak şekilde bölünme gerçekleşmiştir.

Sol düğüm, AST değişkenine göre iki alt gruba ayrılırken; sağ düğüm, PLT değişkenine göre iki alt düğüme ayrılmıştır. Sol düğümü iki alt düğüme ayıran AST değişkeni için kesim değeri 81.05 olarak bulunurken, sağ düğümü ayıran PLT değişkeni için ise kesim değeri 1631 olarak bulunmuştur.

PLT değeri 1631 den küçük olan soldaki düğüm, ALT değişkenine göre yeniden iki alt düğüme bölünürken, sağdaki düğüm, 4.5 rölatif hataya sahip 2 ilçe ile terminal düğüm olmuştur. ALT için buradaki kesim noktası 45.64 olarak gerçekleşmiş ve buna göre ayrılan iki alt gruptan soldakinde 14.5 rölatif hata ile 4 ilçe yer almış, sağdakinde ise 55.3 rölatif hata ile 3 ilçe yer almıştır.

AST değeri 81.05 den küçük olan soldaki düğüm, AST değişkenine göre yeniden iki alt düğüme bölünürken, sağdaki düğüm, 9.33 rölatif hata değerine sahip 3 ilçe ile terminal düğüm olmuştur. AST değişkeni için buradaki kesim noktası ise 64.44 olmuştur. AST değeri 64.44 den büyük olan sağdaki düğüm, AST değişkenine göre yeniden iki alt düğüme bölünürken, soldaki düğümdede, 2 rölatif hata ile 2 ilçe yer almıştır.

AST değişkeni için buradaki kesim noktası ise 74.97 olmuştur. AST değeri 74.97 den küçük olan soldaki düğüm, WBC değişkenine göre yeniden iki alt düğüme bölünürken, sağdaki düğüm, 9.33 rölatif hataya sahip 3 ilçe ile terminal düğüm olmuştur. WBC değişkeni için buradaki kesim noktası 6203 olmuş ve buna göre ayrılan iki alt gruptan soldakinde 1.33 rölatif hata ile 3 ilçe yer almış, sağdakinde ise 0.5 rölatif hata ile 2 ilçe yer almıştır.

Böylece toplam olarak ağaç 8 terminal düğümden oluşmuştur. Bu modele ait R^2 değeri % 61.7'dir.

Modele ait çapraz geçerlik rölatif hata 2.19 olurken; standart hata değeri 0.535 olarak hesaplanmıştır.

Tablo 10. ALB, HB ve GLB deęişkenleri modelden çıkarıldığında oluşturulan tablo

	1. Düğüm	2. Düğüm	3. Düğüm	4. Düğüm	5. Düğüm	6. Düğüm	7. Düğüm
	ALT <52.84 ise R	AST <81.06 ise L	PLT <1631.47 ise L	AST <64.44 ise L		ALT <45.64 ise L	
	PLT <1759 ise L	WBC <5867.86 ise L	YAS <52 ise R	ALT <66.3 ise L		PLT <1580.7 ise L	
	WBC <8000.6 ise R	ALT <63.6 ise L	AST <89.5 ise R	PLT <1559.4 ise L		YAS <52 ise R	
	YAS <60.4 ise R	PLT <1559.4 ise R	ALT <45.6 ise L	WBC <5934.2 ise L		AST <69.7 ise R	
	AST <78 ise L	YAS <47.67 ise L	WBC <6324.7 ise L	YAS <46.8 ise R		WBC <5577.7 ise R	
Gözlem Sayısı	21	12	9	9	3	7	2
Ortalama	1.19, 2.5, 0.86	0.83, 1.5, 0.67	1.67, 3.89, 1.11	0.56, 0.89, 0.89	1.67, 3.33, 0	1.86, 2.71, 0.71	1, 8, 2.5
Toplanmış Hata Kareler Ortalaması	11.2	3.78	17.31	2	3.1	14.53	2.25

Tablo 10. (Devam)

	8. Düğüm	9. Düğüm	10. Düğüm	11. Düğüm	12. Düğüm	13. Düğüm	14. Düğüm	15. Düğüm
		AST <74.97 ise L			WBC <6203.1 ise L			
		WBC <5586.8 ise L			YAS <46.8 ise R			
		ALT <66.3 ise L			ALT <85.37 ise R			
		PLT <1610.8 ise L			PLT <1693.75 ise L			
		YAS <46.8 ise R			AST <70.4 ise L			
Gözlem Sayısı	2	7	4	3	5	2	3	2
Ortalama	0, 2, 0	0.71, 0.57, 1.14	0.75, 1.25, 0.5	3.33, 4.67, 1	0.4, 0.8, 1.2	1.5, 0, 1	0, 0.33, 1.33	1, 1.5, 1
Toplanmış Hata Kareler Ortalaması	1	1.43	3.6	18.4	0.96	1.25	0.44	0.25

Tablo 11. ALB, HB ve GLB değişkenleri modelden çıkarıldığında hesaplanan hata çizelgesi

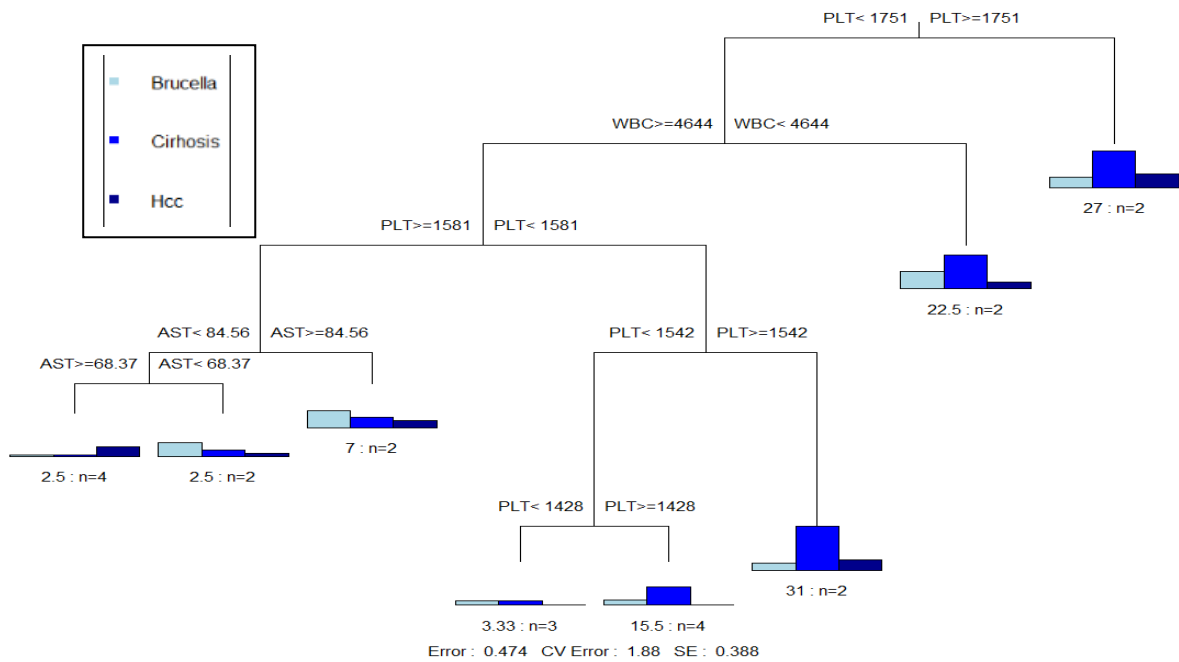
	Maliyet Karmaşası	n bölünme	Rölatif Hata	Çapraz Geçerlik Hatası	Çapraz Geçerlik Standart Hatası
1	0.18	0	1.00	1.11	0.32
2	0.14	2	0.64	1.84	0.50
3	0.08	3	0.51	2.06	0.51
4	0.02	4	0.43	2.13	0.53
5	0.01	5	0.41	2.14	0.53
6	0.00	7	0.38	2.13	0.53

Tablo12. ALB, HB ve GLB değişkenleri modelden çıkarıldığında ağaç bölümleri ve tüm ağaca ait sapma (deviance) değerleri (%)

	ALT<52.84	AST<81.06	AST<64.44	AST<74.97	WBC<6203	PLT<1631	ALT<45.64
Brucella	10.51	15.43	13.23	64.02	40.44	-	35.88
Cirhosis	86.48	74.69	52.91	33.86	55.06	87.68	62.77
Hcc	-	-	33.86	-	-	10.01	-
Toplam Tür Sapması	96.99	90.12	100.00	97.88	95.50	97.69	98.65

Uygulama 5: ALB, HB, GLB ve ALT değişkenlerinin modelden çıkarılması ile oluşan çok değişkenli regresyon ağacı

ALB, HB, GLB ve ALT değişkenlerinin modelden çıkarılması ile oluşan çok değişkenli regresyon ağacı Şekil 6’da, bu ağaca ilişkin sonuçlar ise Tablo 16, 17 ve 18’ de verilmiştir.



Şekil 5. ALB, HB, GLB ve ALT değişkenleri modelden çıkarıldığında oluşan çok değişkenli regresyon ağacı

Şekil 5' te görüldüğü üzere; ALB, HB, GLB ve ALT değişkenleri modelden çıkarıldığında, ilk bölünmenin PLT değişkenine göre olduğu görülür. Buna göre PLT için kesim değeri 1751 olarak bulunmuş olup, bu değerden küçük olanlar sol düğüme, büyük olanlar ise sağ düğüme ayrılacak şekilde bölünme gerçekleşmiştir.

PLT değeri 1751 den küçük olan soldaki düğüm, WBC değişkenine göre yeniden iki alt düğüme bölünürken, sağdaki düğüm, 27 rölatif hataya sahip 2 ilçe ile terminal düğüm olmuştur.

Sol düğümü ayıran WBC değişkeni için kesim değeri 4644 olarak bulunmuştur. WBC değeri 4644 den küçük olan sağdaki düğüme 22.5 rölatif hata ile 2 ilçe yer alırken, WBC değeri 4644 den büyük olan soldaki düğüm, PLT değişkenine göre yeniden iki alt düğüme ayrılmıştır. PLT için buradaki kesim değeri ise 1581 olmuştur. Buna göre ayrılan iki alt gruptan soldaki düğüm, AST değişkenine göre yeniden iki alt düğüme bölünürken; sağdaki düğüm de PLT değişkenine göre yeniden iki alt düğüme ayrılmıştır.

Sol düğümü iki alt düğüme ayıran AST değişkeni için kesim değeri 84.56 olarak bulunurken, sağ düğümü ayıran PLT değişkeni için ise kesim değeri 1542 olarak bulunmuştur. AST değeri 84.56'dan küçük olan soldaki düğüm, AST değişkenine göre yine iki alt gruba ayrılırken; sağdaki düğüm 7 rölatif hataya sahip 2 ilçe ile terminal düğüm olmuştur. AST için buradaki kesim değeri ise 68.37 olarak bulunmuştur. AST değeri 68.37 den büyük olan soldaki düğüme 2.5 rölatif hata ile 4 ilçe yer alırken, sağdaki düğüme de 2.5 rölatif hata ile 2 ilçe yer almıştır.

PLT değeri 1542 den küçük olan soldaki düğüm, PLT değişkenine göre tekrar bölünürken; PLT değeri 1542 den büyük olan sağdaki düğüm, 31 rölatif hata ile 2 ilçeye sahiptir. PLT değişkeni için yeni kesim noktası 1428 olmuştur ve buna göre ayrılan iki alt gruptan soldakinde 3.33 rölatif hata ile 3 ilçe yer almış sağdakinde ise 15.5 rölatif hata ile 4 ilçe yer almıştır.

Böylece toplam olarak ağaç 8 terminal düğümden oluşmuştur. Bu modele ait R^2 değeri % 52.6'dır.

Modele ait çapraz geçerlik rölatif hata 1.88 olurken; standart hata değeri 0.388 olarak hesaplanmıştır.

Tablo 13. ALB, HB, GLB ve ALT deęişkenleri modelden çıkarıldığında oluşturulan tablo

	1. Düğüm	2. Düğüm	3. Düğüm	4. Düğüm	5. Düğüm	6. Düğüm	7. Düğüm
	PLT <1750.9 ise L	WBC <4644 ise R		PLT <1581 ise R		AST <84.56 ise L	PLT <1542.3 ise L
	WBC <8000.7 ise R	PLT <1660.8 ise R		AST <64.9 ise R		YAS <51.03 ise L	YAS <49.8 ise R
	YAS <60.37 ise R	YAS <47.7 ise L		WBC <5867.9 ise L		PLT <1660.8 ise R	WBC <5975.1 ise L
	AST <78 ise L	AST <77.1 ise R		YAS <49.8 ise R		WBC <7798.8 ise L	AST <69.7 ise R
Gözlem Sayısı	21	19	2	17	2	8	9
Ortalama	1.19, 2.52, 0.86	1.16, 2.2, 0.74	1.5, 5.5, 2	1, 1.89, 0.71	2.5, 5, 1	1.25, 0.75, 1.13	0.78, 2.89, 0.33
Toplanmış Hata Kareler Ortalaması	11.2	9.76	13.50	8.3	11.25	2.99	10.5

Tablo 13. (Devam)

	8. Düğüm	9. Düğüm	10. Düğüm	11. Düğüm	12. Düğüm	13. Düğüm	14. Düğüm	15. Düğüm
	AST <68.4 ise R		PLT <1427.88 ise L					
	PLT <1622.7 ise R		YAS <49.6 ise R					
	WBC <5669.1 ise L		WBC <5776.4 ise L					
	YAS <51.03 ise L		AST <81.3 ise L					
Gözlem Sayısı	6	2	7	2	4	2	3	4
Ortalama	0.83, 0.5, 1.17	2.5, 15, 1	0.7, 1.86, 0	1, 6.5, 1.5	0.25, 0.25, 1.5	2, 1, 0.5	0.67, 0.6 7, 0	0.75, 2.75, 0
Toplanmış Hata Kareler Ortalaması	1.86	3.5	3.76	15.5	0.63	1.25	1.1	3.88

Tablo 14. ALB, HB, GLB ve ALT değişkenleri modelden çıkarıldığında hesaplanan hata çizelgesi

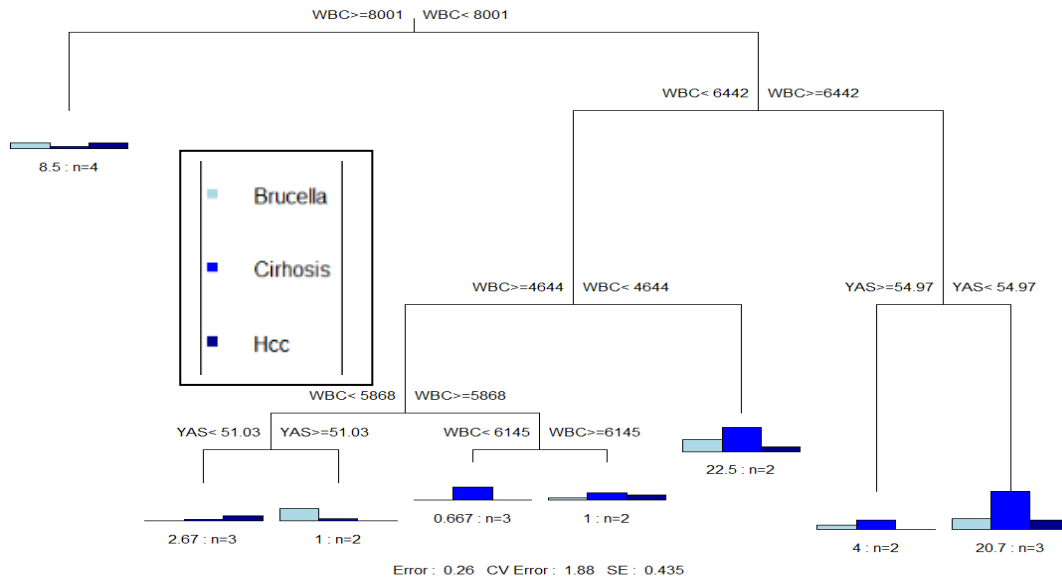
	Maliyet Karmaşası	n bölünme	Rölatif Hata	Çapraz Geçerlik Hatası	Çapraz Geçerlik Standart Hatası
1	0.11	0	1.00	1.11	0.32
2	0.03	4	0.56	1.98	0.45
3	0.02	5	0.52	1.94	0.40
4	0.00	7	0.47	1.95	0.40

Tablo 15. AST, HB, GLB ve ALT değişkenleri modelden çıkarıldığında, ağaç bölümleri ve tüm ağaca ait sapma (deviance) değerleri (%)

	PLT<1751	WBC<4644	PLT<1581	AST<84.56	AST<68.37	PLT<1542	PLT<1428
Brucella	-	18.66	-	73.00	66.22	-	-
Cirrhosis	86.34	80.62	84.33	26.28	12.16	90.24	99.84
Hcc	12.73	-	11.56	-	21.62	-	-
Toplam Tür Sapması	99.07	99.28	95.89	99.28	100.00	90.24	99.84

Uygulama 6: ALB, HB, GLB, ALT ve PLT değişkenleri modelden çıkarıldığında oluşan çok değişkenli regresyon ağacı

ALB, HB, GLB, ALT ve PLT değişkenleri modelden çıkarıldığında oluşan çok değişkenli regresyon ağacı Şekil 6'da, bu ağaca ilişkin sonuçlar ise Tablo 13, 14 ve 15' te verilmiştir.



Şekil 6. ALB, HB, GLB, ALT ve PLT değişkenleri modelden çıkarıldığında oluşan çok değişkenli regresyon ağacı

Şekil 6' da görüldüğü üzere; ALB, HB, GLB, ALT ve PLT değişkenleri analizden çıkarıldığında, ilk bölünmenin WBC değişkenine göre olduğu görülür. Buna göre WBC değişkeni için kesim değeri 8001 olarak bulunmuş olup, bu değerden büyük olanlar sol düğüme, küçük olanlar ise sağ düğüme ayrılacak şekilde bölünme gerçekleşmiştir. Sol düğüm, 8.5 rölatif hata değerine sahip 4 ilçe ile terminal düğüm olurken; sağ düğüm, yeniden WBC değişkenine göre iki alt düğüme ayrılmıştır.

Sağ düğümü iki alt düğüme ayıran WBC değişkeni için kesim değeri 6442 olarak bulunmuştur. WBC değeri 6442'den büyük olan sağdaki düğüm, YAS değişkenine göre yeniden iki alt düğüme bölünmüştür. YAS değişkeni için kesim noktası 54.97 olarak gerçekleşmiştir ve YAS değeri 54.97 den küçük olan sağdaki düğüm, 20.7 rölatif hata değerine sahip 3 ilçe ile terminal düğüm olurken, soldaki düğüm, 4 rölatif hata değerine sahip 2 ilçe ile terminal düğüm olmuştur.

WBC değeri 4644 ten büyük olan soldaki düğüm, WBC değişkenine göre tekrar bölünürken, WBC değeri 4644'den küçük olan sağdaki düğüm, 22.5 rölatif hata ile 2 ilçeye sahiptir. WBC değişkeni için kesim değeri 5868 dir. WBC değeri 5868'den küçük olan soldaki düğüm, YAS değişkenine göre yeniden iki alt düğüme bölünürken, sağdaki düğüm, WBC değişkenine göre yeniden bölünmüştür.

YAS değişkeni için kesim noktası 51.03 olarak gerçekleşmiş ve buna göre ayrılan iki alt gruptan soldaki düğüm, 2.67 rölatif hata değerine sahip 3 ilçe ile terminal düğüm olurken, sağdaki düğüm, 1 rölatif hata değerine sahip 2 ilçe ile terminal düğüm olmuştur.

WBC için buradaki kesim değeri ise 6145 olmuştur. WBC değeri 6145'den küçük olan soldaki düğüm, 0.667 rölatif hata değerine sahip 3 ilçe ile terminal düğüm olurken, sağdaki düğüm, 1 rölatif hata değerine sahip 2 ilçe ile terminal düğüm olmuştur.

Böylece ağaç, toplam olarak 7 terminal düğümden oluşmuştur. Bu modele ait R^2 değeri ise %74 olarak bulunmuştur.

Modele ait çapraz geçerlik rölatif hata bir önceki ağaca benzer şekilde 1.88 olurken; standart hata değeri 0.435 olarak hesaplanmıştır.

Tablo 16. ALB, HB, GLB, ALT ve PLT deęişkenleri modelden çıkarıldığında oluşturulan tablo

	1. Düğüm	2. Düğüm	3. Düğüm	4. Düğüm	5. Düğüm	6. Düğüm	7. Düğüm
	WBC <8000.67 ise R		WBC <6442.4 ise L	WBC <4643.7 ise R	YAS <54.97 ise R	WBC <5867.9 ise L	
	YAS <60.36 ise R		YAS <60.4 ise R	YAS <54.5 ise L	WBC <6918.9 ise R	AST <64.4 ise R	
	AST <78 ise L		AST <77.98 ise L	AST <71.7 ise L	AST <81.3 ise L	YAS <54.5 ise L	
Gözlem Sayısı	21	4	17	12	5	10	2
Ortalama	1.19, 2.52, 0.86	1.25, 0.5 ,1.25	1.18, 3, 0.76	0.92, 1.92, 0.58	1.8, 5.6, 1.2	0.6, 1.3, 0.5	2.5, 5, 1
Toplanmış Hata Kareler Ortalaması	11.2	2.13	12.09	6.56	15	2.7	11.25

Tablo 17. ALB, HB, GLB, ALT ve PLT deęişkenleri modelden çıkarıldığında hesaplanan hata çizelgesi

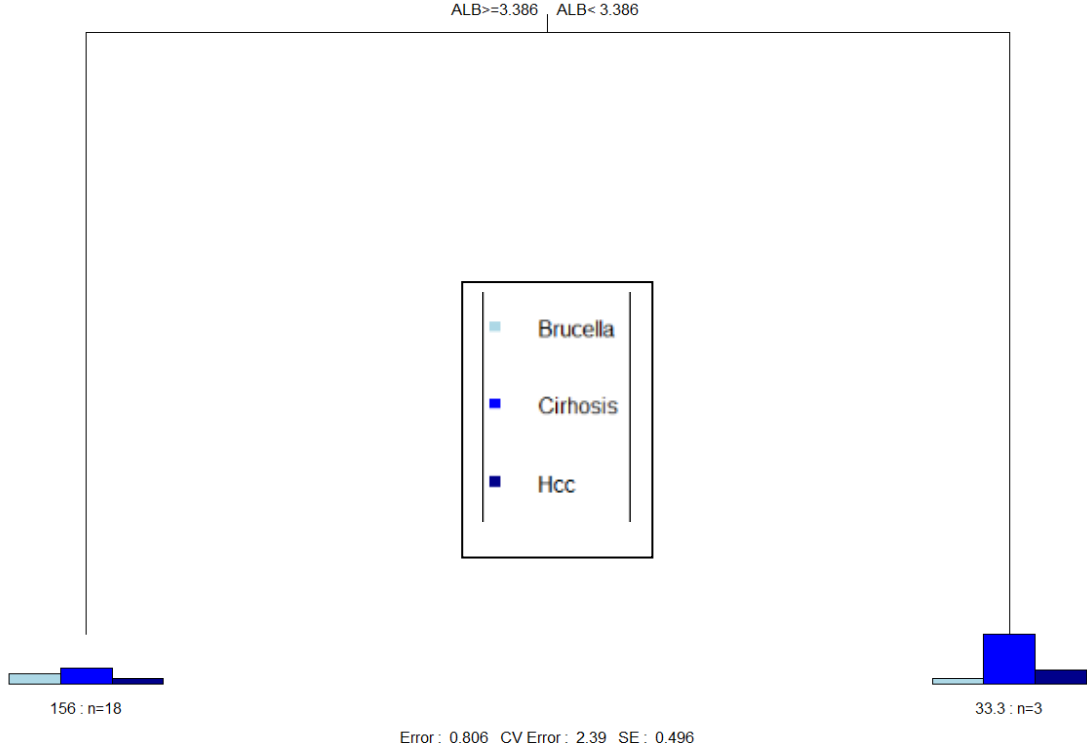
	Maliyet Karmaşası	N bölünme	Rölatif Hata	Çapraz Geçerlik Hatası	Çapraz Geçerlik Standart Hatası
1	0.16	0	1.00	1.11	0.32
2	0.12	3	0.48	1.63	0.42
3	0.04	4	0.35	1.80	0.45
4	0.04	5	0.31	1.83	0.43
5	0.03	6	0.27	1.83	0.43
6	0.00	7	0.26	1.84	0.43

Tablo18. ALB, HB, GLB, ALT ve PLT deęişkenleri modelden çıkarıldığında, ağaç bölünmeleri ve tüm ağaca ait sapma (deviance) deęerleri (%)

	WBC <8001	WBC <6442	WBC <4644	WBC <5868	YAS <51.03	WBC <6145	YAS <54.97
Brucella	-	-	30.19	16.33	85.87	-	-
Cirrhosis	96.29	92.11	-	82.65	-	52.13	86.17
Hcc	-	-	67.92	-	13.74	38.30	-
Toplam Tür Sapması	96.29	92.11	98.11	98.98	99.61	90.43	86.17

Uygulama 7: Tüm deęişkenlerin modele dahil edilmesi ile 1SE kuralına göre oluřan çok deęişkenli regresyon aęacı

Tüm deęişkenlerin modele dahil edilmesi ile 1SE kuralına göre oluřan çok deęişkenli regresyon aęacı Őekil 7’de, bu aęaca iliřkin sonular ise Tablo 19, 20 ve 21’ de verilmiřtir.



Őekil 7. Tüm deęişkenlerin modele dahil edilmesi ile 1SE kuralına göre oluřan çok deęişkenli regresyon aęacı

Őekil 7’ de grldę üzere; modele tm deęişkenler dahil edildięinde, 1SE kuralına göre oluřan aęata ilk blnmenin ALB deęişkenine gre olduęu grlr. Buna gre ALB iin kesim deęeri 3.386 olarak bulunmuř olup, bu deęerden kk olanlar sol dęme, byk olanlar ise saę dęme ayrılacak Őekilde blnme gerekleřmiřtir. Sol dęm, 156 rlatif hata ile 18 ileye sahip olarak terminal dęm olurken; saę dęm 33.3 rlatif hata ile 3 ileye sahip olarak terminal dęm olmuřtur.

Bylece toplam olarak aęa 2 terminal dęmden oluřmuřtur. Bu modele ait R^2 deęeri % 19’dur.

Modele ait apraz geerlik rlatif hata 2.39 olurken; standart hata deęeri 0.496 olarak hesaplanmıřtır.

Tablo 19. Tüm deęişkenler analize dahil edildiğinde 1SE kuralına göre oluşan tablo

	1. Düğüm	2. Düğüm	3. Düğüm
	ALB < 3.385 ise R		
	HB <14.55 ise L		
	GLB < 3.91 ise L		
	ALT <52.84 ise R		
	PLT <1750.92 ise L		
Gözlem Sayısı	21	18	3
Ortalama	1.19, 2.52, 0.86	1.28, 1.94, 0.72	0.67 ,6, 1.67
Toplanmış Hata Kareler Ortalaması	11.19	8.68	11.11

Tablo 20. Tüm deęişkenler analize dahil edildiğinde 1SE kuralına göre hesaplanan hata çizelgesi

	Maliyet Karmaşası	n bölünme	Rölatif Hata	Çapraz Geçerlik Hatası	Çapraz Geçerlik Standart Hatası
1	0.19	0	1.00	1.10	0.32
2	0.16	1	0.80	2.39	0.50

Tablo 21. Tüm deęişkenler analize dahil edildiğinde, 1SE kuralına göre ağaç bölünmeleri ve tüm ağaç tarafından hesaplanan sapma (%)

	ALB<3.386
Brucella	-
Cirrhosis	92.85
Hcc	-
Toplam Tür Sapması	92.85

5. TARTIŞMA ve SONUÇ

Çalışmada, Çok değişkenli regresyon ağacı incelenerek, sağlık alanında bir veri seti ile uygulama yapılmıştır. Uygulamada, 8 adet açıklayıcı değişkenden her defasında bölünmede en önemli olan değişken çıkarılarak, değişken sayısı azaltılmış ve buna göre ağaç modelleri oluşturulmuştur.

Çalışmada toplam 7 adet uygulama yapılmıştır. Uygulamalarda regresyon ağaçları, analiz prosedüründe her bir ağaç için 100 kez tekrarlama yapıldıktan sonra önerilen ağaç yapıları dikkate alınarak oluşturulmuştur. İlk 6 uygulamada, ağaç boyutları CVRE'yi en aza indiren noktalar dikkate alınarak oluşturulurken, yedinci uygulamada tüm değişkenler modele alınarak, "1SE" kuralına göre ağaç boyutu belirlenmiştir.

Uygulama 1 ve Uygulama 7'de her iki ağacın da kök düğümü ALB değişkeni olup, kesim noktası 3.386' dır. Uygulama 1'de ALB değişkeninin değeri 3.386'dan küçük olduğunda, sağ düğüm 33.3 rölatif hata ve 3 ilçe ile terminal düğüm haline gelirken, benzer şekilde Uygulama 7'de de ALB değişkeninin değeri 3.386'dan küçük olduğunda, sağ düğüm 33.3 rölatif hata ve 3 ilçe ile terminal düğüm haline gelmiştir. Uygulama 1'de kök düğümünden sonraki sol alt bölünmenin, ALB değişkenine göre olduğu görülürken, Uygulama 7'de terminal düğüm haline gelmiştir.

Buradan hareketle, Uygulama 7'de elde edilen ağacın daha sade ve daha az bilgi verdiği buna karşılık, Uygulama 1'de elde edilen ağacın ise daha karmaşık ve daha ayrıntılı bilgi verdiği söylenebilir. Ayrıca, aşırı uyumdan kaçınmak için önerilen "1SE" kuralının avantajının ise daha geniş veri setlerinde ortaya çıkabileceği söylemek mümkündür. Diğer yandan, Uygulama 1'den elde edilen rölatif hataya bağlı olarak %56 R^2 değeri elde edilirken, Uygulama 7 den elde edilen R^2 değeri %19 olmuştur. Buradan hareketle, aşırı uyum söz konusu olmadığında, ağacın karmaşıklığının artmasıyla birlikte, açıklanan varyans oranının da artma eğilimde olduğu söylenebilir. Uygulama 1' de ağacın CVRE değeri 2.6 olurken, Uygulama 7'de bu değer 2.39 olmuştur. Standart hata ise Uygulama 1'de 0.485 iken, Uygulama 7'de 0.496 olmuştur. Ağaç boyutu seçmede her ne kadar "1SE" kuralı öneriliyor olsa da oluşan ağaç için açıklanan varyans ve hata değerleri dikkate alındığında, CVRE'yi en aza indiren ağaç boyutu seçme kuralının daha uygun olabileceği söylenebilir. Diğer bir ifade ile Çok değişkenli regresyon ağacı yönteminde, araştırmacının yapabileceği en basit seçim, CVRE'yi en aza indiren ağaç boyutunu seçmektir. Ancak, bu istatistiklerin değişkenliği nedeniyle, ağaç

oluşumunda boyutların farklı olması durumu gözardı edilmemelidir. Diğer yandan, Breiman ve ark. (1984), CVRE'yi en aza indiren çözümün seçilmesi yerine minimum “1SE” içinde kalan en küçük ağacın seçilmesi için bir başka ayarlamayı önermektedir.

Sınıflandırma ve regresyon ağacı çerçevesinde, parametrik olmayan hesaplama yöntemlerinden biri olarak, çok değişkenli veri setlerinin analizinde kullanılan yöntemlerden birisi de çok değişkenli regresyon ağacıdır. Çok değişkenli regresyon ağacında ana kavram, veri seti her bir adımda bağımsız değişkenlerden birine göre kullanılan kesim değeri dikkate alınarak, her bir adımda alt gruplara bölünmesidir. Her adımda bölünme yapılırken, veri setinde yer alan cevap değişkeni için saf olmama değeri olarak bilinen (impurity) kriteri minimize edilmeye çalışılır. Ağacın büyütülmesi tamamlandığında, aşırı uyum (overfit) durumunun ortaya çıkması muhtemeldir. Böyle bir durumu ortadan kaldırmak üzere, optimum büyüklükteki ağaç yapısını elde edebilmek için çapraz geçerlilik kuralına göre dallar birleştirilerek budama işlemi yapılır (Stowell ve Plumbley, 2010).

De'ath (2002) tarafından belirtildiği gibi çok değişkenli regresyon ağaçları ortalamadan sapma kareler toplamını kullandığı için Öklid metodları ailesine dahildir. Ayrıca çalışmada çok değişkenli regresyon ağacı yönteminin, sıfır değeri içeren tablolarda uygulanabilmesi için verilere transformasyon yapılabileceği belirtilmiştir.

Çok değişkenli regresyon ağacında, elde edilen verilere ilişkin herhangi bir varsayım yoktur. Yöntem, açıklama tanımlama ve tahminleme amaçlı olarak kullanılmaktadır. Böylece, çok değişkenli regresyon ağacı, çok değişkenli regresyon analizi yönteminin açıklayıcı, tahmin edici ve araştırmacı hali olarak da düşünülebilir (De'ath, 2002).

Çok değişkenli regresyon ağacı, tek değişkenli regresyon ağacının birden fazla cevap değişkeni için genişletilmiş halidir. Çok değişkenli regresyon ağacı, herhangi bir düğüm içerisinde yer alan bağımlı değişkenlerin her birisi için düğüm içerisindeki varyasyonu minimize edecek şekilde, bunlara etkili olan faktörleri, bu faktörlerin seviyelerini ve bunlar arasındaki interaksiyonları belirlemede kullanılabilecek uygun analiz yöntemlerinden birisidir.

Tek ve çok değişkenli regresyon ağaçlarından hangisinin kullanılacağına karar vermede cevap değişkeni sayısı önemlidir. Tek bir cevap değişkenine açıklayıcı değişkenlerin etkileri ve bu açıklayıcı değişkenler arasındaki etkileşimlerin (interaksiyon) incelenmesi isteniyorsa tek değişkenli regresyon ağacının kullanılması uygun iken, birden fazla cevap değişkeni olması durumunda çok değişkenli regresyon ağacının kullanılması daha uygun olacaktır. Birden çok

cevap deęişkeni olması durumunda, her bir cevap deęişkeni için tek deęişkenli regresyon ağacı modeli, ayrı ayrı oluşturulduğunda, bu cevap deęişkenlerinin birbirinden bağımsız olduğu varsayılır. Böylece, cevap deęişkenlerinin birbiri ile olan ilişkileri yok sayılmış olur. Oysaki gerçekte cevap deęişkenleri az da olsa birbiri ile ilişkilidir. Örneğin cevap deęişkenlerinin türler olması durumunda, bazı türler birbirleri ile birlikte bulunurken, bazıları aynı ortamda yer almayabilir. Benzer şekilde, cevap deęişkenleri hastalıkların varlığı olması durumunda, bazı hastalıklar birlikte bulunabilmektedir. Bu gibi durumlarda, çok deęişkenli regresyon ağacı yönteminin kullanılması ile cevap deęişkenleri arasındaki ilişkiler de dikkate alınarak, bu deęişkenlerin birlikte bulunması durumunda birbirlerine olan etkileri de belirlenebilmiş olmaktadır. Diğer yandan, cevap deęişkeni sayısı çok olması durumunda, her bir cevap deęişkeni için ayrı ayrı yapılacak tek deęişkenli regresyon ağaçları yerine, çok deęişkenli regresyon ağacının kullanılması ile tek bir regresyon ağacı elde edilecek ve sonuçların yorumlanması da daha kolay olacaktır.

Larsen ve Speckman (2004), yeni gözlemlerin regresyon ağacı yöntemi ile tanımlanması veya sınıflandırılmasında regresyon ağacı yöntemlerinin zayıf veya yetersiz olduğunu belirterek, bunun için Bagging Tree ya da Random Forest yöntemlerinin geliştirildiğini belirtmişlerdir. Ancak bu prosedürlere ilişkin sonuçların ve ağaç yapısının basit olmaması nedeniyle, sınıflandırma uygulamalarında yaygın olarak kullanılmadığına dikkat çekmişlerdir. Larsen ve Speckman (2004) ayrıca, verilerin ağırlıklandırılması veya verilere karekök transformasyonu uygulanması durumunda, regresyon ağaçlarını denemişler ve bu tip transformasyonların uygulanması durumunda, makul ve mantıklı regresyon ağaçları elde edilmekle birlikte, ağaç yapısında ve yapıyı oluşturan deęişkenlerin sırasında önemli farklılıklar olduğunu vurgulamışlardır.

Son zamanlarda Questier ve ark. (2005), tarafından Oto İlişkili (Auto-Associative) Çok Deęişkenli Regresyon Ağaçları önerilmektedir. Bu ağaçlar standart çok deęişkenli regresyon ağaçlarından farklıdır. Zira bu ağaç yapılarında açıklayıcı deęişkenler ve cevap deęişkenleri aynıdır. Genel olarak değerlendirildiğinde çok deęişkenli regresyon ağaçları denetimli (supervised) uygulamalar için kullanılmaktadır. Oto ilişkili çok deęişkenli regresyon ağaçlarında aynı veri seti hem çok deęişkenli cevap deęişkeni hem de çok deęişkenli tahminçiler olarak kullanılmaktadır. Ayrıca Oto İlişkili (Auto-Associative) Çok Deęişkenli Regresyon Ağaçları yalnızca açıklayıcı veri analizinde kullanılmaktadır.

Çok deęişkenli regresyon analizi yönteminin avantajlarını aőaęıdaki gibi özetlemek mümkündür. 1) Parametrik olmayan yöntemler arasında yer aldığından, verilerin dağılımı hakkında herhangi bir varsayım ya da ön şart gerektirmez. 2) Açıklayıcı deęişkenlere transformasyon yapılabilir. 3) Basit ve kolay yorumlanabilir grafiksel yöntemler arasında yer alan çok deęişkenli analiz yöntemlerinden birisidir. 4) Kayıp gözlemler içeren veri setleri için uygundur. 5) Gürültülü cevap ve açıklayıcı deęişkenlere karşı sağlam (robust) bir yöntemdir. 6) Aşırı uç deęerlere karşı sağlam bir testtir ve sınıflara veya düğümlere bölünme işlemini hızlı yapar. 7) Model seçiminde çapraz geçerlik testini kullanır (Questier ve ark., 2005).

Sonuç olarak, çok deęişkenli regresyon ağacı yönteminin; parametrik olmayan bir test yöntemi olması, verilerin dağılımı hakkında herhangi bir önşart gerektirmemesi, kayıp gözlemlere ve aşırı uç deęerlere karşı sağlam bir test olması ve sonuçları kolay yorumlanabilir şekilde grafiksel olarak sunabilmesi gibi nedenlerle, başta tür-çevre ilişkilerini belirlemek olmak üzere, birden çok cevap deęişkeni ile birden çok açıklayıcı deęişken arasındaki ilişkileri belirlemede ve dięer bir çok alanda rahatlıkla kullanılabilceęi söylenebilir.

ÖZET

Korkmaz D, Çok deęişkenli regresyon ağaçları ve uygulaması. Van Yüzüncü Yıl Üniversitesi, Sağlık Bilimleri Enstitüsü, Biyoistatistik Anabilim Dalı, Yüksek Lisans Tezi, Van 2018. Çok deęişkenli istatistik analiz yöntemleri, karmaşık veri setlerini analiz etmede yaygın olarak kullanılan yöntemlerdir. Bu yöntemler, çok sayıda bağımsız ve/veya bağımlı deęişkenle ve tüm bu deęişkenlerin deęişik düzeylerde birbiri ile ilişkili olduęu durumlarda, birkaç tek deęişkenli veya iki deęişkenli analiz yapmak yerine tek bir analiz yapma imkanı sağlar. Bu tekniklerden birisi de çok deęişkenli regresyon ağacı yöntemidir. Bu yöntem, varsayımların esneklięi ve sonuçların; doęru ve güvenilir olması ayrıca kolay yorumlanabilirlięi ve interaksyonları göstermesi bakımından avantajlıdır. Bu çalışmada, Çok deęişkenli regresyon ağacı yöntemi incelenerek, bir uygulama yapılmıştır. Uygulamada, 21 ilçeden 95 bireye ait veriler kullanılmıştır. Brucella, HCC ve Siroz (Cirrhosis) hastalıklarının varlıęı (bolluęu) cevap deęişkeni, buna etkili olabileceęi düşünölen Yaş (yıl), Beyaz Küre [WBC, (K/mm³)], Hemoglobün [HB, (g/dL)], Alanin Aminotransferaz [(ALT, (g/dL)], Aspartat Aspartat Aminotransferaz [AST (g/dL)], Trombosit [PLT, (K/mm³)], Albümin [ALB, (g/dL) ve Globölin [GLB, (g/dL)] özellikleri ise açıklayıcı deęişken olarak alınmıştır. Sonuç olarak, yöntemin sahip olduęu bazı avantajlar nedeniyle, çok deęişkenli veri setlerinin analizinde kullanılabilirlięi vurgulanmıştır.

Anahtar kelimeler: Çapraz geçerlik rölatif hatası, düęüm, ekoloji, maliyet karmaşası, sapma, sentroid

SUMMARY

Korkmaz D, Multivariate Regression Trees and Its Application. Yuzuncu Yil University, Institute of Health Sciences, Department of Biostatistics, Master Thesis, Van 2018. Multivariate statistical analysis methods are widely used to analyze complex data sets. These methods provide the ability to perform a single analysis, rather than performing several univariate or bivariate analyzes, with a large number of independent and / or dependent variables, when the variables are correlated at different levels. One of these techniques is the multivariate regression tree method. This method is flexible for the assumptions, has accurate and reliable results as well as to examine interactions and easy interpretations. In this study, a multivariate regression tree method was examined and an application was performed. In application, data of 95 individuals from 21 provinces were used. The abundance of Brucella, HCC and Cirhosis were considered as response variables. Age (year), White Blood Cell [WBC, (K/mm³)], Hemoglobin [HB, (g/dL)], Alanin Aminotransferaz [(ALT, (g/dL)], Aspartat Aminotransferaz [AST (g/dL)], Trombosit [PLT, (K/mm³)], Albumin [ALB, (g/dL) and Globulin [GLB, (g/dL)] that likely to be effect on these response variables were taken as explanatory variables. As a result, it is emphasized that the method can be used for analysing of the sets of multivariate data.

Key words: Cross validation relative error, node, ecology, cost complexity, deviance, centroid

KAYNAKLAR

- Anderson T W (1958). *An Introduction to Multivariate Statistical Analysis, Vol 2*, New York, Wiley.
- Breiman L, Friedman JH, Olshen RA and Stone CJ (1984). *Classification and Regression Trees*, Berkeley.
- Cannon AJ (2012). Koppen versus the computer: comparing Koppen-Geiger and multivariate regression tree climate classifications in terms of climate homogeneity, *Hydrology and Earth System Sciences*, 16, 1, 217.
- Craig AW (2004). Primer on multiple regression coding: Common forms and the additional case of repeated contrasts, *Understanding Statistics*, 3(1)
- Curran Jr WJ, Scott CB, Horton J, Nelson J S, Weinstein AS, Fischelbach AJ, Nelson DF (1993). Recursive partitioning analysis of prognostic factors in three radiation therapy oncology group malignant gliomas, *JNCI: Journal of the National Cancer Institute*, 85, 9, 704-710.
- De'ath G, Fabricius KE (2000). Classification and regression trees: a powerful yet simple technique for ecological data analysis, *Ecology*, 81, 11, 3178-3192.
- De'Ath G (2002). Multivariate regression trees: a new technique for modeling species–environment relationships, *Ecology*, 83, 4, 1105-1117.
- Ebell MH, Hansen JG (2017). Proposed clinical decision rules to diagnose acute rhinosinusitis among adults in primary care, *The Annals of Family Medicine*, 15, 4, 347-354.
- Falconer JA, Naughton BJ, Dunlop DD, Roth EJ, Strasser DC, Sinacore JM (1994). Predicting stroke in patient rehabilitation outcome using a classification tree approach, *Archives of Physical Medicine and Rehabilitation*, 75, 6, 619-625.
- Fonarow GC, Adams KF, Abraham WT, Yancy CW, Boscardin WJ, Adhere Scientific Advisory Committee (2005). Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis, *Jama*, 293, 5, 572-580.
- Hamann A, Gylander T, Chen P Y (2011). Developing seed zones and transfer guidelines with multivariate regression trees, *Tree Genetics and Genomes*, 7, 2, 399-408.
- Hess KR, Abbruzzese MC, Lenzi R, Raber MN, Abbruzzese JL (1999). Classification and regression tree analysis of 1000 consecutive patients with unknown primary carcinoma, *Clinical Cancer Research*, 5, 11, 3403-3410.
- Hsiao WC, Shih YS (2007). Splitting variable selection for multivariate regression trees, *Statistics Probability Letters*, 77, 3, 265-271.
- Larsen DR, Speckman PL (2004). Multivariate regression trees for analysis of abundance data, *Biometrics*, 60, 2, 543-549.

- Lemon SC, Roy J, Clark MA, Friedmann PD, Rakowski W (2003). Classification and regression tree analysis in public health: methodological review and comparison with logistic regression, *Annals of Behavioral Medicine*, 26, 3, 172-181.
- Liaw A, Wiener M (2002). Classification and regression by random forest, *R news*, 2, 3, 18-22.
- Loh WY (2002). Regression trees with unbiased variable selection and interaction detection, *Statistica Sinica*, 361-386.
- Loh W Y (2011). Classification and regression trees, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 1, 1, 14-23.
- Moisen GG (2008). Classification and regression trees, *Encyclopedia of Ecology*, Elsevier, UK, 1, 582-588
- Prasad AM, Iverson LR, Liaw A (2006). Newer classification and regression tree techniques: bagging and random forest for ecological prediction, *Ecosystems*, 9, 2, 181-199.
- Questier F, Put R, Coomans D, Walczak B, VanderHeyden Y (2005). The use of CART and multivariate regression trees for supervised and unsupervised feature selection, *Chemometrics and Intelligent Laboratory Systems*, 76, 1, 45-54.
- Reibnegger G, Weiss G, Werner-Felmayer G, Judmaier G, Wachter H (1991). Neural networks as a tool for utilizing laboratory information: comparison with linear discriminant analysis and with classification and regression trees, *Proceedings of the National Academy of Sciences*, 88, 24, 11426-11430.
- Ross EH, Kearney CA (2017). Posttraumatic symptoms among maltreated youth using classification and regression tree analysis, *Child Abuse and Neglect*, 69, 177-187.
- Ruppert JL, Fortin MJ, Rose GA, Devillers R (2010). Environmental mediation of Atlantic cod on fish community composition: an application of multivariate regression tree analysis to exploited marine ecosystems, *Marine Ecology Progress Series*, 411, 189-201.
- Singh YK (2006). *Fundamental of Research Methodology and Statistics*, New International, ABD.
- Skipper PT (2014). *Exploratory Clusters of Student Technology Participation with Multivariate Regression Trees*, Yüksek Lisans Tezi, University of California, Los Angeles, ABD.
- Smyth C, Coomans D, Everingham Y (2006). Clustering noisy data in a reduced dimension space via multivariate regression trees, *Pattern Recognition*, 39,3, 424-431.
- Speybroeck N (2012). Classification and regression trees, *International Journal of Public Health*, 57, 1, 243-246.
- Strobl C, Malley J, Tutz G (2009). An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forest, *Psychological Methods*, 14, 4, 323.

Stowell D, Plumbley MD, (2010). Birdsong and C4DM: A survey of UK birdsong and machine recognition for music researchers, Centre for Digital Music, Queen Mary University of London, UK.

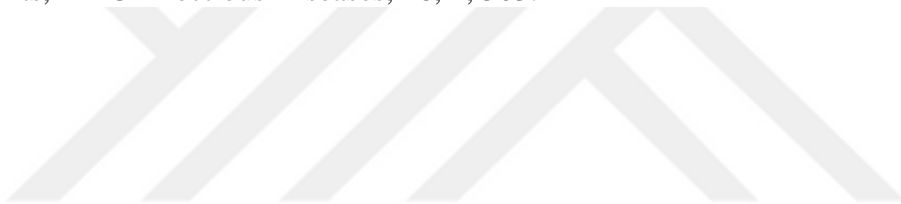
Stowell D, Plumbley MD, (2011). Learning Timbre Analogies from Unlabelled Data by Multivariate Tree Regression, a Queen Mary University of London, UK.

Suchetana B, Rajagopalan B, Silverstein J (2017). Assessment of wASTE water treatment facility compliance with decreasing ammonia discharge limits using a regression tree model, Science of the Total Environment, 598, 249-257.

Suvak B, Dulger AC, Suvak O, Yesilyurt AÖ, Gultepe B, Guducuoglu H (2017). The prevalence and impact of brucellosis in patients with hepatitis delta virus infection: inside the Brucella outbreak with cirrhosis, Arch Med Sci.

Timofeev R (2004). Classification and regression trees (CART) theory and applications, Yüksek Lisans Tezi, Humboldt University, Berlin, Almanya.

Zimmerman RK, Balasubramani GK, Nowalk MP, Eng H, Urbanski L, Jackson ML, Malosh RE (2016). Classification and regression tree (CART) analysis to predict influenza in primary care patients, BMC Infectious Diseases, 16, 1, 503.



ÖZGEÇMİŞ

Adana ili Seyhan ilçesinde 1991 yılında doğdu. İlk, orta ve lise öğrenimini Adana'da tamamladı. 2014 yılında Fırat Üniversitesi, Fen Fakültesi, İstatistik Bölümü'nden bölüm birincisi olarak mezun oldu. 2015 yılında Yüzüncü Yıl Üniversitesi, Tıp Fakültesi, Biyostatistik Anabilim Dalı'nda yüksek lisans öğrenimine başladı. 2017 yılında Erasmus Programı kapsamında Polonyada staj eğitimi gördü.



EK 1. AĞAÇ OLUŞUMU İÇİN R KODLARI

```
yeni.mrt10 <- mvpart(as.matrix(yanit) ~.,hepsi,legend=FALSE,margin=0.01,cp=0,  
xv="pick",xval=nrow(yanit), xvmult=100, which=4)
```

```
summary(yeni.mrt10)
```

```
mrt.wrap1<-MRT(yeni.mrt10,percent=10,species=colnames(yanit))
```

```
summary(mrt.wrap1)
```

```
*****  
*****
```

```
yeni.mrt11 <-mvpart(as.matrix(yanit)~.,hepsialbcik,legend=FALSE,margin=0.01, cp=0,  
xv="pick",xval=nrow(yanit), xvmult=100, which=4)
```

```
summary(yeni.mrt11)
```

```
mrt.wrap2<-MRT(yeni.mrt11,percent=10,species=colnames(yanit))
```

```
summary(mrt.wrap2)
```

```
*****  
*****
```

```
yeni.mrt12 <- mvpart(as.matrix(yanit) ~.,hepsialbhbcik,legend=FALSE, margin=0.01, cp=0,  
xv="pick", xval=nrow(yanit), xvmult=100, which=4)
```

```
mrt.wrap3<-MRT(yeni.mrt12,percent=10,species=colnames(yanit))
```

```
summary(mrt.wrap3)
```

```
summary(yeni.mrt12)
```

```
*****  
*****
```

```
yeni.mrt13 <-mvpart(as.matrix(yanit)~.,hepsialbhbglicik,legend=FALSE, margin=0.01, cp=0,  
xv="pick",xval=nrow(yanit), xvmult=100, which=4)
```

```
summary(yeni.mrt13)
```

```
mrt.wrap4<-MRT(yeni.mrt13,percent=10,species=colnames(yanit))
```

```
summary(mrt.wrap4)
```

```
*****  
*****
```

```
yeni.mrt14<-  
mvpart(as.matrix(yanit)~.,hepsialbhbglbaltcik,legend=FALSE,margin=0.01,cp=0,xv="pick",x  
val=nrow(yanit), xvmult=100, which=4)
```

```
summary(yeni.mrt14)
```

```
mrt.wrap5<-MRT(yeni.mrt14,percent=10,species=colnames(yanit))
```

```
summary(mrt.wrap5)
```

```
*****  
*****
```

```
yeni.mrt15 <- mvpart(as.matrix(yanit) ~. ,hepsialbhbglbaltpltcik,legend=FALSE,  
margin=0.01, cp=0, xv="pick",xval=nrow(yanit), xvmult=100, which=4)
```

```
summary(yeni.mrt15)
```

```
mrt.wrap6<-MRT(yeni.mrt15,percent=10,species=colnames(yanit))
```

```
summary(mrt.wrap6)
```


EK 2. ETİK KURUL BELGESİ



T.C.
YÜZÜNCÜ YIL ÜNİVERSİTESİ
Girişimsel Olmayan Klinik Araştırmalar
Etik Kurul Başkanlığı

Sayı : B.30.2.YYU.0.01.00.00/24

Tarih: 21/02/2018

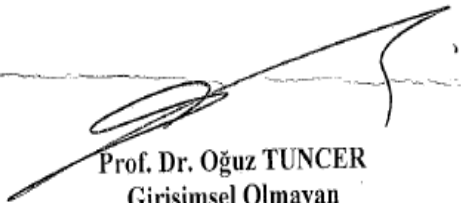
Konu : Etik Kurul Başvurunuz

Sn. Prof.Dr. Sıddık KESKİN

İlgi: 08.02.2018 tarih ve bila sayılı yazınız.

İlgi yazı ile Etik Kurulumuza sunulan “Çok Değişkenli Regresyon Ağaçları ve Uygulaması ” isimli proje özetiniz 16.02.2018 tarihinde yapılan Etik Kurul toplantımızda görüşüldü. Yapılan görüşmede çalışmanın proje özetinin içeriği tamamiyle istatistiksel analiz değerlendirmesi olduğundan adı geçen çalışmanın etik kurulla doğrudan ilişkisi saptanmamıştır.

Bilgilerinize rica ederim.


Prof. Dr. Oğuz TUNCER
Girişimsel Olmayan
Klinik Araştırmalar
Etik Kurul Başkanı

EK3. TEZ ORJİNALLİK RAPORU

YÜZÜNCÜ YIL ÜNİVERSİTESİ SAĞLIK BİLİMLERİ ENSTİTÜSÜ LİSANSÜSTÜ TEZ ORJİNALLİK RAPORU	
Tarih: 24/07/2018	
<p style="text-align: center;">Tez Başlığı / Konusu: Çok Değişkenli Regresyon Ağaçları ve Uygulaması</p> <p>Yukarıda başlığı/konusu belirlenen tez çalışmamın Kapak sayfası, Giriş, Ana bölümler ve Sonuç bölümlerinden oluşan toplam 36 sayfalık kısmına ilişkin, 24/07/2018 tarihinde şahsım/tez danışmanım tarafından TURNITIN intihal tespit programından aşağıda belirtilen filtreleme uygulanarak alınmış olan orijinallik raporuna göre, tezim benzerlik oranı % 3 (üç) dir.</p> <p>Uygulanan filtreler aşağıda verilmiştir:</p> <ul style="list-style-type: none">- Kabul ve onay sayfası hariç,- Teşekkür hariç,- İçindekiler hariç,- Simge ve kısaltmalar hariç,- Gereç ve yöntemler hariç,- Kaynakça hariç,- Alıntılar hariç,- Tezden çıkan yayınlar hariç,- 7 kelimedenden daha az örtüşme içeren metin kısımları hariç (Limit match size to 7 words) <p>Yüzüncü Yıl Üniversitesi Lisansüstü Tez Orijinallik Raporu Alınması ve Kullanılmasına İlişkin Yönergeyi inceledim ve bu yönergede belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.</p> <p>Gereğini bilgilerinize arz ederim.</p> <p style="text-align: center;">24.07.2018</p>	
Adı Soyadı: Duygu KORMAZ Öğrenci No: 149302050 Anabilim Dalı: Biyoistatistik Programı: TIP Statüsü: Y.Lisans x Doktora <input type="checkbox"/>	
DANIŞMAN ONAYI UYGUNDUR	ENSTİTÜ ONAYI UYGUNDUR
Prof. Dr. Sıddık KESKİN	