TÜRKİYE CUMHURİYETİ

ANKARA ÜNİVERSİTESİ

SAĞLIK BİLİMLERİ ENSTİTÜSÜ

RASCH MODELİNE UYGUN YANIT DESENLERİNDE BİLGİSAYAR UYARLAMALI TEST YÖNTEMİ İLE MAKİNE ÖĞRENMESİ YÖNTEMLERİNİN TAHMİN PERFORMANSLARININ BENZETİM ÇALIŞMASI İLE KARŞILAŞTIRILMASI

Emrah Gökay ÖZGÜR

BİYOİSTATİSTİK ANABİLİM DALI

DOKTORA TEZİ

DANIŞMAN

Doç. Dr. Beyza DOĞANAY ERDOĞAN

Ankara

2020

ETİK VE BEYAN

Ankara Üniversitesi

Sağlık Bilimleri Enstitüsü Müdürlüğü' ne

Doktora Tezi olarak hazırlayıp sunduğum "Rasch Modeline Uygun Yanıt Desenlerinde Bilgisayar Uyarlamalı Test Yöntemi ile Makine Öğrenmesi Yöntemlerinin Tahmin Performanslarının Benzetim Çalışması ile Karşılaştırılması " başlıklı tez; bilimsel ahlak ve değerlere uygun olarak tarafımdan yazılmıştır. Tezimin fikir/ hipotezi tümüyle tez danışmanım ve bana aittir. Tezde yer alan araştırma tarafımdan yapılmış olup, tüm cümleler ve yorumlar bana aittir.

Yukarıda belirtilen hususların doğruluğunu beyan ederim.

Öğrencinin Adı Soyadı: Emrah Gökay ÖZGÜR

Tarih:29.06.2020

İmza:

İÇİNDEKİLER

Etik ve Beyan ii

Kabul ve Onay iii

İçindekiler iv

Önsöz vi

Simgeler ve Kısaltmalar vii

Şekiller ix

Çizelgeler x

1.GİRİŞ 1
1.1. Madde Yanıt Teorisi ve Rasch Modelleri 2
1.2. Bilgisayar Uyarlamalı Test 3
1.2.1. Bilgisayar Uyarlamalı Testin İşleyişi 4
1.2.2. Bilgisayar Uyarlamalı Test Yönteminin Avantajları 6
1.2.3. Bilgisayar Uyarlamalı Test Yönteminin Dezavantajları 7
1.3.1. Classification and Regression Trees 7
1.3.1.1. CART Nasıl Çalışır? 8
1.3.1.2. CART Terminolojisi 9
1.3.1.3. Ağaç Tabanlı Modeller Doğrusal Modellerden Daha mı İyidir? 11
1.3.2. Regresyon Ağacı 12
1.3.2.1. Regresyon Ağacı ve Sınıflandırma Ağacı 12
1.3.2.2. Ayırma Sürecinde Kullanılan Kriterler 14
1.3.2.2.1. Varyans Azaltma Kriteri 14
1.3.2.2.2. Ağaç Modellemenin Anahtar Parametreleri Nelerdir ve Karar Ağaçlarında Aşırı Uyumu Nasıl Önlenebilir? 15
1.3.2.3. Budama 16
1.3.2.4. Optimal Ağaç Seçimi 17
1.3.2.5. Çapraz Geçerlilik ve Tahmin Doğruluğu 18
1.3.3. Bagging 18
1.3.3.1. Bagging Dışı Hata Tahmini 20
1.3.3.2. Değişken Önemlilik Ölçüleri 21
1.3.4. Random Forest 21
1.3.5. Gradient Boosting 24
1.3.5.1. Kayıp İşlevi 25
1.3.5.2. Zayıf Öğrenici 25
1.3.5.3. Katkı Modeli 25
1.3.5.4. Adaboost ve Gradient Boosting Arasındaki Farklar 26
1.3.6. Extreme Gradient Boosting 27
1.3.6.1. Gradient Boosting ile Extreme Gradient Boosting Arasındaki Farklar 29
2.GEREÇ VE YÖNTEM 30

3.BULGULAR 35
3.1. Kategori Sayısına Göre Senaryolara ait Yöntem Sonuçları 35
3.2. Karar Ağacı Sonuçları 54
3.3. Senaryolara ait RMSE Değerlerine Göre Yöntem Karşılaştırması 57
3.4. Senaryolara ait Ortalama Madde Sayısına Göre Yöntem Karşılaştırması 60
3.5. Senaryolara ait SKK Değerlerine Göre Yöntem Karşılaştırması 63
4.TARTIŞMA 66
5.SONUÇ VE ÖNERİLER 68
ÖZET 70
SUMMARY 71
KAYNAKLAR 72
ÖZGEÇMİŞ 74

ÖNSÖZ

Bu tez çalışmasında, Rasch modeline uygun yanıt desenlerinde Bilgisayar Uyarlamalı Test Yöntemi(BUT) ile Makinesi Öğrenmesi yöntemlerinin tahmin performansları benzetim çalışması ile incelenmiş ve Makine Öğrenmesi yöntemlerinin BUT yöntemine alternatif olarak kullanılıp kullanılamayacağı araştırılmıştır.

Doktora eğitimim ve tez çalışması süresince ilgisini, bilgisini ve desteğini esirgemeyen değerli danışman hocam Doç. Dr. Beyza DOĞANAY ERDOĞAN'a,

Tanıdığım günden itibaren her konuda bana destek olan, ilgi ve alakasını esirgemeyen değerli hocam Doç. Dr. Derya Gökmen'e

Ankara Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı Öğretim Üyeleri Prof. Dr. Yasemin YAVUZ, Prof. Dr. Atilla Halil ELHAN ve Doç.Dr. Serdal Kenan KÖSE' ye hem asistanlık hem de doktora eğitimim boyunca bana öğrettikleri her şey için,

Lisansüstü eğitimim boyunca desteğini esirgemeyen ve üzerimde büyük emeği olan Marmara Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı Öğretim Üyesi Prof. Dr. G. Nural BEKİROĞLU'na,

Asistan arkadaşlarım Serhat HAYME'ye, Funda Seher ÖZALP ATEŞ'e, Zeynep YAVUZ'a, Batuhan BAKIRARAR'a ve İrem KAR'a sabırları ve destekleri için,

Ayrıca Başkent Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı Araştırma Görevlisi Uğur TOPRAK'a tezim boyunca destekleri için,

Ankara Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı personelleri Ayhan TOPÇU ve Yaşar TOPSAKAL'a,

Her zaman beni destekleyen ve bu mesleği seçmemde büyük katkısı olan canım anneme, babama, abime, her konuda desteğini esirgemeyen değerli eşim Ebru ÖZGÜR'e ve doktora eğitimim sırasında dünyaya gelen canım oğlum Barış Tunga ÖZGÜR'e sonsuz teşekkürlerimi sunarım.

SİMGELER ve KISALTMALAR

a	Karmaşıklık Parametresi
β	Beta
β_i	Madde Zorluk Parametresi
β_{Jk}	Eşik Zorluğu
B	Eğitim Seti Sayısı
BD	Bagging Dışı
BUT	Bilgisayar Uyarlamalı Test
CART	Classification and Regression Trees
CAT	Computerized Adaptive Testing
EAP	Expectation a Posteriori (Sonsal Beklenti)
$f(x)$	Tahmin Değeri
$f_{avg}(x)$	Ortalama Tahmin Değeri
$f^b(x)$	b. Eğitim Seti Tahmin Değeri
$\widehat{f^{*b}(x)}$	b. Eğitim Seti Ortalama Tahmin Değeri
$\widehat{f_{bag}(x)}$	Bagging Yöntemi Ortalama Tahmin Değeri
GD	Gradient Descent
GBM	Gradient Boosting
HKT	Hata Kareler Toplamı
KKM	Kısmi Kredi Modeli
m	Tahminci Sayısı
MAE	Mean Absolute Error
MAP	Maximum a Posteriori (Sonsal Maksimum)
MC	Monte Carlo
ML	Maximum Likelihood (En Çok Olabilirlik)
MYT	Madde Yanıt Teorisi
n	Kişi Sayısı
θ	Teta
θ^2	Varyans
p	Değişken Sayısı

P	Olasılık
R^2	Açıklayıcılık Katsayısı
RF	Random Forest
RMSE	Root Mean Squared Error
S_i	i. Senaryo
SKK	Sınıf İçi Korelasyon Katsayısı
$\bar{X}$	Gözlem Değerinin Ortalaması
X	Gözlem Değeri
XGBoost	Extreme Gradient Boosting
τ_{Jk}	Eşik Değeri

ŞEKİLLER

Şekil 1.1. BUT İşleyiş Algoritması 4
Şekil 1.2. CART Algoritması 9
Şekil 1.3. Karar Ağacı Yapısı 10
Şekil 1.4. Random Forest Algoritması 23
Şekil 3.1. İki Kategorili S_3 İçin Karar Ağacı 54
Şekil 3.2. Üç Kategorili S_3 İçin Karar Ağacı 55
Şekil 3.3. Beş Kategorili S_3 İçin Karar Ağacı 56
Şekil 3.4. İki Kategorili Verilerde RMSE Değerlerine Göre Yöntem Karşılaştırması 57
Şekil 3.5. Üç Kategorili Verilerde RMSE Değerlerine Göre Yöntem Karşılaştırması 58
Şekil 3.6. Beş Kategorili Verilerde RMSE Değerlerine Göre Yöntem Karşılaştırması 59
Şekil 3.7. İki Kategorili Verilerde Ortalama Madde Sayısına Göre Yöntem Karşılaştırması 60
Şekil 3.8. Üç Kategorili Verilerde Ortalama Madde Sayısına Göre Yöntem Karşılaştırması 61
Şekil 3.9. Beş Kategorili Verilerde Ortalama Madde Sayısına Göre Yöntem Karşılaştırması 62
Şekil 3.10. İki Kategorili Verilerde SKK Değerlerine Göre Yöntem Karşılaştırması 63
Şekil 3.11. Üç Kategorili Verilerde SKK Değerlerine Göre Yöntem Karşılaştırması 64
Şekil 3.12. Beş Kategorili Verilerde SKK Değerlerine Göre Yöntem Karşılaştırması 65

ÇİZELGELER

Çizelge 2.1. Senaryolara Ait Bilgiler 31
Çizelge 3.1. Kategori Sayılarına Göre S_1 Sonuçları 36
Çizelge 3.2. Kategori Sayılarına Göre S_2 Sonuçları 38
Çizelge 3.3. Kategori Sayılarına Göre S_3 Sonuçları 40
Çizelge 3.4. Kategori Sayılarına Göre S_4 Sonuçları 42
Çizelge 3.5. Kategori Sayılarına Göre S_5 Sonuçları 44
Çizelge 3.6. Kategori Sayılarına Göre S_6 Sonuçları 46
Çizelge 3.7. Kategori Sayılarına Göre S_7 Sonuçları 48
Çizelge 3.8. Kategori Sayılarına Göre S_8 Sonuçları 50
Çizelge 3.9. Kategori Sayılarına Göre S_9 Sonuçları 52

1.GİRİŞ

Herhangi bir sağlık sisteminin merkezi "hasta" olarak kabul edilebilir. Son yıllarda hasta merkezli sağlık sistemleri giderek artan önem kazanmaktadır. Hastaya uygulanan klinik bir müdahalenin hasta tarafından bildirilen sonuçlarının, klinik, fizyolojik ya da hastaya bakan kişinin raporu gibi diğer sonuçlara göre gelecekte daha önemli olacağı öngörülmektedir. Fiziksel fonksiyon, semptomlar, global sağlık değerlendirmesi, psikolojik iyi hal, sosyal iyi hal, bilişsel fonksiyon, sağlığa bağlı yaşam kalitesi gibi özelliklere ilişkin ölçümler hasta tarafından bildirilen sonuçlar kullanılarak yapılabilir (Deshpande ve ark., 2011).

Özellikle kanser gibi hastalıklarda, yaşam kalitesinin hastalığın ilerlemesine bağlı olduğu durumlar olabilir. Kanserin ilerlemesine bağlı olarak hastalarda sıklıkla birden fazla semptom görülmesi, hastalığın getirdiği ekonomik yük, ev yaşamındaki olumsuz etkiler ve moral düzeyinin olumsuz etkilenmesi hastaların yaşam kalitesini etkileyen faktörlerdir. Bu gibi durumlarda hasta tarafından bildirilen sonuçlar kanser hastalarında yaşam kalitesinin belirlenmesinde yardımcıdır.

Bu örneklerdeki gibi, sağlık alanında doğrudan ölçüm yapılamayan örtük (latent) değişkenlerin ölçülmesinde ölçekler kullanılır. Ölçekler uygun maddelerden oluşan, güvenilir, geçerli ve değişime duyarlı ölçme araçlarıdır. Ölçeklerden elde edilen sonuçlar, bireylerin incelenen özellik (θ) düzeyleri hakkında değer yargılarına varmada kullanılır.

Ölçekler klasik olarak kağıt kalem kullanılarak ölçekteki tüm soruların kişi tarafından yanıtlanması biçiminde uygulanabileceği gibi, bilgisayar uyarlamalı test (BUT) tekniği ile de uygulanabilir. BUT, bilgisayar yardımıyla uygulanan ve testin uygulandığı kişinin θ düzeyini en isabetli biçimde ölçecek soruların seçilmesi üzerine adapte olan bir test biçimidir.

Bu tez çalışmasında, BUT uygulamasında yaygın olarak kullanılan algoritma (algoritma içerisinde olasılıksal model olarak Rasch modeli ve soru seçim kriteri olarak da bilgi kriterine dayalı soru seçimi) ile BUT uygulamasına yeni bir yaklaşım olarak makine öğrenmesi algoritmalarının performansları incelenecektir.

Makine öğrenmesi algoritmalarından Classification and Regression Tree (CART), Random Forest (RF), Gradient Boosting Machines (GBM) ve Extreme Gradient Boosting Machines (XGBoost) algoritmaları kullanılacaktır.

1.1. Madde Yanıt Teorisi ve Rasch Modelleri

Ölçek ya da soru bankası geliştirmede başvurulan iki teori, klasik test teorisi (classical test theory) ve madde yanıt teorisi (item response theory)'dir. Klasik test teorisinde (KTT), bir bireyin bir özelliğe sahip oluş derecesi, ölçme aracının maddelerine verilen cevapların toplamı ile belirlenir. Modern psikometrik analiz yöntemlerinden olan madde yanıt teorisi (MYT) ise, bireyin bir özelliğe sahip oluş derecesini, bireyin ölçek maddesine verdiği yanıt ile ölçek maddelerine ait parametreler arasındaki ilişkiye dayalı matematiksel bir model olarak açıklar (Kaskatı, 2011).

Georg Rasch tarafından iki sonuçlu (evet/hayır, var/yok, katılıyorum/katılmıyorum vs.) maddeler için geliştirilen Rasch modelinde, bireyin bir maddeye spesifik bir yanıt verme olasılığı kişinin yeteneği ile maddenin zorluğu arasındaki farklılığın lojistik fonksiyonu olarak hesaplanır. İki sonuçlu maddeler, tek basamaklı maddeler olarak da adlandırılır. Eğer kişi bu tek-basamağı başarı ile gerçekleştirirse 1 puan, gerçekleştiremezse 0 puan alır. Bu modelde madde zorluk parametresinin (β_i) yanı sıra bireyin θ düzeyi de tahmin edilir.

$$P_i(y = 1) = \frac{e^{(\theta - \beta_i)}}{1 + e^{(\theta - \beta_i)}} \qquad 1.1$$

θ, eğitim alanında "yetenek/bilgi düzeyi" olarak; sağlık alanında kullanılan ölçeğe ya da soru bankasına bağlı olarak "işlevsellik, özürlülük, tükenmişlik, depresyon ya da yaşam kalitesi düzeyi" olarak adlandırılabilir. β, bireyin maddeye doğru yanıt verme

olasılığının 0,50'ye karşılık gelen θ düzeyine eşittir. Rasch modellerinde amaç, verinin özel ölçüm özellikleri tanımlanan modele uyum göstermesini sağlamaktır. Madde ya da birey modele uyum sağlamazsa, bunların nedenlerinin araştırılması gerekmektedir.

İki sonuçlu ve çok sonuçlu yanıt kategorileri için geliştirilmiş farklı Rasch modelleri bulunmaktadır. Soru bankasında yer alan maddeler, iki sonuçlu olduğunda ikili Rasch (dichotomous model) modeli kullanılırken; ikiden fazla sonuçlu olduğunda kısmi kredi modeli (partial credit model) ya da sıralı ölçekli model (rating scale model) kullanılmaktadır (Kaskatı, 2011).

1.2. Bilgisayar Uyarlamalı Test

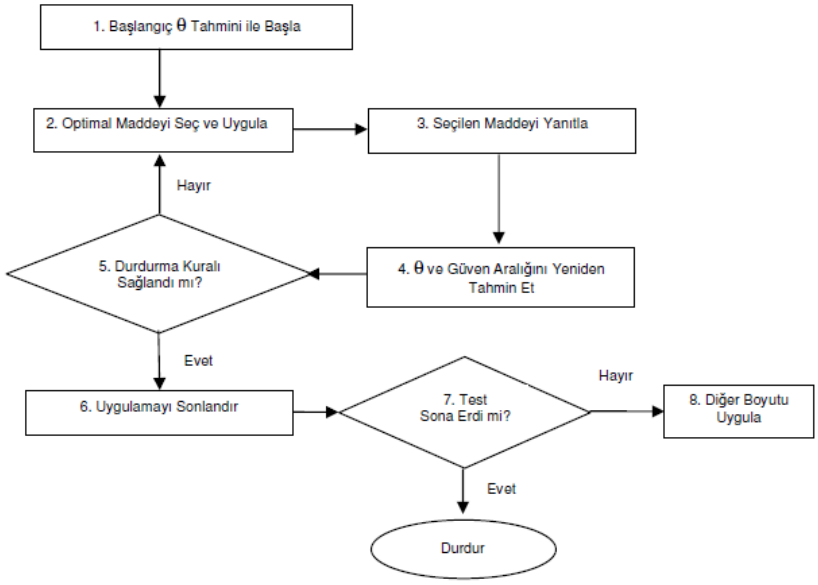
Geleneksel veya sabit formlu testte bir kağıt - kalem formu veya bir bilgisayar kullanılarak her kişiye aynı madde seti uygulanır. Uyarlamalı bir test, farklı kişilere aynı madde kavramının dışına çıkarak her kişiye göre zorluğu ayarlar ve kişilerin performanslarını değerlendirmek için farklı zorlukta maddeler uygulanır. Bu, BUT' u diğer geleneksel yaklaşımlardan ayıran en önemli özelliktir.

Uyarlamalı bir testin temellerinden biri, test deneyiminin dinamik ve test katılımcısının performansına duyarlı olmasıdır (Rezaie ve Golshan, 2015). Uyarlanabilir test fikri yeni değildir ve ilk olarak Binet tarafından 1905 yılında zeka testi ile uygulanmıştır (Öztuna, 2008). Binet bu testte birbirini izleyen her soruyu daha önceki sorulardaki performanslarına göre seçerek uygulamış ve IQ testinin hem etkili hem de yanıt verenin becerisine uygun olmasını sağlamıştır. Burada amaç gruptan ziyade herkes için performansına göre ayrı sorular seçerek kişilerin θ düzeyini tahmin edebilmektir (Öztuna, 2008).

1.2.1. Bilgisayar Uyarlamalı Testin İşleyişi

BUT, ölçme aracından seçilen maddeler kullanılarak uygulanır. BUT yönteminde kişinin ilk maddeye verdiği cevap, sonraki maddelerin sıralamasının belirlenmesinde

önemli rol oynar (Öztuna, 2008). Eğer ilk maddeye doğru cevap verildiyse, bir sonraki madde daha zor, yanlış cevap verildiyse bir sonraki madde daha kolay olacaktır. Bu yaklaşımın arkasındaki mantık, çok kolay ya da çok zor maddelerden bireyin θ düzeyi hakkında bilgi edinilemeyeceği, dolayısıyla maddelerin bireyin θ düzeyi hakkında en fazla bilgiyi sağlayacak şekilde seçildiğidir (Kaskatı, 2011). Böylelikle birey, zor maddeden daha fazla puan alırken, kolay maddeden daha düşük puan almış olur. Bireyin θ düzeyi BUT yöntemi ile test süreci boyunca iteratif olarak tahmin edilir ve maddeler elde edilen tahminlere dayalı olarak seçilir (Öztuna, 2008). BUT yönteminde madde seçimi düşük standart hata [SE(θ)] ve yüksek kesinlikle elde edilinceye kadar devam eder.



Şekil 1.1. BUT İşleyiş Algoritması (Öztuna, 2008)

Bir BUT yöntemi aşağıdaki bileşenlerden oluşur:

- Kalibre edilmiş bir soru bankası
- Başlangıç sorusunun seçilme kuralı

- Madde seçim kuralı

- θ düzeyinin kestirim yöntemi

- Durdurma kuralı

Kalibre edilmiş soru bankası kullanılarak testin uygulanacağı her kişi için belirli bir kurala bağlı olarak seçilen başlangıç sorusu ile test başlar ve kişinin θ düzeyi kestirimi yapılır. Sonrasında belirlenen madde seçim kuralına göre yeni soru seçimi yapılarak yine θ kestirimi yapılır. Soru seçimi ve θ kestirimi basamakları belirlenen durdurma kriteri sağlanana kadar iteratif olarak devam eder. Durdurma kuralı karşılandığında test biter ve en son yapılan θ kestirimi kişinin θ düzeyi kestirimi olarak alınır. Bütün BUT algoritmaları bu temel formata dayalıdır. Ancak soru seçme kuralları gibi bazı detaylar değişkenlik gösterir.

θ düzeyi kestirimi ve soru seçiminde, klasik istatistik ve Bayesci istatistik yaklaşımları kullanılabilir. θ düzeyi için en sık kullanılan kestirim yöntemleri, klasik istatistik çerçevesinde geliştirilen En Çok Olabilirlik (Maximum Likelihood, ML) ve Bayesci istatistik çerçevesinde geliştirilen Sonsal Beklenti (Expectation a Posteriori, EAP) ile Sonsal Maksimum (Maximum a Posteriori, MAP) kestirim yöntemleridir. Soru seçme yöntemleri o anki θ kestiriminde bilgi kriterini maksimum yapan sorunun seçilmesine dayalıdır. En çok kullanılan klasik istatistik soru seçme kriteri Fisher'in en çok bilgi kriteri (Fisher's Maximum Information Criterion) ve Bayesci istatistik kriteri ise sonsal dağılım ağırlıklı bilgi kriteri (Maximum Posterior Weighted Information) yöntemleridir. Bu yöntemlere ek olarak, özellikle eğitim alanında, aynı soruya maruziyet, test kapsamındaki her içerikten soru gelmesi zorunluluğu gibi pratikte göz önünde bulundurulması gereken durumlar da vardır. Ancak sağlık alanında soru seçiminde bu tip pratikte göz önüne alınacak durumlar bulunmamaktadır.

BUT uygulamasındaki θ kestiriminde ve bilgi kriteri içinde yer alan olasılıksal model MYT modelleridir.

1.2.2. Bilgisayar Uyarlamalı Test Yönteminin Avantajları

Bilim insanları BUT yönteminin kullanım kolaylığı sağlayan avantajlarından bahsetmişlerdir. BUT yönteminde test esnektir, puanlar hemen elde edilebilir ve bundan dolayı sınava girenleri motive edebilir (Rudner, 1998). Maddeler kendi seviyelerine uygun olacağı için test kaygıları azaltılabilir (Rezaie ve Golshan, 2015). Ayrıca, verimlilik BUT yönteminin ana avantajıdır (Rezaie ve Golshan, 2015). BUT geleneksel yöntemlere göre maliyet tasarrufu sağlar. Geleneksel yöntemlerde, farklı öğrenci gruplarına farklı test grupları verilmelidir ve farklı test gruplarının hazırlanması zaman alıcı olabilir.

BUT kullanımı, test hazırlama ve işaretleme sürelerini kısaltır ve sonuçların tutarlılığını artırabilir (Rezaie ve Golshan, 2015). BUT uygulamasında testlerin uygulanması kağıt ve kalem kullanılarak yapılan testlere göre daha fazla zaman tasarrufu sağlar. Ayrıca, BUT'da kullanılan madde sayısı ve bu maddelerin cevaplanması için gereken süre geleneksel kağıt ve kalem kullanılarak yapılan testlerden daha kısadır (Rezaie ve Golshan, 2015). BUT yönteminde geleneksel yönteme göre testi değerlendiren kişi ölçüm hatasına neden olan bir faktör olarak ortadan kalkmış olur (Rudner, 1998).

Her test, sınava giren her bir kişinin seviyesine göre belirlendiğinden geleneksel testlere göre BUT'da daha fazla bilgi elde edilebilir. BUT yönteminin önemli avantajlarından biri de ölçüm hassasiyetidir. Bu ölçüm hassasiyeti farklı düzeydeki maddelerin kullanılmasından kaynaklanmaktadır. Sınava giren kişilerin puanı hem doğru cevaplanan soru yüzdesi hem de soruların zorluk düzeylerine göre belirlenmektedir. Geleneksel yöntem ortalama yetenekteki kişiler için daha doğru sonuçlar verirken, BUT geniş bir yetenek aralığında performans gösterir (Rudner, 1998).

1.2.3. Bilgisayar Uyarlamalı Test Yönteminin Dezavantajları

BUT yönteminin avantajlarının yanı sıra çeşitli dezavantajları vardır ve bunlar çeşitli teknik ve yöntemsel sorunları meydana getirmektedir.

BUT yöntemi bütün madde çeşitleri için kullanılamayan MYT modeline dayanmaktadır. Bundan dolayı, BUT yöntemi açık uçlu sorular ve kalibre edilemeyen maddeler için uygun değildir (Rezaie ve Golshan, 2015). Ayrıca, BUT'un diğer bir önemli dezavantajı, sınava giren kişinin cevapladığı sorulara tekrar dönmesine ve cevapları değiştirmesine izin verilmemesidir, çünkü bir sonraki maddeler önceki cevaplanan maddelere göre seçilir (Rudner, 1998).

Madde kalibrasyonu BUT'un tahmin başarısını etkileyen önemli bir faktördür. Maddeler zorluk ve yetenek ölçeğinde uygun şekilde kalibre edilmezse, test ne geçerli ne de güvenilir olacaktır (Rudner, 1998). Ayrıca, her zorluk seviyesinde tekrarlanan ölçümleri mümkün kılmak için birkaç madde kullanılmalıdır. Sonuç olarak, büyük bir soru bankası oluşturulmalıdır (Rezaie ve Golshan, 2015).

1.3.Makine Öğrenmesi Yöntemleri

Bu tez çalışması kapsamında makine öğrenmesi yöntemlerinden Classification and Regression Tree (CART), Random Forest (RF), Gradient Boosting Machines (GBM) ve Extreme Gradient Boosting Machine (XGBoost) yöntemleri kullanılacaktır.

1.3.1.Classification and Regression Tree (CART)

CART, hem sınıflandırma hem de regresyon görevlerini gerçekleştirebilen çok yönlü makine öğrenmesi algoritmasıdır (Michel ve ark., 2018). Bunlar karmaşık veri setleri için çözüm oluşturan çok güçlü algoritmalardır. Ayrıca CART, günümüzde

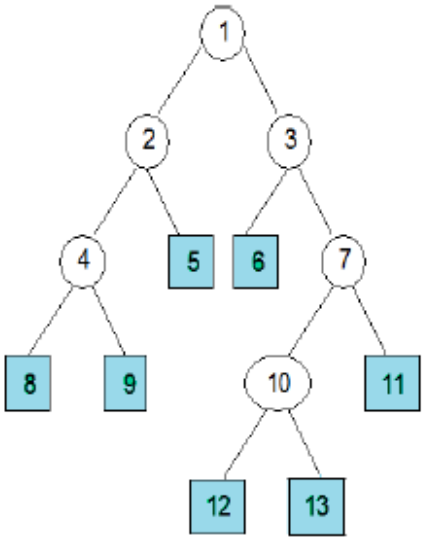
mevcut en güçlü makine öğrenimi algoritmaları arasında olan RF yönteminin temel bileşenleridir.

Ağaç tabanlı öğrenme algoritmaları en iyi ve en çok kullanılan denetimli öğrenme yöntemlerinden biri olarak kabul edilir (James ve ark., 2014, s:304.). Ağaç tabanlı yöntemler, kestirim amaçlı kullanılan modelleri yüksek doğruluk, kararlılık ve yorumlama kolaylığı ile güçlendirir. Doğrusal modellerin aksine, doğrusal olmayan ilişkileri de oldukça iyi gösterir. Eldeki her türlü problemi (sınıflandırma veya regresyon) çözmek için uyarlanabilirler (Loh, 2011).

CART, RF, Boosting gibi yöntemler her türlü veri bilimi probleminde popüler olarak kullanılmaktadır. Bu nedenle, her veri bilimcisi için bu algoritmaları öğrenmek ve bunları modelleme için kullanmak gün geçtikçe önemli hale gelmiştir.

1.3.1.1. CART Nasıl Çalışır?

CART, çoğunlukla sınıflandırma problemlerinde kullanılan ve önceden tanımlanmış bir bağımlı değişkene sahip olan bir tür denetimli öğrenme algoritmasıdır. Hem kategorik hem de nicel, bağımlı ve bağımsız değişkenler için çalışır. Bu teknikte, popülasyon veya örneklem, bağımsız değişkenlerindeki en önemli ayırt edici değişken temel alınarak iki veya daha fazla homojen kümeye (veya alt popülasyonlara) ayrılır.



Şekil 1.2. CART Algoritması (Sevimli Saitoğlu, 2015)

CART, bağımlı değişkenin türüne göre (kategorik ya da nicel) ikiye ayrılır. Bağımlı değişken kategorik olduğunda kullanılacak CART, sınıflama ağacı olarak adlandırılır. Örneğin kalp krizi geçirme durumu cinsiyet, yaş ve beslenme tipine bağlı olarak incelendiği durumda sınıflama karar ağaçlarından faydalanılır. Bu durumda bağımlı değişken kalp krizi geçirme durumu (evet/hayır) biçiminde kategorilere sahiptir. Bağımlı değişken nicel olduğu durumda ise regresyon karar ağaçları kullanılmaktadır. Örneğin kanser tanısında kullanılan bir biyo-belirteç düzeyi, lenfosit yüzdesine, cinsiyete ve hücre sayısına bağlı olarak kestirilmek istendiğinde regresyon ağaçları kullanılabilir.

1.3.1.2. CART Terminolojisi

CART'da kullanılan karar ağaçlarında sıklıkla geçen terimler aşağıda verilmiştir.

- **Kök (Root Node):** Tüm popülasyonu veya örneklemi temsil eder ve kök sonrasında iki veya daha fazla homojen alt kümeye bölünür.

- **Ayırma – Bölünme (Splitting):** Bir düğümü iki ya da daha fazla alt düğüme bölme sürecidir.

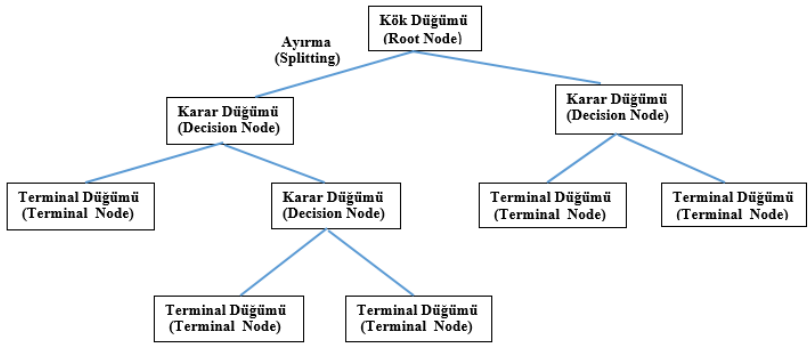
- **Karar Düğümü (Decision Node):** Bir alt düğüm başka alt düğümlere ayrıldığında, karar düğümü adı verilir.

- **Yaprak / Terminal Düğüm (Leaf/ Terminal Node**): Bölünmeyen düğümlere Yaprak veya Terminal düğümü denir.

- **Budama (Pruning):** Bir karar düğümünün alt düğümlerini kaldırdığımızda, bu işleme budama denir. Ayırma işleminin tersi olarak da düşünülebilir.

- **Dal (Branch) / Alt Ağaç (Sub-Tree):** Tüm ağacın bir alt bölümüne dal veya alt ağaç denir.

- **Üst / Ebeveyn(Parent) ve Alt / Çocuk Düğüm (Child Node):** Alt düğümlere bölünmüş bir düğüme üst düğüm, bölünen alt düğümlere ise alt düğüm denir.



Şekil 1.3. Karar Ağacı Yapısı

Bunlar genellikle karar ağaçları için sıklıkla kullanılan terimlerdir. Her algoritmanın avantaj ve dezavantajları vardır.

Karar ağaçlarının en belirgin avantajları (Dhiraj, 2019),

1. Sonuçların kolay anlaşılır olması
2. Verideki ilişki yapılarının incelenmesinde hızlı çözüm sağlaması ve veri setindeki çok sayıda değişken olduğu durumda bile, bağımlı değişken üzerinde en etkili değişkeni bulmada kolaylık sağlaması
3. Diğer yöntemlere göre daha az veri temizleme gerektirmesi, uç değerlerden ve kayıp gözlemlerden belli bir düzeye kadar etkilenmemesi
4. Parametrik olmayan bir yöntem olması, böylelikle popülasyon dağılımı hakkında ve sınıflama yapısı hakkında bir varsayıma sahip olmaması
5. Hem kategorik hem de nicel bağımlı değişkenler için çözüm sağlaması

olarak sayılabilir.

Bu avantajların yanında karar ağaçlarının taşıdığı bazı dezavantajlar da mevcuttur:

1. Aşırı uyum karar ağacı modelleri için en sık karşılaşılan zorluklardan biridir. Bu sorun, model parametreleri ve budama üzerinde kısıtlamalar konularak çözülebilir.
2. Nicel bağımlı değişkenler ile çalışırken karar ağacı, değişkenleri farklı kategorilerde sınıflandırdığında bilgi kaybı ortaya çıkmaktadır.

1.3.1.3. Ağaç Tabanlı Modeller ve Doğrusal Modeller

1. Bağımlı ve bağımsız değişken arasında doğrusal bir ilişki varsa doğrusal regresyon ağaç tabanlı modele göre daha iyi sonuç verecektir.
2. Bağımlı ve bağımsız değişkenler arasında doğrusal olmayan ve karmaşık bir ilişki varsa, ağaç modeli, klasik bir regresyon yönteminden daha iyi performans gösterir.
3. Açıklanması kolay bir model kurmak gerekirse, karar ağacı modeli her zaman doğrusal bir modelden daha iyi sonuç verir. Karar ağacı modellerinin yorumlanması doğrusal regresyondan daha kolaydır.

1.3.2. Regresyon Ağacı

Bağımlı değişken nicel olduğunda kullanılacak CART modeli regresyon ağacıdır (Breimann ve ark., 1998, s:204). Regresyon ağacı karmaşık veri setlerinde daha anlaşılır sonuçlar vermektedir (Breimann ve ark., 1998, s:204). Bu yöntemde heterojen veri setleri, belirlenmiş bağımlı değişkene göre homojen alt gruplara ayrılır. Alt gruplara ayırma işlemi belirli bir takım karar verme kuralları ile yapılmaktadır. Dallanma, kök düğümden terminal düğümlere doğru saflık ölçümleri yapılarak gerçekleştirilir. Entropi, ki-kare vb. değerler ile oluşan düğümlerdeki homojenlik yapısı incelenerek bir dallanma olup olmayacağı kontrol edilir (Dangeti, 2017, s:30). Terminal düğümde ortaya çıkan değerler bağımlı değişkenin tahmini değerleridir.

Regresyon ağaçlarında aşırı uyum sorunu ortaya çıkabilmektedir. Ayırma işlemleri ve dallanmalar ne kadar fazla olursa o kadar aşırı uyum meydana gelecektir. Bu durumda karmaşıklık parametresi sayesinde hata belirlenen bir değere gelindiğinde dallanma durur ve böylece aşırı uyum sorunu ortadan kalkmış olur (Michel ve ark., 2018).

Regresyon ağaçları, kullanılan veri için mükemmel sonuçlar vermesine rağmen bu sonucu genellemek hatalara yol açabilmektedir. Bir veri seti için oluşturulan modeli başka bir veri setine kullanmak hatalı tahminlere yol açabilmektedir. Bu nedenle her kurulan modeli kendi veri setinde değerlendirmek regresyon ağacı yöntemi için daha sağlıklı sonuçlar ortaya çıkmasını sağlar.

1.3.2.1.Regresyon Ağacı ve Sınıflandırma Ağacı

Regresyon ağacı ve sınıflandırma ağacı arasındaki farklar şu şekildedir (Brownlee, 2019).

1. Regresyon ağaçları, bağımlı değişken nicel olduğu durumda kullanılır. Sınıflandırma ağaçları ise daha önce de belirtildiği gibi, bağımlı değişken kategorik olduğunda kullanılır.

2. Regresyon ağacında, eğitim verilerinde terminal düğümlere karşılık gelen değer, o bölgeye (yani, bağımsız değişkenlerinin değerlerine göre oluşturulmuş kümelerden birine, örn. 40 yaş altı kadınlar) düşen gözlemlerin ortalama değeridir. Dolayısıyla, daha önce ortaya çıkmamış bir gözlem bu bölgeye düşerse, bağımlı değişkeninin kestirimi bu ortalama değer olarak alınır.

3. Sınıflandırma ağacında, eğitim verilerinde terminal düğümüne karşılık gelen bağımlı değişkenin kategorisi, o bölgeye düşen gözlemin tepe değeridir. Bu nedenle, daha önce ortaya çıkmamış gözlem o bölgeye düşerse, bağımlı değişkeninin kestirimi bu tepe değeri olarak alınır.

4. Her iki ağaç tipi de, bağımsız değişkenleri farklı ve birbiriyle çakışmayan bölgelere ayırır.

5. Her iki ağaç tipi de yinelemeli iki sonuçlu bölünme olarak bilinen yukarıdan aşağıya doğru bir yaklaşımı takip eder. Bu yaklaşımda, en yukarıda tüm gözlemler ile ağaç başlar ve oluşturulacak yeni bölme için en iyi bağımsız değişken aranır. En iyi bağımsız değişkeninin belirlenmesi daha ileri ki bölme işlemleri düşünülmeden, sadece mevcut bölme işlemi için yapılır.

6. Bu bölme işlemi, kullanıcı tanımlı bir durdurma kriterine ulaşılıncaya kadar devam eder. Örneğin: algoritmanın, düğüm başına gözlem sayısı 50'den az olduğunda durması söylenebilir.

7. Her iki durumda da, ayırma işlemi durdurma kriter(ler)ine ulaşılana kadar tamamen büyütülmüş ağaçlarla sonuçlanır. Ancak, tam olarak büyümüş ağacın verilere aşırı uyumu muhtemeldir. Bu da henüz gözlenmemiş verilerde düşük

doğruluğa sahip sonuçlara neden olur. Böyle bir sorun 'budama' yoluyla aşılabilir. Budama, doğruluğu artırmak için gerekli olmayan düğümleri silen tekniklerden biridir.

1.3.2.2. Ayırma Sürecinde Kullanılan Kriterler

Ağaçtaki düğümlerin nasıl oluşturulacağı, bir ağacın doğruluğunu büyük ölçüde etkiler. Düğüm oluşturmada kullanılan kriterler, sınıflandırma ve regresyon ağaçları için farklıdır. Karar ağaçları, bir düğümü iki veya daha fazla alt düğüme bölmek için birden fazla algoritma kullanır. Alt düğümlerin oluşturulması ya da ayırma işlemi, sonuçta ortaya çıkacak yeni alt düğümlerin homojenliğini artırır. Başka bir deyişle, düğümün saflığının (homojenliğinin) bağımlı değişkene göre arttığı söylenebilir. Karar ağacı, düğümleri mevcut tüm değişkenlere böler ve daha sonra en yüksek homojenliğe neden olan bölmeyi seçer. Algoritma seçimi aynı zamanda bağımlı değişkenlerin türüne göre de değişiklik gösterir. Karar ağacında en sık kullanılan beş algoritmadan Gini İndeksi, Bilgi Kazanımı, Entropi ve Ki-kare sınıflama ağaçlarında kullanılırken, Varyans Azaltma Kriteri, regresyon ağaçları için kullanılmaktadır (Dangeti, 2017, s:30).

1.3.2.2.1. Varyans Azaltma Kriteri

Varyan azaltma kriteri, bağımlı değişken nicel olduğu durumda (regresyon problemleri için) kullanılan bir algoritmadır. Bu algoritma, en iyi bölünmeyi seçmek için standart varyans formülünü kullanır. Düşük varyansa sahip bölme, popülasyonu bölme ölçütleri olarak seçilir:

$$\text{Varyans} = \frac{\sum(X-\bar{X})^2}{n} \qquad 1.2$$

1. Her düğüm için varyans hesaplanır.

2. Her bölüme ait varyans, her düğüm varyansının ağırlıklı varyansı olarak hesaplanır.

1.3.2.2.2 Ağaç Modellemenin Anahtar Parametreleri Nelerdir ve Karar Ağaçlarında Aşırı Uyum Nasıl Önlenebilir?

Aşırı uyum, karar ağaçları modellenirken karşılaşılan en önemli zorluklardan biridir. Bir karar ağacı için bir sınır yoksa eğitim setinde % 100 doğruluk verecektir çünkü en kötü durumda, her gözlem için bir yaprak yapacaktır. Bu nedenle, bir karar ağacını modellerken aşırı uyumu önlemek çok önemlidir ve iki şekilde yapılabilir:

1. Ağaç boyutunda sınırlamalar yapma
2. Ağaç budama

Ağaç boyutunda sınırlama yapmak için çeşitli sonlandırma koşulları kullanılabilir. Bunlar;

- Bir düğüm bölünmesi için minimum gözlem sayısı:

Belirli bir düğümde bölme işlemi için gerekli olan minimum gözlem sayısıdır. Bu değer, modelin veriye aşırı uyumunu kontrol etmek için kullanılır. Belirlenecek yüksek gözlem değeri, modelin kullanılan veriye özgü olabilecek ilişkileri öğrenmesini önler. Ancak çok yüksek bir değer belirlenmesi durumunda modelin veriye uyumu kötü yönde etkilenecektir. Bu nedenle bu değer çapraz geçerlilik (ÇG) kullanılarak bulunabilir.

- Bir terminal düğümü için minimum örneklem

Bir terminal düğümünde gerekli olan minimum gözlem sayısıdır. Bu değer de modelin veriye aşırı uyumunu kontrol etmek için kullanılır. Dengesiz sınıf problemleri

için genel olarak daha düşük değerler seçilmelidir çünkü azınlık sınıfının çoğunlukta olacağı bölgeler çok küçük olacaktır.

- Maksimum ağaç derinliği

Bir ağacın maksimum derinliğidir. Derinlik aşırı uyumu kontrol etmek için kullanıldığından, modelin belirli bir gözleme özgü ilişkileri öğrenmesini sağlayacaktır. ÇG kullanılarak ayarlanmalıdır.

- Maksimum terminal düğüm sayısı

Bir ağaçta bulunan maksimum terminal düğüm veya yaprak sayısıdır. Maksimum derinlik yerine de tanımlanabilir. İkili ağaçlar oluşturulduğundan, 'n' derinliği maksimum 2^n yaprak üretecektir.

- Ayırma için dikkat edilmesi gereken maksimum özellikler

En iyi bölünme aranırken dikkat edilmesi gereken özelliklerin sayısıdır. Bunlar rastgele seçilir. Temel bir kural olarak, toplam özellik sayısının karekökü mükemmel çalışır ancak toplam özellik sayısının % 30 ya da %40'ını kontrol etmek gerekmektedir. Daha yüksek değerler aşırı uyuma neden olabilir.

1.3.2.3. Budama

Bir ağaç oluşturulurken erken durdurma kriteri olmadığında ağaç fazladan oluşabilir. Karar ağaçlarında son altkümeler sonuç değişkeni homojen olana kadar tekrar tekrar daha küçük alt kümelere bölünecektir. Uygulamada bu genellikle son altkümelerin (ağacın yaprakları olarak bilinir) her birinin yalnızca bir veya birkaç veri noktasından oluştuğu anlamına gelir. Ağaç, verileri tam olarak öğrenmiştir, ancak çok az farklılık gösteren yeni bir veri noktası varsa bu iyi tahmin edilemeyebilir.

Karar ağacı oluşturulurken, final modeline en yakın ve basit modelleri kurmak için "Cost-Complexity Pruning" metodu kullanılır. Bu metot budama sürecinde kademeli olarak artan bir karmaşıklık parametresine (α) dayanır. Oluşan çocuk düğümlerin budandıktan sonra tahmini, yanlış sınıflama değerinin, ağacın derinliğinden α kat kadar küçük olmasını temel alır. Bu yüzden α, düğümlerin ayrılmasıyla derinliği azalan modelin ne kadar doğruluk kazancı sağlayacağının bir ölçüsüdür. Ayrıca α büyüdükçe daha fazla düğüm budanır ve daha basit ağaç modelleri elde edilir.

Ağaç budamak için 2 yöntem kullanılır:

- Minimum Hata

Ağaç, ÇG hatasının minimum olduğu noktaya kadar budanır. ÇG, verilerin çoğuna sahip bir ağaç oluşturma ve daha sonra karar ağacının doğruluğunu test etmek için verilerin kalan kısmını kullanma işlemidir.

- En Küçük Ağaç

Ağaç minimum hatadan biraz daha geriye çekilir. Teknik olarak budama, minimum hatanın 1 standart hatası içerisinde çapraz doğrulama hatası olan bir karar ağacı oluşturur. Küçük ağaç, hatada küçük bir artış olsa bile daha anlaşılırdır.

1.3.2.4. Optimal Ağaç Seçimi

Eğitim setinden oluşturulan maksimum ağaç (her bir düğümde 1 gözlem kalana kadar) bize en yüksek doğruluğu içeren modeli verir. Aslında bu yüksek performans, bir aşırı uyum problemidir. Bu problemin üstesinden gelmek için en iyi karmaşıklık parametresi değeri aranır.

1.3.2.5. Çapraz Geçerlilik ve Tahmin Doğruluğu

ÇG, modelin doğru sınıflama performansını bağımsız bir veri setine gerek olmadan test etmek için kullanılan ve hesaplama açısından oldukça yoğun bir yöntemdir. Burada veri, rastgele olarak K adet bölüme ayrılır. Parçalar benzer dağılım gösteren alt kümelere ayrılmıştır. Bu K adet parçanın K-1 tanesi eğitim seti olması için birleştirilir ve geriye kalan parça ise test seti olarak saklanır. Test seti üzerinde tahminler ve doğruluk oranları hesaplanır. Tüm model kurma süreci K defa tekrarlanır. Modelin doğruluk oranı K tane doğruluk oranının ortalaması kadar olur. Böylece ağaç inşası ve budama süreçleri K defa tekrar edilir. Böylece model defalarca test edilmiş olur.

1.3.3. Bagging

Bootstrap, istatistikte sıklıkla kullanılan önemli bir yöntemdir. İlgilenilen niceliğin standart sapmasını doğrudan hesaplamanın zor veya imkansız olduğu birçok durumda kullanılır. Karar ağaçları gibi istatistiksel öğrenme metotlarını geliştirmek için ise bootstrap tamamen farklı bir şekilde kullanılmaktadır.

Karar ağaçları, yüksek varyans değerinden dolayı farklı sonuçlar göstermektedir. Bu, eğer eğitim verilerini rastgele iki parçaya bölersek ve her iki parçaya da bir karar ağacı uydurursak, elde ettiğimiz sonuçların oldukça farklı olabileceği anlamına gelir. Buna karşılık, düşük varyanslı bir yöntem, farklı veri setlerine tekrar tekrar uygulandığında benzer sonuçlar verecektir. Örneklem büyüklüğü/parametre sayısı (n/p) oranının orta derecede büyük olması durumunda doğrusal regresyon düşük varyansa sahip olma eğilimindedir (James ve ark., 2014, s:316). Bootstrap ve Bagging, istatistiksel bir öğrenme yönteminin varyansını azaltmak amacıyla kullanılan yöntemlerdir (James ve ark., 2014, s:316).

Her biri σ^2 varyanslı K_1,……,K_n n tane bağımsız gözlem olsun. Gözlemlerin $\bar{k}$ ortalamasının varyansı σ^2/n şeklinde ifade edilir. Başka bir deyişle, bir dizi gözlemin

ortalaması alınarak varyans azalır. Dolayısıyla, varyansı azaltmanın ve istatistiksel bir öğrenme yönteminin tahmin doğruluğunu artırmanın doğal bir yolu, popülasyondan birçok eğitim seti alıp her eğitim setini kullanarak ayrı bir tahmin modeli oluşturmak ve elde edilen tahminlerin ortalamasını almaktır (James ve ark., 2014, s:316). Yani B tane ayrılmış eğitim setini kullanan $\widehat{f^1}(x), \widehat{f^2}(x), \ldots\ldots, \widehat{f^B}(x)$ hesaplanır ve düşük varyanslı istatistiksel bir öğrenme modeli elde etmek için bunların ortalaması alınır (James ve ark.,2014,s:316).

$$\widehat{f_{avg}}(x) = \frac{1}{B}\sum_{b=1}^{B} \widehat{f^b(x)} \qquad 1.3.$$

Ancak bu pratik bir yöntem değildir, çünkü genellikle çoklu eğitim setlerine ulaşmak mümkün olmamaktadır. Bunun yerine, tek eğitim veri setinden tekrarlanan örnekler alarak bootstrap yapılır. Bu yaklaşımda B tane farklı bootstrap edilmiş eğitim veri setleri üretilir. $\widehat{f^{*b}}(x)$ elde etmek için b. Boostrap eğitim setine yöntem uygulanır ve son olarak bütün tahminlerin ortalaması alınır (James ve ark., 2014, s:317).

$$\widehat{f_{bag}}(x) = \frac{1}{B}\sum_{b=1}^{B} \widehat{f^{*b}(x)} \qquad 1.4.$$

Bu da Bagging olarak adlandırılır.

Bagging birçok regresyon yöntemi için tahminleri iyileştirebilirken, karar ağaçları için özellikle kullanılmaktadır. Bagging yöntemini regresyon ağaçlarına uygulamak için, B tane bootstrap'tan türetilmiş eğitim setlerini kullanarak B tane regresyon ağacı oluşturulur ve elde edilen tahminlerin ortalaması alınır. Bu ağaçlarda budama işlemi yapılmaz. Bu nedenle her bir ağaç yüksek varyansa sahip olur ancak yanlılıkları düşüktür. Bu B tane ağacın ortalaması varyansı azaltır. Bagging yöntemi çok sayıda ağacı tek bir yöntemde birleştirerek kesinliği arttırdığı gösterilmiştir (James ve ark.,2014,s:317).

Nicel bir sonucu tahmin etmek için, regresyon yöntemlerinden Bagging yöntemi kullanılır. Bagging yönteminin bağımlı değişkenin nitel olduğu bir sınıflandırma probleminde kullanılabilmesi için birkaç olası yaklaşım vardır, ancak en çok kullanılan yöntemde belirli bir test gözlemi için, B tane ağaçtan her biri tarafından tahmin edilen sınıf kaydedilir ve en çok seçilen sınıf B tahminleri arasında en sık görülen sınıf olmaktadır.

Test hata oranı, bootstrap ile elde edilen eğitim veri setleri kullanılarak oluşturulan B tane ağacın bir fonksiyonu olarak gösterilir. Bagging test hata oranı bu durumda tek bir ağaçtan elde edilen test hata oranından biraz daha düşük çıkmaktadır. Ağaç sayısı bagging yönteminde kritik bir parametre değildir. Çok büyük bir B değeri kullanmak aşırı uyuma neden olmaz. Uygulamada, hatanın kabul edilebilir olduğu düzeyde B değeri kullanılır.

1.3.3.1. Bagging Dışı Hata Tahmini

Çapraz doğrulama veya doğrulama seti yaklaşımına ihtiyaç duyulmadan, bagging ile elde edilmiş bir modelin test hatasını tahmin etmenin sıkça kullanılan yöntemleri vardır. Bagging yönteminde, ağaçlar tekrar tekrar gözlemlerin bootstrap alt kümelerine uydurulur.

Genel olarak bagging ile elde edilen her ağaç gözlemlerin yaklaşık üçte ikisini kullanır (James ve ark., s:318). Belirli bir bagging ile elde edilen ağaçta kullanılmayan üçte birlik gözlemlere bagging dışı (BD) gözlemler denir. Gözlemin BD olduğu ağaçların her biri kullanılarak i'nci gözlemin yanıtını tahmin edilebilir. Bu, i'nci gözlem için B/3 tane tahminini verecektir. i'nci gözlem için tek bir tahmin elde etmek amacıyla, bu öngörülen yanıtların ortalaması alınabilir (regresyon) veya çoğunluk oyu alınabilir (sınıflandırma). Bu, i'inci gözlem için tek bir BD tahminine yol açar. Bir BD tahmini, genel BD gözlemlerin hata kareler ortalaması (HKO) veya sınıflandırma hatasının hesaplanabildiği n gözlemin her biri için bu şekilde elde edilebilir.

Her bir gözlemin yanıtı elde edilmemiş ağaçlardan tahmin edildiği için BD hatası bagging yöntemi ile elde edilmiş model hatasının geçerli bir tahminidir (James ve ark., s:318). Test hatasını tahmin etmek için BD yaklaşımı, çapraz doğrulamanın hesaplama açısından zahmetli olacağı büyük veri setlerinde bagging yaparken özellikle uygundur (James ve ark., s:318).

1.3.3.2.Değişken Önemlilik Ölçüleri

Bagging yöntemi genel olarak tek bir ağaç kullanarak tahmin üzerine daha fazla doğruluk sağlar. Ancak sonuç olarak ortaya çıkan modeli yorumlamak zor olabilir. Karar ağaçlarının bagging yöntemine göre en büyük avantajlarından biri yorumlamasının kolayca yapılmasını sağlayan diyagramlar ortaya çıkarmasıdır. Ancak çok sayıda ağaca bagging yöntemini uyguladığımızda, ortaya çıkan istatistiksel öğrenme yöntemini tek bir ağaç kullanarak temsil etmek mümkün olmaz ve hangi değişkenlerin yöntem için en önemli olduğu açıkça belirlenememektedir. Bu yüzden bagging yönteminde yorumlanabilirlik iyi olmamasına rağmen tahmin doğruluğu yüksektir.

Bagging yöntemine göre oluşturulan ağaçların yorumlanması tek bir ağaçtan çok daha zor olsa da, hata kareler toplamı (HKT) (regresyon problemi için) veya Gini endeksini (sınıflandırma problemi için) kullanarak her bir maddenin öneminin genel bir özetini elde etmek mümkündür (Hastie ve ark., s:593). Bagging yönteminde B tane ağacın ortalaması alındığında belirli bir madde için HKT azalır. Bundan dolayı büyük değerler o maddenin öneminin daha fazla olduğunu gösterir (James ve ark., s:318).

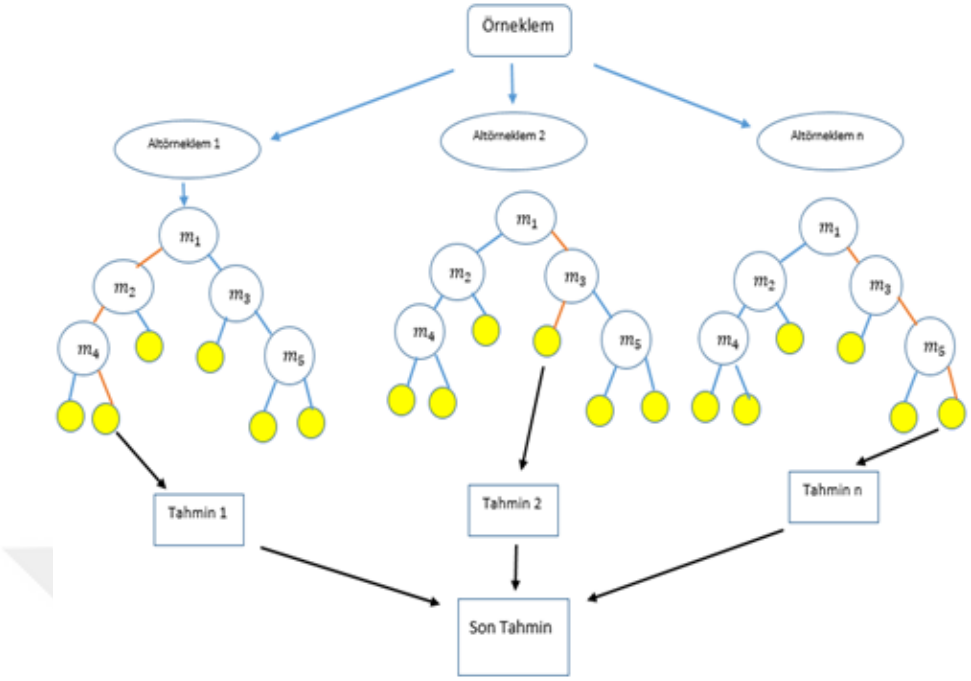
1.3.4. Random Forest

Topluluk öğrenme yöntemlerinden olan RF, Bagging ve Random Subspace (RS) yöntemlerinin birleştirilmesiyle meydana gelir. Bagging yönteminde sadece örnek seçimi rastgele yapılırken Random Forest yönteminde hem örnek hem de değişken

seçiminde rastgelelik olur (Keskin, 2018). Random Forest yönteminde örnek seçimi bootstrap ile değişken seçimi RS ile yapılmaktadır.

RF yönteminde Bagging yönteminde olduğu gibi, bootstrap eğitim örnekleri üzerinde bir dizi karar ağacı oluşturulur. Ancak bu karar ağaçlarını oluştururken, bir ağaçtaki her bir bölünme düşünüldüğünde, tüm p tahmin ediciler yerine, bu tahmin edicilerin m büyüklüğünde rastgele bir altkümesi, bölünmede kullanılacak aday tahmin ediciler olarak seçilir. Bölme, bu m tahmin edicilerin sadece birini kullanmasına izin verir. Her bölünmede yeni bir m tahmincisi örneği alınır ve tipik olarak sınıflama problemlerinde $m \approx \sqrt{p}$, regresyon problemlerinde $m = p/3$ seçilir (James ve ark., s:318). Yani, her bir bölünmede dikkate alınan tahminci sayısı sınıflama problemlerinde yaklaşık olarak toplam tahminci sayısının karekökü, regresyon problemlerinde ise üçte birine eşittir.

Başka bir deyişle, RF uygulanırken ağaçtaki her bölünmede, algoritmanın mevcut maddelerin çoğunluğunun dikkate almasına izin verilmez (James ve ark. s:320). Bu durum şu şekilde açıklanabilir. Veri kümesinde, bir dizi orta derecede güçlü belirleyiciyle birlikte çok güçlü bir belirleyicinin olduğunu varsayalım. Daha sonra bagging yönteminden elde edilen ağaçların koleksiyonunda, ağaçların çoğu veya tamamı bu güçlü belirleyiciyi üst kısımda kullanacaktır. Sonuç olarak, bagging yönteminde elde edilen ağaçların hepsi birbirine oldukça benzer görünecektir. Bu nedenle, bagging yönteminden elde edilen ağaçlardan gelen tahminler yüksek oranda ilişkili olacaktır. Çok yüksek ilişkili niceliklerin ortalaması, birçok ilişkisiz niceliklerin ortalaması kadar varyansta bir azalmaya yol açmaz (James ve ark. s:320). Özellikle, bagging için bu ayardaki tek bir ağaç üzerinde varyansta önemli bir azalmaya yol açmayacağı anlamına gelir.



Şekil 1.4. Random Forest Algoritması

RF her bir bölünmeyi yalnızca tahmincilerin bir alt kümesini düşünmeye zorlayarak bu sorunun üstesinden gelir. Bu nedenle, bölünmelerin ortalama (p - m) / p'si güçlü tahminciyi bile dikkate almayacak ve bu nedenle diğer tahmincilerin şansı daha fazla olacaktır (James ve ark. s:320). Bu işlemi ağaçların ilişkilendirilmesi olarak düşünebilir, böylece ortaya çıkan ağaçların ortalaması daha az değişken, daha düşük varyans ile elde edilir ve sonuçlar daha güvenilir olur.

Bagging ve RF arasındaki temel fark, tahminci altküme boyutu m'nin seçilmesidir. Örneğin, m = p kullanılarak RF uygulanırsa, bu sadece Bagging anlamına gelir. m = p kullanan RF, Bagging yönteminde hem test hatasını hem de BD hatasını azaltır.

RF oluşturmak için küçük bir m değeri kullanmak, genellikle çok sayıda ilişkili tahminciye sahip olunduğunda yardımcı olacaktır.

1.3.5. Gradient Boosting Machines

Boosting yönteminde zayıf öğrenicilerden güçlü öğreniciler oluşturulmaya çalışılır. Bu işlem iteratif olarak yapılır. Yanıt değişkeninin daha kesin tahminini elde edebilmek için yeni modeller kurulur (Natekin ve Knoll, 2013). Boosting algoritmaları arasındaki fark genellikle zayıf öğrenicilerin eksikliğini tanımlama şekli ile ortaya çıkmaktadır (Muratlar, 2020a).

GBM' de öncelikli olarak ilk yaprak oluşturulur (Muratlar, 2020a). Daha sonra tahmin hataları göz önüne alınarak yeni ağaçlar oluşturulur. Bu durum karar verilen ağaç sayısına ulaşılınca ya da modelden daha fazla gelişme kaydedilemeyinceye kadar devam eder (Natekin ve Knoll, 2013).

Tahmin edilecek değişkenin ortalaması alınır. Bu sayı ilk tahmin girişimi olan ilk yapraktır. Bu değer ile bağımlı değişken karşılaştırılarak ne kadar hatalı tahmin yapıldığı gözlemlenir. İlk ağaç, ilk yaprak sonucunda elde edilen hataları tahmin eden bir model olarak kurulacaktır. İlk tahmin ve ilk ağaçtan çıkan sonuçlar toplanırsa bağımlı değişken %100 doğru tahmin edilebilir (Singh, 2018). Ancak bu durum aşırı öğrenme olarak ifade edilir (Singh, 2018). GBM, bu sorunu aşmak için ağaçlara öğrenme oranı (learning rate) ekler (Muratlar, 2020a) Bu değer 0 ile 1 arasında değişmektedir. Bir sonraki adım olarak doğru tahmine daha çok yaklaşabilmek için bir önceki ağacın hataları baz alınarak yeni ağaç kurulur. Yeni ağaç üretimi, daha önce belirtildiği gibi, belirlenen ağaç sayısına ulaşılınca ya da ağaç eklenmesi anlamlı olmayıncaya kadar devam edecektir (Natekin ve Knoll, 2013).

GBM üç temel unsur içerir (Brownlee, 2019):

- Optimize edilecek bir kayıp fonksiyonu
- Tahmin yapmak için zayıf bir öğrenici
- Kayıp fonksiyonu en aza indirmek için zayıf öğreniciler eklemek için bir katkı modeli

1.3.5.1. Kayıp Fonksiyonu

Kullanılan kayıp fonksiyonu çözülecek olan sorunun türüne bağlıdır. Farklılaştırılabilir olmalıdır, fakat birçok standart kayıp fonksiyonu desteklenir. Örneğin, regresyon hata kareleri kullanabilirken sınıflandırma da logaritmik kayıp kullanabilir.

GBM' nin yararlarından biri kullanılmak istenebilecek her kayıp fonksiyonu için yeni bir boosting algoritmasının türetilmesinin gerekmemesidir. Bunun yerine, herhangi bir farklılaştırılabilir kayıp fonksiyonunun kullanılabileceği yeterince genel bir yöntemdir.

1.3.5.2. Zayıf Öğrenici

Karar ağaçları GBM' de zayıf öğrenici olarak kullanılır. Özellikle ayırmalar için gerçek değerler veren ve çıktıları birlikte eklenebilen regresyon ağaçları kullanılır, daha sonraki model çıktılarının eklenmesine ve tahminlerde kalanların düzeltilmesine izin verilir.

Ağaçlar açgözlü (greedy) bir şekilde inşa edilir, kaybı en aza indirmek için Gini gibi saflık puanlarına göre en iyi bölünme noktalarını seçer.

Zayıf öğrenicileri, maksimum sayıda katman, düğüm, bölme veya yaprak düğümü gibi belirli şekillerde kısıtlamak yaygındır. Bunun amacı öğrenicilerin zayıf kalmasını sağlamaktır, ancak yine de açgözlü bir şekilde inşa edilebilir.

1.3.5.3. Katkı Modeli

Ağaçlar birer birer eklenir ve modeldeki mevcut ağaçlar değiştirilmez. Ağaç eklerken kaybı en aza indirmek için Gradient Descent (GD) yöntemi kullanılır.

Geleneksel olarak, GD bir regresyon denklemindeki katsayılar veya yapay sinir ağındaki ağırlıklar gibi bir dizi parametreyi en aza indirmek için kullanılır. Hata veya kayıp hesaplandıktan sonra, bu hatayı en aza indirmek için ağırlıklar güncellenir.

Parametreler yerine, zayıf öğrenen alt modeller veya daha spesifik olarak karar ağaçları vardır. Kayıp hesapladıktan sonra, GD yöntemini gerçekleştirmek için, modele kaybı azaltan bir ağaç eklenmelidir. Bu ağacın parametreleri değiştirilir ve kalan kayıp azaltılır. Genellikle bu yaklaşıma fonksiyonel GD denir. Daha sonra, yeni ağacın çıktısı, modelin nihai çıktısını düzeltmek veya geliştirmek amacıyla mevcut ağaç dizisinin çıktısına eklenir. Harici bir doğrulama, veri kümesinde kayıp kabul edilebilir bir düzeye ulaştığında veya artık iyileşmediğinde sabit sayıda ağaç eklenir veya eğitim durur.

GBM algoritmasını iyileştirmek için dört yöntem kullanılmaktadır (Brownlee, 2019):

- Ağaç Kısıtlamaları (Tree Constraints)
- Öğrenme Oranı (Shrinkage)
- Rasgele Örnekleme (Random Sampling)
- Cezalandırılmış Öğrenme (Penalized Learning)

1.3.5.4.AdaBoost ve GBM Arasındaki Farklar

AdaBoost, üstel kayıpların olduğu kayıp fonksiyonu açısından GBM' in özel bir durumu olarak kabul edilir.

- Hem AdaBoost hem de GBM, temel zayıf bir öğrenici kullanır ve odağı tekrar tekrar tahmin edilmesi zor, sorunlu gözlemlere kaydırarak zayıf bir öğrenicinin performansını artırmaya çalışır. Sonunda, zayıf bir öğrenici ilavesiyle (veya ağırlıklı ilavesiyle) güçlü bir öğrenici oluşturulur.

- AdaBoost'ta değişim, daha önce yanlış sınıflandırılan gözlemlerin ağırlıklandırılmasıyla yapılır. GBM'de önceki yinelemelerde hesaplanan büyük artıkların zor gözlemlerini tanımlar.
- AdaBoost'ta "eksiklikler" yüksek ağırlık veri noktaları ile tanımlanır. GBM'de "eksiklikler" gradyanlar tarafından tanımlanır.
- AdaBoost'un üstel kaybı, daha kötü uyum sağlayan örnekler için daha fazla ağırlık verir. GBM daha fazla açıklama getirmek için hata bileşenlerini daha fazla inceleyecektir.

1.3.6. Extreme Gradient Boosting

XGBoost, GBM algoritmasının optimize edilmiş yüksek performanslı halidir. Tianqi Chen ve Carlos Guestrin'in 2016 yılında yayınladıkları "XGBoost: A Scalable Tree Boosting System" adlı makale ile kullanılmaya başlanmıştır. Algoritmanın en önemli özellikleri hızlı bir şekilde yüksek tahmin gücü elde edebilmesi, aşırı öğrenmenin önüne geçebilmesi ve eksik verileri yönetebilmesidir. Tianqi'ye göre XGBoost diğer popüler algoritmalardan 10 kat daha hızlı çalışmaktadır. Daha az kaynak kullanarak üstün sonuçlar elde etmek için yazılım ve donanım optimizasyon teknikleri uygulanmıştır. Karar ağacı tabanlı algoritmaların en iyisi olarak gösterilmektedir.

Çalışma mantığı Gradient Boosting ile oldukça benzerdir. XGBoost'ta ilk adım ilk tahmini yapmaktır. Bu tahmin, bundan sonraki adımlarda yapılacak işlemler ile yakınsayarak doğru sonuca ulaşılacağı için herhangi bir sayı olabilir (Muratlar, 2020b). Bu sayı için varsayılan değer 0,5'tir.

Yapılacak bu tahminin ne kadar iyi olduğunu anlamak için bir sonraki aşamada GBM'de olduğu gibi hataları tahminleyen karar ağacı kurulur. Burada amaç hataları öğrenip doğru tahmine yaklaşmaktır.

Oluşturulan ağacın her bir dalı için benzerlik skoru hesaplanır. Benzerlik skorları hesaplandıktan sonraki soru, daha iyi bir tahmin yapılıp yapılamayacağıdır. Bu soruyu cevaplamak için olabilecek tüm olasılıklardaki ağaçlar kurulur. Hepsi için benzerlik skorları hesaplanır. Hangi ağacın daha iyi olduğuna karar vermek için bilgi kazancı hesaplanacaktır. "Benzerlik skoru" ile dallar değerlendirilirken, "kazanç" ile bütün ağaç değerlendirilmektedir (Muratlar, 2020b)

Hangi ağacın daha yüksek kazanç değerine sahip olduğu belirlenip, ağacın kullanılmasına karar verildikten sonra budama işlemi başlayacaktır (Tianqi, 2016). Budama için "gamma" adı verilen değer seçilir. Gamma, bilgi kazancı skoruna getirilen bir değerlendirmedir. Kazanç skoru, gamma skorundan düşük olan dallar budanacaktır. Dolayısıyla gamma değerini artırmak sadece değerli dalların ağaçta kalmasına ve aşırı öğrenmenin önüne geçilmesine yardımcı olur.

Varsayılan gamma değeri 0'dır. Gamma değerinin sıfır olması, bilgi kazancı değerinin eksi değer alması durumunda dalların budanacağı anlamına gelmektedir. Eğer dalların hepsi budanmışsa en iyi tahmin en başta yapılan 0,5'tir.

Lambda değeri modeli hassaslaştırır. Lambda değeri arttıkça hesaplanan benzerlik skoru, dolayısıyla bilgi kazancı değeri düşer. Bunun sonucunda model, dal budamada daha cömert davranır. Sadece yüksek skorlu dallar ağaçta kalabilir, budamadan kurtulabilir.

Lambda değeri arttıkça bağımlı değişkenin değeri olması gerekenden az gelecektir. Yani lambda ne kadar yüksekse doğru tahmine o kadar fazla iterasyon ile ulaşılacaktır.

1.3.6.1. GBM ile XGBoost Arasındaki Farklar

GBM ve XGBoost aynı prensiple çalışmaktadır. XGBoost, farklı teknikler kullanarak daha yüksek tahmin başarısı gösterir ve büyük veri setlerinde çalışmak için optimize edilmiştir.

XGBoost' un GBM' den ayrıştığı başlıca konular şunladır.

- Regülarizasyon
- Budama
- Eksik Veriler ile Çalışılabilmesi
- Sistem Optimizasyonu

2. GEREÇ VE YÖNTEM

Bu tez çalışmasının amacı Rasch modeline dayalı olarak türetilen veri üzerinde BUT uygulamasında yaygın olarak kullanılan algoritma ile BUT uygulamasına yeni bir yaklaşım olarak makine öğrenmesi yöntemlerinin performanslarının incelenip BUT algoritmasına bir alternatif olup olmayacağını incelemektir.

Tez kapsamında Rasch modelinden veri türetme; madde sayısı, bireylerin θ düzeyi, madde zorlukları ve madde yanıt seçeneklerinin kategori sayısına göre farklı senaryolarda yapılmıştır.

Türetilen veri seti için madde sayısı 50, 150 ve 250, kişi sayısı 100, her maddeye ait yanıt kategori sayısı 2, 3 ve 5 olarak belirlenmiştir. Çalışma kapsamında 1000 Monte Carlo (MC) tekrarı yapılmıştır.

Kişilere ait (θ) düzeyinin belirlenmesi:

Kişilere ait parametreleri içeren vektörü, minimum -4 logit, maksimum 4 logit olacak şekilde $(\theta_{maksimum} - \theta_{minimum})/(kişi\ sayısı - 1)$ birimlik artışlarla oluşturulmuştur.

Madde parametrelerinin (β) belirlenmesi:

Madde zorlukları, minimum -4 logit, maksimum 4 logit; madde sayısı da 50,150 ve 250 olacak şekilde $(\beta_{maksimum} - \beta_{minimum})/(kişi\ sayısı - 1)$ birimlik artışlarla oluşturulmuştur. Maddelere ait eşik zorluklarının ortalaması madde zorluğuna eşittir. Eşik zorlukları (β_{jk}) eşik değerlerinin (τ_{jk}) madde zorluğuna (β_i) eklenmesi ile elde edilir. Bu nedenle eşik değerleri toplamlarının sıfır olması gerekir. Her madde için kategori sayısına göre 3 kategorili için Uniform (0,2) ve 5 kategori için Uniform (0,4) dağılımından eşik sayısı kadar rastgele sayı türetilerek, türetilen bu rastgele sayı madde zorluklarına eklenip çıkartılması ile eşik zorlukları elde edilmiştir.

- Madde sayısı: 50, 150, 250
- Bireylerin θ düzeyi ve madde parametreleri için:

➢ θ_1: -4,5;4,5 β_1: -4,5;4,5 (madde zorlukları dağılımı ile bireylerin incelenen özellik dağılımı örtüşüyor.)

➢ θ_2: -4,5;2,5 β_2: -2,5;4,5 (madde zorlukları dağılımı ile bireylerin incelenen özellik dağılımı örtüşmüyor.)

➢ θ_3: -2,5;4,5 β_3: -4,5;2,5 (madde zorlukları ve bireyin incelenen özellikleri örtüşmüyor.)

- Bağımlı değişken kategori sayısı: 2 (İki Sonuçlu Rasch Modeli), 3(Kısmi Kredi Modeli (KKM)), 5 (Kısmi Kredi Modeli (KKM))

olmak üzere 9 senaryo türetilmiştir.

Çizelge 2.1. Senaryolara ait Bilgiler

	Madde Sayısı	Kategori Sayısı	Teta	Beta
S_1	50	2,3,5	-4,5;4,5	-4,5;4,5
S_2	150	2,3,5	-4,5;4,5	-4,5;4,5
S_3	250	2,3,5	-4,5;4,5	-4,5;4,5
S_4	50	2,3,5	-4,5;2,5	-2,5;4,5
S_5	150	2,3,5	-4,5;2,5	-2,5;4,5
S_6	250	2,3,5	-4,5;2,5	-2,5;4,5
S_7	50	2,3,5	-2,5;4,5	-4,5;2,5
S_8	150	2,3,5	-2,5;4,5	-4,5;2,5
S_9	250	2,3,5	-2,5;4,5	-4,5;2,5

Senaryolarda yer alan bağımsız değişkenlerden faydalanarak bireylerin toplam skorları(bağımlı değişken) elde edilmiştir. Bağımlı değişken iki kategorili (0, 1) olduğunda S_1, S_4 ve S_7 için toplam skor en az 0 ve en fazla 50'dir. S_2, S_5 ve S_8 için toplam skor en az 0 ve en fazla 150'dir. S_3, S_6 ve S_9 için toplam skor en az 0 ve en fazla 250'dir. Bağımlı değişken üç kategorili (0, 1, 2) olduğunda S_1, S_4 ve S_7 için toplam skor en az 0 ve en fazla 100'dür. S_2, S_5 ve S_8 için toplam skor en az 0 ve en fazla 300'dür. S3, S_6 ve S_9 için toplam skor en az 0 ve en fazla 500' dür. Bağımlı değişken beş kategori (0, 1, 2, 3, 4) olduğunda S_1, S_4 ve S_7 için toplam skor en az 0 ve en fazla 200'dür. S_2, S_5 ve S_8 için toplam skor en az 0 ve en fazla 600' dür. S_3, S_6 ve S_9 için toplam skor en az 0 ve en fazla 1000'dir. Her bir senaryo için bu toplam skorlar

BUT, CART, RF, GBM ve XGBoost yöntemleri ile tahmin edilmiştir. Bu tahminlerden elde edilen ortalama madde sayısı, hata kareler ortalamasının karekökü (root mean squared error, RMSE), Sınıf içi Korelasyon Katsayısı (SKK) ve ortalama tahmin değeri (toplam skor) üzerinden yöntemlerin performansları karşılaştırılmıştır.

Bu tez çalışması kapsamında BUT algoritması için incelenen özellik kestirimi sonsal dağılımın beklenen değeri (Expected a Posteriori, EAP) ile soru seçimi ise sonsal dağılım ağırlıklı bilgi kriteri (Maximum Posterior Weighted Information, MPWI) yöntemi ile gerçekleştirilmiştir. Algoritmada durdurma kriteri olarak θ düzeyi kestiriminin standart hatası 0.33 olarak belirlenmiştir. BUT algoritması SmartCat programı (Ankara Üniversitesi Tıp Fak., Biyoistatistik AD, Psikometrik Lab.) ve R v.3.6.3 programında gerçekleştirilmiştir.

BUT algoritması ile bulunan Rasch skorlarının toplam skora dönüştürülmesi için 1.5'deki formül kullanılmıştır.

y = m + (s * Bireyin Yetenek Düzeyi) 1.5.

s = (İstenen Aralık) / (Kullanılan Aralık) 1.6.

m = (İstenen minimum değer) – (Kullanılan minimum değer * s) 1.7.

Makine öğrenmesi yöntemlerinin analizi R v3.6.3 programında rpart, rpart.plot, caret, randomForest, gbm ve xgboost kütüphaneleri ile gerçekleştirilmiştir. Ayrıca tahminler ile toplam skor arasındaki sınıf içi korelasyon katsayısı (SKK) için irr kütüphanesi kullanılmıştır.

Türetilen veri seti %80 eğitim verisi ve %20 test verisi olarak ayrılmıştır. Eğitim setinden elde edilen modeller üzerinde test verisinin sonuçları elde edilmiştir. BUT yöntemi ile makine öğrenmesi yöntemlerinin karşılaştırılmasında bütünlük olması bakımından BUT yöntemi de test verisi için uygulanmıştır.

Makine öğrenmesi yöntemlerinde seçilen parametreler farklı değerler aralığında denemiş olup en optimum sonuçları veren değerler model oluşturmak için kullanılmıştır.

CART yönteminde minsplit: 10, max.depth: 20 ve cp: 0,01 olarak alınmıştır.

Minsplit: Her bir bölünmedeki minimum gözlem sayısı

Max. Depth: Son düğümdeki maksimum derinlik

cp: Karmaşıklık parametresi

GBM yönteminde n.trees:100, interaction. depth: 3 ve shrinkage: 0,1 olarak alınmıştır.

n.trees: Ağaç sayısı, İterasyon Sayısı

Interaction. Depth: Her bir ağaçtaki maksimum düğüm sayısı

Shrinkage: Öğrenme Oranı

XGBoost yönteminde max.depth: 3, eta: 0,1 nthread:2 ve nrounds:100 olarak alınmıştır.

Max.Depth: Maksimum Ağaç Derinliği

Eta: Öğrenme Oranı

Nthread: Paralel olarak çalıştırılacak işlemci sayısı

Nrounds: Ağaç sayısı, İterasyon Sayısı

3.BULGULAR

Senaryolar için yöntemlere ait sonuçlar çizelge ve şekillerle verilmiştir. Her bir senaryo için kategori sayısına göre yöntem sonuçları çizelge olarak, ortalama madde sayısına göre, RMSE değerine göre ve SKK değerine göre yöntemlerin karşılaştırılması şekil üzerinde yapılmıştır.

3.1. Kategori Sayısına Göre Senaryolara ait Yöntem Sonuçları

Bu bölümde kategori sayılarına göre senaryolara ait yöntemlerden elde edilen RMSE değeri, ortalama madde sayısı, tahmin değerleri ve SKK değerleri çizelge olarak verilmiştir ve karşılaştırmalar bu değerlere göre yapılmıştır.

Çizelge 3.1. Kategori Sayılarına Göre S_1 Sonuçları

	2 Kategori					3 Kategori					5 Kategori				
S_1	RMSE	Madde	Tahmin	Gerçek	SKK	RMSE	Madde	Tahmin	Gerçek	SKK	RMSE	Madde	Tahmin	Gerçek	SKK
BUT	1,50	49,99	23,81	25,03	0,98	4,84	19,19	50,37	50,03	0,99	9,58	8,89	100,37	100,06	0,99
CART	4,47	7,24	24,99	25,03	0,74	10,47	7,70	50,04	50,03	0,95	16,80	7,61	100,11	100,06	0,97
RF	2,50	16	25,01	25,03	0,92	3,38	16	49,87	50,03	0,99	4,98	16	99,57	100,06	0,99
GBM	1,96	18	2501	25,03	0,96	3,62	42	50,03	50,03	0,99	6,67	43	100,16	100,06	0,99
XGBoost	2,52	48	24,54	25,03	0,91	1,97	50	49,96	50,03	0,99	2,14	50	100,19	100,06	0,99

RMSE: Hata Kareler Ortalamasının Karekökü, Gerçek: Ortalama Toplam Skor, Tahmin: Tahmin Edilen Ortalama Skor, SKK: Sınıf İçi Korelasyon Katsayısı

S_1'in iki kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin Değerleri, Gerçek Değerler ve SKK değerlerine bakılmıştır. S_1'de toplam 50 madde yer almaktadır. En düşük ortalama madde sayısı 7,24 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 49,99 ile BUT yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 1,50 ile BUT yönteminde elde edilmiştir. En yüksek RMSE değeri ise 4,47 ile CART yönteminden elde edilmiştir. SKK değerlerine bakıldığında en yüksek SKK değeri 0,98 ile BUT yönteminden elde edilirken en düşük SKK değeri 0,74 ile CART yönteminde elde edilmiştir.

S_1'in üç kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin Değerleri, Gerçek Değerler ve SKK değerlerine bakılmıştır. S_1'de toplam 50 madde yer almaktadır. En düşük ortalama madde sayısı 7,70 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 50 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 1,97 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 10,47 ile CART yönteminden elde edilmiştir. SKK değerlerine bakıldığında ise bütün yöntemlere ait SKK değerleri 0,95 ve üzerindedir.

S_1'in beş kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin Değerleri, Gerçek Değerler ve SKK değerlerine bakılmıştır. S_1'de toplam 50 madde yer almaktadır. En düşük ortalama madde sayısı 7,61 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 50 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 2,14 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 16,80 ile CART yönteminden elde edilmiştir. SKK değerlerine bakıldığında ise bütün yöntemlere ait SKK değerleri 0,97 ve üzerindedir.

Çizelge 3.2. Kategori Sayılarına Göre S_2 Sonuçları

	2 Kategori					3 Kategori					5 Kategori				
S_2	RMSE	Madde	Tahmin	Gerçek	SKK	RMSE	Madde	Tahmin	Gerçek	SKK	RMSE	Madde	Tahmin	Gerçek	SKK
BUT	5,372	38,24	74,51	74,94	0,97	16,16	16,96	150,95	150,33	0,99	29,97	8,85	301,19	299,42	0,99
CART	13,40	7,72	74,98	74,94	0,70	32,39	7,86	150,42	150,33	0,95	51,61	7,64	298,65	299,42	0,97
RF	7,53	50	74,93	74,94	0,90	10,27	50	150,04	150,33	0,99	14,46	50	298,21	299,42	0,99
GBM	5,22	56	74,90	74,94	0,96	11,35	93	150,44	150,33	0,99	20,36	101	300,27	299,42	0,99
XGBoost	3,96	150	74,16	74,94	0,97	3,18	150	150,44	150,33	0,99	7,34	150	300,18	299,42	0,99

RMSE: Hata Kareler Ortalamasının Karekökü, Gerçek: Ortalama Toplam Skor, Tahmin: Tahmin Edilen Ortalama Skor, SKK: Sınıf İçi Korelasyon Katsayısı

S_2'nin iki kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_2'de toplam 150 madde yer almaktadır. En düşük ortalama madde sayısı 7,72 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 150 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 3,96 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 13,40 ile CART yönteminden elde edilmiştir. SKK değerlerine bakıldığında en yüksek SKK değeri 0,97 ile XGBoost ve BUT yöntemlerinden elde edilirken en düşük SKK değeri 0,70 ile CART yönteminde elde edilmiştir.

S_2'nin üç kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_2'de toplam 150 madde yer almaktadır. En düşük ortalama madde sayısı 7,86 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 150 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 3,18 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 32,39 ile CART yönteminden elde edilmiştir. SKK değerlerine bakıldığında ise bütün yöntemlere ait SKK değerleri 0,95 ve üzerindedir.

S_2'nin beş kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_2'de toplam 150 madde yer almaktadır. En düşük ortalama madde sayısı 7,64 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 150 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 7,34 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 51,61 ile BUT yönteminden elde edilmiştir. SKK değerlerine bakıldığında ise bütün yöntemlere ait SKK değerleri 0,97 ve üzerindedir.

Çizelge 3.3. Kategori Sayılarına Göre S_3 Sonuçları

	2 Kategori					3 Kategori					5 Kategori				
S_3	RMSE	Madde	Tahmin	Gerçek	SKK	RMSE	Madde	Tahmin	Gerçek	SKK	RMSE	Madde	Tahmin	Gerçek	SKK
BUT	10,06	36,61	124,50	124,98	0,96	27,59	16,92	251,40	250,31	0,99	50,28	8,85	503,03	499,59	0,99
CART	22,48	7,83	124,83	124,98	0,70	54,19	7,84	249,87	250,31	0,95	85,70	7,64	499,82	499,59	0,97
RF	12,63	83	124,89	124,98	0,90	16,93	83	249,77	250,31	0,99	23,58	83	497,85	499,59	0,99
GBM	8,78	74	124,83	124,98	0,96	19,33	126	250,33	250,31	0,99	34,05	122	501,56	499,59	0,99
XGBoost	5,01	248	123,98	124,98	0,98	5,40	250	250,98	250,313	0,99	14,70	250	500,85	499,59	0,99

RMSE: Hata Kareler Ortalamasının Karekökü, Gerçek: Ortalama Toplam Skor, Tahmin: Tahmin Edilen Ortalama Skor, SKK: Sınıf İçi Korelasyon Katsayısı

S_3'ün iki kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_3'de toplam 250 madde yer almaktadır. En düşük ortalama madde sayısı 7,83 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 248 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 5,01 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 22,48 ile CART yönteminden elde edilmiştir. SKK değerlerine bakıldığında en yüksek SKK değeri 0,98 ile XGBoost ve BUT yöntemlerinden elde edilirken en düşük SKK değeri 0,70 ile CART yönteminde elde edilmiştir.

S_3'ün üç kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_3'de toplam 250 madde yer almaktadır. En düşük ortalama madde sayısı 7,84 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 250 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 5,40 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 54,19 ile CART yönteminden elde edilmiştir. SKK değerlerine bakıldığında ise bütün yöntemlere ait SKK değerleri 0,95 ve üzerindedir.

S_3'ün beş kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_3'de toplam 250 madde yer almaktadır. En düşük ortalama madde sayısı 7,64 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 250 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 14,70 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 85,70 ile CART yönteminden elde edilmiştir. SKK değerlerine bakıldığında ise bütün yöntemlere ait SKK değerleri 0,97 ve üzerindedir.

Çizelge 3.4. Kategori Sayılarına Göre S_4 Sonuçları

	2 Kategori					3 Kategori					5 Kategori				
S_4	RMSE	Madde	Tahmin	Gerçek	SKK	RMSE	Madde	Tahmin	Gerçek	SKK	RMSE	Madde	Tahmin	Gerçek	SKK
BUT	4,70	49,99	23,39	18,70	0,86	13,91	17,00	27,13	27,94	0,91	28,47	9,15	80,83	55,48	0,91
CART	5,27	7,41	18,69	18,70	0,72	10,33	7,29	27,57	27,94	0,93	16,43	7,46	55,34	55,48	0,95
RF	2,88	16	18,58	18,70	0,92	3,85	16	27,32	27,94	0,99	5,64	16	54,50	55,48	0,99
GBM	2,08	29	18,73	18,70	0,96	3,49	30	27,92	27,94	0,99	6,20	39	55,47	55,48	0,99
XGBoost	2,98	50	18,43	18,7	0,91	2,21	49	27,90	27,94	0,99	1,96	50	55,55	55,48	0,99

RMSE: Hata Kareler Ortalamasının Karekökü, Gerçek: Ortalama Toplam Skor, Tahmin: Tahmin Edilen Ortalama Skor, SKK: Sınıf İçi Korelasyon Katsayısı

S_4'ün iki kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin Değerleri, Gerçek Değerler ve SKK değerlerine bakılmıştır. S_4'de toplam 50 madde yer almaktadır. En düşük ortalama madde sayısı 7,41 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 50 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 2,08 ile GBM yönteminde elde edilmiştir. En yüksek RMSE değeri ise 5,27 ile CART yönteminden elde edilmiştir. SKK değerlerine bakıldığında en yüksek SKK değeri 0,96 ile GBM yönteminden elde edilirken en düşük SKK değeri 0,72 ile CART yönteminde elde edilmiştir.

S_4'ün üç kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin Değerleri, Gerçek Değerler ve SKK değerlerine bakılmıştır. S_4'de toplam 50 madde yer almaktadır. En düşük ortalama madde sayısı 7,29 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 49 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 2,21 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 13,91 ile BUT yönteminden elde edilmiştir. SKK değerlerine bakıldığında ise bütün yöntemlere ait SKK değerleri 0,91 ve üzerindedir.

S_4'ün beş kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin Değerleri, Gerçek Değerler ve SKK değerlerine bakılmıştır. S_4'de toplam 50 madde yer almaktadır. En düşük ortalama madde sayısı 7,46 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 50 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 1,96 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 28,47 ile BUT yönteminden elde edilmiştir. SKK değerlerine bakıldığında ise bütün yöntemlere ait SKK değerleri 0,91 ve üzerindedir.

Çizelge 3.5. Kategori Sayılarına Göre S_5 Sonuçları

	2 Kategori					3 Kategori					5 Kategori				
S_5	RMSE	Madde	Tahmin	Gerçek	SKK	RMSE	Madde	Tahmin	Gerçek	SKK	RMSE	Madde	Tahmin	Gerçek	SKK
BUT	20,87	37,10	76,13	55,87	0,68	44,01	16,91	122,62	83,90	0,90	74,63	8,21	231,19	165,61	0,92
CART	15,81	7,76	56,00	55,87	0,71	31,25	7,73	83,45	8,390	0,93	50,29	7,78	164,75	165,61	0,95
RF	8,65	50	74,93	74,94	0,90	11,62	50	82,22	83,90	0,99	16,65	50	162,99	165,61	0,99
GBM	5,99	62	55,92	55,87	0,96	11,19	81	83,80	83,90	0,99	19,41	87	165,72	165,61	0,99
XGBoost	4,79	149	55,30	55,87	0,97	3,51	148	83,71	83,90	0,99	4,61	150	165,85	165,61	0,99

RMSE: Hata Kareler Ortalamasının Karekökü, Gerçek: Ortalama Toplam Skor, Tahmin: Tahmin Edilen Ortalama Skor, SKK: Sınıf İçi Korelasyon Katsayısı

S_5'in iki kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_5'de toplam 150 madde yer almaktadır. En düşük ortalama madde sayısı 7,76 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 149 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 4,79 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 20,87 ile BUT yönteminden elde edilmiştir. SKK değerlerine bakıldığında en yüksek SKK değeri 0,97 ile XGBoost ve BUT yöntemlerinden elde edilirken en düşük SKK değeri 0,68 ile BUT yönteminde elde edilmiştir.

S_5'in üç kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_5'de toplam 150 madde yer almaktadır. En düşük ortalama madde sayısı 7,73 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 148 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 3,51 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 44,01 ile BUT yönteminden elde edilmiştir. SKK değerlerine bakıldığında ise bütün yöntemlere ait SKK değerleri 0,90 ve üzerindedir.

S_5'in beş kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_5'de toplam 150 madde yer almaktadır. En düşük ortalama madde sayısı 7,78 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 150 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 4,61 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 74,63 ile BUT yönteminden elde edilmiştir. SKK değerlerine bakıldığında ise bütün yöntemlere ait SKK değerleri 0,92 ve üzerindedir.

Çizelge 3.6. Kategori Sayılarına Göre S_6 Sonuçları

	2 Kategori					3 Kategori					5 Kategori				
S_6	RMSE	Madde	Tahmin	Gerçek	SKK	RMSE	Madde	Tahmin	Gerçek	SKK	RMSE	Madde	Tahmin	Gerçek	SKK
BUT	33,89	36,15	125,50	92,93	0,68	72,55	16,53	202,91	139,07	0,90	146,09	8,14	404,87	275,84	0,90
CART	26,83	7,80	92,70	92,93	0,68	53,54	7,77	137,59	139,07	0,92	84,32	7,87	273,97	275,84	0,95
RF	14,33	83	92,33	92,93	0,91	19,30	83	136,48	139,07	0,99	27,77	83	271,61	275,84	0,99
GBM	10,04	89	93,00	92,93	0,96	18,65	120	139,08	139,07	0,99	31,88	122	276,00	275,84	0,99
XGBoost	5,96	245	92,19	92,93	0,98	4,51	249	138,78	139,07	0,99	9,66	249	276,62	275,84	0,99

RMSE: Hata Kareler Ortalamasının Karekökü, Gerçek: Ortalama Toplam Skor, Tahmin: Tahmin Edilen Ortalama Skor, SKK: Sınıf İçi Korelasyon Katsayısı

S_6'nın iki kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_6'da toplam 250 madde yer almaktadır. En düşük ortalama madde sayısı 7,80 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 245 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 5,96 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 33,89 ile BUT yönteminden elde edilmiştir. SKK değerlerine bakıldığında en yüksek SKK değeri 0,98 ile XGBoost yöntemlerinden elde edilirken en düşük SKK değeri 0,68 ile CART ve BUT yönteminde elde edilmiştir.

S_6'nın üç kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_6'da toplam 250 madde yer almaktadır. En düşük ortalama madde sayısı 7,87 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 249 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 4,51 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 72,55 ile BUT yönteminden elde edilmiştir. SKK değerlerine bakıldığında ise bütün yöntemlere ait SKK değerleri 0,90 ve üzerindedir.

S_6'nın beş kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_6'da toplam 250 madde yer almaktadır. En düşük ortalama madde sayısı 7,64 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 250 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 14,70 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 146,09 ile CART yönteminden elde edilmiştir. SKK değerlerine bakıldığında ise bütün yöntemlere ait SKK değerleri 0,90 ve üzerindedir.

Çizelge 3.7. Kategori Sayılarına Göre S_7 Sonuçları

	2 Kategori					3 Kategori					5 Kategori				
S_7	RMSE	Madde	Tahmin	Gerçek	SKK	RMSE	Madde	Tahmin	Gerçek	SKK	RMSE	Madde	Tahmin	Gerçek	SKK
BUT	4,42	49,99	26,63	31,34	0,86	11,41	27,59	61,67	72,06	0,94	28,77	9,90	118,91	144,50	0,91
CART	5,27	7,41	31,32	31,34	0,72	10,35	7,22	72,22	72,06	0,93	16,36	7,43	144,89	144,50	0,95
RF	2,864	16	31,46	31,34	0,92	3,83	16	72,43	72,06	0,99	5,55	16	144,77	144,50	0,99
GBM	2,06	28	31,32	31,34	0,96	3,57	32	72,13	72,06	0,99	6,23	38	144,64	144,50	0,99
XGBoost	2,49	50	30,80	31,34	0,94	1,28	50	71,79	72,06	0,99	1,43	50	144,39	144,50	0,99

RMSE: Hata Kareler Ortalamasının Karekökü, Gerçek: Ortalama Toplam Skor, Tahmin: Tahmin Edilen Ortalama Skor, SKK: Sınıf İçi Korelasyon Katsayısı

S_7'nin iki kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_7'de toplam 50 madde yer almaktadır. En düşük ortalama madde sayısı 7,41 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 50 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 2,06 ile GBM yönteminde elde edilmiştir. En yüksek RMSE değeri ise 5,27 ile CART yönteminden elde edilmiştir. SKK değerlerine bakıldığında en yüksek SKK değeri 0,96 ile GBM yönteminden elde edilirken en düşük SKK değeri 0,72 ile CART yönteminde elde edilmiştir.

S_7'nin üç kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_7'de toplam 50 madde yer almaktadır. En düşük ortalama madde sayısı 7,22 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 50 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 1,28 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 11,41 ile BUT yönteminden elde edilmiştir. SKK değerlerine bakıldığında ise bütün yöntemlere ait SKK değerleri 0,93 ve üzerindedir.

S_7'nin beş kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_7'de toplam 50 madde yer almaktadır. En düşük ortalama madde sayısı 7,43 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 50 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 1,43 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 28,77 ile BUT yönteminden elde edilmiştir. SKK değerlerine bakıldığında ise bütün yöntemlere ait SKK değerleri 0,91 ve üzerindedir.

Çizelge 3.8. Kategori Sayılarına Göre S_8 Sonuçları

	2 Kategori					3 Kategori					5 Kategori				
S_8	RMSE	Madde	Tahmin	Gerçek	SKK	RMSE	Madde	Tahmin	Gerçek	SKK	RMSE	Madde	Tahmin	Gerçek	SKK
BUT	20,11	37,42	74,75	94,20	0,70	42,62	17,86	178,75	216,42	0,91	87,71	9,23	356,98	434,29	0,90
CART	15,89	7,77	94,16	94,20	0,69	32,13	7,64	217,21	216,42	0,92	50,56	7,81	435,18	434,29	0,95
RF	8,58	50	94,59	94,2	0,91	11,54	50	217,67	216,42	0,99	16,47	50	435,25	434,29	0,99
GBM	5,94	74	94,18	94,20	0,96	11,16	83	216,52	216,42	0,99	18,91	82	434,95	434,29	0,99
XGBoost	3,64	150	93,41	94,20	0,98	2,15	150	216,42	216,42	0,99	7,78	150	434,42	434,29	0,99

RMSE: Hata Kareler Ortalamasının Karekökü, Gerçek: Ortalama Toplam Skor, Tahmin: Tahmin Edilen Ortalama Skor, SKK: Sınıf İçi Korelasyon Katsayısı

S_8'in iki kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_8'de toplam 150 madde yer almaktadır. En düşük ortalama madde sayısı 7,76 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 149 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 3,64 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 20,11 ile BUT yönteminden elde edilmiştir. SKK değerlerine bakıldığında en yüksek SKK değeri 0,98 ile XGBoost ve BUT yöntemlerinden elde edilirken en düşük SKK değeri 0,69 ile BUT yönteminde elde edilmiştir.

S_8'in üç kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_8'de toplam 150 madde yer almaktadır. En düşük ortalama madde sayısı 7,64 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 150 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 2,15 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 42,62 ile BUT yönteminden elde edilmiştir. SKK değerlerine bakıldığında ise bütün yöntemlere ait SKK değerleri 0,91 ve üzerindedir.

S_8'in beş kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_8'de toplam 150 madde yer almaktadır. En düşük ortalama madde sayısı 7,78 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 150 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 4,61 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 87,71 ile BUT yönteminden elde edilmiştir. SKK değerlerine bakıldığında ise bütün yöntemlere ait SKK değerleri 0,90 ve üzerindedir.

Çizelge 3.9. Kategori Sayılarına Göre S_9 Sonuçları

	2 Kategori					3 Kategori					5 Kategori				
S_9	RMSE	Madde	Tahmin	Gerçek	SKK	RMSE	Madde	Tahmin	Gerçek	SKK	RMSE	Madde	Tahmin	Gerçek	SKK
BUT	30,56	36,55	127,82	156,95	0,74	66,87	17,56	302,27	360,79	0,91	151,49	9,17	590,24	723,96	0,90
CART	26,63	7,82	157,13	156,95	0,69	53,08	7,78	362,00	360,79	0,92	83,75	7,87	724,87	723,96	0,95
RF	14,44	83	157,62	156,95	0,91	19,23	83	362,84	360,79	0,99	27,06	83	725,43	723,96	0,99
GBM	10,12	80	156,93	156,95	0,96	18,80	107	361,26	360,79	0,99	31,19	104	724,58	723,96	0,99
XGBoost	4,44	250	156,05	156,95	0,99	5,09	250	361,20	360,79	0,99	15,56	250	724,28	723,96	0,99

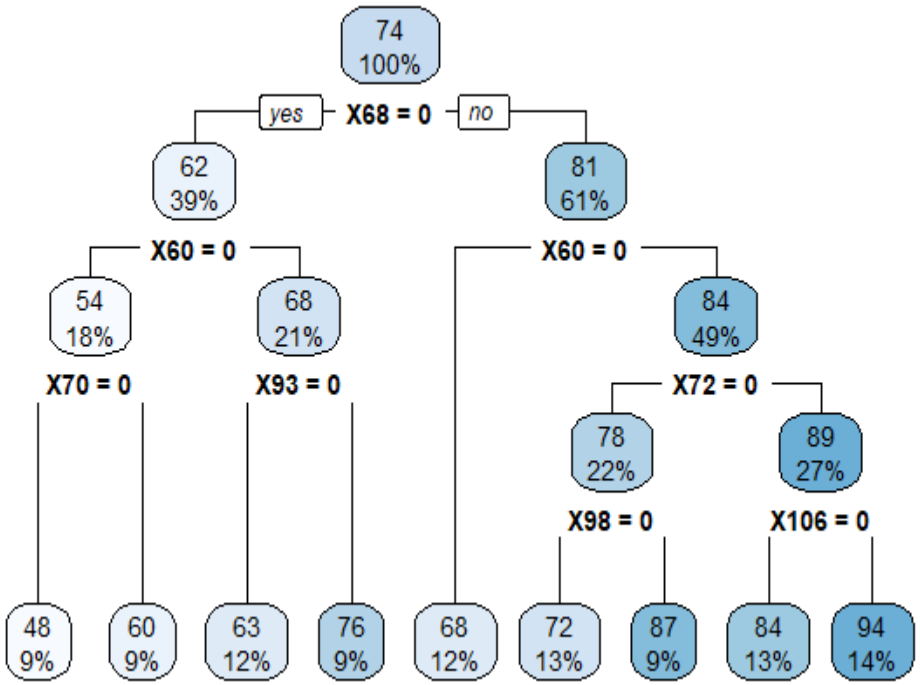
RMSE: Hata Kareler Ortalamasının Karekökü, Gerçek: Ortalama Toplam Skor, Tahmin: Tahmin Edilen Ortalama Skor, SKK: Sınıf İçi Korelasyon Katsayısı

S_9'un iki kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_9'da toplam 250 madde yer almaktadır. En düşük ortalama madde sayısı 7,82 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 250 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 4,44 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 30,56 ile BUT yönteminden elde edilmiştir. SKK değerlerine bakıldığında en yüksek SKK değeri 0,99 ile XGBoost yöntemlerinden elde edilirken en düşük SKK değeri 0,69 ile CART yönteminde elde edilmiştir.

S_9'un üç kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_9'da toplam 250 madde yer almaktadır. En düşük ortalama madde sayısı 7,78 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 250 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 5,09 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 66,87 ile BUT yönteminden elde edilmiştir. SKK değerlerine bakıldığında ise bütün yöntemlere ait SKK değerleri 0,91 ve üzerindedir.

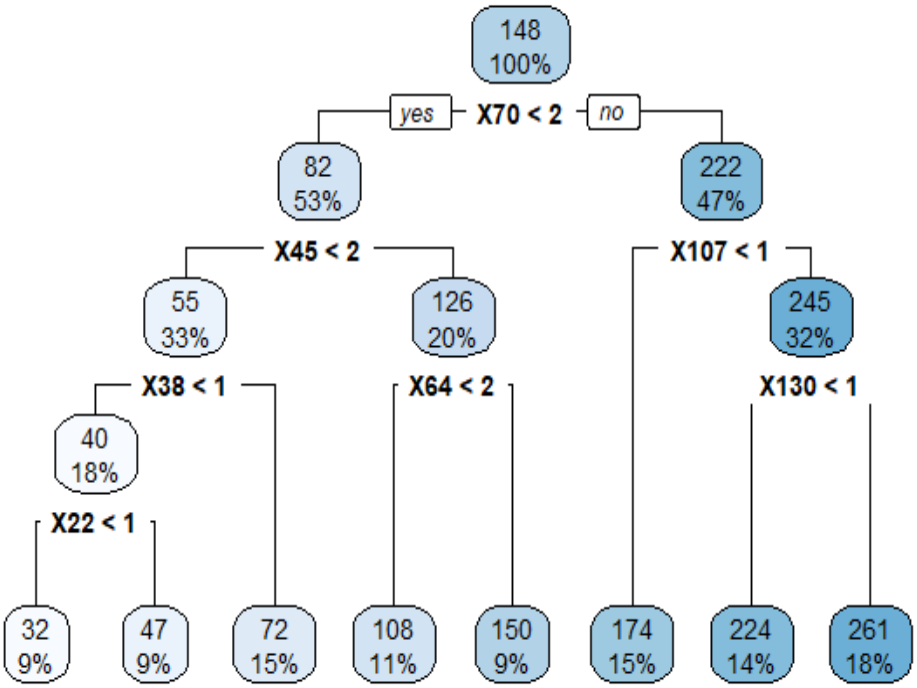
S_9'un beş kategorili verilerinde RMSE, Ortalama Madde Sayısı, Tahmin değerleri, Gerçek değerler ve SKK değerlerine bakılmıştır. S_9'da toplam 250 madde yer almaktadır. En düşük ortalama madde sayısı 7,87 ile CART yönteminden elde edilmiştir. En yüksek ortalama madde sayısı 250 ile XGBoost yönteminden elde edilmiştir. RMSE değerlerine bakıldığında en düşük RMSE değeri 15,56 ile XGBoost yönteminde elde edilmiştir. En yüksek RMSE değeri ise 151,49 ile CART yönteminden elde edilmiştir. SKK değerlerine bakıldığında ise bütün yöntemlere ait SKK değerleri 0,90 ve üzerindedir.

3.2. Karar Ağacı Sonuçları



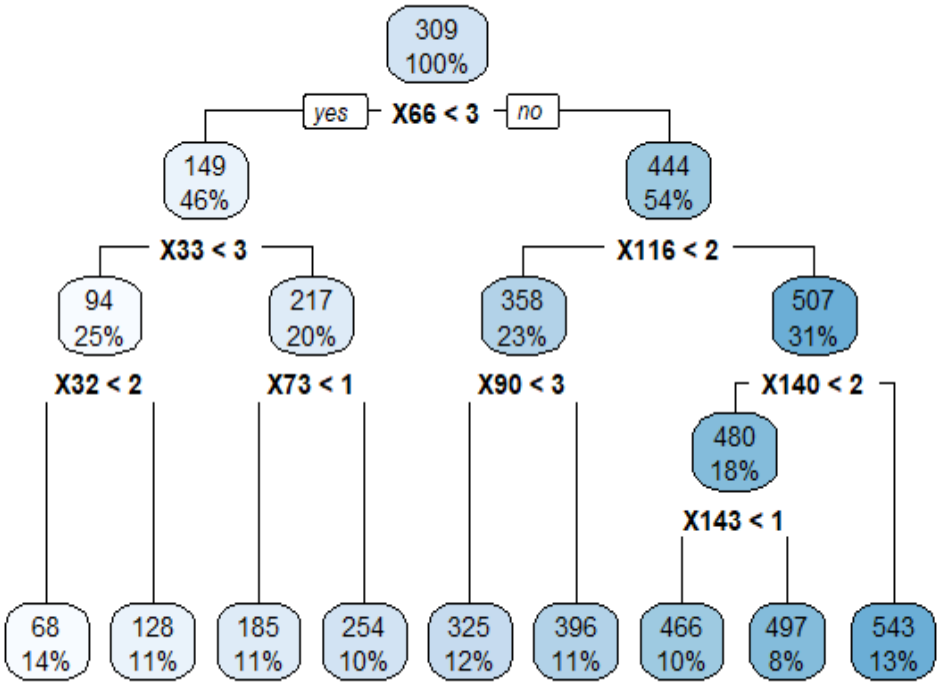
Şekil 3.1. İki Kategorili S_3 için Karar Ağacı

Yanıt kategori sayısı iki olan maddeleri içeren S_3 için karar ağacına bakıldığında ilk ayrımı gerçekleştiren madde 68'dir. 68, 60 ve 70. maddeye sıfır yanıtını veren kişilerin toplam skoru 48'dir. 68, 60, 72 ve 106. maddeye bir yanıtını veren kişilerin toplam skoru 94'dür.



Şekil 3.2. Üç Kategorili S_3 için Karar Ağacı

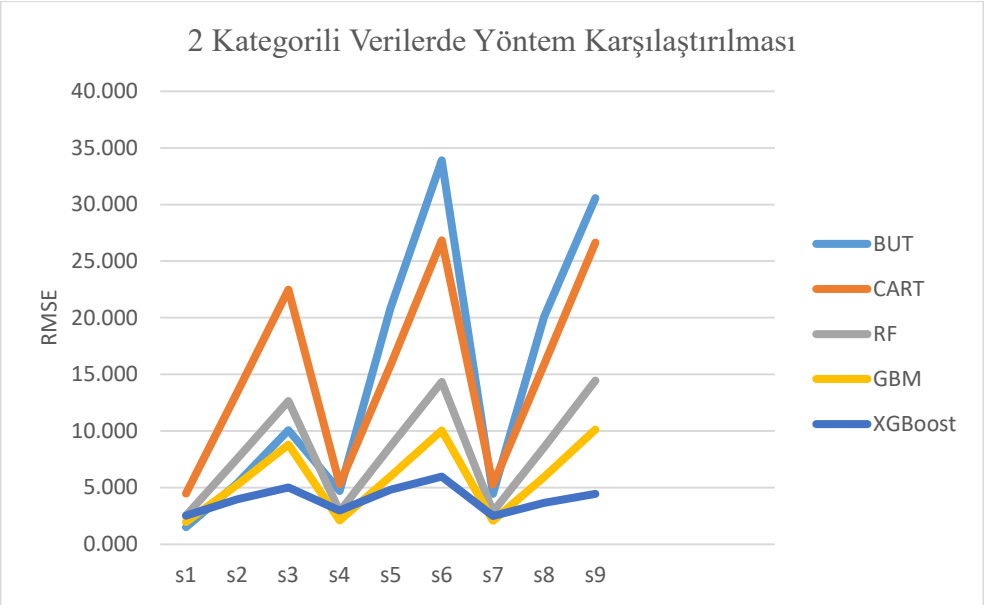
Yanıt kategori sayısı üç olan maddeleri içeren S_3 için karar ağacına bakıldığında ilk ayrımı gerçekleştiren madde 70' dir. 70 ve 60. maddelere ikiden az (0, 1), 38. ve 22. maddelere sıfır yanıtını veren kişilerin toplam skoru 32'dir. 70. maddeye iki, 107. Maddeye bir ya da 2, 130. Maddeye bir ya da iki yanıtını verenlerin toplam skoru 261'dir.



Şekil 3.3. Beş Kategorili S_3 için Karar Ağacı

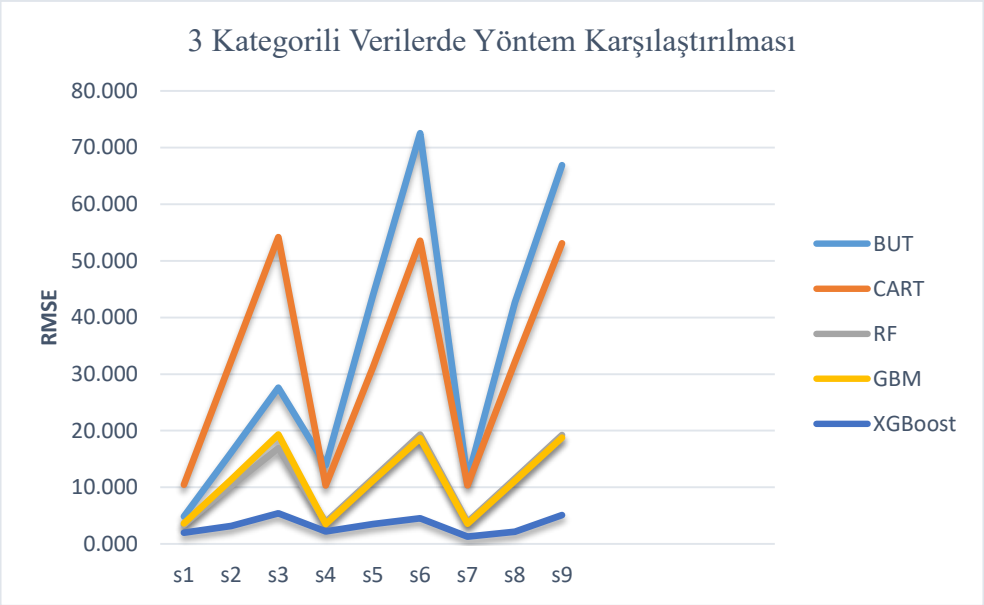
Yanıt kategori sayısı 5 olan maddeleri içeren S_3 için karar ağacına bakıldığında ilk ayrımı gerçekleştiren madde 66'dır. 66. ve 33. maddelere üçten az (0, 1, 2), 32. maddeye ise ikiden az (0, 1) yanıtını veren kişilerin toplam skoru 68'dir. 66. maddeye üç ve daha fazla (3, 4), 116. ve 140. maddelere iki ve daha fazla (2, 3, 4) yanıtını verenlerin toplam skoru 543'dür.

3.3. Senaryolara ait RMSE Değerlerine Göre Yöntem Karşılaştırması



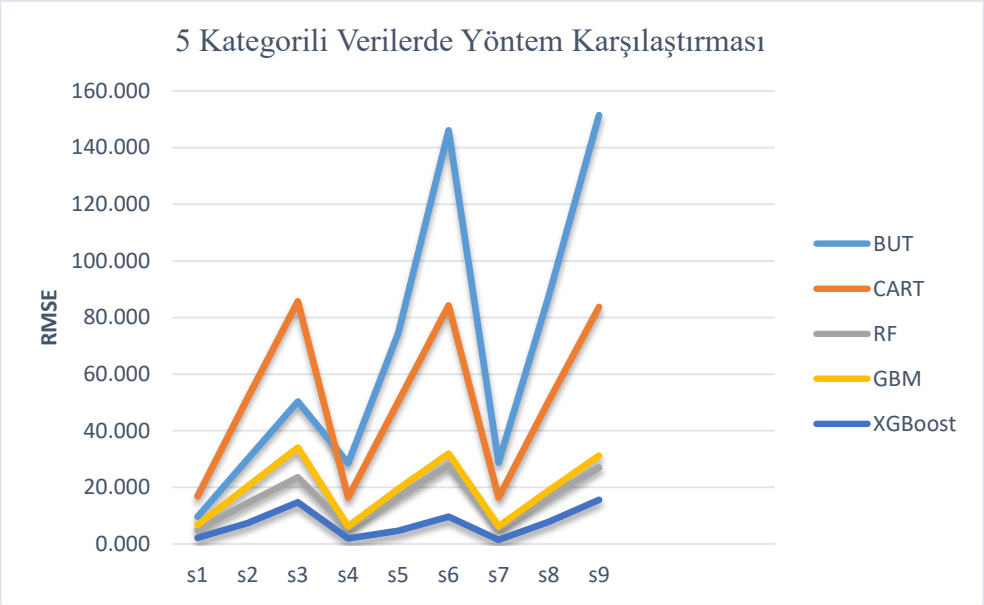
Şekil 3.4. İki Kategorili Verilerde RMSE Değerlerine Göre Yöntem Karşılaştırması

Şekil 3.4'e göre iki kategorili verilerde makine öğrenmesi yöntemlerinde toplam madde sayıları arttıkça RMSE değerlerinin arttığı görülmektedir. Yöntemler karşılaştırıldığında S_1' de en düşük RMSE değeri BUT yönteminden, S_2 ve S_3' de GBM yönteminden, diğer yöntemlerde ise XGBoost yönteminden en düşük RMSE değeri elde edilmiştir.



Şekil 3.5. Üç Kategorili Verilerde RMSE Değerlerine Göre Yöntem Karşılaştırması

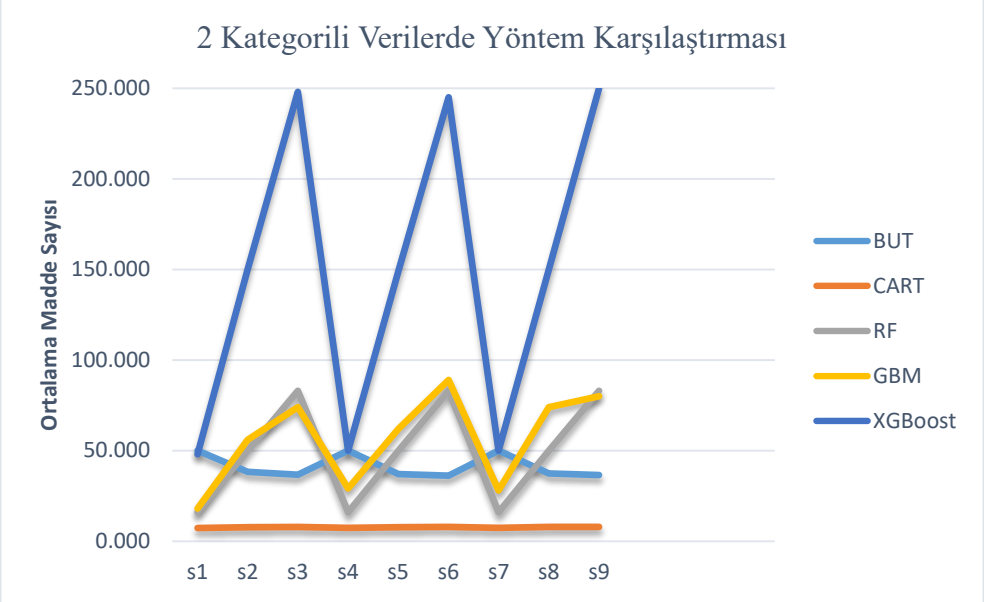
Şekil 3.5'e göre üç kategorili verilerde makine öğrenmesi yöntemlerinde toplam madde sayıları arttıkça RMSE değerlerinin arttığı görülmektedir. Yöntemler karşılaştırıldığında bütün senaryolarda en düşük RMSE değeri XGBoost yönteminden elde edilmiştir. Bütün senaryolarda RF ve GBM yöntemlerinin birbirine oldukça yakın RMSE değerleri verdiği görülmektedir.



Şekil 3.6. Beş Kategorili Verilerde RMSE Değerlerine Göre Yöntem Karşılaştırması

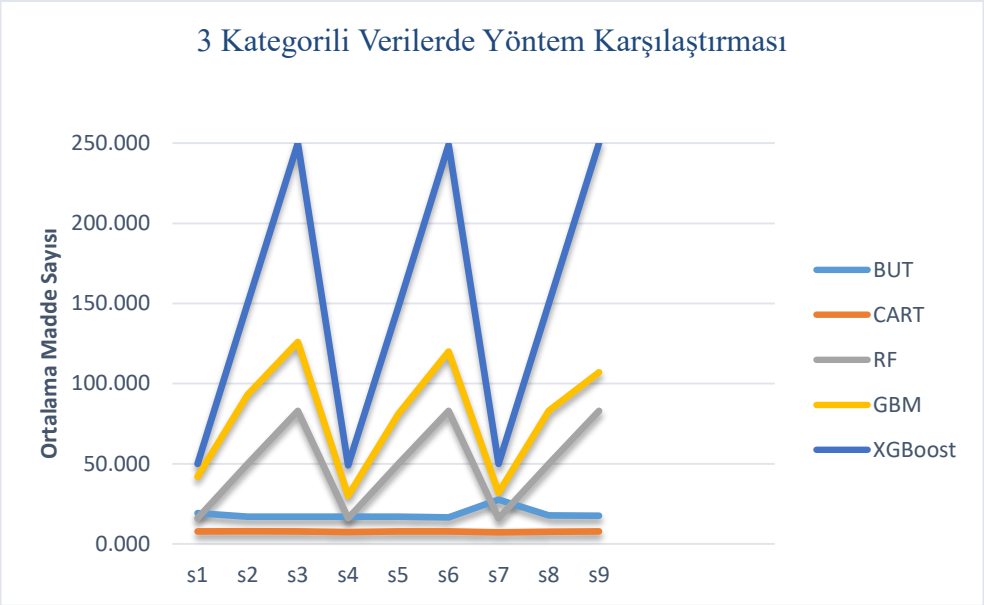
Şekil 3.6'ya göre beş kategorili verilerde makine öğrenmesi yöntemlerinde madde sayıları arttıkça RMSE değerlerinin arttığı görülmektedir. Yöntemler karşılaştırıldığında bütün senaryolarda en düşük RMSE değeri XGBoost yönteminden elde edilmiştir. GBM ve RF yöntemleri birbirine yakın RMSE değerleri verirken özellikle S_3, S_6 ve S_9' da CART ve BUT yöntemlerinin RMSE değerleri diğer yöntemlerden oldukça farklıdır.

3.4. Senaryolara ait Ortalama Madde Sayısına Göre Yöntem Karşılaştırması



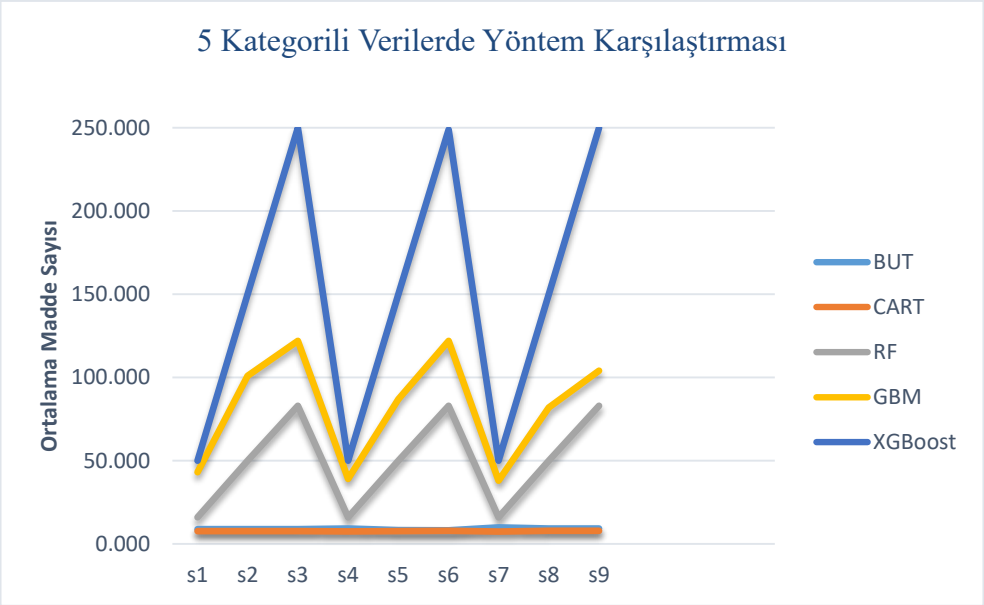
Şekil 3.7. İki Kategorili Verilerde Ortalama Madde Sayısına Göre Yöntem Karşılaştırması

Şekil 3.7'ye göre iki kategorili verilerde en düşük ortalama madde sayısı CART yönteminde kullanılırken en yüksek madde XGBoost yönteminde kullanılmıştır. Toplam madde sayısı arttıkça CART yönteminde kullanılan madde sayısında bir değişim görülmemektedir. BUT yönteminde ise kullanılan madde sayısının azaldığı görülmektedir. XGBoost, GBM ve RF yöntemlerinde toplam madde sayıları arttıkça kullanılan madde sayıları artış göstermektedir.



Şekil 3.8. Üç Kategorili Verilerde Ortalama Madde Sayısına Göre Yöntem Karşılaştırması

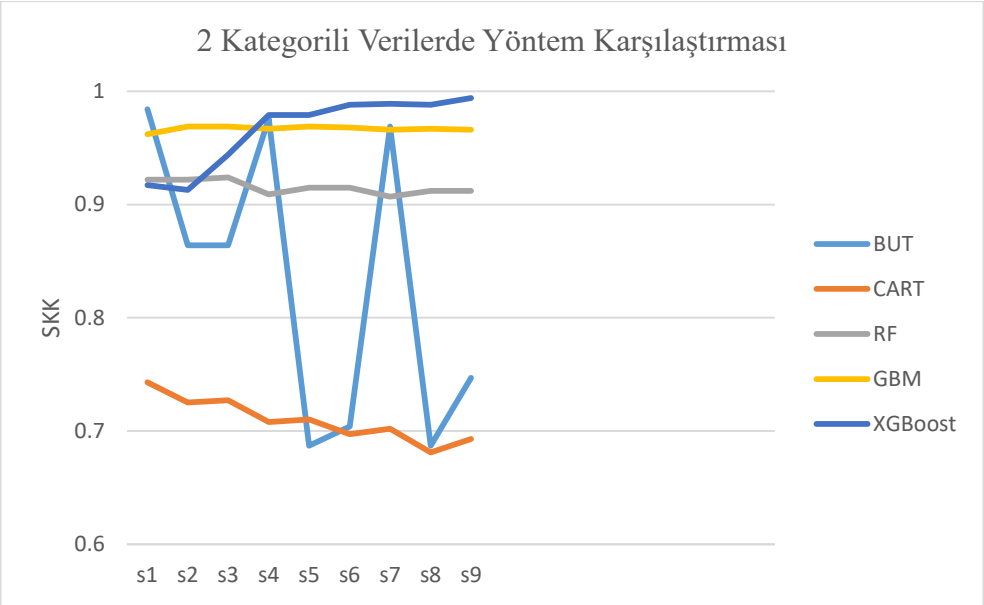
Şekil 3.8'e göre üç kategorili verilerde en düşük ortalama madde sayısı CART yönteminde kullanılırken en yüksek madde XGBoost yönteminde kullanılmıştır. Toplam madde sayısı arttıkça CART yönteminde kullanılan madde sayısında bir değişim görülmemektedir. BUT yönteminde ise kullanılan madde sayısı S_7 hariç benzer bir dağılım göstermektedir. RF, GBM ve XGBoost yöntemlerinde toplam madde sayıları arttıkça kullanılan madde sayıları da artış göstermektedir.



Şekil 3.9. Beş Kategorili Verilerde Ortalama Madde Sayısına Göre Yöntem Karşılaştırması

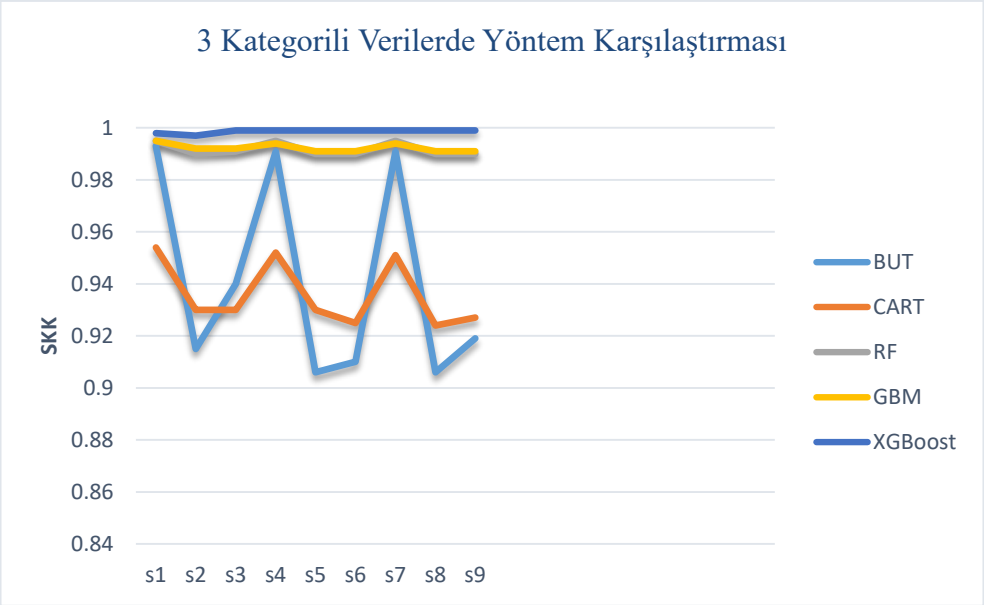
Şekil 3.9'a göre beş kategorili verilerde en düşük ortalama madde sayısı CART yönteminde kullanılırken en yüksek madde XGBoost yönteminde kullanılmıştır. CART ve BUT yöntemlerinin ortalama madde sayıları birbirine oldukça yakındır. Toplam madde sayısı arttıkça CART ve BUT yöntemlerinde kullanılan madde sayısında bir değişim görülmemektedir. RF, GBM ve XGBoost yöntemlerinde toplam madde sayıları arttıkça kullanılan madde sayıları da artış göstermektedir.

3.5. Senaryolara ait SKK Değerlerine Göre Yöntem Karşılaştırması



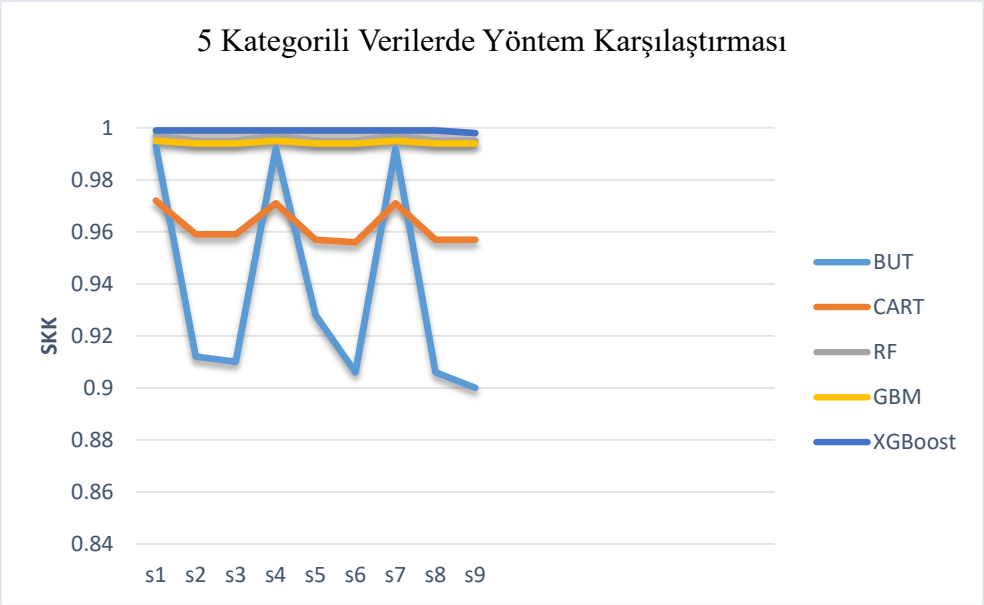
Şekil 3.10. İki Kategorili Verilerde SKK' ya Göre Yöntem Karşılaştırması

Şekil 3.10'a göre iki kategorili verilerde en düşük SKK değeri S_5 hariç CART yönteminde iken en yüksek SKK değeri S_4' den sonra XGBoost yönteminden elde edilmiştir. BUT yönteminde θ ve β' nın örtüştüğü senaryolarda yüksek SKK elde edilirken, diğer durumlarda SKK değerleri düşmüştür. RF ve GBM yöntemlerinden elde edilen SKK değerleri senaryolar arasında büyük değişiklikler göstermemektedir. RF, GBM ve XGBoost yöntemlerinin SKK değerleri 0,90'ın üstündedir. Bu yöntemler için tahmin edilen değerler ile toplam skor değerlerinin arasında mükemmel derecede uyum olduğu görülmektedir.



Şekil 3.11. Üç Kategorili Verilerde SKK' ya Göre Yöntem Karşılaştırması

Şekil 3.11'e göre üç kategorili verilerde en düşük SKK değeri BUT yönteminde iken en yüksek SKK değeri XGBoost yönteminden elde edilmiştir. RF, GBM ve XGBoost yöntemlerinde senaryolar arasındaki SKK değerlerinde fazla değişim görülmemektedir. Özellikle θ ve β' nın örtüşmediği senaryolarda BUT ve CART yöntemlerinden elde edilen SKK değerlerinin diğer yöntemlerden farklı olduğu görülse de genel olarak bütün yöntemlerin SKK değeri 0,90'ın üstündedir. Karşılaştırılan yöntemlerden elde edilen tahmin değerleri ile toplam skorlar arasında mükemmel derecede uyum olduğu görülmektedir.



Şekil 3.12. Beş Kategorili Verilerde SKK' ya Göre Yöntem Karşılaştırması

Şekil 3.12'ye göre beş kategorili verilerde en düşük SKK değeri BUT yönteminde iken en yüksek SKK değeri XGBoost yönteminden elde edilmiştir. RF, GBM ve XGBoost yöntemlerinin SKK değerleri birbirine oldukça yakındır. BUT ve CART yöntemlerinde SKK değerleri bakımından senaryolar arasında farklılık olduğu görülse de genel olarak bütün yöntemlerin SKK değeri 0,90'ın üstündedir. Karşılaştırılan yöntemlerden elde edilen tahmin değerleri ile toplam skorlar arasında mükemmel derecede uyum olduğu görülmektedir.

4.TARTIŞMA

Sağlık alanında doğrudan ölçüm yapılamayan örtük (latent) değişkenlerin değerlendirilmesinde kullanılan ölçekler, bireylerin θ düzeyleri hakkında bilgi sahibi olmak için kullanılır. Ölçekler, klasik kağıt-kalem yöntemiyle uygulanmasının yanısıra BUT yöntemi ile de uygulanabilir. Bu şekilde bireylerin θ düzeyleri daha kısa sürede ve daha az sayıda madde kullanarak elde edilebilir. Günümüzde bu alanda yapılan çalışmaların büyük çoğunluğunda BUT uygulamaları kullanılmaktadır.

Son yıllarda giderek önemli bir hale gelen makine öğrenmesi yöntemleri hemen her alanda kullanılmaya başlanmıştır. Sağlık alanında da önemini giderek arttırmaktadır. Bu tez çalışmasında ölçeklerden bireylerin θ düzeylerini elde etmek için kullanılan BUT yöntemine alternatif olarak makine öğrenmesi yöntemlerinin kullanılıp kullanılamayacağı detaylı bir biçimde incelenmiştir.

Michel ve ark. 2016 yılında yaptığı çalışmada CART ve BUT yöntemlerinin tahmin performanslarını karşılaştırmıştır. Bu çalışmada özellikle skorların tahmin değerleri arasında bir uyum tespit edilmiştir. CART ile bireylerin toplam skorunu 50,09, BUT ile ise 50.00 olarak tahmin etmiştir. Ayrıca bu iki yöntemin RMSE değerlerine bakıldığında ise her iki yöntem için de en iyi olduğu modellerin RMSE değerleri CART için 0,16, BUT için ise 0,22 olarak bulunmuştur. Sonuç olarak Michel ve ark. CART yönteminin BUT yöntemine göre alternatif olarak kullanabileceğini göstermiştir (Michel ve ark., 2016). Bu tez çalışmasında Michel ve ark.'nın yaptığı çalışmayı daha da geliştirerek farklı makine öğrenmesi yöntemlerini de analiz edildi. Bu analiz sonucunda farklı makine öğrenmesi yöntemlerinin de psikometrik model tabanlı BUT algoritmasına alternatif olarak kullanılabileceği görülmüştür.

Squires ve ark. 2019 yılında yaptığı çalışmada diyabetik lomber omurga cerrahi hastalarında fiziksel fonksiyonların değerlendirilmesi için BUT yöntemini kullanmışlardır (Squires ve ark., 2019).

Boudreaux ve ark. 2019 yılında yaptığı çalışmada acil serviste intihar ilişkisi taramasını BUT yöntemi ile yapmışlardır. 19 maddenin yer aldığı ölçek ile çalışan Boudreaux ve ark. BUT yöntemi ile 4 madde ile istedikleri sonuçlara ulaşmışlardır (Boudreaux ve ark., 2019). Yapmış olduğumuz tez çalışmasında da 250 maddelik ölçek BUT yöntemi ile 9 maddeye indirgenmiştir.

Edmond ve ark. 2019 yılında yaptığı çalışmada kronik bel ağrılı hastaların yön tercihi, bilişsel davranışsal girişimleri ve sonuçlarını BUT yöntemi analiz etmişlerdir (Edmond ve ark., 2019).

Chester ver ark. 2019 yılında yaptığı çalışmada öz yeterlilik ve kalıcı omuz ağrısı riskini CART yöntemi ile analiz etmişlerdir (Chester ve ark., 2019).

Yangın 2019 yılında yaptığı tez çalışmada diyabet hastaları üzerinde makine öğrenmesi yöntemlerinin performanslarını karşılaştırmıştır. CART, RF, GBM ve XGBoost yöntemlerini karşılaştıran Yangın, XGBoost yönteminin diğer yöntemlere daha iyi sonuç verdiğini ortaya koymuştur (Yangın, 2019).

5.SONUÇ VE ÖNERİLER

Yapılan tez çalışmasında makine öğrenmesi yöntemlerinin psikometrik model tabanlı klasik BUT algoritmasına alternatif olarak kullanılıp kullanılmayacağı araştırılmıştır. Makine öğrenmesi yöntemlerinden CART, RF, GBM ve XGBoost yöntemleri kullanılmıştır. Rasch modelinden farklı senaryolarda türetilen yapay veride bu yöntemlerin performansları değerlendirilmiştir. Bu senaryolarda madde sayıları ile (θ) ve (β) değerleri değişkenlik göstermektedir.

İki kategorili yanıta sahip maddeleri içeren soru bankasında toplam madde sayısı arttıkça BUT yönteminde kullanılan madde sayısında azalma görülmektedir. CART yöntemi yaklaşık 8 madde ile en az sayıda madde ile tahmin yapan yöntemdir. BUT yönteminde kullanılan ortalama madde sayısı yaklaşık 37 ile 50 arasında değişmektedir. RF yönteminde kullanılan ortalama madde sayısı 16 ile 83 arasında değişirken GBM yönteminde madde sayısı 18 ve 80 arasında değişmektedir. RMSE değerlerine bakıldığında özellikle XGBoost yönteminde RMSE değeri diğer yöntemlere göre daha düşüktür. SKK değerlerine bakıldığında özellikle RF, GBM ve XGBoost yöntemlerinin SKK değerleri oldukça yüksektir.

Üç kategorili yanıta sahip maddeleri içeren soru bankasında bütün yöntemlerin SKK değerleri oldukça yüksektir. Toplam skorları yüksek uyumla tahmin etmiştirler. Kullanılan madde sayılarına bakıldığında yine en düşük madde sayısı CART yönteminde kullanılmıştır. Ancak CART yönteminin RMSE ve MAE değerleri diğer yöntemlere göre oldukça yüksektir.

Beş kategorili maddeleri içeren soru bankası sonuçlarına bakıldığında BUT yönteminde kullanılan madde sayısının önemli derecede azaldığı görülmüştür. Kategori sayısı ve toplam madde sayısı arttıkça BUT yönteminde kullanılan madde sayısında önemli derecede azalma olduğu görülmektedir. Ayrıca SKK değerlerinin de bütün yöntemler için 0,90 değerinin üzerindedir. Bu da tahmin edilen değerler ile toplam skorlar arasında mükemmel derecede uyum olduğunu göstermektedir.

Bu tez çalışmasında makine öğrenmesi yöntemlerinin farklı göstergeler bazında BUT yöntemi için alternatif olarak kullanılabileceği görülmüştür. Çalışma yapan kişinin amacına bağlı olarak düşük RMSE, düşük madde sayısı ya da yüksek SKK için farklı makine öğrenmesi yöntemleri BUT yöntemine alternatif olarak kullanılabilir. Düşük ortalama madde sayısı ile ilgilenildiğinde özellikle CART yöntemi BUT yöntemine alternatif olarak kullanılabilir. Düşük RMSE ile ilgilenildiğinde XGBoost yöntemi BUT yöntemine alternatif olarak kullanılabilir ancak XGBoost yönteminde kullanılan ortalama madde sayısı yüksek olduğu için RF ve GBM yöntemleri de bu durumda kullanılabilir. SKK ile ilgilenildiğinde ise büyük madde sayılarında bütün yöntemlerin BUT yöntemine alternatif olarak kullanılabileceği önerilmektedir.

ÖZET

Rasch Modeline Uygun Yanıt Desenlerinde Bilgisayar Uyarlamalı Test Yöntemi ile Makine Öğrenmesi Yöntemlerinin Tahmin Performanslarının Benzetim Çalışması İle Karşılaştırılması

Sağlık alanında doğrudan ölçüm yapılamayan örtük (latent) değişkenlerin değerlendirilmesinde kullanılan ölçekler bireylerin θ düzeyleri hakkında bilgi sahibi olmak için kullanılır. Ölçekler klasik kalem-kağıt uygulaması yanısıra BUT yöntemi ile de uygulanabilir.

Bu tez çalışmasında, farklı senaryolarda Rasch modelinden türetilen yapay verilerde BUT uygulamasında yaygın olarak kullanılan algoritma (algoritma içerisinde olasılıksal model olarak Rasch modeli ve soru seçim kriteri olarak da bilgi kriterine dayalı soru seçimi) ile BUT uygulamasına alternatif yeni bir yaklaşım olarak makine öğrenmesi algoritmalarının (CART, RF, GBM ve XGBoost) performansları incelenmiştir.

Farklı parametreler için değerlendirilen yöntemlerin birbirilerinden üstün olduğu farklı parametreler bulunmuştur. Madde sayıları dikkate alındığında CART yönteminin en az madde ile tahminleri yaptığı sonucu ortaya çıkmıştır. RMSE değerlerine bakıldığında ise GBM ve XGBoost yöntemlerinin düşük RMSE değerlerine sahip olduğu görülmektedir. Karşılaştırılan yöntemler toplam skorları tahmininde iyi bir SKK değerine sahiptir.

Genel olarak yapılan karşılaştırmalar sonucunda, yeni bir çalışma planlayan bir kişinin amacına uygun olarak BUT yöntemine alternatif olarak makine öğrenmesi yöntemlerini kullanması önerilmektedir.

Anahtar Kelimeler: BUT, Makine Öğrenmesi, Ölçek, Rasch,

SUMMARY

Comparison of Prediction Performance of Computer Adaptive Testing with Machine Learning Methods in Response Patterns Suitable for Rasch Model with Simulation Study

The scales used to measure latent variables, which cannot be measured directly in the field of health, are used to have information about the characteristics (θ) levels of individuals. Scales can be answered with classical methods as well as CAT method.

In this thesis, the performance of machine learning algorithms (CART, RF, GBM and XGBoost) as a new approach to CAT application with the algorithm routinely used in CAT application (Rasch model as a probabilistic model in the algorithm and information criterion as a question selection criterion) in an simulation data derived from different scenarios from the Rasch model.

Different parameters were found in which the methods evaluated for different parameters are superior to each other. When the number of items is taken into account, it is revealed that the CART method makes predictions with the least number of items. When RMSE values are analyzed, it is seen that GBM and XGBoost methods have low RMSE values. The methods compared have a good ICC value in estimating total scores.

As a result of general comparisons, it is recommended that a person planning a new study use machine learning methods as an alternative to the CAT method in accordance with its purpose.

Keywords: CAT, Machine Learning, Rasch, Scale

KAYNAKLAR

BREIMAN L, FRIEDMAN JH, OLSHEN RA, STONE CJ. 1984. Classification and Regression Tree. 1st Ed.

BOUDREAUX ED, DE BEURS DP, NGUYEN TH, HASKİNS BL, LARKİN C, BARTON B.2019. Applying Computer Adaptive Testing Methods to Suicide Risk Screening in the Emergency Department. *Suicide Life Threat Behav*. **49(4)**:917-927.

BROWNLEE J (2016). A Gentle Introduction to the Gradient Boosting Algorithm for Machine Learning. Erişim Adresi: [https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/]Erişim Tarihi: 08/06/2020.

BROWNLEE J (2019). Difference Between Classification ans Regression in Machine Learning. Erişim Adresi: [https://machinelearningmastery.com/classification-versus-regression-in-machine-learning/]. Erişim Tarihi: 08/06/2020.

CHESTER R, KHONDOKER M, SHEPSTONE L, LEWIS JS, JEROSCH-HEROLD C.2019. Self-efficacy and Risk of Persistent Shoulder Pain: Results of a Classification and Regression Tree (CART) Analysis. *Br J Sports Med*. 53(13):825-834

DANGETİ P. 2017. Statistical for Machine Learning.1st Ed.

DESHPANDE GC, RAO SC, KEİL AD, PATOLE SK. 2011. Evidence-based Guidelines for Use of Probiotics in Preterm Neonates. *BMC Med.* **2**: 9-92.

DHİRAJ K (2019). Top 5 advantages and disadvantages of Decision Tree Algorithm. Erişim Adresi: [https://medium.com/@dhiraj8899/top-5-advantages-and-disadvantages-of-decision-tree-algorithm-428ebd199d9a]. Erişim Tarihi: 08/06/2020.

EDMOND SL, WERNEKE MW, YOUNG M, GRİGSBY D, MCGİLL T, MCCLENAHAN B. 2019. Directional Preference, Cognitive Behavioural Interventions, and Outcomes Among Patients With Chronic Low Back Pain. *Physiother Res Int*. **24(3)**: e1773.

HEALY BC, ZURAWSKİ J, GONZALEZ CT, CHİTNİS T, WEİNER HL, GLANZ BI. 2019. Assessment of Computer Adaptive Testing Version of the Neuro-QQL for People With Multiple Sclerosis. *Mult. Scler.* **25(13)**:1791-1799.

JAMES G, EİTTEN D, HASTİE T, TİBSHİRANİ R. 2014. An Introduction to Statistical Learning with Applications in R. 4th Ed.

KASKATI OT (2011). Rasch Modelleri Kullanarak Romatoid Artiritli Hastaları Özürlülük Değerlendirimi İçin Bilgisayar Uyarlamalı Test Yönteminin Geliştirilmesi. Doktora Tezi, Ankara Üniversitesi, Sağlık Bilimleri Enstitüsü

KESKİN MV (2018). Büyük Veride Makine Öğrenmesi Uygulaması. Yükseklisans Tezi, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü.

LOH WY 2011. Classification and Regression Trees. *Wires Data Min Knowl*. **1**: 14-23

MİCHEL P, BAUMSTARCK K, LOUNDOU A, GHATTAS B, AUQUİER P, BOYER L.2018. Computerized adaptive testing with decision regression trees: an alternative to item response thetory for quality of life measurements in multiple sclerosis. *Patient Prefer Adher*.**12**: 1043-1053.

MURATLAR ER (2020a). Gradient Boosted Regresyon Ağaçları. Erişim Adresi: [https://www.veribilimiokulu.com/gradient-boosted-regresyon-agaclari/]. Erişim Tarihi: 08/06/2020.

MURATLAR ER (2020b). XGBoost Nasıl Çalışır? Neden İyi Performans Gösterir? Erişim Adresi: [https://www.veribilimiokulu.com/xgboost-nasil-calisir/]. Erişim Tarihi: 08/06/2020.

NATEKIN A, KNOLL A. 2013. Gradient Boosting Machines, a Tutorial. *Front Neurorobotics*. **7**:1-21.

ÖZTUNA D (2008). Kas İskelet Sisteminin Sorunlarının Özürlülük Değerlendirmesinde Bilgisayar Uyarlamalı Test Yönteminin Uygulanması. Doktora Tezi, Ankara Üniversitesi, Sağlık Bilimleri Enstitüsü

REZAİE M, GOLSHAN M,2015. Computer Adaptive Test(CAT): Advantages and Limitations. *IJEI*. **2(5)**: 128-137.

RUDNER L.M. (1998). An On-Line, Interactive, Computer Adaptive Testing Tutorial. Erişim Adresi: [http://edres.org/scripts/cat]. Erişim Tarihi: 20/03/2020.

SEVİMLİ SAİTOĞLU Y.(2015). Sınıflama ve Regresyon Ağaçları. Doktora Tezi. Marmara Üniversitesi Sosyal Bilimler Enstitüsü

SINGH H (2018). Understanding Gradient Boosting Machines. Erişim Adresi: [https://towardsdatascience.com/understanding-gradient-boosting-machines-9be756fe76ab]. Erişim Tarihi: 08/06/2020.

SQUİRES MD, BRODKE DS, NEESE A, ZHANG Y, SPİKER WR, LAWRENCE B SPİNA N. 2019. Physical Function Computer Adaptive Test Outcomes in Diabetic Lumbar Spine Surgical Patientes. *Spine J*. **19(6)**:1048-1056.

YANGIN G (2019). XGBoost ve Karar Ağacı Tabanlı Algoritmanın Diyabet Veri Setleri Üzerine Uygulaması. Yükseklisans Tezi. Mimar Sinan Güzel Sanatlar Üniversitesi. Fen Bilimleri Enstitüsü

ÖZGEÇMİŞ

I- Bireysel Bilgiler

Adı ve Soyadı : Emrah Gökay ÖZGÜR

Doğum Yeri ve Tarihi : Düzce 21/03/1986

Uyruğu :T.C

Medeni Durumu :Evli

Askerlik Durumu :Yaptı. 30/09/2013.

İletişim Adresi ve Telefonu : Ankara Üniversitesi Tıp Fakültesi
Biyoistatistik Anabilim Dalı
Sıhhiye/Altındağ/ANKARA
0312 595 8196

II- Eğitimi

Yüksek Lisans: Marmara Üniversitesi Sağlık Bilimleri Enstitüsü
Biyoistatistik Anabilim Dalı (İstanbul)

Lisans : Fırat Üniversitesi Fen- Edebiyat Fakültesi
İstatistik Bölümü (Elazığ)

Lisans : Anadolu Üniversitesi İktisat Fakültesi
İktisat Bölümü (Eskişehir)

Lise : Çağrıbey Anadolu Lisesi (İstanbul)

Yabancı Dil

YÖKDİL : İngilizce 87,50 17/03/2018

III- Ünvanları

2015- Araştırma Görevlisi

2013-2015 İstatistikçi

IV- Mesleki Deneyimi

2016- :Araştırma Görevlisi. Ankara Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı

2015-2016 : Araştırma Görevlisi. Kocaeli Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı

2013-2015 : İstatistikçi. Mersin Halk Sağlığı Müdürlüğü

V- Bilimsel İlgi Alanları

Yayınların Listesi

Uluslararası Hakemli Dergilerde Yayımlanan Makaleler

Akdemir AO, Karabakan M, Aktas BK, Bozkurt A, **Özgür EG**, Akdoğan N, Yarıs M (2018). Visceral Adiposity Index Is Useful for Evaulating Obesity Effect on Erectile Dysfunction. Andrologia. DOİ: 10.1111/and.13282

Baydemir C, **Özgür EG**, Balci S (2018). Evaluation of adherence to Mediterranean diet in medical students at Kocaeli University, Turkey. Journal of International Medical Research, 46(4), 1585-1594., Doi: 10.1177/0300060518757158

Ulusal Hakemli Dergilerde Yayımlanan Makaleler

Özgür EG, Bekiroğlu GN, Baydemir C (2017). Sağlık Bilimlerinde Çoklu Uyum Analizi ve Uygulaması. Kocaeli Üniversitesi Sağlık Bilimleri Dergisi, 3(2), 9-18. (Kontrol No: 3810867)

Ulusal ve Uluslararası Bilimsel Toplantılarda Sunulan Poster ve Bildiriler

1. Toprak U, **Özgür EG,** Hayme S, Doğanay Erdoğan B (2018). Comparing Prediction Abilities of Classification Algortihms on Determining Specific

Chemical Substances. XX. Ulusal ve III. Uluslararası Biyoistatistik Kongresi (Tam Metin Bildiri/ Sözlü Sunum), Yayın No: 4463589.

2. **Özgür EG**, Bekiroğlu GN (2017). Çoklu Uyum Analizi (MCA): Üniversite Öğrencilerinde Yeme Alışkanlıkları. 19. Ulusal ve 2. Uluslararası Biyoistatistik Kongresi (Özet Bildiri/ Poster), Yayın No: 3810948.
3. **Özgür EG**, Günel Karadeniz P, Bekiroğlu GN (2010). Biyoistatistikçilerin Bilgisayar Kullanımına Bağlı Fiziksel Rahatsızlıkları Nordic Ölçeği. 12.Ulusal Biyoistatistik Kongresi (Özet Bildiri/ Poster), Yayın No: 2223186.
4. Gümüştepe E, Akyürek S, Arslan Y, **Özgür EG**, Gökçe ŞG (2018). Evre 3 Küçük Hücre Dışı Akciğer Karsinomunda Sarkopeni ve İnflamatuar Belirteçlerin Prognostik Önemi. 13.Ulusal Radyasyon Onkoloji Kongresi.
5. Mungan C, **Özgür EG**, Can A (2018). Miyokard Enfarktüsü Geçiren ve Kalp Kasına Kök Hücre Yapılan Bireylerin Yaşam Kalitesinin Değerlendirilmesi. 14. Ulusal Histoloji ve Embriyoloji Kongresi.

Kitap Bölümü

1. Yavuz Z, **Özgür EG**, Gökmen D (2018). Analitik Araştırmalar – Bilimsel Araştırma Yöntemleri Kitabı. AÜ Tıp Fakültesi Yayın No: 468