T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

# FİNANSAL TEKNOLOJİ ALANINDA KULLANICI DENEYİMLERİNİN MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE İNCELENMESİ

**Gizem ZENGİN**

YÜKSEK LİSANS TEZİ

İstatistik Anabilim Dalı

İstatistik Programı

Danışman

Prof. Dr. Ersoy ÖZ

Haziran, 2022

T.C.

YILDIZ TEKNİK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

# FİNANSAL TEKNOLOJİ ALANINDA KULLANICI DENEYİMLERİNİN MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE İNCELENMESİ

Gizem ZENGİN tarafından hazırlanan tez çalışması 26/05/2022 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, İstatistik Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Prof. Dr. Ersoy Öz
Yıldız Teknik Üniversitesi
Danışman

**Jüri Üyeleri**

Prof. Dr. Ersoy Öz, Danışman
Yıldız Teknik Üniversitesi _____

Prof. Dr. Semra Erpolat Taşabat, Üye
Mimar Sinan Güzel Sanatlar Üniversitesi _____

Doç. Dr. Öyküm Esra Yiğit, Üye
Yıldız Teknik Üniversitesi _____

Danışmanım Prof. Dr. Ersoy ÖZ sorumluluğunda tarafımca hazırlanan Finansal Teknoloji Alanında Kullanıcı Deneyimlerinin Makine Öğrenmesi Yöntemleri İle İncelenmesi başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Gizem ZENGİN

İmza

*Sevgili aileme*

# TEŞEKKÜR

Yüksek lisans ve lisans eğitimim süresince desteğini esirgemeyen, gerek eğitimim boyunca gerekse bu araştırmanın tüm süreçlerinde bana yol gösterip sabırla ışık tutan değerli danışman hocam Prof. Dr. Ersoy ÖZ'e akademik hayatıma katkılarından dolayı çok teşekkür ederim. Ayrıca, beni her zaman destekleyen, cesaret veren ve motive eden, her koşulda yanımda olan biricik aileme ve sevgilime teşekkür ederim.

Gizem Zengin

# İÇİNDEKİLER

SİMGE LİSTESİ ..... vii

KISALTMA LİSTESİ ..... viii

ŞEKİL LİSTESİ ..... ix

TABLO LİSTESİ ..... x

ÖZET ..... xiii

ABSTRACT ..... xv

1 GİRİŞ ..... 1

1.1 Literatür Özeti ..... 2

1.2 Tezin Amacı ..... 4

1.3 Hipotez ..... 5

2 FİNANS TEKNOLOJİSİ VE KULLANICI DENEYİMİ ..... 7

2.1 Finans Teknolojisi ..... 7

2.2 Kullanıcı Deneyimi ..... 9

3 MAKİNE ÖĞRENMESİ ..... 10

3.1 Gözetimli Öğrenme ..... 10

3.2 Gözetimsiz Öğrenme ..... 11

3.3 Pekiştirmeli Öğrenme ..... 11

3.4 Araştırmada Kullanılan Makine Öğrenmesi Algoritmaları ..... 12

3.4.1 Lojistik Regresyon (LR) ..... 12

3.4.2 Karar Ağaçları (KA) ..... 12

3.4.3 Destek Vektör Makineleri (DVM) ..... 13

3.4.4 Rasgele Orman (RO) ..... 17

3.4.5 Gradyan Arttırma Makineleri (GAM) ..... 17

3.4.6 Uyarlanabilir Arttırma (AdaBoost) ..... 18

3.4.7 Hafif Gradyan Arttırma Makineleri (HGAM) ..... 18
3.4.8 Ekstrem Gradyan Arttırma (XGBoost) ..... 19
3.4.9 Kategorik Arttırma (CatBoost) ..... 19
3.4.10 K En Yakın Komşu (KNN) ..... 20
3.5 Başarı Metrikleri ..... 20
3.6 Öznitelik Seçim Yöntemleri ..... 21
3.6.1 Feature Importance Fonksiyonu ..... 22
3.6.2 Mutual Info Yöntemi ..... 22
3.6.3 Shap Fonksiyonu ..... 22
3.6.4 Borda Yöntemi ..... 22
3.7 Veri Seti Bölme ..... 23
4 UYGULAMA ..... 24
4.1 Veri Seti ..... 24
4.2 Veri Ön İşleme ..... 26
4.3 Veri Setine Ait Tanımlayıcı İstatistikler ..... 27
4.4 Veri Modelleme ..... 29
4.5 Tüm Veri Seti Modelleri için Uygulama Sonuçları ..... 30
4.5.1 Tüm Özniteliklerle Gerçekleştirilen Sınıflandırma Sonuçları ..... 31
4.5.2 Açıklayıcılığı En Yüksek 15 Öznitelik ile Gerçekleştirilen Sınıflandırma Sonuçları .. 48
4.5.3 Açıklayıcılığı En Yüksek 10 Öznitelik ile Gerçekleştirilen Sınıflandırma Sonuçları .. 66
4.5 Sonuçların Değerlendirilmesi ..... 84
5 SONUÇ VE ÖNERİLER ..... 89
KAYNAKÇA ..... 92
A VERİ TOPLAMA ÇALIŞMASI ..... 96
TEZDEN ÜRETİLMİŞ YAYINLAR ..... 100

# SİMGE LİSTESİ

| | |
|---|---|
| $Q_1, \dots, Q_n$ | Ağaç Vektörleri |
| $w$ | Ağırlık Vektörü |
| $n$ | Bağımsız Değişken Katsayısı |
| $\lambda$ | Bozulma Faktörü |
| $C$ | Düzenleme Parametresi |
| $b$ | Eğilim Değeri |
| $z$ | Fayda Fonksiyonu |
| $x_i$ | Girdi Değeri |
| $x_{max}$ | Girdi Setindeki En Büyük Değer |
| $x_{min}$ | Girdi Setindeki En Küçük Değer |
| $y_1, \dots, y_i$ | Gözlenen Değerler |
| $\varepsilon$ | Hata Terimi |
| $\alpha$ | Öğrenme Oranı |
| $k$ | Örnek Sayısı |
| $\varphi$ | Özellik Haritası |
| $\beta_1, \dots, \beta_n$ | Regresyon Katsayıları |
| $\beta_0$ | Regresyon Sabiti |
| $y$ | Sınıf Etiketi |
| $y_i^p$ | Tahmin Edilen Değer |
| $\xi_i$ | Yapay Değişken |

# KISALTMA LİSTESİ

| | |
|---|---|
| AO | Ağırlıklı Ortalama |
| DVM | Destek Vektör Makineleri |
| DO | Doğrukuk Oranı |
| DUO | Duyarlılık Oranı |
| XGBoost | Ekstrem Gradyan Arttırma (Extreme Gradient Boosting) |
| FinTech | Finansal Teknoloji (Financial Technology) |
| GN | Gerçek Negatif |
| GP | Gerçek Pozitif |
| GS | Gerçek Sınıf |
| GAM | Gradyan Arttırma Makineleri |
| HGAM | Hafif Gradyan Arttırma Makineleri |
| HO | Hata Oranı |
| KA | Karar Ağaçları |
| CatBoost | Kategorik Arttırma (Categoric Boosting) |
| KNN | K En Yakın Komşu |
| KO | Kesinlik Oranı |
| LR | Lojistik Regresyon |
| MO | Makro Ortalama |
| ÖS | Örnek Sayısı |
| RBF | Radyal Tabanlı Fonksiyon (Radial Basis Function) |
| RO | Rasgele Orman |
| TES | Tahmin Edilen Sınıf |
| AdaBoost | Uyarlanabilir Arttırma (Adaptive Boosting) |
| YN | Yanlış Negatif |
| YP | Yanlış Pozitif |

# ŞEKİL LİSTESİ

**Şekil 3.1** Optimum hiperdüzlemlerin olası konumları [34] ..... 14
**Şekil 3.2** Doğrusal olarak ayrılabilen veri setleri için hiperdüzlem [33] ..... 15
**Şekil 4.1** Model 1.1'e göre sınıflandırma algoritmalarının doğruluk oranları .... 34
**Şekil 4.2** Model 1.2'ye göre sınıflandırma algoritmalarının doğruluk oranları .. 38
**Şekil 4.3** Model 1.3'e göre sınıflandırma algoritmalarının doğruluk oranları .... 41
**Şekil 4.4** Model 1.4'e göre sınıflandırma algoritmalarının doğruluk oranları .... 45
**Şekil 4.5** Model 1.5'e göre sınıflandırma algoritmalarının doğruluk oranları .... 48
**Şekil 4.6** Model 2.1'e göre sınıflandırma algoritmalarının doğruluk oranları .... 52
**Şekil 4.7** Model 2.2'ye göre sınıflandırma algoritmalarının doğruluk oranları .. 56
**Şekil 4.8** Model 2.3'e göre sınıflandırma algoritmalarının doğruluk oranları .... 59
**Şekil 4.9** Model 2.4'e göre sınıflandırma algoritmalarının doğruluk oranları .... 63
**Şekil 4.10** Model 2.5'e göre sınıflandırma algoritmalarının doğruluk oranları .. 66
**Şekil 4.11** Model 3.1'e göre sınıflandırma algoritmalarının doğruluk oranları .. 70
**Şekil 4.12** Model 3.2'ye göre sınıflandırma algoritmalarının doğruluk oranları 73
**Şekil 4.13** Model 3.3'e göre sınıflandırma algoritmalarının doğruluk oranları .. 76
**Şekil 4.14** Model 3.4'e göre sınıflandırma algoritmalarının doğruluk oranları .. 80
**Şekil 4.15** Model 3.5'e göre sınıflandırma algoritmalarının doğruluk oranları .. 83
**Şekil 4.16** Veri setinin %50 test olarak kullanılması durumunda DO değerlerine göre tüm sınıflandırma sonuçları..... 84
**Şekil 4.17** Veri setinin %40 test olarak kullanılması durumunda DO değerlerine göre tüm sınıflandırma sonuçları..... 85
**Şekil 4.18** Veri setinin %30 test olarak kullanılması durumunda DO değerlerine göre tüm sınıflandırma sonuçları..... 86
**Şekil 4.19** Veri setinin %20 test olarak kullanılması durumunda DO değerlerine göre tüm sınıflandırma sonuçları..... 87
**Şekil 4.20** Veri setinin %10 test olarak kullanılması durumunda DO değerlerine göre tüm sınıflandırma sonuçları..... 88

# TABLO LİSTESİ

**Tablo 2.1** Finansal teknoloji gelişme dönemleri ..... 7
**Tablo 3.1** DVM'ler için temel çekirdek fonksiyonları ve formülleri ..... 16
**Tablo 3.2** Karmaşıklık matrisi ..... 20
**Tablo 4.1** Araştırma soruları ve değişken türlerine göre yanıtlar ..... 25
**Tablo 4.2** Veri setine ait tanımlayıcı istatistikler ..... 27
**Tablo 4.3** Finansal teknoloji ve kredi risk yaklaşımları için tanımlayıcı istatistikler ..... 28
**Tablo 4.3** Finansal teknoloji ve kredi risk yaklaşımları için tanımlayıcı istatistikler (devamı) ..... 29
**Tablo 4.4** Veri seti modelleri ..... 30
**Tablo 4.5** Model 1.1 için gerçekleştirilen sınıflandırma sonuçları ..... 31
**Tablo 4.5** Model 1.1 için gerçekleştirilen sınıflandırma sonuçları (devamı) ..... 33
**Tablo 4.6** Model 1.1'e ait sonuç karşılaştırma matrisi ..... 33
**Tablo 4.6** Model 1.1'e ait sonuç karşılaştırma matrisi (devamı) ..... 34
**Tablo 4.7** Model 1.2 için gerçekleştirilen sınıflandırma sonuçları ..... 35
**Tablo 4.7** Model 1.2 için gerçekleştirilen sınıflandırma sonuçları (devamı) ..... 36
**Tablo 4.8** Model 1.2'ye ait sonuç karşılaştırma matrisi ..... 37
**Tablo 4.9** Model 1.3 için gerçekleştirilen sınıflandırma sonuçları ..... 38
**Tablo 4.9** Model 1.3 için gerçekleştirilen sınıflandırma sonuçları (devamı) ..... 39
**Tablo 4.9** Model 1.3 için gerçekleştirilen sınıflandırma sonuçları (devamı) ..... 40
**Tablo 4.10** Model 1.3'e ait sonuç karşılaştırma matrisi ..... 40
**Tablo 4.10** Model 1.3'e ait sonuç karşılaştırma matrisi (devamı) ..... 41
**Tablo 4.11** Model 1.4 için gerçekleştirilen sınıflandırma sonuçları ..... 42
**Tablo 4.11** Model 1.4 için gerçekleştirilen sınıflandırma sonuçları (devamı) ..... 43
**Tablo 4.12** Model 1.4'e ait sonuç karşılaştırma matrisi ..... 44
**Tablo 4.13** Model 1.5 için gerçekleştirilen sınıflandırma sonuçları ..... 45
**Tablo 4.13** Model 1.5 için gerçekleştirilen sınıflandırma sonuçları (devamı) ..... 46
**Tablo 4.13** Model 1.5 için gerçekleştirilen sınıflandırma sonuçları (devamı) ..... 47
**Tablo 4.14** Model 1.5'e ait sonuç karşılaştırma matrisi ..... 47
**Tablo 4.15** Model 2.1 için gerçekleştirilen sınıflandırma sonuçları ..... 49
**Tablo 4.15** Model 2.1 için gerçekleştirilen sınıflandırma sonuçları (devamı) ..... 50
**Tablo 4.15** Model 2.1 için gerçekleştirilen sınıflandırma sonuçları (devamı) ..... 51
**Tablo 4.16** Model 2.1'e ait sonuç karşılaştırma matrisi ..... 51

**Tablo 4.16** Model 2.1'e ait sonuç karşılaştırma matrisi (devamı) ..... 52
**Tablo 4.17** Model 2.2 için gerçekleştirilen sınıflandırma sonuçları..... 53
**Tablo 4.17** Model 2.2 için gerçekleştirilen sınıflandırma sonuçları (devamı) .... 54
**Tablo 4.18** Model 2.2'ye ait sonuç karşılaştırma matrisi..... 55
**Tablo 4.19** Model 2.3 için gerçekleştirilen sınıflandırma sonuçları..... 56
**Tablo 4.19** Model 2.3 için gerçekleştirilen sınıflandırma sonuçları (devamı) .... 57
**Tablo 4.19** Model 2.3 için gerçekleştirilen sınıflandırma sonuçları (devamı) .... 58
**Tablo 4.20** Model 2.3'e ait sonuç karşılaştırma matrisi..... 58
**Tablo 4.20** Model 2.3'e ait sonuç karşılaştırma matrisi (devamı) ..... 59
**Tablo 4.21** Model 2.4 için gerçekleştirilen sınıflandırma sonuçları..... 60
**Tablo 4.21** Model 2.4 için gerçekleştirilen sınıflandırma sonuçları (devamı) .... 61
**Tablo 4.22** Model 2.4'e ait sonuç karşılaştırma matrisi..... 62
**Tablo 4.23** Model 2.5 için gerçekleştirilen sınıflandırma sonuçları..... 63
**Tablo 4.23** Model 2.5 için gerçekleştirilen sınıflandırma sonuçları (devamı) .... 64
**Tablo 4.23** Model 2.5 için gerçekleştirilen sınıflandırma sonuçları (devamı) .... 65
**Tablo 4.24** Model 2.5'e ait sonuç karşılaştırma matrisi..... 65
**Tablo 4.25** Model 3.1 için gerçekleştirilen sınıflandırma sonuçları..... 67
**Tablo 4.25** Model 3.1 için gerçekleştirilen sınıflandırma sonuçları (devamı) .... 68
**Tablo 4.26** Model 3.1'e ait sonuç karşılaştırma matrisi..... 69
**Tablo 4.27** Model 3.2 için gerçekleştirilen sınıflandırma sonuçları..... 70
**Tablo 4.27** Model 3.2 için gerçekleştirilen sınıflandırma sonuçları (devamı) .... 71
**Tablo 4.27** Model 3.2 için gerçekleştirilen sınıflandırma sonuçları (devamı) .... 72
**Tablo 4.28** Model 3.2'ye ait sonuç karşılaştırma matrisi..... 72
**Tablo 4.29** Model 3.3 için gerçekleştirilen sınıflandırma sonuçları..... 73
**Tablo 4.29** Model 3.3 için gerçekleştirilen sınıflandırma sonuçları (devamı) .... 74
**Tablo 4.29** Model 3.3 için gerçekleştirilen sınıflandırma sonuçları (devamı) .... 75
**Tablo 4.30** Model 3.3'e ait sonuç karşılaştırma matrisi..... 75
**Tablo 4.30** Model 3.3'e ait sonuç karşılaştırma matrisi (devamı) ..... 76
**Tablo 4.31** Model 3.4 için gerçekleştirilen sınıflandırma sonuçları..... 77
**Tablo 4.31** Model 3.4 için gerçekleştirilen sınıflandırma sonuçları (devamı) .... 78
**Tablo 4.32** Model 3.4'e ait sonuç karşılaştırma matrisi..... 79
**Tablo 4.33** Model 3.5 için gerçekleştirilen sınıflandırma sonuçları..... 80
**Tablo 4.33** Model 3.5 için gerçekleştirilen sınıflandırma sonuçları (devamı) .... 81
**Tablo 4.33** Model 3.5 için gerçekleştirilen sınıflandırma sonuçları (devamı) .... 82

**Tablo 4.34** Model 3.5'e ait sonuç karşılaştırma matrisi..... 82
**Tablo 4.34** Model 3.5'e ait sonuç karşılaştırma matrisi (devamı) ..... 83

# ÖZET

## Finansal Teknoloji Alanında Kullanıcı Deneyimlerinin Makine Öğrenmesi Yöntemleri ile İncelenmesi

Gizem ZENGİN

İstatistik Anabilim Dalı

Yüksek Lisans Tezi

Danışman: Prof. Dr. Ersoy ÖZ

Finans teknolojisi, finans alanındaki faaliyetleri geliştirmek ve iyileştirmek için teknolojiyi kullanan, dinamik, gelişmekte olan ve aynı zamanda geleneksel finans yöntemleri ile rekabet içerisinde olan bir sektördür. İnternetin hayatımızın her alanında yarattığı değişiklikler ve kolaylıkların finans sektörüne yansıması olarak ele alınabilecek bu sektör, özellikle 2008 küresel finans krizinden sonra büyüme fırsatı yakalamıştır. Küresel kriz ile aynı yıllarda hayatımıza giren akıllı telefonlar finansal teknolojiler için ciddi bir piyasa açığı oluşturmuştur. Kriz sonrası bankalara duyulan güvenin azalması ve geleneksel bankacılık sisteminde artan sorunlar bu gelişimi tetikleyen sebeplerden sadece bazılarıdır. Sektördeki bu açığı fark eden girişimciler finansal hizmetler ve teknolojiyi birleştirerek, maliyetler ile fiyatları azaltarak finansal teknoloji piyasasına girmişlerdir. Krizle beraber olgunlaşmaya başlayan bu sektör, son on yılda hızla büyüyerek gelişmeye devam etmektedir. Finansal teknoloji alanındaki bu hızlı ivmelenmeyle beraber bankalar teknolojiye ayak uydurmak zorunda kalmış ve geleneksel bankacılığın yerini alacak teknolojiler geliştirmeye başlamışlardır. Günümüzde bu teknolojilerin başında gelen mobil uygulamalar ve internet şubeler kullanıcıların bulundukları

yerden, en az efor ve maliyetle bireysel veya kurumsal birçok bankacılık işlemini uçtan uca kontrol edebileceği bir seviyeye taşımıştır. Dijital bankacılık uygulamaları günümüzde birçok kullanıcının güvenle ve kolaylıkla bankacılık işlemlerini gerçekleştirdiği bir araç olarak görülürken, bazı kullanıcı grupları tarafından geleneksel bankacılık yöntemleri güvenilirlik veya kullanım alışkanlıkları açısından halen tercih edilebilmektedir. Bu araştırma; İstanbul'da ikamet eden on sekiz yaşından büyük bireylerin dijital bankacılık uygulamalarına karşı tutumlarını, kullanıcı gruplarını ve kullanım alışkanlıklarını, sosyoekonomik ve demografik açılardan ele alarak makine öğrenmesi yöntemleri ile incelemektedir. Veri setinin farklı eğitim ve test bölünmeleri için sınıflandırmaya yönelik Kategorik Arttırma, Hafif Gradyan Arttırma Makineleri, Ekstrem Gradyan Arttırma, Gradyan Arttırma, Uyarlanabilir Arttırma, K En Yakın Komşu, Karar Ağaçları, Rasgele Orman, Lojistik Regresyon ve Radyal Tabanlı Fonksiyon Çekirdekli Destek Vektör Makineleri algoritmaları ile uygulamalar gerçekleştirilmiştir. Araştırmadan elde edilecek çıktılar, dijital bankacılık uygulamalarını kullanım düzeylerini belirlemeye yönelik sınıflandırma algoritmalarının karşılaştırılması ve bu sınıflandırmada etkili olan önemli değişkenlerin belirlenmesi ile finansal teknoloji kullanımına yönelik değerlendirmeler içermektedir.

**Anahtar Kelimeler:** Finansal teknoloji, fintek, dijital bankacılık, kullanıcı deneyimi, makine öğrenmesi.

# ABSTRACT

## Examining User Experiences in the Field of Financial Technology with Machine Learning Methods

Gizem ZENGİN

Department of Statistics

Master's Thesis

Supervisor: Prof. Dr. Ersoy ÖZ

Fintech is a dynamic and emerging industry that uses technology to develop and improve operations in finance while competing with traditional financial methods. This sector can be considered as, the reflection of the changes and conveniences created by the internet in all areas of our lives, on the finance sector. Especially after the 2008 global financial crisis, it caught the opportunity to grow. Smartphones, which entered our lives in the same years as the global crisis, created a serious market deficit for financial technologies. The decrease of trust in banks after the crisis and the increasing problems in the traditional banking system are just some of the reasons that triggered this development. Realizing this gap in the sector, entrepreneurs entered the fintech market by combining financial services and technology, reducing costs and prices. This sector, which started to mature with the crisis, continues to develop rapidly in the last ten years. With this rapid acceleration in the field of financial technology, banks had to keep up with technology and started to develop technologies to replace traditional banking. Today, mobile applications and internet branches, which are at the forefront of these technologies, have brought users to a level where they can control many individual or corporate banking transactions end-to-end with minimum effort and

cost. While digital banking applications are seen as a tool that many users can safely and easily perform their banking transactions today, traditional banking methods are still preferred by some user groups in terms of reliability or usage habits. This research; examines the attitudes, user groups and usage habits of individuals over the age of eighteen residing in Istanbul, to digital banking applications, by considering socio-economic and demographic aspects, with machine learning methods. For different versions of training and test divisions of the data set, classifications were made with Categoric Boosting, Light Gradient Boosting Machines, Extreme Gradient Boosting, Gradient Boosting Machines, Adaptive Boosting, K Nearest Neighbours, Decision Tree, Random Forest, Logistic Regression and Radial Basis Function Kernel Support Vector Machines algorithms. The outputs to be obtained from the research include the comparison of classification algorithms to determine the level of use of digital banking applications, the determination of the important variables that are effective in this classification, and evaluations for the use of financial technology.

**Keywords:** Financial technology, fintech, digital banking, user experience, machine learning.

# 1
# GİRİŞ

Gelişen teknoloji ve değişen trendler neticesinde insan hayatında önemli bir yere sahip olan akıllı telefonlarla birlikte alınan hizmetlerin de dijitalleşmesi birçok kullanıcıyı barındıran farklı pazarlar oluşturmaya başlamıştır. Tarihçesi oldukça geçmişe dayanan finansal teknolojilerin günümüzde aldığı yapıya ulaşması pek kolay olmamıştır. Günümüz şartlarında bankacılık işlemlerinin neredeyse tamamının mobil bir hal alması, fiziksel efora ihtiyaç duymaksızın bulunduğu yerden gerçekleştirilebilmesi artık sıradan bir hal alsa da finansal teknolojiler her geçen gün gelişmeye ve değişmeye devam etmektedir. Bununla birlikte geleneksel bankacılık yöntemleri yaşamını sürdürse de bankaların fiziksel işlem hacminin azaldığı ve yerini dijitale bıraktığı aşikardır. Dijital kanalların bu kadar geliştiği ve değiştiği bir dönemde, bankaların müşterilerine sunduğu dijital hizmetler üzerinden rekabet ettiği, kullanıcı odaklı dijital ürünlerin değer kazandığı bir atmosfer oluşmuştur. Bu doğrultuda sunulan dijital ürünlerin kullanıcı tarafından nasıl karşılandığı, kullanım alışkanlıkları, neden tercih etme veya etmeme eğiliminde olunduğu, kullanılma veya kullanılmama motivasyonlarının neler olduğu gibi hedef kitlenin beklentilerinin belirlendiği çeşitli faktörler bir finansal teknoloji uygulamasının ana yaşam döngüsünde oldukça önemli bir yere sahiptir. Ürünün yaşam döngüsünü devam ettirebilmesi ve güncelliğini koruması bakımından hedef kitlelerin çok iyi tanımlanması ve beklentilerini karşılamaya yönelik geliştirmeler yapılması oldukça önemlidir.

Bu araştırmada, farklı kullanıcı gruplarının finansal teknoloji uygulamalarına olan tutum ve yaklaşımlarını makine öğrenmesi yöntemleri ile araştırılmış, 15 farklı model üzerinde uygulanan yöntemler doğruluk oranlarına göre değerlendirilerek analiz edilmiştir.

Çalışmanın ilk bölümünde literatür araştırmaları, tezin amacı ve hipotez sunulmuştur. İkinci bölümünde finans teknolojisi ve kullanıcı deneyimi kavramları ele alınmıştır. Ardından üçüncü bölümde araştırmanın yöntemi olan makine öğrenmesi yöntemleri ve araştırmada kullanılan diğer algoritmalara yer

verilmiştir. Dördüncü bölümde ise araştırmanın uygulama kısmı, veri seti ve yapılan analizler sonucundaki çıktılar yorumlanmıştır. Beşinci ve son bölümde ise araştırmanın sonucu ve genel değerlendirmelere yer verilmiştir.

## 1.1 Literatür Özeti

Bu araştırma kapsamında literatür taraması gerçekleştirilirken, öncelikle makine öğrenmesi alanında çeşitli araştırmalar incelenmiş, ardından finansal teknolojiler alanındaki araştırmalar ele alınmıştır. Daha sonra ise makine öğrenmesi yöntemlerinden yararlanılarak finansal teknoloji alanında gerçekleştirilen çalışmalar araştırılmıştır.

Makine öğrenmesi terimi ilk olarak 1959 yılında Amerikan bilgisayar bilimcisi Arthur Samuel tarafından öne sürülmüştür [1]. Astronomi, biyoloji ve genetik alanında yaygın olarak kullanılan büyük verinin günümüzde başta finans ve bankacılık sektörü olmak üzere, sağlık sektörü, pazarlama, yapay zekâ ve tahmin konusunda çalışmaların ilerlemesi ile birçok farklı alanda belirleyici rol oynamaya başladığı görülmektedir [2]. 1995 yılında Russel ve Norvig tarafından yayınlanan "Artificial Intelligence: A Modern Approach" isimli makalelerinde yapay zekâ üzerine modern bir yaklaşımda bulunarak, makine öğrenmesi algoritmalarının hedeflenen sonuca göre birkaç sınıfa ayrıldığına değinmişlerdir. Bu sınıflar; gözetimli öğrenme, gözetimsiz öğrenme ve pekiştirmeli öğrenmedir [3].

Teknolojiye en çok yatırım yapan sektörlerin başında gelen bankacılık sektörü müşterilerini mobil ve internet bankacılığına yönlendirmektedir. Türkiye'de internet bankacılığının tarihi 1997 yılına dayanmaktadır ve bu hizmeti ilk sunmaya başlayanlar Garanti Bankası ile Türkiye İş Bankası'dır. 2019 yılında Fettahoğlu ve Kıldıze tarafından, İstanbul'da ikamet etmekte olan 102 katılımcı ile finansal okuryazarlık seviyelerine göre finansal teknolojilere karşı tutumlarının değerlendirildiği bir araştırma gerçekleştirilmiştir. Araştırmanın sonucunda, katılımcıların finansal ürün ve hizmetlerin çoğunluğu hakkında bilgi sahibi olduğu ancak daha fazla bilgi gerektiren ürün ve hizmetleri kullanmaktan kaçınıldığı gözlenmiştir. Ayrıca aracısız işlem yapma ve algılanan fayda faktörlerinin kullanıcıların bahsi geçen ürünleri kullanmaları üzerinde pozitif bir etkiye sahip olduğu ortaya çıkmıştır [4].

Emerson ve diğerlerinin 2019 yılında hazırlamış oldukları kantitatif finans alanında makine öğreniminin eğilimleri ve uygulamaları konulu araştırmalarında, kantitatif yatırım geleneksel olarak önde gelen bir alan olmakla birlikte yeni tekniklerin benimsenmesi açısından literatürde en yaygın kullanım örneklerine rastlanmıştır [5].

Zhang ve diğerlerinin 2020 yılında makine öğrenmesi yöntemlerinden yararlanarak bir çevrimiçi kredi yatırım platformu örneği ile fintek riskini analiz ettikleri araştırmalarında, çift yönlü uzun ve kısa vadeli belleğe dayalı soru-cevap metinlerinden oluşan risk tanıma modeli geliştirmişlerdir. 3 adımdan oluşan metinsel analiz ve değerlendirme sonrasında kullanıcının yanıtlarına göre ilgili yatırım platformunun risk seviyesi değerlendirilmektedir. Geliştirilen modelin deneysel sonuçlarına göre oldukça başarılı F1 skor ve doğruluk oranı elde edilmiş olup, modelin çevrimiçi yatırım platformlarında yayınlanan metinlerden risklerin belirlenmesi için uygulanabilir ve yatırımcılara rehberlik edebilir bir araç olarak kullanılabilirliği gösterilmiştir [6].

Chen ve Yeh tarafından 2019 yılında kitle fonlaması fintek projesinin başarı tahminlemesi için makine öğrenimi yaklaşımıyla geliştirdikleri araştırma sonuçlarına göre; kitlesel fonlama projelerinin başarısının sosyal medya faaliyetleri, fon sağlayıcıların insan sermayesi ve çevrimiçi proje sunumlarının büyük verileri analiz edilerek proje başarısının tahmin edilebilir olduğunu ortaya koymuşlardır [7].

Gambacorta ve diğerleri, 2017 yılı Mayıs ve Eylül ayları arasında Çin'in önde gelen bir fintek firmasına ait verilerden yararlanarak, makine öğrenimi yöntemlerine dayalı kredi puanlama modellerinin tahmin gücü ile geleneksel kayıp ve temerrüt modellerinin karşılaştırıldığı araştırma gerçekleştirmişlerdir. Bu araştırmanın sonucuna göre makine öğrenimine ve geleneksel olmayan verilere dayalı modelin, toplam kredi arzına olumsuz bir şok olması durumunda, geleneksel modellere göre kayıpları ve temerrütleri daha iyi bulduğu tespit edilmiştir [8].

Lacruz ve Saniie tarafından gerçekleştirilen fintek kredi kartı dolandırıcılık tespitinde makine öğrenimi konulu arşatırmaya göre; makine öğrenmesi ve yapay zeka tahminlerinden elde edilen veriler finansal kurumların operasyonel

maliyetlerini 2030 yılına kadar %22 düşürmelerine olanak tanıyacağı öngörülmüştür. Ayrıca yüksek doğrulukta dolandırıcılık tespiti için yarı denetimli öğrenme ve lojistik regresyon olmak üzere uygulanan iki yöntem için de %94 KO ile umut verici sonuçlar elde edilmiştir [9].

Vives'in 2017'de yayımladığı araştırma makalesinde, fintek firmalarının büyük veriye dayalı gelişmiş istatistiksel modeller kullanarak çeşitli veri kombinasyonarı gerçekleştirebilir, geleneksel bankaların erişemediği verilere ulaşabilir, potansiyel borçluları daha iyi tanıyabilir olunduğunu ortaya koymuştur [10].

ABD'de uygulanan bir araştırmaya göre Y kuşağının neredeyse yarısının kredi puanlarının kendilerini etkilediğini düşünmeleri veya geleneksel finans kurumlarının kredi geçmişi olmayan bireylere yüksek faiz oranları uygulamasının sonucu olarak bireylerin krediden yoksun bırakılmıştır. Agarwal ve diğerlerinin 2020 yılında gerçekleştirdiği araştırma, kredi başvuru sonuçlarını tahmin etmede bireysel borçluların mobil ayak izlerinin etkisi analiz edilmiştir. Y kuşağı bireylerinin yeterli kredi geçmişi ve kredi puanı olmadığı durumlarda, fintek kuruluşu alternatif bir kredi tarama süreci olarak bireylerin mobil ayak izini kullanmaktadır. Bir bireyin hem kaba mobil ayak iznin hem de arama günlüklerine dayalı daha derin sosyal ayak izinin, büyük veriden yararlanılarak, temerrütleri tahmin etmede kredi puanına sahip bir modelden önemli ölçüde daha iyi performans gösterdiği basit tahmine dayalı bir model tespit edilmiştir [11].

Makine öğrenmesi yöntemlerinin finans veya finansal teknoloji alanında literatürde çeşitli uygulamaları mevcuttur, bunlar arasından bazılarına bu kısımda yer verilmiştir. Ancak kullanıcı deneyimi ve finansal teknolojilerin makine öğrenmesi ile bir araya geldiği araştırmalara literatürde pek rastlanmamıştır.

## 1.2 Tezin Amacı

Bu araştırmada, Türkiye'de faaliyet gösteren bankaların dijital bankacılık uygulamalarının (mobil uygulama ve internet şube) kullanımını incelemek üzere İstanbul'da ikamet eden on sekiz yaşından büyük bireylerin tutumları araştırılmıştır. Araştırmada sorular, kullanıcıların demografik yapısı, sosyoekonomik seviyeleri, eğitim durumları, kredi kullanım alışkanlıkları, risk algıları gibi çeşitli değişkenlere yanıt oluşturacak şekilde hazırlanmıştır.

Dijital ürünlerde kullanıcı deneyimi, ürünün hayatta kalabilmesi ve yatırımcısına gelir veya prestij sağlaması açısından son derece kritiktir. Kullanıcı deneyiminin iyileştirilmesi için birçok önemli nokta bulunmaktadır ancak bunların öğrenilmesi ve geliştirilmesi için öncelikle hedef kitlenin çok iyi tanımlanması gerekmektedir. Hedef kitle, yani ürüne talep gösterecek kullanıcıların ürünü tercih etmesi için öncelikle hedef kitle alışkanlıklarının doğru bir şekilde analiz edilmesi oldukça kritik bir aşamadır. Bu nedenle, bu araştırma kapsamında katılımcıların çeşitli yönlerden alışkanlıkları analiz edilerek, dijital bankacılık uygulamalarını kullanım düzeyleri hakkında bilgi toplanmıştır. Dijital bankacılık uygulamalarını sıklıkla kullanan kitle ve bu kullanıcı kitlesinin uygulamalara olan yaklaşımının yanı sıra dijital bankacılık uygulamalarını kullanmayı tercih etmeyen veya az kullanmayı tercih eden kitleler de veri sağlayıcı olarak ele alınmıştır. Dijital bankacılık uygulamalarının geliştirilmesinde, farklı kullanıcı gruplarına hizmet edebilecek yeni fonksiyonlar geliştirilmesinde veya bu uygulamaları kullanmayan kitlelere kullanım alışkanlıkları sağlamada izlenebilecek alternatif yollara ışık tutabilecek sonuçlar elde edilmesi oldukça önemlidir.

Bu tezin amacı, İstanbul ilinde ikamet eden 18 yaş üstü bireylerin finans sektörüne yönelik tutumları ve demografik bilgilerini kullanarak finansal teknoloji kullanım düzeylerine yönelik makine öğrenmesi algoritmaları ile sınıflandırma algoritmalarını kullanarak modellemeler gerçekleştirmektir. Ayrıca sınıflandırma başarısını etkileyen önemli değişkenlerin (özniteliklerin) belirlenmesi de amaçlanmaktadır. Elde edilecek modeller sayesinde finans sektörüne yönelik finansal teknoloji kullanım düzeyleri noktasında değerlendirmeler yapılabilecektir.

## 1.3 Hipotez

Finansal teknoloji uygulamalarında kullanıcı deneyimi ürünün tercih edilebilirliği ve sürekliliği açısından önemli olmakla birlikte insan teknoloji etkileşiminin özgünlüğünü ve ihtiyaca dayalı olmasını sağlamaya odaklıdır. Bu araştırma kapsamında kullanıcıların finansal teknoloji uygulamalarına karşı tutumları ele alınmıştır. Makine öğrenmesi yöntemlerinden yararlanılarak kullanıcılardan elde edilen verilere göre finansal teknoloji uygulamalarına karşın tutumları belirli

modellemelerle gerçekleştirilecektir. Bunun sonucunda kullanıcıların dijital bankacılık uygulamalarına yönelik tutumları ile alakalı değişkenlerin (özniteliklerin) değerleri biliniyorken makine öğrenmesi sınıflandırma algoritmalarından yararlanarak başarılı sınıflandırmalar yapılabileceği ortaya konacaktır.

# 2
# FİNANS TEKNOLOJİSİ VE KULLANICI DENEYİMİ

## 2.1 Finans Teknolojisi

Finans teknolojisi veya fintek, finansal hizmetlerin sunumunda geleneksel finansal yöntemlerle rekabet etmeyi amaçlayan bir teknoloji ve yeniliktir. Bir başka deyişle; finansal hizmetlerin daha hızlı, kolay ve kaliteli bir şekilde kullanıcılara sunulması için finans ve teknolojinin bir araya gelmesinden oluşan yeni akım "FinTech" olarak tanımlanmaktadır [12].

Finansal teknolojiler, maliyetleri azaltma, daha fazla şeffaflık sağlama, aracıları azaltma ve finansal bilgiye erişimi sağlama konusunda bireylere fırsatlar sağlar [13].

Finans ve teknoloji her ne kadar son yıllarda ortaya çıkan ve gelişmekte olan bir birleşme olarak görülse de aslında tarihçesi oldukça uzun yıllara dayanmaktadır.

**Tablo 2.1** Finansal teknoloji gelişme dönemleri

| Dönem | Fintek 1.0 | Fintek 2.0 | Fintek 3.0 | Fintek 3.5 |
|---|---|---|---|---|
| | 1866 – 1967 | 1967 – 2008 | 2008 – Devam etmekte | |
| Anahtar Oyuncular | Altyapı | İnternet | Start-up'lar | |

Fintek 1.0 dönemi olarak adlandırılan ve 1866 – 1967 yıllarını kapsayan dönem finansal küreselleşmenin başladığı yıllardır. Bu dönemin teknolojileri arasında yer alan telgraf, transatlantik kablosu gibi gelişmeler o dönemde ülkeler arasındaki finansal ilişkilerin gelişmesine katkıda bulunmuştur [14].

Fintek 2.0'ın başlangıcı 1967 yılında ilk ATM'nin (Automatic Teller Machine) kullanılması ile gerçekleşmiş, 2008 yılına kadar sürecek bir dönemin başlangıcı olmuştur. 2008 küresel krizine kadar devam edecek bu dönem içerisinde birçok

yenilik barındırmaktadır. İnternet bankacılığı, ATM, kredi kartlarının kullanıma sunulması bu dönemde sektöre kazandırılan yeniliklerden sadece bazılarıdır. Yine bu dönemde finansal hizmetler küreselleşerek döneme göre daha dijital bir hal almaya devam etmiştir. 1987'de yaşanan "Black Monday" olarak adlandırılan borsa çöküşü bu küreselleşme ve teknolojinin sağladığı entegrasyonun bir örneği olarak görülmüştür, çünkü Amerika'da yaşanan borsa çöküşü dünya genelini etkilemiştir. İlk cep telefonu, internetin kullanılması gibi tarihte önemli gelişmeler yine Fintek 2.0 döneminde yaşanmıştır [15].

2008 küresel finans krizi yeni bir dönemin başlangıcı olmuş ve bununla birlikte fintek kavramı günümüzde algılandığı yapıya ulaşmıştır. Fintek 3.0 dönemi beraberinde küresel ekonomik kriz ve bankacılık sisteminin büyük zarar görmesi sektördeki en büyük gelişmelerin başlangıcı olmuştur. İlk akıllı telefonun hayatımıza girdiği bu yıllarda, internetin mobil hal alması, taşınabilir teknolojilerin hızlı gelişimi ve mobil uygulama kavramının hayatımıza girmesiyle birlikte sektör baş döndürücü bir gelişim sürecine girmiştir. Geleneksel bankacılık sistemine duyulan güvenin azalması, kriz sonrası bankaların karlılığının düşmesi ve buna karşın fintek girişimlerinin bu fırsatı görerek düşük maliyetlerle topluma ulaşmaları neticesinde büyümelerinde ivme kazanmışlardır. Ardından 2009 yılında Bitcoin ve kripto para kavramlarının hayatımıza girmesi, mobil ödeme sistemlerinin gelişmesi gibi gelişmeler yaşanmıştır [15].

Finans teknolojisindeki bu büyümeyle birlikte özellikle bankacılık sektöründe şubeleşme ve istihdam daralması yaşanması beklenmektedir. Bununla birlikte sektörde olmayan yeni roller ve pozisyonlar doğacak, yeni unvanlar oluşturulacak ve dijital çağı içselleştirmiş, dijital çağı benimseyerek yetiştirilmiş genç istihdamın artması ön görülmektedir. Bu dijital dönüşüme ayak uydurabilen kurumların büyüyerek sektörde devam etmesi, dönüşüme ayak uydurmakta başarılı olamayan firmaların ise sektörden kaybolması olasıdır. Finans teknolojisi, dijital dönüşüm anlamında geçirmiş olduğu bu büyük gelişimler paralelinde kendisiyle birlikte birçok sektöre de liderlik etmeye devam etmektedir [15].

## 2.2 Kullanıcı Deneyimi

Kullanıcı deneyimi literatürde çeşitli şekillerde tanımlansa da temelde ürün kullanıcı etkileşimine dayanmaktadır. Kullanıcının dijital bir ürün veya servise karşı verdiği tepki ve davranışına ve bunları etkileyen faktörlere odaklanmaktadır. Gelişen teknoloji ve internetin hayatımızın her alanında yer almasıyla birlikte kullanıcı deneyiminin insan faktörü üzerine yoğunlaşması ve insan ile makineler arasındaki bağlantının güçlendirilmesi oldukça önem kazanmıştır. Teknolojik ürünlerin kullanılabilirliğinin arttırılması amacıyla tasarlanan kullanıcı merkezli ara yüzler tercih edilmeye başlanmıştır. Günümüzde en az dijital ürünün varlığı kadar önemli olan kolay etkileşim, erişilebilirlik ve kişiselleştirme olguları dijital bir ürünün olmazsa olmazları arasında yerini almaktadır [16].

Özellikle son yıllarda rekabetin oldukça arttığı finansal teknoloji alanında bankalara karşın küçük start-up şirketlerin yatırım bulması ve doğrudan sektörün önde gelenleri ile rekabet edebilir bir hale gelmesi bu alandaki yeniliklerin hız kesmeden devam ediyor olmasını sağlamaktadır. Bu yeni şirketler, dijitale doğmuş olmanın avantajını yaşarken, geleneksel bankacılık fiziksel ve yüz yüze etkileşimlerden vazgeçme veya kaynak yönetimine odaklanma sorunları yaşayabilmektedirler. Bu noktada hızlı ve çevik olmak, dijital dönüşümleri takip edebilmek bankacılık sektörü için oldukça önemli bir hal almıştır. İşte bu noktada, kullanıcı deneyimi ve etkileşimi, gelenekselden dijitale geçiş ve sürdürülebilirlik aşamalarında oldukça önem kazanmaktadır [17]. Kullanıcı odaklı ürünler yaratmak, kullanıcı deneyimini iyileştirmek ve geliştirmek dijital bankacılık sektörünün hızlı dinamiklerine ayak uydurmak tüm dünyada olduğu gibi Türkiye'de de sektördeki tüm markaların öncelikleri arasında yer almaktadır.

# 3
# MAKİNE ÖĞRENMESİ

Makine öğrenmesi, büyük veriden öğrenerek, verilerin hatasız ve nispeten daha hızlı işlenmesine imkân verirken, gelecekte olabilecek olaylar hakkında tahminlerde bulunarak ve gerçekleşmesi olası senaryolara karşı doğru kararlar alınabilmesini sağlar. Kullanım alanı yıllar içerisinde gelişerek oldukça fazla alanda kullanılmaya devam etmektedir.

Makine öğrenimi, bilgisayarların açıkça programlanmadan, belirli görevleri yerine getirmeleri için sağlanan verilerden yararlanarak öğrenmelerini ve görevleri nasıl yerine getirebileceklerini keşfetmelerini içerir [18].

Tom M. Mitchell'ın 1997 yılında yayınlanan kitabında yer vermiş olduğu tanıma göre; bir bilgisayar programının performans ölçüsü "P" ve bazı "T" görev sınıflarında gösterdiği performans "E" deneyimi ile iyileşmekteyse, "P" ölçüsünün "E" deneyiminden öğrendiği söylenmektedir [19].

Makine öğrenmesi yöntemlerinde kullanılan veriler etiketli veriler ve etiketsiz veriler olmak üzere iki tipten oluşmaktadır. Etiketli veriler bir algoritmanın eğitilmesi için kullanılırken, etiketsiz veriler ise eğitilen algoritmanın test sürecinde kullanılmaktadır. Makine öğrenmesi yöntemlerinde genellikle üç farklı öğrenme modeli kullanılmaktadır; pekiştirmeli öğrenme, gözetimli öğrenme ve gözetimsiz öğrenme [20].

## 3.1 Gözetimli Öğrenme

Bu yöntemde etiketli veri setindeki kalıplar bulunarak öğrenme gerçekleştirilir [21]. İlk olarak girdi veriler etiketlenir ve çıktı değişkenler tanımlanır ve sonrasında makine öğrenmesi algoritması girdi verilerle çıktı veriler arasındaki ilişkiyi bulmak üzere eğitilir. Algoritma eğitimi tamamlandıktan sonra işlem yeni verilere uygulanır [22]. Modele girdi verileri sağlandıkça, model uygun oluncaya kadar ağırlıklarını düzenler. Bu durum, modelde aşırı uyum veya yetersiz uyum yaşanmaması için bir çapraz doğrulama aşaması sağlamaktadır [23]. Eğitim seti

kullanılarak öğrenilen ve tahmine dayalı performansı karşılamak için görünmeyen verilere uygulanabilen bir fonksiyon oluşturmayı gerektiren bu yöntemin amacı, işlevini hiç görmediği veriler üzerinde iyi bir şekilde genelleştirerek oluşturmaktır [24]. Bir örnek ile açıklamak gerekirse; bir bankanın geçmiş dönemlerde müşterilerine sağlamış olduğu kredilere ait veriler bankanın veri tabanında saklanmaktadır. Burada girdiler kredi alan müşterilerin çeşitli açılardan özellikleridir, çıktı ise kredinin geri ödenip ödenmemesidir. Elde olan veri setine en uygun model oluşturularak test edilir, en başarılı performansı sunan algoritma daha sonraki kredi başvuru taleplerinde kullanılmak üzere başvuru sahibinin krediyi ödeyip ödeyemeyeceğini tahminlemede kullanılır [25].

Makine öğrenmesinde çıktı değişkenler nümerik veya kategorik olabilmektedir. Nümerik çıktı değişken olması halinde regresyon algoritmalarından, kategorik çıktı değişken olması halinde sınıflandırma algoritmalarından yararlanılmaktadır.

## 3.2 Gözetimsiz Öğrenme

Bu yöntemde gözetimli öğrenmede uygulanan etiketleme işlemi gerçekleştirilmeden tüm veriler üzerinden ilişki araştırılır. Gözetimsiz öğrenme tekniği gözlemlere dayalı öğrenme gerçekleştirmektedir. Yani; etiketleme işlemi yapılmadığından, veri setindeki ham verilerin tamamı üzerinde öğrenme işlemi gerçekleştirilir ve doğrulamak istediği bir hedef değişken bulunmadığından minimize etmeyi hedeflediği bir hata bulunmaz [26]. Bu yöntemde amaç tahminleme yapmak değil, veriyle ilgili çıkarımlar yapmaktır. İsminden de anlaşılacağı gibi bu yöntem bir eleştirmenden yoksundur ve performansını ölçmenin bir yolu yoktur [24].

## 3.3 Pekiştirmeli Öğrenme

Pekiştirmeli öğrenme yöntemi, gözetimli öğrenme ile benzerlik gösteren ancak algoritmanın basit veri ile eğitilmediği bir davranışsal makine öğrenmesi yöntemidir [23]. Belirli bir sorun için en iyi öneri geliştirilmek üzere başarılı sonuçlar dizisi pekiştirilir. Bu model deneme yanılma yöntemiyle öğrenmektedir. Pekiştirmeli öğrenme, bir ortamdaki durumlar ve belli durumlarda var olan eylemler bağlamı ile mevcuttur [24].

## 3.4 Araştırmada Kullanılan Makine Öğrenmesi Algoritmaları

### 3.4.1 Lojistik Regresyon (LR)

Diğer standart regresyon modellerinde olduğu gibi lojistik regresyon kullanılarak bağımlı ve bağımsız değişken arasındaki ilişki araştırılmaktadır. Çoklu regresyon yöntemi olan LR modelinde bağımlı değişkenin ikili sınıftan oluşurken, bağımsız değişkenlerin ise kategorik veya sürekli değişken türündedir [27]. Lojistik regresyon bağımlı değişkenin stokastik bir olay olduğunu varsayarak, bağımlı ve bağımsız değişkenler arasındaki ilişkiyi minimum değişken kullanılarak maksimum uyumu elde edecek sekilde tanımlamayı amaçlar [28]. Bahsi geçen ilişkiler $y = f(x)$ şeklinde basit bir fonksiyon ile, $x$ bağımsız değişkeni ve $y$ bağımlı değişkeni temsil edecek şekilde tanımlanabilir. İstatistiksel olarak $x$ değeri biliniyorken, $y$ değeri belirli bir hata oranı ile tahmin edilebilirdir. Lojistik kümülatif dağılım fonksiyonunu tanımlamak için Denklem 3.1 kullanılmaktadır.

$$y = f(x) = \frac{1}{1+e^{-z}} \tag{3.1}$$

Denklemde yer alan $z$, "Fayda Fonksiyonu" olarak tanımlanmaktadır ve denklem 3.2'deki gibi gösterilmektedir.

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon \tag{3.2}$$

Fayda fonksiyonunun formülüne göre $\beta_0$ değeri sabit, $\beta_1, \ldots, \beta_n$ değerleri ise tahmin değeriyle çarpılan regresyon katsayılarını temsil etmektedir. Veri setinde yer alan bağımsız değişken sayısı $n$ değeri ile temsil edilirken, $\varepsilon$ değeri hata terimini temsil etmektedir. LR kesin meydana gelen bir durumun olasılığını tahminlemek için kullanıldığında, maksimum olabilirlik yaklaşımı, çoklu iterasyonlar değerlendirilerek, gözlenen ve tahmin edilen sonuçlar arasındaki olsaı en küçük sapmayı ele alır [29].

### 3.4.2 Karar Ağaçları (KA)

Çok sayıda gözlem içeren bir veri setini, bir dizi karar kuralları uygulayarak daha küçük veri kümelerine bölmeye yarayan bir yöntemdir. KA, önceden tanımlanmış bir hedef değişkene sahiplerdir ve yapıları itibari ile en tepeden en aşağıya inen hiyerarşik bir stratejileri bulunmaktadır [30]. KA karar ve bitiş düğümlerinden

oluşmaktadır. Bir düğümü iki veya daha fazla alt düğüme bölmek için birden fazla algoritma kullanılmaktadır. Alt düğümlerin oluşturulması düğümlerin homojenliğini arttırdığı gibi hedef değişkene göre saflığını da arttırmaktadır. Bu yöntemin en büyük avantajı anlaşılabililiği ve yorumlanması basit oluşudur. Bu yöntem hem kategorik hem de nümerik veri işleyebilmektedir. Karar Ağaçları sınıflandırma amacıyla kullanıldığında, farklı dallara ayırma yöntemlerinden yararlanır. Denklem 3.3'te bilgi kazanımını ölçmede kullanılan düzensizlik (entropy) ya da Denklem 3.4'te GINI kirliliği (impurity) fonksiyonları günümüzde sıklıkla kullanılmaktadır [31].

$$H(X) = -\sum_x P(x) \log_2 P(x) \tag{3.3}$$

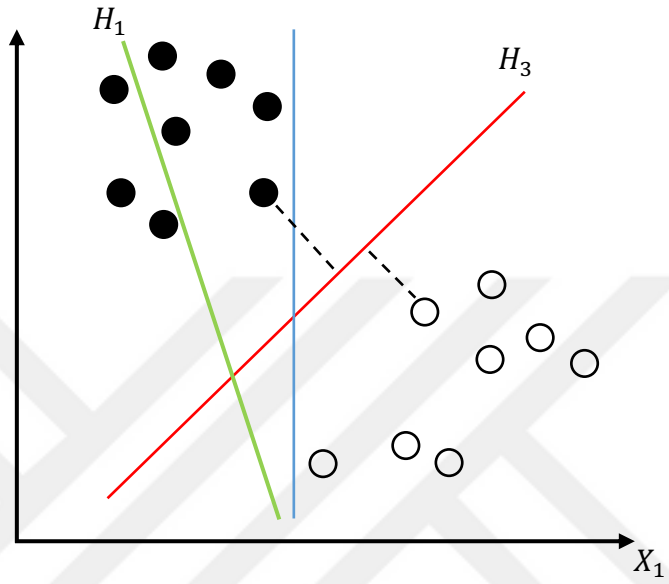
$$GINI(X) = 1 - \sum_x P(x)^2 \tag{3.4}$$

### 3.4.3 Destek Vektör Makineleri (DVM)

1963 yılında Vladimir Vapnik ve Alexey Chervonenkis tarafından temelleri atılan DVM, istatistiksel öğrenme teorisine dayalı bir gözetimli öğrenme tekniğidir. Temelleri 60'lı yıllara dayansa da, 1995 yılında Vladimir Vapnik, Berhard Boser ve Isabelle Guyon tarafından geliştirilmiştir [32]. DVM hem doğrusal hem de doğrusal olmayan problemlerde kullanılabilir olmalarıyla birlikte, çok fazla değişken olduğu durumlarda ve karmaşık yapılandırılmış verilerle baş etmede oldukça başarılıdır. Yeni Kernel (çekirdek) teknolojisi ile birlikte yüksek derecede doğrusal olmayan sınıflandırıcıları çok başarılı bir şekilde oluşturdukları gözlenmektedir [33].

Doğrusal DVM ile sınıflandırmada, iki sınıfa ait örneklerin doğrusal olarak dağıldığı varsayılırsa, eğitim verisi kullanılarak elde edilen bir karar fonksiyonu aracılığıyla iki sınıfın birbirinden ayrılması amaçlanmaktadır. Veri kümesinin ikiye ayıran doğru karar doğrusu olarak adlandırılmaktadır. Sonsuz tane karar doğrusu çizilebilmesine karşın asıl önemli olan en optimal karar doğrusunun ve hiperdüzlemin belirlenmesidir. Bazı durumlarda ise sınıflandırmanın doğrusal olarak gerçekleştirilebilmesi mümkün değildir. Eğitim verilerinin çoğunluğunun hiperdüzlemin diğer tarafında kalmasıyla yaşanan bu sorun, pozitif yapay değişken eklenerek çözümlenmektedir.

DVM'ler için optimum hiperdüzlemlerin olası konumları Şekil 3.1'de yer almaktadır. $H_1$ hiperdüzlemi iki sınıfı ayırma özelliğine sahip değilken, $H_2$ ve $H_3$ ile gösterilen hiperdüzlemler iki sınıfı ayırıcı özelliğe sahiptir. Sınıflar arasındaki mesafe dikkate alınarak bu iki hiperdüzlem incelendiğinde, $H_3$ hiperdüzlemi $H_2$'ye göre daha optimum görünmektedir.




**Şekil 3.1** Optimum hiperdüzlemlerin olası konumları [34]

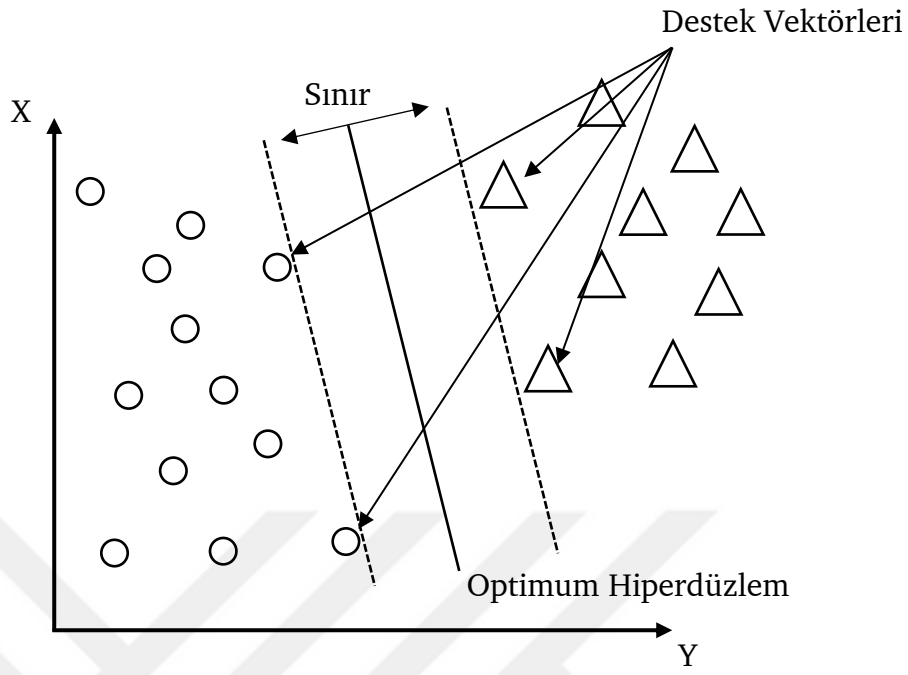
Doğrusal olarak ayrılabilen ve doğrusal olarak ayrılamayan veriler için DVM'ler iki farklı şekilde incelenmektedir.

Ele alınan sınıflandırma probleminde veriler doğrusal şekilde ayrılabiliyorsa, DVM'nin öğrenmesi için $k$ örnek yer alan eğitim veri setinin $\{x_i, y_i\}, i = 1,2,3, \dots, k$ şeklinde olduğu kabul edilerek optimum hiperdüzlem eşitsizlikleri Denklem 3.5'teki gibi yazılmaktadır.

$$\begin{cases} w\,x_i + b \geq 1, & y = 1 \\ w\,x_i + b \leq 1, & y = -1 \end{cases} \tag{3.5}$$

Burada, $w$ ağırlık vektörünü (hiperdüzlemin normalini), $x \in R^N$ olup N boyutlu uzayı, $y \in \{-1, +1\}$ sınıf etiketlerini, $b$ değeri ise eğilim değerini simgelemektedir. Optimum hiperdüzlemin ortaya koyulabilmesi için bu hiperdüzleme paralel iki hiperdüzlem belirlenmesi gerekmektedir, bu paralel hiperdüzlemleri belirleyen noktalara ise destek vektörler olarak isimlendirilir. Bu düzlemler, $w\,x_i + b = \pm 1$

şeklinde ifade edilmektedirler. Şekil 3.2'de optimum hiperdüzlem ve destek vektörlerine yer verilmiştir [33].




**Şekil 3.2** Doğrusal olarak ayrılabilen veri setleri için hiperdüzlem [33]

Optimum hiperdüzlemin belirlenmesi destek vektörlerin sınırının maksimum değer alması ve *w* değerinin minimize edilmesi gerekmektedir. Bu durum Denklem 3.6'daki kısıtlı optimizasyon probleminin çözümü ile mümkündür.

$$min\left[\frac{1}{2}\|w\|^2\right] \tag{3.6}$$

Ele alınan problemin sınıflandırması bazen doğrusal olarak mümkün olmayabilir, bu durumda eğitim verilerinin çoğunluğunun optimum hiperdüzlemin diğer tarafında kalması sorunu yaşanmaktadır. Bu sorun, pozitif yapay değişken $\xi_i$'nin eklenmesiyle giderilmektedir. Hiperparametre sınırının maksimize edilmesi ile yanlış sınıflandırmaların minimize edilmesi arasındaki denge, $(0,\infty)$ aralığındaki bir düzenleme parametresi olan ve $C$ parametresi ile kontrol edilmektedir [33]. Bu tarz verilere doğrusal ayrım yapılamayan durumlar için optimizasyon problemi Denklem 3.7'de verilmektedir.

$$min\left[\frac{1}{2}\|w\|^2 + C\sum_{i=1}^{r}\xi_i\right] \tag{3.7}$$

DVM'ler Denklem 3.8'de matematiksel ifadesi yer alan çekirdek fonksiyonu yardımıyla doğrusal olmayan dönüşümler yapabilmektedir.

$$K(x_i, x_j) = \varphi(x).\varphi(x_j) \tag{3.8}$$

Çekirdek fonksiyonu verilerin yüksek boyutta doğrusal olarak ayrılmasını sağlamaktadır. Böylece, çekirdek fonksiyonu yardımıyla doğrusal olarak ayrılmayan iki sınıflı bir problemin çözümü için karar kural Denklem 3.9'da verilmiştir [33].

$$f(x) = sign(\sum_i \alpha_i y_i \varphi(x).\varphi(x_i) + b) \tag{3.9}$$

Literatürde en sık kullanılan çekirdek fonksiyonları ve formülleri parametreleri ile birlikte Tablo 3.1'de verilmiştir [33].

**Tablo 3.1** DVM'ler için temel çekirdek fonksiyonları ve formülleri

| **Çekirdek fonksiyonu** | **Matematiksel ifadesi** | **Parametre** |
|---|---|---|
| Polinom | $K(x,y) = ((x.y)+1)^d$ | Polinom derecesi (d) |
| Normalleştirilmiş polinom | $K(x,y) = \frac{((x.y)+1)^d}{\sqrt{((x.x)+1)^d((y.y)+1)^d}}$ | Polinom derecesi (d) |
| Radyal tabanlı fonksiyon | $K(x,y) = e^{-\gamma\|(x-x_i)\|^2}$ | Çekirdek boyutu ($\gamma$) |
| Pearson VII | $\frac{1}{\left[1+\left(\frac{2.\sqrt{\|x-y_i\|^2\sqrt{2^{(1/\omega)}-1}}}{\sigma}\right)^2\right]^\omega}$ | Pearson genişliği parametreleri ($\sigma, \omega$) |

Tablo 3.1'de de görüleceği üzere, çekirdek fonksiyonlarının düzenleme parametrelerinin modeli kuran kişi tarafından belirlenmesi gerekmektedir. Bu parametrelerin aldığı değer olması gerekenden çok küçük veya büyük olması durumunda optimum hiperdüzlem doğru bir şekilde belirlenemez ve sınıflandırma algoritmasının DO düşer. Çekirdek fonksiyonları karşılaştırıldığında, polinom ve

radyal tabanlı çekirdek fonksiyonlarının daha sade ve anlaşılır olduğu ifade edilebilir [33]. Bu çalışmada DVM için radyal tabanlı fonksiyon (RBF) kullanılacaktır.

### 3.4.4 Rasgele Orman (RO)

Leo Breiman tarafından 2001 yılında geliştirilen bu algoritma, ağaç tipi sınıflandırıcılar topluluğudur. Ağaç sayısı arttıkça elde edilecek sonucun kesinliği de artmaktadır. Ancak RO algoritmasında türetilen ağaç sayısı arttırıldığında model performansının yükselmesi beklenmekle birlikte ağaç sayısının arttırılmasının her zaman daha yüksek performans garantisi vermemektedir [34]. Karar ağacı algoritması ile arasındaki en temel fark, rasgele orman algoritmasında karar düğümlerini bulma ve bitiş düğümlerini bölme işlemlerinin tamamen rasgele gerçekleştiriliyor olmasıdır. Breiman bu yöntemi makalesinde açıklarken şöyle bahsetmiştir; $k$'ıncı ağaç için geçmiş rasgele $Q_1, Q_2, \dots, Q_{k-1}$ vektörlerinden bağımsız fakat aynı dağılıma sahip $Q_k$ rasgele vektörü oluşturulur. $Q_k$ kullanılarak $x$'in giriş vektörü olduğu $h(x, Q_k)$ sınıflandırıcısının sonucu olarak bir ağaç oluşturulur [35]. RO karar ağaçları için tanımlanmış olmasına rağmen tüm sınıflandırıcılar için geçerli bir yaklaşımdır. En önemli avantajlarından biri çok fazla girdi değişkeni ele alabilmesidir [36].

### 3.4.5 Gradyan Arttırma Makineleri (GAM)

Arttırma algoritmalarının temel mantığı güçlü bir öğrenici oluşturmak için zayıf öğrenicileri bir araya getirerek yükseltme yöntemi kullanmaktır. Temel öğrenici olarak regresyon ağaçları seriler halinde her biri bir öncekinin hataları üzerine kurulu devam etmektedir. Makine öğrenmesinde sıklıkla kullanılan bir topluluk algoritması olan GAM, veri madenciliği ve makine öğrenmesinde oldukça iyi performanslar elde etmektedir [38]. Bu yöntemde ilk olarak, gerçekleştirilen ilk iterasyonda bir tahmin fonksiyonu oluşturulmaktadır, daha sonrasında tahminler ve elde edilen gözlemler arasındaki farkların hesaplanması ile elde edilen bir kayıp fonksiyonu meydana gelmektedir. İkinci iterasyona geçildiğinde, tahmin ve kayıp fonksiyonları birleştirilerek ilk adımdaki ile aynı şekilde tekrar tahminler ve gözlemler arasındaki farklar hesaplanmaktadır. Birbiri üzerine kurulu bir düzende devam eden bu süreçte tahmin fonksiyonunun performansının iyileştirilmesi

amaçlanmaktadır. Amaç tahminler ile gözlemler arasındaki fark ile tanımlanan kayıp fonksiyonunun sıfıra mümkün olduğunca yaklaştırılmasıdır. Kayıp fonksiyonu için Denklem 3.10'da, gözlenen değer için Denklem 3.11'de ve tahmin değeri için Denklem 3.12'de formüllerine yer verilmiştir.

$$L(y_i, y_i^p) = \frac{1}{N}\sum(y_i - y_i^p)^2 \tag{3.10}$$

$$y_i^p = y_i^p + \alpha * \delta \sum \frac{(y_i - y_i^p)^2}{\delta y_i^p} \tag{3.11}$$

$$y_i^p = y_i^p - \alpha * 2 * \sum(y_i - y_i^p) \tag{3.12}$$

### 3.4.6 Uyarlanabilir Arttırma (AdaBoost)

AdaBoost algoritması ilk boosting algoritması olarak kabul edilmektedir. Bu yöntem ilk olarak eğitim setinin önce bir zayıf öğrenici ile eğitilmesi ile başlar. Eğitim sonrası yanlış tahmin edilen örnekler bu algoritma için önemlidir. Bir sonraki eğitimde ilk tahminlemede yanlış öğrenilen eğitim verilerine öncelik verilerek ağırlıkları arttırılarak tekrar eğitilir. Kümülatif bir yaklaşım olduğundan sıralama önem kazanmaktadır. Bu algoritmaya ait matematiksel yaklaşımları açıklamak için öncelikle her bir $x_i$ (i = 1, … , t) değişkeninin zayıf sınıflandırıcıları Denklem 3.13'ten yararlanarak tespit edilmektedir [38].

$$H(x) = sign(\sum_{t=1}^{T} \alpha_t h_t(x)) \tag{3.13}$$

Burada $h_t(x)$ zayıf sınıflandırıcı çıktısıdır ve $\alpha_t$ ise sınıflandırıcının ağırlığıdır. Denklem 3.14'te $E$ hata terimini içeren, $\alpha_t$'nin hesaplandığı denklem sunulmaktadır.

$$\alpha_t = \frac{1}{2} ln \frac{1-E}{E} \tag{3.14}$$

Zayıf sınıflandırıcıların eğitilmesinin ardından her bir eğitim verisinin ağırlıkları Denklem 3.15'ten yararlanılarak güncellenmektedir [38].

$$\alpha_{t+1}(i) = \frac{\alpha_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t} \tag{3.15}$$

### 3.4.7 Hafif Gradyan Arttırma Makineleri (HGAM)

Bu algoritma Microsoft DMTK (Distributed Machine Learning Toolkit) projesi kapsamında 2017 yılında geliştirilmiştir. Gradyan arttırma tekniklerinden biri

olan HGAM algoritması histogram tabanlı bir algoritma olup, sürekli değişkenleri kesikli hale getirerek hesaplama maliyetini azaltmaktadır. Histogram tabanlı algoritmalar, optimum bölünme noktalarına karar verirken değişkenlerin histogramlarından yararlanmaktadır. HGAM ve XGBoost algoritmaları kıyaslanarak farklı veri tipindeki veri setlerinde karşılaştırılmıştır, gerçekleştirilen uygulamaların hepsinde hem tüketilen toplam hem de doğruluk oranları açısından en iyi sonucu veren algoritma olmuştur [39].

### 3.4.8 Ekstrem Gradyan Arttırma (XGBoost)

XGBoost, en etkin ve yaygın arttırma algoritmalarından biri olarak kabul edilmektedir. Yüksek tahmin gücü ve aşırı uyumu önleyen performans iyileştirmeleri içermektedir. Diğer arttırma algoritmalarına göre on kat daha hızlıdır. Çalışma mantığı GAM ile benzer olmakla birlikte ilk olarak bir tahmin skoru ile başlatılmaktadır. Yapılan tahmin ilerleyen adımlarda yakınsanarak doğru sonuca ulaşmaya yarayacaktır. Bu algoritmada yapılan en önemli gelişmelerden biri, algoritma çalıştırıldığında yalnızca eksik olmayan verilerin ziyaret ediliyor olmasıdır. Eksik değerleri tespit ederek ele almak ve bu değerleri işlemek için en iyi yönü öğrenmektedir [40].

### 3.4.9 Kategorik Arttırma (CatBoost)

Category ve Boosting kelimelerinden türetilen ismi ile bu algoritma, 2017 yılında Yandex firması tarafından geliştirilmiş ve açık kaynak kodları ile yayımlanmıştır. Diğer arttırımlı algoritmalardan en büyük farkı veri ön işleme evresini kısaltması, boş verilerle başa çıkabilir ve kategorik verilere kodlama yapabilir olmasıdır. Derin ağaçlar kurmak yerine simetrik ağaçlar kurarak yüksek tahmin oranı yakalarken bir yandan da aşırı öğrenme sorununu ortadan kaldırmaktadır.

Sıralı boosting ve kategorik değişkenlerin işlenmesine yönelik geliştirilen yapısı, CatBoost ile sunulan iki kritik ilerleme olarak kabul edilmektedir. Mevcutta bulunan gradyan arttırma algoritmalarında yaşanan tahmin kayması sorunlarıyla mücadele etmek ve hedefe ulaşmada yaşanan kayıpları gidermek adına geliştirilmiştir [41].

Tahmin kayması, bulunmak istenen değişkende tahmin edilen değerin veri setinde yer alan değere yaklaşıp gerçek değerden uzaklaşmasıyla gerçekleşmektedir. Tahmin edilen değer ile gözlenen değer arasındaki fark artık olarak tanımlanmaktadır. CatBoost algoritması temelde artıklarda kaymayı önlemeye çalışmaktadır [41].

### 3.4.10 K En Yakın Komşu (KNN)

1967 yılında ortaya konulan bu algoritma, eski, basit ve gürültülü veri kümelerine karşın etkili olması sebebiyle en popüler sınıflandırma algoritmalarından bir tanesidir. Sınıfları belli bir veri setinden yararlanılarak veri setine eklenecek yeni değişkenin mevcut verilere olan uzaklığı hesaplanılarak k tane en yakın komşuluğuna bakılmaktadır. Bu k tane en yakın komşu çoğunlukla hangi sınıfa dahil edilmişse yeni örnek de o sınıfa dahil edilecektir [42]. Veriler arasıdaki mesafe değerleri hesaplanırken genellikle Öklid ölçütü kullanılmaktadır. Bu ölçüte göre iki nokta arasındaki mesafe Denklem 3.16'daki gibi hesaplanmaktadır.

$$d = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \tag{3.16}$$

## 3.5 Başarı Metrikleri

Sınıflandırma algoritmalarında model başarılarının ölçümlenmesinde kullanılan çeşitli metrikler bulunmaktadır. Bu metrikler karmaşıklık matrisi içerisinde yer alan gerçek değerler ve tahmin edilen değerler arasındaki bağlantılardan formüle edilmektedir. Yani; Tablo 3.2'de verilen karmaşıklık matrisi, gerçek değerlerin bilinmekte olduğu bir grup test verisi üzerinde bir sınıflandırma modelinin performansını tanımlamak için kullanılmaktadır.

**Tablo 3.2** Karmaşıklık matrisi

| | | Gerçek Değerler | |
|---|---|---|---|
| | | Pozitif Tahminler | Negatif Tahminler |
| Tahmin Değerler | Pozitif Tahminler | GP | YP |
| | Negatif Tahminler | YN | GN |

Doğruluk Oranı (DO): Doğru tahminlenen örnek sayısının (GP+GN), toplam örnek sayısına (GP+GN+YP+YN) oranıdır.

$$DO = \frac{GP+GN}{GP+GN+YP+YN} \tag{3.17}$$

Hata Oranı (HO): Doğruluk oranını 1'e tamamlayandır, veya yanlış tahmin edilmiş (YN+YP) örnek sayısının toplam örnek sayısına (GP+GN+YP+YN) oranıdır.

$$HO = \frac{YN+YP}{GP+GN+YP+YN} \tag{3.18}$$

Kesinlik Oranı (KO): Doğru olan örneğin sınıfını doğru olarak tahmin edilmiş (GP) örnek sayısının, sınıfı doğru olarak tahmin edilmiş tüm örnek sayısına (GP+YP) oranıdır.

$$KO = \frac{GP}{GP+YP} \tag{3.19}$$

Duyarlılık Oranı (DUO): Doğru olan örneğin sınıfını doğru olarak tahmin edilmiş (GP) gerçekte doğru olan toplam pozitif örnek sayısına (GP+YN) oranıdır.

$$DUO = \frac{GP}{GP+YN} \tag{3.20}$$

F1 skor: KO ve DUO metriklerinin harmonik ortalamasıdır ve sınıflandırıcının ne kadar iyi performans gösterdiğinin bir ölçüsüdür.

$$F1 = 2 * \frac{KO*DUO}{KO+DUO} \tag{3.21}$$

Kappa: Gözlenen ve beklenen değerler arasındaki uyumu gösteren metriktir, yani gerçek değerlerle tahmin edilen değerler arasındaki uyum olarak da ifade edilmektedir. -1 ile +1 arasında değer alabilen Kappa değeri 1'e yaklaştıkça belirtilen değerler arasındaki uyumun güçlendiğini göstermektedir.

## 3.6 Öznitelik Seçim Yöntemleri

Araştırma kapsamında ele alınan 30 öznitelik, öznitelik seçim yöntemlerinden yararlanılarak en iyi açıklayıcılığa sahip öznitelikler belirlenmiştir. Belirlenen öznitelikler daha sonra veri modelleme aşamasında alternatif veri modelleri oluşturulurken yararlanılmak üzere kullanılmıştır.

### 3.6.1 Feature Importance Fonksiyonu

Bu yöntem; bir hedef değişkeni tahmin etme başarısına göre girdi özniteliklerine puan atama tekniğidir. Veri kümesine ilişkin içgörü sağlamak için kullanılan bu teknik aracılığıyla hangi özniteliklerin hedefle en alakalı, hangi özniteliklerin hedefe en az alakalı olduğu saptanmaktadır. Bu fonksiyon içerisinde; XGBoost, HGAM, CatBoost, AdaBoost, GAM, LR ve RO algoritmaları kullanılarak özniteliklere ait skorlar elde edilmiştir [43].

### 3.6.2 Mutual Info Yöntemi

Mutual Info, aynı verinin iki etiketi arasındaki benzerliğin bir ölçüsüdür. Bir diğer deyişle, gözlemlenen birlikte gerçekleşme sıklığının beklediğimizden ne ölçüde farklı olduğunu ifade etmektedir. Bir örneklem içerisinde çifti birlikte gözlemleme sayısına karşı çifti ayrı ayrı görme sayısına bağlı olarak hesaplanır [44].

David J. C. Mackay'in kısa tanımına göre ise; bir değişkenin değerinin öğrenilmesinden kaynaklanan bir diğer değişken hakkındaki belirsizlikteki ortalama azalmayı ölçer, veya tam tersi bir değişkenin bir diğer değişken hakkında ilettiği ortalama bilgi miktarını ölçmektedir [45].

### 3.6.3 Shap Fonksiyonu

1953 yılında Lloyd Shapley tarafından öne sürülen bu yaklaşım, orijinalinde işbirlikçi oyunlar için geliştirilen Shapley Değerleri ile doğrudan ilişkilidir. Çok sayıda olası sıralamalar üzerinden bir değişkenin nitelik değerinin ortalamasını alma fikrine dayanmaktadır. Her bir özniteliğin tahmine katkısını hesaplayarak herhangi bir modelin tahminini açıklamak için kullanılabilmektedir [46]. Bu fonksiyon içerisinde; XGBoost, HGAM ve CatBoost algoritmaları kullanılarak özniteliklere ait skorlar elde edilmiştir.

### 3.6.4 Borda Yöntemi

Borda kuralı veya bir diğer deyişle borda sayımı, 1770 yılında Fransız matematikçi ve siyaset bilimci Jean-Charles de Borda tarafından geliştirimiş, tek kazananın olduğu bir seçim yöntemidir. Bu yönteme göre, tüm alternatifler çiftler halinde oylanarak, her bir alternatifin diğer alternatife karşı aldığı oy bir matris üzerinde

gösterilir. Bu ikili karşılaştırmalar sonucunda en çok oyu alan alternatif kazanır [47]. Uygulanabilirliği oldukça kolay olan bu yöntem, her bir alternatifi ayni önem seviyesinde tutması açısından da iyi bir performans sunmaktadır. Tüm bunlara ilave olarak Borda sayım yöntemi iki veya daha fazla sıralama tekniğini daha rasyonel tek bir sıralama haline getiren veri birleştirme yöntemlerinden biridir [48].

Yöntemin uygulanma şeklinden kısaca bahsetmek gerekirse; n farklı alternatifin bulunduğu bir sınıflandırma durumunda en az tercih edilen alternatife 0 puan olmak üzere, bir sonrakine 1 puan ve en çok tercih edilene (n-1) puan verilmektedir. Daha sonra alternatifler borda skorlarına göre küçükten büyüğe doğru sıralanmaktadırlar.

Bu araştırmada, veri setinde yer alan 30 öznitelikle beraber veri setine uygulanan ve yukarıda açıklanan her bir öznitelik seçim algoritması için borda sayım tekniğinden yararlanılarak sıralama yapılmıştır.

## 3.7 Veri Seti Bölme

Makine öğrenmesi algoritmaları kullanılırken, veri seti eğitim ve test olmak üzere ikiye ayrılır. Algoritmalar eğitim veri seti ile eğitildikten sonra test veri seti ile test gerçekleştirilmektedir. Bu bölme işlemi modelleme aşamasında belirli oranlara göre yapılabilmektedir (%50 eğitim - %50 test, %60 eğitim - %40 test, %70 eğitim - %30 test, %80 eğitim - %20 test, %90 eğitim - %10 test vb.). Ayrıca k katlı çapraz doğrulama yöntemi de kullanılabilir. K katlı çapraz doğrulama yönteminde veri seti k parçaya ayrılmaktadır. k-1 parça ile eğitim, 1 parça ile test gerçekleştirilmektedir. Ardından bu işlem her parçayı test verisi olarak kullanmak şartıyla sürdürülmektedir. Tüm test verilerinin değerlendirilmesi ile algoritma performansı elde edilmektedir.

# 4
# UYGULAMA

Bu bölümde çalışmanın veri toplama, veri ön işleme ve veri setinin modellenmesi aşamalarından bahsedilecek, ardından modeller üzerinden gerçekleştirilen analizler ve sonuçları yorumlamalarıyla birlikte ele alınarak modellerin gösterdiği başarılar doğruluk oranları ve diğer metrikler üzerinden değerlendirilecektir. Başarı sonuçlarına göre elde edilen en iyi sınıflandırma sonucunu veren model ve elde edilen bulgular detaylı olarak değerlendirilecektir.

Bu tez çalışmasında gerçekleştirilen sınıflandırma işlemleri ve diğer program tabanlı işlemler Phyton 3.8.5 programında gerçekleştirilmiştir.

Çalışmalar sonucunda elde edilen sınıflandırma sonuçlarına göre temel değerlendirmeler doğruluk oranları baz alınarak gerçekleştirilirken, diğer metrikler de araştırma boyunca yorumlanmıştır. Tüm veri seti modelleri için sınıflandırma sonuçları, sonuç karşılaştırma matrisleri ve doğruluk oranları grafikleri üzerinden değerlendirilmiştir.

## 4.1 Veri Seti

Bu araştırma kapsamında veriler, İstanbul'da ikamet eden on sekiz yaşından büyük bireylerin dijital bankacılık uygulamalarına karşın tutumları demografik yapı, sosyoekonomik seviye, eğitim durumu, kredi kullanım alışkanlığı, risk algısı gibi çeşitli faktörlere yanıt verecek bir çalışma aracılığıyla elde edilmiştir. Yanıtları çoktan seçmeli veya açık uçlu olmak üzere çalışma toplam 30 sorudan oluşmaktadır.

Hazırlanan çalışma Yıldız Teknik Üniversitesi Sosyal ve Beşerî Bilimler Araştırmaları Etik Kurulu tarafından değerlendirilmiş, etiğe aykırı bir durum bulunmamıştır (Toplantı tarihi: 25.01.2022, Toplantı no: 2022.01).

Çalışmada yer alan sorular ve değişken türlerine göre yanıtlar Tablo 4.1'de yer almaktadır. Ayrıca uygulanan çalışma sorularına ve yanıtlara Ek-A'da detaylı olarak yer verilmiştir.

Araştırma kapsamında kullanılan sorular hazırlanırken, Kim ve diğerlerinin 2008 yılında gerçekleştirmiş oldukları e-ticarette güvene dayalı tüketici karar verme modelleri ile ilgili araştırmaları [49], Lee'nin 2009 yılında gerçekleştirmiş olduğu internet bankacılığının benimsenmesini etkileyen faktörlerle ilgili araştırması [50] ve Ryu'nun 2017 yılında kullanıcıları fintek kullanmaya istekli veya çekimser kılan faktörleri [51] araştırdığı çalışmalarından yararlanılmıştır.

**Tablo 4.1** Araştırma soruları ve değişken türlerine göre yanıtlar

| **Soru** | **Yanıt / Değişken** |
|---|---|
| Yaş | Sürekli Değişken |
| Cinsiyet | Kategorik Değişken |
| Eğitim durumu | Kategorik Değişken |
| Medeni hal | Kategorik Değişken |
| Çocuk sayısı | Kategorik Değişken |
| Bir günde ekran başında geçirilen süre | Sürekli Değişken |
| Cep telefonunda aktif kullanılan uygulama sayısı | Kategorik Değişken |
| Çalışma durumu | Kategorik Değişken |
| Meslek | Kategorik Değişken |
| Aylık net gelir | Sürekli Değişken |
| Bireysel emeklilik sözleşmesi | Kategorik Değişken |
| İkamet edilen ilçe | Kategorik Değişken |
| Konut sahipliği | Kategorik Değişken |
| İkamet edilen konut tipi | Kategorik Değişken |
| İkamet edilen konutun en yakın AVM uzaklığı | Kategorik Değişken |
| Kira geliri var/yok | Kategorik Değişken |
| Şahsi araç | Kategorik Değişken |
| Araç kasko bedeli | Kategorik Değişken |
| İkamet edilen konutun müşterisi olunan banka ATM uzaklığı | Kategorik Değişken |
| Yatırım tercihleri | Kategorik Değişken |

**Tablo 4.1** Araştırma soruları ve değişken türlerine göre yanıtlar (devamı)

| | |
|---|---|
| Kredi yaklaşımı | Kategorik Değişken |
| Mevcutta ödenmekte olan kredi sayısı | Kesikli Değişken |
| Aylık birikim/gelir oranı | Sürekli Değişken |
| Müşterisi olunan banka sayısı | Kesikli Değişken |
| Geleneksel/dijital bankacılık tercihi | Kategorik Değişken |
| Sık kullanılan bankacılık işlemleri | Kategorik Değişken |
| Kredi kartı sahipliği | Kategorik Değişken |
| Kripto para yatırımı | Kategorik Değişken |
| Mobil veya internet bankacılığı kullanımı | Kategorik Değişken |
| Dijital bankacılık kullanım düzeyi | Kategorik Değişken |

Bu çalışmada veri toplama süreci 01.02.2022 ve 15.03.2022 tarihleri arasında gerçekleştirilmiş olup, çalışma dijital bir platform üzerinden online olarak gerçekleştirilerek 372 katılımcı tarafından veri sağlanmıştır. Hedef kitle olarak belirlenen gruba ikamet ettiği şehirlere göre sorular sorulmuştur. Çalışma içerisinde yer alan her bir soru araştırma kapsamında bir öznitelik olarak değerlendirilecek olup, kullanıcıların bu özniteliklere karşılık gelen yanıtları analiz edilecektir. Özetle, 30 öznitelik ve 372 örnekten oluşan bir veri seti ile çalışma gerçekleştirilecektir. Elde edilen veriler makine öğrenmesi algoritmalarının dijital teknoloji kullanım alanında bir uygulama olarak kullanılacaktır. Tablo 4.1'de yer alan değişkenlerden dijital bankacılık kullanım düzeyi tahmin edilecek özniteliktir. Bu öznitelik bağımlı değişken olarak da düşünülebilir ve kategorileri "hiç kullanmıyorum", "haftada dört defa veya daha az kullanıyorum", "günde 1 defa veya daha fazla kullanıyorum" şeklinde 3 kategoriden oluşmaktadır. Bu kategoriler algoritmalara yönelik sonuçlarda sırasıyla 0, 1 ve 2 şeklinde gösterilmiştir. Diğer tüm öznitelikler bağımsızdır.

## 4.2 Veri Ön İşleme

Veri ön işleme aşamasında, katılımcılardan elde edilen değerlendirmelerin tamamı incelenerek eksik veya hatalı veriler içeren örnekler veri setinden çıkartılmıştır. Ardından sürekli değişken türündeki yanıtlar sınıflandırılarak

kategorik değişkenler haline getirilmiştir. 372 örnekten oluşan örneklem, veri ön işleme aşamasından sonra eksik ve hatalı verilerin elenmesiyle 300 örneğe düşürülmüş ve araştırma bu 300 veri ile gerçekleştirilmiştir.

## 4.3 Veri Setine Ait Tanımlayıcı İstatistikler

Çalışmada ele alınan veri setine ait 300 örnek için tanımlayıcı istatistiklere Tablo 4.2'de yer verilmiştir. Gerçekleştirilen çalışmada katılımcıların %56 oranında çoğunlukla kadın bireylerden oluştuğu, eğitim düzeyinin %60 oranla lisans seviyesinde olduğu, katılımcıların %60'ının bekar olduğu, %73'ünün ise hiç çocuğu olmadığı, %63 oranı ile çoğunluğun özel sektörde çalışmakta olduğu ve %56 oranında katılımcıların İstanbul Avrupa yakasında ikamet etmekte oldukları söylenebilir.

**Tablo 4.2** Veri setine ait tanımlayıcı istatistikler

| Tanımlayıcı İstatistik | Gruplar | Frekans | % |
|---|---|---|---|
| **Cinsiyet** | Kadın | 169 | %56 |
| | Erkek | 131 | %44 |
| **Öğrenim Düzeyi** | İlk – orta öğrenim | 22 | %7 |
| | Lisans | 179 | %60 |
| | Yüksek lisans veya doktora | 99 | %33 |
| **Medeni Hal** | Evli | 121 | %40 |
| | Bekar | 179 | %60 |
| **Çocuk Durumu** | Çocuğum var | 219 | %27 |
| | Çocuğum yok | 81 | %73 |
| **Çalışma Durumu** | Özel sektör | 188 | %63 |
| | Kamu | 53 | %18 |
| | Kendi işinin sahibi | 15 | %5 |
| | Çalışmıyor | 44 | %15 |
| **Yaşadığı İlçe** | İstanbul Avrupa | 167 | %56 |
| | İstanbul Anadolu | 133 | %44 |

Çalışmada yer alan finansal teknoloji kullanımı, kredi ve risk yaklaşımını ele alan sorular için elde edilen yanıtlara ait tanımlayıcı istatistiklere Tablo 4.3'te yer verilmiştir.

**Tablo 4.3** Finansal teknoloji ve kredi risk yaklaşımları için tanımlayıcı istatistikler

| Tanımlayıcı İstatistik | Gruplar | Frekans | % |
|---|---|---|---|
| Müşterisi olduğu banka sayısı | 1 | 45 | %15 |
| | 2 | 101 | %34 |
| | 3 | 82 | %27 |
| | 4 | 24 | %8 |
| | 5 veya daha fazla | 48 | %16 |
| Bankacılık işlemleri için tercih edilen kanal | Banka | 13 | %4 |
| | ATM | 29 | %10 |
| | İnternet veya mobil şube | 258 | %86 |
| Kredi kartı sahipliği | Var | 243 | %81 |
| | Yok | 57 | %19 |
| Kripto para yatırımı | Var | 95 | %32 |
| | Yok | 205 | %68 |
| Yatırım risk algısı | Yüksek getiri için yüksek riskli yatırımlar tercih eder | 23 | %8 |
| | Orta riskli yatırımlarla dengeli gelir elde etmeyi tercih eder | 109 | %37 |
| | Risk almayı sevmez | 105 | %35 |
| | Yatırım yapmaz | 62 | %21 |
| Mevcutta ödenen kredi sayısı | Kredim yok | 148 | %49 |
| | 1 | 87 | %29 |
| | 2 | 32 | %11 |
| | 3 veya daha fazla | 33 | %11 |

**Tablo 4.3** Finansal teknoloji ve kredi risk yaklaşımları için tanımlayıcı istatistikler (devamı)

| Tanımlayıcı İstatistik | Gruplar | Frekans | % |
|---|---|---|---|
| **Mobil veya internet bankacılığı kullanımı** | Sadece mobil | 103 | %34 |
| | Sadece internet | 14 | %5 |
| | Her ikisini de kullanan | 160 | %53 |
| | Her ikisini de kullanmayan | 23 | %8 |
| **Dijital bankacılık kullanım düzeyi** | Hiç kullanmıyorum | 23 | %8 |
| | Haftada 4 defa veya daha az | 140 | %47 |
| | Günde 1 defa veya daha fazla | 137 | %46 |

Tablo 4.3'te yer alan tanımlayıcı istatistiklere göre; katılımcıların %34'ü 2 farklı banka müşterisiyken, %86'sı bankacılık işlemlerini mobil veya internet şubeden gerçekleştirmektedir, araştırmada yer alan katılımcılar arasında işlemlerini bankaya giderek gerçekleştiren kullanıcı oranı %4'tür. Ayrıca, katılımcıların %81'i en az bir kredi kartına sahiptir, katılımcıların %32'si kripto para yatırımı yaparken %68 oranındaki çoğunluk kripto para yatırımına sahip değildir. Katılımcıların %37 oranındaki çoğunluğu orta riskli yatırımlarla dengeli gelir elde etmeyi tercih etmekteyken %35 oranında ise hiç risk almayan kullanıcılar gözlenmektedir. Tablo 4.3'e göre katılımcıların %49'unun herhangi aktif bir kredi borcu bulunmamaktadır. %53 oranıyla katılımcıların en çok hem mobil hem internet şubeyi beraber kullanma eğiliminde oldukları, %34 oranındaki diğer çoğunluğun ise yalnızca mobil şube kullandıkları söylenebilir. Katılımcıların dijital bankacılık kanallarını kullanım düzeyi ise %47 oranında haftada 4 defa veya daha az kullanmakta oldukları, %46 oranında ise günde en az bir defa kullandıkları araştırma sonucunda ortaya çıkan istatistiklerle elde edilmiştir.

## 4.4 Veri Modelleme

Veri seti modellerine göre öznitelik kullanımı, eğitim seti ve test seti ayrımları Tablo 4.4'te verilmektedir. 300 örnek ve 30 öznitelikten oluşan veri seti 15 farklı modele göre sınıflandırılmış ve tüm sonuçlar incelenmiştir.

Veri setinde yer alan 300 örnek için öncelikle tüm özniteliklerle gerçekleştirilen sınıflandırma sonuçları incelenmiştir. Buna ek olarak öznitelik seçimi yapılmış ve en yüksek açıklayıcılığa sahip ilk 10 öznitelik ve ilk 15 öznitelik belirlenerek, bu öznitelikler için de aynı sınıflandırma çalışmaları gerçekleştirilmiştir.

En iyi açıklayıcılığa sahip öznitelikler Borda Sayım Yöntemi kullanılarak belirlenmiştir. Borda yöntemi birden fazla sınıflayıcı tarafından oluşturulmuş sıralamaları birleştirip tek bir sıralama sunmaktadır [52].

**Tablo 4.4** Veri seti modelleri

| Veri Seti Modeli | Öznitelik Kullanımı | Eğitim Seti | Test Seti |
|---|---|---|---|
| **Model 1.1** | Tüm Öznitelikler | %50 | %50 |
| **Model 1.2** | | %60 | %40 |
| **Model 1.3** | | %70 | %30 |
| **Model 1.4** | | %80 | %20 |
| **Model 1.5** | | %90 | %10 |
| **Model 2.1** | Açıklayıcılığı en yüksek ilk 15 öznitelik | %50 | %50 |
| **Model 2.2** | | %60 | %40 |
| **Model 2.3** | | %70 | %30 |
| **Model 2.4** | | %80 | %20 |
| **Model 2.5** | | %90 | %10 |
| **Model 3.1** | Açıklayıcılığı en yüksek ilk 10 öznitelik | %50 | %50 |
| **Model 3.2** | | %60 | %40 |
| **Model 3.3** | | %70 | %30 |
| **Model 3.4** | | %80 | %20 |
| **Model 3.5** | | %90 | %10 |

## 4.5 Tüm Veri Seti Modelleri için Uygulama Sonuçları

Bu kısımda Tablo 4.4'te yer alan veri seti modelleri için uygulanan sınıflandırma algoritmaları ve uygulama çalışmaları sonucunda elde edilen tüm metriklerin yer aldığı sonuç tablolarına yer verilmiştir. Veri seti modellerine uygulanan

algoritmalar; XGBoost, HGAM, CatBoost, GAM, AdaBoost, KNN, KA, RO, LR ve DVM RBF Çekirdek şeklindedir. Her bir model için gerçekleştirilen sınıflandırma sonuçları, karşılaştırma matrisi ve doğruluk oranları grafiği yorumlanarak bu bölümde yer almıştır.

### 4.5.1 Tüm Özniteliklerle Gerçekleştirilen Sınıflandırma Sonuçları

Bu bölümde, herhangi bir öznitelik seçim algoritması kullanılmadan, tüm özniteliklerin eğitim ve test gruplarına ayrılarak gerçekleştirilen sınıflandırma sonuçlarına yer verilmiştir. Sınıflandırma sonuçları tablolarında Bölüm 3.5'te açıklanan KO, DUO, F1 Skor, DO ve Kappa metriklerinin değerleri yer almaktadır. Ayrıca bu metriklere yönelik makro ortalama (MO) ve örnek sayısı (ÖS) değerlerinin dikkate alınması ile hesaplanan ağırlıklı ortalama (AO) değerleri yer almaktadır.

**Tablo 4.5** Model 1.1 için gerçekleştirilen sınıflandırma sonuçları

| Algoritma | Sınıf | KO | DUO | F1 Skor | ÖS | DO | Kappa |
|---|---|---|---|---|---|---|---|
| **XGBoost** | 0 | 0,92 | 1,00 | 0,96 | 11 | 0,740 | 0,542 |
| | 1 | 0,71 | 0,74 | 0,72 | 69 | | |
| | 2 | 0,74 | 0,70 | 0,72 | 70 | | |
| MO | | 0,79 | 0,81 | 0,80 | 150 | | |
| AO | | 0,74 | 0,74 | 0,74 | 150 | | |
| **HGAM** | 0 | 1,00 | 1,00 | 1,00 | 11 | 0,773* | 0,599* |
| | 1 | 0,72 | 0,83 | 0,77 | 69 | | |
| | 2 | 0,80 | 0,69 | 0,74 | 70 | | |
| MO | | 0,84 | 0,84 | 0,84 | 150 | | |
| AO | | 0,78 | 0,77 | 0,77 | 150 | | |
| **CatBoost** | 0 | 1,00 | 0,91 | 0,95 | 11 | 0,767 | 0,586 |
| | 1 | 0,71 | 0,83 | 0,77 | 69 | | |
| | 2 | 0,8 | 0,69 | 0,74 | 70 | | |
| MO | | 0,84 | 0,81 | 0,82 | 150 | | |
| AO | | 0,77 | 0,77 | 0,77 | 150 | | |

**Tablo 4.5** Model 1.1 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **GAM** | 0 | 0,92 | 1,00 | 0,96 | 11 | 0,720 | 0,507 |
| | 1 | 0,67 | 0,78 | 0,72 | 69 | | |
| | 2 | 0,75 | 0,61 | 0,68 | 70 | | |
| MO | | 0,78 | 0,80 | 0,78 | 150 | | |
| AO | | 0,73 | 0,72 | 0,72 | 150 | | |
| **AdaBoost** | 0 | 1,00 | 1,00 | 1,00 | 11 | 0,633 | 0,352 |
| | 1 | 0,59 | 0,67 | 0,63 | 69 | | |
| | 2 | 0,62 | 0,54 | 0,58 | 70 | | |
| MO | | 0,74 | 0,74 | 0,74 | 150 | | |
| AO | | 0,64 | 0,63 | 0,63 | 150 | | |
| **KNN** | 0 | 0,2 | 0,09 | 0,13 | 11 | 0,493 | 0,079 |
| | 1 | 0,50 | 0,57 | 0,53 | 69 | | |
| | 2 | 0,51 | 0,49 | 0,5 | 70 | | |
| MO | | 0,40 | 0,38 | 0,38 | 150 | | |
| AO | | 0,48 | 0,49 | 0,48 | 150 | | |
| **KA** | 0 | 1,00 | 1,00 | 1,00 | 11 | 0,713 | 0,493 |
| | 1 | 0,68 | 0,72 | 0,70 | 69 | | |
| | 2 | 0,71 | 0,66 | 0,68 | 70 | | |
| MO | | 0,79 | 0,79 | 0,79 | 150 | | |
| AO | | 0,71 | 0,71 | 0,71 | 150 | | |
| **RO** | 0 | 1,00 | 0,91 | 0,95 | 11 | 0,767 | 0,586 |
| | 1 | 0,71 | 0,83 | 0,77 | 69 | | |
| | 2 | 0,80 | 0,69 | 0,74 | 70 | | |
| MO | | 0,84 | 0,81 | 0,82 | 150 | | |
| AO | | 0,77 | 0,77 | 0,77 | 150 | | |

**Tablo 4.5** Model 1.1 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | |
|---|---|---|---|---|---|---|
| **LR** | 0 | 0,56 | 0,91 | 0,69 | 11 | |
| | 1 | 0,59 | 0,58 | 0,58 | 69 | 0,613 0,337 |
| | 2 | 0,66 | 0,60 | 0,63 | 70 | |
| MO | | 0,60 | 0,70 | 0,63 | 150 | |
| AO | | 0,62 | 0,61 | 0,61 | 150 | |
| **DVM RBF** | 0 | 0,00 | 0,00 | 0,00 | 11 | |
| | 1 | 0,46 | 0,94 | 0,62 | 69 | |
| | 2 | 0,50 | 0,07 | 0,12 | 70 | 0,467 0,012 |
| MO | | 0,32 | 0,34 | 0,25 | 150 | |
| AO | | 0,45 | 0,47 | 0,34 | 150 | |

* Algoritmaların DO ve kappa değerleri içerisinde en yüksek değere sahip olanı göstermektedir.

Tablo 4.5'te yer alan sonuçlar Model 1.1'e ait olup, veri setinin %50 eğitim ve %50 test olarak ayrılmasıyla elde edilmiştir. Yapılan sınıflandırmada doğruluk oranı değerlerine göre en iyi sınıflandırma HGAM algoritması tarafından gerçekleştirilmiştir. Tablo 4.5 incelendiğinde KNN ve DVM RBF Çekirdek algoritmaları ile yapılan sınıflandırmaların doğruluk oranı değerlerinin 0,5'in altında kalarak en kötü sınıflandırma performansını gerçekleştirmişlerdir. Kappa değerlerine göre Tablo 4.5 incelendiğinde, HGAM algoritması 0,599 ile en iyi performansı sergilemiştir. Buna ek olarak HGAM algoritması doğruluk oranı ve Kappa metrikleri dışındaki KO, DUO, F1 Skor metriklerine göre de en yüksek başarıyı göstermiştir.

**Tablo 4.6** Model 1.1'e ait sonuç karşılaştırma matrisi

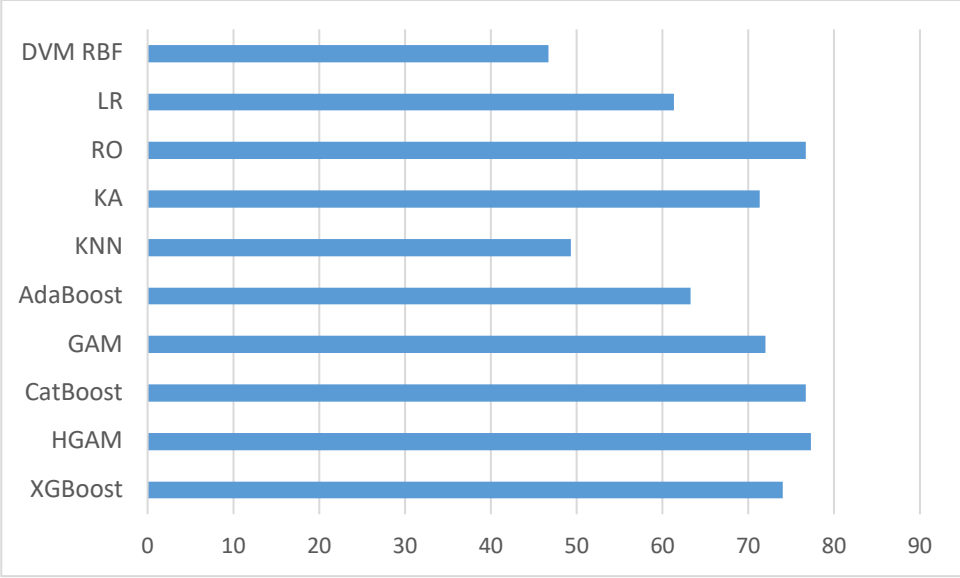
| **Algoritma** | | **XGB** | | | **HGAM** | | | **CatBoost** | | | **GAM** | | | **AdaBoost** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **TES →** | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| **GS** | 0 | 11 | 0 | 0 | 11 | 0 | 0 | 10 | 1 | 0 | 11 | 0 | 0 | 11 | 0 | 0 |
| | 1 | 1 | 51 | 17 | 0 | 57 | 12 | 0 | 57 | 12 | 1 | 54 | 14 | 0 | 46 | 23 |
| | 2 | 0 | 21 | 49 | 0 | 22 | 48 | 0 | 22 | 48 | 0 | 27 | 43 | 0 | 32 | 38 |

**Tablo 4.6** Model 1.1'e ait sonuç karşılaştırma matrisi (devamı)

| Algoritma | | KNN | | | KA | | | RO | | | LR | | | DVM RBF | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TES → | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| GS | 0 | 1 | 6 | 4 | 11 | 0 | 0 | 10 | 1 | 0 | 10 | 1 | 0 | 0 | 10 | 1 |
| | 1 | 1 | 39 | 29 | 0 | 50 | 19 | 0 | 57 | 12 | 7 | 40 | 22 | 0 | 65 | 4 |
| | 2 | 3 | 33 | 34 | 0 | 24 | 46 | 0 | 22 | 48 | 1 | 27 | 42 | 0 | 65 | 5 |

TES: Tahmin edilen sınıf, GS: Gerçek sınıf

Tablo 4.6'da yer alan Model 1.1'e ait sonuç karşılaştırma matrisi incelendiğinde HGAM algoritmasının toplam 150 örnek içerisinden 116 örneği doğru, 34 örneği ise hatalı tahmin ederek en başarılı sonucu verdiği söylenebilir. Doğru tahmin edilen sınıflandırma sonuçları dikkate alındığında CatBoost algoritması HGAM algoritmasından bir fazla yanlış tahmin ile ikinci en iyi sonuç veren algoritma olmuştur. Şekil 4.1'de yer alan modellerin doğruluk oranları grafiği incelendiğinde, en başarılı algoritma HGAM olurken, CatBoost ve RO algoritmaları HGAM algoritmasına yakın sonuç vermiştir.




**Şekil 4.1** Model 1.1'e göre sınıflandırma algoritmalarının doğruluk oranları

**Tablo 4.7** Model 1.2 için gerçekleştirilen sınıflandırma sonuçları

| Algoritma | Sınıf | KO | DUO | F1 Skor | ÖS | DO | Kappa |
|---|---|---|---|---|---|---|---|
| **XGBoost** | 0 | 1,00 | 0,89 | 0,94 | 9 | 0,708 | 0,485 |
| | 1 | 0,64 | 0,81 | 0,72 | 54 | | |
| | 2 | 0,77 | 0,58 | 0,66 | 57 | | |
| MO | | 0,80 | 0,76 | 0,77 | 120 | | |
| AO | | 0,73 | 0,71 | 0,71 | 120 | | |
| **HGAM** | 0 | 1,00 | 0,89 | 0,94 | 9 | 0,700 | 0,470 |
| | 1 | 0,63 | 0,80 | 0,70 | 54 | | |
| | 2 | 0,75 | 0,58 | 0,65 | 57 | | |
| MO | | 0,79 | 0,75 | 0,77 | 120 | | |
| AO | | 0,72 | 0,70 | 0,70 | 120 | | |
| **CatBoost** | 0 | 1,00 | 0,89 | 0,94 | 9 | 0,733 * | 0,529* |
| | 1 | 0,66 | 0,83 | 0,74 | 54 | | |
| | 2 | 0,80 | 0,61 | 0,69 | 57 | | |
| MO | | 0,82 | 0,78 | 0,79 | 120 | | |
| AO | | 0,75 | 0,73 | 0,73 | 120 | | |
| **GAM** | 0 | 1,00 | 1,00 | 1,00 | 9 | 0,675 | 0,429 |
| | 1 | 0,61 | 0,78 | 0,68 | 54 | | |
| | 2 | 0,71 | 0,53 | 0,61 | 57 | | |
| MO | | 0,77 | 0,77 | 0,76 | 120 | | |
| AO | | 0,69 | 0,68 | 0,67 | 120 | | |
| **AdaBoost** | 0 | 1,00 | 1,00 | 1,00 | 9 | 0,608 | 0,311 |
| | 1 | 0,56 | 0,65 | 0,60 | 54 | | |
| | 2 | 0,60 | 0,51 | 0,55 | 57 | | |
| MO | | 0,72 | 0,72 | 0,72 | 120 | | |
| AO | | 0,61 | 0,61 | 0,61 | 120 | | |
| **KNN** | 0 | 0,00 | 0,00 | 0,00 | 9 | 0,475 | 0,039 |
| | 1 | 0,47 | 0,61 | 0,53 | 54 | | |
| | 2 | 0,50 | 0,42 | 0,46 | 57 | | |
| MO | | 0,32 | 0,34 | 0,33 | 120 | | |
| AO | | 0,45 | 0,47 | 0,46 | 120 | | |

**Tablo 4.7** Model 1.2 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **KA** | 0 | 1,00 | 1,00 | 1,00 | 9 | 0,708 | 0,485 |
| | 1 | 0,67 | 0,69 | 0,68 | 54 | | |
| | 2 | 0,70 | 0,68 | 0,69 | 57 | | |
| MO | | 0,79 | 0,79 | 0,79 | 120 | | |
| AO | | 0,71 | 0,71 | 0,71 | 120 | | |
| **RO** | 0 | 1,00 | 0,89 | 0,94 | 9 | 0,717 | 0,498 |
| | 1 | 0,67 | 0,74 | 0,70 | 54 | | |
| | 2 | 0,73 | 0,67 | 0,70 | 57 | | |
| MO | | 0,80 | 0,77 | 0,78 | 120 | | |
| AO | | 0,72 | 0,72 | 0,72 | 120 | | |
| **LR** | 0 | 0,67 | 0,89 | 0,76 | 9 | 0,575 | 0,265 |
| | 1 | 0,53 | 0,61 | 0,57 | 54 | | |
| | 2 | 0,61 | 0,49 | 0,54 | 57 | | |
| MO | | 0,60 | 0,66 | 0,62 | 120 | | |
| AO | | 0,58 | 0,57 | 0,57 | 120 | | |
| **DVM RBF** | 0 | 0,00 | 0,00 | 0,00 | 9 | 0,433 | -0,033 |
| | 1 | 0,44 | 0,93 | 0,60 | 54 | | |
| | 2 | 0,33 | 0,04 | 0,06 | 57 | | |
| MO | | 0,26 | 0,32 | 0,22 | 120 | | |
| AO | | 0,36 | 0,43 | 0,30 | 120 | | |

* Algoritmaların DO ve kappa değerleri içerisinde en yüksek değere sahip olanı göstermektedir.

Veri setinin %60 eğitim %40 test olarak ayrılmasıyla çalışılan Model 1.2'ye ait sınıflandırma sonuçları Tablo 4.7'de yer almaktadır. Sınıflandırma sonuçları doğruluk oranı değerlerine göre incelendiğinde, en iyi sınıflandırma %73,3 doğruluk oranı ile CatBoost algoritması tarafından gerçekleştirilmiştir. Tablo 4.7 incelendiğinde KNN ve DVM RBF Çekirdek algoritmaları ile yapılan sınıflandırmaların doğruluk oranı değerlerinin 0,5'in altında kalarak en kötü sınıflandırma performansını gerçekleştirmişlerdir. Tablo 4.5 Kappa değerlerine

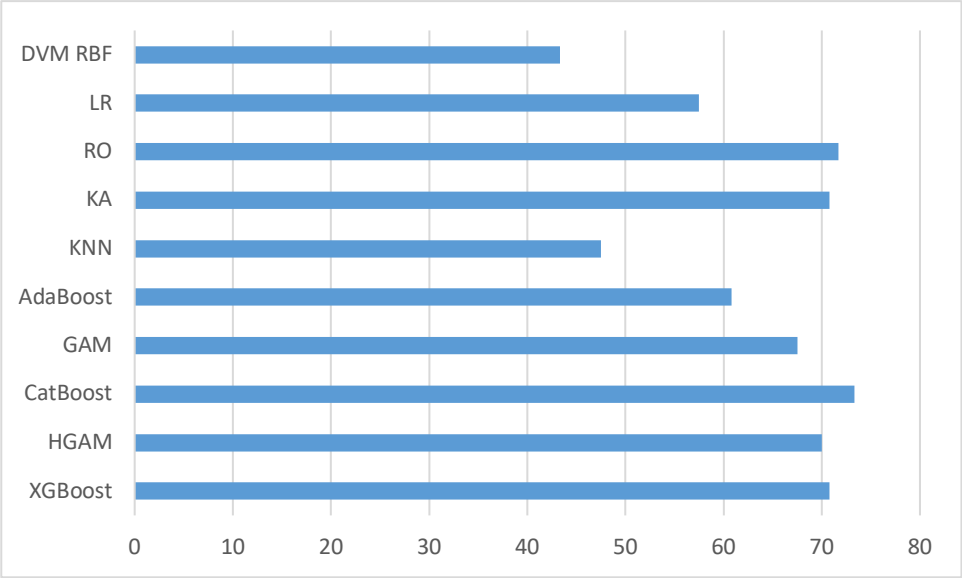
göre incelendiğinde, CatBoost algoritması 0,529 ile en iyi performansı sergilemiştir. Ayrıca, CatBoost algoritması doğruluk oranı ve Kappa metrikleri dışındaki KO, DUO, F1 Skor metriklerine göre de en yüksek başarıyı göstermiştir.

**Tablo 4.8** Model 1.2'ye ait sonuç karşılaştırma matrisi

| **Algoritma** | | **XGBoost** | | | **HGAM** | | | **CatBoost** | | | **GAM** | | | **AdaBoost** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **TES →** | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| **GS** | 0 | 8 | 1 | 0 | 8 | 1 | 0 | 8 | 1 | 0 | 9 | 0 | 0 | 9 | 0 | 0 |
| | 1 | 0 | 44 | 10 | 0 | 43 | 11 | 0 | 45 | 9 | 0 | 42 | 12 | 0 | 35 | 19 |
| | 2 | 0 | 24 | 33 | 0 | 24 | 33 | 0 | 22 | 35 | 0 | 27 | 30 | 0 | 28 | 29 |
| **Algoritma** | | **KNN** | | | **KA** | | | **RO** | | | **LR** | | | **DVM RBF** | | |
| **TES →** | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| **GS** | 0 | 0 | 6 | 3 | 9 | 0 | 0 | 8 | 1 | 0 | 8 | 1 | 0 | 0 | 9 | 0 |
| | 1 | 0 | 33 | 21 | 0 | 37 | 17 | 0 | 40 | 14 | 3 | 33 | 18 | 0 | 50 | 4 |
| | 2 | 2 | 31 | 24 | 0 | 18 | 39 | 0 | 19 | 38 | 1 | 28 | 28 | 0 | 55 | 2 |

TES: Tahmin edilen sınıf, GS: Gerçek sınıf

Model 1.1'e ait sonuç karşılaştırma matrisi Tablo 4.8'de yer almaktadır. Tablo 4.8'e göre CatBoost algoritmasının toplam 120 örnek içerisinden 88 örneği doğru, 32 örneği ise hatalı tahmin ederek en başarılı sonucu verdiği söylenebilir. Doğru tahmin edilen sınıflandırma sonuçları dikkate alındığında RO algoritması CatBoost algoritmasından iki fazla yanlış tahmin ile ikinci en iyi sonuç veren algoritma olmuştur. Şekil 4.2'de yer alan modellerin doğruluk oranları grafiği incelendiğinde, en başarılı algoritma CatBoost olurken, KNN ve DVM RBF Çekirdek algoritmaları %50'nin altında doğruluk oranı ile en kötü sonuçları sergilemiştir.




**Şekil 4.2** Model 1.2'ye göre sınıflandırma algoritmalarının doğruluk oranları

**Tablo 4.9** Model 1.3 için gerçekleştirilen sınıflandırma sonuçları

| Algoritma | Sınıf | KO | DUO | F1 Skor | veri | DO | Kappa |
|---|---|---|---|---|---|---|---|
| **XGBoost** | 0 | 1,00 | 0,86 | 0,92 | 7 | | |
| | 1 | 0,65 | 0,82 | 0,73 | 39 | | |
| | 2 | 0,80 | 0,64 | 0,71 | 44 | 0,733 | 0,531 |
| MO | | 0,82 | 0,77 | 0,79 | 90 | | |
| AO | | 0,75 | 0,73 | 0,73 | 90 | | |
| **HGAM** | 0 | 1,00 | 1,00 | 1,00 | 7 | | |
| | 1 | 0,62 | 0,74 | 0,67 | 39 | | |
| | 2 | 0,72 | 0,59 | 0,65 | 44 | 0,689 | 0,456 |
| MO | | 0,78 | 0,78 | 0,77 | 90 | | |
| AO | | 0,70 | 0,69 | 0,69 | 90 | | |
| **CatBoost** | 0 | 1,00 | 0,86 | 0,92 | 7 | | |
| | 1 | 0,67 | 0,87 | 0,76 | 39 | | |
| | 2 | 0,85 | 0,64 | 0,73 | 44 | 0,756 * | 0,571* |
| MO | | 0,84 | 0,79 | 0,80 | 90 | | |
| AO | | 0,78 | 0,76 | 0,75 | 90 | | |

**Tablo 4.9** Model 1.3 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **GAM** | 0 | 1,00 | 0,71 | 0,83 | 7 | 0,689 | 0,449 |
| | 1 | 0,61 | 0,77 | 0,68 | 39 | | |
| | 2 | 0,75 | 0,61 | 0,67 | 44 | | |
| MO | | 0,79 | 0,70 | 0,73 | 90 | | |
| AO | | 0,71 | 0,69 | 0,69 | 90 | | |
| **AdaBoost** | 0 | 1,00 | 1,00 | 1,00 | 7 | 0,622 | 0,321 |
| | 1 | 0,62 | 0,33 | 0,43 | 39 | | |
| | 2 | 0,58 | 0,82 | 0,68 | 44 | | |
| MO | | 0,73 | 0,72 | 0,70 | 90 | | |
| AO | | 0,63 | 0,62 | 0,60 | 90 | | |
| **KNN** | 0 | 0,50 | 0,14 | 0,22 | 7 | 0,489 | 0,086 |
| | 1 | 0,48 | 0,77 | 0,59 | 39 | | |
| | 2 | 0,52 | 0,30 | 0,38 | 44 | | |
| MO | | 0,50 | 0,40 | 0,40 | 90 | | |
| AO | | 0,50 | 0,49 | 0,46 | 90 | | |
| **KA** | 0 | 1,00 | 0,86 | 0,92 | 7 | 0,722 | 0,512 |
| | 1 | 0,65 | 0,82 | 0,73 | 39 | | |
| | 2 | 0,77 | 0,61 | 0,68 | 44 | | |
| MO | | 0,81 | 0,76 | 0,78 | 90 | | |
| AO | | 0,74 | 0,72 | 0,72 | 90 | | |
| **RO** | 0 | 1,00 | 0,86 | 0,92 | 7 | 0,744 | 0,549 |
| | 1 | 0,67 | 0,79 | 0,73 | 39 | | |
| | 2 | 0,79 | 0,68 | 0,73 | 44 | | |
| MO | | 0,82 | 0,78 | 0,79 | 90 | | |
| AO | | 0,76 | 0,74 | 0,75 | 90 | | |
| **LR** | 0 | 0,67 | 0,86 | 0,75 | 7 | 0,611 | 0,330 |
| | 1 | 0,56 | 0,64 | 0,60 | 39 | | |
| | 2 | 0,67 | 0,55 | 0,60 | 44 | | |
| MO | | 0,63 | 0,68 | 0,65 | 90 | | |
| AO | | 0,62 | 0,61 | 0,61 | 90 | | |

**Tablo 4.9** Model 1.3 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | |
|---|---|---|---|---|---|---|
| **DVM RBF** | 0 | 0,00 | 0,00 | 0,00 | 7 | |
| | 1 | 0,43 | 0,95 | 0,59 | 39 | |
| | 2 | 0,33 | 0,02 | 0,04 | 44 | 0,422 -0,023 |
| MO | | 0,25 | 0,32 | 0,21 | 90 | |
| AO | | 0,35 | 0,42 | 0,28 | 90 | |

* Algoritmaların DO ve kappa değerleri içerisinde en yüksek değere sahip olanı göstermektedir.

Tablo 4.9'da yer alan sonuçlar Model 1.3'e ait olup, veri setinin %70 eğitim ve %30 test olarak ayrılmasıyla elde edilmiştir. Yapılan sınıflandırmada doğruluk oranı değerlerine göre en iyi sınıflandırma CatBoost algoritması tarafından gerçekleştirilmiştir. Tablo 4.9 incelendiğinde, diğer tasarım modellerinde de olduğu gibi KNN ve DVM RBF Çekirdek algoritmaları ile yapılan sınıflandırmaların doğruluk oranı değerleri 0,5'in altında kalarak en kötü sınıflandırma performansını gerçekleştirmişlerdir. Kappa değerlerine göre Tablo 3.6'da yer alan veriler incelendiğinde CatBoost algoritması 0,571 ile en iyi performansı sergilemiştir. Buna ek olarak CatBoost algoritması doğruluk oranı ve Kappa metrikleri dışındaki KO, DUO, F1 Skor metriklerine göre de en yüksek başarıyı göstermiştir.

**Tablo 4.10** Model 1.3'e ait sonuç karşılaştırma matrisi

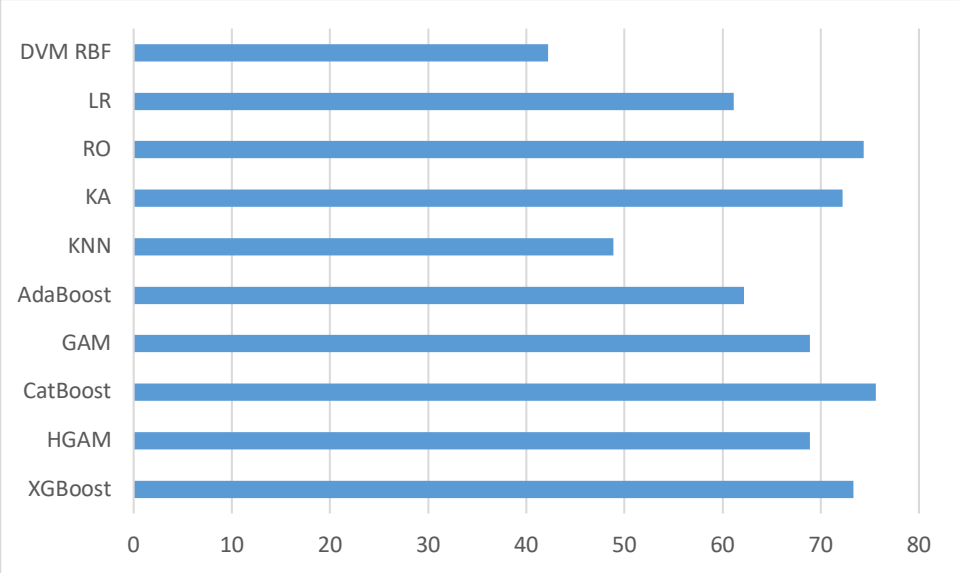
| **Algoritma** | | **XGBoost** | | | **HGAM** | | | **CatBoost** | | | **GAM** | | | **AdaBoost** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **TES →** | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| **GS** | 0 | 6 | 1 | 0 | 7 | 0 | 0 | 6 | 1 | 0 | 5 | 2 | 0 | 7 | 0 | 0 |
| | 1 | 0 | 32 | 7 | 0 | 29 | 10 | 0 | 34 | 5 | 0 | 30 | 9 | 0 | 13 | 26 |
| | 2 | 0 | 16 | 28 | 0 | 18 | 26 | 0 | 16 | 28 | 0 | 17 | 27 | 0 | 8 | 36 |

**Tablo 4.10** Model 1.3'e ait sonuç karşılaştırma matrisi (devamı)

| Algoritma | | KNN | | | KA | | | RO | | | LR | | | DVM RBF | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TES → | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| GS | 0 | 1 | 3 | 3 | 6 | 0 | 1 | 6 | 1 | 0 | 6 | 1 | 0 | 0 | 7 | 0 |
| | 1 | 0 | 30 | 9 | 0 | 32 | 7 | 0 | 31 | 8 | 2 | 25 | 12 | 0 | 37 | 2 |
| | 2 | 1 | 30 | 13 | 0 | 17 | 27 | 0 | 14 | 30 | 1 | 19 | 24 | 0 | 43 | 1 |

TES: Tahmin edilen sınıf, GS: Gerçek sınıf

Model 1.3'e ait sonuç karşılaştırma matrisi Tablo 4.10'da yer almaktadır. Tablo 4.10'a göre CatBoost algoritmasının toplam 90 örnek içerisinden 68 örneği doğru, 22 örneği ise hatalı tahmin ederek en başarılı sonucu verdiği söylenebilir. Doğru tahmin edilen sınıflandırma sonuçları dikkate alındığında RO algoritması CatBoost algoritmasından bir fazla yanlış tahmin ile ikinci en iyi sonuç veren algoritma olmuştur. Şekil 4.3'te yer alan modellerin doğruluk oranları grafiği incelendiğinde, en başarılı algoritma CatBoost olurken, KNN ve DVM RBF Çekirdek algoritmaları %50'nin altında doğruluk oranı ile en kötü sonuçları sergilemiştir.




**Şekil 4.3** Model 1.3'e göre sınıflandırma algoritmalarının doğruluk oranları

**Tablo 4.11** Model 1.4 için gerçekleştirilen sınıflandırma sonuçları

| Algoritma | Sınıf | KO | DUO | F1 Skor | ÖS | DO | Kappa |
|---|---|---|---|---|---|---|---|
| **XGBoost** | 0 | 1,00 | 0,67 | 0,80 | 3 | 0,750 | 0,535 |
| | 1 | 0,71 | 0,83 | 0,76 | 29 | | |
| | 2 | 0,79 | 0,68 | 0,73 | 28 | | |
| MO | | 0,83 | 0,72 | 0,76 | 60 | | |
| AO | | 0,79 | 0,75 | 0,75 | 60 | | |
| **HGAM** | 0 | 1,00 | 1,00 | 1,00 | 3 | 0,767 | 0,572 |
| | 1 | 0,73 | 0,83 | 0,77 | 29 | | |
| | 2 | 0,79 | 0,68 | 0,73 | 28 | | |
| MO | | 0,84 | 0,84 | 0,83 | 60 | | |
| AO | | 0,77 | 0,77 | 0,77 | 60 | | |
| **CatBoost** | 0 | 1,00 | 0,67 | 0,80 | 3 | 0,817* | 0,659 |
| | 1 | 0,75 | 0,93 | 0,83 | 29 | | |
| | 2 | 0,91 | 0,71 | 0,80 | 28 | | |
| MO | | 0,89 | 0,77 | 0,81 | 60 | | |
| AO | | 0,84 | 0,82 | 0,81 | 60 | | |
| **GAM** | 0 | 1,00 | 1,00 | 1,00 | 3 | 0,817* | 0,663* |
| | 1 | 0,76 | 0,90 | 0,83 | 29 | | |
| | 2 | 0,87 | 0,71 | 0,78 | 28 | | |
| MO | | 0,88 | 0,87 | 0,87 | 60 | | |
| AO | | 0,83 | 0,82 | 0,81 | 60 | | |
| **AdaBoost** | 0 | 0,50 | 1,00 | 0,67 | 3 | 0,600 | 0,301 |
| | 1 | 0,73 | 0,28 | 0,40 | 29 | | |
| | 2 | 0,58 | 0,89 | 0,70 | 28 | | |
| MO | | 0,60 | 0,72 | 0,59 | 60 | | |
| AO | | 0,65 | 0,60 | 0,56 | 60 | | |
| **KNN** | 0 | 0,00 | 0,00 | 0,00 | 3 | 0,517 | 0,088 |
| | 1 | 0,54 | 0,69 | 0,61 | 29 | | |
| | 2 | 0,50 | 0,39 | 0,44 | 28 | | |
| MO | | 0,35 | 0,36 | 0,35 | 60 | | |
| AO | | 0,49 | 0,52 | 0,50 | 60 | | |

**Tablo 4.11** Model 1.4 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **KA** | 0 | 1,00 | 1,00 | 1,00 | 3 | 0,750 | 0,541 |
| | 1 | 0,71 | 0,83 | 0,76 | 29 | | |
| | 2 | 0,78 | 0,64 | 0,71 | 28 | | |
| MO | | 0,83 | 0,82 | 0,82 | 60 | | |
| AO | | 0,76 | 0,75 | 0,75 | 60 | | |
| **RO** | 0 | 1,00 | 0,67 | 0,80 | 3 | 0,817 * | 0,659 |
| | 1 | 0,75 | 0,93 | 0,83 | 29 | | |
| | 2 | 0,91 | 0,71 | 0,80 | 28 | | |
| MO | | 0,89 | 0,77 | 0,81 | 60 | | |
| AO | | 0,84 | 0,82 | 0,81 | 60 | | |
| **LR** | 0 | 1,00 | 0,67 | 0,80 | 3 | 0,717 | 0,474 |
| | 1 | 0,69 | 0,76 | 0,72 | 29 | | |
| | 2 | 0,73 | 0,68 | 0,70 | 28 | | |
| MO | | 0,81 | 0,70 | 0,74 | 60 | | |
| AO | | 0,72 | 0,72 | 0,72 | 60 | | |
| **DVM RBF** | 0 | 0,00 | 0,00 | 0,00 | 3 | 0,422 | -0,023 |
| | 1 | 0,86 | 0,21 | 0,33 | 29 | | |
| | 2 | 0,53 | 1,00 | 0,69 | 28 | | |
| MO | | 0,46 | 0,40 | 0,34 | 60 | | |
| AO | | 0,66 | 0,57 | 0,48 | 60 | | |

* Algoritmaların DO ve kappa değerleri içerisinde en yüksek değere sahip olanı göstermektedir.

Veri setinin %80 eğitim %20 test olarak ayrılmasıyla çalışılan Model 1.4'e ait sınıflandırma sonuçları Tablo 4.11'de yer almaktadır. Sınıflandırma sonuçları doğruluk oranı değerlerine göre incelendiğinde CatBoost, RO ve GAM algoritmalarının üçü de %81,7 doğruluk oranı ile en iyi sınıflandırma sonuçlarını gerçekleştirmişlerdir. Tablo 4.11 incelendiğinde KNN algoritmasına ait sınıflandırma sonucu doğruluk oranı değerlerinin 0,5'in altında kalarak en kötü sınıflandırma performansını gerçekleştirmiştir. Tablo 4.11 Kappa değerlerine göre

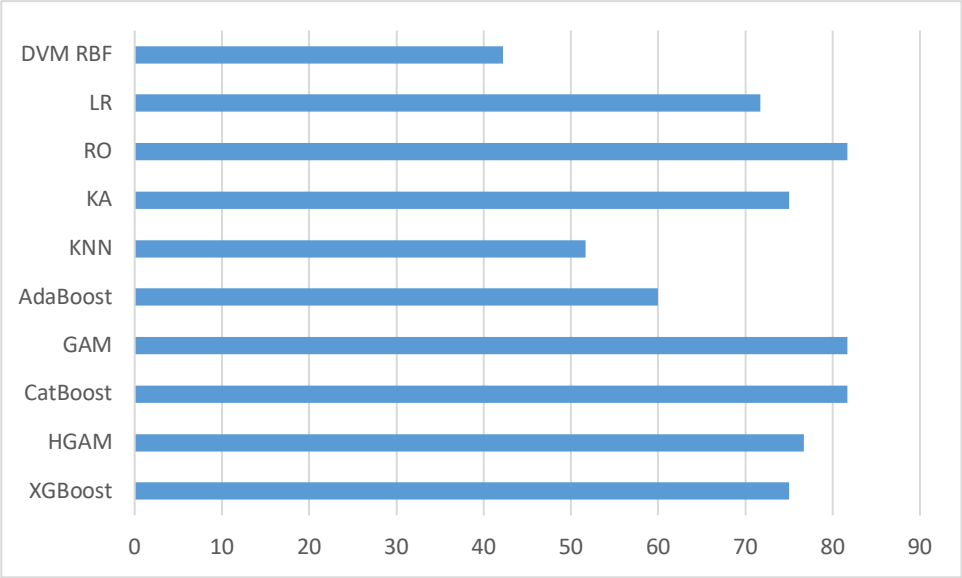
incelendiğinde, GAM algoritması 0,663 ile en iyi performansı sergilemiştir, CatBoost ve RO algoritmaları yüksek doğruluk oranları elde etmesine rağmen Kappa istatistiğinde 0,659 ile GAM algoritmasının ardından ikinci sırada yer almışlardır.

**Tablo 4.12** Model 1.4'e ait sonuç karşılaştırma matrisi

| Algoritma | | XGBoost | | | HGAM | | | CatBoost | | | GAM | | | AdaBoost | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TES → | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| GS | 0 | 2 | 1 | 0 | 3 | 0 | 0 | 2 | 1 | 0 | 3 | 0 | 0 | 3 | 0 | 0 |
| | 1 | 0 | 24 | 5 | 0 | 24 | 5 | 0 | 27 | 2 | 0 | 26 | 3 | 3 | 8 | 18 |
| | 2 | 0 | 9 | 19 | 0 | 9 | 19 | 0 | 8 | 20 | 0 | 8 | 20 | 0 | 3 | 25 |
| Algoritma | | KNN | | | KA | | | RO | | | LR | | | DVM RBF | | |
| TES → | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| GS | 0 | 0 | 1 | 2 | 3 | 0 | 0 | 2 | 1 | 0 | 2 | 1 | 0 | 0 | 1 | 2 |
| | 1 | 0 | 20 | 9 | 0 | 24 | 5 | 0 | 27 | 2 | 0 | 22 | 7 | 0 | 6 | 23 |
| | 2 | 1 | 16 | 11 | 0 | 10 | 18 | 0 | 8 | 20 | 0 | 9 | 19 | 0 | 0 | 28 |

TES: Tahmin edilen sınıf, GS: Gerçek sınıf

Model 1.4'e ait sonuç karşılaştırma matrisi Tablo 4.12'de yer almaktadır. Tablo 4.12'ye göre CatBoost, GAM ve RO algoritmaları toplam 60 örnek içerisinden 49 örneği doğru, 11 örneği ise hatalı tahmin ederek en başarılı sonucu verdikleri söylenebilir. DVM RBF Çekirdek algoritması 0,5 doğruluk oranının altında kalarak oldukça başarısız bir sınıflandırma sonucu vermesine karşın iki numaralı sınıftaki tüm örnekleri doğru tahminlemesi Tablo 4.12'de dikkat çekmektedir. Şekil 4.4'te yer alan modellerin doğruluk oranları grafiği incelendiğinde, en başarılı algoritmalar CatBoost, GAM ve RO algoritmaları olarak öne çıkmaktadır.




**Şekil 4.4** Model 1.4'e göre sınıflandırma algoritmalarının doğruluk oranları

**Tablo 4.13** Model 1.5 için gerçekleştirilen sınıflandırma sonuçları

| Algoritma | Sınıf | KO | DUO | F1 Skor | ÖS | DO | Kappa |
|---|---|---|---|---|---|---|---|
| **XGBoost** | 0 | 0,00 | 0,00 | 0,00 | 1 | 0,767 | 0,550 |
| | 1 | 0,71 | 0,86 | 0,77 | 14 | | |
| | 2 | 0,85 | 0,73 | 0,79 | 15 | | |
| MO | | 0,52 | 0,53 | 0,52 | 30 | | |
| AO | | 0,75 | 0,77 | 0,75 | 30 | | |
| **HGAM** | 0 | 1,00 | 1,00 | 1,00 | 1 | 0,767 | 0,563 |
| | 1 | 0,71 | 0,86 | 0,77 | 14 | | |
| | 2 | 0,83 | 0,67 | 0,74 | 15 | | |
| MO | | 0,85 | 0,84 | 0,84 | 30 | | |
| AO | | 0,78 | 0,77 | 0,76 | 30 | | |
| **CatBoost** | 0 | 0,00 | 0,00 | 0,00 | 1 | 0,767 | 0,550 |
| | 1 | 0,71 | 0,86 | 0,79 | 14 | | |
| | 2 | 0,85 | 0,73 | 0,79 | 15 | | |
| MO | | 0,52 | 0,53 | 0,52 | 30 | | |
| AO | | 0,75 | 0,77 | 0,75 | 30 | | |

**Tablo 4.13** Model 1.5 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **GAM** | 0 | 1,00 | 1,00 | 1,00 | 1 | 0,800* | 0,627* |
| | 1 | 0,72 | 0,93 | 0,81 | 14 | | |
| | 2 | 0,91 | 0,67 | 0,77 | 15 | | |
| MO | | 0,88 | 0,87 | 0,86 | 30 | | |
| AO | | 0,82 | 0,80 | 0,80 | 30 | | |
| **AdaBoost** | 0 | 1,00 | 1,00 | 1,00 | 1 | 0,633 | 0,299 |
| | 1 | 0,71 | 0,36 | 0,48 | 14 | | |
| | 2 | 0,59 | 0,87 | 0,70 | 15 | | |
| MO | | 0,77 | 0,74 | 0,73 | 30 | | |
| AO | | 0,66 | 0,63 | 0,61 | 30 | | |
| **KNN** | 0 | 0,00 | 0,00 | 0,00 | 1 | 0,500 | 0,041 |
| | 1 | 0,47 | 0,64 | 0,55 | 14 | | |
| | 2 | 0,55 | 0,40 | 0,46 | 15 | | |
| MO | | 0,34 | 0,35 | 0,34 | 30 | | |
| AO | | 0,49 | 0,50 | 0,49 | 30 | | |
| **KA** | 0 | 1,00 | 1,00 | 1,00 | 1 | 0,667 | 0,372 |
| | 1 | 0,64 | 0,64 | 0,64 | 14 | | |
| | 2 | 0,67 | 0,67 | 0,67 | 15 | | |
| MO | | 0,77 | 0,77 | 0,77 | 30 | | |
| AO | | 0,67 | 0,67 | 0,67 | 30 | | |
| **RO** | 0 | 0,00 | 0,00 | 0,00 | 1 | 0,767 | 0,550 |
| | 1 | 0,71 | 0,86 | 0,77 | 14 | | |
| | 2 | 0,85 | 0,73 | 0,79 | 15 | | |
| MO | | 0,52 | 0,53 | 0,52 | 30 | | |
| AO | | 0,75 | 0,77 | 0,75 | 30 | | |
| **LR** | 0 | 0,00 | 0,00 | 0,00 | 1 | 0,633 | 0,287 |
| | 1 | 0,62 | 0,57 | 0,59 | 14 | | |
| | 2 | 0,65 | 0,73 | 0,69 | 15 | | |
| MO | | 0,42 | 0,43 | 0,43 | 30 | | |
| AO | | 0,61 | 0,63 | 0,62 | 30 | | |

**Tablo 4.13** Model 1.5 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | |
|---|---|---|---|---|---|---|
| **DVM RBF** | 0 | 0,00 | 0,00 | 0,00 | 1 | 0,600 | 0,207 |
| | 1 | 0,75 | 0,21 | 0,33 | 14 | | |
| | 2 | 0,58 | 1,00 | 0,73 | 15 | | |
| MO | | 0,44 | 0,40 | 0,36 | 30 | | |
| AO | | 0,64 | 0,60 | 0,52 | 30 | | |

* Algoritmaların DO ve kappa değerleri içerisinde en yüksek değere sahip olanı göstermektedir.

Tablo 4.13'te yer alan sonuçlar Model 1.5'e ait olup, veri setinin %90 eğitim ve %10 test olarak ayrılmasıyla elde edilmiştir. Yapılan sınıflandırmada doğruluk oranı değerlerine göre en iyi sınıflandırma GAM algoritması tarafından gerçekleştirilmiştir. Kappa değerlerine göre Tablo 4.13'te yer alan veriler incelendiğinde GAM algoritması 0,627 ile en iyi performansı sergilemiştir. Buna ek olarak GAM algoritması doğruluk oranı ve Kappa metrikleri dışındaki KO, DUO, F1 Skor metriklerine göre de en yüksek başarıyı göstermiştir.

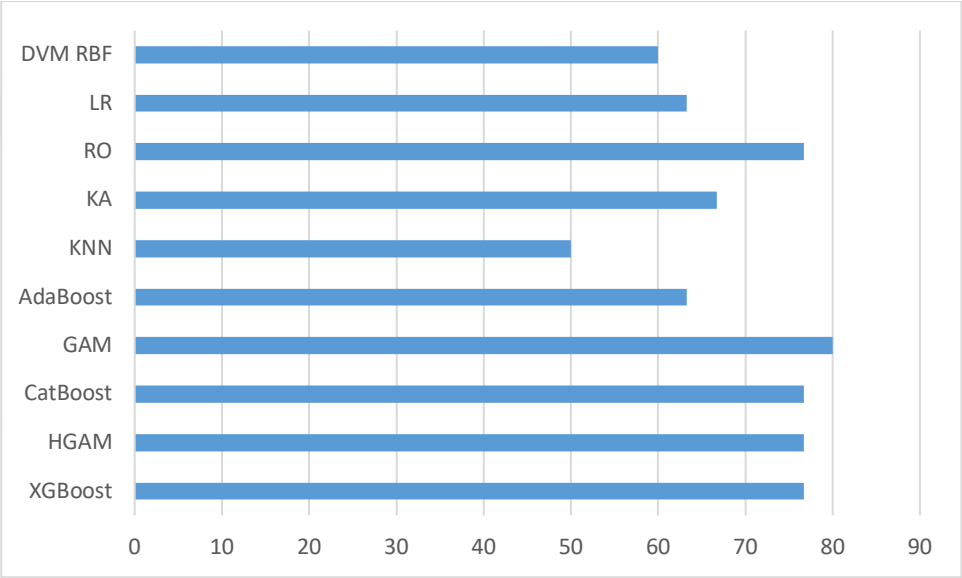
**Tablo 4.14** Model 1.5'e ait sonuç karşılaştırma matrisi

| **Algoritma** | | **XGBoost** | | | **HGAM** | | | **CatBoost** | | | **GAM** | | | **AdaBoost** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **TES →** | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| **GS** | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| | 1 | 0 | 12 | 2 | 0 | 12 | 2 | 0 | 12 | 2 | 0 | 13 | 1 | 0 | 5 | 9 |
| | 2 | 0 | 4 | 11 | 0 | 5 | 10 | 0 | 4 | 11 | 0 | 5 | 10 | 0 | 2 | 13 |

| **Algoritma** | | **KNN** | | | **KA** | | | **RO** | | | **LR** | | | **DVM RBF** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **TES →** | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| **GS** | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| | 1 | 0 | 9 | 5 | 0 | 9 | 5 | 0 | 12 | 2 | 0 | 8 | 6 | 0 | 3 | 11 |
| | 2 | 0 | 9 | 6 | 0 | 5 | 10 | 0 | 4 | 11 | 0 | 4 | 11 | 0 | 0 | 15 |

TES: Tahmin edilen sınıf, GS: Gerçek sınıf

Model 1.5'e ait sonuç karşılaştırma matrisi Tablo 4.14'te yer almaktadır. Tablo 4.14'e göre GAM algoritmasının toplam 30 örnek içerisinden 24 örneği doğru, 6 örneği ise hatalı tahmin ederek en başarılı sonucu verdiği söylenebilir. Doğru tahmin edilen sınıflandırma sonuçları dikkate alındığında XGBoost, HGAM, CatBoost ve RO algoritmaları GAM algoritmasından bir fazla yanlış tahmin ile ikinci en iyi sonuç veren algoritmalar olmuşlardır. Şekil 4.5'te yer alan modellerin doğruluk oranları grafiği incelendiğinde, en başarılı algoritma GAM olurken, KNN algoritması 0,5 doğruluk oranı ile en kötü sonucu sergilemiştir.




**Şekil 4.5** Model 1.5'e göre sınıflandırma algoritmalarının doğruluk oranları

### 4.5.2 Açıklayıcılığı En Yüksek 15 Öznitelik ile Gerçekleştirilen Sınıflandırma Sonuçları

Bu bölümde, öznitelik seçim algoritmalarından yararlanılarak belirlenmiş olan açıklayıcılığı en yüksek ilk 15 öznitelik kullanılarak gerçekleştirilen sınıflandırma sonuçlarına yer verilecektir.

Araştırmada ele alınan öznitelikler önem dereceleri belirlenerek Borda yönteminden yararlanılarak sıralanmıştır. Sıralama işlemi için Borda yönteminden yararlanılmasının amacı genel bir sıralama ortaya koymaktır, her bir algoritmanın

kendi değer aralıklarında verdikleri sonuçtan yararlanılması hatalı bir sıralama ve hatalı sonuçlar doğuracağından sıralama işlemi için bu yöntem tercih edilmiştir. 3.6.4 Bölümünde anlatılan Borda yöntemini kullanarak Feature Importance, Mutual Info ve Shap tabanlı, araştırmada incelenen algoritmaları da kullanarak tüm öznitelikler önem derecelerine göre sıralanmıştır. Öznitelik seçim algoritmalarından yararlanılarak belirlenen en iyi açıklayıcılığa sahip ilk 15 öznitelik; bir günde ekran başında geçirilen süre, geleneksel/dijital bankacılık tercihi, mobil veya internet bankacılığı kullanımı, mevcutta ödenmekte olan kredi sayısı, müşterisi olunan banka sayısı, yaş, aylık birikim/gelir oranı, aylık net gelir, sık kullanılan bankacılık işlemleri, yatırım tercihleri, meslek, cinsiyet, konut sahipliği, ikamet edilen konutun müşterisi olunan banka ATM uzaklığı ve kredi yaklaşımı şeklindedir.

**Tablo 4.15** Model 2.1 için gerçekleştirilen sınıflandırma sonuçları

| Algoritma | Sınıf | KO | DUO | F1 Skor | ÖS | DO | Kappa |
|---|---|---|---|---|---|---|---|
| **XGBoost** | 0 | 0,85 | 1,00 | 0,92 | 11 | 0,727 | 0,521 |
| | 1 | 0,68 | 0,77 | 0,72 | 69 | | |
| | 2 | 0,76 | 0,64 | 0,70 | 70 | | |
| MO | | 0,76 | 0,80 | 0,78 | 150 | | |
| AO | | 0,73 | 0,73 | 0,72 | 150 | | |
| **HGAM** | 0 | 1,00 | 1,00 | 1,00 | 11 | 0,733 | 0,529 |
| | 1 | 0,67 | 0,83 | 0,74 | 69 | | |
| | 2 | 0,78 | 0,60 | 0,68 | 70 | | |
| MO | | 0,82 | 0,81 | 0,81 | 150 | | |
| AO | | 0,74 | 0,73 | 0,73 | 150 | | |
| **CatBoost** | 0 | 1,00 | 1,00 | 1,00 | 11 | 0,760 | 0,576 |
| | 1 | 0,70 | 0,84 | 0,76 | 69 | | |
| | 2 | 0,80 | 0,64 | 0,71 | 70 | | |
| MO | | 0,83 | 0,83 | 0,83 | 150 | | |
| AO | | 0,77 | 0,76 | 0,76 | 150 | | |

**Tablo 4.15** Model 2.1 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **GAM** | 0 | 0,92 | 1,00 | 0,96 | 11 | 0,720 | 0,507 |
| | 1 | 0,67 | 0,78 | 0,72 | 69 | | |
| | 2 | 0,75 | 0,61 | 0,68 | 70 | | |
| MO | | 0,78 | 0,80 | 0,78 | 150 | | |
| AO | | 0,73 | 0,72 | 0,72 | 150 | | |
| **AdaBoost** | 0 | 0,79 | 1,00 | 0,88 | 11 | 0,620 | 0,335 |
| | 1 | 0,65 | 0,38 | 0,48 | 69 | | |
| | 2 | 0,58 | 0,80 | 0,67 | 70 | | |
| MO | | 0,67 | 0,73 | 0,68 | 150 | | |
| AO | | 0,63 | 0,62 | 0,60 | 150 | | |
| **KNN** | 0 | 0,00 | 0,00 | 0,00 | 11 | 0,480 | 0,050 |
| | 1 | 0,49 | 0,57 | 0,52 | 69 | | |
| | 2 | 0,50 | 0,47 | 0,49 | 70 | | |
| MO | | 0,33 | 0,35 | 0,34 | 150 | | |
| AO | | 0,46 | 0,48 | 0,47 | 150 | | |
| **KA** | 0 | 1,00 | 1,00 | 1,00 | 11 | 0,713 | 0,493 |
| | 1 | 0,70 | 0,65 | 0,68 | 69 | | |
| | 2 | 0,68 | 0,73 | 0,70 | 70 | | |
| MO | | 0,79 | 0,79 | 0,79 | 150 | | |
| AO | | 0,71 | 0,71 | 0,71 | 150 | | |
| **RO** | 0 | 1,00 | 1,00 | 1,00 | 11 | 0,767* | 0,588* |
| | 1 | 0,71 | 0,84 | 0,77 | 69 | | |
| | 2 | 0,81 | 0,66 | 0,72 | 70 | | |
| MO | | 0,84 | 0,83 | 0,83 | 150 | | |
| AO | | 0,78 | 0,77 | 0,76 | 150 | | |
| **LR** | 0 | 0,60 | 0,82 | 0,69 | 11 | 0,607 | 0,317 |
| | 1 | 0,58 | 0,57 | 0,57 | 69 | | |
| | 2 | 0,63 | 0,61 | 0,62 | 70 | | |
| MO | | 0,60 | 0,67 | 0,63 | 150 | | |
| AO | | 0,61 | 0,61 | 0,61 | 150 | | |

**Tablo 4.15** Model 2.1 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | |
|---|---|---|---|---|---|---|
| **DVM RBF** | 0 | 0,00 | 0,00 | 0,00 | 11 | |
| | 1 | 0,46 | 0,94 | 0,62 | 69 | |
| | 2 | 0,50 | 0,07 | 0,12 | 70 | 0,467 0,012 |
| MO | | 0,32 | 0,34 | 0,25 | 150 | |
| AO | | 0,45 | 0,47 | 0,34 | 150 | |

* Algoritmaların DO ve kappa değerleri içerisinde en yüksek değere sahip olanı göstermektedir.

Tablo 4.15'te yer alan sonuçlar Model 2.1'e ait olup, en iyi açıklayıcılığa sahip ilk 15 öznitelik seçilerek, veri setinin %50 eğitim ve %50 test olarak ayrılmasıyla elde edilmiştir. Yapılan sınıflandırmada doğruluk oranı değerlerine göre en iyi sınıflandırma RO algoritması tarafından gerçekleştirilmiştir. Tablo 4.15'e göre KNN ve DVM RBF Çekirdek algoritmaları ile yapılan sınıflandırmaların doğruluk oranı değerleri 0,5'in altında kalarak en kötü sınıflandırma performansını gerçekleştirmişlerdir. Kappa değerlerine göre Tablo 4.15'te yer alan veriler incelendiğinde GAM algoritması 0,627 ile en iyi performansı sergilemiştir. Buna ek olarak RO algoritması doğruluk oranı ve Kappa metrikleri dışındaki KO, DUO, F1 Skor metriklerine göre de en yüksek başarıyı göstermiştir.

**Tablo 4.16** Model 2.1'e ait sonuç karşılaştırma matrisi

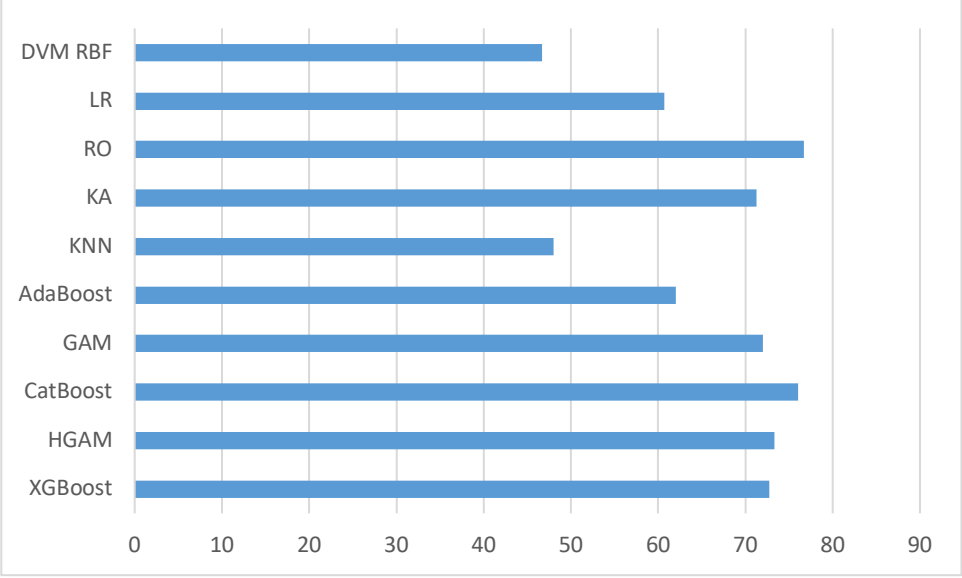
| **Algoritma** | | **XGBoost** | | | **HGAM** | | | **CatBoost** | | | **GAM** | | | **AdaBoost** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **TES →** | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| **GS** | 0 | 11 | 0 | 0 | 11 | 0 | 0 | 11 | 0 | 0 | 11 | 0 | 0 | 11 | 0 | 0 |
| | 1 | 2 | 53 | 14 | 0 | 57 | 12 | 0 | 58 | 11 | 1 | 54 | 14 | 3 | 26 | 40 |
| | 2 | 0 | 25 | 45 | 0 | 28 | 42 | 0 | 25 | 45 | 0 | 27 | 43 | 0 | 14 | 56 |

**Tablo 4.16** Model 2.1'e ait sonuç karşılaştırma matrisi (devamı)

| Algoritma | | KNN | | | KA | | | RO | | | LR | | | DVM RBF | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TES → | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| GS | 0 | 0 | 7 | 4 | 11 | 0 | 0 | 11 | 0 | 0 | 9 | 2 | 0 | 0 | 10 | 1 |
| | 1 | 1 | 39 | 29 | 0 | 45 | 24 | 0 | 58 | 11 | 5 | 39 | 25 | 0 | 65 | 4 |
| | 2 | 3 | 34 | 33 | 0 | 19 | 51 | 0 | 24 | 46 | 1 | 26 | 43 | 0 | 65 | 5 |

TES: Tahmin edilen sınıf, GS: Gerçek sınıf

Model 2.1'e ait sonuç karşılaştırma matrisi Tablo 4.16'da yer almaktadır. Tablo 4.16'ya göre RO algoritmasının toplam 150 örnek içerisinden 115 örneği doğru, 35 örneği ise hatalı tahmin ederek en başarılı sonucu verdiği söylenebilir. Doğru tahmin edilen sınıflandırma sonuçları dikkate alındığında CatBoost algoritması RO algoritmasından bir fazla yanlış tahmin ile ikinci en iyi sonuç veren algoritma olmuştur. Şekil 4.6'da yer alan modellerin doğruluk oranları grafiği incelendiğinde, en başarılı algoritma RO algoritması olarak öne çıkmaktadır.




**Şekil 4.6** Model 2.1'e göre sınıflandırma algoritmalarının doğruluk oranları

**Tablo 4.17** Model 2.2 için gerçekleştirilen sınıflandırma sonuçları

| Algoritma | Sınıf | KO | DUO | F1 Skor | ÖS | DO | Kappa |
|---|---|---|---|---|---|---|---|
| **XGBoost** | 0 | 1,00 | 0,89 | 0,94 | 9 | 0,700 | 0,470 |
| | 1 | 0,63 | 0,80 | 0,70 | 54 | | |
| | 2 | 0,75 | 0,58 | 0,65 | 57 | | |
| MO | | 0,79 | 0,75 | 0,77 | 120 | | |
| AO | | 0,72 | 0,70 | 0,70 | 120 | | |
| **HGAM** | 0 | 1,00 | 0,89 | 0,94 | 9 | 0,708 | 0,484 |
| | 1 | 0,64 | 0,80 | 0,71 | 54 | | |
| | 2 | 0,76 | 0,60 | 0,67 | 57 | | |
| MO | | 0,80 | 0,76 | 0,77 | 120 | | |
| AO | | 0,72 | 0,71 | 0,71 | 120 | | |
| **CatBoost** | 0 | 1,00 | 0,89 | 0,94 | 9 | 0,750* | 0,558* |
| | 1 | 0,67 | 0,87 | 0,76 | 54 | | |
| | 2 | 0,83 | 0,61 | 0,71 | 57 | | |
| MO | | 0,83 | 0,79 | 0,80 | 120 | | |
| AO | | 0,77 | 0,75 | 0,75 | 120 | | |
| **GAM** | 0 | 1,00 | 1,00 | 1,00 | 9 | 0,717 | 0,502 |
| | 1 | 0,65 | 0,80 | 0,72 | 54 | | |
| | 2 | 0,76 | 0,60 | 0,67 | 57 | | |
| MO | | 0,80 | 0,80 | 0,79 | 120 | | |
| AO | | 0,73 | 0,72 | 0,71 | 120 | | |
| **AdaBoost** | 0 | 1,00 | 1,00 | 1,00 | 9 | 0,617 | 0,321 |
| | 1 | 0,59 | 0,48 | 0,53 | 54 | | |
| | 2 | 0,58 | 0,68 | 0,63 | 57 | | |
| MO | | 0,72 | 0,72 | 0,72 | 120 | | |
| AO | | 0,62 | 0,62 | 0,61 | 120 | | |
| **KNN** | 0 | 0,00 | 0,00 | 0,00 | 9 | 0,492 | 0,063 |
| | 1 | 0,48 | 0,61 | 0,54 | 54 | | |
| | 2 | 0,52 | 0,46 | 0,49 | 57 | | |
| MO | | 0,33 | 0,36 | 0,34 | 120 | | |
| AO | | 0,46 | 0,49 | 0,47 | 120 | | |

**Tablo 4.17** Model 2.2 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **KA** | 0 | 1,00 | 1,00 | 1,00 | 9 | 0,675 | 0,427 |
| | 1 | 0,63 | 0,69 | 0,65 | 54 | | |
| | 2 | 0,67 | 0,61 | 0,64 | 57 | | |
| MO | | 0,77 | 0,77 | 0,77 | 120 | | |
| AO | | 0,68 | 0,68 | 0,67 | 120 | | |
| **RO** | 0 | 1,00 | 1,00 | 1,00 | 9 | 0,725 | 0,517 |
| | 1 | 0,66 | 0,81 | 0,73 | 54 | | |
| | 2 | 0,77 | 0,60 | 0,67 | 57 | | |
| MO | | 0,81 | 0,80 | 0,80 | 120 | | |
| AO | | 0,74 | 0,72 | 0,72 | 120 | | |
| **LR** | 0 | 0,70 | 0,78 | 0,74 | 9 | 0,600 | 0,300 |
| | 1 | 0,56 | 0,65 | 0,60 | 54 | | |
| | 2 | 0,64 | 0,53 | 0,58 | 57 | | |
| MO | | 0,63 | 0,65 | 0,64 | 120 | | |
| AO | | 0,61 | 0,60 | 0,60 | 120 | | |
| **DVM RBF** | 0 | 0,00 | 0,00 | 0,00 | 9 | 0,433 | -0,033 |
| | 1 | 0,44 | 0,93 | 0,60 | 54 | | |
| | 2 | 0,33 | 0,04 | 0,06 | 57 | | |
| MO | | 0,26 | 0,32 | 0,22 | 120 | | |
| AO | | 0,36 | 0,43 | 0,30 | 120 | | |

* Algoritmaların DO ve kappa değerleri içerisinde en yüksek değere sahip olanı göstermektedir.

Veri seti içerisinde en iyi açıklayıcılığa sahip ilk 15 öznitelik belirlenerek, %60 eğitim ve %40 test olarak ayrılmasıyla çalışılan Model 2.2'ye ait sınıflandırma sonuçları Tablo 4.17'de yer almaktadır. Sınıflandırma sonuçları doğruluk oranı değerlerine göre incelendiğinde CatBoost algoritması %75 doğruluk oranı en iyi sınıflandırma sonucunu gerçekleştirmiştir. Tablo 4.17 incelendiğinde KNN ve DVM RBF Çekirdek algoritmaları sınıflandırma sonuçlarının doğruluk oranı değerleri 0,5'in altında kalarak en kötü sınıflandırma performansını

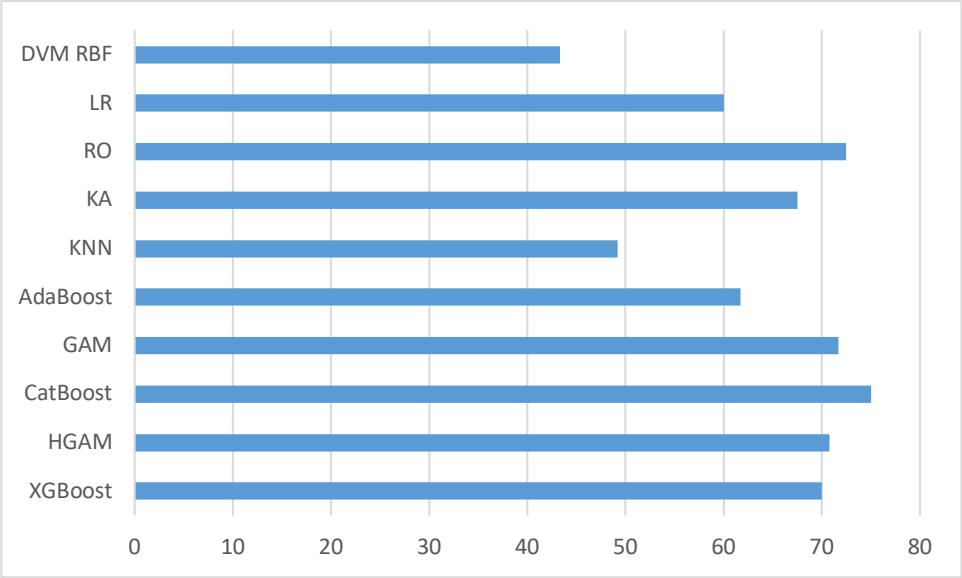
gerçekleştirmiştir. Tablo 4.17 Kappa değerlerine göre incelendiğinde, DVM RBF Çekirdek algoritmaları sıfırın altında kalarak en kötü sonucu göstermiştir, CatBoost algoritması ise 0,558 Kappa değeri ile en iyi performansı sergilemiştir.

**Tablo 4.18** Model 2.2'ye ait sonuç karşılaştırma matrisi

| **Algoritma** | | **XGBoost** | | | **HGAM** | | | **CatBoost** | | | **GAM** | | | **AdaBoost** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **TES→** | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| **GS** | 0 | 8 | 1 | 0 | 8 | 1 | 0 | 8 | 1 | 0 | 9 | 0 | 0 | 9 | 0 | 0 |
| | 1 | 0 | 43 | 11 | 0 | 43 | 11 | 0 | 47 | 7 | 0 | 43 | 11 | 0 | 26 | 28 |
| | 2 | 0 | 24 | 33 | 0 | 23 | 34 | 0 | 22 | 35 | 0 | 23 | 34 | 0 | 18 | 39 |
| **Algoritma** | | **KNN** | | | **KA** | | | **RO** | | | **LR** | | | **DVM RBF** | | |
| **TES →** | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| **GS** | 0 | 0 | 6 | 3 | 9 | 0 | 0 | 9 | 0 | 0 | 7 | 2 | 0 | 0 | 9 | 0 |
| | 1 | 0 | 33 | 21 | 0 | 37 | 17 | 0 | 44 | 10 | 2 | 35 | 17 | 0 | 50 | 4 |
| | 2 | 1 | 30 | 26 | 0 | 22 | 35 | 0 | 23 | 34 | 1 | 26 | 30 | 0 | 55 | 2 |

TES: Tahmin edilen sınıf, GS: Gerçek sınıf

Tablo 4.18'de yer alan Model 2.2'ye ait sonuç karşılaştırma matrisi incelendiğinde CatBoost algoritmasının toplam 120 örnek içerisinden 90 örneği doğru, 30 örneği ise hatalı tahmin ederek en başarılı sonucu verdiği söylenebilir. Şekil 4.7'de yer alan modellerin doğruluk oranları grafiği incelendiğinde, en başarılı algoritma CatBoost olurken, en iyi ikinci doğruluk oranının RO algoritması tarafından gerçekleştirildiği gözlenmektedir.




**Şekil 4.7** Model 2.2'ye göre sınıflandırma algoritmalarının doğruluk oranları

**Tablo 4.19** Model 2.3 için gerçekleştirilen sınıflandırma sonuçları

| Algoritma | Sınıf | KO | DUO | F1 Skor | ÖS | DO | Kappa |
|---|---|---|---|---|---|---|---|
| **XGBoost** | 0 | 1,00 | 0,86 | 0,92 | 7 | 0,733 | 0,532 |
| | 1 | 0,65 | 0,85 | 0,73 | 39 | | |
| | 2 | 0,82 | 0,61 | 0,70 | 44 | | |
| MO | | 0,82 | 0,77 | 0,79 | 90 | | |
| AO | | 0,76 | 0,73 | 0,73 | 90 | | |
| **HGAM** | 0 | 1,00 | 1,00 | 1,00 | 7 | 0,700 | 0,477 |
| | 1 | 0,62 | 0,79 | 0,70 | 39 | | |
| | 2 | 0,76 | 0,57 | 0,65 | 44 | | |
| MO | | 0,79 | 0,79 | 0,78 | 90 | | |
| AO | | 0,72 | 0,70 | 0,70 | 90 | | |
| **CatBoost** | 0 | 1,00 | 0,86 | 0,92 | 7 | 0,711 | 0,494 |
| | 1 | 0,62 | 0,85 | 0,72 | 39 | | |
| | 2 | 0,81 | 0,57 | 0,67 | 44 | | |
| MO | | 0,81 | 0,76 | 0,77 | 90 | | |
| AO | | 0,74 | 0,71 | 0,71 | 90 | | |

**Tablo 4.19** Model 2.3 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **GAM** | 0 | 1,00 | 0,71 | 0,83 | 7 | 0,722 | 0,508 |
| | 1 | 0,64 | 0,82 | 0,72 | 39 | | |
| | 2 | 0,80 | 0,64 | 0,71 | 44 | | |
| MO | | 0,81 | 0,72 | 0,75 | 90 | | |
| AO | | 0,75 | 0,72 | 0,72 | 90 | | |
| **AdaBoost** | 0 | 1,00 | 1,00 | 1,00 | 7 | 0,622 | 0,321 |
| | 1 | 0,62 | 0,33 | 0,43 | 39 | | |
| | 2 | 0,58 | 0,82 | 0,68 | 44 | | |
| MO | | 0,73 | 0,72 | 0,70 | 90 | | |
| AO | | 0,63 | 0,62 | 0,60 | 90 | | |
| **KNN** | 0 | 0,00 | 0,00 | 0,00 | 7 | 0,478 | 0,063 |
| | 1 | 0,47 | 0,72 | 0,57 | 39 | | |
| | 2 | 0,54 | 0,34 | 0,42 | 44 | | |
| MO | | 0,33 | 0,35 | 0,33 | 90 | | |
| AO | | 0,46 | 0,48 | 0,45 | 90 | | |
| **KA** | 0 | 1,00 | 1,00 | 1,00 | 7 | 0,678 | 0,432 |
| | 1 | 0,62 | 0,64 | 0,63 | 39 | | |
| | 2 | 0,67 | 0,66 | 0,67 | 44 | | |
| MO | | 0,77 | 0,77 | 0,77 | 90 | | |
| AO | | 0,68 | 0,68 | 0,68 | 90 | | |
| **RO** | 0 | 1,00 | 1,00 | 1,00 | 7 | 0,744* | 0,557* |
| | 1 | 0,65 | 0,90 | 0,75 | 39 | | |
| | 2 | 0,86 | 0,57 | 0,68 | 44 | | |
| MO | | 0,84 | 0,82 | 0,81 | 90 | | |
| AO | | 0,78 | 0,74 | 0,74 | 90 | | |
| **LR** | 0 | 0,56 | 0,71 | 0,63 | 7 | 0,622 | 0,351 |
| | 1 | 0,56 | 0,69 | 0,62 | 39 | | |
| | 2 | 0,73 | 0,55 | 0,62 | 44 | | |
| MO | | 0,62 | 0,65 | 0,62 | 90 | | |
| AO | | 0,64 | 0,62 | 0,62 | 90 | | |

**Tablo 4.19** Model 2.3 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | |
|---|---|---|---|---|---|---|
| **DVM RBF** | 0 | 0,00 | 0,00 | 0,00 | 7 | |
| | 1 | 0,43 | 0,95 | 0,59 | 39 | |
| | 2 | 0,33 | 0,02 | 0,04 | 44 | 0,422 -0,023 |
| MO | | 0,25 | 0,32 | 0,21 | 90 | |
| AO | | 0,35 | 0,42 | 0,28 | 90 | |

* Algoritmaların DO ve kappa değerleri içerisinde en yüksek değere sahip olanı göstermektedir.

Tablo 4.19'da yer alan sonuçlar Model 2.3'e ait olup, en iyi açıklayıcılığa sahip ilk 15 öznitelik seçilerek, veri setinin %70 eğitim ve %30 test olarak ayrılmasıyla elde edilmiştir. Yapılan sınıflandırmada doğruluk oranı değerlerine göre en iyi sınıflandırma RO algoritması tarafından gerçekleştirilmiştir. Tablo 4.19 incelendiğinde KNN ve DVM RBF Çekirdek algoritmaları ile yapılan sınıflandırmaların, doğruluk oranı değerleri 0,5'in altında olduğundan, en kötü sınıflandırma performansını gerçekleştirdikleri gözlenmektedir. Kappa değerlerine göre Tablo 4.19'da yer alan veriler incelendiğinde RO algoritması 0,557 ile en iyi performansı sergilemiştir. Buna ek olarak RO algoritması doğruluk oranı ve Kappa metrikleri dışındaki KO, DUO, F1 Skor metriklerine göre de en yüksek başarıyı göstermiştir.

**Tablo 4.20** Model 2.3'e ait sonuç karşılaştırma matrisi

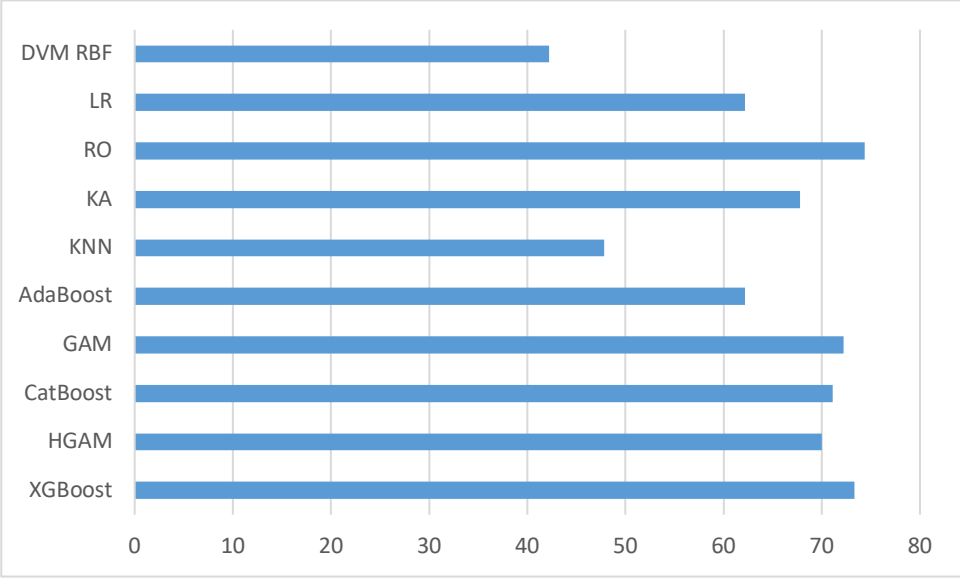
| **Algoritma** | | **XGBoost** | | | **HGAM** | | | **CatBoost** | | | **GAM** | | | **AdaBoost** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **TES →** | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| **GS** | 0 | 6 | 1 | 0 | 7 | 0 | 0 | 6 | 1 | 0 | 5 | 2 | 0 | 7 | 0 | 0 |
| | 1 | 0 | 33 | 6 | 0 | 31 | 8 | 0 | 33 | 6 | 0 | 32 | 7 | 0 | 13 | 26 |
| | 2 | 0 | 17 | 27 | 0 | 19 | 25 | 0 | 19 | 25 | 0 | 16 | 28 | 0 | 8 | 36 |

**Tablo 4.20** Model 2.3'e ait sonuç karşılaştırma matrisi (devamı)

| Algoritma | | KNN | | | KA | | | RO | | | LR | | | DVM RBF | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TES → | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| GS | 0 | 0 | 4 | 3 | 7 | 0 | 0 | 7 | 0 | 0 | 5 | 2 | 0 | 0 | 7 | 0 |
| | 1 | 1 | 28 | 10 | 0 | 25 | 14 | 0 | 35 | 4 | 3 | 27 | 9 | 0 | 37 | 2 |
| | 2 | 1 | 28 | 15 | 0 | 15 | 29 | 0 | 19 | 25 | 1 | 19 | 24 | 0 | 43 | 1 |

TES: Tahmin edilen sınıf, GS: Gerçek sınıf

Model 2.3'e ait sonuç karşılaştırma matrisi Tablo 4.20'de yer almaktadır. Tablo 4.20'ye göre RO algoritmasının toplam 90 örnek içerisinden 67 örneği doğru, 23 örneği ise hatalı tahmin ederek en başarılı sonucu verdiği söylenebilir. Şekil 4.8'de yer alan modellerin doğruluk oranları grafiği incelendiğinde, en başarılı algoritma RO algoritması olarak öne çıkmaktadır. Doğruluk oranları 0,5'in altında kalan KNN ve DVM RBF Çekirdek algoritmaları ise en kötü performanslar olarak grafikte gözlenmektedir.




**Şekil 4.8** Model 2.3'e göre sınıflandırma algoritmalarının doğruluk oranları

**Tablo 4.21** Model 2.4 için gerçekleştirilen sınıflandırma sonuçları

| Algoritma | Sınıf | KO | DUO | F1 Skor | ÖS | DO | Kappa |
|---|---|---|---|---|---|---|---|
| **XGBoost** | 0 | 1,00 | 0,67 | 0,80 | 3 | 0,817 | 0,659 |
| | 1 | 0,76 | 0,90 | 0,83 | 29 | | |
| | 2 | 0,88 | 0,75 | 0,81 | 28 | | |
| MO | | 0,88 | 0,77 | 0,81 | 60 | | |
| AO | | 0,83 | 0,82 | 0,82 | 60 | | |
| **HGAM** | 0 | 1,00 | 1,00 | 1,00 | 3 | 0,817 | 0,663 |
| | 1 | 0,76 | 0,90 | 0,83 | 29 | | |
| | 2 | 0,87 | 0,71 | 0,78 | 28 | | |
| MO | | 0,88 | 0,87 | 0,87 | 60 | | |
| AO | | 0,83 | 0,82 | 0,81 | 60 | | |
| **CatBoost** | 0 | 1,00 | 0,67 | 0,80 | 3 | 0,783 | 0,597 |
| | 1 | 0,72 | 0,90 | 0,80 | 29 | | |
| | 2 | 0,86 | 0,68 | 0,76 | 28 | | |
| MO | | 0,86 | 0,75 | 0,79 | 60 | | |
| AO | | 0,80 | 0,78 | 0,78 | 60 | | |
| **GAM** | 0 | 1,00 | 1,00 | 1,00 | 3 | 0,817 | 0,664 |
| | 1 | 0,78 | 0,86 | 0,82 | 29 | | |
| | 2 | 0,84 | 0,75 | 0,79 | 28 | | |
| MO | | 0,87 | 0,87 | 0,87 | 90 | | |
| AO | | 0,82 | 0,82 | 0,82 | 90 | | |
| **AdaBoost** | 0 | 0,50 | 1,00 | 0,67 | 3 | 0,600 | 0,301 |
| | 1 | 0,73 | 0,28 | 0,40 | 29 | | |
| | 2 | 0,58 | 0,89 | 0,70 | 28 | | |
| MO | | 0,60 | 0,72 | 0,59 | 60 | | |
| AO | | 0,65 | 0,60 | 0,56 | 60 | | |
| **KNN** | 0 | 0,00 | 0,00 | 0,00 | 3 | 0,533 | 0,133 |
| | 1 | 0,56 | 0,62 | 0,59 | 29 | | |
| | 2 | 0,54 | 0,50 | 0,52 | 28 | | |
| MO | | 0,37 | 0,37 | 0,37 | 60 | | |
| AO | | 0,52 | 0,53 | 0,53 | 60 | | |

**Tablo 4.21** Model 2.4 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **KA** | 0 | 1,00 | 1,00 | 1,00 | 3 | 0,733 | 0,510 |
| | 1 | 0,68 | 0,86 | 0,76 | 29 | | |
| | 2 | 0,80 | 0,57 | 0,67 | 28 | | |
| MO | | 0,83 | 0,81 | 0,81 | 60 | | |
| AO | | 0,75 | 0,73 | 0,73 | 60 | | |
| **RO** | 0 | 1,00 | 1,00 | 1,00 | 3 | 0,833* | 0,695* |
| | 1 | 0,81 | 0,86 | 0,83 | 29 | | |
| | 2 | 0,85 | 0,79 | 0,81 | 28 | | |
| MO | | 0,88 | 0,88 | 0,88 | 60 | | |
| AO | | 0,83 | 0,83 | 0,83 | 60 | | |
| **LR** | 0 | 0,67 | 0,67 | 0,67 | 3 | 0,683 | 0,419 |
| | 1 | 0,67 | 0,76 | 0,71 | 29 | | |
| | 2 | 0,71 | 0,61 | 0,65 | 28 | | |
| MO | | 0,68 | 0,68 | 0,68 | 60 | | |
| AO | | 0,69 | 0,68 | 0,68 | 60 | | |
| **DVM RBF** | 0 | 0,00 | 0,00 | 0,00 | 3 | 0,567 | 0,185 |
| | 1 | 0,86 | 0,21 | 0,33 | 29 | | |
| | 2 | 0,53 | 1,00 | 0,69 | 28 | | |
| MO | | 0,46 | 0,40 | 0,34 | 60 | | |
| AO | | 0,66 | 0,57 | 0,48 | 60 | | |

* Algoritmaların DO ve kappa değerleri içerisinde en yüksek değere sahip olanı göstermektedir.

Veri seti içerisinde en iyi açıklayıcılığa sahip ilk 15 öznitelik belirlenerek, %80 eğitim ve %20 test olarak ayrılmasıyla çalışılan Model 2.4'e ait sınıflandırma sonuçları Tablo 4.21'de yer almaktadır. Sınıflandırma sonuçları doğruluk oranı değerlerine göre incelendiğinde RO algoritması %83,3 doğruluk oranı en iyi sınıflandırma sonucunu gerçekleştirmiştir. Tablo 4.21 Kappa değerlerine göre incelendiğinde, RO algoritması 0,695 değeri ile en iyi performansı sergilemiştir.

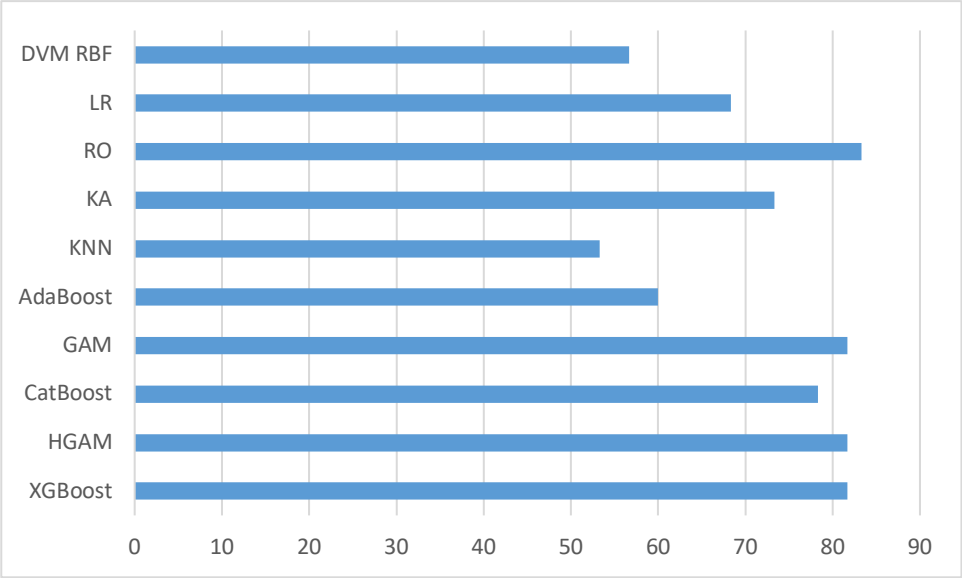
Buna ek olarak RO algoritması doğruluk oranı ve Kappa metrikleri dışındaki KO, DUO, F1 Skor metriklerine göre de en yüksek başarıyı göstermiştir.

**Tablo 4.22** Model 2.4'e ait sonuç karşılaştırma matrisi

| Algoritma | | XGBoost | | | HGAM | | | CatBoost | | | GAM | | | AdaBoost | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TES → | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| GS | 0 | 2 | 1 | 0 | 3 | 0 | 0 | 2 | 1 | 0 | 3 | 0 | 0 | 3 | 0 | 0 |
| | 1 | 0 | 26 | 3 | 0 | 26 | 3 | 0 | 26 | 3 | 0 | 25 | 4 | 3 | 8 | 18 |
| | 2 | 0 | 7 | 21 | 0 | 8 | 20 | 0 | 9 | 19 | 0 | 7 | 21 | 0 | 3 | 25 |
| Algoritma | | KNN | | | KA | | | RO | | | LR | | | DVM RBF | | |
| TES → | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| GS | 0 | 0 | 1 | 2 | 3 | 0 | 0 | 3 | 0 | 0 | 2 | 1 | 0 | 0 | 1 | 2 |
| | 1 | 1 | 18 | 10 | 0 | 25 | 4 | 0 | 25 | 4 | 0 | 22 | 7 | 0 | 6 | 23 |
| | 2 | 1 | 13 | 14 | 0 | 12 | 16 | 0 | 6 | 22 | 1 | 10 | 17 | 0 | 0 | 28 |

TES: Tahmin edilen sınıf, GS: Gerçek sınıf

Tablo 4.22'de yer alan Model 2.4'e ait sonuç karşılaştırma matrisi incelendiğinde RO algoritmasının toplam 60 örnek içerisinden 50 örneği doğru, 10 örneği ise hatalı tahmin ederek en başarılı sonucu gösterdiği gözlenmektedir. Şekil 4.9'da yer alan modellerin doğruluk oranları grafiği incelendiğinde, en başarılı algoritma RO olurken, en iyi ikinci doğruluk oranının XGBoost, HGAM ve GAM algoritmaları tarafından gerçekleştirildiği gözlenmektedir.




**Şekil 4.9** Model 2.4'e göre sınıflandırma algoritmalarının doğruluk oranları

**Tablo 4.23** Model 2.5 için gerçekleştirilen sınıflandırma sonuçları

| Algoritma | Sınıf | KO | DUO | F1 Skor | ÖS | DO | Kappa |
|---|---|---|---|---|---|---|---|
| **XGBoost** | 0 | 1,00 | 1,00 | 1,00 | 1 | 0,833* | 0,687* |
| | 1 | 0,80 | 0,86 | 0,83 | 14 | | |
| | 2 | 0,86 | 0,80 | 0,83 | 15 | | |
| MO | | 0,89 | 0,89 | 0,89 | 30 | | |
| AO | | 0,84 | 0,83 | 0,83 | 30 | | |
| **HGAM** | 0 | 1,00 | 1,00 | 1,00 | 1 | 0,800 | 0,625 |
| | 1 | 0,75 | 0,86 | 0,80 | 14 | | |
| | 2 | 0,85 | 0,73 | 0,79 | 15 | | |
| MO | | 0,87 | 0,86 | 0,86 | 30 | | |
| AO | | 0,81 | 0,80 | 0,80 | 30 | | |
| **CatBoost** | 0 | 0,00 | 0,00 | 0,00 | 1 | 0,800 | 0,615 |
| | 1 | 0,72 | 0,93 | 0,81 | 14 | | |
| | 2 | 0,92 | 0,73 | 0,81 | 15 | | |
| MO | | 0,55 | 0,55 | 0,54 | 30 | | |
| AO | | 0,80 | 0,80 | 0,79 | 30 | | |

**Tablo 4.23** Model 2.5 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **GAM** | 0 | 1,00 | 1,00 | 1,00 | 1 | 0,733 | 0,502 |
| | 1 | 0,67 | 0,86 | 0,75 | 14 | | |
| | 2 | 0,82 | 0,60 | 0,69 | 15 | | |
| MO | | 0,83 | 0,82 | 0,81 | 30 | | |
| AO | | 0,75 | 0,73 | 0,73 | 30 | | |
| **AdaBoost** | 0 | 1,00 | 1,00 | 1,00 | 1 | 0,567 | 0,172 |
| | 1 | 0,57 | 0,29 | 0,38 | 14 | | |
| | 2 | 0,55 | 0,80 | 0,65 | 15 | | |
| MO | | 0,71 | 0,70 | 0,68 | 30 | | |
| AO | | 0,57 | 0,57 | 0,54 | 30 | | |
| **KNN** | 0 | 0,00 | 0,00 | 0,00 | 1 | 0,467 | -0,026 |
| | 1 | 0,44 | 0,57 | 0,50 | 14 | | |
| | 2 | 0,50 | 0,40 | 0,44 | 15 | | |
| MO | | 0,31 | 0,32 | 0,31 | 30 | | |
| AO | | 0,46 | 0,47 | 0,46 | 30 | | |
| **KA** | 0 | 1,00 | 1,00 | 1,00 | 1 | 0,700 | 0,439 |
| | 1 | 0,65 | 0,79 | 0,71 | 14 | | |
| | 2 | 0,75 | 0,60 | 0,67 | 15 | | |
| MO | | 0,80 | 0,80 | 0,79 | 30 | | |
| AO | | 0,71 | 0,70 | 0,70 | 30 | | |
| **RO** | 0 | 1,00 | 1,00 | 1,00 | 1 | 0,767 | 0,562 |
| | 1 | 0,73 | 0,79 | 0,76 | 14 | | |
| | 2 | 0,79 | 0,73 | 0,76 | 15 | | |
| MO | | 0,84 | 0,84 | 0,84 | 30 | | |
| AO | | 0,77 | 0,77 | 0,77 | 30 | | |
| **LR** | 0 | 0,00 | 0,00 | 0,00 | 1 | 0,667 | 0,353 |
| | 1 | 0,64 | 0,64 | 0,64 | 14 | | |
| | 2 | 0,69 | 0,73 | 0,71 | 15 | | |
| MO | | 0,44 | 0,46 | 0,45 | 30 | | |
| AO | | 0,64 | 0,67 | 0,65 | 30 | | |

**Tablo 4.23** Model 2.5 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | |
|---|---|---|---|---|---|---|
| **DVM RBF** | 0 | 0,00 | 0,00 | 0,00 | 1 | 0,600 | 0,207 |
| | 1 | 0,75 | 0,21 | 0,33 | 14 | | |
| | 2 | 0,58 | 1,00 | 0,73 | 15 | | |
| MO | | 0,44 | 0,40 | 0,36 | 30 | | |
| AO | | 0,64 | 0,60 | 0,52 | 30 | | |

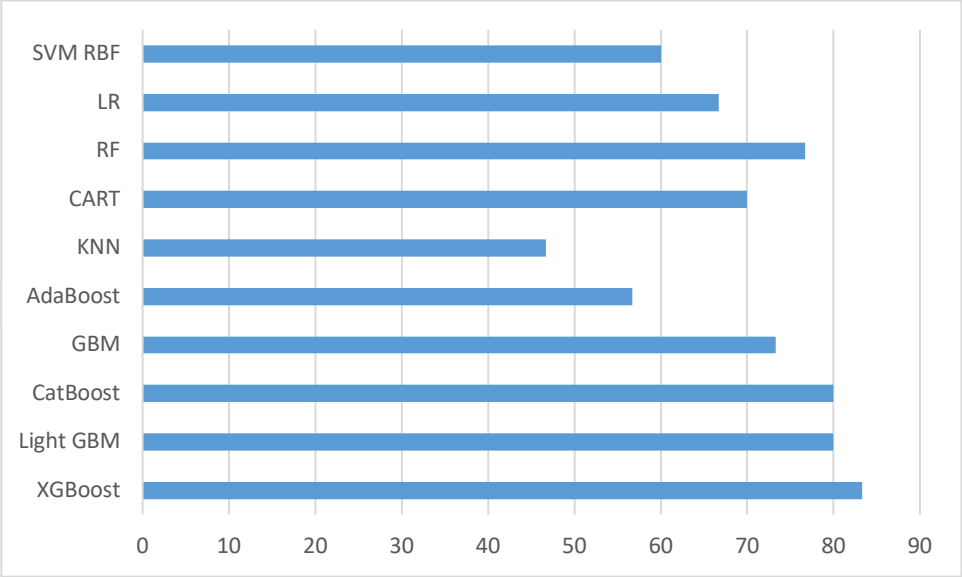
* Algoritmaların DO ve kappa değerleri içerisinde en yüksek değere sahip olanı göstermektedir.

Tablo 4.23'te yer alan sonuçlar Model 2.5'e ait olup, en iyi açıklayıcılığa sahip ilk 15 öznitelik seçilerek, veri setinin %90 eğitim ve %10 test olarak ayrılmasıyla elde edilmiştir. Yapılan sınıflandırmada doğruluk oranı değerlerine göre en iyi sınıflandırma %83,3 doğruluk oranı ile XGBoost algoritması tarafından gerçekleştirilmiştir. Kappa değerlerine göre Tablo 4.23'te yer alan veriler incelendiğinde XGBoost algoritması 0,687 değeri ile en iyi performansı sergilemiştir. Buna ek olarak XGBoost algoritması doğruluk oranı ve Kappa metrikleri dışındaki KO, DUO, F1 Skor metriklerine göre de en yüksek başarıyı göstermiştir.

**Tablo 4.24** Model 2.5'e ait sonuç karşılaştırma matrisi

| **Algoritma** | | **XGBoost** | | | **HGAM** | | | **CatBoost** | | | **GAM** | | | **AdaBoost** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **TES →** | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| **GS** | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| | 1 | 0 | 12 | 2 | 0 | 12 | 2 | 0 | 13 | 1 | 0 | 12 | 2 | 0 | 4 | 10 |
| | 2 | 0 | 3 | 12 | 0 | 4 | 11 | 0 | 4 | 11 | 0 | 6 | 9 | 0 | 3 | 12 |
| **Algoritma** | | **KNN** | | | **KA** | | | **RO** | | | **LR** | | | **DVM RBF** | | |
| **TES →** | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| **GS** | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| | 1 | 0 | 8 | 6 | 0 | 11 | 3 | 0 | 11 | 3 | 0 | 9 | 5 | 0 | 3 | 11 |
| | 2 | 0 | 9 | 6 | 0 | 6 | 9 | 0 | 4 | 11 | 0 | 4 | 11 | 0 | 0 | 15 |

Model 2.5'e ait sonuç karşılaştırma matrisi Tablo 4.24'te yer almaktadır. Tablo 4.24'e göre XGBoost algoritmasının toplam 30 örnek içerisinden 25 örneği doğru, 5 örneği ise hatalı tahmin ederek en başarılı sonucu verdiği söylenebilir. Şekil 4.10'da yer alan modellerin doğruluk oranları grafiği incelendiğinde, en başarılı algoritma XGBoost algoritması olarak öne çıkmaktadır. Doğruluk oranı 0,5'in altında kalan KNN algoritmasının ise en kötü performansı sergilediği görülmüştür.




**Şekil 4.10** Model 2.5'e göre sınıflandırma algoritmalarının doğruluk oranları

### 4.5.3 Açıklayıcılığı En Yüksek 10 Öznitelik ile Gerçekleştirilen Sınıflandırma Sonuçları

Bu bölümde, öznitelik seçim algoritmalarından yararlanılarak belirlenmiş olan açıklayıcılığı en yüksek ilk 10 öznitelik kullanılarak gerçekleştirilen sınıflandırma sonuçlarına yer verilecektir.

Öznitelik seçim algoritmalarından yararlanılarak belirlenen en iyi açıklayıcılığa sahip ilk 10 öznitelik şöyle sıralanabilir; bir günde ekran başında geçirilen süre, geleneksel/dijital bankacılık tercihi, mobil veya internet bankacılığı kullanımı, mevcutta ödenmekte olan kredi sayısı, müşterisi olunan banka sayısı, yaş, aylık birikim/gelir oranı, aylık net gelir, sık kullanılan bankacılık işlemleri, yatırım tercihleri şeklindedir.

**Tablo 4.25** Model 3.1 için gerçekleştirilen sınıflandırma sonuçları

| Algoritma | Sınıf | KO | DUO | F1 Skor | ÖS | DO | Kappa |
|---|---|---|---|---|---|---|---|
| **XGBoost** | 0 | 0,85 | 1,00 | 0,92 | 11 | 0,713 | 0,498 |
| | 1 | 0,67 | 0,74 | 0,70 | 69 | | |
| | 2 | 0,74 | 0,64 | 0,69 | 70 | | |
| MO | | 0,75 | 0,79 | 0,77 | 150 | | |
| AO | | 0,71 | 0,71 | 0,71 | 150 | | |
| **HGAM** | 0 | 1,00 | 1,00 | 1,00 | 11 | 0,700 | 0,470 |
| | 1 | 0,64 | 0,78 | 0,71 | 69 | | |
| | 2 | 0,73 | 0,57 | 0,64 | 70 | | |
| MO | | 0,71 | 0,70 | 0,70 | 150 | | |
| AO | | 0,81 | 0,80 | 0,80 | 150 | | |
| **CatBoost** | 0 | 1,00 | 1,00 | 1,00 | 11 | 0,747 | 0,552 |
| | 1 | 0,69 | 0,81 | 0,75 | 69 | | |
| | 2 | 0,78 | 0,64 | 0,70 | 70 | | |
| MO | | 0,82 | 0,82 | 0,82 | 150 | | |
| AO | | 0,75 | 0,75 | 0,74 | 150 | | |
| **GAM** | 0 | 0,92 | 1,00 | 0,96 | 11 | 0,733 | 0,531 |
| | 1 | 0,69 | 0,75 | 0,72 | 69 | | |
| | 2 | 0,75 | 0,67 | 0,71 | 70 | | |
| MO | | 0,79 | 0,81 | 0,80 | 150 | | |
| AO | | 0,73 | 0,73 | 0,73 | 150 | | |
| **AdaBoost** | 0 | 0,79 | 1,00 | 0,88 | 11 | 0,620 | 0,335 |
| | 1 | 0,65 | 0,38 | 0,48 | 69 | | |
| | 2 | 0,58 | 0,80 | 0,67 | 70 | | |
| MO | | 0,67 | 0,73 | 0,68 | 150 | | |
| AO | | 0,63 | 0,62 | 0,60 | 150 | | |
| **KNN** | 0 | 0,00 | 0,00 | 0,00 | 11 | 0,460 | 0,018 |
| | 1 | 0,47 | 0,54 | 0,50 | 69 | | |
| | 2 | 0,48 | 0,46 | 0,47 | 70 | | |
| MO | | 0,32 | 0,33 | 0,32 | 150 | | |
| AO | | 0,44 | 0,46 | 0,45 | 150 | | |

**Tablo 4.25** Model 3.1 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **KA** | 0 | 1,00 | 1,00 | 1,00 | 11 | 0,687 | 0,446 |
| | 1 | 0,66 | 0,65 | 0,66 | 69 | | |
| | 2 | 0,66 | 0,67 | 0,67 | 70 | | |
| MO | | 0,77 | 0,77 | 0,77 | 150 | | |
| AO | | 0,69 | 0,69 | 0,69 | 150 | | |
| **RO** | 0 | 1,00 | 1,00 | 1,00 | 11 | 0,793* | 0,635* |
| | 1 | 0,74 | 0,84 | 0,79 | 69 | | |
| | 2 | 0,82 | 0,71 | 0,76 | 70 | | |
| MO | | 0,85 | 0,85 | 0,85 | 150 | | |
| AO | | 0,80 | 0,79 | 0,79 | 150 | | |
| **LR** | 0 | 0,77 | 0,91 | 0,83 | 11 | 0,653 | 0,392 |
| | 1 | 0,63 | 0,64 | 0,63 | 69 | | |
| | 2 | 0,66 | 0,63 | 0,64 | 70 | | |
| MO | | 0,68 | 0,73 | 0,70 | 150 | | |
| AO | | 0,65 | 0,65 | 0,65 | 150 | | |
| **DVM RBF** | 0 | 0,00 | 0,00 | 0,00 | 11 | 0,467 | 0,012 |
| **Çekirdek** | 1 | 0,46 | 0,94 | 0,62 | 69 | | |
| | 2 | 0,50 | 0,07 | 0,12 | 70 | | |
| MO | | 0,32 | 0,34 | 0,25 | 150 | | |
| AO | | 0,45 | 0,47 | 0,34 | 150 | | |

* Algoritmaların DO ve kappa değerleri içerisinde en yüksek değere sahip olanı göstermektedir.

Tablo 4.25'te yer alan sonuçlar Model 3.1'e ait olup, en iyi açıklayıcılığa sahip ilk 10 öznitelik seçilerek, veri setinin %50 eğitim ve %50 test olarak ayrılmasıyla elde edilmiştir. Yapılan sınıflandırmada doğruluk oranı değerlerine göre en iyi sınıflandırma %79,3 doğruluk oranı ile RO algoritması tarafından gerçekleştirilmiştir. Tablo 4.25 incelendiğinde RO algoritmasından sonra en iyi doğruluk oranına sahip algoritma %74,7 doğruluk oranı ile CatBoost algoritması olduğu görülmektedir. Kappa değerlerine göre Tablo 4.25'te yer alan veriler

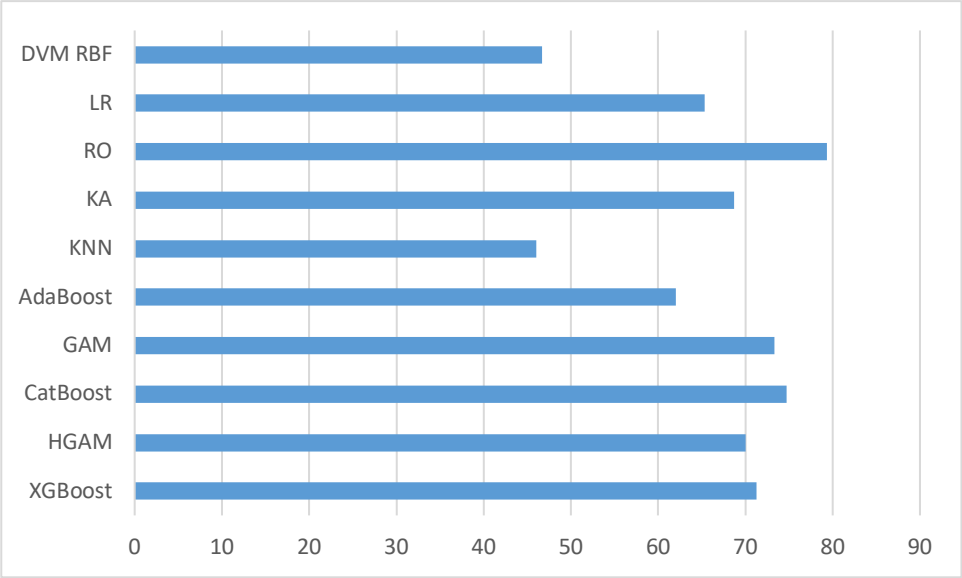
incelendiğinde RO algoritması 0,635 değeri ile en iyi performansı sergilemiştir. Buna ek olarak RO algoritması doğruluk oranı ve Kappa metrikleri dışındaki KO, DUO, F1 Skor metriklerine göre de en yüksek başarıyı göstermiştir.

**Tablo 4.26** Model 3.1'e ait sonuç karşılaştırma matrisi

| **Algoritma** | | **XGBoost** | | | **HGAM** | | | **CatBoost** | | | **GAM** | | | **AdaBoost** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **TES →** | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| **GS** | 0 | 11 | 0 | 0 | 11 | 0 | 0 | 11 | 0 | 0 | 11 | 0 | 0 | 11 | 0 | 0 |
| | 1 | 2 | 51 | 16 | 0 | 54 | 15 | 0 | 56 | 13 | 1 | 52 | 16 | 3 | 26 | 40 |
| | 2 | 0 | 25 | 45 | 0 | 30 | 40 | 0 | 25 | 45 | 0 | 23 | 47 | 0 | 14 | 56 |
| **Algoritma** | | **KNN** | | | **KA** | | | **RO** | | | **LR** | | | **DVM RBF** | | |
| **TES →** | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| **GS** | 0 | 0 | 7 | 4 | 11 | 0 | 0 | 11 | 0 | 0 | 10 | 1 | 0 | 0 | 10 | 1 |
| | 1 | 2 | 37 | 30 | 0 | 45 | 24 | 0 | 58 | 11 | 2 | 44 | 23 | 0 | 65 | 4 |
| | 2 | 3 | 35 | 32 | 0 | 23 | 47 | 0 | 20 | 50 | 1 | 25 | 44 | 0 | 65 | 5 |

TES: Tahmin edilen sınıf, GS: Gerçek sınıf

Tablo 4.26'da yer alan Model 3.1'e ait sonuç karşılaştırma matrisi incelendiğinde RO algoritmasının toplam 150 örnek içerisinden 119 örneği doğru, 31 örneği ise hatalı tahmin ederek en başarılı sonucu gösterdiği gözlenmektedir. Şekil 4.11'de yer alan modellerin doğruluk oranları grafiği incelendiğinde, en başarılı algoritma RO olurken, en iyi ikinci doğruluk oranının CatBoost algoritması tarafından gerçekleştirildiği gözlenmektedir.




**Şekil 4.11** Model 3.1'e göre sınıflandırma algoritmalarının doğruluk oranları

**Tablo 4.27** Model 3.2 için gerçekleştirilen sınıflandırma sonuçları

| Algoritma | Sınıf | KO | DUO | F1 Skor | ÖS | DO | Kappa |
|---|---|---|---|---|---|---|---|
| **XGBoost** | 0 | 1,00 | 0,89 | 0,94 | 9 | | |
| | 1 | 0,64 | 0,76 | 0,69 | 54 | | |
| | 2 | 0,73 | 0,61 | 0,67 | 57 | 0,700 | 0,469 |
| MO | | 0,79 | 0,75 | 0,77 | 120 | | |
| AO | | 0,71 | 0,70 | 0,70 | 120 | | |
| **HGAM** | 0 | 1,00 | 0,89 | 0,94 | 9 | | |
| | 1 | 0,61 | 0,76 | 0,68 | 54 | | |
| | 2 | 0,71 | 0,56 | 0,63 | 57 | 0,675 | 0,425 |
| MO | | 0,77 | 0,74 | 0,75 | 120 | | |
| AO | | 0,69 | 0,68 | 0,67 | 120 | | |
| **CatBoost** | 0 | 1,00 | 1,00 | 1,00 | 9 | | |
| | 1 | 0,68 | 0,83 | 0,75 | 54 | | |
| | 2 | 0,80 | 0,63 | 0,71 | 57 | 0,750 | 0,560 |
| MO | | 0,83 | 0,82 | 0,82 | 120 | | |
| AO | | 0,76 | 0,75 | 0,75 | 120 | | |

**Tablo 4.27** Model 3.2 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **GAM** | 0 | 1,00 | 0,89 | 0,94 | 9 | 0,733 | 0,528 |
| | 1 | 0,67 | 0,80 | 0,73 | 54 | | |
| | 2 | 0,77 | 0,65 | 0,70 | 57 | | |
| MO | | 0,81 | 0,78 | 0,79 | 120 | | |
| AO | | 0,74 | 0,73 | 0,73 | 120 | | |
| **AdaBoost** | 0 | 1,00 | 1,00 | 1,00 | 9 | 0,633 | 0,346 |
| | 1 | 0,69 | 0,33 | 0,45 | 54 | | |
| | 2 | 0,58 | 0,86 | 0,69 | 57 | | |
| MO | | 0,76 | 0,73 | 0,71 | 120 | | |
| AO | | 0,66 | 0,63 | 0,61 | 120 | | |
| **KNN** | 0 | 0,00 | 0,00 | 0,00 | 9 | 0,458 | 0,014 |
| | 1 | 0,46 | 0,57 | 0,51 | 54 | | |
| | 2 | 0,49 | 0,42 | 0,45 | 57 | | |
| MO | | 0,32 | 0,33 | 0,32 | 120 | | |
| AO | | 0,44 | 0,46 | 0,44 | 120 | | |
| **KA** | 0 | 1,00 | 1,00 | 1,00 | 9 | 0,708 | 0,484 |
| | 1 | 0,69 | 0,65 | 0,67 | 54 | | |
| | 2 | 0,68 | 0,72 | 0,70 | 57 | | |
| MO | | 0,79 | 0,79 | 0,79 | 120 | | |
| AO | | 0,71 | 0,71 | 0,71 | 120 | | |
| **RO** | 0 | 1,00 | 1,00 | 1,00 | 9 | 0,767* | 0,589* |
| | 1 | 0,70 | 0,83 | 0,76 | 54 | | |
| | 2 | 0,81 | 0,67 | 0,73 | 57 | | |
| MO | | 0,84 | 0,83 | 0,83 | 120 | | |
| AO | | 0,78 | 0,77 | 0,77 | 120 | | |
| **LR** | 0 | 0,70 | 0,78 | 0,74 | 9 | 0,617 | 0,320 |
| | 1 | 0,57 | 0,69 | 0,62 | 54 | | |
| | 2 | 0,67 | 0,53 | 0,59 | 57 | | |
| MO | | 0,65 | 0,66 | 0,65 | 120 | | |
| AO | | 0,63 | 0,62 | 0,61 | 120 | | |

**Tablo 4.27** Model 3.2 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **DVM RBF** | 0 | 0,00 | 0,00 | 0,00 | 9 | 0,467 | 0,012 |
| | 1 | 0,44 | 0,93 | 0,60 | 54 | | |
| | 2 | 0,33 | 0,04 | 0,06 | 57 | | |
| MO | | 0,26 | 0,32 | 0,22 | 120 | | |
| AO | | 0,36 | 0,43 | 0,30 | 120 | | |

* Algoritmaların DO ve kappa değerleri içerisinde en yüksek değere sahip olanı göstermektedir.

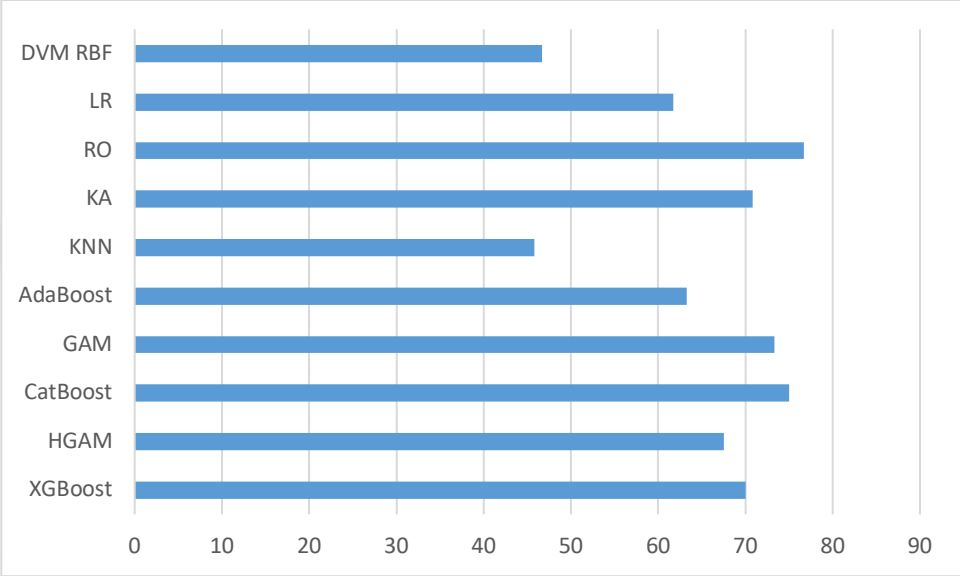
Veri seti içerisinde en iyi açıklayıcılığa sahip ilk 10 öznitelik belirlenerek, %60 eğitim ve %40 test olarak ayrılmasıyla çalışılan Model 3.2'ye ait sınıflandırma sonuçları Tablo 4.27'de yer almaktadır. Sınıflandırma sonuçları doğruluk oranı değerlerine göre incelendiğinde RO algoritması %76,7 doğruluk oranı en iyi sınıflandırma sonucunu gerçekleştirmiştir. Tablo 4.27 Kappa değerlerine göre incelendiğinde, RO algoritması 0,589 değeri ile en iyi performansı sergilemiştir. Buna ek olarak RO algoritması doğruluk oranı ve Kappa metrikleri dışındaki KO, DUO, F1 Skor metriklerine göre de en yüksek başarıyı göstermiştir.

**Tablo 4.28** Model 3.2'ye ait sonuç karşılaştırma matrisi

| **Algoritma** | | **XGBoost** | | | **HGAM** | | | **CatBoost** | | | **GAM** | | | **AdaBoost** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **TES →** | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| **GS** | 0 | 8 | 1 | 0 | 8 | 1 | 0 | 9 | 0 | 0 | 8 | 1 | 0 | 9 | 0 | 0 |
| | 1 | 0 | 41 | 13 | 0 | 41 | 13 | 0 | 45 | 9 | 0 | 43 | 11 | 0 | 18 | 36 |
| | 2 | 0 | 22 | 35 | 0 | 25 | 32 | 0 | 21 | 36 | 0 | 20 | 37 | 0 | 8 | 49 |
| **Algoritma** | | **KNN** | | | **KA** | | | **RO** | | | **LR** | | | **DVM RBF** | | |
| **TES →** | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| **GS** | 0 | 0 | 6 | 3 | 9 | 0 | 0 | 9 | 0 | 0 | 7 | 2 | 0 | 0 | 9 | 0 |
| | 1 | 1 | 31 | 22 | 0 | 35 | 19 | 0 | 45 | 9 | 2 | 37 | 15 | 0 | 50 | 4 |
| | 2 | 2 | 31 | 24 | 0 | 16 | 41 | 0 | 19 | 38 | 1 | 26 | 30 | 0 | 55 | 2 |

TES: Tahmin edilen sınıf, GS: Gerçek sınıf

Model 3.2'ye ait sonuç karşılaştırma matrisi Tablo 4.28'de yer almaktadır. Tablo 4.28'e göre RO algoritmasının toplam 120 örnek içerisinden 92 örneği doğru, 28 örneği ise hatalı tahmin ederek en başarılı sonucu verdiği söylenebilir. Şekil 4.12'de yer alan modellerin doğruluk oranları grafiği incelendiğinde, en başarılı algoritma RO algoritması olarak öne çıkmaktadır. Doğruluk oranı 0,5'in altında kalan KNN ve DVM SBF Çekirdek algoritmaları ise en kötü performansları sergilemiştir.




**Şekil 4.12** Model 3.2'ye göre sınıflandırma algoritmalarının doğruluk oranları

**Tablo 4.29** Model 3.3 için gerçekleştirilen sınıflandırma sonuçları

| Algoritma | Sınıf | KO | DUO | F1 Skor | ÖS | DO | Kappa |
|---|---|---|---|---|---|---|---|
| **XGBoost** | 0 | 1,00 | 0,71 | 0,83 | 7 | | |
| | 1 | 0,67 | 0,87 | 0,76 | 39 | | |
| | 2 | 0,85 | 0,66 | 0,74 | 44 | 0,756 | 0,568 |
| MO | | 0,84 | 0,75 | 0,78 | 90 | | |
| AO | | 0,78 | 0,76 | 0,76 | 90 | | |
| **HGAM** | 0 | 1,00 | 1,00 | 1,00 | 7 | | |
| | 1 | 0,62 | 0,85 | 0,72 | 39 | | |
| | 2 | 0,80 | 0,55 | 0,65 | 44 | 0,711 | 0,498 |
| MO | | 0,81 | 0,80 | 0,79 | 90 | | |
| AO | | 0,74 | 0,71 | 0,71 | 90 | | |

**Tablo 4.29** Model 3.3 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **CatBoost** | 0 | 1,00 | 0,86 | 0,92 | 7 | 0,722 | 0,514 |
| | 1 | 0,63 | 0,87 | 0,73 | 39 | | |
| | 2 | 0,83 | 0,57 | 0,68 | 44 | | |
| MO | | 0,82 | 0,77 | 0,78 | 90 | | |
| AO | | 0,76 | 0,72 | 0,72 | 90 | | |
| **GAM** | 0 | 1,00 | 0,71 | 0,83 | 7 | 0,733 | 0,527 |
| | 1 | 0,65 | 0,82 | 0,73 | 39 | | |
| | 2 | 0,81 | 0,66 | 0,73 | 44 | | |
| MO | | 0,82 | 0,73 | 0,76 | 90 | | |
| AO | | 0,75 | 0,73 | 0,73 | 90 | | |
| **AdaBoost** | 0 | 1,00 | 0,71 | 0,83 | 7 | 0,622 | 0,309 |
| | 1 | 0,62 | 0,33 | 0,43 | 39 | | |
| | 2 | 0,59 | 0,86 | 0,70 | 44 | | |
| MO | | 0,74 | 0,64 | 0,66 | 90 | | |
| AO | | 0,64 | 0,62 | 0,60 | 90 | | |
| **KNN** | 0 | 0,33 | 0,14 | 0,20 | 7 | 0,511 | 0,127 |
| | 1 | 0,49 | 0,72 | 0,58 | 39 | | |
| | 2 | 0,57 | 0,39 | 0,46 | 44 | | |
| MO | | 0,46 | 0,42 | 0,41 | 90 | | |
| AO | | 0,52 | 0,51 | 0,49 | 90 | | |
| **KA** | 0 | 1,00 | 1,00 | 1,00 | 7 | 0,678 | 0,436 |
| | 1 | 0,61 | 0,72 | 0,66 | 39 | | |
| | 2 | 0,70 | 0,59 | 0,64 | 44 | | |
| MO | | 0,77 | 0,77 | 0,77 | 90 | | |
| AO | | 0,69 | 0,68 | 0,68 | 90 | | |
| **RO** | 0 | 1,00 | 1,00 | 1,00 | 7 | 0,767* | 0,594* |
| | 1 | 0,67 | 0,90 | 0,77 | 39 | | |
| | 2 | 0,87 | 0,61 | 0,72 | 44 | | |
| MO | | 0,85 | 0,84 | 0,83 | 90 | | |
| AO | | 0,80 | 0,77 | 0,76 | 90 | | |

**Tablo 4.29** Model 3.3 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **LR** | 0 | 0,62 | 0,71 | 0,67 | 7 | 0,689 | 0,459 |
| | 1 | 0,64 | 0,72 | 0,67 | 39 | | |
| | 2 | 0,76 | 0,66 | 0,71 | 44 | | |
| MO | | 0,67 | 0,70 | 0,68 | 90 | | |
| AO | | 0,70 | 0,69 | 0,69 | 90 | | |
| **DVM RBF** | 0 | 0,00 | 0,00 | 0,00 | 7 | 0,422 | -0,023 |
| | 1 | 0,43 | 0,95 | 0,59 | 39 | | |
| | 2 | 0,33 | 0,02 | 0,04 | 44 | | |
| MO | | 0,25 | 0,32 | 0,21 | 90 | | |
| AO | | 0,35 | 0,42 | 0,28 | 90 | | |

* Algoritmaların DO ve kappa değerleri içerisinde en yüksek değere sahip olanı göstermektedir.

Tablo 4.29'da yer alan sonuçlar Model 3.3'e ait olup, en iyi açıklayıcılığa sahip ilk 10 öznitelik seçilerek, veri setinin %70 eğitim ve %30 test olarak ayrılmasıyla elde edilmiştir. Yapılan sınıflandırmada doğruluk oranı değerlerine göre en iyi sınıflandırma %76,7 doğruluk oranı ile RO algoritması tarafından gerçekleştirilmiştir. Tablo 4.29 incelendiğinde RO algoritmasından sonra en iyi doğruluk oranına sahip algoritma %75,6 doğruluk oranı ile XGBoost algoritması olduğu görülmektedir. Kappa değerlerine göre Tablo 4.29'da yer alan veriler incelendiğinde RO algoritması 0,594 değeri ile en iyi performansı sergilemiştir. Buna ek olarak RO algoritması doğruluk oranı ve Kappa metrikleri dışındaki KO, DUO, F1 Skor metriklerine göre de en yüksek başarıyı göstermiştir.

**Tablo 4.30** Model 3.3'e ait sonuç karşılaştırma matrisi

| **Algoritma** | | **XGBoost** | | | **HGAM** | | | **CatBoost** | | | **GAM** | | | **AdaBoost** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **TES →** | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| **GS** | 0 | 5 | 2 | 0 | 7 | 0 | 0 | 6 | 1 | 0 | 5 | 2 | 0 | 5 | 2 | 0 |
| | 1 | 0 | 34 | 5 | 0 | 33 | 6 | 0 | 34 | 5 | 0 | 32 | 7 | 0 | 13 | 26 |
| | 2 | 0 | 15 | 29 | 0 | 20 | 24 | 0 | 19 | 25 | 0 | 15 | 29 | 0 | 6 | 38 |

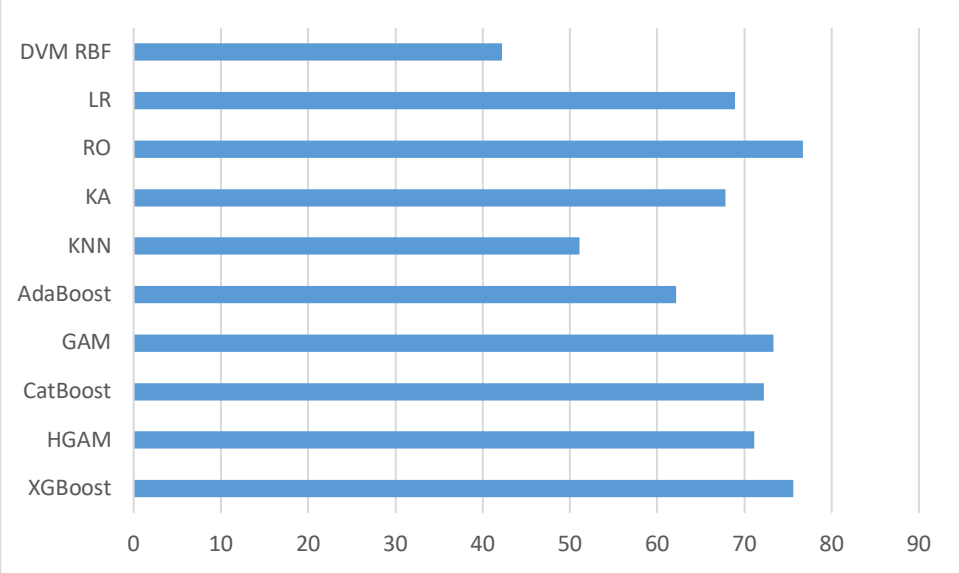
**Tablo 4.30** Model 3.3'e ait sonuç karşılaştırma matrisi (devamı)

| Algoritma | | KNN | | | KA | | | RO | | | LR | | | DVM RBF | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TES → | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| GS | 0 | 1 | 3 | 3 | 7 | 0 | 0 | 7 | 0 | 0 | 5 | 2 | 0 | 0 | 7 | 0 |
| | 1 | 1 | 28 | 10 | 0 | 28 | 11 | 0 | 35 | 4 | 2 | 28 | 9 | 0 | 37 | 2 |
| | 2 | 1 | 26 | 17 | 0 | 18 | 26 | 0 | 17 | 27 | 1 | 14 | 29 | 0 | 43 | 1 |

TES: Tahmin edilen sınıf, GS: Gerçek sınıf

Tablo 4.30'da yer alan Model 3.3'e ait sonuç karşılaştırma matrisi incelendiğinde RO algoritmasının toplam 90 örnek içerisinden 69 örneği doğru, 21 örneği ise hatalı tahmin ederek en başarılı sonucu gösterdiği gözlenmektedir.

Şekil 4.13'te yer alan modellerin doğruluk oranları grafiği incelendiğinde, en başarılı algoritma RO olurken, en iyi ikinci doğruluk oranının XGBoost algoritması tarafından gerçekleştirildiği gözlenmektedir.




**Şekil 4.13** Model 3.3'e göre sınıflandırma algoritmalarının doğruluk oranları

**Tablo 4.31** Model 3.4 için gerçekleştirilen sınıflandırma sonuçları

| Algoritma | Sınıf | KO | DUO | F1 Skor | ÖS | DO | Kappa |
|---|---|---|---|---|---|---|---|
| **XGBoost** | 0 | 1,00 | 1,00 | 1,00 | 3 | 0,783 | 0,603 |
| | 1 | 0,79 | 0,76 | 0,77 | 29 | | |
| | 2 | 0,76 | 0,79 | 0,77 | 28 | | |
| MO | | 0,85 | 0,85 | 0,85 | 60 | | |
| AO | | 0,76 | 0,78 | 0,78 | 60 | | |
| **HGAM** | 0 | 1,00 | 1,00 | 1,00 | 3 | 0,783 | 0,603 |
| | 1 | 0,75 | 0,83 | 0,79 | 29 | | |
| | 2 | 0,80 | 0,71 | 0,75 | 28 | | |
| MO | | 0,85 | 0,85 | 0,85 | 60 | | |
| AO | | 0,79 | 0,78 | 0,78 | 60 | | |
| **CatBoost** | 0 | 1,00 | 0,67 | 0,80 | 3 | 0,800 | 0,628 |
| | 1 | 0,74 | 0,90 | 0,81 | 29 | | |
| | 2 | 0,87 | 0,71 | 0,78 | 28 | | |
| MO | | 0,87 | 0,76 | 0,80 | 60 | | |
| AO | | 0,81 | 0,80 | 0,80 | 60 | | |
| **GAM** | 0 | 1,00 | 1,00 | 1,00 | 3 | 0,817 | 0,664 |
| | 1 | 0,78 | 0,86 | 0,82 | 29 | | |
| | 2 | 0,84 | 0,75 | 0,79 | 28 | | |
| MO | | 0,87 | 0,87 | 0,87 | 60 | | |
| AO | | 0,82 | 0,82 | 0,82 | 60 | | |
| **AdaBoost** | 0 | 0,50 | 1,00 | 0,67 | 3 | 0,600 | 0,301 |
| | 1 | 0,73 | 0,28 | 0,40 | 29 | | |
| | 2 | 0,58 | 0,89 | 0,70 | 28 | | |
| MO | | 0,60 | 0,72 | 0,59 | 60 | | |
| AO | | 0,65 | 0,60 | 0,56 | 60 | | |
| **KNN** | 0 | 0,00 | 0,00 | 0,00 | 3 | 0,533 | 0,133 |
| | 1 | 0,56 | 0,62 | 0,59 | 29 | | |
| | 2 | 0,54 | 0,50 | 0,52 | 28 | | |
| MO | | 0,37 | 0,37 | 0,37 | 60 | | |
| AO | | 0,52 | 0,53 | 0,53 | 60 | | |

**Tablo 4.31** Model 3.4 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **KA** | 0 | 1,00 | 1,00 | 1,00 | 3 | 0,783 | 0,603 |
| | 1 | 0,77 | 0,79 | 0,78 | 29 | | |
| | 2 | 0,78 | 0,75 | 0,76 | 28 | | |
| MO | | 0,85 | 0,85 | 0,85 | 60 | | |
| AO | | 0,78 | 0,78 | 0,78 | 60 | | |
| **RO** | 0 | 1,00 | 1,00 | 1,00 | 3 | 0,850* | 0,725* |
| | 1 | 0,81 | 0,90 | 0,85 | 29 | | |
| | 2 | 0,88 | 0,79 | 0,83 | 28 | | |
| MO | | 0,90 | 0,89 | 0,89 | 60 | | |
| AO | | 0,85 | 0,85 | 0,85 | 60 | | |
| **LR** | 0 | 0,67 | 0,67 | 0,67 | 3 | 0,733 | 0,512 |
| | 1 | 0,72 | 0,72 | 0,72 | 29 | | |
| | 2 | 0,75 | 0,75 | 0,75 | 28 | | |
| MO | | 0,71 | 0,71 | 0,71 | 60 | | |
| AO | | 0,73 | 0,73 | 0,73 | 60 | | |
| **DVM RBF** | 0 | 0,00 | 0,00 | 0,00 | 3 | 0,567 | 0,185 |
| | 1 | 0,86 | 0,21 | 0,33 | 29 | | |
| | 2 | 0,53 | 1,00 | 0,69 | 28 | | |
| MO | | 0,46 | 0,40 | 0,34 | 60 | | |
| AO | | 0,66 | 0,57 | 0,48 | 60 | | |

* Algoritmaların DO ve kappa değerleri içerisinde en yüksek değere sahip olanı göstermektedir.

Veri seti içerisinde en iyi açıklayıcılığa sahip ilk 10 öznitelik belirlenerek, %80 eğitim ve %20 test olarak ayrılmasıyla çalışılan Tasarım Modeli 3.4'e ait sınıflandırma sonuçları Tablo 4.31'de yer almaktadır. Sınıflandırma sonuçları doğruluk oranı değerlerine göre incelendiğinde RO algoritması %85 doğruluk oranı ile tüm tasarım modelleri arasında en iyi sınıflandırma sonucunu gerçekleştirmiştir. Tablo 4.31 Kappa değerlerine göre incelendiğinde, Random

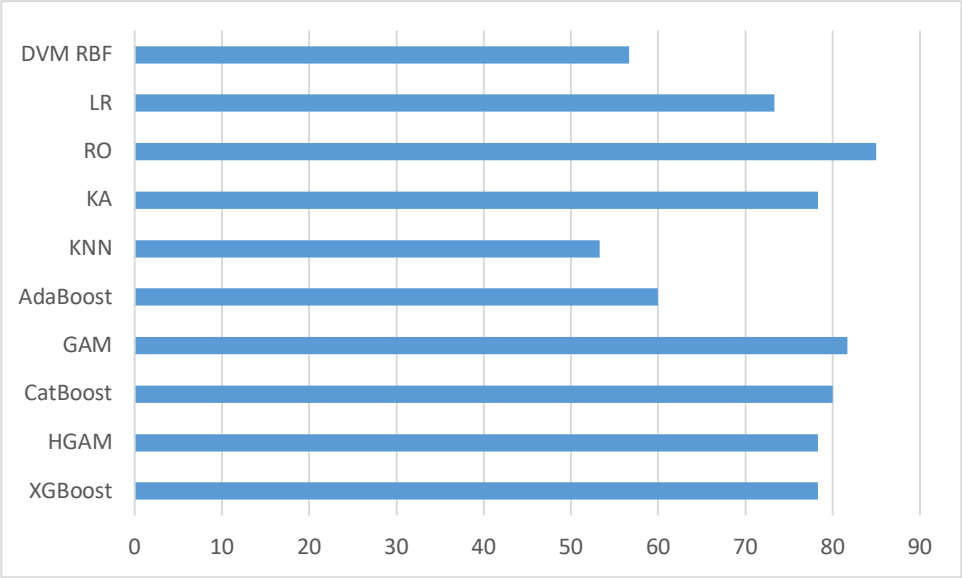
Forest algoritması 0,725 değeri ile en iyi performansı sergilemiştir. Buna ek olarak RO algoritması doğruluk oranı ve Kappa metrikleri dışındaki KO, DUO, F1 Skor metriklerine göre de tüm çalışma kapsamında en yüksek başarıyı göstermiştir.

**Tablo 4.32** Model 3.4'e ait sonuç karşılaştırma matrisi

| Algoritma | | XGBoost | | | HGAM | | | CatBoost | | | GAM | | | AdaBoost | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TES → | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| GS | 0 | 3 | 0 | 0 | 3 | 0 | 0 | 2 | 1 | 0 | 3 | 0 | 0 | 3 | 0 | 0 |
| | 1 | 0 | 22 | 7 | 0 | 24 | 5 | 0 | 26 | 3 | 0 | 25 | 4 | 3 | 8 | 18 |
| | 2 | 0 | 6 | 22 | 0 | 8 | 20 | 0 | 8 | 20 | 0 | 7 | 21 | 0 | 3 | 25 |
| Algoritma | | KNN | | | KA | | | RO | | | LR | | | DVM RBF | | |
| TES → | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| GS | 0 | 0 | 1 | 2 | 3 | 0 | 0 | 3 | 0 | 0 | 2 | 1 | 0 | 0 | 1 | 2 |
| | 1 | 1 | 18 | 10 | 0 | 23 | 6 | 0 | 26 | 3 | 1 | 21 | 7 | 0 | 6 | 23 |
| | 2 | 1 | 13 | 14 | 0 | 7 | 21 | 0 | 6 | 22 | 0 | 7 | 21 | 0 | 0 | 28 |

TES: Tahmin edilen sınıf, GS: Gerçek sınıf

Model 3.4'e ait sonuç karşılaştırma matrisi Tablo 4.32'da yer almaktadır. Tablo 4.32'ye göre RO algoritmasının toplam 60 örnek içerisinden 51 örneği doğru, 9 örneği ise hatalı tahmin ederek en başarılı sonucu verdiği söylenebilir. Şekil 4.14'te yer alan modellerin doğruluk oranları grafiği incelendiğinde, en başarılı algoritma RO algoritması olarak öne çıkmaktadır.




**Şekil 4.14** Model 3.4'e göre sınıflandırma algoritmalarının doğruluk oranları

**Tablo 4.33** Model 3.5 için gerçekleştirilen sınıflandırma sonuçları

| Algoritma | Sınıf | KO | DUO | F1 Skor | ÖS | DO | Kappa |
|---|---|---|---|---|---|---|---|
| **XGBoost** | 0 | 1,00 | 1,00 | 1,00 | 1 | | |
| | 1 | 0,80 | 0,86 | 0,83 | 14 | | |
| | 2 | 0,86 | 0,80 | 0,83 | 15 | 0,833* | 0,687* |
| MO | | 0,89 | 0,89 | 0,89 | 30 | | |
| AO | | 0,84 | 0,83 | 0,83 | 30 | | |
| **HGAM** | 0 | 1,00 | 1,00 | 1,00 | 1 | | |
| | 1 | 0,67 | 0,71 | 0,69 | 14 | | |
| | 2 | 0,71 | 0,67 | 0,69 | 15 | 0,700 | 0,436 |
| MO | | 0,79 | 0,79 | 0,79 | 30 | | |
| AO | | 0,70 | 0,70 | 0,70 | 30 | | |
| **CatBoost** | 0 | 1,00 | 1,00 | 1,00 | 1 | | |
| | 1 | 0,75 | 0,86 | 0,80 | 14 | | |
| | 2 | 0,85 | 0,73 | 0,79 | 15 | 0,800 | 0,625 |
| MO | | 0,87 | 0,86 | 0,86 | 30 | | |
| AO | | 0,81 | 0,80 | 0,80 | 30 | | |

**Tablo 4.33** Model 3.5 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **GAM** | 0 | 1,00 | 1,00 | 1,00 | 1 | 0,767 | 0,563 |
| | 1 | 0,71 | 0,86 | 0,77 | 14 | | |
| | 2 | 0,83 | 0,67 | 0,74 | 15 | | |
| MO | | 0,85 | 0,84 | 0,84 | 30 | | |
| AO | | 0,78 | 0,77 | 0,76 | 30 | | |
| **AdaBoost** | 0 | 1,00 | 1,00 | 1,00 | 1 | 0,600 | 0,234 |
| | 1 | 0,67 | 0,29 | 0,40 | 14 | | |
| | 2 | 0,57 | 0,87 | 0,68 | 15 | | |
| MO | | 0,74 | 0,72 | 0,69 | 30 | | |
| AO | | 0,63 | 0,60 | 0,56 | 30 | | |
| **KNN** | 0 | 0,00 | 0,00 | 0,00 | 1 | 0,500 | 0,041 |
| | 1 | 0,47 | 0,64 | 0,55 | 14 | | |
| | 2 | 0,55 | 0,40 | 0,46 | 15 | | |
| MO | | 0,34 | 0,35 | 0,34 | 30 | | |
| AO | | 0,49 | 0,50 | 0,49 | 30 | | |
| **KA** | 0 | 1,00 | 1,00 | 1,00 | 1 | 0,733 | 0,498 |
| | 1 | 0,71 | 0,71 | 0,71 | 14 | | |
| | 2 | 0,73 | 0,73 | 0,73 | 15 | | |
| MO | | 0,82 | 0,82 | 0,82 | 30 | | |
| AO | | 0,73 | 0,73 | 0,73 | 30 | | |
| **RO** | 0 | 1,00 | 1,00 | 1,00 | 1 | 0,833* | 0,688 |
| | 1 | 0,76 | 0,93 | 0,84 | 14 | | |
| | 2 | 0,92 | 0,73 | 0,81 | 15 | | |
| MO | | 0,89 | 0,89 | 0,88 | 60 | | |
| AO | | 0,85 | 0,83 | 0,83 | 60 | | |
| **LR** | 0 | 0,00 | 0,00 | 0,00 | 1 | 0,700 | 0,417 |
| | 1 | 0,69 | 0,64 | 0,67 | 14 | | |
| | 2 | 0,71 | 0,80 | 0,75 | 15 | | |
| MO | | 0,47 | 0,48 | 0,47 | 30 | | |
| AO | | 0,68 | 0,70 | 0,69 | 30 | | |

**Tablo 4.33** Model 3.5 için gerçekleştirilen sınıflandırma sonuçları (devamı)

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **DVM** | 0 | 0,00 | 0,00 | 0,00 | 1 | | |
| **RBF** | 1 | 0,75 | 0,21 | 0,33 | 14 | | |
| | 2 | 0,58 | 1,00 | 0,73 | 15 | 0,600 | 0,207 |
| MO | | 0,44 | 0,40 | 0,36 | 30 | | |
| AO | | 0,64 | 0,60 | 0,52 | 30 | | |

* Algoritmaların DO ve kappa değerleri içerisinde en yüksek değere sahip olanı göstermektedir.

Tablo 4.33'te yer alan sonuçlar Model 3.5'e ait olup, en iyi açıklayıcılığa sahip ilk 10 öznitelik seçilerek, veri setinin %90 eğitim ve %10 test olarak ayrılmasıyla elde edilmiştir. Yapılan sınıflandırmada doğruluk oranı değerlerine göre en iyi sınıflandırma %83,3 DO ile XGBoost ve RO algoritmaları tarafından gerçekleştirilmiştir. Tablo 4.33 incelendiğinde XGBoost ve RO algoritmasından sonra en iyi doğruluk oranına sahip algoritma %80 doğruluk oranı ile CatBoost algoritması olduğu görülmektedir. Kappa değerlerine göre Tablo 4.33'te yer alan veriler incelendiğinde RO algoritması 0,688 değeri ile en iyi performansı sergilemiştir, XGBoost algoritması ise Kappa değerinde ikinci sırada ve 0,687 değerini göstermiştir. Buna ek olarak RO algoritması doğruluk oranı ve Kappa metrikleri dışındaki KO, DUO, F1 Skor metriklerine göre de en yüksek başarıyı göstermiştir.

**Tablo 4.34** Model 3.5'e ait sonuç karşılaştırma matrisi

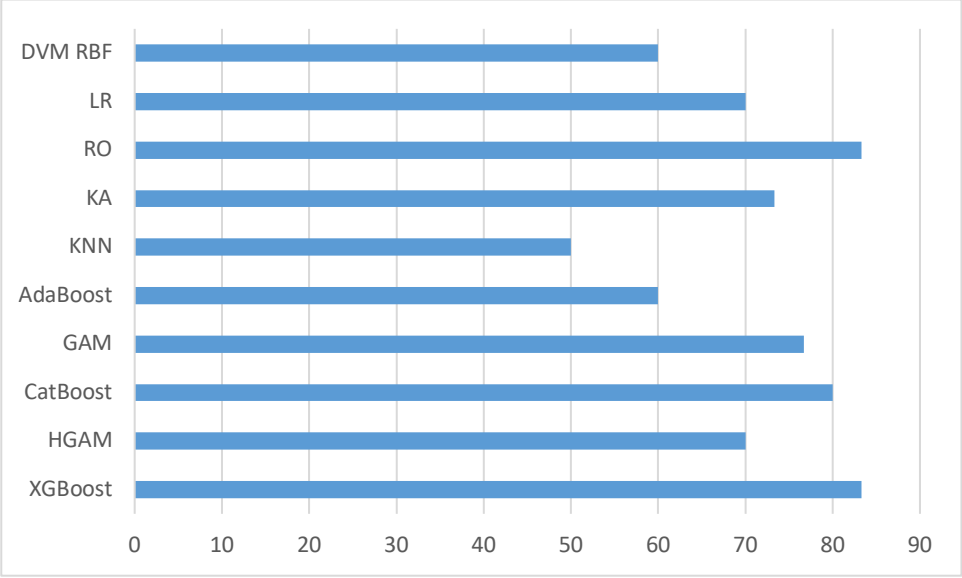
| **Algoritma** | | **XGBoost** | | | **HGAM** | | | **CatBoost** | | | **GAM** | | | **AdaBoost** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **TES →** | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| **GS** | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| | 1 | 0 | 12 | 2 | 0 | 10 | 4 | 0 | 12 | 2 | 0 | 12 | 2 | 0 | 4 | 10 |
| | 2 | 0 | 3 | 12 | 0 | 5 | 10 | 0 | 4 | 11 | 0 | 5 | 10 | 0 | 2 | 13 |

**Tablo 4.34** Model 3.5'e ait sonuç karşılaştırma matrisi (devamı)

| Algoritma | | KNN | | | KA | | | RO | | | LR | | | DVM RBF | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| TES → | | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 | 0 | 1 | 2 |
| GS | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| | 1 | 0 | 9 | 5 | 0 | 10 | 4 | 0 | 13 | 1 | 0 | 9 | 5 | 0 | 3 | 11 |
| | 2 | 0 | 9 | 6 | 0 | 4 | 11 | 0 | 4 | 11 | 0 | 3 | 12 | 0 | 0 | 15 |

TES: Tahmin edilen sınıf, GS: Gerçek sınıf

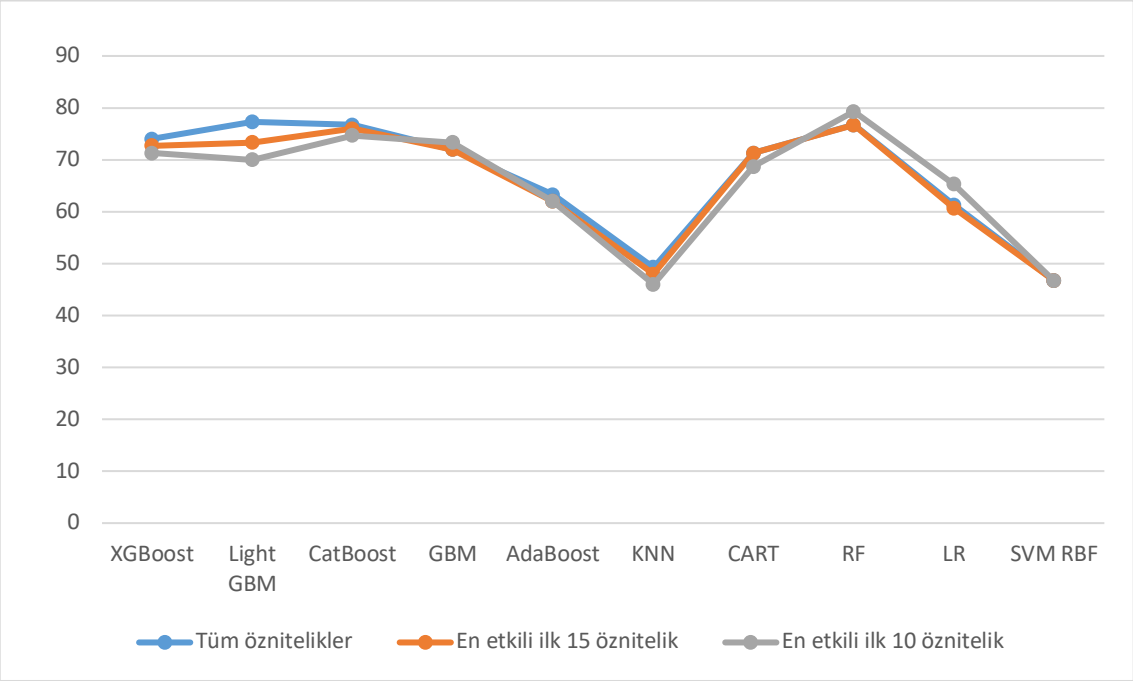
Tablo 4.34'te yer alan Model 3.5'e ait sonuç karşılaştırma matrisi incelendiğinde RO ve XGBoost algoritmalarının toplam 30 örnek içerisinden 25 örneği doğru, 5 örneği ise hatalı tahmin ederek en başarılı sonucu gösterdikleri gözlenmektedir. Şekil 4.15'te yer alan modellerin doğruluk oranları grafiği incelendiğinde, en başarılı algoritma RO ve XGBoost algoritmaları olurken, tarafından gerçekleştirildiği gözlenmektedir.




**Şekil 4.15** Model 3.5'e göre sınıflandırma algoritmalarının doğruluk oranları

## 4.5 Sonuçların Değerlendirilmesi

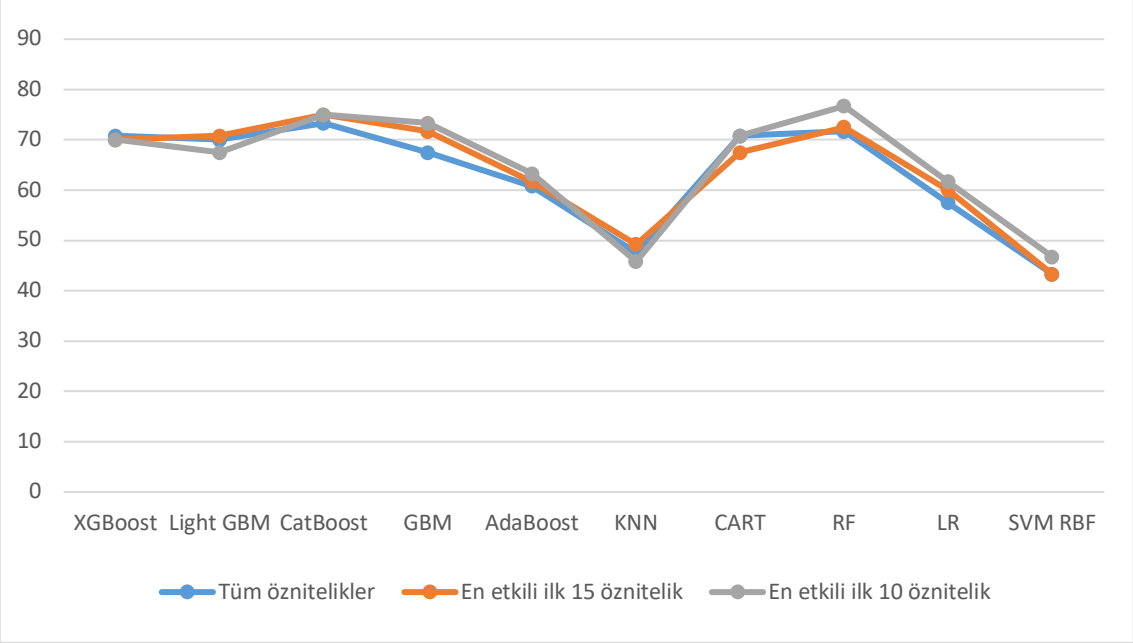
Bu kısımda, çalışmada ele alınan modeller tüm öznitelikler, en etkili ilk on beş öznitelik ve en etkili ilk on öznitelik için uygulanan eğitim ve test bölünmelerine göre ayrı ayrı grafiklerle ele alınarak kıyaslanmıştır.




**Şekil 4.16** Veri setinin %50 test olarak kullanılması durumunda DO değerlerine göre tüm sınıflandırma sonuçları

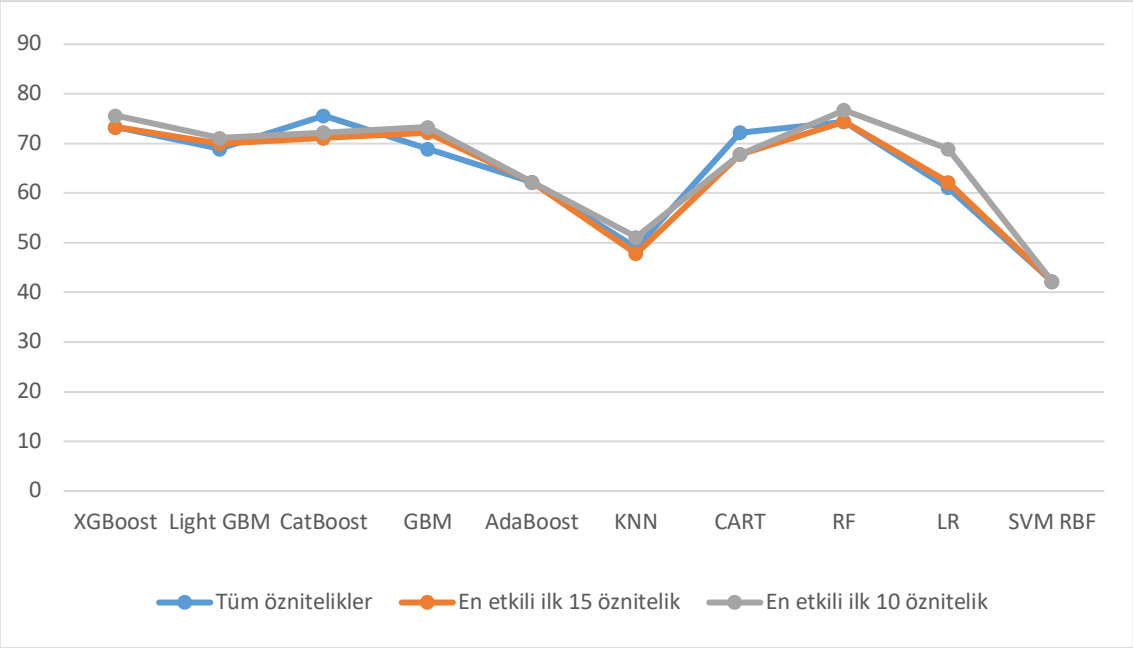
Şekil 4.16'da veri setinin %50 eğitim %50 test olarak ayrılarak çalışıldığı senaryoda üç farklı öznitelik grubu için elde edilen doğruluk oranları grafiği yer almaktadır. KNN ve DVM RBF Çekirdek algoritmalarının tüm öznitelik grupları için bu eğitim modelinde diğer modellere göre daha başarısız olduğu ve doğruluk oranının 0,5 değerinin altında kaldığı gözlenmektedir. RO algoritması, en iyi 10 öznitelik ile gerçekleştirilen analiz sonucunda %50 eğitim %50 test senaryoları içerisinde %79,3 doğruluk oranı ile en başarılı sonucu göstermiştir. Ardından HGAM algoritması tüm özniteliklerle yapılan analizde %77,3 doğruluk oranı ile ikinci sırayı almaktadır. CatBoost ve RO algoritmaları tüm özniteliklerle

gerçekleştirilen analizler sonucunda %76,7 doğruluk oranı ile üçüncü en iyi sınıflandırma sonucunu elde etmişlerdir.



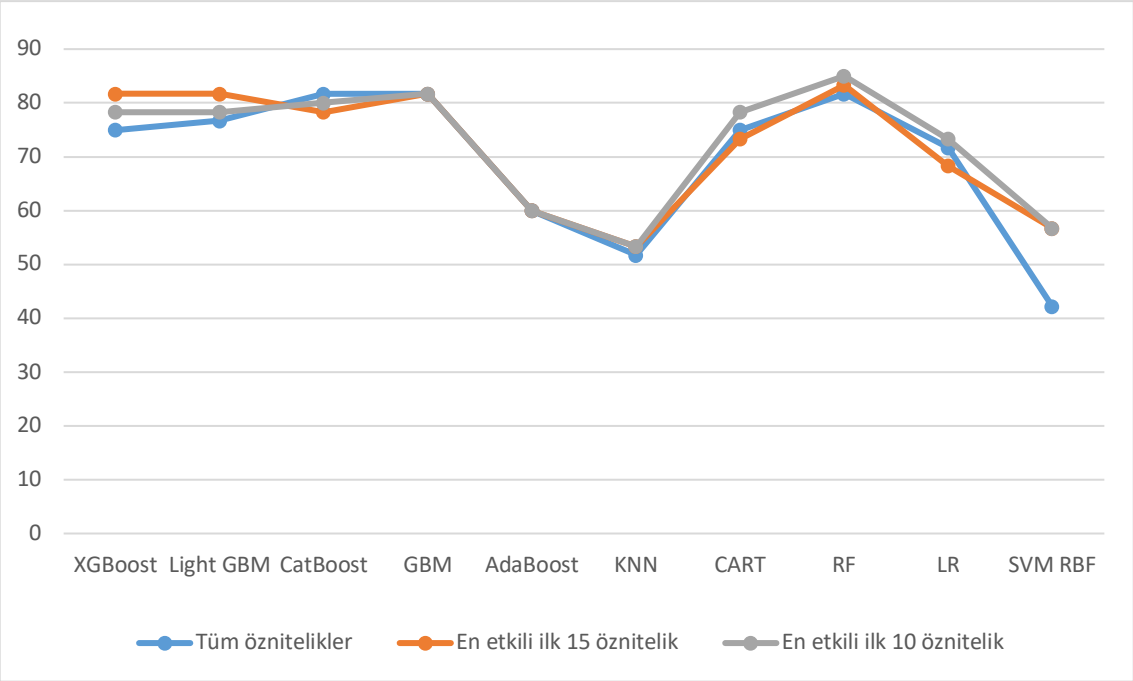

**Şekil 4.17** Veri setinin %40 test olarak kullanılması durumunda DO değerlerine göre tüm sınıflandırma sonuçları

Veri setinin %60 eğitim ve %40 test olarak ele alındığı analize ait doğruluk oranları Şekil 4.17'de yer almaktadır. Grafik incelendiğinde, en iyi sınıflandırma sonucunun bir önceki modelde olduğu gibi RO algoritması tarafından sağlandığı gözlenmektedir. En iyi 10 öznitelik ile gerçekleştirilen çalışmada RO algoritması %76,7 doğruluk oranı ile en iyi sınıflandırma sonucunu vermiş olmasına karşın doğruluk oranı %50 eğitim %50 test modelinin altında kalmıştır. KNN ve DVM RBF Çekirdek algoritmalarının %60 eğitim modelinde de öncekine benzer şekilde oldukça başarısız olduğu gözlenmektedir.



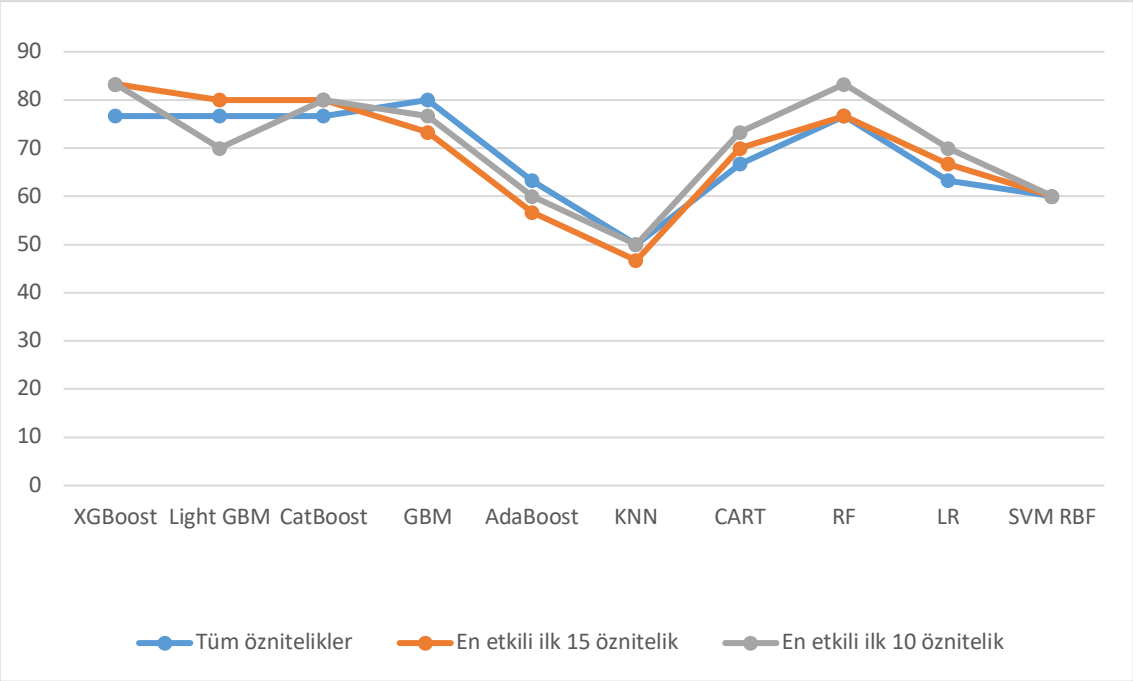

**Şekil 4.18** Veri setinin %30 test olarak kullanılması durumunda DO değerlerine göre tüm sınıflandırma sonuçları

Veri seti %70 eğitim %30 test olarak analiz edildiği senaryoda elde edilen doğruluk oranları grafiği Şekil 4.18'de yer almaktadır. Grafik incelendiğinde, RO algoritmasının en iyi 10 öznitelik ile analiz edildiği modelde %76,7 doğruluk oranı ile en iyi, CatBoost algoritmasının ise tüm özniteliklerle analiz edildiği modelde %75,6 doğruluk oranı ile ikinci en iyi sınıflandırma sonuçlarını verdikleri görülmektedir. DVM RBF Çekirdek algoritması daha önceki modellerde de olduğu gibi %50'nin altında kalan doğruluk oranı ile en kötü performansı sergilerken, KNN algoritmasının doğruluk oranı diğer modellere göre biraz artış göstermiş olsa da en kötü ikinci performans olarak gözlenmiştir.




**Şekil 4.19** Veri setinin %20 test olarak kullanılması durumunda DO değerlerine göre tüm sınıflandırma sonuçları

Veri setinin %80 eğitim %20 test olarak ayrılarak yapılan sınıflandırmaya ait doğruluk oranları grafiği Şekil 4.19'da yer almaktadır. Bu eğitim modelinde tüm algoritmaların doğruluk oranlarının önceki bölünmelere göre artış gösterdiği gözlenmektedir. RO algoritması en iyi 10 öznitelik ile ele alındığı analiz sonucunda %85 doğruluk oranı ile araştırma boyunca elde edilen en iyi sınıflandırma performansını sergilemiştir. İkinci en iyi sınıflandırma sonucu %83,3 olarak yine RO algoritması tarafından en iyi 15 öznitelik kullanılarak gerçekleştirilen analiz sonucunda elde edilmiştir. XGBoost ve HGAM algoritmaları en iyi 15 öznitelik biliniyorken %81,7 doğruluk oranı ile üçüncü en iyi sınıflandırma sonuçlarını sergilemişlerdir.




**Şekil 4.20** Veri setinin %10 test olarak kullanılması durumunda DO değerlerine göre tüm sınıflandırma sonuçları

Şekil 4.20'de veri setinin %90 eğitim %10 test olarak çalışıldığı sınıflandırma sonuçları yer almaktadır. XGBoost algoritması en iyi 10 öznitelik ve en iyi 15 öznitelik kullanılarak %83,3 doğruluk oranı, RO algoritması ise en iyi 15 öznitelik kullanılarak %83,3 doğruluk oranı ile en iyi sınıflandırma performansını sergilemişlerdir. Ancak bu performans yine de %80 eğitim bölünmesi ile yapılan sınıflandırmanın önüne geçememiştir. DVM RBF Çekirdek algoritması araştırma boyunca sergilediği en iyi performansı göstererek %60 doğruluk oranına bu modelde ulaşmıştır, ancak bu oran yine de en kötü ikinci performans olarak görülmektedir.

# 5
# SONUÇ VE ÖNERİLER

Bu araştırmada kullanıcıların finansal teknoloji uygulamalarına karşın tutum ve davranışları, İstanbul'da ikamet etmekte olan on sekiz yaşından büyük bireylerin değerlendirmeleri ile analiz edilmiştir. Araştırma katılımcıların demografik yapısı, sosyoekonomik statüsü, finansal risk yaklaşımları gibi değişkenler baz alınarak hazırlanmış çalışmadan elde edilen verilerle yürütülmüştür ve makine öğrenmesi yöntemleri ile analiz edilmiştir. Finansal teknoloji uygulamaları ve kullanıcı deneyimi alanındaki gelişimlere ve tarihçelerine bu araştırma kapsamında detaylı olarak yer verilmiştir.

Çalışmada makine öğrenmesi hakkında teorik ve temel bilgilerin yanı sıra tarihsel gelişimi hakkında bilgiler sunulmuştur. Araştırmanın konusunu oluşturan finansal teknoloji uygulamaları ve kullanıcı deneyimi tarihçelerine yer verilerek detaylı olarak açıklanmış ve önemi vurgulanmıştır.

Bu çalışmada Borda yöntemi ile, Feature Importance, Mutual Info ve Shap tabanlı bir öznitelik çalışması gerçekleştirilmiş ve ele alınan öznitelikler en etkiliden en etkisize olacak şekilde sıralandırılmıştır. Daha sonra veri setinin yarısına karşılık gelen 15 öznitelik ile sınıflandırma performansları incelenmiştir. Ardından öznitelik sayısı 10 özniteliğe kadar düşürülerek sınıflandırma performansları takip edilmiştir. 15 öznitelikten 10 özniteliğe kadar azaltılma sürecinde sınıflandırma performansları arasında ciddi farklılıklar gözlenmediğinden, araştırma sonuçlarında 15 öznitelik için gerçekleştirilen sınıflandırma sonuçlarının ardından 10 öznitelik ile gerçekleştirilen sınıflandırmaya yer verilmiştir. Öznitelikler 10'un altına düşürüldüğünde sınıflandırma başarılarında azalmalar meydana geldiği gözlemlendiğinden, bu sonuçlara yer verilmemiştir.

Araştırma sonucunda, çeşitli modeller üzerinde uygulanan sınıflandırma algoritmalarının sonuçları detaylı olarak uygulama bölümünde yorumlanmıştır. Uygulanan sınıflandırma algoritmaları sonucunda; veri seti %80 eğitim ve %20 test olarak ele alındığı modelde, en iyi açıklayıcılığa sahip ilk on öznitelik

biliniyorken RO algoritması %85 doğruluk oranı ile en iyi performansı sergilemiştir. Bu sonuç göstermektedir ki; belirtilen model kullanılıyorken kullanıcıların finansal teknoloji uygulamalarını kullanım düzeyi %85 oranında tahmin edilebilirdir. Burada %85 doğruluk oranı elde edilirken kullanılan değişkenler; bir günde ekran başında geçirilen süre, geleneksel/dijital bankacılık tercihi, mobil veya internet bankacılığı kullanımı, mevcutta ödenmekte olan kredi sayısı, müşterisi olunan banka sayısı, yaş, aylık birikim/gelir oranı, aylık net gelir, sık kullanılan bankacılık işlemleri, yatırım tercihleri şeklindedir. Dolayısıyla çalışmada incelenen veri sağlayıcı gruba yönelik bu on özniteliğin değeri biliniyorken RO algoritması kullanılarak dijital bankacılık kullanım düzeyini %85 oranında doğru sınıflandırabileceği söylenebilir.

Fettahoğlu ve Kıldıze'nin dijital finansal okuryazarlık ve bireylerin finansal teknoloji kullanma konusundaki tutumlarını ele alan araştırmalarında kullanıcıların finansal teknolojiye dayalı ürünlerin büyük çoğunluğu hakkında bilgiye, farkındalığa sahip olduğu ancak söz konusu finansal ürün ve hizmetleri kullanım açısından mobil ve internet bankacılığı kullanımının yoğunlukta olduğu ve buna karşın daha fazla bilgi birikimi gerektiren ürünleri kullanmaktan çekindikleri ortaya koyulmuştur. Cinsiyet, yaş ve aylık gelir düzeylerine göre finansal teknolojiye dayalı ürün ve hizmetlerden yararlanma eğiliminin anlamlı olarak değişkenlik göstermesi, bizim araştırmamızda yer alan bağımlı değişkeni etkileyen ve öznitelik seçim algoritmaları ile belirlenmiş açıklayıcılığı yüksek öznitelikleri destekleyici olduğu gözlenmektedir [4].

Araştırmadan elde edilen sonuçlar, bankacılık sektöründe teknoloji kullanımına yönelik müşteri davranışlarının ve bu davranışları etkileyen önemli unsurların belirlenmesinde kullanılabilir. Bununla birlikte, bankacılık sektöründe dijial kanalların kullanımına yönelik teşvikler yaratmak için hedef kitle çalışmalarında, dijital ürünlerin pazarlama stratejilerinde, yeni dijital müşteri kazanımında veya benzer çalışmalarda uygulanabilir olduğunu göstermiştir. Dijital bir ürün piyasaya kullanıcıların hizmetine sunulmadan önce farklı profillerden oluşan kitleler bu yöntem kullanılarak uygulamanın sağlayacağı kullanım oranı, etkileşim miktarı gibi metrikler bir uygulama üretilmeden öngörülebilir ve yatırım getirisi değerlendirilerek buna göre fizibilite çalışmaları gerçekleştirilebilir.

Bu araştırma yalnızca İstanbul'da ikamet etmekte olan 300 kişilik katılımcı grubu ile gerçekleştirilmiş olup, araştırmalarda katılımcı sayısının artırılması gerçekleştirilecek sınıflandırma başarılarını artırabilecektir.

# KAYNAKÇA


[1] Samuel A., “Some Studies in Machine Learning Using the Game of Checkers”, IBM Journal of Research and Development, Temmuz 1959.

[2] Çağlayan Akay E., “Ekonometride Yeni Bir Ufuk: Büyük Veri ve Makine Öğrenmesi”, Social Sciences Research Journal, Vol. 7, Issue 2, 2018.

[3] Russel S.ve Norvig P., “Artificial Intelligence: A Modern Approach (2.Bas)”, Prentice Hall, 1995-2003.

[4] Fettahoğlu S. Ve Kıldıze D., “Dijital Finansal Okuryazarlık ve Bireylerin Finansal Teknoloji Kullanma Konusundaki Tutumları”, International Journal of Society Researches, Year 9, Vol. 12, ss. 872 - 883, Ağustos 2019.

[5] Emerson S., Kennedy R., O’Seha L. ve O’Brien J., “Trends and Applications of Machine Learning in Quantitative Finance”, 8th International Conference on Economics and Finance Research, 2019.

[6] Zhang J. Z., Liu J. ve Xia H., “Identifying Fintech Risk through Machine Learning: analyzing the Q&A text of an online loan investment platform”, Annals of Operations Research, 2021.

[7] Chen C.H. ve Yeh Y.J., “A machine learning approach to predict the success of crowfunding fintech Project”, Journal of Enterprise Information Management, 2020.

[8] Gambacorta L., Huang Y., Qiu H. ve Wang J., “How do machine learning and non-traditional data effect credit scoring? New evidence from a Chinese fintech firm”, Bank for International Settlements Working Paper no.834, 2020.

[9] Lacruz F. ve Saniie J., “Applications of Machine Learning in Fintech Credit Card Fraud Detection”, IEEE International Conference on Electo Information Technology, USA, 2021.

[10] Vives X., “The impact of fintech on banking”, European Economy – Banks, Regulation and the Real Sector, Issue 2, 2017.

[11] Agarwal S., Alok S., Ghosh P. ve Gupta S., “Financial inclusion and alternate credit scoring for millenials: role of big data and machine learning in fintech”, National University Singapore Working Paper, 2020.

[12] Fettahoğlu S., “E-Finans”, Umuttepe Yayınları, ss. 53, 2017.

[13] Zavolokina L., Dolata M. ve Schwabe G., “Fintech-what’s in a name?”, Proceedings in International Conference on Information Systems, ss. 1, Dublin, 2016.

[14] Arner D. W., Barberis J. N. ve Buckley R. P., “The Evolution of Fintech: A New Post-Crisis Paradigm?”, Social Science Research Network, 2015.

[15] Burak Aktürk, “Fintek Tarihi ve Türk Bankacılık Sektöründeki Yeri”, https://katilimfinans.com.tr/fintech-ve-dijitallesme/fintek-tarihi-ve-turk-bankacilik-sektorundeki-yeri-h13215.html, 2021, [Erişim tarihi: 15 Nisan 2022].

[16] Abdillah L.A., “FinTech E-Commerce Payment Application User Experience Analysis during Covid-19 Pandemic”, Scientific Journal of Informatics, vol.7, no.2, Kasım 2020.

[17] Deloitte ve Fintech İstanbul, “FinTech Ecosystem in Turkey”, www.deloitte.com/tr/en/pages/finance/articles/fintech-ecosystem-in-turkey.html, Nisan 2017, [Erişim tarihi: 12 Nisan 2022].

[18] Alpaydın E., “Introduction to Machine Learning (4.bas)”,MIT, ss. xix, 1-3, 13-18, 2020.

[19] Mitchell Tom M., “Machine Learning”, McGraw-Hill, ss. 2, 1997.

[20] Gentleman R., Huber W. ve Carey V. J., “Supervised Machine Learning in Bioconductor Case Studies”, ss. 120-140, 2008.

[21] Cielen, Davy, Arno D.B., ve Mohamed A., “Introducing Data Science”, Manning Publications, 2016.

[22] Chui, Michael, Vishnu K. ve McCarthy B., “An Executive’s Guide To AI”, 2018.

[23] IBM, “Makine Öğrenmesi IBM Cloud Education”, https://www.ibm.com/tr-tr/cloud/learn/machine-learning#toc-makine-ren-7-mHabo_, Temmuz 2020, [Erişim Tarihi: 12 Nisan 2022].

[24] Jones M.T., “Models for Machine Learning”, https://developer.ibm.com/articles/cc-models-machine-learning/#reinforcement-learning, 4 Aralık 2017, [Erişim Tarihi: 12 Nisan 2022].

[25] Onat A., “Veri Madenciliğinin Web Tabanlı Uygulamalarda İnsan Uyumluluklarının Tespiti Üzerine Bir Çalışma”, Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya, 2008.

[26] Hinton G. ve Sejnowski T.J., “Unsupervised Learning and Map Formation: Foundations of Neural Computation”, MIT Press, 1999.

[27] Field A., “Discovering Statistics Using SPSS(2. Bas.)”, Sage Publications, 2005.

[28] Çokluk Ö., “Lojistik Regresyon Analizi: Kavram ve Uygulama”, Kuram ve Uygulamada Eğitim Bilimleri, cilt 10, sayı 3, ss. 1350-1410, 2010.

[29] McCullagh P. ve Nelder J.A., “Generalized Linear Models”, vol.37, CRC Press, 1989.

[30] Çelik U., Akçetin E. ve Gök M., “Rapidminer ile Uygulamalı Veri Madenciliği”, Pusula Yayıncılık, 2017.

[31] Alpaydın E., “Yapay Öğrenme”, Boğaziçi Üniversitesi Yayınevi, Türkiye, 2011.

[32] Akpınar H., “DATA – Veri Madenciliği”, Papatya Bilim, Eylül 2014.

[33] Bao H. T., “Çekirdek Methods and Support Vector Machines”, Japan Advance Institute of Science and Technology John von Neumann Institute, VNU- HCM.

[34] Breiman L., “Random Forests”, https://www.stat.berkeley.edu/~breiman /randomforest2001.pdf, Statistics Department University of California, Ocak 2001, Erişim tarihi [2 Nisan 2022].

[35] Breiman, L., “Random forests”, Machine Learning Kluwer Academic Publishers, vol.45, ss.5–32, Hollanda, 2001.

[36] Skurichina, M. ve Duin, R. P. W., “Bagging, boosting and the random subspace method for linear classifiers”, Pattern Analysis and Applications, vol.5, kısım 2, ss.121–135, 2002.

[37] Johnson, R. ve Zhang, T., “Learning Nonlinear Functions Using Regularized Greedy Forest”, Technical Report, 2012.

[38] Patel S., “Chapter 6: AdaBoost Classifier”, https://medium.com/machine-learning-101/https-medium-com-savanpatel-chapter-6-adaboost-classifier-b945f330af06 , [Erişim Tarihi: 19 Nisan 2022].

[39] Readthedocs, “Light GBM Documentation”, https://lightgbm.readthedocs.io/en/latest/index.html, [Erişim Tarihi: 22 Nisan 2022].

[40] Readthedocs, “XGBoost Documentation”, https://xgboost.readthedocs.io/en/latest/tutorials/index. html, [Erişim Tarihi : 20 Nisan 2022].

[41] Prokherenkova L., Gusev G., Vorobev A., Dorogush A.V. ve Gulin A., “CatBoost: unbiased boosting with categorical features”, 32nd Conference on Neural Information Processing Systems, Canada, 2018.

[42] Sökün H., Kalkan H. ve Cetişli B., “Classification of physical activities using accelerometer signals”, 20th Signal Processing and Communications Applications Conference, Turkey, 2012.

[43] Brownlee J., “How to Calculate Feature Importance with Python”, https://machinelearningmastery.com/calculate-feature-importance-with-python, Mart 2020, [Erişim Tarihi: 20 Mart 2022].

[44] Scikit Learn, “Sklearn Metrics Mutual Info Score”, https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mutual_info_score.html, [Erişim Tarihi: 15 Nisan 2022].

[45] Mackay D.J.C., “Information Theory, Inference & Learning Algorithms”, Cambridge University Press, ss.139, New York, 2002

[46] Shapley L., “A value for n-pearson games”, Princeton University Press, ss. 301-317, Princeton, 1953.

[47] Kirmanoğlu H., “Kamu Ekonomisi Analizi (4.bas.)”, Beta Yayıncılık, ss.251, İstanbul, 2012.

[48] Nuray R. ve Can F., “Automatic Ranking of Infirmation Retrieval Systems Using Data Fusion”, Information Processing and Management: an International Journal, vol.42, kısım 3, ss. 595,614, 2006.

[49] Kim D.J., Ferrin D.L. ve Rao H.R., “A trust based consumer decision-making model in electronic commerce: The role of trust, perceived risk and their antecedents”, Decision Support Systems, vol.44, ss.544-564, 2008.

[50] Lee M.C., “Factors influencing the adoption of internet banking: An integration of TAM and TPB with perceived risk and perceived benefits”, Electronic Commerce Research and Applications, vol.8, kısım 3, ss.130-141, 2009.

[51] Ryu H.S., “What makes users willing or hesitant to use fintech? The moderating effect of user type”, Industrial Management & Data Systems, Vol. 118, kısım 3, ss.541-569, 2017.

[52] Wu W. W., “Beyond Travel & Tourism competitiveness ranking using DEA, GST, ANN and Borda Count”, Expert Systems with Applications, 2011.

# A
# VERİ TOPLAMA ÇALIŞMASI

S1: Yaş

S2: Cinsiyet

1. Kadın
2. Erkek

S3: Eğitim durumu

1. İlk-orta öğrenim veya lise
2. Lisans
3. Yüksek lisans veya doktora

S4: Medeni Hal

1. Evli
2. Bekar

S5: Çocuk Sayısı

1. Çocuğum yok
2. Çocuğum var

S6: Bir günde bilgisayar başında geçirilen süre (saat olarak belirtiniz)

S7: Cep telefonunda aktif kullanılan uygulama sayısı

1. 5 veya daha az uygulama
2. 6 veya daha fazla uygulama

S8: Çalışma durumu

1. Özel sektörde çalışıyorum
2. Kamuda çalışıyorum
3. Kendi işimin sahibiyim
4. Çalışmıyorum

S9: Meslek grubu

1. Bilişim&Araştırma sektörü
2. Eğitim sektörü
3. Finans sektörü
4. Hizmet sektörü
5. Mühendis
6. Öğrenci
7. Özel Sektör Diğer
8. Sağlık Sektörü

9. Emekli

S10: Aylık net gelir

S11: Bireysel emeklilik sözleşmesi

1. Var(kendim yatırıyorum)
2. Var(işveren katkılı)
3. Yok

S12: Yaşadığı ilçe

1. Avrupa yakası
2. Anadolu yakası

S13: Yaşadığı yer

1. Kirada yaşıyorum
2. Kendi evimde yaşıyorum
3. Bir başkasının yanında kira vermeden yaşıyorum
4. Diğer

S14: Yaşadığı konut tipi

1. Apartman dairesi
2. Residance
3. Site içi
4. Müstakil
5. Diğer

S15: İkamet ettiği konutun en yakın AVM'ye uzaklığı

1. 0-500m
2. 500m-1km
3. 1km-3km
4. 3km'den fazla

S16: Kira geliri

1. Var
2. Yok

S17: Şahsi aracınız

1. Var
2. Yok

S18: Aracın değeri

1. Aracım yok
2. 1TL-100,000TL
3. 100,001TL-200,000TL
4. 200,001TL-400,000TL
5. 400,001TL veya üzeri

S19: İkamet ettiği konutun müşterisi olduğu banka ATM'sine uzaklığı

1. 0-500m
2. 500m-1km
3. 1km-3km
4. 3km'den fazla

S20: Yatırım risk algısı

1. Yüksek getiri elde etmek için riskli yatırımlar tercih eder
2. Orta riskli yatırımlarla dengeli gelir elde etmeyi tercih eder(kısmen yüksek, kısmen düşük riskli)
3. Risk almayı sevmem, ancak düzenli getiri elde edebileceğim yatırımlar tercih eder
4. Yatırım yapmaz

S21: Kredi yaklaşımı

1. İhtiyaçlarımı kredilendirerek karşılamayı tercih ederim, birikimlerimi kullanmam
2. İhtiyaçlarımı birikimlerimle karşılamayı tercih ederim, borçlanmayı sevmem
3. İhtiyaçlarımı karşılamak için birikimlerim yetersizse kredi kullanabilirim

S22: Mevcutta ödenmekte olan kredi sayısı

1. Kredim yok
2. 1 kredi
3. 2 kredi
4. 3 veya daha fazla kredi

S23: Aylık gelirinizin % kaçı ile birikim yapıyor (% olarak belirtiniz)

S24: Müşterisi olduğunuz banka sayısı

1. 1
2. 2
3. 3
4. 4
5. 5 veya daha fazla

S25: Bankacılık işlemleri için tercih edilen kanal

1. Bankadan işem
2. ATM'den işlem
3. Dijital kanallardan işlem (internet veya mobil şube)

S26: Sıklıkla kullanılan bankacılık işlemleri (birden fazla seçim yapılabilir)

1. Maaş işlemleri
2. Yatırım/birikim işlemleri
3. Kredi kartı işlemleri

4. Para transfer işlemleri
5. Fatura/kredi borcu ödeme işlemleri

S27: Kredi kartı sahipliği

1. Var
2. Yok

S28: Kripto para yatırımı

1. Var
2. Yok

S29: Mobil veya internet bankacılığı kullanımı

1. Sadece mobil bankacılık
2. Sadece internet bankacılığı
3. Hem mobil hem internet bankacılığı
4. Her ikisini de kullanmıyorum

S30: Dijital bankacılık kullanım düzeyi

0. Hiç kullanmıyorum
1. Haftada 4 defa veya daha az
2. Günde 1 defa veya daha fazla

# TEZDEN ÜRETİLMİŞ YAYINLAR

## Konferans Bildirileri

1. Zengin G. ve Öz E., “Makine Öğrenmesi Yöntemleri ile Finansal Teknoloji Uygulamalarının Kullanımına Yönelik Bir Uygulama”, Akdeniz Uluslararası Uygulamalı Bilimler Kongresi, Antalya, Nisan 2022