

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

**MAKİNE ÖĞRENMESİ TEKNİKLERİ İLE TAHSİLAT DAVRANIŞI
TAHMİNİ: TELEKOMÜNİKASYON SEKTÖRÜ ÖRNEĞİ**

YÜKSEK LİSANS TEZİ

Elif EKİZ

Endüstri Mühendisliği Anabilim Dalı

Endüstri Mühendisliği Programı

HAZİRAN 2019

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ

**MAKİNE ÖĞRENMESİ TEKNİKLERİ İLE TAHSİLAT DAVRANIŞI
TAHMİNİ: TELEKOMÜNİKASYON SEKTÖRÜ ÖRNEĞİ**

YÜKSEK LİSANS TEZİ

**Elif EKİZ
(507161106)**

Endüstri Mühendisliği Anabilim Dalı

Endüstri Mühendisliği Programı

Tez Danışmanı: Doç. Dr. Başar ÖZTAYŞI

HAZİRAN 2019

İTÜ, Fen Bilimleri Enstitüsü'nün 507161106 numaralı Yüksek Lisans Öğrencisi Elif EKİZ, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı “MAKİNE ÖĞRENMESİ TEKNİKLERİ İLE TAHSİLAT DAVRANIŞI TAHMİNİ: TELEKOMÜNİKASYON SEKTÖRÜ ÖRNEĞİ” başlıklı tezini aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

Tez Danışmanı : **Doç. Dr. Başar ÖZTAYŞI**
İstanbul Teknik Üniversitesi

Jüri Üyeleri : **Doç. Dr. Sezi Çevik ONAR**
İstanbul Teknik Üniversitesi

Prof. Dr. Selçuk ÇEBİ
Yıldız Teknik Üniversitesi

Teslim Tarihi : 03 Mayıs 2019
Savunma Tarihi : 12 Haziran 2019





Aileme,



ÖNSÖZ

Yüksek lisans tez teslim sürecinde bana yol göstererek, benden her türlü yardımını ve desteğini esirgemeyen, bu projede çalışmama imkan sağladığı için değerli hocam Doç. Dr. Başar ÖZTAYŞI'ye teşekkürü borç bilirim.

Ayrıca Doç. Dr. Sezi Çevik'e ve Prof. Cengiz Kahraman'a da destek ve yardımlarından ötürü şükranlarımı sunarım.

Bu çalışma sırasında veri temini sürecinin kolaylıkla ilerleyebilmesi için büyük özveride bulunan ve değerli fikirlerini ve yorumlarını esirgemeyen Ersin Öztürk ve Aykut Şahin'e özellikle hızlı geri dönüşleri için teşekkür ederim. Yöneticim Yavuz Topçu'ya ve çalışma arkadaşlarıma gösterdikleri anlayış için teşekkür ederim.

5170012 numaralı proje ile çalışmalarımı destekleyen ve bana burs sağlayan TÜBİTAK'a teşekkür ederim.

Aileme bu süreçteki destekleri ve bana gösterdikleri sabır için teşekkür ederim.

Haziran 2019

Elif Ekiz
(Bilgisayar Mühendisi)



İÇİNDEKİLER

Sayfa

ÖNSÖZ.....	vii
İÇİNDEKİLER	ix
KISALTMALAR	xiii
SEMBOLLER	xv
ÇİZELGE LİSTESİ.....	xvii
ŞEKİL LİSTESİ.....	xxi
ÖZET.....	xxiii
SUMMARY	xxv
1. GİRİŞ	1
1.1 Telekomünikasyon Şirketinin Yürüttüğü Süreç	1
1.2 Tezin Amacı	2
1.3 Çalışılan Şirket Bilgileri	3
1.4 Tezin Yapısı	4
2. LİTERATÜR ARAŞTIRMASI	7
2.1 Konu Olarak Benzer Çalışmalar	7
2.1.1 Telekomünikasyon sektörü ile ilgili çalışmalar	8
2.1.2 Kredi geri ödeme risk tahmini ile ilgili çalışmalar	8
2.2 Öznitelik Seçimi ile İlgili Diğer Çalışmalar	10
2.3 Doğruluk Artırıcı Yöntem Olarak Benzer Çalışmalar	12
2.4 Değerlendirme Metrikleri ile İlgili Kredi Puanlama Çalışmaları.....	12
3. KULLANILAN YÖNTEMLER	15
3.1 K-En Yakın Komşu (K-Nearest Neighbour).....	15
3.2 Karar Ağaçları	17
3.2.1 Entropi.....	18
3.2.2 Bilgi kazancı	19
3.2.3 Gini indeksi	19
3.3 Naive Bayes.....	21
3.4 Lojistik Regresyon	22
3.5 Yapay Sinir Ağları	24
3.6 Topluluk Yöntemleri (Ensemble).....	28
3.6.1 Yerine koyarak artırma (Bagging – Bootstrap aggregating).....	30
3.6.1.1 Rastgele orman ağaçları (Random forest trees)	31
3.6.1.2 Son derece rastgele ağaçlar (Extra trees (Extremely randomized trees))	31
3.6.2 Artırıcı (Boosting).....	32
3.6.2.1 Adaboosting (Adaptive boosting)	33
3.6.2.2 Gradyen artırma (Gradient boosting)	34
3.6.2.3 Fazlasıyla gradyen artırma (XGBoost (Extreme gradient boosting)..	35
4. MÜŞTERİNİN ÖDEME TAHMİNİ SIRASINDAKİ SÜREÇLER.....	37
4.1 Uygulama Verisinin Toplanması	37

4.2 Uygulama Verisi Üzerinde Yapılan İşlemler	49
4.2.1 Veri temizleme	50
4.2.2 Veri indirgeme.....	51
4.2.3 Veri genişletme	53
4.2.4 Veri kategorize hale dönüştürme (Mapping)	54
4.2.5 Normalizasyon	58
4.2.6 Veri tabanı boyutu.....	60
4.2.7 Kodlayıcı (Encoding)	60
4.2.8 K kat çapraz doğrulama.....	60
4.2.9 Veriyi eğitim ve test olarak bölme oranları.....	61
4.2.10 Veri kümesi dengeleme	61
4.2.11 Veri kümesi rastgeleliği	62
4.3 İşlemler Sonucu Veri Kümesinin Son Hali	62
4.4 Öznitelik Seçim Metriklerine Göre Öznitelik Altkümesinin Belirlenmesi	70
4.5 Öznitelik Seçimi ve Öznitelik Seçiminde Kullanılan Yöntemler.....	70
4.5.1 Öznitelik seçimi ve sıralanması.....	71
4.5.2 Ağaç temelli yöntemlerden elde edilen öznitelik önem sırası.....	72
4.5.3 Temel bileşen analizine göre öznitelik önem sırası	73
4.5.4 Özyinelemeli öznitelik sıralaması	75
4.5.5 Tek değişkenli seçim.....	77
5. MÜŞTERİNİN ÖDEME TAHMİNİNİN BELİRLENMESİ	
UYGULAMASI	81
5.1 Sistem Tasarımı	81
5.2 Performans Analizi Metrikleri.....	81
5.2.1 Kesinlik (Precision).....	82
5.2.2 Hassaslık (Recall).....	82
5.2.3 Doğruluk (Accuracy).....	83
5.2.4 F ölçütü.....	83
5.2.5 ROC istatistiği	83
5.2.6 Karmaşıklık matrisi (Confusion matrix)	84
5.3 Sınıflandırıcı Modellerin Sonuçlarının Yorumlanması	85
5.3.1 Ağaç temelli yöntemler ile oluşturulmuş alt kümeler	85
5.3.2 Temel bileşen analizi (PCA) yöntemiyle oluşturulmuş alt kümeler	96
5.3.3 Lojistik regresyon ile özyinelemeli yöntem ile oluşturulmuş alt kümeler	106
5.3.4 Adaboost ile özyinelemeli yöntem ile oluşturulmuş alt kümeler	116
5.3.5 Ki-kare temelli yöntemler ile oluşturulmuş alt kümeler	126
5.4 Öznitelik Altkümelerinin En İyi Sonuçlarının Karşılaştırılması	138
5.4.1 Ağaç modellerinin öznitelik veri kümesi en iyi alt kümesinin modeller	138
5.4.2 Temel bileşen analizi yöntemiyle öznitelik veri kümesinin en iyi alt	139
5.4.3 Lojistik regresyon ile özyinelemeli öznitelik veri kümesi en iyi alt kümenin	141
5.4.4 Adaboost ile özyinelemeli öznitelik veri kümesi en iyi alt kümenin modeller	142
5.4.5 Ki-kare temelli öznitelik veri kümesi en iyi alt kümenin modeller üzerinde	144
5.5 Tüm Öznitelik Seçim Yöntemlerinin En İyi Model Performanslarının Birbirleri	145

6. SONUÇLAR VE ÖNERİLER	147
KAYNAKLAR	153
EKLER.....	157
ÖZGEÇMİŞ.....	161





KISALTMALAR

YSA	: Yapay Sinir Ağları
MLP	: Multi-Layer Perceptron - Çok Katmanlı Algılayıcılar
KA	: Karar Ağaçları
RO	: Rastgele Orman
XGB	: Extra Gradient Boosting
ROC	: Receiver Operating Characteristic Curve
TBA	: Temel Bileşen Analizi (PCA)
DHO	: Dengelenmiş Hata Oranı (ing. Balanced Error Rate)
DVM	: Destek Vektör Makineleri
LR	: Lojistik Regresyon
DSY	: Doğru Sınıflandırılan Vakaların Yüzdesi (PCC–Percentage Correctly Classified Cases)
CART	: Classification and Regression Trees)(Sınıflandırma ve regresyon ağaçları)
KOBİ	: Küçük ve Orta Büyüklükteki İşletmeler
GA	: Genetik Algoritma
YP	: Yanlış Pozitif
DP	: Doğru Pozitif
YN	: Yanlış Negatif
DN	: Doğru Negatif
K-EYK	: K-en yakın komşu
Extra trees	: Extremely Randomized Trees
Ark.	: Arkadaşları
Diğ.	: Diğerleri



SEMBOLLER

E_D	: D Veri Kümesinin Entropisi
$BK(K, X)$	: X Özniteliğinin Düğüm Olması Durumundaki Bilgi Kazancı
G	: Gini İndeksi
P_i	: i Sınıfının Olasılığı
Y	: Tahmin Edilecek Değer
W_x	: Özniteliğin Ağırlığı/Etkisi
X_i	: Öznitelikler ($i = 1, 2, \dots, 34$)
B	: Bias (Sapma Miktarı)
$P(A \setminus B)$	: A'nın B'ye koşullu olasılığı
$\Lambda(\lambda)$	: Özdeğer



ÇİZELGE LİSTESİ

Sayfa

Çizelge 4.1 : Telekomünikasyon şirketinden alınan öznitelik listesi.	37
Çizelge 4.2 : Müşterilerin uyruklarına göre dağılımı.	45
Çizelge 4.3 : Müşterilerin yaşadığı coğrafik bölgelerin dağılımı.	55
Çizelge 4.4 : Uyruk bilgisi.	55
Çizelge 4.5 : Yaş aralıkları I.	56
Çizelge 4.6 : Yaş aralıkları II.	56
Çizelge 4.7 : Medeni durum I.	57
Çizelge 4.8 : Medeni durum II.	57
Çizelge 4.9 : Veri kümesi öznitelikleri ve açıklamaları.	63
Çizelge 4.10 : Ağaç temelli yöntemler ile öznitelik sıralaması.	72
Çizelge 4.11 : Temel bileşene göre öznitelik seçimi.	74
Çizelge 4.12 : Lojistik regresyona göre özyinelemeli öznitelik sıralaması.	75
Çizelge 4.13 : Adaboost'a göre özyinelemeli öznitelik sıralaması.	76
Çizelge 4.14 : Ki-kare'ye göre öznitelik sıralaması.	78
Çizelge 5.1 : Karmaşıklık matrisi.	84
Çizelge 5.2 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – Lojistik regresyon.	85
Çizelge 5.3 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – Naive Bayes sınıflayıcısı.	86
Çizelge 5.4 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – Karar ağaçları sınıflayıcısı.	87
Çizelge 5.5 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – K en yakın komşu sınıflayıcısı.	88
Çizelge 5.6 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – Gradyen boosting sınıflayıcısı.	89
Çizelge 5.7 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – Adaboost sınıflandırıcısı.	90
Çizelge 5.8 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – Yerine koyma sınıflandırıcısı.	91
Çizelge 5.9 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – Rastgele orman sınıflayıcısı.	92
Çizelge 5.10 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – YSA sınıflayıcısı.	93
Çizelge 5.11 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – XGB sınıflayıcısı.	94
Çizelge 5.12 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – Extra Trees sınıflayıcısı.	95
Çizelge 5.13 : Temel bileşen yöntemlerle oluşturulmuş öznitelik önemleri – Lojistik Regresyon sınıflayıcısı.	96

Çizelge 5.14 : Temel bileşen yöntemlerle oluşturulmuş öznitelik önemleri – Naive Bayes sınıflayıcısı.	97
Çizelge 5.15 : Temel bileşen yöntemlerle oluşturulmuş öznitelik önemleri – Karar ağaçları sınıflayıcısı.	98
Çizelge 5.16 : Temel bileşen yöntemlerle oluşturulmuş öznitelik önemleri – K-en yakın komşu sınıflayıcısı.	99
Çizelge 5.17 : Temel bileşen yöntemlerle oluşturulmuş öznitelik önemleri – Gradyen Boosting sınıflayıcısı.	100
Çizelge 5.18 : Temel bileşen yöntemlerle oluşturulmuş öznitelik önemleri – AdaBoost sınıflayıcısı.	101
Çizelge 5.19 : Temel bileşen yöntemlerle oluşturulmuş öznitelik önemleri – Yerine koyma (Bagging) sınıflayıcısı.	102
Çizelge 5.20 : Temel bileşen yöntemiyle oluşturulmuş öznitelik önemleri – Rastgele orman sınıflayıcısı.	103
Çizelge 5.21 : Temel bileşen yöntemiyle oluşturulmuş öznitelik önemleri – YSA sınıflayıcısı.	103
Çizelge 5.22 : Temel bileşen yöntemiyle oluşturulmuş öznitelik önemleri – XGB sınıflayıcısı.	104
Çizelge 5.23 : Temel bileşen yöntemiyle oluşturulmuş öznitelik önemleri – Extra Trees sınıflayıcısı.	105
Çizelge 5.24 : Özyinelemeli yöntemle oluşturulmuş öznitelik önemleri – Lojistik regresyon sınıflayıcısı.	106
Çizelge 5.25 : Özyinelemeli yöntemle oluşturulmuş öznitelik önemleri – Naive Bayes sınıflayıcısı.	107
Çizelge 5.26 : Özyinelemeli yöntemle oluşturulmuş öznitelik önemleri – Karar ağaçları sınıflayıcısı.	108
Çizelge 5.27 : Özyinelemeli yöntemle oluşturulmuş öznitelik önemleri – K en yakın komşu sınıflayıcısı.	109
Çizelge 5.28 : Özyinelemeli yöntemle oluşturulmuş öznitelik önemleri – Gradyen boosting sınıflayıcısı.	110
Çizelge 5.29 : Özyinelemeli yöntemle oluşturulmuş öznitelik önemleri – AdaBoost sınıflayıcısı.	111
Çizelge 5.30 : Özyinelemeli yöntemle oluşturulmuş öznitelik önemleri – Bagging sınıflayıcısı.	112
Çizelge 5.31 : Özyinelemeli yöntemle oluşturulmuş öznitelik önemleri – Rastgele orman sınıflayıcısı.	113
Çizelge 5.32 : Özyinelemeli yöntemle oluşturulmuş öznitelik önemleri – YSA sınıflayıcısı.	114
Çizelge 5.33 : Özyinelemeli yöntemle oluşturulmuş öznitelik önemleri – XGB sınıflayıcısı.	115
Çizelge 5.34 : Özyinelemeli yöntemle oluşturulmuş öznitelik önemleri – Extra trees sınıflayıcısı.	116
Çizelge 5.35 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznitelik önemleri – Lojistik regresyon sınıflayıcısı.	117
Çizelge 5.36 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznitelik önemleri – Naive Bayes sınıflandırıcısı.	117
Çizelge 5.37 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznitelik önemleri – Karar ağaçları sınıflandırıcısı.	118
Çizelge 5.38 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznitelik önemleri – K en yakın komşu sınıflayıcısı.	119

Çizelge 5.39 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznitelik önemleri – Gradyen boosting sınıflayıcısı.	120
Çizelge 5.40 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznitelik önemleri – AdaBoost sınıflayıcısı.	121
Çizelge 5.41 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznitelik önemleri – Bagging sınıflayıcısı.	122
Çizelge 5.42 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznitelik önemleri – Rastgele orman sınıflayıcısı.	123
Çizelge 5.43 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznitelik önemleri – YSA sınıflayıcısı.	124
Çizelge 5.44 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznitelik önemleri – XGB sınıflayıcısı.	125
Çizelge 5.45 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznitelik önemleri – Extra trees sınıflayıcısı.	126
Çizelge 5.46 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – Lojistik Regresyon sınıflayıcısı.	127
Çizelge 5.47 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – Naive Bayes sınıflayıcısı.	128
Çizelge 5.48 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – Karar ağaçları sınıflayıcısı.	129
Çizelge 5.49 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – K en yakın komşu sınıflayıcısı.	130
Çizelge 5.50 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – Gradyen boosting sınıflayıcısı.	131
Çizelge 5.51 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – AdaBoost sınıflayıcısı.	132
Çizelge 5.52 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – Bagging sınıflayıcısı.	133
Çizelge 5.53 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – Rastgele orman sınıflayıcısı.	134
Çizelge 5.54 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – YSA sınıflayıcısı.	135
Çizelge 5.55 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – XGB sınıflayıcısı.	136
Çizelge 5.56 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – Extra trees sınıflayıcısı.	137
Çizelge 5.57 : Ağaç modellerinin karşılaştırılması.	138
Çizelge 5.58 : Ağaç öznitelik sıralaması ile karmaşıklık matrisi.	139
Çizelge 5.59 : Temel bileşen analizi modellerinin karşılaştırılması.	139
Çizelge 5.60 : TBA öznitelik sıralaması ile karmaşıklık matrisi.	140
Çizelge 5.61 : Özyinelemeli lojistik regresyon modellerinin karşılaştırılması.	141
Çizelge 5.62 : Özyinelemeli lojistik reg. öznitelik sıralaması ile karmaşıklık matrisi.	142
Çizelge 5.63 : Adaboost ile özyinelemeli modellerin karşılaştırılması.	142
Çizelge 5.64 : Özyinelemeli adaboost öznitelik sıralaması ile karmaşıklık matrisi.	144
Çizelge 5.65 : Ki-kare modellerinin karşılaştırılması.	144
Çizelge 5.66 : Ki-kare öznitelik sıralaması ile karmaşıklık matrisi.	145
Çizelge 5.67 : Tüm özniteliklerin en iyi performansları.	145
Çizelge 6.1 : Yaş ve medeni durum çok seçmeli özniteliklerinin seçim tercihleri.	150
Çizelge A.1 : Ağaç modellerinin çalıştırılmasıyla elde edilen sonuçlar.	158

Çizelge A.2 : Ağaç modellerinin normalize edilmiş ve ortalamalarının alınmış hali.	
.....	159



ŞEKİL LİSTESİ

Sayfa

Şekil 3.1 : K-en yakın komşu algoritması $k=3$ ve $k=5$ için.	16
Şekil 3.2 : Düşük ve yüksek entropi ile hedef değişken tahmin şekli.	18
Şekil 3.3 : Ağaç görselleştirimi I.	20
Şekil 3.4 : Ağaç görselleştirimi I kesitinin büyütülmüş hali.	20
Şekil 3.5 : Lineer ve lojistik model karşılaştırması.	24
Şekil 3.6 : Sinir ağı modelinin çalışma prensibi.	25
Şekil 3.7 : Sinir ağı katmanları şekli.	26
Şekil 3.8 : Aktivasyon fonksiyonlarının grafikleri.	28
Şekil 3.9 : Topluluk yöntemleri bileşenleri.	29
Şekil 3.10 : Yerine koyma (bagging) ve artırıcı (boosting) topluluk yöntemleri.	29
Şekil 3.11 : Aşırı öğrenme model öğrenme şekli.	30
Şekil 3.12 : Boosting algoritması görseli.	32
Şekil 3.13 : Adaboost modelinin çalışma prensibi.	34
Şekil 3.14 : Gradyen Boosting algoritması çalışma prensibi.	35
Şekil 4.1 : Müşterinin yaşadığı şehir görseli.	38
Şekil 4.2 : Durum bilgisi dağılımı.	39
Şekil 4.3 : Durum tipi dağılımı.	40
Şekil 4.4 : Cinsiyet dağılımı.	40
Şekil 4.5 : İletişim durumu dağılımı.	41
Şekil 4.6 : Toplam Borç içerisinde alacaklıların dağılımı.	42
Şekil 4.7 : Medeni durum dağılımı.	43
Şekil 4.8 : Medeni durum ve cinsiyet dağılımı.	43
Şekil 4.9 : Büro şehirleri.	44
Şekil 4.10 : Taksit durumu dağılımı.	44
Şekil 4.11 : Transfer tarihi grafiği.	45
Şekil 4.12 : Müşterilerin yasal takip adedi dağılımı.	47
Şekil 4.13 : Doğum tarihi dağılımı.	47
Şekil 4.14 : Doğum tarihi dağılımı (uç noktalardan arındırılmış).	48
Şekil 4.15 : Bürolara atanmış yıllık toplam dosya sayısı.	49
Şekil 4.16 : Büro bazında 1 yıllık toplam tahsilat tutarı.	49
Şekil 4.17 : Meslek dağılımı.	52
Şekil 4.18 : Yaş aralıkları I.	56
Şekil 4.19 : Çapraz doğrulama.	61
Şekil 4.20 : Veri kümesi dengeleme.	62
Şekil 4.21 : Korelasyon matrisi.	65
Şekil 4.22 : Veri dağılımı I.	66
Şekil 4.23 : Veri dağılımı II.	67
Şekil 4.24 : Kategorik özniteliklerin dağılımları.	69
Şekil 4.25 : Öznitelik seçimi.	71
Şekil 4.26 : Ağaç temelli yöntemler ile öznitelik sıralaması.	73

Şekil 4.27 : Temel bileşene göre öznitelik sıralaması.	75
Şekil 4.28 : Ki-kare’ye göre öznitelik sıralaması.	79
Şekil 5.1 : Hassaslık ve kesinlik ilişki grafiği.....	82
Şekil 5.2 : Ağaç öznitelik sıralaması ile XGB modeli ROC grafiği.	139
Şekil 5.3 : TBA öznitelik sıralaması ile Gradyen Boosting modeli ROC grafiği....	140
Şekil 5.4 : Özyinelemeli lojistik reg. öznitelik sıralaması ile Gradyen Boosting modeli ROC grafiği.	142
Şekil 5.5 : Özyinelemeli adaboost öznitelik sıralaması ile Gradyen Boosting modeli ROC grafiği.	143
Şekil 5.6 : Ki-kare öznitelik sıralaması ile Gradyen Boosting modeli ROC grafiği.	145
Şekil B.1 : Özniteliklerin korelasyon matrisinin sayısal verileriError! Bookmark not defined.	



MAKİNE ÖĞRENMESİ TEKNİKLERİ İLE TAHSİLAT DAVRANIŞI

TAHMINİ: TELEKOMÜNİKASYON SEKTÖRÜ ÖRNEĞİ

ÖZET

Şirketlerin varlıklarını sürdürebilmeleri için, müşterilerine sağladıkları hizmetlerin karşılıklarını almaları gerekmektedir. Böylece kendi finansmanlarını sağlayıp, gelişmeye devam edebilirler. Ancak zamanında ödenmeyen faturalar şirketlere ciddi bir yük getirmektedir. Şirketler gelişmeye devam etmek ve yeni ürünler ortaya koymak yerine, bu alacakları tahsil etmeye çalışarak kendi bünyelerinde performans kaybı yaşamaktadırlar, gelişmeye ve büyümeye devam edememektedirler. Bu performans kaybını önlemek için bazı şirketler borçlu müşterilerinden, alacaklarının tahsilatını dış kaynaklarla yürütmektedirler.

Bu çalışma kapsamında da, Türkiye’deki bir telekomünikasyon şirketinin borçlu müşterilerinden alacağını tahsil edebilmesi için izlediği yolun optimize edilmesine çalışılmıştır. Bu kapsamda, şirket borçlu müşterilerini arayarak ilk olarak borcu kendisi tahsil etmeye çalışmaktadır. Ulaşamadığı veya tahsilat gerçekleştiremediği müşterilerini avukatlık bürolarına yönlendirmektedir. Daha sonrasında bu avukatlık büroları, tahsil edebildikleri tutar üzerinden bir performans sistemine dahil edilmektedirler. Bu performans sisteminden çıkan sonuç, büronun bir sonraki ay telekomünikasyon şirketinden devir alacağı müşteri sayısında etkili olmaktadır. Bu kapsamda, şirket, daha doğru değerlendirme yapabilmek için, bu bürolara atanmış olan dosyaların homojen zorlukta olmasını istemektedir. Eşit zorlukta yapılan atamalar sayesinde bürolar adil ve şeffaf bir şekilde değerlendirilebileceklerdir. Böylece hem bürolar için daha şeffaf bir yapı kurulmuş olacak hem de şirket doğru atamalar yaparak elde edeceği tahsilat tutarını artırmış olacaktır.

Yapılan tez çalışmasında, müşteri örnekleri incelenerek müşterinin borcunu ödeme potansiyeli öngörülecek sonrasında her müşteri için borcunu geri ödeme ihtimali hesaplanacaktır. Sonrasında ise bu ihtimal değerleri kullanılarak büro bazında homojen zorlukta müşteri ataması yapılması planlanmıştır.

Bu çalışma kapsamında sınıflandırma yapılmıştır. Sınıflandırma algoritmalarına göre olasılığı 0.5’ten büyük olanlar öder, yani 1, 0.5’ten küçük olanlar ödemez, yani 0 olarak işaretlenir. Bu çalışmada doğruluğu ölçmek için sınıflandırma algoritmaları kullanılmış olsa da, telekomünikasyon şirketinin atamaları gerçekleştireceği platforma sağlanacak veriler, ödeme olasılıklarının ortalaması kullanılarak atamaların homojen olması yönünde geliştirilmiştir.

Çalışma toplam altı alt bölümden oluşmaktadır. Birinci bölümde, genel çerçeveyi aktaran giriş bölümünde problem incelenecek ve tez çalışmasının amacından bahsedilecektir. Sonrasında, bu alanda yapılmış çalışmalar incelenecektir. Üçüncü bölümde, teknik alt yapı hakkında bilgi aktarımı gerçekleşecektir. Dördüncü bölümde, makine öğrenmesi işlemlerinden en doğru sonuçları elde edebilmek amacıyla veri üzerinde yapılan işlemlere değinilecektir. Ayrıca yine bu bölümde değinilen öznetelik seçimi için kullanılan farklı teknikler ile veri altkümeleri yaratılacaktır. Son bölümde ise veri kümesi üzerinde gerekli düzeltmeler yapıldıktan sonra, oluşturulmuş bu alt veri kümelerinin modeller üzerindeki performansları karşılaştırılacaktır. Son bölüm olan Bölüm 6’da ise tez çalışmasından elde edilen sonuçlar yorumlanmıştır.

İlk olarak, telekomünikasyon şirketinden temin edilen ~327K farklı müşteri içeren 2018 yılına ait veri kümesi üzerinde çalışma yapılmıştır. Veri temizlenip, boş değerler doldurulup, gereksiz veriler elenip, en son aşamada ilgili yeni öznitelikler de eklendikten sonra, birkaç farklı yöntem ile öznitelik önem sıralaması yapılacaktır. Bu sıralama ile veriler makine öğrenmesi modellerinde uygulandıktan sonra, en iyi doğruluk sonucunu hangi öznitelik alt kümesi ile belirlendiği ortaya konacaktır. En az sayıda öznitelik ile en doğru sonucu veren model bulunmaya çalışılmıştır. Telekomünikasyon şirketinden alınan 18 öznitelik, çalışma kapsamında 38 özniteliğe kadar türetilmiştir. Öznitelik seçim yöntemleri olarak, ağaç temelli seçim yöntemlerinin ortalama değerleri, Temel Bileşen Analizi, Özyinelemeli Lojistik Regresyon, Özyinelemeli Adaboost, ki-kare testlerinden elde edilen öznitelik önem listeleri en etkiliden en az etkiliye doğru sırasıyla, öznitelik sayısı artırılarak modellerde işlenmiştir. Çalışmada kullanılan modeller, Lojistik Regresyon (LR), Karar Ağacı (KA), Naive Bayes, K-en yakın komşu, Adaboost, Bagging, Rastgele Ormanlar, Gradyen Boosting, XGB (Extra Gradyen Boosting), Extra Ağaçlar ve Yapay Sinir Ağı (YSA)'dır. En iyi sonuç; 38 özniteliğin ağaç modellerinin uygulanmasıyla elde edilmiş önem sırasına göre ilk 33 öznitelik ile gradyen boosting modellemesiyle elde edilmiştir Bu değerin de 74,28 olduğu gözlenmiştir.

PREDICTION OF DEBT COLLECTION BEHAVIOUR WITH MACHINE LEARNING TECHNIQUES: A CASE STUDY ON TELECOMMUNICATION COMPANY CUSTOMERS

SUMMARY

In order for companies to maintain their assets, they must receive the provisions of the services they provide to their customers. Therefore, they can get their own funding and keep improving.

However, invoices that are not paid on time place a serious burden on companies. Instead of continuing to develop and putting forward new products, companies are trying to collect these receivables and are experiencing a loss of performance within their own structure and are unable to continue to develop and grow.

The burden imposed by the customer who does not pay the debt should not be considered solely as the inability to collect the debt. At the same time, the company's resources for collection must be taken into account. The company is also exposed to both its employee effort loss and company assets (telephone, electricity, computers, etc.) expenses. For this reason, some companies outsource their collection to customers who do not pay their debt. In addition to the customer's debt to the telecommunications company, the outsourcing office tries to collect the debt from the customer by adding its own amount.

Within the scope of this thesis, the telecommunication company is also carrying out the collection of unpaid invoices with external sources. The company tries to collect from its customers who do not pay their debts and who are in legal proceedings with the help of law offices.

The current operation of the process that the company wants to optimize is to try to reach the customer by the employees of the company before dispatching the debtor to external sources and make efforts to collect the debt. Customers who do not pay their debt for 120 days are referred to law offices. First, these offices try to reach the customer themselves and make an effort to collect the debt, and for those who do not pay their debts, a lawsuit is filed and thus the process is started.

Each month, the company transfers its customers who do not pay their debts in different amounts to the law offices for the collection to take place. Monthly average ~25K customers follow the law. This ~25K customer is distributed to offices in line with their performance. On a monthly basis, the company determines the performance of the offices according to the amount the offices manage to collect, and in line with this performance, the number of customer files to be assigned to that office in the following month is determined.

The company regularly collects collection information from the offices every month. In this direction, it performs a performance calculation, determining the number of customers to be assigned the following month. The offices prefer to call among new customers first, as new customers arrive each month. Because they don't want to lose a new customer who is easy to collect by taking care of a customer they didn't get, even though they tried to get paid earlier. They are able to return to the past months after completing their new customer calls that month. The following month, ~25K new customers will be assigned to the offices again. This process continues until the end of

the year. There are fewer appointments in December compared to other months. At the end of the year, all customer debts that have not been collected by the offices are recovered and redistributed.

In addition, offices are included in a performance system based on the amount they can collect from customers who have been transferred to them. The result from this performance system is effective in the number of customers that the office will take over from the telecommunications company next month. In this context, the company wants the customers assigned to these offices to be homogenous in order to make a more accurate assessment.

However, during this assignment process, the randomization of customer assignment can lead to subjective. For example, having high debt amounts assigned to an office or being clustered in a geographic area of Turkey where customers live may have an impact on the amount of collection.

Offices will be evaluated in a fair and transparent manner thanks to equally difficult and homogeneous assignments.

In addition, this performance account is required to be accurate both for the bureau to be able to continue to get new customers to make more money and for the company to increase the amount of collection it will obtain by making accurate assignments. If the company makes the right assignments, it can increase the amount of collection that it will get more customers to right office that is able to make the high collection. In order for this assignment to be fair, customers sent to offices must be of equal difficulty. The word difficulty meaning here is the possibility of collecting the customer's debt.

However, the company wants to make efforts for the most optimized and fair progress of this process.

In the thesis study, the customer samples will be examined and the potential to pay the customer's debt will be foreseen and the probability of paying back the debt will be calculated for each customer. After this, it is planned to assign customers in uniform difficulty on the basis of office by using these probability values.

Classification was made within the scope of this study. According to classification algorithms, those whose probability are greater than 0.5 that means customer pays, target variable is set to 1, those less than 0.5 that means customer doesn't pay, target variable is marked as 0. Although classification algorithms were used to measure accuracy in this study, data to be provided to the platform on which the telecommunications company would perform the assignments was developed to make the assignments homogeneous using the average of the payout probabilities.

First, work has been done on the dataset for 2018, which includes ~327K different customers sourced from the telecommunications company.

The 18 attributes taken from the telecommunications company are derived up to 38 attributes within the scope of the study.

After deciding the models to be applied on the data, the procedures for using the data to train the models are listed below. First, the content of the data received from the company will be mentioned and they will be visualized. Then, with the interpretation of the data, the operations on the data will be decided. Data cleanup, data demotion, data expansion, categorization, normalizing continuous data, coding categorized data has been performed. To further increase the accuracy rate, the operations of K-fold crossover, decomposing the dataset as training and testing, and random reordering of

the cluster were run on the data. In order to sort the attributes from the most important to the least effective in the next model-ready, cleaned dataset, the charts obtained by using attribute importance methods are mentioned. In attribute selection methods, the values obtained by normalizing and averaging the results obtained from attribute importance lists first obtained by tree-based and ensemble-based methods are listed. Then the basic component analysis method was applied to the dataset. The logistic regression and adaboost algorithms were then given to the recursively operating model, and their results are listed below, respectively. Finally, the severity of the attributes is listed according to the chi-square test.

On data set, features obtained from attribute inference and attribute selection stages from the most effective to the least effective, respectively, are increased and processed in models.

Decision trees, nearest neighbor, naive Bayes, logistic regression, artificial neural networks and ensemble methods are frequently preferred in the literature using supervised learning methods to estimate the behavior of customers on payment will be tried. In this section, models will be examined comparatively in terms of their performance in forecasting achievements.

After the model was trained with the training dataset, measurements were made so that the consistency and accuracy of the model could be observed by applying it to the test dataset as well. As a result of these measurements, decision trees and tree-based boosting methods and selected attribute subsets cannot achieve high success rates on Decision Trees models, while the subset that gives the best results on the boosting models has been. The best result was an extra gradient Boost algorithm for attribute ranking using tree-based methods, which achieved an accuracy rate of 74.28. Instead of using 38 attributes, a higher accuracy value was reached than we would get with 33 attributes and 38 attributes. Both the running time of the model has been shortened and its accuracy has increased. Attributes not included are '*transfer_maas_yatma_araligi*', '*transfer_gunu_araligi*', '*bayram_var*', '*legal_faiz*', '*legal_faiz_bolu_toplam*'. The meaning of these attributes is, in order of effect, whether the assignment of the client is in the period in which the salary will be paid, which period of the month the assignment comes to, whether it is feast or not, legal interest amount operated by the state over the amount of debt, the effect of the legal interest amount on the total debt. The model that gives the most accurate result with the least number of attributes has been attempted.

The study consists of a total of six subdivisions. In the first section, the problem will be examined in the introduction section which conveys the general framework and the purpose of the thesis study will be mentioned. Later, studies in this field will be examined. In the third section, information about the technical infrastructure will be transferred. In the fourth section, the processes performed on the data will be discussed in order to obtain the most accurate results from machine learning processes. In addition, data subsets will be created with the different techniques used for attribute selection mentioned in this section. In the last section, the performance of these created sub-datasets on the models will be compared after the necessary corrections are made on the dataset. In the last chapter, Section 6, the results from the thesis study were interpreted.



1. GİRİŞ

Tüm sektörlerde olduğu gibi telekomünikasyon sektöründe de rekabetin artması yeni müşteriler edinmenin zorlaşmasına neden olmuştur. Bu nedenle telekomünikasyon ve bankacılık sektöründeki şirketler makine öğrenmesi teknikleri ile müşterilerine özel kampanyalar sunarak müşteri sayılarını arttırmaya veya eldeki müşteri tutmaya çalışmaktadırlar. Ancak makine öğrenmesinin uygulandığı tek alan müşteri kazanma veya müşteri sadakati konusundaki çalışmalar değildir. Şirketlerin gelişimlerini devam ettirebilmeleri için kazandıkları müşterilerden kazanç sağlamaları gerekmektedir. Eğer müşteri borcunu ödemiyorsa, müşteri kazanmak şirket için karlı bir durum olmaktan çıkıp, şirkete fazladan yük getirmektedir. Bu yük, sadece borcun tahsil edilememesi olarak düşünülmemelidir. Aynı zamanda tahsilat için şirketin ayıracağı kaynaklar da hesaba katılmalıdır. Şirket hem çalışanlarını hem de şirket varlıklarını (telefon, elektrik, bilgisayar vb.) bu kanala sevk ederek zarara uğrayacaktır. Bu nedenle bazı şirketler borcunu ödemeyen müşterilerinden tahsilat gerçekleştirmek için dış kaynak kullanmaktadırlar. Dış kaynak (büro), müşterinin telekomünikasyon şirketine olan borcuna ilave olarak, kendi alacağı tutarı da ekleyerek müşteriden borcu tahsil etmeye çalışır.

Bu tez kapsamında incelenen telekomünikasyon şirketi de ödenmeyen faturaların tahsilatını dış kaynaklar ile gerçekleştirmektedir. Şirket, borcunu ödemeyen ve yasal takibe girmiş müşterilerinden avukatlık bürolarının yardımıyla tahsilat yapmaya çalışmaktadır.

1.1 Telekomünikasyon Şirketinin Yürüttüğü Süreç

Şirketin optimize etmek istediği sürecin mevcut işleyişi; alacaklı olunan müşteriyi dış kaynaklara sevk etmeden önce, müşteriye şirket çalışanları tarafından ulaşılmaya çalışılıp borcun tahsilatı için çaba harcanır. Bu süreye idari takip süresi denir. 120 gün boyunca borcunu ödemeyen müşteriler, avukatlık bürolarına yönlendirilir. İlk olarak, bu bürolar müşteriye kendileri ulaşmaya çalışıp, borcun tahsilatı için çaba harcarlar, borçlarını ödemeyenler için ise dava açılır ve böylece yasal takip süreci başlamış olur.

Şirket, her ay, farklı tutarlarda borcunu ödemeyen müşterilerini avukatlık bürolarına tahsilat işleminin gerçekleşmesi için devretmektedir. Aylık olarak ortalama ~25K müşteri yasal takibe girmektedir. Bu ~25K müşteri, bürolara performansları doğrultusunda dağıtılmaktadır. Şirket, aylık bazda, büroların tahsil etmeyi başardığı tutara göre büroların performanslarını belirlemekte ve bu performans doğrultusunda, o büro için, bir sonraki ay atanacak dosya sayısı belirlenmektedir.

Şirket, her ay düzenli olarak, bürolardan yaptıkları tahsilat bilgisini topluyor. Bu doğrultuda, bir performans hesabı yaparak, bir sonraki ay atanması gereken dosya sayısını belirliyor. Bürolar, her ay yeni müşteriler geldiği için, ilk olarak yeni müşteriler arasında arama gerçekleştirmeyi tercih ediyorlar. Çünkü tahsilat yapması kolay olan yeni bir müşteriye, daha önce ödeme almayı denedikleri halde alamadıkları bir müşteri ile ilgilenerek kaybetmiş olmak istemiyorlar. O ayki yeni müşteri aramalarını tamamladıktan sonra geçmiş aylara geri dönebiliyorlar. Bir sonraki ay, tekrar ~25K yeni müşteri bürolara atanıyor. Bu süreç yılsonuna kadar böyle devam ediyor. Aralık ayında diğer aylara oranla daha az sayıda atama yapılıyor. Yılsonunda bürolar tarafından tahsil edilememiş tüm müşteri dosyaları geri alınır ve yeniden dağıtım sağlanır.

1.2 Tezin Amacı

Avukatlık büroları, borçlu müşterilerden yaptıkları tahsilat tutarına göre bir performans sistemine dahil edilmektedirler. Bu performans doğrultusunda, büroya, bir sonraki ay atanacak müşteri sayısı belirlenmektedir. Müşterilerin bürolara dağıtımı, büronun performansı dolaylı olarak da geliri üzerinde doğrudan etkilidir. Bu dağıtım işlemi sırasında, müşteri dağıtımının rassallığı, subjektifliğe neden olabilir. Örneğin bir büroya borç tutarı yüksek müşterilerin atanmış olması veya müşterilerin yaşadığı bölgenin Türkiye'nin coğrafi bir bölgesinde kümelenmiş olması tahsilat miktarı üzerinde etkili olabilir.

Ayrıca bu performans hesabının doğru olması hem büronun gelir elde etmeye devam edebilmesi için hem de şirket doğru atamalar yaparak elde edeceği tahsilat miktarını artırabilmesi için gereklidir. Şirket doğru atamalar yaparsa, yüksek tahsilat yapabilecek büroya çok sayıda müşteri dosyası, yüksek tahsilat yapamayacak büroya da o doğrultuda müşteri dosyası ile elde edeceği tahsilat miktarını artırabilir. Bu atama işleminin adil olması için, bürolara gönderilen dosyaların eşit zorlukta olması

gerekmektedir. Burada zorluk kelimesi ile kastedilen, müşterinin borcunu kapama ihtimalidir.

Müşterilerin dağıtımının olabildiğince objektif olması, bürolara müşterilerin atamasını yapan şirket çalışanının etkisinin ortadan kaldırılması, bürolar açısından daha şeffaf bir sisteme dahil edilmiş olmak ve müşteri atama süresinin kısaltılması sayesinde tahsilat sürecinin daha hızlı başlayacak olması müşterilerin puanlanması ile sağlanacaktır. Ayrıca şirket açısından, tahsilat alması kolay olan müşteriler sayesinde yüksek performans notu alan büroya, sonraki ay atanacak daha fazla tutardaki müşterilerden alınması olası tahsilat gerçekleştirilemeyebilir. Çünkü büro gerçek performansını yansıtmayan bir puan almıştır. Bir sonraki ay, yüksek tahsilat beklentisi ile büro üzerinde baskı oluşacak ve belki de hedefi tutturmak için daha agresif bir tutum sergilemeleriyle şirketin itibarının müşteri gözünde zedelenmesiyle sonuçlanabilecek zararlar meydana gelebilir. Bunun tam tersi de mümkündür. Yapılmış yanlış atamalar nedeniyle, düşük performans puanı almış bir büroya, daha az müşterinin atanması, tahsilat potansiyeli yüksek olan bir bürodan yeteri kadar faydalanamamak şirketin gelirlerinde olumsuz etkiye neden olacaktır.

Bu çalışma ile yasal takibe girmiş müşterilerin borçlarını ödeme ihtimalleri üzerine makine öğrenmesi yöntemleri ile tahmin gerçekleştirilecektir. Bürolara atanan müşterilerin ortalama bir puanda olması ile adil bir atama sistemi kurulmaya çalışılacaktır.

1.3 Çalışılan Şirket Bilgileri

Çalışılan şirketin bilgilerine bu tez kapsamında yer verilmesinin nedeni ise; şirketten alınan verilerin güvenilirliği ve şirketin pazar hakimiyetinin gösterilmek istenmesidir. Pazara hakim bir şirket ile çalışarak farklı özellik gösteren pek çok müşteri verisini inceleme fırsatı elde edilmiştir. Şirketin kurumsal web sayfası üzerinden verdiği bilgilere göre, Türkiye’de mobil iletişim, 1994 yılında Turkcell’in hizmet vermeye başlamasıyla hayata geçmiştir. 1998 yılında T.C. Ulaştırma Bakanlığı ile 25 yıllık Telekomünikasyon lisans anlaşması imzalayan Turkcell, müşterilerine sunduğu hizmetlerin çeşitliliğini, kalitesini ve buna bağlı olarak müşteri sayısını da artırarak, 35 milyonu aşkın müşteri sayısına ulaşmıştır. Şirket, Türkiye pazarında lider olmasına rağmen, gelişimini sürdürmeye devam etmektedir. Her alanda geliştirdiği projelerde en optimize süreçler ile müşterilerini mutlu etmeye devam etmektedir.

Hisseleri 11 Temmuz 2000’de Borsa İstanbul (BİST) ve New York Stock Exchange’de (NYSE) eşzamanlı olarak işlem görmeye başlayan Turkcell, NYSE’ye kote olan tek Türk şirketi unvanına sahiptir. Bu çalışma ile de Turkcell, süreçleri beraber yürüttüğü paydaşları ile sürdürdüğü çalışma yöntemini optimize etmeye çalışmaktadır.

1.4 Tezin Yapısı

Bu çalışmada, makine öğrenmesi teknikleri ile müşterilerden tahsil edilememiş borçların alınıp alınamayacağına dair bir ihtimal hesaplanarak, bürolara tahsilat için atanan müşterilerin ödeme ihtimallerinin ortalamalarının aynı olması hedeflenmiştir. Böylece, performans hesaplanması sisteminin daha adil bir hale getirilmesine çalışılmıştır. Bu çalışma kapsamında sınıflandırma yapılmıştır. Sınıflandırma algoritmalarına göre olasılığı 0.5’ten büyük olanlar öder, yani 1, 0.5’ten küçük olanlar ödemez, yani 0 olarak işaretlenir. Bu çalışmada doğruluğu ölçmek için sınıflandırma algoritmaları kullanılmış olsa da, telekomünikasyon şirketinin atamaları gerçekleştireceği platforma sağlanacak veriler, olasılıkların ortalamasının homojen olması yönünde geliştirilmiştir.

Çalışma toplam altı alt bölümden oluşmaktadır. Birinci bölümde, genel çerçeveyi aktaran giriş bölümünde problem incelenecek ve tez çalışmasının amacından bahsedilecektir. Sonrasında, bu alanda yapılmış çalışmalar incelenecektir. Üçüncü bölümde, teknik alt yapı hakkında bilgi aktarımı gerçekleştirilecektir. Dördüncü bölümde, makine öğrenmesi işlemlerinden en doğru sonuçları elde edebilmek amacıyla veri üzerinde yapılan işlemlere değinilecektir. Ayrıca yine bu bölümde değinilen öznitelik seçimi için kullanılan farklı teknikler ile veri altkümeleri yaratılacaktır. Son bölümde ise veri kümesi üzerinde gerekli düzeltmeler yapıldıktan sonra, oluşturulmuş bu alt veri kümelerinin modeller üzerindeki performansları karşılaştırılacaktır. Son bölüm olan Bölüm 6’da ise tez çalışmasından elde edilen sonuçlar yorumlanmıştır.

Bu amaçla oluşturulmuş veri kümesinde ~327K farklı müşteri için 17 öznitelik bilgisi telekomünikasyon şirketinden temin edilmiştir. Bu çalışma ile incelenen veri kümesinin tamamı yasal takibe girmiş müşterilerden oluşmaktadır. Eğitim veri kümesindeki müşteriler üzerinde yapılan modelleme çalışmalarından sonra, test veri kümesi üzerinde en iyi sonucu veren modelleme algoritması ile en iyi algoritmaya karar verilecektir. Böylece, yasal takibe yeni düşmüş bir müşterinin en iyi sonucu veren modelleme algoritması ile değerlendirilmesi sonucunda ödeme olasılığı

hesaplanacaktır. Tüm müşteriler için bu hesaplama yapıldığında (~25K) bürolara homojen bir şekilde atamalarının yapılması sağlanacaktır.





2. LİTERATÜR ARAŞTIRMASI

Bu bölümde, bu çalışmaya benzer literatür çalışmaları incelenecektir. Literatürdeki çalışmalar bu kapsamda dörde ayrılmıştır. İlk olarak, bu tez çalışması ile konu olarak benzer olan müşterinin risk tahmini üzerine çalışmalar incelenecektir. Devamında, bu çalışmada da uygulanan öznitelik alt kümesi seçimi ile ilgili çalışmalar incelenecektir. Sonrasında bu çalışma kapsamında kullanılmış metotların kullanımının ne amaçlarla ve nasıl kullanıldığı anlatılacaktır. En son olarak da, bu çalışmada kullanılmış değerlendirme metrikleri ile ilgili literatür çalışmalarından bahsedilecektir.

Ayrıca, literatürde, müşterilerin geri ödemelerine dair çalışmalar çoğunlukla bankalar için kredi puanlaması amacıyla kullanıldığı görülmüştür. Telekomünikasyon sektörü için yapılan çalışma sayısı literatürde oldukça azdır. Bu nedenle, incelenecek örnekler çoğunlukla bankacılık ağırlık olacaktır.

Telekomünikasyon sektörüne ait az çalışma olmasının bir sebebi de, başvuru esnasında müşterilerden talep edilen bilgilerin yetersizliğidir. Bir kişi, banka kredisi almak için başvuruda bulunduğu anda kendisine ait pek çok bilgiyi, mal varlığını, çalışma durumunu, mesleğini beyan etmektedir. Bilginin doğruluğu, güvenilirliği ispatlamaktadır. Bu nedenler sebebiyle bankacılık sektörü ile ilgili daha fazla çalışma vardır.

2.1 Konu Olarak Benzer Çalışmalar

Bu tez kapsamında yapılan çalışma ile bir telekomünikasyon şirketinin borçlu müşterilerinin borçlarını ödeme ihtimalleri ile ilgili tahmin gerçekleştirilecektir. Bu çalışma hedefine benzer bir işlemi bankalar kredi talep eden müşterileri için, kişinin risk analizini yaparak geri ödeme yapıp yapmayacağı konusunda bir tahminde bulunmaktadırlar. Literatür incelemesinde görülmüştür ki, bankalar bu tahmin için makine öğrenmesi tekniklerinden sıklıkla faydalanmışlardır. Kullanmış oldukları bu teknikler, bu çalışmada da kullanılmış olduğundan dolayı, aşağıda, öncelikli olarak telekomünikasyon sektörü, sonrasında bankacılık sektörü örnekler ile ele alınacaktır.

2.1.1 Telekomünikasyon sektörü ile ilgili çalışmalar

Madeira ve arkadaşları tarafından 2003 yılında yapılan çalışmada, Portekiz'deki bir telekomünikasyon şirketinin müşterilerini düşük, orta ve yüksek olmak üzere üç risk grubuna ayırmışlardır. Çalışmada, müşterilerin altı aylık verisi kullanılarak, sonraki üç ay içerisinde hangi risk grubunda yer alacağı tahmin edilmeye çalışılmıştır. Altı aylık verisi bulunmayan yeni müşteriler için ise, son üç aya ait verileri kullanılarak, bir sonraki ay için tahminde bulunulmuştur. Tahminler için çoklu lojistik regresyon, yapay zeka ve karar ağaçları algoritmalarından faydalanılmıştır. Sonuçların doğruluğu ve yorum yapma kolaylığı nedeniyle karar ağaçları modeli en uygun model olarak seçilmiştir. Sonrasında, müşteriler hakkında daha açıklayıcı bilgiler bulunduran öznitelikler de eklenerek yeniden karar ağacı algoritması ile modellemişlerdir. Bu iki aşamalı işlem ile müşterilere ait borçlu duruma düşme risk tahminleri yapılmıştır. Bu iki aşamalı sistemin, ilk kullandıkları sisteme göre %15 daha doğru sonuçlar ürettiğini görmüşlerdir.

2.1.2 Kredi geri ödeme risk tahmini ile ilgili çalışmalar

Müşterilerin kredilerini geri ödemeleri ile ilgili çalışmalar, müşterilerin faturalarını geri ödeme ihtimallerinin puanlanması ile benzerdir.

Geri ödeme ile ilgili risk analizi konularında çoğunlukla bankalar için yapılmış literatür çalışmalarına rastlanmıştır. Bu çalışmalarda, kredi verme konusunda karar almaya yardımcı olacak sistem geliştirilmesi amaçlanmıştır.

İtalya'da yapılan bir çalışmada Liberati ve ark. (2017) demografik veriler ile birlikte finansal veriler kullanılarak 3 ayrı metot ile analiz yapılmıştır. Yapılan analizler sonucunda lojistik regresyon modelinin tahmin gücü %79.70, kernel diskiriminant analizinin tahmin gücü ise %74.30, diskiriminant analizinin tahmin gücü %58.08 olarak belirlenmiştir.

Zhu ve diğerleri (2016) ise Çin'deki bir bankanın müşterileri üzerinde yaptıkları çalışmada Adaboost yöntemini kullanarak, müşterileri puanlamaya çalışmışlardır. Bu çalışmadaki esas amaçları; temel eğitim veri kümesinden rastgele seçilmiş ve kendinden bir önceki sınıflama metodunun yanlış hesapladıklarını içeren yeni bir veri kümesi ile sınıflandırıcıyı eğitmişlerdir. Bunun ile sistemin aşırı öğrenmesinin önüne geçmeyi hedeflemişlerdir. Deneyler sonucunda en iyi ölçüm olarak %86,74'ü elde edebilmişlerdir.

2010 yılında Gurny tarafından yapılan bir başka çalışmada, bankaların geri ödeme performansı modellenmeye çalışılmıştır. 298 bankaya ait veriler kullanılarak, bankaların ödeme güçlüğüne düşüp düşmeyeceği tahmin edilmeye çalışılmış. Bunu tahmin edebilmek için de t-1 dönemindeki verilerle, t dönemindeki durum tahmin edilmeye çalışılmıştır. Kurulan modelde lojistik ve probit yöntem kullanılmıştır. Lojistik regresyonun tahmin gücü %98.47, probit yöntemin tahmin gücü %98.27 olarak belirlenmiştir.

Bir başka çalışmada ise Mihalovic, lojistik regresyon ve diskriminant analizi ile model oluşturularak, modellerin tahmin gücü karşılaştırılmıştır. 236 şirkete ait 2014 yılı verileri kullanılmış olup, bu şirketlerin 118'i iyi, 118'i ise kötü olarak sınıflandırılmış şirketlerdir. Analizde 5 adet bağımsız öznitelik olarak kullanılmıştır. Diskriminant analizi ile oluşturulan modelin tahmin gücü %64.41, lojistik regresyon ile oluşturulan modelin tahmin gücü ise %68.64 olarak belirlenmiştir. Modelin tahmin gücüne katkı yapan en anlamlı öznitelikler net gelir, toplam varlıklar, cari oran, yükümlülükler ve toplam varlıklar olarak saptanmıştır.

Moares ve arkadaşlarının yaptığı çalışmada, müşterileri profillerini, kredi vermek konusunda puanlamak için iyi veya kötü olarak sınıflandırmışlardır. Bu sınıflandırma için kullandıkları yöntem, C4.5 karar ağacı ve YSA olmuştur. Yapılan çalışmada, C4.5 karar ağacı ile %90.07 doğruluk oranı elde ederlerken, YSA ise %95.58'lik bir doğruluk oranı elde etmişlerdir.

Kim ve Sohn, teknoloji kredisine başvuran müşterilerine kredilerin sağlanıp sağlanamayacağına objektif bir şekilde karar vermek için destek vektör makineleri yöntemine başvurmuşlardır. DVM'nin yanı sıra yapay sinir ağları ve lojistik regresyon metotlarına da başvurarak yöntemlerin çıktılarını karşılaştırmışlardır. Müşterilerinin şirket finansal oranlarını, teknoloji özelliklerini ve ekonomik faktörlerini öznitelik olarak kullanarak, en iyi yöntem olarak %66.16 ile DVM'ye karar vermişlerdir. İkinci olarak en iyi performansı da YSA yöntemiyle %64.23'ü hesaplarlarken, son olarak LR ile de %64.16 sonucunu elde etmişlerdir.

Chen, Xiang, Liu ve Wang, 2012 yılında yaptıkları çalışmayla, krediye başvurmuş yeni bir müşteriye kredi verilip verilmeyeceğine dair çalışmışlardır. En doğru puanlama modeli oluşturmak için ilk olarak veri kümesini kümeleme sonrasında da sınıflandırma modellerinde incelemişlerdir. Yeni başvuru sahiplerini kümeleme

algoritmaları sayesinde homojen kümeler halinde gruplandırdıktan sonra kümeden uzak olan örnekler silinmiş, tutarsız veriler tekrar etiketlenmiştir. Sınıflandırma modeli olarak destek vektör makineleri tercih edilmiş olup, ayrıca kredi tahmininde daha doğru sonuçlara ulaşmak için iki sınıflandırma yapmak yerine dört farklı sınıfa ayırmışlardır.

Koh ve Low, 2004 yılında yaptıkları araştırmada 1980-1987 yılları arasında faaliyet göstermiş 165 başarılı ve 165 başarısız işletmeyi lojistik regresyon, karar ağacı ve yapay sinir ağı modellerini kullanarak analiz etmişlerdir. Çalışma neticesinde finansal başarısızlık tahmininde en etkili modelin karar ağacı modeli, en az etkili olan modelin ise lojistik regresyon modeli olduğu sonucuna varmışlardır.

2.2 Öznitelik Seçimi ile İlgili Diğer Çalışmalar

Wang, 2012 yılında kredi başvurusunda bulunan KOBİ şirketleri değerlendirirken sadece şirketlerin ekonomik durumlarını değil, aynı zamanda kurucularının öznitelik özelliklerinin de geri ödeme tahmininde etkili olduğunu savunmuştur. Bu nedenle, şirketlerin finansal durumlarına ek olarak, kurucularının kredi özelliklerini de modelinde kullanmıştır. LR ile elde ettiği sonuçlara göre, KOBİ şirketlerinin kurucularının yaşı ve kişisel kredi bilgileriyle ilgili özniteliklerin risk analizi üzerinde büyük bir etkiye sahip olduğunu ispatlamıştır.

Grunert ve Norden Almanya'nın önde gelen bir bankasının 160 ticari müşterisine ait 1992-1996 yılları arasındaki verileri ile finansal olmayan özniteliklerin tahmin gücüne nasıl etki ettiğini tespit etmek amacıyla bir model oluşturulmuş, yalnızca finansal verilerle yapılan analizin tahmin gücü %88,24 olurken, finansal olmayan veriler de dahil edildiğinde modelin tahmin gücü %91,42'ye yükselmiştir.

Vladimir Bugera, Hiroshi Konno ve Stanislav Uryasev yaptıkları ortak çalışmada, 1995-1996 yılı içerisinde Yunanistan Ulusal Bankası'na yapılmış olan 150 kredi kartı başvurusunu matematiksel programlama algoritmaları ile inceleyip başvuranlar arasında bir puanlama çalışması yapmışlardır. Ellerindeki veri kümesini 25 öznitelikten 7 özniteliğe düşürerek fayda fonksiyonunun en optimize sonuç üretmesine çalışmışlardır. Çalışmada başvuran kişiler 3 grup halinde incelenmiştir: kabul, red, belirsiz. Çalışmada ikinci dereceden olan doğrusal fayda fonksiyonunun çeşitli kısıtlar ile farklı sınıfların performansları karşılaştırılmıştır. Karşılaştırma sonucunda, en iyi örneklem dışı öznitelikleri bulabilmek, fayda fonksiyonunun katsayıları üzerindeki

pozitiflik sınırları ile birlikte ikinci dereceden fayda fonksiyonları tarafından mümkündür.

Šušteršič ve arkadaşları (2009) ise çalışmalarında YSA modelleri ile finansal kuruluşlar için tüketici kredilerini puanlamaya üzerine yoğunlaşmışlardır. Veri kümelerinde bulunan özniteliklerin birbirleri ile ilişkili olmasından ve kredi puanlaması yapılması aşamasında kullanılıp kullanılmamasından emin olamadıkları için modeli oluşturmadan önce öznitelik altküme seçiminin gerekliliğini belirtmişlerdir. Altküme seçimi için genetik algoritmaya başvurmuşlardır. İlk olarak veri kümesinde 67 öznitelik varken, ilk olarak 21'e düşürmeyi başarmışlardır. Sonrasında ise YSA ve LR modellerinin kredi puanlaması üzerindeki performanslarını karşılaştırmışlardır. Bunun sonucunda GA ile öznitelik boyutunu azalttıkları veri kümeleri üzerinde YSA modeli en iyi sonucu vermiştir.

Chen and Lin (2006), DVM ile birden fazla özellik seçim yöntemi kullanımının performans üzerindeki etkisini araştırmışlardır. Çalışmalarında, beş farklı veri kümesi kullanılmaktadır. F-Ölçütü, Rastgele Orman, DVM yöntemleriyle belirledikleri özellik alt kümeleri ile DVM kullanımında, beş veri kümesinden sadece dördünde performans artışı sağlanmıştır. Performans ölçümünde '*dengelemiş hata oranı*' kullanılmaktadır. DHO, yanlış tahmin edilmiş pozitif örneklerin tüm pozitif örneklere oranı ile yanlış tahmin edilmiş negatif örneklerin tüm negatif örneklere oranının toplamının yarısına eşittir.

Masud ve ark. (2011) sınıflanması istenen nesneler oluşturulurken birden fazla öznitelik tanımlanabilmektedir. Bu özelliklerin bazıları sınıflama açısından gereksiz veya konuyla tamamen ilgisiz olabilmektedir. Bu tür özelliklere gürültü denmektedir. Hızlı ve verimli bir sınıflama için, gürültü niteliğindeki özelliklerin ayıklanarak ayırıcı niteliğe sahip özelliklerin bulunması gerekmektedir. Özellik seçimi denen 3 bu işlem, özniteliklerin en uygun alt kümesinin bulunmasına dair bir araştırma konusudur.

Jha ve Ragha (2013), geliştirdikleri modelde, k-ortalamlar algoritması kullanılarak belirlenen öznitelikleri kullanarak modeli eğitmişlerdir. Önerilen modelde eğitim ve test süresinde azalma ve isabet oranında artış gözlemlenmiştir.

Salem tarafından 2012 yılında, gelişmiş bir öznitelik seçim yöntemi önerilmiştir. Bu yöntemle kullandığı sınıflandırma yöntemlerinde isabet oranında artış ve hatalı tespitlerde azalma sağlamıştır.

2.3 Doğruluk Artırıcı Yöntem Olarak Benzer Çalışmalar

Bu başlık altında, çalışma sırasında sonucun doğruluğunu artırmak için kullanılan yöntemlerden bahseden çalışmalara değinilecektir.

Wang ve diğerleri, 2009 yılında yaptıkları araştırmada özellik normalizasyonunun tespit performansını, k-en yakın komşu, temel bileşen analizi ve DVM teknikleri üzerinde artırdığını ortaya koymaktadır.

Olsen ve arkadaşları, iflas tahmini yapabilmek için veri kümesini, K tane ayrışık ve yaklaşık olarak eşit büyüklükte alt gruplara ayırarak, bu gruplardan biri, test bölümü olarak ayrıldıktan sonra, geriye kalan K-1 grup üzerinden modeli eğitmişlerdir. Eğitim bittikten sonra doğruluk performansı ayrılmış olan test grubu üzerinden hesaplanmıştır. Bu işlem diğer tüm parçaların da en az bir kere test veri kümesi olacağı şekilde K kere tekrarlanmaktadır. Son olarak, K tane farklı doğruluk değerinin ortalaması alınmaktadır. 10-kat çapraz doğrulamadan elde edilen bir tahminin, yaygın olarak tek bir doğrulama kümesi kullanılarak uygulanan bir tahminden daha güvenilir olduğunu ortaya koymuşlardır.

2.4 Değerlendirme Metrikleri ile İlgili Kredi Puanlama Çalışmaları

Baesens ve arkadaşlarının üzerinde durduğu konu ise, modellerin farklı veri kümeleri için farklı sonuçlar üretmesi üzerine olmuştur. Bir veri kümesi üzerine uygulanan modellerde, bir model diğer modellere oranla büyük bir doğruluk oranına sahip olsa da, başka bir veri kümesi üzerinde uygulandığında başka bir model en yüksek doğruluk oranına sahip olmaktadır. 8 farklı veri kümesi üzerinde yaptıkları çalışmada uyguladıkları makine öğrenmesi teknikleriyle finansal kuruluşlara ait kredi puanlaması çalışmasını yürütmüşlerdir. Uyguladıkları modeller; lojistik regresyon, lineer diskriminant analizi, quadratik diskriminant analizi, lineer programlama, dört farklı destek vektör makinesi uygulaması, yapay zeka, naive bayes, ağaç yapısı eklenmiş naive bayes, dört farklı karar ağacı algoritması, iki farklı k-en yakın komşuluk sınıflandırıcılarıdır. Sınıflandırma metotlarının performansını ROC altında kalan alan ve doğru sınıflandırılan vakaların yüzdesi (DSY) ile değerlendirilmiştir. Kullanılan modellerin çoğu kabul edilebilir sonuçlar üretse de destek vektör makineleri ve yapay zeka modelleri hem DSY hem ROC için en yüksek performansı göstermişlerdir.

Arunasakthi ve diğ. 2014 yılında, çalışmalarında üç farklı veri kümesi üzerinde DVM makinesinin performansının F-Ölçüt analiz yöntemi ile belirlenen öznelik alt kümesinin kullanılmasıyla geliştirilmesini incelemişlerdir. Çalışma sonucunda, daha az sayıda öznelik ile DVM kullanarak, üç veri kümesinden ikisinde isabet oranının arttığı, üçüncü veri kümesinde ise DVM'nin yalnız kullanımına eşit isabet oranı elde edildiği görülmektedir. Çalışma süresindeki değişimi dikkate almamışlardır.





3. KULLANILAN YÖNTEMLER

Bu bölümde, müşterinin borcunu kapatma ihtimaline yönelik tahmin çalışması sırasında, bu çalışmada kullanılan yöntemlerin, çalışma prensipleri hakkında genel bilgi verilmiştir. Bu çalışma kapsamında kullanılan gözetimli öğrenme yöntemleri; komşuluk yöntemi olarak k-en yakın komşu yöntemi, karar ağaçları yöntemi, Naive Bayes yöntemi, yapay sinir ağları yöntemi, lojistik regresyon yöntemi ve topluluk yöntemleridir.

3.1 K-En Yakın Komşu (K-Nearest Neighbour)

K-en yakın komşu sınıflandırıcısı, ilk olarak Cover and Hart tarafından 1967 yılında önerilmiştir. K en yakın komşu sınıflandırma algoritmasında, eğitim kümesindeki örnekler n-boyutlu nümerik öznitelikler ile belirtilmektedir. Eğitim kümesindeki her bir örnek, n-boyutlu uzayda, bir noktayı belirtecek şekilde konumlandırılır. Uzay, eğitim kümesindeki örnekleri, hedef değişkenine göre bölgelere ayrılır. Hedef sınıfı bilinmeyen bir örnek için, eğitim kümesindeki verilerle benzerliği çeşitli yöntemlerle hesaplanarak, yeni örneğe en yakın olan k verinin ortalaması alınır ve belirlenen eşik değere göre hangi sınıfta yer alacağına karar verilir. Bu nedenle, k-en yakın komşu yöntemi “birbirine yakın olan nesneler aynı sınıfa aittir” prensibine göre çalışmaktadır.

Algoritmanın adımları şu şekildedir:

- 1-) Tüm veriler sayısal değerlere dönüştürülür.
- 2-) K sayısı yani en yakın kaç tane komşuya bakılacağı belirlenir.
- 3-) Sınıflandırılacak verinin, eğitim kümesindeki örneklere olan uzaklığı seçilen algoritmayla hesaplanır.
- 4-) Minimum uzaklıktaki en yakın k adet komşu bulunur ve ortalaması alınarak eşik değeri belirlenir.
- 5-) Belirlenen eşik değerine göre sınıflandırılacak olan verinin sınıf ataması yapılır (Han, 2001).

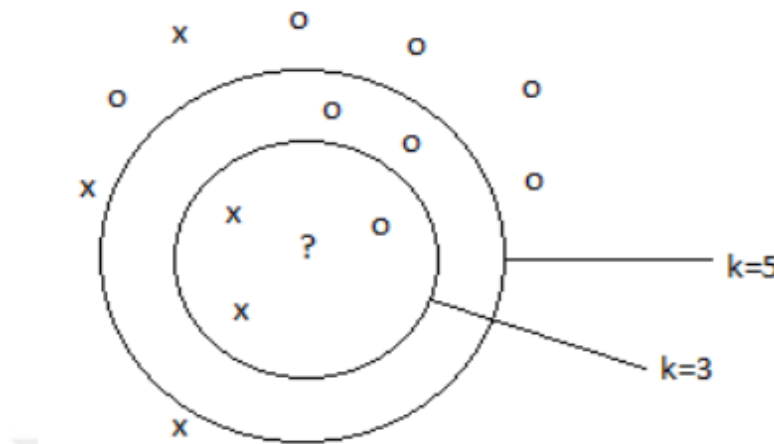
Yöntemin performansı üzerinde etkili parametreler; öznitelik sayısı, eğitim kümesindeki verilerin sayısı, k en yakın komşu sayısı, hedef sınıf ile ilişkisiz öznitelikler, uzaklık ölçütü ve ağırlıklandırma yöntemi sonucu etkilemektedir.

Öznitelik sayısının artması boyut artışına sebep olduğundan dolayı, boyutun kapladığı bölgeye düşen nokta sayısı azalacaktır. Bundan dolayı sınıflandırma başarısı oldukça düşecektir. Öznitelik sayısının model başarısı üzerindeki etkisi gibi, eğitim kümesindeki veri sayısı da önemlidir. Noktaların bölgedeki dağılımı üzerinde değişiklik yaratacağından, eğitim kümesinin az veri içerdiği durumlarda, model etkili sonuç üretmemektedir.

En yakın komşu sayısı parametresi değeri 1 olarak atanırsa, yeni gelen örnek en yakın komşusunun sınıfına dahil edilecektir. K değerinin ataması, kritik öneme sahiptir. Çünkü yeni verinin atanacağı sınıfa karar verilmesinde doğrudan etkilidir.

Ancak özellikle k değerinin 1 atandığı durumda model, gürültülü veya aşırı uç verilere karşı zayıf performans gösterecektir. K değerinin büyük atanması durumunda ise, modelin hesaplama yükünün artmasına ve tek bir sınıf olarak görünmesine sebep olmaktadır.

K sayısı ile ilgili olan bir diğer unsur ise k sayısının çift değer almamasıdır. Eşitlik söz konusu olduğunda örneğin hangi sınıfa dahil edilmesi ile karmaşa yaratacaktır. Bir veri sadece bir sınıfa dahil olabilir (Thomas, Edelman, & Crook, 2002). Bu tez çalışması kapsamında k değeri 5 olarak atanmıştır.



Şekil 3.1 : K-en yakın komşu algoritması $k=3$ ve $k=5$ için.

Şekil 3.1’de, k değerinin seçimi modelin performansı üzerinde doğrudan etkili olduğu görülmektedir. Şekilde, k değeri 3 alındığında yeni gelen veri “x” sınıfına; 5 alındığında ise “o” sınıfına dahil olacağı görülmektedir.

Hedef sınıf ile ilişkisiz öznitelikler, verilerin uzayda konumlandırılması üzerinde etki sahibi olacaklar ancak bu etki, noktanın ait olacağı sınıf ile ilişkisiz olacaktır. Noktaların uzaydaki dağılımını etkilediğinden dolayı modelin başarı oranında düşüşe neden olacaktır (Bhatia and Mantana, 2010).

Yeni gelen veriye en yakın komşunun uzaklığını belirleme faktörü olarak ise Manhattan uzaklığı, Öklid uzaklığı, Minkowski uzaklığı gibi ölçüm kriterleri uygulanmaktadır. Bu faktörler de modelin ürettiği sonuçların doğruluğu üzerinde etkilidir (Keller, Gray, Givens, 1985). Bu çalışma kapsamında, uzaklık ölçütü olarak denklem 3.3’teki Minkowski uzaklık yöntemi kullanılmıştır.

$$d_{öklid} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (3.1)$$

$$d_{Manhattan} = \sum_{i=1}^n |x_i - y_i| \quad (3.2)$$

$$d_{Minkowski} = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}} \quad (3.3)$$

Yöntemin performansını etkileyen faktörlerden birisi de ağırlıklandırmadır. Veri, kendisine yakın olan komşuya fazla ağırlık, uzak olana düşük ağırlık vererek, ait olduğu sınıfın en doğru şekilde bulunmasına çalışılır. En çok kullanılan ağırlıklandırma yöntemleri, d komşular arası uzaklık olmak üzere her bir komşunun ağırlığının $1/d$ ya da $1/d^2$ olarak alınmasıdır.

3.2 Karar Ağaçları

Karar ağaçları, bir ağaç yapısında olan, sınıflama ve tahmin etme amacıyla kullanılan güçlü bir modeldir. Modelin çalışma prensibinde, veri kümesi üzerinde herhangi bir kayıp yaratmadan, ağacın kökünden yaprak düğüme ulaşınca kadar ilerleyerek bir gözlemi sınıflandırmak için kullanılırlar. Bu düğümler, iç içe geçmiş eğer/diğer

(if/else) kurallar dizisinden oluşmaktadır. Bu kurallar entropi, bilgi kazancı, gini indeksi veya ki-kare testi gibi istatistiksel yöntemler ile düğümlere ve eşik değerlerine karar vererek veri kümesinin sınıflandırılması algoritmalarıdır.

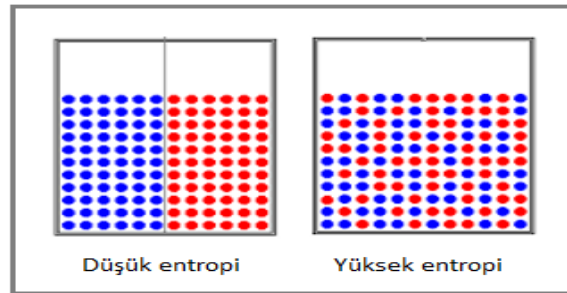
Ancak, en doğru sonuçları alabilmek için bir ağacın doğru boyutlandırılması gerekmektedir. Aşırı boyutlandırılmış ağaç, eğitim verisi üzerindeki doğruluk değeri çok yüksek olmasına rağmen, test verisi üzerinde doğruluk oranının azalmasına neden olur. Buna aşırı ezberleme (over-fitting) denir. Doğru dengeleme yaparak, ağacı budayarak, eğitim verisi için modelin doğruluk oranlarının düşürülmesi, test verisi üzerindeki doğruluğu artıracaktır (Ogryczak, Ruszczyński, 1999).

Karar ağacı yapısı oluşturulurken temel prensip; verilerle ilgili bir takım sorular sorulması ve elde edilen cevaplara göre, ağacın dalları üzerinde hareket edilerek en kısa sürede sonuca ulaşılmasıdır. Karar ağacı yapısında, sorulan sorular düğüm olarak, cevaplar dal olarak, kararın verildiği sınıf ise yaprak olarak gösterilmektedir.

Karar ağacı oluşturulurken en önemli adım, dallanma için karar verilecek en ayırt edici özneliliğin belirlenmesidir. Literatürde bu durum için geliştirilmiş çeşitli yaklaşımlar bulunmaktadır. Bu yaklaşımların en önemlileri; bilgi kazancı, Gini indeksi, entropi ve ki-kare olasılık tablo istatistiği yaklaşımlarıdır. Bu çalışmada kullanılan karar ağacı metodu olarak CART algoritması kullanılmıştır. Kırılım noktalarının sırası, gini indeksi ile belirlenmiştir.

3.2.1 Entropi

Entropi, olayların olma olasılıklarıyla ilişkili olup belirsizliğin ölçülmesi için kullanılan bir ölçüttür. Olayın rastgeleliğinin indikatörüdür. Entropi [0,1] arası değerler alır ve 1 değerine yaklaştıkça belirsizliğin arttığını gösterir. Şekil 3.2’de de görüldüğü üzere yüksek entropiye sahip modellerin sınıflandırmalarında karmaşıklık yaşanacaktır.



Şekil 3.2 : Düşük ve yüksek entropi ile hedef değişken tahmin şekli.

Entropinin hesaplanması denklem 3.4'te gösterilmiştir. P_i ; D veri kümesindeki i sınıfının olasılığı olarak varsayılırsa, bu değer, i sınıfına denk düşen veri sayısının tüm veri kümesindeki toplam örnek sayısına bölünmesi ile elde edilmiş olduğunu varsayalım. Buradan aşağıdaki denklem ile entropiler hesaplanır.

$$E_D = - \sum_{k=1}^m p_i \log_2(p_i) \quad (3.4)$$

Karar ağaçları kurulurken amaç, veri kümesinin entropisini her kırımla düşürerek, yaprak düğüme ulaşıldığında entropinin sıfır olmasını sağlamaktır. Bir başka deyişle, ağacın entropiyi azaltacak şekilde bölünmesini amaçlanmaktadır. Entropi yüksek ise, bu durum farklı sınıfların karışımına sahip olduğu anlamına gelir.

3.2.2 Bilgi kazancı

Bilgi kazancı, karar ağacı algoritmalarında eğitim veri kümesindeki özniteliğin etkinliğinin ölçüm değeri olarak kullanılır. Bilgi kazancı hesaplanırken, öncelikle veri kümesi alt bölümlere ayrılmadan önce entropisi bulunur, daha sonra her alt bölümün ayrı ayrı entropisi hesaplanır. Bu iki değer arasındaki farkın en yüksek olduğu öznitelik, en iyi dallara ayırmayı sağlayacak öznitelik olarak seçilir ve bölünmeye ondan başlanarak ağaç oluşturulmaya başlanır. X özniteliği üzerinden n tane bölünme sonucunda elde edilecek olan bilgi kazancı aşağıdaki şekilde bulunur.

$$Bilgi\ Kazancı(D, X) = E(D) - \sum_{i=1}^n p(D_i)E(D_i) \quad (3.5)$$

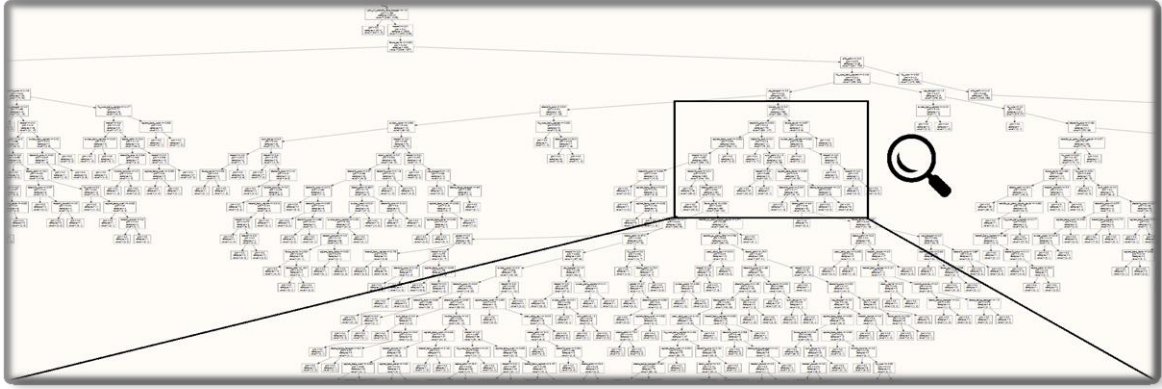
Burada, $E(D)$, veri kümesinin bölünmeden önceki entropisi; $E(D_i)$, i alt bölümünün X özniteliği üzerinden bölündükten sonraki entropisi ve $p(D_i)$, i alt bölünmesinin X özniteliğine göre bölünme sonrası olasılığıdır.

3.2.3 Gini indeksi

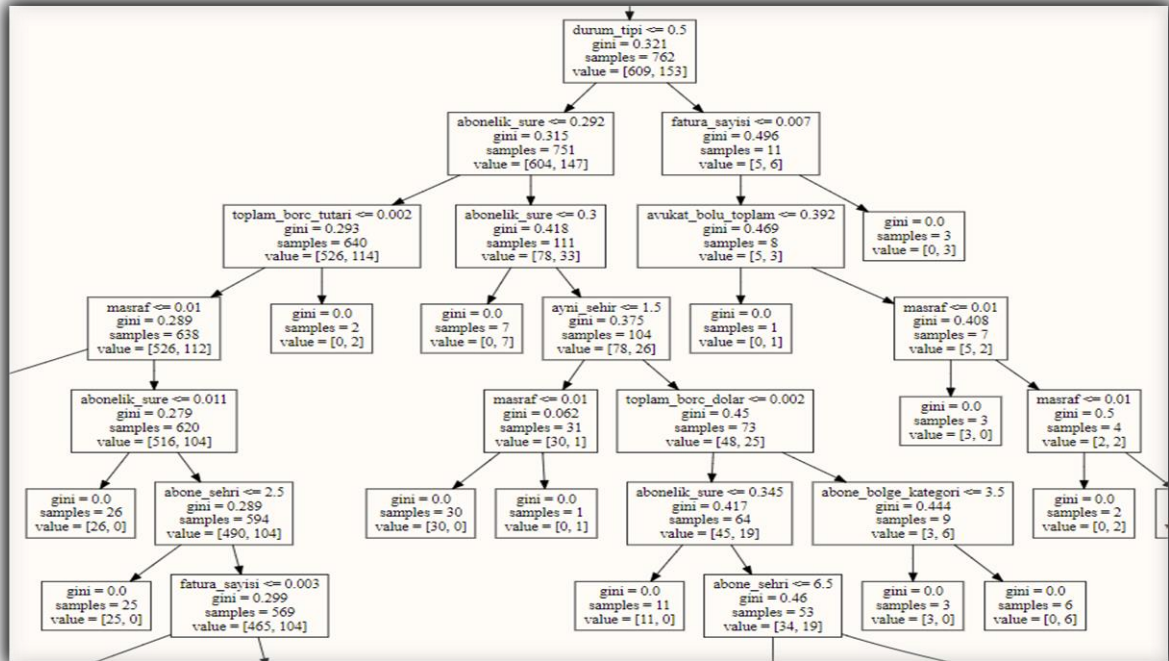
Karar ağacı algoritmalarında, ağacın oluşumu gini indeksine göre karar verilerek de dallandırılabilir. Eğitim veri kümesindeki özniteliğin etkinliğini ölçmek için kullanılır. Gini indeksi, 0 ile 1 arasında değerler almaktadır. Bu değer 1'e ne kadar yakınsa, Şekil 3.2'de gösterildiği gibi, belirsizlik o kadar fazladır. Gini indeksinin hesabı denklem 3.5'te verilmiştir (Zhau, Sun, Sang, Yang, Jia, 2012).

$$G(N) = \sum_{i=1}^n p_i(1 - p_i) \quad (3.5)$$

Bu tez çalışmasında kullanılmış olan karar ağacının görselleştirilmiş şeklinin ufak bir kesiti Şekil 3.3'teki gibidir. Bu kesiti daha okunabilir yapmak adına daha da küçük bir kesit alanını Şekil 3.4'teki gibi inceleyecek olursak ağacın kendi kurallarını, gini indeksini kullanarak oluşturduğu karar düğümleri ile yaprak düğüme kadar ilerlediği görülmektedir.



Şekil 3.3 : Ağaç görselleştirimi I.



Şekil 3.4 : Ağaç görselleştirimi I kesitinin büyütülmüş hali.

3.3 Naive Bayes

Naive Bayes sınıflandırıcısı, Bayes teoremi uygulayarak koşullu olasılıklar arasında bağıntı gerçekleştirir.

$P(A|B)$ 'nin anlamı, A'nın B'ye koşullu olasılığı olarak ifade edilmektedir. Yani B olayının bilindiği durumda A olayının meydana gelme olasılığıdır.

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\text{A'nın B içinde olma olasılığı}}{\text{B'nin olma olasılığı}} \quad (3.6)$$

Denklem 3.6'da $P(A \cap B)$ 'nin olasılığı yerine denklem 3.7 yazılırsa;

$$P(A \cap B) = P(A|B) \cdot P(B) \quad (3.7)$$

Denklem 3.6'da verilen koşullu olasılık ifadesini kullanarak B'nin A'ya koşullu olasılığı yazılacak olursak;

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{\text{B'nin A içinde olma olasılığı}}{\text{A'nin olma olasılığı}} \quad (3.8)$$

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)} \quad (3.9)$$

$B_1 \cap B_2 = \emptyset$ ve $B_1 \cup B_2 = B$ olacak şekilde B_1 ve B_2 kümelerini ele alalım. A ise B içerisinde bir olay olsun. Bu durumda A'nın B_1 ya da B_2 içerisinde olma durumunu bulmak için öncelikle A olayının B_1 ve B_2 cinsinden eşiti denklem 3.10'daki gibi bulunur.

$$A = (B_1 \cap A) \cup (B_2 \cap A) \quad (3.10)$$

Bu durumda A olayının olasılığı denklem 3.11'de verilmiştir.

$$P(A) = (B_1 \cap A) + (B_2 \cap A) \quad (3.11)$$

$$P(B_1|A) = \frac{P(A|B_1) \cdot P(B_1)}{P(A)} = \frac{P(A|B_1) \cdot P(B_1)}{(B_1 \cap A) + (B_2 \cap A)} \quad (3.12)$$

Formül genelleştirildiğinde; $B_1 \cup B_2 \cup B_3 \cup B_4 \cup \dots \cup B_n = B$ ve $\forall i,j \leq n$ için $B_1 \cap B_2 = \emptyset$ olmak üzere A, B kümesi içinde bir olay ise A olayının B_j sınıfında olma olasılığı aşağıdaki denklemden şöyle bulunur.

$$P(B|A) = \frac{P(A \cap B).P(B)}{P(A)} \quad (3.13)$$

$$P(A) = \sum_{i=1}^n P(A \cap B_i) \quad (3.14)$$

Denklem 3.14, A'nın B_i kümeleri kullanılarak olasılığını vermektedir. Denklem 3.14, denklem 3.13'te yerine konulursa Bayes kuralı elde edilmiş olacaktır.

$$P(B|A) = \frac{P(A \cap B).P(B)}{\sum_{i=1}^n P(A \cap B_i)} \quad (3.15)$$

Özet olarak bu sınıflayıcı özniteliklerin istatistiksel bağımsızlığına ve Bayes teoremine dayanır. Naive Bayes yönteminde, öznitelikler bağımsız kabul edildiğinden dolayı bu sınıflandırıcı büyük verilere kolayca uygulanarak başarılı sonuçlar elde edebilmektedir. Her ne kadar gerçek hayatta, özniteliklerin istatistiksel bağımsızlığı söz konusu değilse de, bu sınıflayıcı yönteminin çok iyi tahminler ürettiği bilinmektedir (Martinez, Sucar, 2006).

3.4 Lojistik Regresyon

Lojistik regresyon modeli, bir tür lineer bir modeldir. Lineer modeller, negatif sonsuzdan pozitif sonsuza kadar değişen doğrusal fonksiyon üretirler.

Lojistik regresyon temelinde bağımlı değişken ve bağımsız değişkenler arasındaki ilişkiyi ölçmek ve bu ölçümü kullanarak ilgili tahminleri yapabilmek yer alır. Tahmin modellerinde bağımlı değişken hedef değişkeni, bağımsız değişkenler ise öznitelikleri temsil etmektedir.

Lojistik regresyon yönteminde verideki gözlemlerin hangi sınıfa ait olduğu bilinmekte ve bu bilgiden yararlanarak yeni gözlemlerin hangi sınıfa dahil olacağı tahmin edilmektedir. En çok olabilirlik kestirimi yöntemiyle katsayılar hesaplanarak kurulan bu model ile iteratif olarak sınıflama yapılır.

Lineer ilişki tespit edilmesi en kolay ilişkilerden biridir. Bir öznitelik değeri artıyorken, diğerinin de artması, biri azalıyorken diğerinin de azalması lineer ilişki olduğunu gösterir. Ancak gerçek hayatta karşılaşılan durumlarda böyle bir duruma rastlamak oldukça nadirdir.

Doğrusal regresyon analizinde amaç, bağımlı değişkenin, yani tahmin etmeye çalıştığımız hedef değişkenin, bağımsız değişkenler karşısında alacağı ortalama değeri bulmaktır. Lojistik regresyon modelinde ise temel amaç, hedef değişken Y'nin π olasılıkla 1 ve $1-\pi$ olasılıkla 0 değerini alabileceğini göstermektir.

Ancak, bu modelin uygulanmasıyla eşik değer olarak 0.5 verildiği takdirde, 0.4 ihtimali olan bir müşteri için borcunu ödemez olarak işaretleneceği ayrıca 0.6 ihtimali olan müşterinin borcunu öder olarak işaretleneceği de değerlendirme sonucunda yorumlanmalıdır (Le, 1998).

Lojistik regresyon doğrusal bir regresyona benzer, ancak eğri olasılıktan ziyade hedef değişkenin “olasılıklarının” doğal logaritması kullanılarak oluşturulur. Ayrıca, tahmin edicilerin her bir grupta normal olarak dağılması veya eşit farklılık göstermesi gerekmediğinden müşteri puanlaması çalışmalarında sıklıkla kullanıldığı literatür incelemesinde görülmüştür (Kleinbaum, 1994).

Lojistik Regresyon modeli ifadesindeki regresyon kısmı, özniteliklerin değerlerinin lineer kombinasyonunun kullanılması ve regresyon katsayılarının varlığından kaynaklanır.

Lineer model formülü:

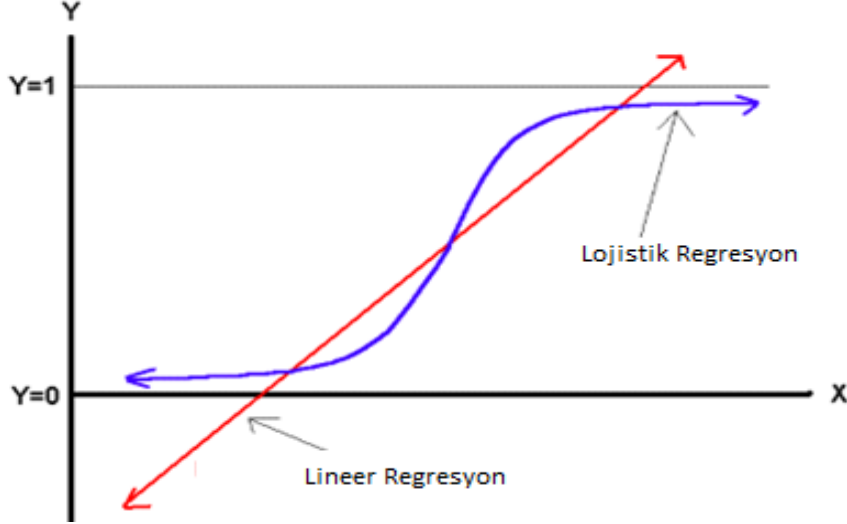
$$Y = W_1X_1 + W_2X_2 + W_3X_3 + \dots + W_nX_n + B \quad (3.16)$$

(Anderson, 2007).

Formülde yer alan özniteliğin önemi, öznitelik ve sapma miktarı kullanılarak hedef değerin eğitim verisi ile hesaplanmasının formülüdür.

Bu lineer denklemin lojistik dönüşümünde; lojistik model, dönüşüm fonksiyonlarını kullanarak [0,1] değerleri arasında değişen sonuçlar üretir. Lineer modelin alabileceği değerler ve bu değerlerin [0,1] aralığına denk düşecek şekilde düzeltme yapılması sonucu grafikleri Şekil 3.5'te görselleştirilmiştir. [0,1] arasında değiştiğinden

sınıflandırma problemleri için uygundur. Bu çalışma ile incelenen müşterinin borcunu geri ödeyip ödemeyeceği tahmini olan hedef değişkeni, 0 veya 1 değerini alan ikili bir değişken olduğundan lojistik regresyon kullanılmıştır.



Şekil 3.5 : Lineer ve lojistik model karşılaştırması.

Lineer modelde olayın gerçekleşme olasılığının $[0,1]$ arasında bir değer aldığını bulabilmek için Sigmoid dönüşüm formülü, lineer model formülü denklem 3.16 uygulandığında;

$$Sigmoid(z) = \frac{1}{1 + e^{-z}} \quad (3.17)$$

X özniteliğinin olasılığı denklem 3.18'deki gibi çıkmaktadır:

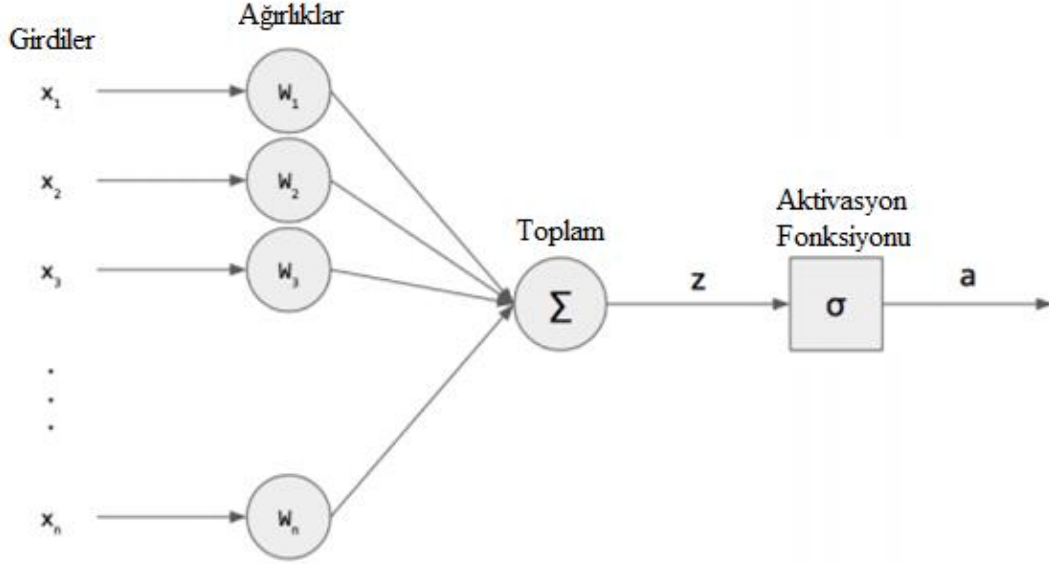
$$p(x) = \frac{1}{1 + e^{-(W_1X_1 + W_2X_2 + W_3X_3 + \dots + W_nX_n + B)}} \quad (3.18)$$

3.5 Yapay Sinir Ağları

Yapay sinir ağları, insan beynindeki sinir ağlarının bilginin işlemesini ve iletmesini modelleme denemelerinden geliştirilmiştir. Dolayısıyla yapay sinir ağı; insan beyninin sinir hücrelerinden oluşmuş katmanlı yapısının tüm fonksiyonlarından esinlenerek elektronik ortamda geliştirilmiş modellenmesidir.

Biyolojik sinir hücresinde, diğer sinir hücrelerinden gelen sinyalleri toplar. Eğer toplam sinyal birikimi belirli bir eşik değerini aştıysa, sinyalin bir başka sinir hücresine

iletimini gerçekleştirir. YSA'da da, biyolojik sinir hücrelerinde olduğu gibi, Şekil 3.6'da da görüldüğü üzere özniteliklerden gelen değerler belirli bir ağırlık oranları ile çarpılarak toplanır. Bu toplam değer, aktivasyon fonksiyonuna sokularak bir sınıflandırma yapılır.



Şekil 3.6 : Sinir ağı modelinin çalışma prensibi.

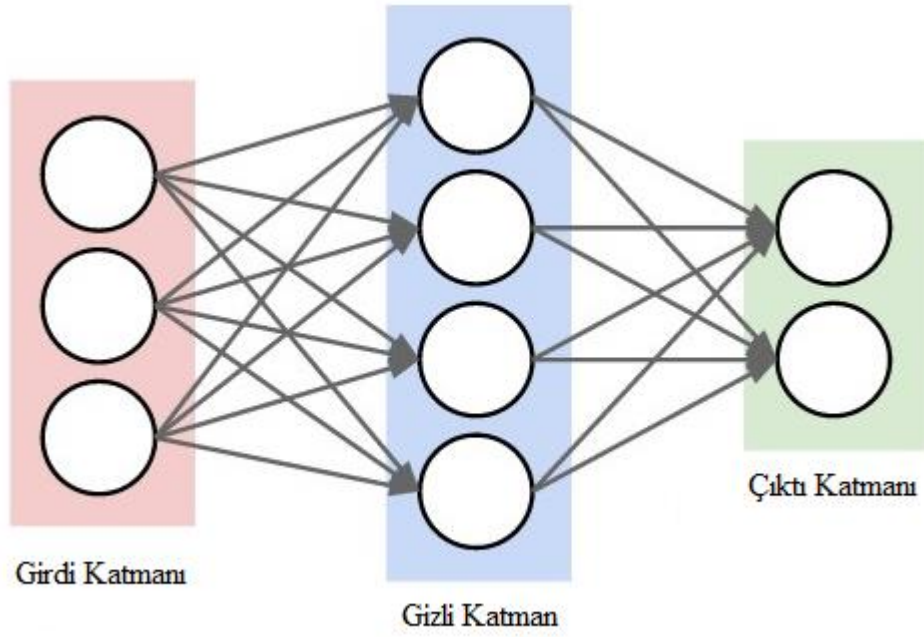
Sinir hekimi Warren McCulloch ile matematikçi Walter Pitts tarafından 1943 yılında ilk yapay sinir ağı modeli gerçekleştirilmiştir. McCulloch ve Pitts, ilk yapay sinir ağı modelini geliştirirken insan beyninin hesaplama yeteneğinden esinlenerek elektrik devreleriyle basit bir sinir ağı modellemesi gerçekleştirmişlerdir. 1949 yılında ise Hebb “Organization of Behavior” adlı eserinde öğrenme ile ilgili temel teoriyi ele almıştır. Frank Rosenblatt’ın 1957 yılında Perceptron’u geliştirmesi YSA ile ilgili çalışmaların hız kazanmasını sağlamıştır.

Rosenblatt, Cornell üniversitesinde insan beyninin çalışma prensibiyle ilgili çalışmalar yaparken perceptron’u bulmuştur. Perceptron, lineer bir sınıflandırıcıdır. Yukarıda bahsedildiği gibi x özneliliğinin bir ağırlık katsayısı ile çarpılıp, bias değerinin eklenmesi sonucunda elde edilen lineer fonksiyon, lineer olmayan aktivasyon metodu ile işleme sokulup ikili sınıflandırma sonucuna ulaşır. Genel gösterim, denklem 3.19’da verilmiştir. Bu nedenle, lojistik regresyon, YSA’nın atasıdır diyebiliriz.

$$Y = wX + B \quad (3.19)$$

İlk olarak, tek katmanlı perceptronlar, lineer olan 'VE', 'VEYA' gibi mantıksal problemleri çözebilmesine rağmen, XOR gibi lineer olmayan problemleri çözmekte yetersiz sonuçlar üretiyordu (Rumelhart, McClelland, 1986).

Ancak çalışmalar katmanların çoklanmasıyla devam edilince, çok katmanlı perceptronların XOR ve XOR'un dışında pek çok lineer olmayan problemi sonucu yüksek doğruluk oranı ile tahmin ettiği görülmüştür.



Şekil 3.7 : Sinir ağı katmanları şekli.

Şekil 3.7’de gösterildiği gibi YSA’lar en az 3 katman içerir. Bunlar; girdi katmanı, çıktı katmanı ve gizli katman. Bu katmanlardaki her bir düğümü insan beynindeki nöron olarak düşünebiliriz. Bu katmanlardaki düğümler, kendisinden bir önceki katmandaki düğümlere tamamen bağlantılı bir şekildedir. Yani, bir katmandaki düğümler, kendisinden bir önceki katmandaki tüm düğümlerle ve kendisinden bir sonraki katmandaki tüm düğümlerle iletişim halindedir.

Bu çalışma için kullanılan yapay sinir ağı, çok katmanlıdır. Yani girdiler ile hedef arasında birden fazla gizli katman bulunmaktadır. Girdi değişkenleri ile hedef değişken arasında ilave gizli katmanlar yaratılır. Gizli katmanın sayısı, gizli katmandaki düğüm sayısı modelin öğrenmesi ve performansı üzerinde etkilidir. Ancak bu parametrelerin aşırı artırılması modelin, eğitim verisini kopyalamasına (over-

fitting) ve eğitim süresinin uzamasına neden olabilir. Eğitim verisinin ve test verisinin doğru oluşturulması YSA sonuçları açısından çok önemlidir (Rosenblatt, 1986).

Şekil 3.7’de gösterilen görselde gizli katmandaki tek bir sinir hücresinin yaptığı işlemi formüle çevirecek olursak, denklem 3.20’yi elde ederiz.

$$\sum_{i=1}^n w_i x_i + b \quad (3.20)$$

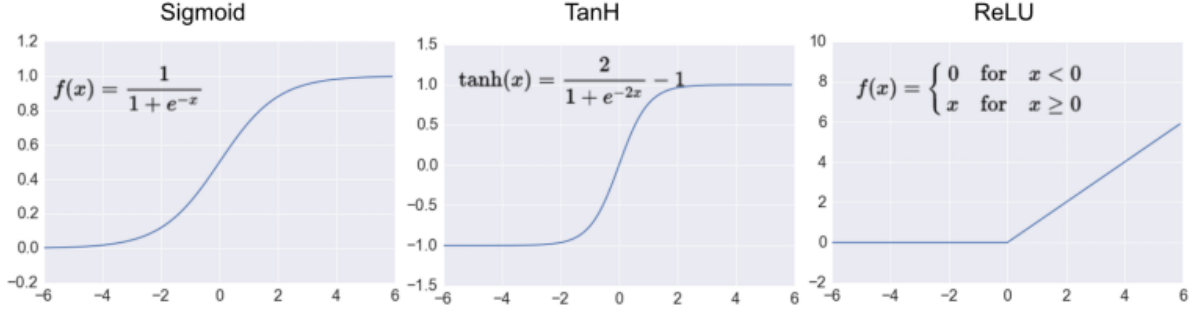
Çok katmanlı algılayıcı (Multi-layer perceptron) modelinin çalışması; her bir gizli katman nöronu, bir önceki katmandaki nöronun çıktılarının ağırlıklandırılmış kombinasyonuna dayalı bir girdi alır. İterasyon ile ilk gizli katman sonucunun doğruluğuna göre bir sonraki katmanda ağırlıklar yeniden belirlenir. Ağırlık değişimi, ancak doğruluk değerinde iyileşme gerçekleşmediğinde durur. Ağırlıklar ilk olarak rastgele atanır. Bu çalışma kapsamında performans için iterasyon sayısı 200 ile sınırlandırılmıştır. Model 200 kez iterasyon yapmasına rağmen yakınsamasını sürdürmeye devam ediyor olsa bile, performans açısından bu işlem yarıda kesilmiştir. Ayrıca, düğümleri ağırlıklandırmak için kullanılan yöntemler olarak quasi-Newton ailesinin iyileştirici bir metodu olan ‘lbfgs’, olasılıksal dereceli azalma olarak ‘sgd’ ve ya Kingma, Diederik ve Jimmy tarafından önerilen olasılıksal derece temelli iyileştirici metod olarak ‘adam’ kullanılmaktadır. Bu çalışma kapsamında ağırlıklandırmak için ‘adam’ yöntemi tercih edilmiştir (Kröse, Smagt, 1986).

Yapay sinir ağlarında yaygın olarak kullanılan bazı aktivasyon fonksiyonları vardır. Her bir nöronun kendi aktivasyon fonksiyonu vardır. Aktivasyon fonksiyonu sayesinde 0-1 çıkıyor. Sigmoid, tanjant hiperbolik ve ReLu aktivasyon fonksiyonları yaygın olarak kullanılmaktadır. Formülleri denklem 3.21-23’te verilmiştir. Aktivasyon formüllerinin grafikleri Şekil 3.8’de gösterilmiştir. Şekil 3.8’de de görüldüğü üzere, Sigmoid fonksiyonu 0 ile 1 arasında; tanjant hiperbolik -1 ile 1 arasında değerler üretmektedir. Aktivasyon fonksiyonu (0,1) veya (-1,1) arasında sonuç üretilmesini sağlar. Bu çalışmada kullanılan aktivasyon fonksiyonu *ReLu* olmuştur.

$$Sigmoid(z) = \frac{1}{1 + e^{-z}} \quad (3.21)$$

$$\text{TanH}(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (3.22)$$

$$\text{Relu } f(z) = \max(0, z) \quad (3.23)$$



Şekil 3.8 : Aktivasyon fonksiyonlarının grafikleri.

(Minsky, Papert, 1996).

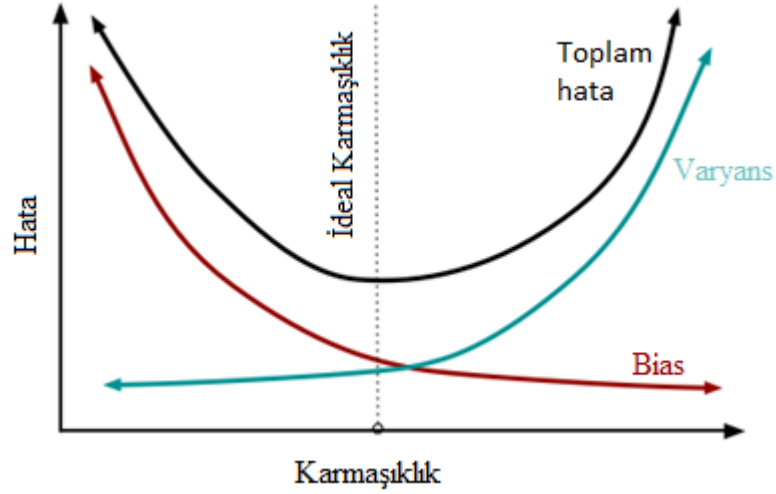
3.6 Topluluk Yöntemleri (Ensemble)

Topluluk yöntemleri, model iyileştirmesi yapmak için, birden fazla yöntemin bir arada kullanılmasıdır. Yani yöntem, birden çok model (kullanılan bu modellere zayıf öğrenici denir) uygulayarak sonrasında bunları birleştirerek yeni ve daha güçlü birleşik tek bir model elde etmeye çalışır. Tahmini bu yeni modele göre yapar ve sonuçları iterasyonlarla iyileştirmeye çalışır.

İterasyonlarda, farklı eğitici alt veri kümeleri üzerinde kullanılan modeller aynı olabileceği gibi, farklı modeller çalıştırarak bunların sonuçlarından bir iyileştirme metodolojisi çıkartılabilir (Zhou, 2012).

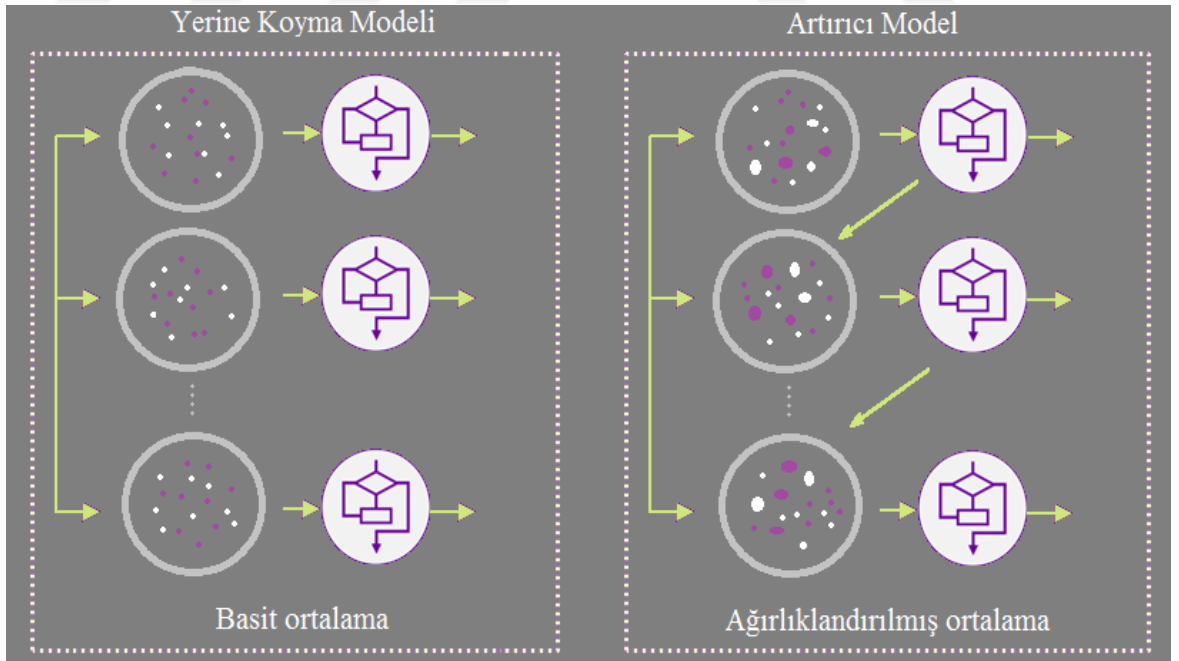
Yaygın olarak kullanılan yöntem, her bir öğrenen, aynı modeli kullanarak tahmin yapar. Buna homojen topluluk yöntemleri denilmektedir.

Sınıflandırma algoritmalarının hatasını belirleyen iki bileşen vardır. Topluluk yöntemleri bu bileşenlere odaklanarak sonuçları optimize etmeye çalışırlar. Bu bileşenler; sınıflandırıcının doğruluğu (bias) ve sınıflandırıcının farklı eğitim kümeleri ile eğitilmesine gösterdiği sapmadır (varyans). Bu iki bileşen arasında ters orantı vardır. Şekil 3.9'da da görüldüğü gibi düşük hatalı sınıflandırıcılar yüksek sapma eğilimi gösterir ve bunun tersi de geçerlidir (Zhang ve Ma, 2012).



Şekil 3.9 : Topluluk yöntemleri bileşenleri.

Diğer taraftan, ortalama almanın, sapmayı azaltan bir etkisi de söz konusudur. Bundan dolayı ensemble yöntemlerinin hedefi nispeten belirlenmiş ya da benzer hatalı birkaç sınıflandırıcı oluşturmak, bunların verilerini bir araya getirmek, ortalamasını tespit etmek ve sapmayı azaltmaktır (Zhang ve Ma, 2012). Ensemble metotların bunu nasıl yaptığını dair görsel Şekil 3.10'de verilmiştir. Herbir metodun detaylı anlatımına kendi başlıklarında yer verilecektir.



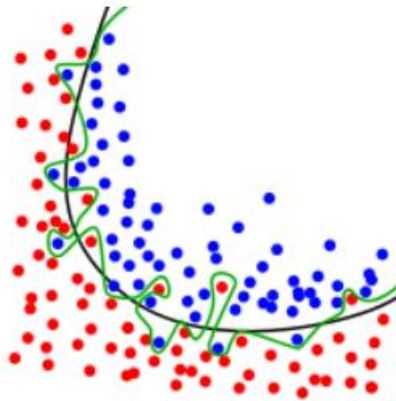
Şekil 3.10 : Yerine koyma (bagging) ve artırıcı (boosting) topluluk yöntemleri.

3.6.1 Yerine koyarak artırma (Bagging – Bootstrap aggregating)

Bagging yöntemi, 1996 yılında, Breiman tarafından ortaya konan modeldir. Tahminlerin doğruluğunu artırmak için, aynı sınıflandırıcı yönteminin, eğitim kümesindeki farklı alt kümeler için iterasyonlarla çalıştırılması temeline dayanır. Temelinde bootstrap mantığıyla çalışan yöntem, bootstrap ile kümelenmiş veri kümelerinden aldığı rastgele alt küme üzerinde ağaç algoritmalarıyla budama ve ağırlıklandırma işlemleri uygulayarak sonuçları optimize etmeye çalışır. Bootstrap ile kümeleme işlemi sırasında bir müşteri birden fazla veri alt kümesinde bulunarak, modelin eğitilmesinde daha fazla rol oynayabileceği gibi, bazı müşteriler de modelin öğrenmesinde hiç rol almayabilir. Doğruluğu artırmak için kullandığı yöntemin amacı varyansı azaltmaktır (Breiman, 1996).

Bagging yöntemler, varyansı azaltmaya yardımcı olur. Paralel olarak üretilen çeşitli karar ağaçları sonuçları, bagging tekniğinin temel öğrenenlerini oluşturur. Elde edilen bu eğitim verilerinin modellenmesi sonucundaki tahmin, tüm modellerin ortalama çıktısıdır. Bu sayede modelin varyansı düşmüş olur. Bu da aşırı öğrenme (over-fitting) durumunu azaltmış olur.

Aşırı Öğrenme: Oluşturulan tahminsel model için aşırı öğrenme (over-fitting) söz konusu olabilir. Bu durumda oluşturulan model, eğitim veri kümesi için %100'e yakın doğru sınıflama sağlarken, yeni gelen veriler için doğruluk oranı çok düşük olabilir. Çünkü eğitim verisini kopyalamış olabilir. Aşırı öğrenmiş bir model, Şekil 3.11'deki yeşil ile çizilmiş gibi bir eğriye sahip olur. Ancak istenen siyah ile çizilmiş grafiğe sahip olmasıdır.



Şekil 3.11 : Aşırı öğrenme model öğrenme şekli.

3.6.1.1 Rastgele orman ağaçları (Random forest trees)

Yerine koyarak artırma (bagging) modellerinden olan rastgele orman ağaçları yöntemi, temelinde karar ağaçlarına dayanır.

Rastgele orman yönteminin temel eğiticileri karar ağaçlarıdır. Temel eğiticiler yine bootstrap ile üretilmiş, eğitim örnekleriyle eğitilirler. Ağaçların oluşturulması aşamasında her ağaçta düğümde kullanılmak için, en iyi öznitelik seçimi yapılırken, tüm öznitelikler arasından değil de rastgele oluşturulan öznitelik alt kümeleri arasından karşılaştırmayla seçilir. Ayrıca bu öznitelik alt kümelerinin boyutu da önceden belirlenebilir. En son aşamada, elde edilen eğitici sonuçları birleştirilir. Birleştirme işlemi, sınıflandırma problemlerinde en çok oyu alan sınıfın seçilmesi; regresyon problemleri için ise ortalamalarının alınması yöntemiyle varyans düşürülmeye çalışılır. Sınıfı bilinmeyen yeni bir veri geldiğinde, her karar ağacı bu veriyi kendi ağacında değerlendirerek kendine göre sınıflandırmasını yapar ve sonuç bu ayrı tahminlerin sonucunda hangi sınıf daha fazla seçilmişse verinin o sınıftan olduğuna karar verilir (Fraiman ve diğ., 2012).

Öznitelik sayısındaki azalış sayesinde, rastgele karar ağacının üretim süresi kısalmış ve karar ağaçlarının farklılığı ile rastgelelik artırılmış olur. Ancak öznitelik sayısının azaltımı çalışma süresini kısaltmış olsa da, ağaç sayısı arttıkça daha iyi başarı elde edilmesi beklenmektedir. Ancak ağaç sayısı arttıkça algoritmanın çalışma süresi artmaktadır. Ek olarak rastgele orman yönteminde ağaç sayısı ne kadar artarsa artsın aşırı öğrenme (over-fitting) durumu ile karşılaşmamaktadır. Breiman (2001)'a göre, rastgele özniteliklerin seçilmesi sınıflandırma doğruluk oranını artırırken, verinin gürültülere karşı hassasiyetini azaltır.

3.6.1.2 Son derece rastgele ağaçlar (Extra trees (Extremely randomized trees))

Pierre Geurts, Damien Ernst ve Louis Wehenkel tarafından “Son derece rastgele ağaçlar”, 2006 tarihinde ortaya atılmıştır. Bu model, eğitim kümesi oluşturmak için bootstrap uygulamazlar. Rastgele ormandan farklı olarak her bir karar ağacı tüm veri kümesini içerir. Karar ağaçlarında düğümün ve kesme eşiğinin belirlenmesi için bilgi kazancı oranına bakmaz. Rastgele bölünmeyi seçer. Yani rastgele bir öznitelik ve rastgele bir eşik değeri seçerek ağacın dallarını oluşturur.

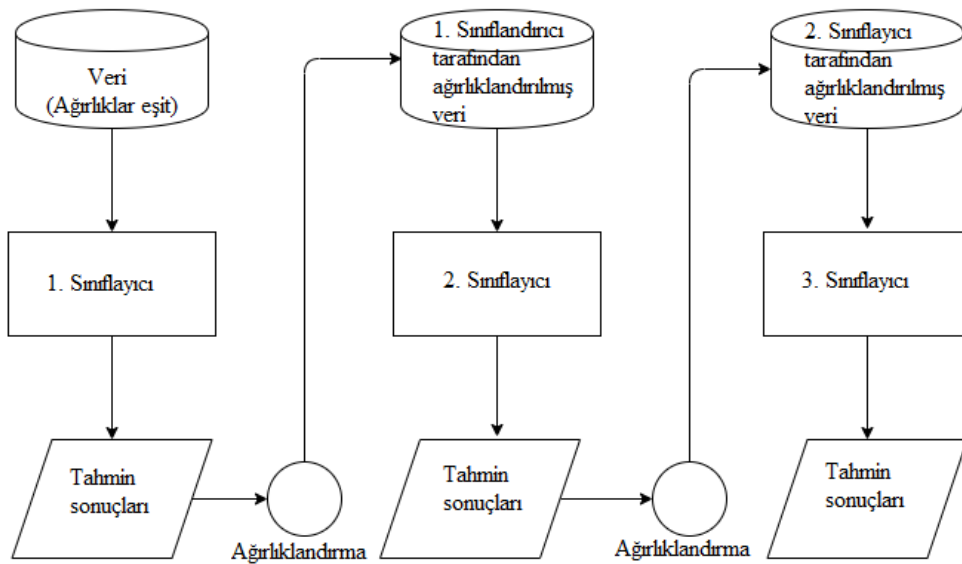
3.6.2 Artırıcı (Boosting)

Artırıcı (boosting) modeller eğitim kümesindeki zayıf örneklerin üzerine yoğunlaşarak doğruluğu artırmayı hedefler. Aşırı öğrenme (over-fitting) ve yetersiz kalma (under-fitting) durumlarının önüne geçilmesi için kullanılır. Hem yanlılık (bias) hem de varyansı düşürerek modelin doğruluğunu artırabilmek için, kullanılan zayıf modeller sıralı olarak kullanılır. Bir modelin çıktısı sonucunda hesaplanan hatalı tahminlerin ağırlığı artırılarak, bir sonraki modelin girdisi olarak kullanılır. Dolayısıyla ağaçların kurulması için, bir önceki modelin işlemini tamamlaması beklenir (Schapire, 1999).

Algoritmanın çalışma mantığı ve adımları şu şekildedir:

- Eğitim kümesinden rastgele bir alt küme alınır.
- Seçilen bu küme, verilere ilk başta eşit ağırlık atayarak modeli eğitir.
- Tüm eğitim kümesi ile test yapılır.
- Her bir müşteri için sınıflandırma modelinin hatası ölçülür. Yanlış sınıflandırma yapılmış müşteriler için, o müşterinin ağırlığı veya bir sonraki iterasyonda kullanılmak üzere veri kümesindeki örnek sayısı artırılır.
- Eğitim kümesinden yeni bir rastgele alt küme oluşturulup yapılmış adımlar tekrar edilir.

Yukarıdaki algoritmanın görselleştirilmiş hali Şekil 3.12’de verilmiştir.



Şekil 3.12 : Boosting algoritması görseli.

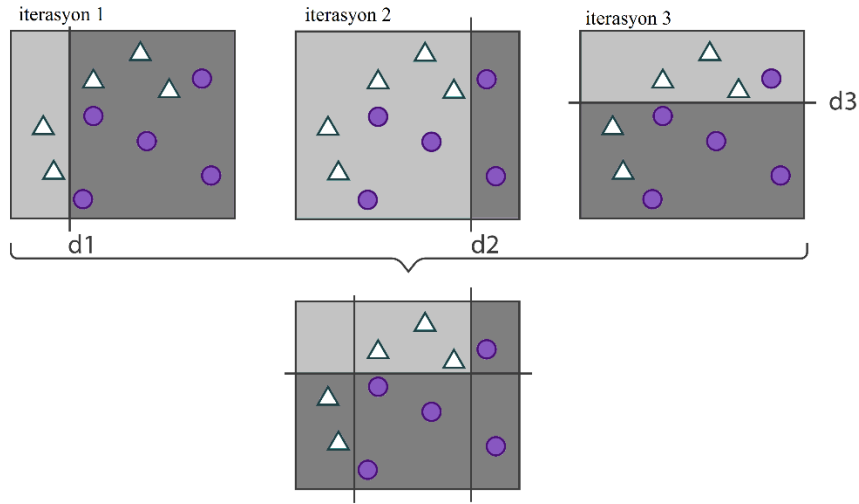
Gao, Wu ve diğ. (2009) yaptıkları çalışmada boosting yöntemleri ile modelin varyans ve yanlılığı düşürülmüş olur ancak bu durumda modelin aşırı öğrenme riski oluşabileceğine değinmişlerdir.

Ağaçların azami ölçüde büyüdüğü rastgele orman gibi torbalama tekniklerinin aksine, artırma daha az bölünmüş ağaçları kullanır. Çok derin olmayan bu küçük ağaçların çok olması, aşırı öğrenmeye neden olabilir. Bu nedenle, artırma için durma ölçütlerini dikkatli seçmek gerekir. İterasyon sayısının, ağacın derinliğinin ve ağaç sayısının doğru belirlenmesi gerekir. Ayrıca çapraz doğrulama yöntemi ile aşırı öğrenmenin önüne geçilebilir (Schapire, 1999).

3.6.2.1 Adaboosting (Adaptive boosting)

Adaboosting sınıflandırıcısı, adını ‘*adaptive*’ yani uyum sağlayan kelimesinin kısaltılmış halinden alır. Yöntem, 1995 yılında, Freund ve Schapire tarafından ortaya atılmıştır (Freund, Schapire, 1997). Model, sonucun başarısını artırabilmek için, her iterasyonda bir önceki iterasyon sonucunda yanlış tahmin yapılmış müşterilerin ağırlığı artırılarak, sınıflandırıcı metot tekrar çalıştırılır. Böylece, yanlış tahminler üzerine odaklanarak, modelin tahminlerinde doğruluk oranı yükselmektedir (Wang, 2012). Ayrıca model, Breiman’ın denemelerine göre, boosting sınıflandırma metotları içerisinde en iyisi olarak kabul edilmektedir. Yöntemin altında yatan temel mantık; rastgele veri kümeleri ile çalışmasından ziyade, aynı veri kümesini tekrarlı olarak çalıştırması ancak hatalı olanların ağırlıklarını değiştirerek kullanılmasıdır. Böylece, bu model, diğer boosting yöntemlerin aksine büyük veri kümesine ihtiyaç duymaz (Zhu, Zou, 2009).

Modelin çalışma prensibinin görselleştirilmiş hali Şekil 3.13’te verilmiştir. İlk olarak her verinin ağırlığı aynıdır, her aşamada yanlış sınıflandırılan verilerin ağırlıkları artırılır yada azaltılır. Böylece sınıflandırılması zor olan bu verilerin üzerine yoğunlaşma imkanı sağlanır (Hastie, Tibshirani, Friedman, 2009).

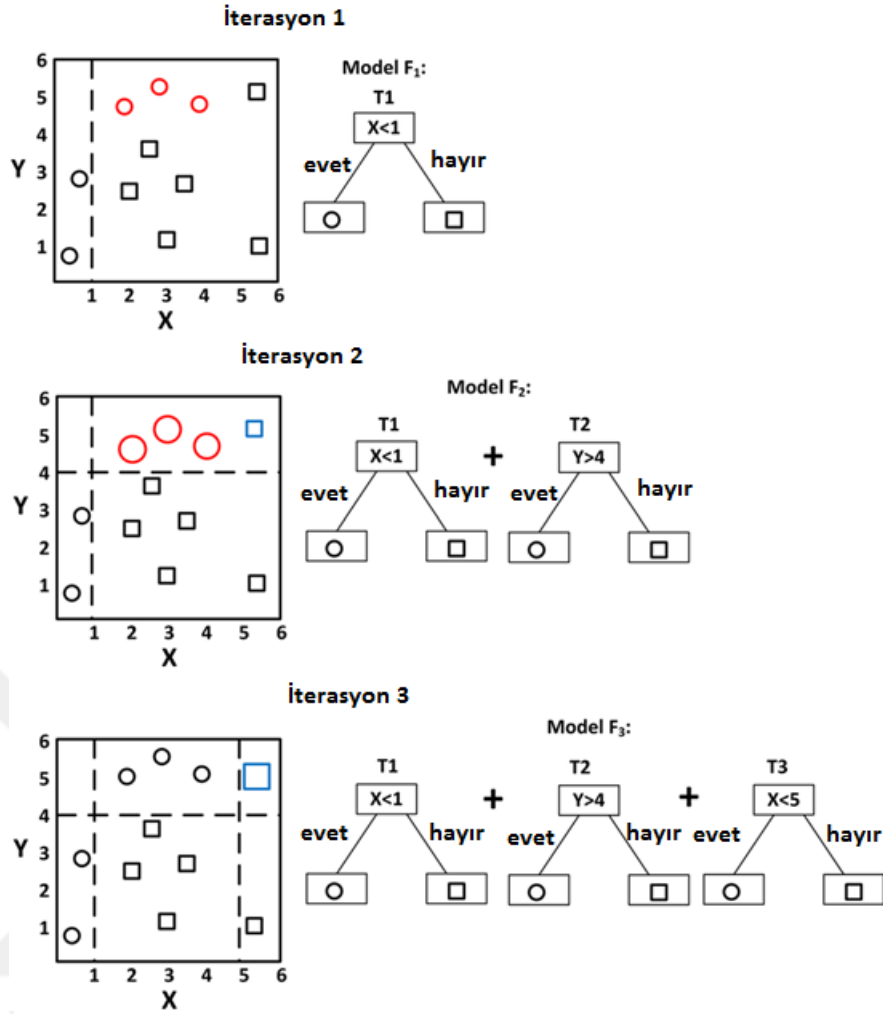


Şekil 3.13 : Adaboost modelinin çalışma prensibi.

3.6.2.2 Gradyen artırma (Gradient boosting)

Gradyen artırma yöntemi, Friedman tarafından 2001 yılında yayınlanmıştır. Yöntemin amacı diğer artırma yöntemlerinde olduğu gibi sınıflandırmanın başarısını arttırmaktır. Algoritmanın çalışma prensibinde; eğitici sınıftaki sınıflandırma yöntemleri için logaritmik kayıp, regresyon yöntemleri için ise kare kayıp fonksiyonları kullanılarak elde edilen kayıp sonuçları optimize edilmeye çalışılır. Yani, iteratif olarak hatayı tahminleyip, iyileştirmeye çalışır. Optimize etme yöntemi ise; kayıp fonksiyon değeri yüksek çıkan test müşterileri birer karar ağacıdır. Her iterasyonda yeni karar ağaçları eklenerek ilerlenir, önceki ağaçta değişiklik yapılmaz. Yeni dallar eklendikçe kayıp fonksiyonunun değeri iyileşir. Yani aslında, model önceki iterasyonda ürettiği değerleri saklayarak, bu değerlere yeni iterasyon sonucunda elde ettiği ağırlıkları da ekleyerek ilerler.

Yönteme, rastgeleliği eklemek hem metodun hızını hem de doğru tahmin yapma olasılığını artırır. Bu nedenle, alt küme, tüm eğitim veri kümesinden rastgele seçilir. Bu seçim, aynı zamanda aşırı uyum sorununa karşı da sağlamlığı artırır (Mayr ve ark., 2014). Gradyen boosting yöntemi, Adaboost yönteminde olduğu gibi örnekleri yeniden ağırlıklandırmaya çalışmaz. Bir önceki iterasyonda bulunan kayıp fonksiyonunu negatif gradyan vektörüne uydurmaya çalışır. Hatayı azaltmak için, dereceli azaltma (gradient descent) sonuçlarını fonksiyona ekler. Bu durum Şekil 3.14'te şematize edilmiştir.



Şekil 3.14 : Gradyen Boosting algoritması çalışma prensibi.

3.6.2.3 Fazlasıyla gradyen artırma (XGBoost (Extreme gradient boosting))

XGBoost modeli, boosting modellerinde anlatılmış olan gradyen boosting modelinin çalışma prensibiyle, ağaç modellerinin bir arada kullanılmış halidir. Bu sayede model, hem sınıflandırma hem de regresyon modelleri için yüksek doğrulukla çalışan bir modeldir. Çünkü aşırı öğrenme'yi kontrol eder. XGBoost için kullanılan ağaç modeli CART modelidir. Bu model, karar ağacı modeli anlatılırken yukarıda detaylandırılmıştır (Bölüm 3.2). Ancak burada değinmek gerekirse, karar ağaçlarında, kırılim noktalarına doğru bir şekilde karar vermek sonuç üzerinde büyük etkiye sahiptir. CART ve dolayısıyla XGBoost algoritması kayıp olmayan verileri kullanarak kırılim değerine karar verdiğinden dolayı doğruluğu daha yüksek sonuçlar elde edilmektedir. Bunun aksine, gradyen boosting ağacı oluştururken negatif kayıp kriterine baktığı için aşağı doğru fazla büyüyen ağaçlar yaratılabiliyor. Ancak, fazlasıyla gradyen boosting modelde, ağacın derinliği baştan belirlenebilir. Eğer ağaç

ařaęı doęru ok bydyse, bunu geri ynde budama yaparak modelin ařırı ęrenmesi nlenmelidir. Ayrıca model, kendi ierisinde, dıřarıdan parametre almaksızın apraz doęrulama yaparak modelin doęruluęunun maksimum olmasını saęlamaktadır.



4. MÜŞTERİNİN ÖDEME TAHMİNİ SIRASINDAKİ SÜREÇLER

Veri üzerinde uygulanacak modellere karar verdikten sonra, veriyi modelleri eğitmede kullanabilmek için yapılan işlemler aşağıda listelenmiştir. İlk olarak, şirketten alınan verilerin içeriğinden bahsedilecek ve bunlar görselleştirilmeye çalışılacaktır. Sonrasında verinin yorumlanmasıyla birlikte, veri üzerinde yapılacak işlemlere karar verilecektir. Veri temizleme, veri indirgeme, veri genişletme, kategorize hale getirme, sürekli verileri normalize etme, kategorize verileri kodlama işlemleri uygulanmıştır. Doğruluk oranını daha da artırmak için k-kat çaprazlama, veri kümesinin eğitim ve test olarak ayrıştırılması ve kümenin rastgele bir şekilde tekrar sıralanması işlemleri veri üzerinde çalıştırılmıştır. Sonrasında elde edilen modellenmeye hazır, temizlenmiş veri kümesindeki öznitelikler en önemliden en az etkiliye doğru sıralayabilmek için öznitelik önem metodlarının kullanılmasıyla elde edilen çizelgelerden bahsedilmiştir. Öznitelik seçme yöntemlerinde ilk olarak ağaç temelli ve ensemble temelli yöntemler ile elde edilmiş öznitelik önem listelerinden elde edilen sonuçların normalize edilip ortalamalarının alınmasıyla elde edilmiş değerler listelenmiştir. Sonrasında temel bileşen analizi yöntemi veri kümesine uygulanmıştır. Daha sonra özyinelemeli olarak çalışan modele lojistik regresyon ve adaboost algoritmaları verilmiş ve sonuçları aşağıda sırasıyla listelenmiştir. Son olarak da ki-kare testine göre özniteliklerin önem dereceleri listelenmiştir.

4.1 Uygulama Verisinin Toplanması

Telekomünikasyon şirketinden alınan öznitelikler Çizelge 4.1’de verilmiştir.

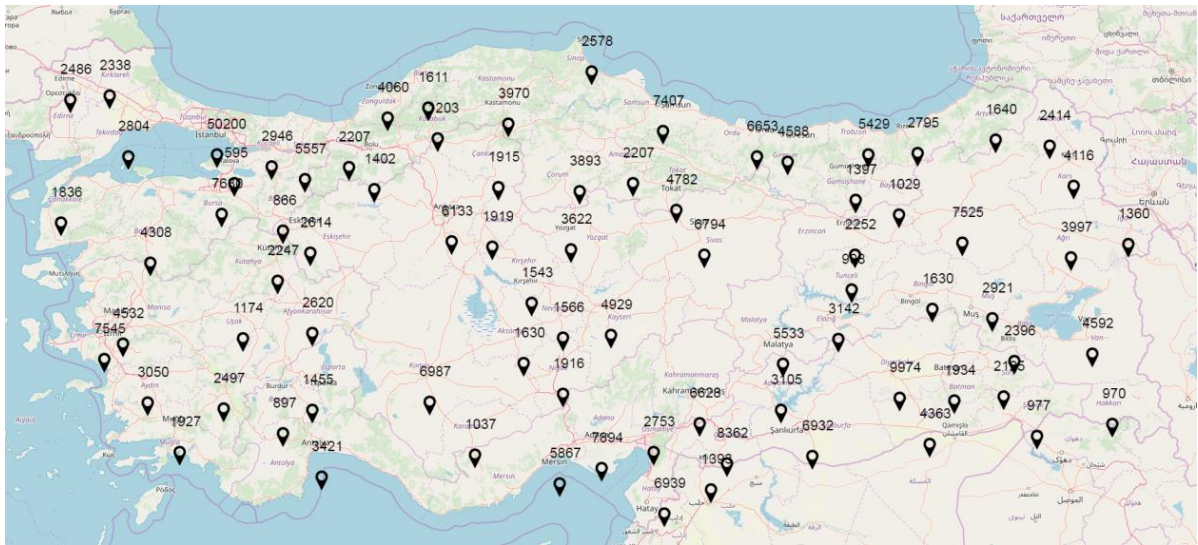
Çizelge 4.1 : Telekomünikasyon şirketinden alınan öznitelik listesi.

Öznitelikler	Boş değer	Sınıf	Öznitelik açıklaması
abone_sehri	Var	kategorik	müşterinin yaşadığı şehir
abonelik_sure	Var	normalize	müşterinin telekomünikasyon şirketine müşterilik süresi
avukat_ucreti	Yok	normalize	müşterinin avukatlık bürosuna olan borcu
cinsiyet	Yok	kategorik	müşterinin cinsiyeti

durum_info	Yok	kategorik	müşterinin borcu ödeyip ödemediği bilgisi
durum_tipi	Yok	kategorik	kurumsal müşteri/bireysel müşteri tipi kategorize
fatura_sayısı	Var	normalize	müşterinin telekomünikasyon ödediği fatura sayısı
iletisim_kuruldu	Yok	kategorik	müşteri ile telekomünikasyon şirketinin iletişim kurduğu bilgisi
ilk_tutar	Yok	normalize	müşterinin telekomünikasyon şirketine olan borcu
masraf	Yok	normalize	takip başlatmanın bedeli
medeni_durum	Var	kategorik	müşterinin medeni durum bilgisi
ofis_sehir	Yok	kategorik	büronun bulunduğu şehrin plaka kodu
taksit_yapildi	Yok	kategorik	müşterinin borcunun taksitlendirilme bilgisi
transfer_tarihi	Yok	date	müşterinin avukatlık bürosuna atandığı tarih
yabanci_uyruk	Yok	kategorik	müşterinin TC vatandaşı olup olmadığı bilgisi
yasal_faiz	Yok	normalize	gecikmeden dolayı yasal olarak belirlenmiş faiz miktarı
yasal_takip_sayi	Yok	normalize	müşterinin hayatı boyunca kaç kez yasal takibe girdiği
doğum_tarihi	Var	date	müşterinin doğduğu tarih bilgisi

- **abone_sehri**

Müşterinin yaşadığı şehrin plaka bilgisidir. Çalışmada, kategorik olarak değerlendirilmiştir.



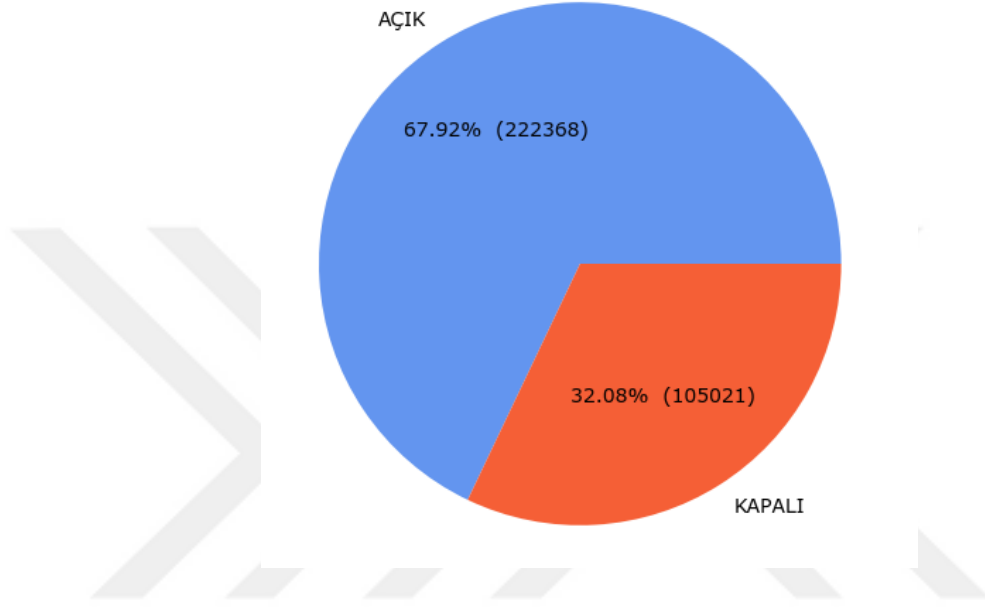
Şekil 4.1 : Müşterinin yaşadığı şehir görseli.

- **abonelik_sure**

Müşterinin telekomünikasyon şirketi ile olan ilişkisinin gün olarak tutulan bilgisidir.

- **durum_info**

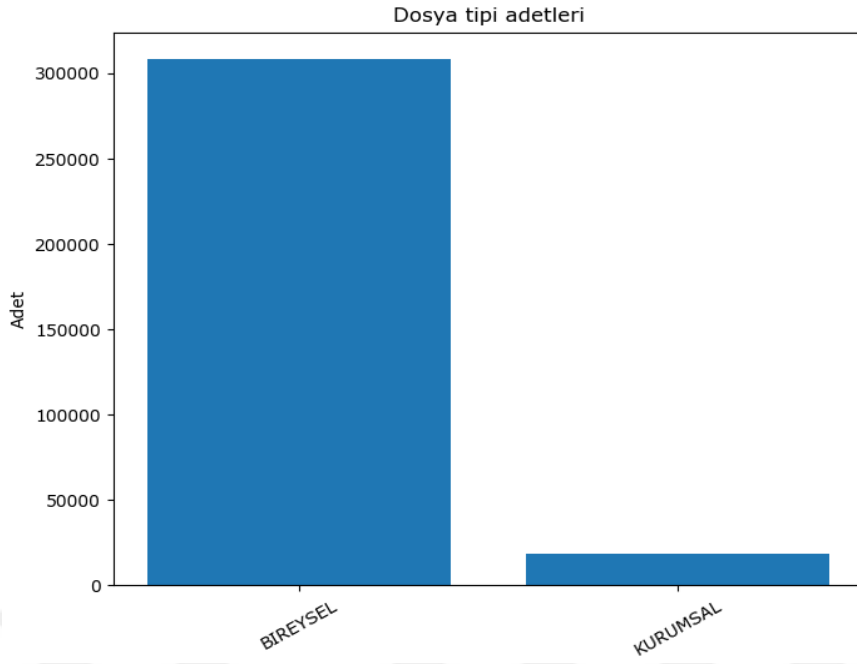
Hedef değişken bilgisidir. Müşterinin büroya devredildikten sonra ödemenin gerçekleşip gerçekleşmediğinin bilgisidir.



Şekil 4.2 : Durum bilgisi dağılımı.

- **durum_tipi**

Müşterinin bireysel müşteri mi yoksa kurumsal müşteri mi olduğunun bilgisidir. Müşteri kurumsal müşteri ise birden fazla hat sahibi olduğundan, borç tutarı yüksek olabiliyor.



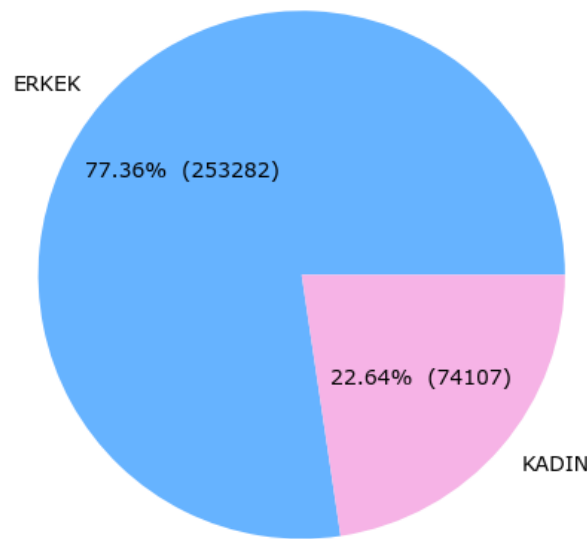
Şekil 4.3 : Durum tipi dağılımı.

- **fatura_sayısı**

Müşterinin, telekomünikasyon şirketinin müşterisi olduğu süre boyunca kaç adet fatura ödediğinin bilgisidir.

- **cinsiyet**

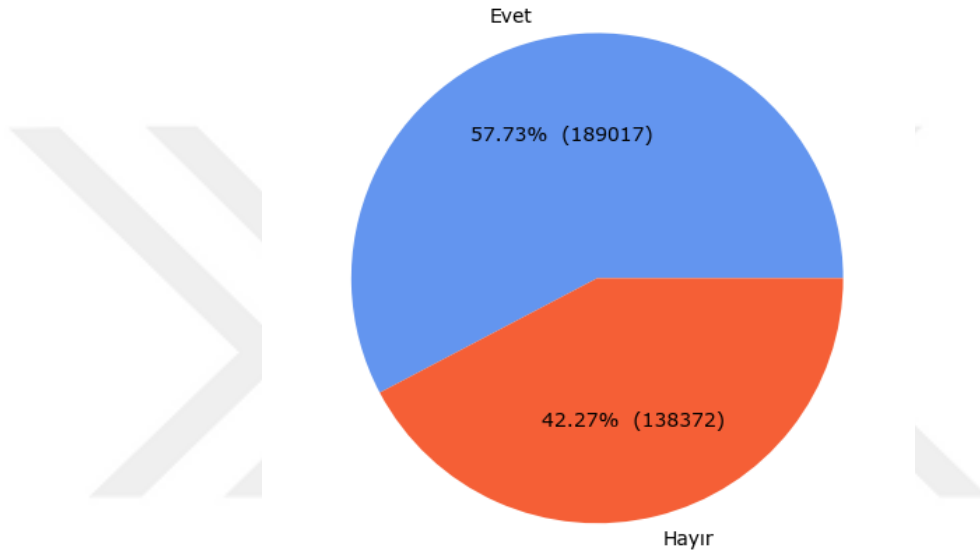
Müşterinin cinsiyeti kadın/erkek tipinden 0 ve 1 olarak kategorik olarak değiştirilmiştir.



Şekil 4.4 : Cinsiyet dağılımı.

- **iletisim_kuruldu**

Müşterinin borcunu ödemediği 120 günlük idari takip süresinde, telekomünikasyon şirketi çalışanları tarafından arandığında ulaşıp ulaşılmadığı bilgisidir. Bu süre içerisinde ulaşıp borç tahsil edildiyse, müşteri avukatlık bürolarına devredilmez. Ancak bazı durumlar da ulaşılsa bile borç tahsil edilemiyor. Bu durumlar için bu alan 1 olarak işaretlenmiş şekilde veri kümesinde yer almıştır. Aşağıdaki grafik telekomünikasyon şirketinin idari takip süresinde müşteriye arama yaparak ulaşabilip ulaşamadığının bilgisi içermektedir.



Şekil 4.5 : İletişim durumu dağılımı.

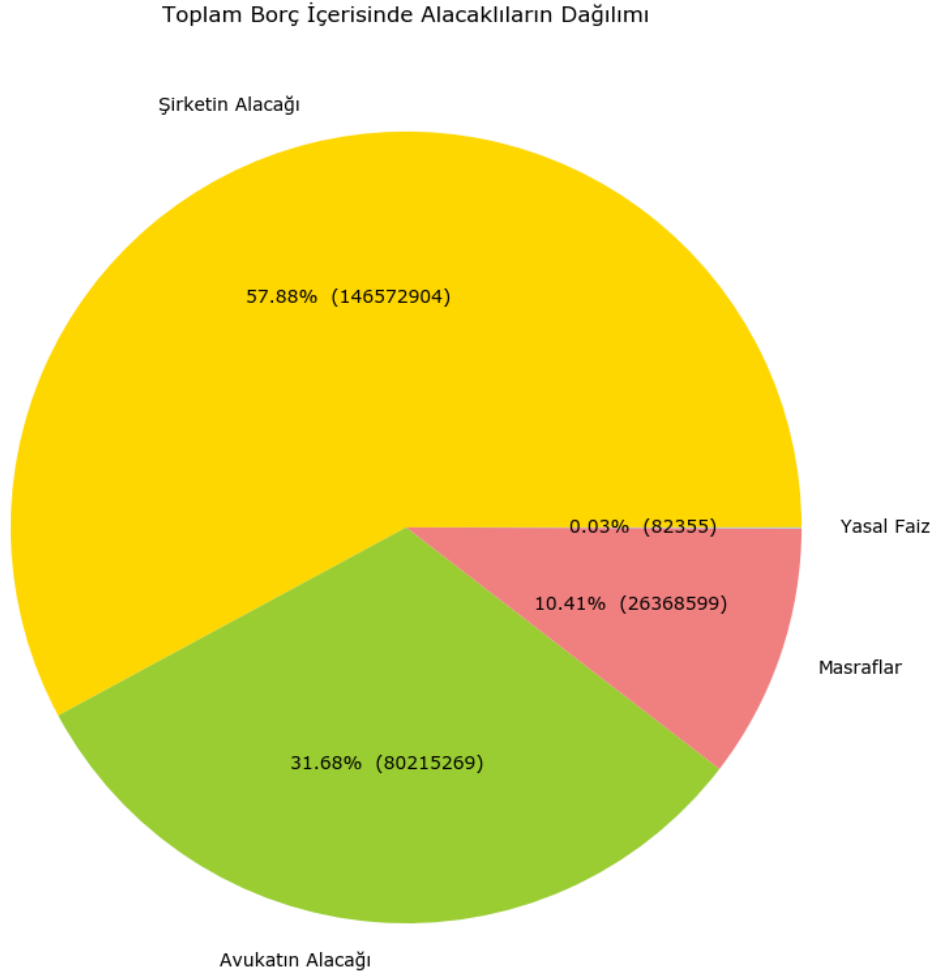
- **ilk_tutar & masraf & avukat_ucreti & yasal_faiz**

İlk tutar; müşterinin telekomünikasyon şirketine olan borcudur. Ödemediği faturaların tutarıdır.

Masraf; yasal takip sürecinin başlaması için, müşteriye karşı bir avukat tarafından dava açılmış olması gerekmektedir. Davanın avukat tarafından açılması zorunluluğu yasal gerekliliktir. Masraf tutarı da bu davanın açılması için ödenen harç bedelidir.

Avukat ücreti; Telekomünikasyon şirketi tarafından, dış kaynağa yönlendirilen tahsilattan kaynaklanan avukat ücreti bedeli müşteriden tahsil edilir. Bu tutar borcun miktarına göre yasalar ile belirlenmiş bir tutardır. Müşterinin borcuna eklenerek müşteriden talep edilir.

Yasal faiz; borcun ödenmeyen kısmının devlet tarafından belirlenmiş bir yüzde ile çarpımına karşılık gelen bedeldir.

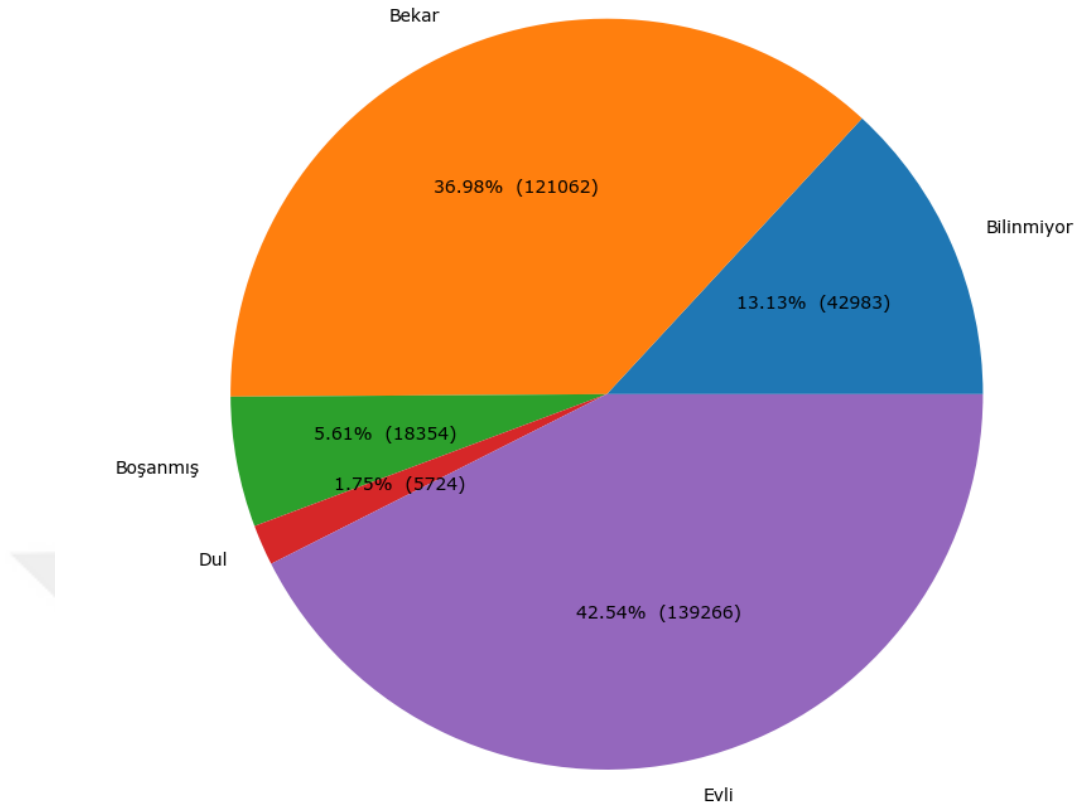


Şekil 4.6 : Toplam Borç içerisinde alacaklıların dağılımı.

- **medeni_durum**

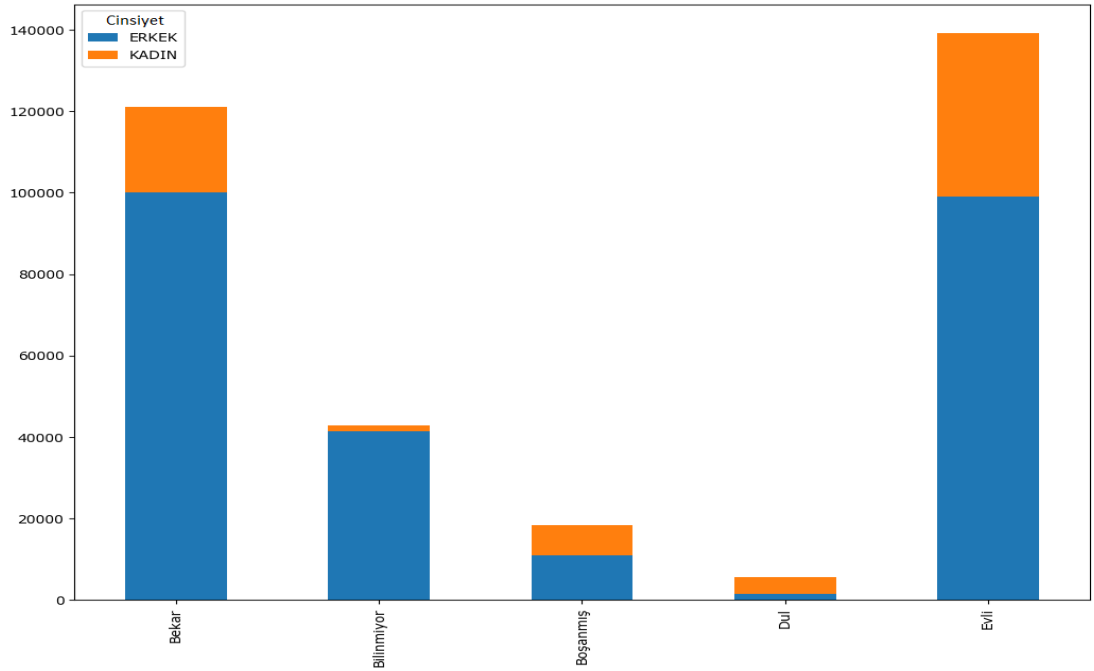
Müşterinin medeni durum bilgisidir. Bu noktada, 4 ayrı kategori bulunmaktadır.

Bunlar; evli, bekar, boşanmış, dul ve bilinmeyen boş değerlerdir.



Şekil 4.7 : Medeni durum dağılımı.

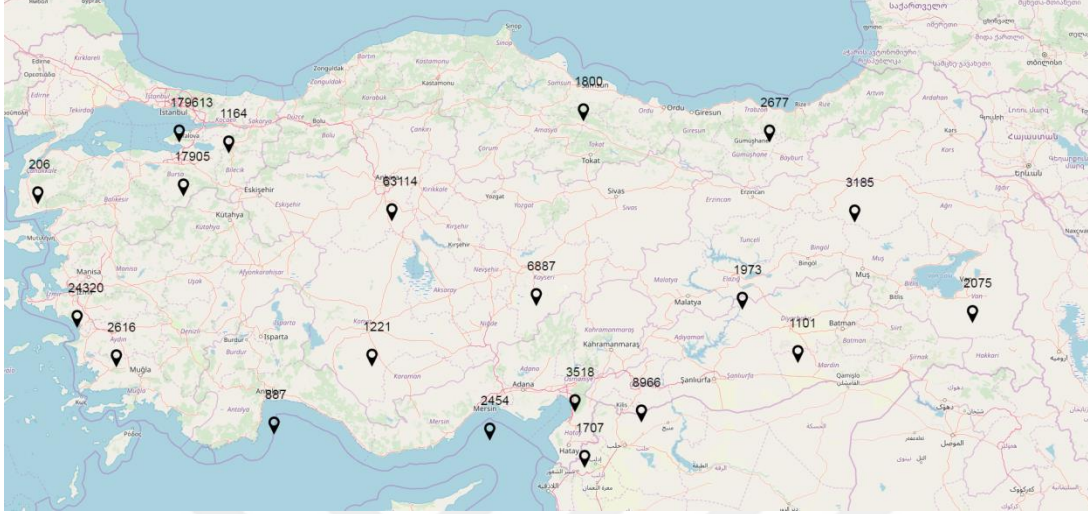
Müşterinin cinsiyetinin medeni durum üzerindeki dağılımı ise;



Şekil 4.8 : Medeni durum ve cinsiyet dağılımı.

- **ofis_sehir**

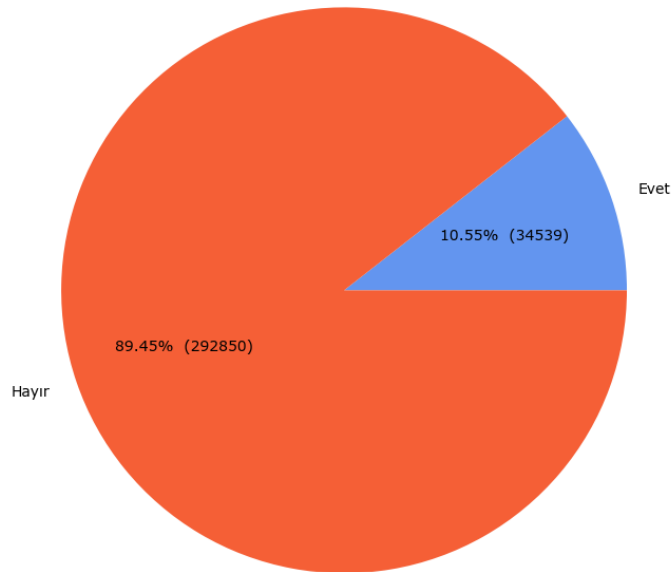
Müşterinin atandığı büronun olduğu ilin plaka kodudur. Telekomünikasyon şirketinin her şehirde, paydaşı büro bulunmamaktadır.



Şekil 4.9 : Büro şehirleri.

- **taksit_yapildi**

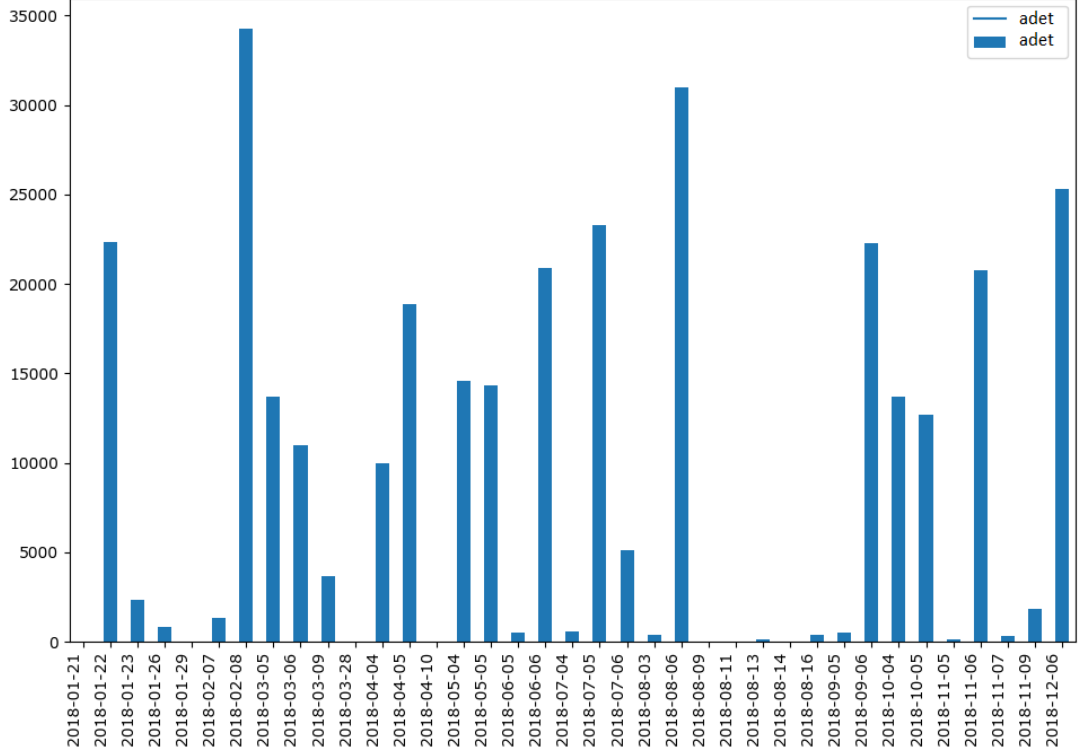
Yapılan aramalarda, müşterilerden bir kısmından gelen talep doğrultusunda borcun taksitlendirilmesi istenebiliyor. Bu süreci, telekomünikasyon şirketi yönetmiyor. Borcu tahsil edebilmek için büro yönetiyor. Toplam müşterilerin ~%10'u taksitlendirmeyi tercih ediyor.



Şekil 4.10 : Taksit durumu dağılımı.

- **transfer_date**

Müşterinin avukatlık bürosuna atandığı tarihtir. Genelde, atama işlemi, ayın ilk günlerinde ve tüm müşteriler bürolara toplu halde atılmaktadır.



Şekil 4.11 : Transfer tarihi grafiği.

- **yabanci_uyruk**

Telekomünikasyon şirketi müşterilerini veri tabanına kaydederken, hangi ülkenin vatandaşı olduklarını da kaydetmektedir. Ancak şirketin izlediği ön-ödemeli politika sebebiyle yabancı uyruklu müşteri sayısı, tüm müşteri sayısına oranla oldukça azdır. Çizelge 4.2’de müşterilerin hangi ülke vatandaşı olduklarına dair bilgiler paylaşılmıştır.

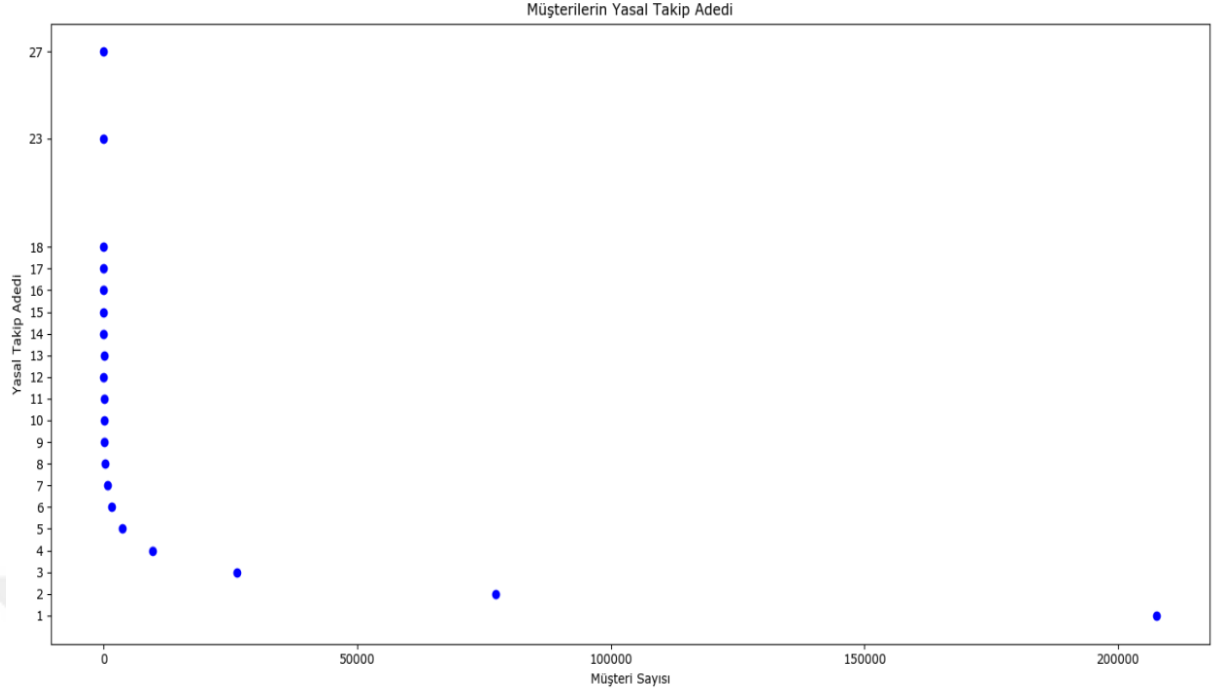
Çizelge 4.2 : Müşterilerin uyruklarına göre dağılımı.

Ülke	Müşteri Sayısı
Afganistan	12
Almanya	18
Andora	1
Arjantin	2
Arnavutluk	1
Avustralya	2
Avusturya	1

Azerbaycan	53
Beyaz Rusya	1
Ermenistan	2
Fas	2
Filipinler	2
Filistin	2
Fransa	3
Güney Afrika	1
Gürcistan	18
Irak	20
Katar	1
Kuveyt	3
Libya	2
Lübnan	1
Moldova	4
Nijerya	1
Rusya	4
Somali	1
Sudan	2
Suriye	148
Suudi Arabistan	3
Tacikistan	3
Tunus	2
Türkmenistan	24
Yeni Zelanda	1
Ürdün	1
Yunanistan	3
Özbekistan	96

- **yasal_takip_sayi**

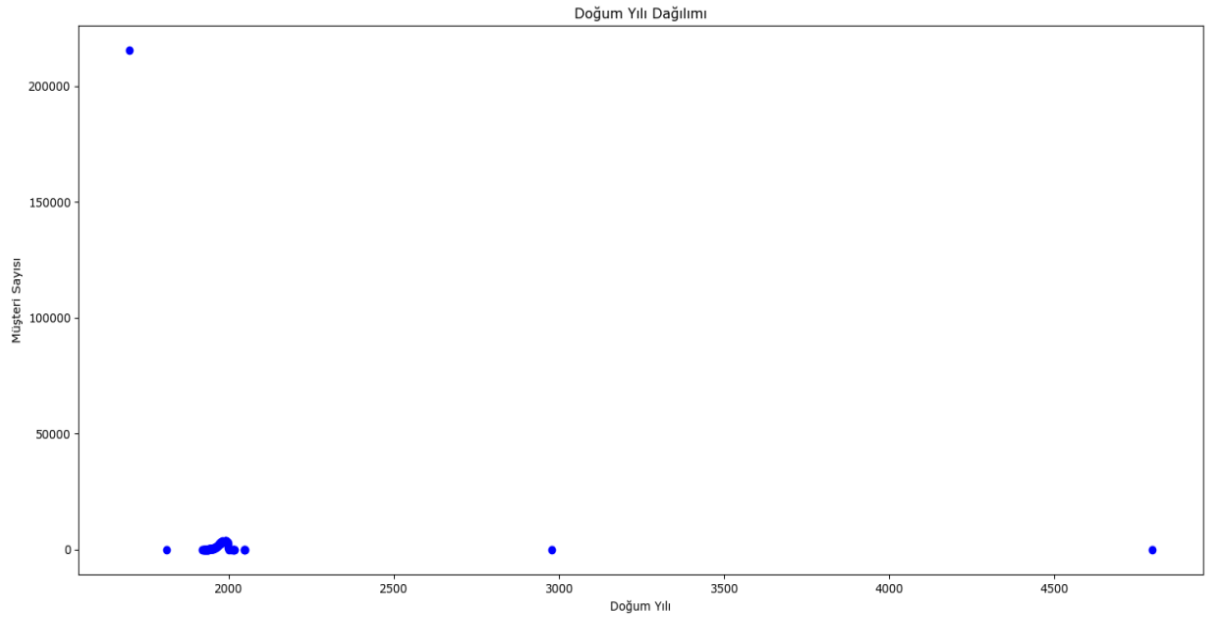
Yasal takibe girmiş olan müşterin hayatı boyunca, başka nedenler veya yine aynı nedenle toplam kaç kez yasal takibe girdiğinin bilgisidir. Müşterinin yasal takip sürecini bilmesi veya bu durumun müşteri alışkanlığı olarak değerlendirilip, sonuç üzerinde fayda sağlayabileceği düşünülerek eklenmiştir.



Şekil 4.12 : Müşterilerin yasal takip adedi dağılımı.

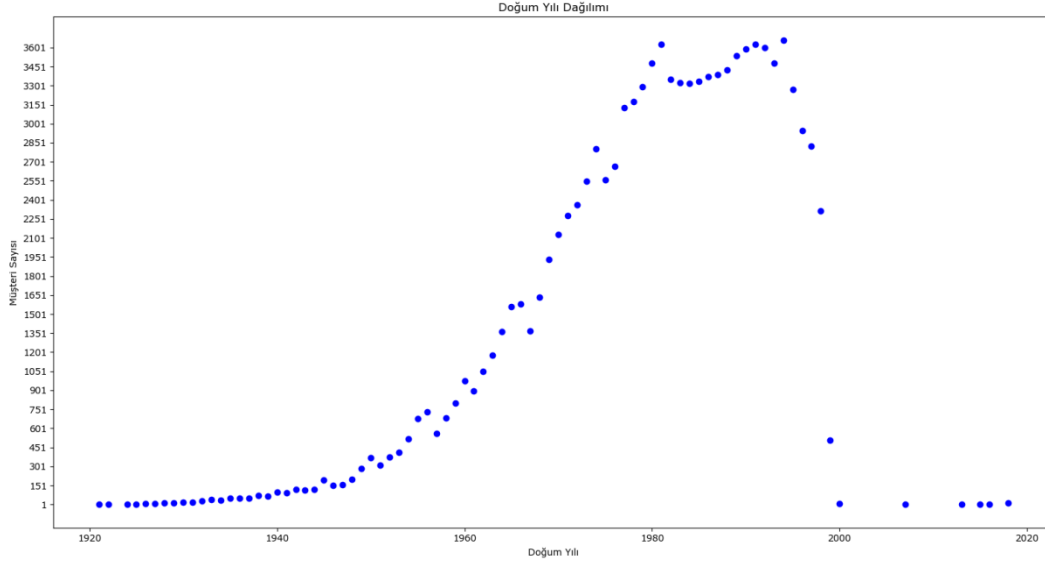
- **dogum_tarihi**

Müşterinin doğduğu tarih bilgisidir.



Şekil 4.13 : Doğum tarihi dağılımı.

Ancak grafikten de görüldüğü üzere dağılım bozuk değerler nedeniyle düzgün incelenememektedir. Dağılımı daha yakından inceleyebilmek için grafiğin uç noktalarındaki bozuk veriyi çıkartarak bir grafik daha çizdirecek olursak:

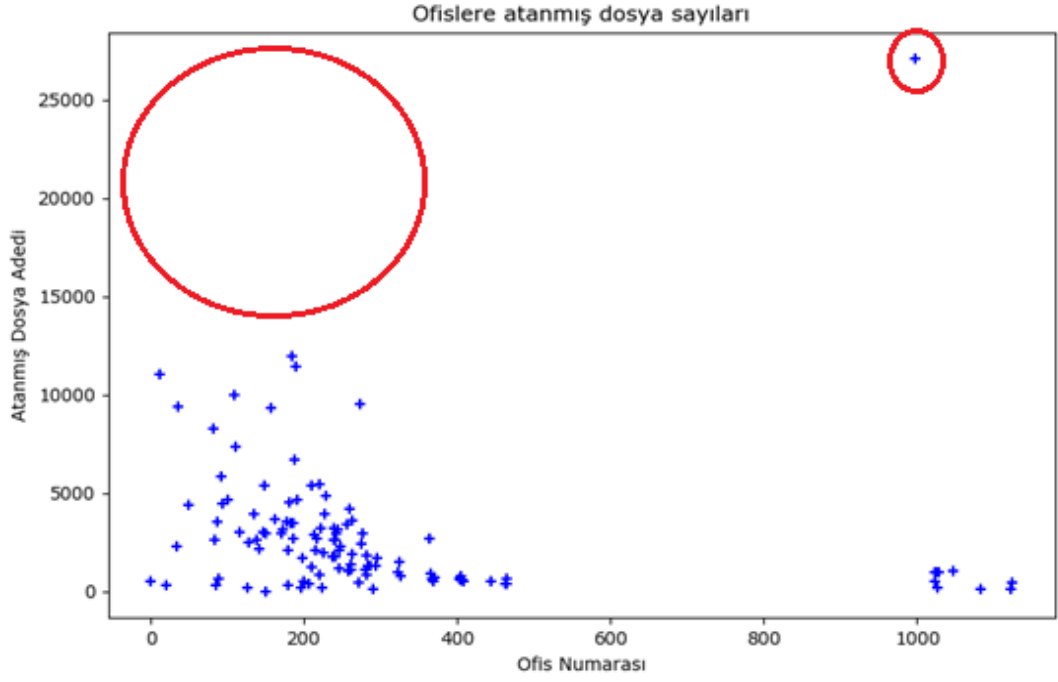


Şekil 4.14 : Doğum tarihi dağılımı (uç noktalardan arındırılmış).

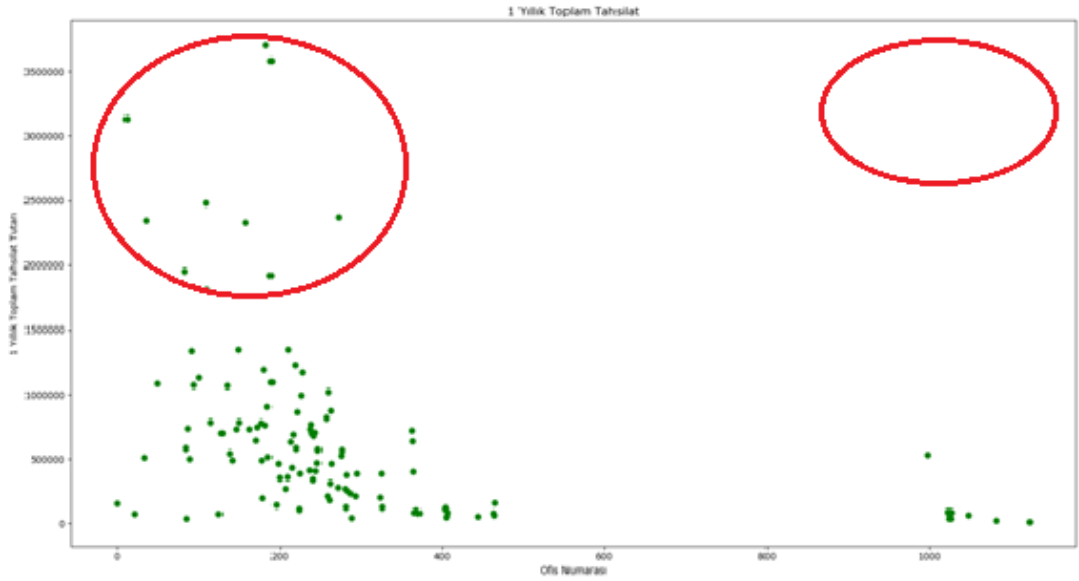
- **Büro_id**

Dosyanın atandığı avukatlık bürosunun kimlik numarasıdır. Dosyanın büro ataması yapıldıktan sonra büro ile çalışmaya devam edildiği sürece o büro üzerinde kaldığı varsayımı ile ilerlenecektir. Büro ile çalışmanın kesilmesi, büronun kapanması gibi nedenlerle devir işlemleri olabiliyor. Ancak bu durum bu çalışma kapsamında incelenen veriden arındırılmıştır.

Ancak verinin görselleştirilmesiyle ortaya çıkan şekiller sayesinde bu çalışmanın gerekliliği de ortaya çıkmış bulunmaktadır. Şekil 4.15'te ve Şekil 4.16'da görüldüğü üzere sağ üst köşede yuvarlak içerisine alınan kısımda da görüldüğü gibi, bir büroya atanan dosya sayısı çok fazla halde, bunun tutar karşılığı oldukça küçüktür. Bu demek oluyor ki, az miktarda borca sahip çok sayıda müşteriden tahsilat gerçekleştirilmesi beklenmektedir. Tam tersi şekilde, sol üst tarafta bir yoğunluk olmamasına rağmen, tutar alanında yoğunluk görülmektedir. Bu durum yüksek borç tutarlarına sahip müşterilerin atandığı bürolar olduğunu göstermektedir.



Şekil 4.15 : Bürolara atanmış yıllık toplam dosya sayısı.



Şekil 4.16 : Büro bazında 1 yıllık toplam tahsilat tutarı.

4.2 Uygulama Verisi Üzerinde Yapılan İşlemler

Kullanılacak veri her zaman ilk alındığı hali ile kullanılmayabilir veya istenilen özellikler olmayabilir. Verinin ilk hali eksik, yanlış veya gürültülü olabilir. Bu durumun önüne geçmek için aşağıdaki adımları uygulayacağız.

4.2.1 Veri temizleme

Veri kümesinde eksik, hatalı veya güncellenmemiş veriler bulunabilir. Bu verileri kullanarak yapılacak olan değerlendirme hatalı ve yanlış sonuçlar çıkmasına neden olabilir. Ayrıca, modelin öğrenme süresini uzatabilir. Bu durumu önlemek için literatürde yapılmış işlemler aşağıda listelenmiştir:

- Eksik değerler veri kümesinden silinebilir.
- Eksik değerler o değerlerin ortalaması ile doldurulabilir.
- Eksik değerler kendisine en yakın hücredeki değer ile doldurulabilir.
- Eksik değer rastgele seçilmiş bir kolondaki değer ile doldurulabilir.
- Eksik değerleri diğer değerlerden ayırmak için ayrı bir değer verilebilir.
- Eksik değer regresyon, bayes veya karar ağaçları gibi yöntemlerle tahmin edilip doldurulabilir.

Bu çalışmada kullanılan veri kümesindeki bozuk veya hatalı değerler veri kalitesinin korunması için veri kümesinden çıkarılmıştır. Güncellenmemiş veriler tespit edilip güncellenmiştir. Bazı değerlere ortalama değer atanmıştır. Üzerinde işlem yapılan veriler aşağıda listelenmiştir.

- **Şirketin alacağı borç tutarı & Avukatın alacağı tutar & Masraflar & Yasal faiz vergi tutarı**

Boş olan değerler ödemenin gerçekleşmediği müşteriler için sıfır olarak güncellenmiştir. Aslında şirket açısından kendi alacağı borcun tahsil edilmiş olması önceliklidir. Müşteri ödeme yaparsa ilk olarak şirkete olan borç kapatılmaktadır. Ancak avukatın alacağı tutar veya masraf fazla ise bu müşterinin ödemeye karşı olan motivasyonunu azaltabileceği düşünülüp incelemeye/araştırmaya katılmıştır.

- **Yaşadığı il**

Yaşadığı il alanı boş olan 38775 adet kayıt bulunmaktadır. Bu değerler İstanbul gibi düşünülecektir. İstanbul, Türkiye'nin en kozmopolit ve en kalabalık şehridir. Diğer illerden göç almaya devam etmektedir. Türkiye nüfusunun ~%20'si İstanbul'da yaşamaktadır. Bu nedenle bu müşteriler için bu alan 34 olarak kabul edilmiştir.

- **Doğum tarihi**

Doğum tarihlerinde bulunduğumuz tarihten ileri tarihli olan 5 kayıt bulunmaktadır. Ayrıca 16 kayıt şu anki tarihe çok yakın olduğundan bu değerlerin de doğru olmadığı

düşüncesiyle ve veri kalitesinin korunması için bu müşteri kayıtları veri kümesinden çıkarılmıştır.

327389 kaydın 215445 tanesinin doğum tarihi bilgisi bulunmamaktadır. Bu boş değerler, dolu olan kolonların ortalama değeri ile değiştirilmiştir. Gün, ay ve yıl olarak ayrı ayrı ortalama alınarak bulunan tarih; “1980-06-12” olmuştur.

- **Borçlu Durumu**

Müşterinin toplam borcu ile müşteriden tahsil edilmiş toplam tahsilat miktarı eşit olan, ancak halen ‘borçlu’ statüsünde kalmış olan 76 adet kayıt görülmüştür. Henüz sisteme yansımamış güncelleme olduğu bilgisi telekomünikasyon şirketinden teyit alınmış ve bu kişilerin durumu ‘ödendi’ olarak güncellenmiştir.

Borçlu durumu ‘ödendi’ olmasına rağmen hiçbir tahsilat kaydı olmayan 1862 adet müşteri tespit edilmiş. Bu müşteriler ile şirket arasında farklı ödeme kanalları kullanılması nedeniyle veri kümesinden çıkarılmıştır.

4.2.2 Veri indirgeme

Elde bulunan veriler içerisindeki niteliksiz, ilgisiz veya tekrarlı veriler veri kümesinden çıkarılarak, hem sistemin performansında iyileştirme yapılabilir hem de daha doğru sonuçlar elde edilebilir. Bu nedenle aşağıdaki alanlar üzerinde düzenleme yapılmıştır.

- **Meslek**

Meslek alanındaki bilgilerin isimlendirilmesinde bazı kayıtlar şirket adı olarak girilmiş olduğu, bazı alanların şehir adı olduğu ve bazı değerlerin anlamsız olduğu görülmüştür. Ayrıca, bu alanda müşterilere ait bilgilerin ~%70’inin boş olması, bilgi paylaşmış olan müşterilerin de bilgilerinin doğruluğunun teyit edilmeden sisteme girilmiş olması ve bu alandaki verilerin güvenilirliği konusunda şüpheye düşüldüğünden dolayı bu alan incelemiden çıkarılmıştır.

4.2.3 Veri genişletme

Verilerin modele aktarımı modelin eğitilmesi üzerinde doğrudan etkili olduğundan modelin başarısını değiştirebilir. Bu nedenle, mevcut özniteliklerin yeterli olmadığı durumda eldeki verilerden faydalanarak yeni öznitelikler üretilmesi modelin öznitelikleri daha iyi anlamasını ve daha doğru sonuçlar üretmesini sağlayabilir.

- **Tarih bilgisi**

Python'da tarih (date) türünde bir veri tipi bulunmamaktadır. Tarih bilgisi modele sayısal değerler olarak aktarılabilir. Bu alana ait bilgi ne kadar detaylı aktarılırsa, modelin performansında da o kadar etkili olacaktır. Dosyanın büroya gönderildiği tarihin;

- Hangi ay olduğu,
- Ayın kaçınıcı günü olduğu,
- Haftanın hangi günü olduğu,
- Yılın hangi çeyreği olduğu,
- Yılın kaçınıcı haftası olduğu,
- Ayın ilk günü/son günü olup olmadığının incelenmesi,
- Ayın kaçınıcı aralığında olduğu (10 günlük aralıklar ile 0-10;11-20;21-31),
- Mevsim bilgisi

Ödemeler ay sonunda yapıldığından ötürü, transfer tarihiyle ilgili daha açıklayıcı bir şekilde tarih bilgisini aktarabilmek için yukarıdaki alanlara ilave olarak aşağıdaki alanlar da sistemde öznitelik olarak belirtilmiştir.

- Transfer tarihinden ay sonuna kadar olan toplam işgünü sayısı,
- Transfer tarihinden ay sonuna kadar olan toplam gün sayısı,
- Transfer tarihinin yıl sonuna olan uzaklığı (kategorik 0-30;30-60;60-90;90-120..)

- **Müşteri ile büronun aynı şehirde olup olmadığı**

Şirketin çalıştığı büroların Türkiye üzerindeki dağılımı incelendiğinde görülmüştür ki, bürolar her şehirde bulunmamaktadır. Bu nedenle bazı müşteriler farklı illerde bulunan bürolara atanmaktadır. Bu durumun müşteriden tahsilat yapılabilmesi konusunda bir etkisinin olup olmadığı konusunda inceleme yapabilmek için bu kontrol de kategorik olarak eklenmiştir (0 -> farklı şehir, 1 -> aynı şehir, 2 -> bilinmiyor).

- **Müşteri ile büronun aynı bölgede olup olmadığı**

Şirketin çalıştığı büroların Türkiye üzerindeki dağılımı her şehirde bulunmamaktadır. Bu nedenle bazı müşteriler farklı coğrafik bölgelerde bulunan bürolara atanmaktadır. Bu durumun müşteriden tahsilat yapılabilmesi konusunda bir etkisinin olup olmadığı konusunda inceleme yapabilmek için eklenmiştir

- **Bayram olup olmadığı**

Bayramlar, Ramazan ve Kurban Bayramı olmak üzere bir yıl içerisinde iki kez dini bayramımız vardır. Bayram süresince 3-4 günlük tatil yapma imkanıyla beraber kişiler dini görevlerini yerine getirebilmektedir veya gelenekleri yaşatmak için akraba / memleket ziyaretleri veya tatil planı yapmaktadırlar. Bu durum neticesinde diğer aylara oranla fazladan harcama yapmaları gerekmektedir. Bu durum bankaların bu dönemlerde, 'bayram kredisi' reklamları yapmalarıyla da kişilerin bu dönemlerde fazladan nakite ihtiyacı olduğu ile ispatlanabilir.

- **Tutarların toplam borç üzerindeki yüzdeleri**

Şirketin alacağı miktarın toplam borç üzerindeki oranı veya masraf bedelinin toplam tutar üzerindeki oranı müşterinin borcu kapatma isteği üzerinde etkili olup olmadığının incelenmesi amacıyla eklenmiş bir öz niteliktir.

- **Toplam borç tutarı**

Müşterinin borç kalemlerinin (şirkete olan borcu, avukat ücreti, masraf bedeli, yasal faiz tutarı) toplanmış halidir.

- **Dolar kuru ve borcun dolar karşılığı**

Borçlu müşterinin büroya devredildiği aya ait dolar kuru değeri ve müşterinin borcunun dolar cinsinden karşılığı da öz nitelik olarak verilmiştir.

4.2.4 Veri kategorize hale dönüştürme (Mapping)

Bazı kelimelerin kategorikleştirilerek sayısallaştırılması işlemine tabi tutulması gerekir. Bu işleme mapping denir. Aşağıda kategorik hale dönüştürülmüş öz nitelikler listelenmiştir.

- **Müşterinin yaşadığı coğrafik bölge**

Türkiye, tabii, beşeri ve ekonomik olarak benzer özellikler gösteren 7 bölgeye ayrılmıştır. Müşteriye ait şehir bilgisi bilindiğinden, daha genel bir çerçeveden inceleyebilmek ve belirli bölgelere ait belirli bir davranış tutumu izleyebilmek adına şehirler bölgelere dönüştürülmüştür. Çizelge 4.3'te bölgelerde yaşayan müşteri sayıları verilmiştir.

Çizelge 4.3 : Müşterilerin yaşadığı coğrafik bölgelerin dağılımı.

Bölge	Müşteri Sayısı
Marmara Bölgesi	81566
Ege Bölgesi	25592
Akdeniz Bölgesi	35854
İç Anadolu Bölgesi	42605
Doğu Anadolu Bölgesi	43756
Güneydoğu Anadolu Bölgesi	39165
Karadeniz Bölgesi	58851

- **Cinsiyet ve durum tipi kategorik karşılıkları**

Cinsiyet için kadın ve erkek 0-1'li yapıya dönüştürülmüştür. Durum tipi alanı için bireysel ve kurumsal 0-1 olarak değiştirilmiştir.

- **Uyruk**

Yukarıda belirtildiği üzere veri kümesinde müşterinin ait ülkenin adı yazmaktaydı ancak müşterinin hangi ülkenin vatandaşı olduğundan ziyade TC vatandaşı olup olmamaları ile ilgilenilmiştir. Yabancı uyruklu müşterilerin Türkiye'de taşınmaz mallarının olmama ihtimalinin yüksek olması nedeniyle mahkeme sürecinde, müşteriye ait bir bağlayıcılık tespit edilemediğinden uyruk bilgisinin önemli olacağı düşünülmüştür. Çizelge 4.3'te müşterilere ait uyruk bilgisi verilmiştir. Uygulamada ikili değer olarak 0 veya 1 olarak kullanılmıştır.

Çizelge 4.4 : Uyruk bilgisi.

Uyruk	Müşteri Sayısı
T.C. Vatandaşı	326948
Yabancı Uyruk	441

- **Yaş aralığı**

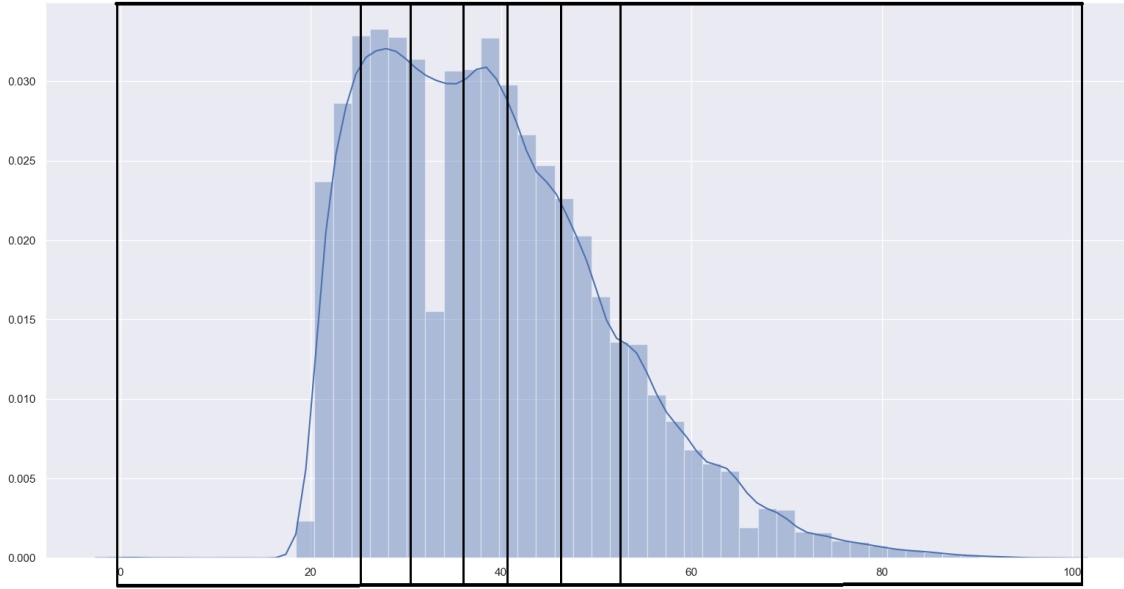
Şimdiki tarihten müşterinin doğum yılının çıkartılarak müşterinin yaşı bulunmuştur. Ancak bu değer bir istatistik elde edebilmek amacıyla aralıklara bölünmüştür. Yaş değeri negatif veya 18 yaşından küçük olan müşteriler veri kalitesinin korunması

amacıyla örnekten çıkartılmıştır. İlk olarak yaş aralıkları eşit aralık olacak şekilde bölünmüştür. Çizelge 4.5’te yaş aralıkları verilmiştir.

Çizelge 4.5 : Yaş aralıkları I.

Yaş aralıkları	Kategori numarası
18-25	1
26-35	2
36-45	3
46-55	4
56-65	5
66-100	6
Doğum tarihi bilinmeyenler	7

Ancak elimizdeki veri kümesinde doğum tarihi bilinmeyen müşteri sayısı toplam veri kümesinin yaklaşık 3’te 2’sine denk gelmektedir. ~115K’lık verinin ortalamasını, geri kalanlara atamak ve bunun üzerinden veriyi kategorize etmek sapmanın büyük olmasına neden olabilir. Bu nedenle aşağıdaki yaş dağılımı da veri üzerinde denenmiştir. Aşağıda Şekil 4.18’de görülen yaş aralıklarını, müşteri sayısı eşit olacak şekilde bölünmüş aralıklarla gösterimidir. Bu dağılım sonucunda oluşan yaş aralıkları Çizelge 4.6’da verilmiştir.



Şekil 4.18 : Yaş aralıkları I.

Çizelge 4.6 : Yaş aralıkları II.

Yaş aralıkları	Kategori numarası
0-24	1

25-31	2
32-36	3
37-41	4
42-46	5
47-52	6
53-98	7

Yaş özniteliği için iki farklı çizelgeden de üretilmiş sonuçlar öznitelik önemi belirleme yöntemlerine uygulanmış. Bu iki çizelgeden daha yüksek öneme sahip olan öznitelik o model için tercih edilmiştir. Bölüm 4.5.2 - 4.5.5'te öznitelik seçim listelerinde seçilmemiş modelin üzeri çizilerek onun seçilmediği belirtilmiştir.

- **Medeni durum**

Müşterilere ait medeni durum bilgisi şirket tarafından evli, bekar, boşanmış, dul ve bilinmeyen olarak 5 farklı kategoride toplanmıştır. Bu değerler modele nümerik olarak aktarımı Çizelge 4.7'de gösterilmiştir. Ancak bu medeni durum bilgileri kişinin şirketin müşterisi olmak için yaptığı başvuru sırasında alınan bilgisidir. Kişi borçlu duruma düştüğündeki medeni durumu değişmiş olabilir. Ayrıca, kaydı tutan personel bu bilgiyi, kişinin kimliğinden teyit etmemiş olabilir.

Çizelge 4.7 : Medeni durum I.

Medeni durum	Kategori numarası
Evli	0
Bekar	1
Boşanmış	2
Dul	3
Bilinmeyen	4

Hem yukarıda bahsedilen yanlış kayıtları düzenleyebilmek için hem de boşanmış ve dul kişilerin de bekar diyebileceği ve yalnız yaşadıklarından benzer davranış gösterme eğilimleri olabileceği düşüncesiyle ikinci bir kategorize grup daha oluşturulmuştur. Çizelge 4.8'de değerler gösterilmiştir. Bu durum öznitelik seçimi metotlarına medeni durum2 olarak verilmiş ve bu iki medeni durum öznitelikleri arasından hedef değişken ile daha fazla ilişkili olan seçilmiştir. Diğerinin üzeri çizilmiştir (Bölüm 4.5.2 - 4.5.5'te).

Çizelge 4.8 : Medeni durum II.

Medeni durum	Kategori numarası
Evli	0
Bekar	1

Boşanmış	1
Dul	1
Bilinmeyen	2

4.2.5 Normalizasyon

Bazı durumlarda veriyi uygulamaya aynen katmak uygun olmayabilir. En alt ve en üst değerlerin çok yüksek olması yani değerler arasındaki farklılığın fazla olması veya tüm verinin belirli bir aralıkta yığılması modellerin üreteceği sonucu etkileyecektir. Diğer özniteliklere olan baskı artacaktır. Bu nedenle bu veriler normalize edilerek aralarında uçurum azaltılmış olur.

Öznitelik seçimi ve sınıflandırma gibi yöntemlerin başarısı yüksek olabilmesi için, kullanılan veri kümesinin kalitesi önemlidir. Bundan dolayı verinin ön işleme sürecinde normalizasyona tabi tutulması çok önemlidir. Verinin kalitesini ve dolayısıyla uygulanan yöntemlerin başarısını artırmaktadır.

Öznitelikler arasında, kategorik veya ikili (binary) olmayan tüm alanlara normalizasyon uygulanmıştır. Bu alanlar; tutar bilgileri (şirkete olan borç, masraflar, avukatlık ücreti, yasal faiz tutarı), yasal takibe girme sayıları, o aydaki iş günü sayısı gibi, transfer yapıldıktan sonra ay sonuna kalan gün sayısı, fatura sayısı, müşterilik süresi, toplam borç miktarı gibi..

Diğer bir kullanılışı ise farklı ölçekleme sistemlerinde bulunan verileri birbiri ile karşılaştırabilme imkanı sunmasıdır. Örneğin bu çalışma kapsamında ağaç tabanlı öznitelik seçimi (feature selection) yöntemleri uygulanmış, bunların çıktılarının ortalaması hesaplanarak model üzerinde ağaç tabanlı modellerin öznitelik seçimi üzerinde ortak etkisi incelenmek istenmiştir. Bu kıyaslama yapılmaya çalışılırken, her bir ağaç modelinin farklı değer aralıklarında sonuçlar ürettiği görülmüştür. Her bir modelin çıktısına ilk olarak min-max normalizasyon uygulanıp hepsinin önem derecesi 0-1 aralığında bir değer aldıktan sonra ortalama hesaplanmıştır. Bu kapsamda uygulanan modeller;

- Karar ağaçları
- Rastgele orman
- Extra Gradyen
- Gradyen Boosting
- AdaBoost (Adaptive Boosting)

- Extra Trees

Ağaç modellerinin ürettiği sonuçlar Ek A'da yer almaktadır. Ayrıca normalizasyon uygulandıktan sonraki değerlere de Çizelge A.2'den ulaşılabilir.

Normalizasyon yapmak için birden fazla yöntem vardır. Bunlardan en yaygın kullanılanı min-max normalizasyondur. Bu yöntemlere ait formüller aşağıda listelenmiştir.

- **Min-Max normalizasyon**

Min-Max normalleştirme uygulanacak öznitelik için minimum değer ve maksimum değer baz alınarak tüm değerleri [0,1] aralığına yerleştirir.

Maksimum değer 1 değerini alır. Minimum değer ise 0 değerini alır.

Formül ise:

$$X^* = \frac{X - \min(X)}{\max(X) - \min(X)} \quad (3.24)$$

- **Ondalık ölçekleme**

Min-Max normalleştirme dışında, ondalık ölçekleme yöntemi de öznitelik değerlerinin normalize edilmesi için kullanılmaktadır. Öznitelik değerlerinin ondalık kısmını değiştirerek verilerin [-1,+1] aralığında değer almaları için kullanılabilir. (Wang & Chen, 2007)

$$X^* = \frac{X}{10^k} \quad (3.25)$$

- **Z - skor**

Bu yöntem ile öznitelik, ortalama değeri 0, standart sapma değeri de 1 olacak şekilde standartlaştırılır. Kendisinden ortalama değer çıkartılarak bulunur.

$$X^* = X - E[X] \quad (3.26)$$

Ayrıca bu çalışmada, veri üzerinde uygulanan normalizasyon işlemi, eğitim verisi ve test verisi olarak ayrıldıktan sonra ayrı ayrı uygulanmıştır. Test verisinin eğitim verisinden bağımsız bir şekilde normalize edilmesi eğitim verisi ve test verisinin birbirinden bağımsız kümeler olduğu durumunda yanlış sonuçlar üretebilir. Bu durumun önlenmesi için, eğitim verisine uygulanan normalleştirme işlemi sonucunda özniteliklere ait ilgili min-max değerleri saklanmıştır. Aynı değerler test verisini

normalize ederken kullanılmıştır. Böylece test verisi, eğitim verisine göre normalize edilmiştir.

4.2.6 Veri tabanı boyutu

Veri tabanı boyutunun yeteri kadar büyük olmaması durumunda model tam öğrenme sağlayamaz ve doğru sonuçlar veremez.

Bu kapsamda, bu çalışma için kullanılan küme ~327K farklı müşteri bilgisinden oluşmaktadır.

4.2.7 Kodlayıcı (Encoding)

Kategorik olan öznitelikler eğer ikili değerden daha fazla sayıda değere sahipse, bu alanlar kodlanmıştır. Kodlama ile, ikili değerler haricindeki kategorik öznitelikler, ikili hale dönüştürülmüştür. Bu dönüştürme işlemi; örneğin 3 değerli (0, 1, 2) bir kategorik öznitelik olduğu durumda, “0 mı?”, “1 mi?”, “2 mi?” şeklinde üç ayrı kolon olarak değerlendirilmiştir. Böylece ikili değere sahip olmayan kategorik öznitelikler de ikili gibi değerlendirilebilmiştir. Kodlama işlemi uyguladıktan sonra modelin performansında ciddi artış gözlenmiştir. Modelden modele değişiyor olsa da, doğruluk metriği üzerinde yaklaşık +[2, 4] arasında artış meydana gelmiştir. Kodlama işlemi uygulanan öznitelikler: büronun bulunduğu şehir, müşterinin şehri, müşterinin yaşadığı coğrafi bölge, mevsim, büro ve müşteri aynı coğrafik bölgede olup olmadığı (0 farklı; 1 aynı; 2 bilinmiyor), büro ve müşteri aynı şehirde olup olmadığı (0 farklı; 1 aynı; 2 bilinmiyor), atamanın gerçekleştiği ay, atamanın ayın hangi aralığında gerçekleştiği (1-10 1; 11-20 2; 21-31 3), atamanın haftanın kaçınıcı gününde yapıldığı, atamanın yılın hangi çeyreğinde ve kaçınıcı haftasında yapıldığı, atamadan yılsonuna kadar olan gün sayısı (0-30 1; 31-60 2; 61-90 3,..), medeni durum bilgisi, yaş kategorileri’dir.

4.2.8 K kat çapraz doğrulama

Algoritmalar, test yöntemi olarak kullanılan ve istatiksel bir yöntem olan “çapraz doğrulama” metodu ile beraber çalıştırılmıştır. Bu yöntemle veri kümesi k adet gruba ayrılır. Uygulamada k değeri 10 olarak verilmiştir. Kullanılan veri kümesi ~327K olduğundan, her bir grupta ~32K lık müşteri örneği olacaktır. Veri kümesinin içerdiği veri miktarına göre bu bölünmenin her zaman eşit olamayacağı açıktır. Sonrasında, her bölüm bir kez test verisi olur ve kalan diğer k-1 bölüm, bu çalışmada 9 bölüm,

öğrenme verisi olarak kullanılır. Bu süreç k defa tekrar eder. Böylece her grup, test kümesi olarak da incelenmiş olur. Çıkan sonuçların ortalaması başarı yüzdesi olarak alınır. Şekil 4.19’da çalışma prensibi verilmiştir.

k=1	TEST	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM
k=2	EĞİTİM	TEST	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM
k=3	EĞİTİM	EĞİTİM	TEST	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM
k=4	EĞİTİM	EĞİTİM	EĞİTİM	TEST	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM
k=5	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	TEST	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM
k=6	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	TEST	EĞİTİM	EĞİTİM	EĞİTİM
k=7	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	TEST	EĞİTİM	EĞİTİM
k=8	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	TEST	EĞİTİM
k=9	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	TEST
k=10	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	EĞİTİM	TEST

Şekil 4.19 : Çapraz doğrulama.

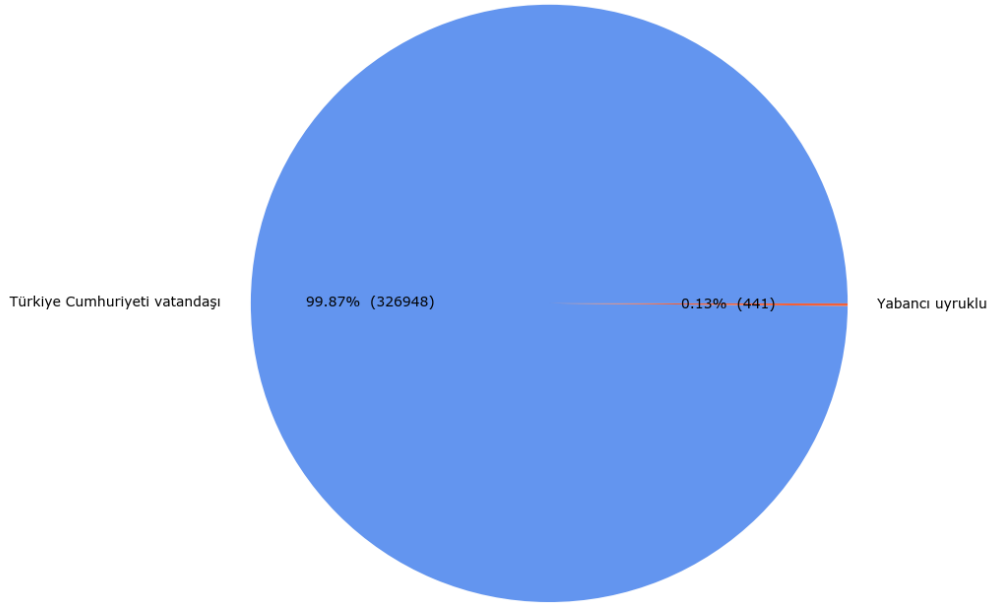
4.2.9 Veriyi eğitim ve test olarak bölme oranları

Modelin eğitilmesi ve test edilmesi için öznitelikler belirli bir oranda ayrıştırılır. Test için ayrılan veriler modelin eğitilmesi için kullanılmaz. Bu çalışmada bölme oranı %80’e %20 olarak belirlenmiştir. Eğitim verisi ~261K örnekten, test verisi ise ~66K’dan oluşmaktadır.

4.2.10 Veri kümesi dengeleme

Veri kümesi dengesizliği belirli bir sınıf sayısının, diğerlerine oranla azınlık kalması veya çoğunluğu oluşturması durumudur. Verilerin dengesiz durumu, makine öğrenme yöntemlerinin öğrenmesini zorlaştırmaktadır. Hatta bazen öğrenmeyi engellemektedir. Bu dengesizliğin giderilmesi için rastgele örnek eleme (under-sampling) veya rastgele aşırı örnekleme (over-sampling) gibi farklı yöntemler vardır. Rastgele örnek eleme yönteminde, çoğunluk sınıflardan rastgele örnekler silinir. Böylece, çoğunluk sınıfları azınlık sınıflarının belirli bir oranına denk getirilir. Rastgele aşırı örnekleme ile azınlık sınıflarının sayısı artırılarak çoğunluk sınıfının belirli bir oranına denk getirilir. Veri kümesinde çok fazla örnek bulunduğu durumda örnek eleme tekniği, yeterli örnek bulunmadığında ise aşırı örnekleme tekniği kullanmak daha uygundur. Bu çalışmada incelenen veri kümesinde yabancı uyruklu olma durumunun sonuç üzerinde etkisi olması beklenmiştir. Bu yöndeki çalışmalar ile literatürde karşılaşılmıştır. Ancak

yukarıda da bahsedildiği üzere, şirketin yabancılara uyguladığı ön ödemeli satış modeli sayesinde, yasal takibe girmiş yabancı uyruklu müşteri sayısı, T.C. vatandaşı sayısına oranla azınlık kalmaktadır. Bu durum Şekil 4.20’de gösterilmiştir. Bu nedenle yukarıda bahsedilen dengeleme yöntemleri, şirketin uyguladığı stratejide bir değişiklik yapma ihtimalinin olmaması ve bu nedenle yabancı uyruklu müşteri sayısında bir artış olmaması nedeniyle veri kümesinde uygulanmamıştır.



Şekil 4.20 : Veri kümesi dengeleme.

4.2.11 Veri kümesi rastgeleliği

Elimizdeki veri kümesi Ocak ayından başlayarak Aralık ayına kadar borçlu müşteri listesinin tarih için sıralı listesidir. Art arda gelen verilerin modelin ezberlemesi (over-fitting) üzerinde etkisi olabileceği, ayrıca k-kat çaprazlama ile veriyi 10 gruba böldüğümüzden, oluşacak grupların, yaklaşık olarak ay periyodlarına denk gelebileceği kaygısıyla, k-kat çaprazlama işlemi öncesinde, veri kümesindeki kayıtların sırası yer değiştirilmiştir.

4.3 İşlemler Sonucu Veri Kümesinin Son Hali

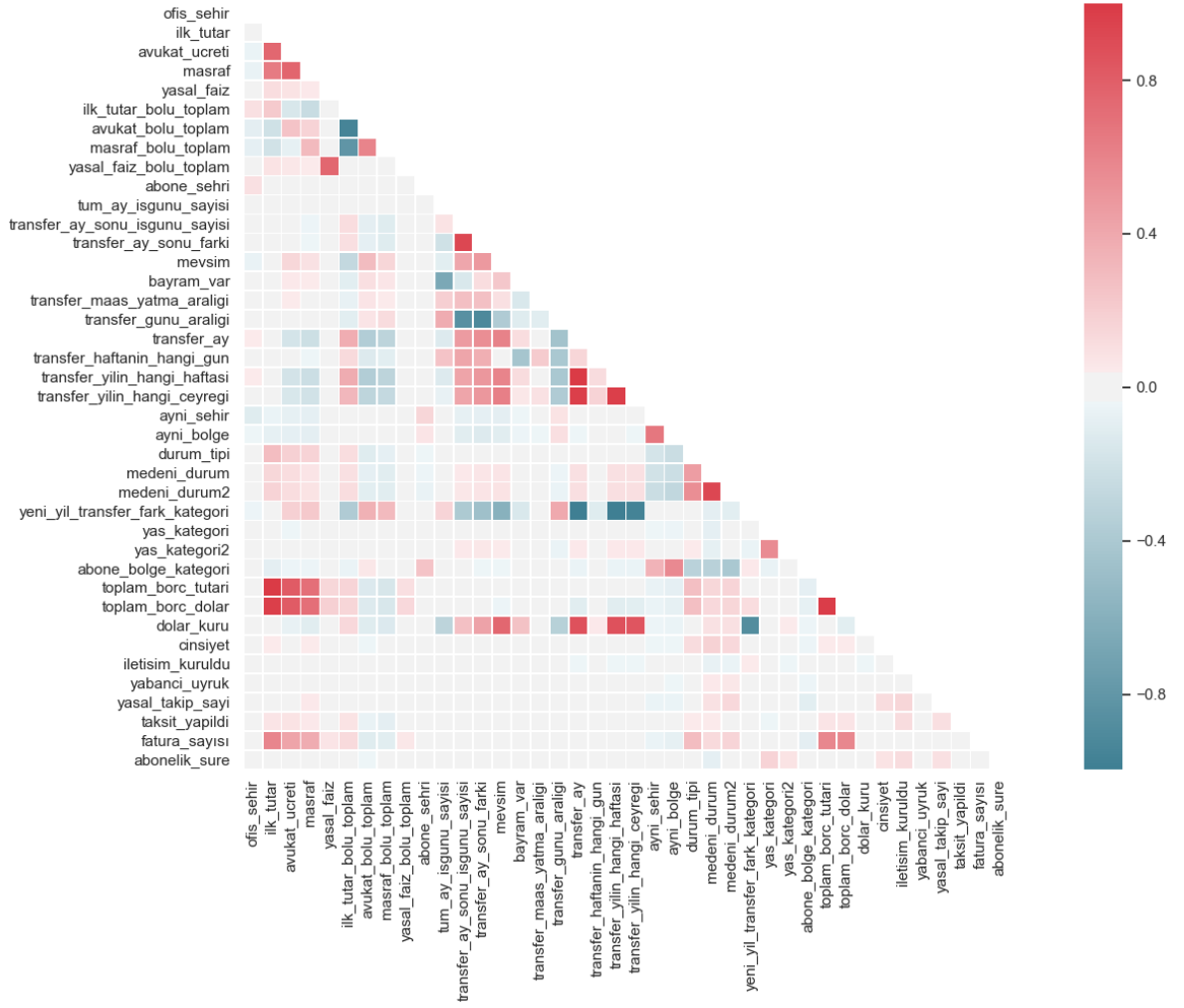
Veriyi iyi anlamak ve iyi analiz yapmış olmak, sonuçları yorumlamakta ve hangi modelin kullanılmasına karar vermekte önemlidir. Eklenen yeni öznitelikler ile birlikte veri kümesinin son hali Çizelge 4.9’da verilmiştir.

Çizelge 4.9 : Veri kümesi öznitelikleri ve açıklamaları.

Öznitelikler	Boş değer	Sınıf	Öznitelik açıklaması
ofis_sehir	yok	kategorik	büronun bulunduğu şehrin plaka kodu
ilk_tutar	yok	normalize	müşterinin telekomünikasyon şirketine olan borcu
avukat_ucreti	yok	normalize	müşterinin avukatlık bürosuna olan borcu
masraf	yok	normalize	takip başlatmanın bedeli
yasal_faiz	yok	normalize	gecikmeden dolayı yasal olarak belirlenmiş faiz miktarı
ilk_tutar_bolu_toplam	yok	-	müşterinin telekomünikasyon şirketine olan borcunun toplam borcuna oranı
avukat_bolu_toplam	yok	-	müşterinin avukatlık bürosuna olan borcunun toplam borcuna oranı
masraf_bolu_toplam	yok	-	takip başlatma bedelinin toplam borca oranı
yasal_faiz_bolu_toplam	yok	-	gecikmeden dolayı yasal olarak belirlenmiş faiz miktarının tüm borca oranı
abone_sehri	var	kategorik	müşterinin yaşadığı şehir
tum_ay_isgunu_sayisi	yok	normalize	dosyanın atandığı aydaki iş günü sayısı
transfer_ay_sonu_isgunu_sayisi	yok	normalize	dosya atandıktan sonra ay sonuna kadar iş günü sayısı
transfer_ay_sonu_farki	yok	normalize	dosya transfer edildikten sonra ay sonuna olan toplam gün sayısı
mevsim	yok	kategorik	dosyanın atandığı mevsim bilgisi(1-4)
bayram_var	yok	kategorik	dosyanın atandığı ay için dini bayram olup olmadığı bilgisi
transfer_maas_yatma_araligi	yok	kategorik	dosya atandığında maaş alım dönemi olan ay başı ay sonu bilgisi
transfer_gunu_araligi	yok	kategorik	dosyanın transferi ayın hangi döneminde yapıldığı 0-10 -> 1; 11-20->2; 21-31->3
transfer_ay	yok	kategorik	dosyanın transfer edildiği ay (0-12)
transfer_haftanin_hangi_gun	yok	kategorik	transfer haftanın kaçınıcı günü (0-7)
transfer_yilin_hangi_haftasi	yok	kategorik	dosya transferi yılın kaçınıcı haftası bilgisi (0-52)
transfer_yilin_hangi_ceyregi	yok	kategorik	transfer yılın hangi çeyreğinde gerçekleşti bilgisi(1-4)

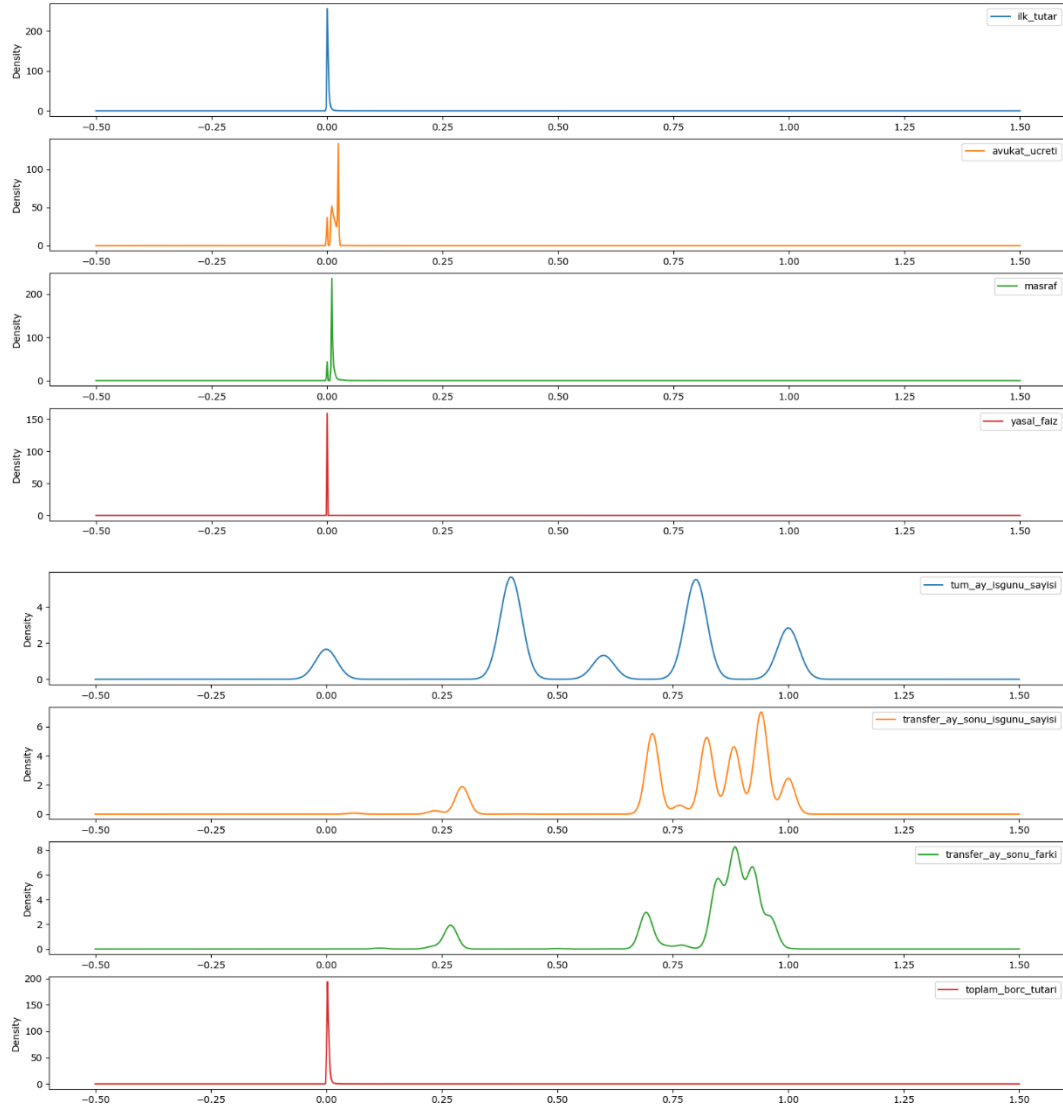
ayni_sehir	yok	kategorik	avukatlık bürosu ve müşteri aynı şehirde olduğu bilgisi(0-1)
ayni_bolge	yok	kategorik	avukatlık bürosu ve müşteri aynı bölgede olduğu bilgisi(0-1)
durum_tipi	yok	kategorik	kurumsal müşteri/bireysel müşteri tipi kategorize
medeni_durum	var	kategorik	müşterinin medeni durum bilgisi
medeni_durum2	var	kategorik	müşterinin medeni durum bilgisi (farklı kategorize)
yeni_yil_transfer_fark_kategori	yok	kategorik	dosya avukatlık bürosundan alınana kadarki geçen süre
yas_kategori	var	kategorik	müşterinin yaş bilgisinin kategorize hali
yas_kategori2	var	kategorik	müşterinin yaş bilgisinin farklı kategorize hali
musteri_bolge_kategori	yok	kategorik	müşterinin yaşadığı şehrin bulunduğu bölge
toplam_borc_tutari	yok	normalize	müşterinin tüm borç çeşitlerinin toplamı
toplam_borc_dolar	yok	normalize	müşterinin borcunun yabancı para cinsinden değeri
dolar_kuru	yok	normalize	dosyanın transfer edildiği aydaki kur bilgisi
cinsiyet	yok	kategorik	müşterinin cinsiyeti
iletisim_kuruldu	yok	kategorik	müşteri ile telekomünikasyon şirketi iletişim kurduğu bilgisi
yabanci_uyruk	yok	kategorik	müşterinin TC vatandaşı olup olmadığı bilgisi
yasal_takip_sayi	yok	normalize	müşterinin hayatı boyunca kaç kez yasal takibe girdiği
taksit_yapildi	yok	kategorik	müşterinin borcunun taksitlendirilme bilgisi
fatura_sayisi	var	normalize	müşterinin telekomünikasyon ödediği fatura sayısı
abonelik_sure	var	normalize	müşterinin telekomünikasyon şirketine müşterilik süresi
durum_info	yok	kategorik	müşterinin borcu ödeyip ödemediği bilgisi

Öznitelikler arasındaki ilişkinin görülebilmesi için, öznitelikler arası korelasyon matrisi oluşturulmuştur. Matrisin sayısal verileri Ek B’de yer almaktadır. Ancak görsel olarak renklendirilmiş hali, ilişkileri daha kolay görmeyi sağlayacağı için renklendirilmiş haline Şekil 4.21’de yer verilmiştir.

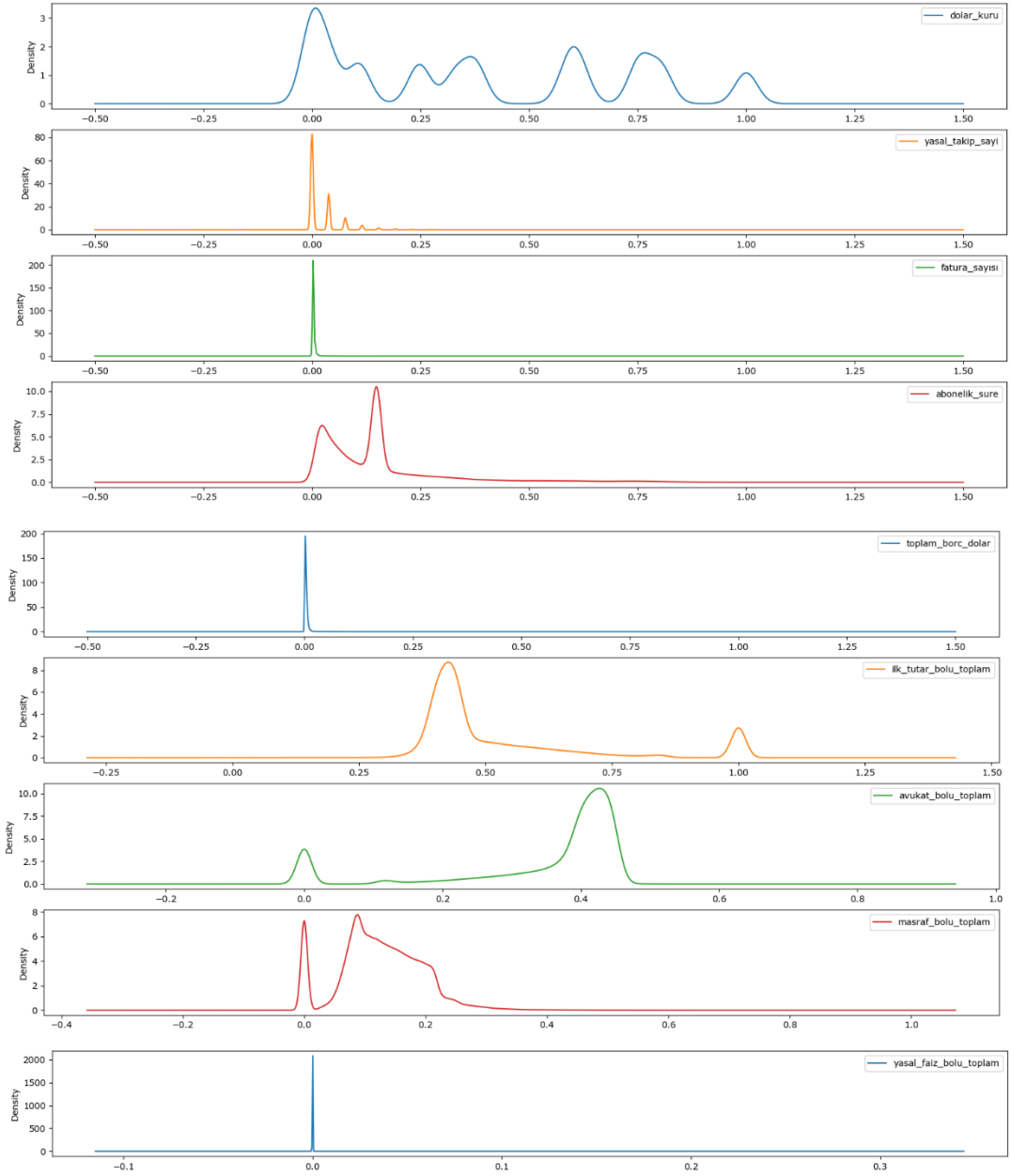


Şekil 4.21 : Korelasyon matrisi.

Normalize edilmiş verilerin, dağılımları Şekil 4.22'deki ve Şekil 4.23'teki grafiklerle gösterilmiştir. Bunlar; ilk tutar (şirketin alacağı tutar), avukat ücreti, masraf, yasal faiz, atama yapıldığı aydaki toplam iş günü sayısı, atama yapıldıktan sonra ay sonuna kadar olan iş günü sayısı, atama yapıldıktan sonra ay sonuna kadarki toplam gün sayısı, toplam borç tutarı, o aydaki dolar kuru, müşterinin yasal takibe girme sayısı, ödediği toplam fatura sayısı, müşterilik süresi, müşterinin toplam borcunun dolar cinsinden karşılığıdır. Bu değerlere ilave olarak normalize edilmesine gerek kalmamış değerlerin dağılımına da aşağıda yer verilmiştir. Bu değerler; şirketin alacağının toplam borca oranı, avukatın alacağının toplam borca oranı, masrafların toplam borca oranı, yasal faiz tutarının toplam borca oranıdır.



Şekil 4.22 : Veri dağılımı I.



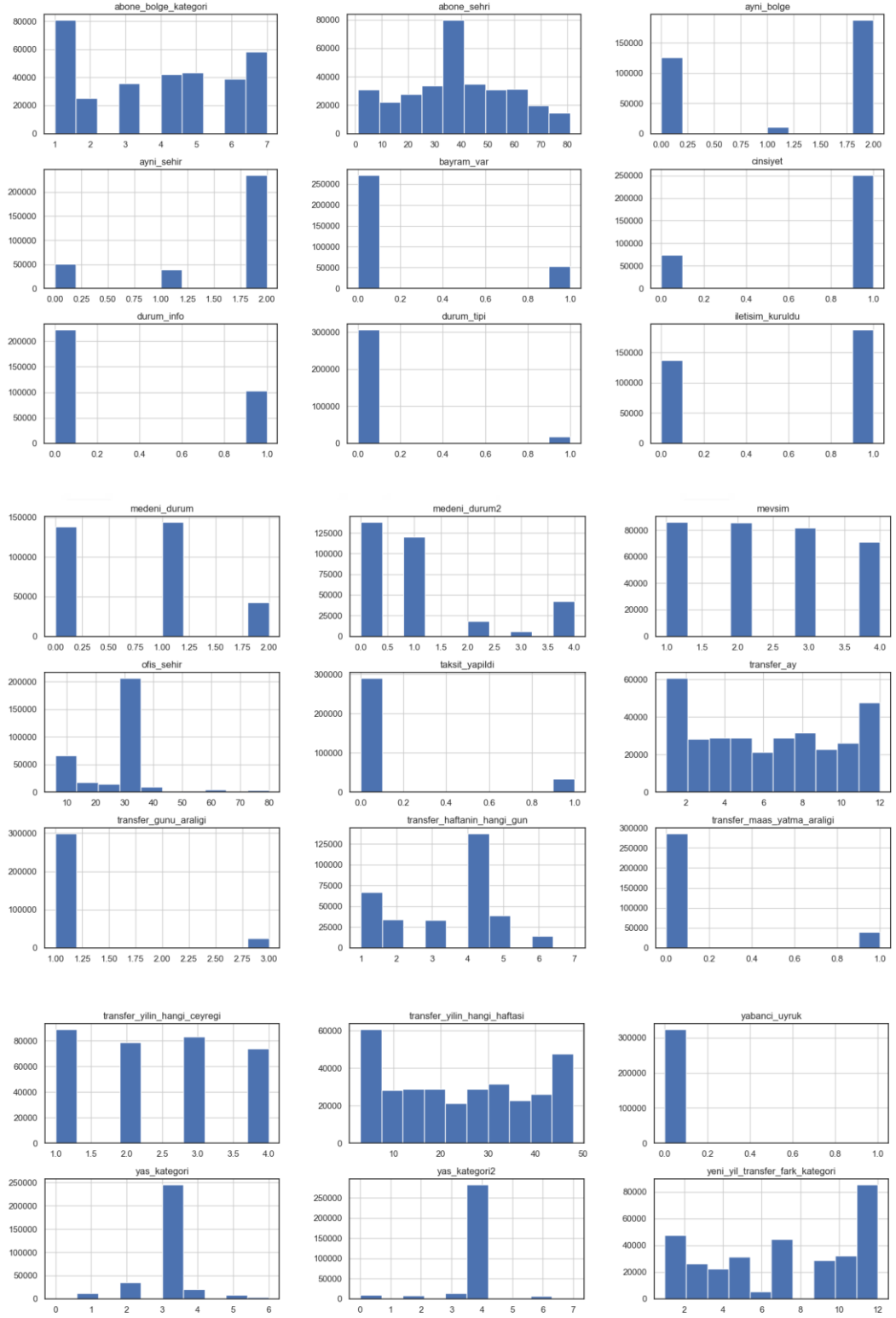
Şekil 4.23 : Veri dağılımı II.

Kategorik özniteliklere ait bar grafikleri Şekil 4.24’de verilmiştir. Bunlar; müşterinin yaşadığı coğrafik bölge (1-7), müşterinin yaşadığı şehir (1-81), müşteri ve büro aynı bölgede (0, 1, 2(bilinmiyor)), müşteri ve büro aynı şehirde (0, 1, 2(bilinmiyor)), atama yapıldığı ayda bayram olup olmadığı, müşterinin cinsiyeti, hedef öznitelik (durum_info), müşterinin kurumsal müşteri mi bireysel müşteri mi olduğu, müşteriye ulaşılabilip ulaşılamadığı bilgisi, müşterinin medeni durumu (0, 1, 2(bilinmiyor)), müşterinin medeni durumu (0, 1, 4(bilinmiyor)), mevsim bilgisi(1-4), büronun bulunduğu şehir bilgisi (1-81), borcun taksitlendirilip taksitlendirilmediği bilgisi,

atamanın yapıldığı ay (1-12), atamanın yapıldığı ayın hangi aralığında (1-10; 11-20; 21-31), atama haftanın hangi günü (1-7), atama maaş yattığı dönemde mi gerçekleşmiş, atama yılın hangi çeyreğinde (1-4)

atama yılın hangi haftasında (1-52), müşteri yabancı uyruklu bilgisi, müşterinin yaşı kategorik olarak, müşterinin yaşı2 kategorik olarak, atama ile yeni yıl arasındaki gün farkı kategorik (0-30 1; 30-60 2; 60-90 3;..)





Şekil 4.24 : Kategorik özniteliklerin dağılımları.

4.4 Öznitelik Seçim Metriklerine Göre Öznitelik Altkümesinin Belirlenmesi

Öznitelik seçme işlemi, makine öğrenmesi sürecinde eğitim işleminin bir parçası olduğu için, elde edilen öznitelik alt kümesi eğitilen modele özel olmaktadır. Bir model için uygun olan alt küme, başka bir model için uygun olmayabilmektedir. Bu nedenle, literatürde sıkça kullanılmış olan birkaç farklı öznitelik seçme metodolojisi karşılaştırılacaktır.

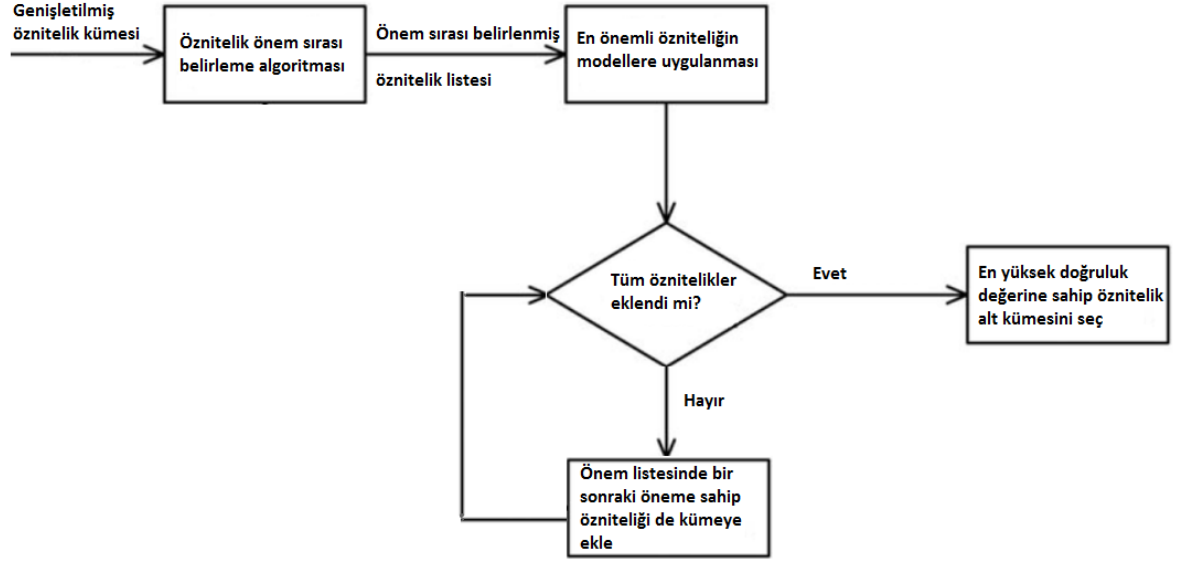
Veri kümesi üzerinde uygulanmış olan öznitelik seçim yöntemlerinden elde edilen listeler, önem sıraları ile öznitelik sayıları sırasıyla en önemli olandan başlayarak ve birer birer artırılabilecektir. Geri beslemeli eleme yönteminde minimum küme 5 olarak verildiğinden, aşağıdaki özyinelemeli model sonuç çizelgeleri, 1 değerine sahip 5'er tane özniteliğe sahip olduklarından bunlara ait minimum alt küme sayısı 5'ten başlanarak artırılabilecektir.

Öznitelik altkümeleri modellerin eğitiminde uygulanarak, test verisi üzerinde doğruluk (accuracy) ve F-ölçütü analizi ile her yöntem için tüm modeller için en iyi sonucu veren alt küme incelenmiştir.

4.5 Öznitelik Seçimi ve Öznitelik Seçiminde Kullanılan Yöntemler

Sınıflandırma, veri kümesini oluşturan özniteliklerin içerdikleri bilgiler ve birbirleri ile olan ilişkileri göz önünde tutularak yapılır. Fakat veri kümeleri ile gerçek hayat birebir benzetilemeyeceği için her zaman doğru ve gerekli bilgileri içermeyebilir. Öznitelikler arasından bağımlı ve gereksiz özellikleri ayırarak modelimizi en iyi açıklayan alt küme ya da alt kümeler seçilmesi önemlidir. Çünkü gereksiz ve fazla özniteliğe sahip veri kümesi öğrenme algoritmalarında karmaşıklığa ve süre kaybına sebep olur.

Öznitelik seçimi modelin doğru sonuç vermesi ve modelin performansı için önemli bir faktördür. Gereksiz verilerin veri kümesinden ayıklanması, modelin karar verirken bu artık verileri dikkate alarak karar vermesini önlemeyi sağlar. Daha az veri ile model daha hızlı çalışır. Ancak, öncesinde verilerin öznitelik seçimi işlemi için harcanacak süre ile programın toplam çalışma süresinde artış olabilir. Fakat uzun vadede daha doğru sonuçlar elde edilmesini sağlayacaktır. Şekil 4.25'te öznitelik seçimi görselleştirilmiştir.



Şekil 4.25 : Öznitelik seçimi.

4.5.1 Öznitelik seçimi ve sıralanması

Veri kümesinde öznitelikler arasında oluşan yüksek korelasyon değerleri, modelin gereksiz bilgiye sahip olduğunu ve modelin aşırı öğrenme (over-fitting) gösterme ihtimalinin olduğunu göstermektedir.

Boyut küçültmenin en kolay yolu veriyi en iyi tanımlayan öznitelikleri kullanmak, geri kalanını öznitelik altkümesinden temizlemektir. Bilgi kaybına neden olacak özniteliklerin kümeden çıkartılması, modelin performansı açısından önemli bir husustur. Öznitelik seçmek için kullanılan bazı teknikler vardır. Bu çalışmada bu tekniklerden elde edilen sonuçlar karşılaştırılacaktır. Sonrasında öznitelikler önem sıralarına göre sırasıyla modellerde eğitilerek elde edilen doğruluk değerleri yardımıyla optimum öznitelik alt kümesi belirlenecektir.

Bölüm 4.2.3'te bahsedilen yaş ve medeni durum kategorik öznitelikleri, iki farklı gruplama yöntemiyle incelenmek istenmiştir. Eklenen kategorik öznitelikler ile ilgili olarak yukarıda anlatıldığı üzere iki farklı yaş kategorisi ve medeni durum bilgisi incelemeye dahil edilmiştir. Bunlardan yüksek öneme sahip olan değer modelin eğitilmesi için incelemeye alınıp, diğer değer öznitelik alt kümeden çıkartılmıştır. Çizelgelerde de daha düşük öneme sahip olan diğer özniteliklerin üzeri çizili olarak gösterilmiştir.

4.5.2 Ağaç temelli yöntemlerden elde edilen öznitelik önem sırası

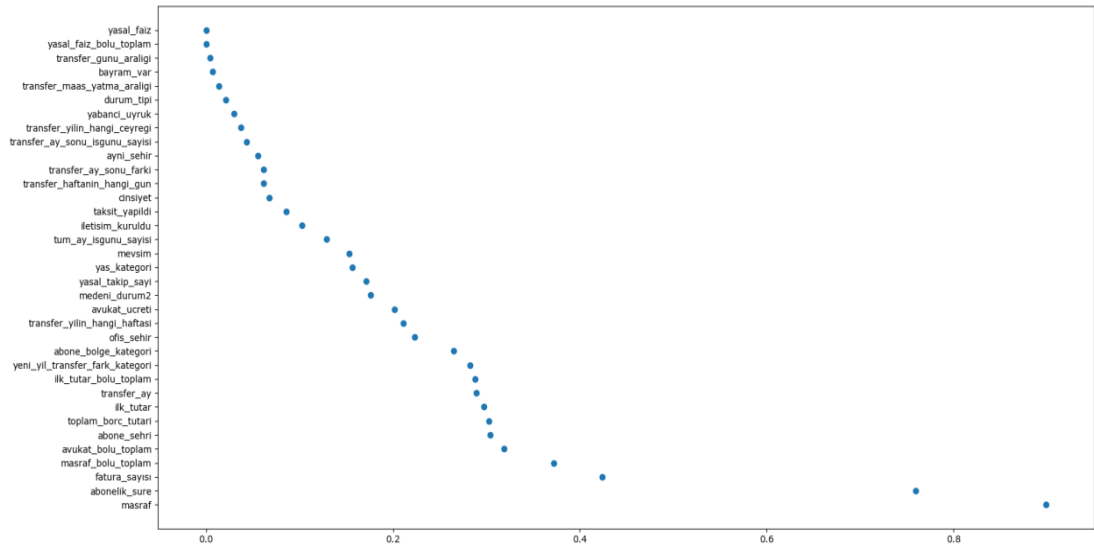
Ağaç temelli yöntemler özniteliklerin önem derecesini belirlemek için Gini önemi olarak da bilinen değeri kullanırlar. Ağaç temelli yöntemler, modelin eğitilmesi sırasında, özniteliklerin önem derecelerini de hesaplar. Böylece, kök özniteliğe karar verirler, kökten sonraki düğümlerde kıyaslanacak veri kırılım sırasını belirlemiş olurlar. Öğrenme sırasında öznitelik önem sırasını çıkartan ağaç algoritmaları üzerinde veri kümesi çalıştırılmıştır. Bu algoritmalar; karar ağaçları, rastgele ormanlar, extra gradyen boosting, gradyen boosting, adaboost ve extra trees'dir (Ek A). Ancak bu önem sıraları arasında küçük farklılıklar olduğu gözlemlenmiştir. Öznitelik sıralamalarında, öznitelikler bazen aynı sırada bazen de sıra olarak birkaç öznitelik daha önde veya daha arkada olduğu sonuç tablosundan incelenmiş olup, bu farklılığın giderilmesi için ağaç modellerinin ortalama değeri kullanılmıştır. Ancak modellerin ürettiği sonuçlar farkı skalalarda olduğundan öncelikle tüm sonuç listeleri kendi içerisinde [0,1] aralığına denk düşecek şekilde normalize edilmiştir (Çizelge A.2). Daha sonrasında ortalama değeri alınmıştır. Değerler Çizelge 4.10'daki gibidir.

Çizelge 4.10 : Ağaç temelli yöntemler ile öznitelik sıralaması.

Ağaç temelli yöntemlere göre öznitelik sıralaması	
0.9488592545191447	masraf
0.7614177898984708	abonelik_sure
0.43687167863765214	fatura_sayisi
0.39016333541757503	masraf_bolu_toplam
0.31788355178877364	toplam_borc_tutari
0.3160666865443594	avukat_bolu_toplam
0.31094619985744365	toplam_borc_dolar
0.3079128430339614	yeni_yil_transfer_fark_kategori
0.3063294103192921	abone_sehri
0.2666454595130235	ilk_tutar
0.26651307710781647	ilk_tutar_bolu_toplam
0.23237586581220246	transfer_ay
0.23031215483620246	musteri_bolge_kategori
0.21822173207375636	ofis_sehir
0.19810135798860307	avukat_ucreti
0.17956840282840916	transfer_yilin_hangi_haftasi
0.17722518893417663	medeni_durum2
0.17698479602578457	yasal_takip_sayi
0.16442713863246824	yas_kategori
0.12105706160154577	tum_ay_isgunu_sayisi
0.10867777510573993	medeni_durum
0.10750304543289087	mevsim
0.09286527077833971	dolar_kuru
0.08844574313968802	taksit_yapildi

0.08161101987398578	iletisim_kuruldu
0.0731196003422117	yas_kategori2
0.07184750032848551	cinsiyet
0.06879885084186153	transfer_yilin_hangi_ceyregi
0.06654275359001004	transfer_haftanin_hangi_gun
0.06566158151833677	transfer_ay_sonu_isgunu_sayisi
0.06384533293794865	transfer_ay_sonu_farki
0.0576200680084454	ayni_bolge
0.050897739048709444	ayni_sehir
0.02989547368705101	yabanci_uyruk
0.012816951747716854	durum_tipi
0.012764023858493451	transfer_maas_yatma_araligi
0.006039066991498689	transfer_gunu_araligi
0.005201749034044731	bayram_var
4.072549604149126e-06	yasal_faiz
3.6285305025146098e-06	yasal_faiz_bolu_toplam

Yukarıdaki çizelgede yer alan değerler arasındaki kırılımların daha iyi anlaşılabilmesi için öznitelik önem sırasına göre çizdirilmiş hali Şekil 4.26'da gösterilmiştir.



Şekil 4.26 : Ağaç temelli yöntemler ile öznitelik sıralaması.

4.5.3 Temel bileşen analizine göre öznitelik önem sırası

Genellikle veri azaltma yöntemi olarak da anılır. Sonuç listesinde yer alacak olan öznitelik sayısını belirleme imkanı sunar.

Bu analiz eldeki veriyi daha az sayıda öznitelikle ifade edebilecek en iyi dönüşümü belirlemeyi amaçlar. Dönüşüm sonrasında elde edilen öznitelik kümesi tüm özniteliklerin temel bileşenleri olarak adlandırılır. İlk temel bileşen varyans değeri en büyük olandır ve diğer temel bileşenler varyans değerleri azalacak şekilde sıralanır.

Diğer bir ifadeyle ilk açıklanan varyans değeri, birinci özdeğer olarak tanımlanıp λ_1 ikinci yükleme λ_2 ve diğer takip eden yüklemeler $\lambda_1 > \lambda_2 > \lambda_3 > \lambda_4 > \dots > \lambda_k$, sırasını bozmayacak şekilde büyükten küçüğe doğru sıralanırlar (Bordi ve Sutura, 2001).

TBA hesaplamalarının son önemli adımı ise döndürme (rotasyon) sürecidir. Elde edilen TBA sonuçları, hesaplanan değerlerin iç yapısı korunacak şekilde döndürülürler. Böylece, tüm öznitelikler birbirine dik bir şekilde uzaya yerleştirilmiş olur (Vicente-Serrano ve diğ., 2004).

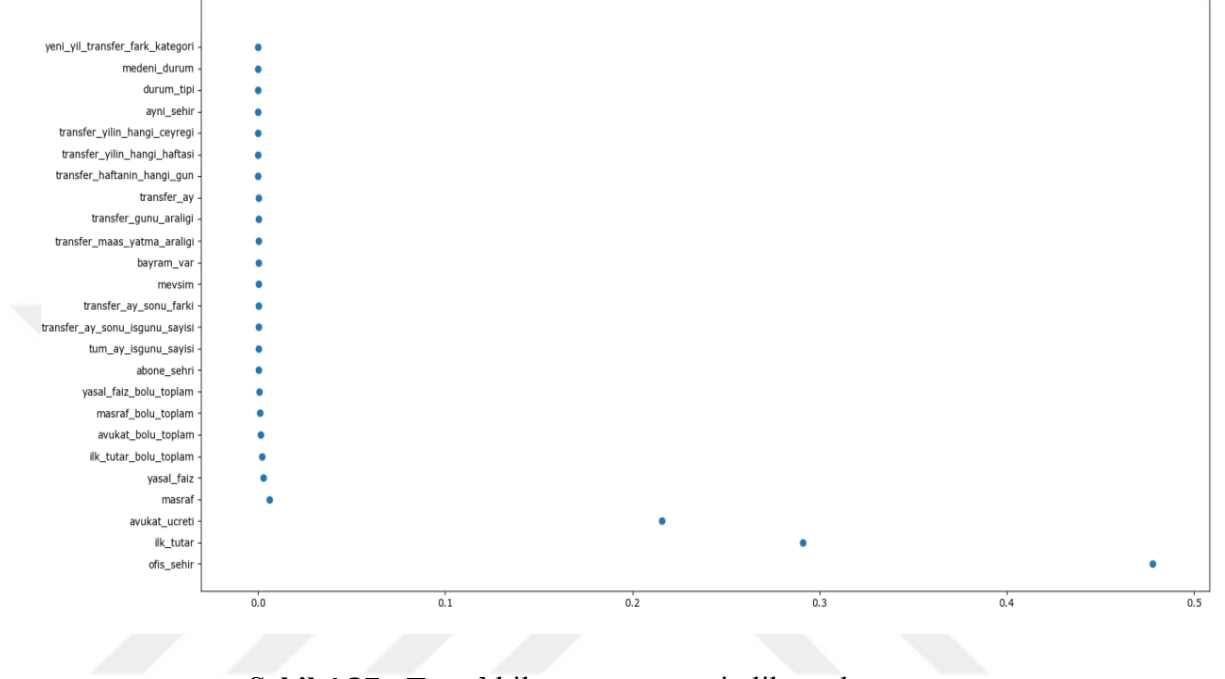
Analiz sonucunun öznitelikler için etki derecelerine göre sıralanmış listesi Çizelge 4.11 verilmiştir.

Çizelge 4.11 : Temel bileşene göre öznitelik seçimi.

Temel Bileşen Analizi'ne göre öznitelik sıralaması	
0.4771416638014263	ofis_sehir
0.2906977561365517	ilk_tutar
0.2154902811293996	avukat_ucreti
0.0066231560196557335	masraf
0.0026909472545090067	yasal_faiz
0.0019812864063467397	ilk_tutar_bolu_toplam
0.0011850718723056845	avukat_bolu_toplam
0.0011401256353485414	masraf_bolu_toplam
0.0008476844047216579	yasal_faiz_bolu_toplam
0.00034073270734418936	abone_sehri
0.00032494802717770673	tum_ay_isgunu_sayisi
0.000292249759452048	transfer_ay_sonu_isgunu_sayisi
0.0002507510351075988	transfer_ay_sonu_farki
0.00022529793558605056	mevsim
0.00020741389279308232	bayram_var
0.00011635581825523457	transfer_maas_yatma_araligi
0.00011100211020815938	transfer_gunu_araligi
8.124913588387743e-05	transfer_ay
6.083820435853466e-05	transfer_haftanin_hangi_gun
5.6403114994049393e-05	transfer_yilin_hangi_haftasi
4.020980794149739e-05	transfer_yilin_hangi_ceyregi
3.1294231724578906e-05	ayni_sehir
2.5678914220554373e-05	ayni_bolge
2.2192952951503678e-05	durum_tipi
6.593479726520078e-06	medeni_durum
3.5657439960157755e-06	medeni_durum2
1.6562980237781036e-06	yeni_yil_transfer_fark_kategori
1.5536083558934986e-06	yas_kategori
1.2376490870278459e-06	yas_kategori2
4.6996115631127933e-07	musteri_bolge_kategori
2.0164761463331874e-07	toplam_borc_tutari
7.259534095251354e-08	toplam_borc_dolar

3.071392077258841e-08	dolar_kuru
1.2751627659549734e-08	cinsiyet
9.424166594858545e-09	iletisim_kuruldu

Yukarıdaki çizelgede yer alan değerler arasındaki kırılımların daha iyi anlaşılabilmesi için öznitelik önem sırasına göre çizdirilmiş hali Şekil 4.27’da gösterilmiştir.



Şekil 4.27 : Temel bileşene göre öznitelik sıralaması.

4.5.4 Özyinelemeli öznitelik sıralaması

Öznitelikler, özyinelemeli olarak öznitelik altkümesinden çıkartılarak hangi öznitelik altkümesinin hedef değişkeninin doğru tahmin edilmesine en fazla katkıda bulunduğu belirler. Bu yöntemin kullanılmasında, başka bir modelden yardım alınmaktadır. Bu noktada, bu model hem lojistik regresyon hem de adaboost model uygulanmıştır. Model çıktısının en az önem derecesine sahip olanından en etkili özniteliklere göre sıralanması aşağıdaki şekildedir. Özyinelemeli modellerde minimum öznitelik sayısı belirtilmelidir. Bu sayı 5 olarak verilmiştir. 1 değerine sahip olan 5 öznitelik, hedef değişkenin belirlenmesinde en etkili olan öznitelikler Çizelge 4.12’de listelenmiştir. Sonuçlar değer olarak değil önem sırası olduğundan dolayı, öznitelik önem grafiği çizdirilmemiştir.

Çizelge 4.12 : Lojistik regresyona göre özyinelemeli öznitelik sıralaması.

Lojistik Regresyon’a göre özyinelemeli öznitelik sıralaması	
36	ofis_sehir

35	abone_sehri
34	musteri_bolge_kategori
33	ayni_sehir
32	ayni_bolge
31	yas_kategori2
30	cinsiyet
29	ilk_tutar_bolu_toplam
28	yasal_faiz
27	iletisim_kuruldu
26	yas_kategori
25	transfer_ay_sonu_igunu_sayisi
24	transfer_haftanin_hangi_gun
23	yasal_faiz_bolu_toplam
22	transfer_maas_yatma_araligi
21	transfer_yilin_hangi_haftasi
20	yeni_yil_transfer_fark_kategori
19	durum_tipi
18	medeni_durum2
17	medeni_durum
16	taksit_yapildi
15	toplam_borc_dolar
14	bayram_var
13	transfer_ay_sonu_farki
12	tum_ay_igunu_sayisi
11	transfer_gunu_araligi
10	toplam_borc_tutari
9	transfer_yilin_hangi_ceyregi
8	dolar_kuru
7	transfer_ay
6	masraf_bolu_toplam
5	mevsim
4	abonelik_sure
3	yasal_takip_sayi
2	yabanci_uyruk
1	masraf
1	ilk_tutar
1	fatura_sayisi
1	avukat_ucreti
1	avukat_bolu_toplam

Özyinelemeli model, bu sefer de Adaboost yöntem kullanılarak öznitelik önem listesi oluşturulmuştur. Liste Çizelge 4.13'te verilmiştir.

Çizelge 4.13 : Adaboost'a göre özyinelemeli öznitelik sıralaması.

Adaboost'a göre özyinelemeli öznitelik sıralaması	
36	transfer_yilin_hangi_haftasi
35	transfer_yilin_hangi_ceyregi

34	mevsim
33	ayni_sehir
32	toplam_borc_dolar
31	yasal_faiz_bolu_toplam
30	bayram_var
29	ayni_bolge
28	cinsiyet
27	yas_kategori2
26	iletisim_kuruldu
25	transfer_ay
24	transfer_gunu_araligi
23	tum_ay_isgunu_sayisi
22	medeni_durum
21	durum_tipi
20	transfer_maas_yatma_araligi
19	avukat_ucreti
18	yasal_faiz
17	ilk_tutar_bolu_toplam
16	ilk_tutar
15	taksit_yapildi
14	yasal_takip_sayi
13	yabanci_uyruk
12	avukat_bolu_toplam
11	abone_sehri
10	transfer_ay_sonu_isgunu_sayisi
9	transfer_ay_sonu_farki
8	transfer_haftanin_hangi_gun
7	ofis_sehir
6	yas_kategori
5	masraf_bolu_toplam
4	dolar_kuru
3	medeni_durum2
2	musteri_bolge_kategori
1	yeni_yil_transfer_fark_kategori
1	toplam_borc_tutari
1	masraf
1	fatura_sayisi
1	abonelik_sure

4.5.5 Tek değişkenli seçim

Bazı istatistiksel testler ile öznitelik seçimine dayanır. ‘*SelectKBest*’ yöntemi ile ki-kare yöntemlerine göre farklı alt kümeler oluşturulmuştur.

Her bir öznitelik için hesaplanan ki-kare değerleri büyükten küçüğe, yani önemliden az önemliye doğru sıralanır. Ki-kare değeri, gözlenen frekanslar ile beklenen frekanslar arasındaki farkın istatistik olarak anlamlı olup olmadığı temeline dayanır.

Özniteliklerin önem değerleri Çizelge 4.14'te, önem grafiği de Şekil 4.28'de gösterilmiştir.

Çizelge 4.14 : Ki-kare'ye göre öznitelik sıralaması.

Ki-kare'ye göre öznitelik sıralaması	
132256.01808469748	transfer_yilin_hangi_haftasi
26914.93338938051	transfer_ay
26693.424938004064	yeni_yil_transfer_fark_kategori
6609.305565673922	transfer_yilin_hangi_ceyregi
1906.51354132095	dolar_kuru
1787.2253410413855	medeni_durum2
667.8347933532617	abone_sehri
551.3708134855292	ilk_tutar_bolu_toplam
499.2065503568489	musteri_bolge_kategori
493.5450246576936	avukat_bolu_toplam
410.0952106138452	transfer_gunu_araligi
270.80063512482275	ofis_sehir
249.95234186386466	medeni_durum
244.3929564634716	durum_tipi
240.57226982941353	mevsim
229.2546274985449	transfer_haftanin_hangi_gun
195.1132680911374	abonelik_sure
186.06132309844855	taksit_yapildi
124.98083308821911	iletisim_kuruldu
121.75586291126723	masraf_bolu_toplam
95.49779271027231	transfer_maas_yatma_araligi
94.50101652815394	yabanci_uyruk
84.15967435420953	yas_kategori
83.389354590487	transfer_ay_sonu_farki
79.45961723934734	transfer_ay_sonu_isgunu_sayisi
74.63294311833054	ayni_bolge
73.24352084118324	yas_kategori2
34.7636733804705	ilk_tutar
26.512823040972133	fatura_sayisi
25.526423169017477	bayram_var
17.584368722625886	toplam_borc_tutari
13.887785619631376	ayni_sehir
12.283481605469955	yasal_takip_sayi
9.23722749737084	toplam_borc_dolar
7.797106174617111	tum_ay_isgunu_sayisi
4.072566298676131	masraf
2.3941055004497342	yasal_faiz
1.846189426251803	yasal_faiz_bolu_toplam
0.3831433456908301	avukat_ucreti
0.11105952938325729	cinsiyet



5. MÜŞTERİNİN ÖDEME TAHMİNİNİN BELİRLENMESİ UYGULAMASI

Çalışmanın bu bölümünde tüm veri kümesi üzerinde yapılan çalışmalardan sonra, öznitelik çıkarımı ve öznitelik seçimi aşamalarından elde ettiğimiz veri kümeleri üzerinde karar ağaçları, en yakın komşu, naive bayes, lojistik regresyon, yapay sinir ağları ve topluluk yöntemleri gibi literatürde sıkça tercih edilen gözetimli öğrenme yöntemleri kullanarak müşterilerin ödeme konusundaki davranışları tahmin edilmeye çalışılacaktır. Bu bölümde, modeller tahmin başarılarındaki performansları açısından karşılaştırmalı olarak incelenecektir.

5.1 Sistem Tasarımı

Modelin eğitilmesi için genişletilmiş olan öznitelik alt kümesi ile veri ön işlemden geçirildikten sonra, öznitelikler eğer sürekli değer ise normalizasyon işlemleri, kategorik öznitelikler ise kodlama ile sisteme bildirilmiştir. Sonrasında veri kümelerinde en çok etkili olan özniteliğin değerlendirilmesi ve önem sırasının belirlenmesi için birkaç farklı yöntem uygulanmıştır. Elde edilen sonuçlardan en yüksek öneme sahip öznitelik, bu kümeye eklendikten sonra, sırasıyla kendisinden daha düşük etkiye sahip olanların birer birer eklenmesiyle tüm veri kümesi eklenene kadar veri kümesi alt kümeleri elde edilir. O alt kümede bulunan öznitelikler arasından kodlama (encoding) yapılması gerekenlere kodlama uygulanmış. Her alt küme kendi içerisinde eğitim verisi ve test verisi olmak üzere %80'e %20 oranıyla ayrılmıştır. Yöntemler, eğitim verisi üzerinde k-kat yöntemi uygulanarak çalıştırılmış ve doğruluk ve F-ölçüt değerleri test verisi üzerinde incelenmiştir. Uygulanan yöntemler: lojistik regresyon, naive bayes, karar ağaçları, k en yakın komşu, gradyan boosting, adaboost, bagging sınıflayıcı, rastgele ormanlar, YSA, XGB'dir. Müşterinin borcunu ödemesi üzerine tahmin gerçekleştirilmiştir.

5.2 Performans Analizi Metrikleri

Modelini sınıflandırma doğruluğu incelenirken önemli bazı kriterler ile değerlendirilir. Modelin doğru sınıflandırılmış gözlemlerin oranı, kesinlik ve hassaslık gibi ölçütlerine bakılır. Bu ölçütler hesaplanırken; doğru pozitif (dp), doğru negatif (dn), yanlış pozitif

(yp) ve yanlış negatif (yn) değerleri kullanılır. Doğru pozitif (dp): ödeyecek olan müşterileri öder olarak doğru sınıflandırılması Yanlış-pozitif (yp): ödemeyecek olan müşteriler öder olarak yanlış sınıflandırılması Doğru negatif (dn): ödemeyecek olan müşterileri ödemez olarak doğru sınıflandırılması Yanlış negatif (yn): ödeyecek müşterileri ödemez olarak yanlış sınıflandırılması durumlarıdır.

5.2.1 Kesinlik (Precision)

Doğru pozitif sayısının, doğru olarak tahminlenmiş gözlem sayısına oranı ile bulunur. Denklem 3.27 ile bulunur.

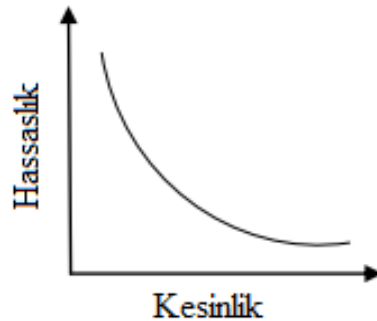
$$K = \frac{\text{doğru pozitif}}{\text{doğru pozitif} + \text{yanlış pozitif}} \quad (3.27)$$

5.2.2 Hassaslık (Recall)

Doğru pozitif sayısının, gerçekte doğru olan gözlem sayısına oranı ile bulunur. Denklem 3.28'deki gibi hesaplanır.

$$H = \frac{\text{doğru pozitif}}{\text{doğru pozitif} + \text{yanlış negatif}} \quad (3.28)$$

Kesinlik ve hassaslık arasındaki ilişki Şekil 5.1'de gösterilmiştir. Bir model, diğer modele oranla daha fazla, daha iyi kesinlik değerine sahip olabilir. Ancak başka bir model, bu modele oranla daha iyi duyarlılık değerine sahip olabilir. Bu durumda F ölçütüne bakılır. Ayrıca, bir model diğer modele kıyasla, daha iyi hassaslık ve kesinlik değerine sahipse diğerine göre daha iyi olduğu söylenebilir.



Şekil 5.1 : Hassaslık ve kesinlik ilişkisi grafiği.

5.2.3 Doğruluk (Accuracy)

Doğru tahminlerin tüm gözlem sayısına oranı doğruluk oranını verir. Modelin toplam yaptığı doğru tahmin oranına Doğru Sınıflandırma Oranı (Accuracy) denir. Buradaki doğruluk; borç tahsilatı yapılabilecek müşteriye, ödeme yapar diye belirlemek, borcunu ödemeyecek müşterinin de 0 yani ödemez şeklinde tahmin edilmeleri kastetmektedir.

Denklem 3.29’da verildiği gibi hesaplanır. Buradaki doğruluk; borç tahsil etmenin kolay olduğu müşterinin kolay, zor olduğu müşterinin de zor olarak tahmin edilmeleri kastedilmektedir.

$$D = \frac{\text{doğru pozitif} + \text{doğru negatif}}{\text{doğru poz.} + \text{doğru neg.} + \text{yanlış poz.} + \text{yanlış neg.}} \quad (3.29)$$

5.2.4 F ölçütü

Hassaslık ve kesinlik metriklerinin beraber incelenerek yapılan ölçüm F ölçütünün değerini verir. Bu ölçüm, kesinlik ve hassaslık değerlerinin harmonik ortalaması alınarak hesaplanır. Harmonik ortalama denklem 3.30’da verildiği gibi hesaplanır.

$$F = \frac{2 \times \text{Kesinlik} \times \text{Hassaslık}}{\text{Kesinlik} + \text{Hassaslık}} \quad (3.30)$$

5.2.5 ROC istatistiği

ROC eğrisi model başarısını değerlendirmede ve model karşılaştırmalarında kullanılan görsel bir tekniktir. ROC eğrisindeki x eksenini; yanlış-pozitif oranını, y eksenini ise doğru-pozitif oranını yani hassaslığı göstermektedir. Buradan da anlaşılacağı gibi, ROC eğrisi dp ile yp arasındaki değişimi gösterir ve altında kalan alan (AUC) modelin performansının ölçütü olarak kullanılmaktadır. Bu değer 0.5 ile 1 arasında yer almaktadır. Model performansının başarısı ROC eğrisi altındaki alanın 1’e yakınlığı ile değerlendirilir. Ayrıca modelin sağlamlığının bir ölçüsü olarak da bilinir. ROC eğrisinde kullanılan yanlış pozitif oranı aşağıdaki denklem ile bulunur.

$$\text{Yanlış pozitif oranı} = \frac{\text{yanlış pozitif}}{\text{doğru negatif} + \text{yanlış pozitif}} \quad (3.31)$$

ROC eğrisi üzerindeki her bir nokta bir kesim noktası olasılığını gösterir. Kesim noktası seçimi duyarlılık ve hassaslık katsayıları arasındaki bir değişimi gösterir. Dik eksendeki değer, doğru tespit edilen ödeyecek müşterilerin kendi içerisindeki yüzdesini verirken, yatay eksen de yanlış tespit edilen ödemeyecek müşterileri gösterir. İdeal olarak hem duyarlılık hem de hassaslık için bu değerlerin yüksek olması istenir; böylece model hem ödeyecekleri hem de ödemeyecekleri daha doğru tahmin edebilir. Düşük bir kesim noktası genel olarak daha fazla yanlış pozitif, yüksek bir kesim noktası da daha fazla yanlış negatif anlamına gelir.

5.2.6 Karmaşıklık matrisi (Confusion matrix)

Karmaşıklık matrisi, uygulanan algoritmanın yaptığı sınıflandırma sonucunda, algoritmanın yaptığı tahmin ve gerçek durum arasındaki bilgileri içeren bir gösterim tablosudur. Performans değerlendirilirken bu matristeki değerler göz önünde bulundurulur. Bu matrisin sütunları ve satırlarından biri gerçek durumu, öteki de tahminin sonucunu ifade eder. Aşağıda iki sınıflı bir problem için sınıflandırma sonucu oluşan karmaşıklık matrisi bulunmaktadır.

Çizelge 5.1'deki matris incelenirse; ilk satırdaki ilk kolon (doğru pozitif), borcunu ödememiş müşterilerin model tarafından ne kadarlık bir başarıyla tespit edildiğini ölçer. İkinci satırdaki ilk kolon (yanlış negatif) ile modelin gerçekte ödeyecek olan müşterilerin ödemez olarak tahmin ettiği müşteri sayısı yer almaktadır. İlk satır ikinci kolon, borcunu ödememiş ancak ödeyecek diye tahmin edilmiş müşterilerin model tarafından hesaplanmış sayısını verir. İkinci satır ikinci kolon ise, borcunu ödemiş müşterilerin model tarafından ne kadarının tahmin edildiğini ölçer.

Hesaplama yapılabilecek hatalar, iki tip olabilir. İlk olarak, borç tahsil etmesi zor olan müşteriyi kolay olarak hesaplamak (yanlış negatif), ikinci olarak borç tahsil etmesi kolay müşteriyi zor olarak hesaplamak (yanlış pozitif) olacaktır.

Çizelge 5.1 : Karmaşıklık matrisi.

Karmaşıklık matrisi	Ödemez (Tahmin)	Öder (Tahmin)
Ödemez	Doğru Negatif	Yanlış Pozitif
Öder	Yanlış Negatif	Doğru Pozitif

5.3 Sınıflandırıcı Modellerin Sonuçlarının Yorumlanması

Sınıflandırıcı seçiminde kullanılan yöntemler; eğitim verisini girdi olarak alır. Eğitim verisini kullanarak, model oluşturulduktan sonra, test verisi kullanılarak modelin ne kadar doğru öğrendiği sınanır. Çeşitli sınıflandırıcı algoritmaların model performansları incelenecektir. Bunlar: adaboost, bagging, karar ağacı, extra trees, gradyen boosting, en yakın komşu, lojistik regresyon, yapay sinir ağları, naive bayes, rastgele orman, son derece rastgele orman ve fazlasıyla gradyen boosting..

Aşağıdaki bölümlerde yer alan çizelgelerde, her bir model için sırasıyla en önemliden başlayarak özniteliklerin birer birer eklendiğindeki doğruluk hesaplarının çizelgeleri verilmiştir. Bu çizelgelerde o model için en yüksek doğruluk oranına sahip model (F-ölçütü ile de kıyaslanarak) koyu renkli olarak belirtilmiştir.

5.3.1 Ağaç temelli yöntemler ile oluşturulmuş alt kümeler

Lojistik regresyon modeli incelendiğinde; bir noktadan sonra yeni öznitelik eklemenin model üzerinde ciddi bir iyileşmeye neden olmadığı görülmüştür. 30 öznitelikten oluşan veri kümesi ile 38 öznitelikten oluşan veri kümesi arasında fark önemsenmeyebilir. Lojistik regresyon, eklenen her yeni öznitelik ile lineer fonksiyona bir yeni bilinmeyen daha eklediğinden, eklenen her öznitelik ile iyileşme gösterdiği görülmüştür. Değerler Çizelge 5.2’de verilmiştir.

Çizelge 5.2 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – Lojistik regresyon.

Lojistik Regresyon	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	70,1266	27,9122	0,71	0,95	0,62	0,18	0,72
altküme_37	70,122	27,8983	0,71	0,95	0,62	0,18	0,72
altküme_36	70,1204	27,908	0,71	0,95	0,61	0,18	0,72
altküme_35	70,1204	27,908	0,71	0,95	0,61	0,18	0,72
altküme_34	70,125	27,9004	0,71	0,95	0,62	0,18	0,72
altküme_33	70,1173	27,922	0,71	0,95	0,61	0,18	0,72
altküme_32	70,1266	27,9122	0,71	0,95	0,62	0,18	0,72
altküme_31	70,1158	27,8352	0,71	0,95	0,62	0,18	0,72
altküme_30	70,1281	27,8166	0,71	0,95	0,62	0,18	0,72
altküme_29	70,1035	27,7786	0,71	0,95	0,61	0,18	0,72
altküme_28	70,1081	27,7924	0,71	0,95	0,61	0,18	0,72
altküme_27	70,105	27,7474	0,71	0,95	0,61	0,18	0,72
altküme_26	70,0742	27,7106	0,71	0,95	0,61	0,18	0,72
altküme_25	70,0696	27,6699	0,71	0,95	0,61	0,18	0,72
altküme_24	70,0727	27,6881	0,71	0,95	0,61	0,18	0,72

altküme_23	70,1712	27,9101	0,71	0,95	0,62	0,18	0,72
altküme_22	69,9541	26,9149	0,71	0,95	0,61	0,17	0,71
altküme_21	69,9572	26,9278	0,71	0,95	0,61	0,17	0,71
altküme_20	69,9526	26,8974	0,71	0,95	0,61	0,17	0,71
altküme_19	69,9526	26,9084	0,71	0,95	0,61	0,17	0,71
altküme_18	69,814	25,8155	0,71	0,95	0,61	0,16	0,71
altküme_17	69,7355	25,3541	0,71	0,95	0,6	0,16	0,71
altküme_16	69,1426	20,1912	0,7	0,96	0,59	0,12	0,71
altküme_15	69,1442	20,1665	0,7	0,96	0,59	0,12	0,71
altküme_14	68,7962	16,2222	0,69	0,97	0,58	0,09	0,70
altküme_13	68,4621	13,8036	0,69	0,97	0,56	0,08	0,70
altküme_12	68,459	13,7879	0,69	0,97	0,56	0,08	0,70
altküme_11	68,4482	13,5437	0,69	0,97	0,55	0,08	0,70
altküme_10	68,4482	13,5437	0,69	0,97	0,55	0,08	0,70
altküme_9	68,4343	13,4729	0,69	0,97	0,55	0,08	0,70
altküme_8	67,9585	8,0672	0,68	0,98	0,5	0,04	0,70
altküme_7	67,7676	3,0476	0,68	0,99	0,42	0,02	0,65
altküme_6	67,763	3,0292	0,68	0,99	0,42	0,02	0,65
altküme_5	67,7769	1,645	0,68	0,99	0,37	0,01	0,61
altküme_4	67,7799	1,6359	0,68	0,99	0,38	0,01	0,61
altküme_3	67,9524	0,096	0,68	1	0,42	0	0,60
altküme_2	67,9355	0,0384	0,68	1	0,17	0	0,58
altküme_1	67,9585	0	0,68	1	0	0	0,58

Naive bayes yönteminde, az sayıdaki öznitelikler ile yüksek doğruluğa ulaşılabildiği Çizelge 5.3 ile görülmüştür. Bu durum çok şaşırtıcı olmasa da F ölçütünün de yüksek çıkmış olması ve ROC değerinin 2-3 öznitelik ile en iyi sonuca ulaşması sürpriz olmuştur. Ağaç öznitelik sıralamasıyla sıralanmış özniteliklerden, ilk sıradaki özniteliklerin tahmin üzerinde en önemli özneliği seçmiş olduğuna karar verilmiştir.

Çizelge 5.3 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – Naive Bayes sınıflayıcısı.

Naive Bayes	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	41,545	52,0045	0,96	0,15	0,35	0,99	0,61
altküme_37	41,9053	52,0445	0,95	0,15	0,35	0,98	0,63
altküme_36	42,7414	52,1846	0,94	0,17	0,36	0,98	0,64
altküme_35	42,7368	52,1826	0,94	0,17	0,36	0,98	0,64
altküme_34	42,6506	52,1709	0,94	0,17	0,36	0,98	0,64
altküme_33	42,6414	52,1657	0,94	0,17	0,36	0,98	0,64
altküme_32	42,6198	52,1649	0,94	0,17	0,36	0,98	0,64
altküme_31	42,666	52,169	0,94	0,17	0,36	0,98	0,64
altküme_30	42,5998	52,1414	0,94	0,17	0,36	0,98	0,64
altküme_29	42,4458	52,1151	0,94	0,16	0,36	0,98	0,64
altküme_28	42,4058	52,1026	0,94	0,16	0,36	0,98	0,64
altküme_27	42,3842	52,1019	0,94	0,16	0,36	0,98	0,64
altküme_26	42,2533	52,0575	0,94	0,16	0,35	0,98	0,64

altküme_25	42,0208	52,0368	0,95	0,16	0,35	0,98	0,64
altküme_24	42,0208	52,0368	0,95	0,16	0,35	0,98	0,64
altküme_23	42,0162	52,036	0,95	0,16	0,35	0,98	0,64
altküme_22	42,0116	52,0304	0,95	0,16	0,35	0,98	0,64
altküme_21	42,0008	52,0306	0,95	0,16	0,35	0,98	0,64
altküme_20	41,9377	52,0096	0,95	0,15	0,35	0,98	0,64
altküme_19	41,93	52,0063	0,95	0,15	0,35	0,98	0,64
altküme_18	41,836	52,0147	0,95	0,15	0,35	0,98	0,63
altküme_17	41,836	52,016	0,95	0,15	0,35	0,98	0,63
altküme_16	41,796	51,9866	0,95	0,15	0,35	0,98	0,63
altküme_15	44,9911	51,7679	0,86	0,23	0,36	0,92	0,63
altküme_14	45,1527	51,8675	0,86	0,23	0,36	0,92	0,63
altküme_13	41,7406	51,8344	0,94	0,15	0,35	0,98	0,64
altküme_12	41,6574	51,8551	0,94	0,15	0,35	0,98	0,64
altküme_11	43,3958	51,4597	0,87	0,2	0,35	0,94	0,63
altküme_10	43,439	51,7287	0,88	0,19	0,36	0,95	0,63
altküme_9	44,1165	51,8552	0,88	0,21	0,36	0,94	0,63
altküme_8	41,7329	52,029	0,96	0,15	0,35	0,99	0,65
altküme_7	44,4306	51,1935	0,84	0,22	0,36	0,91	0,62
altküme_6	45,0634	51,3407	0,84	0,24	0,36	0,9	0,62
altküme_5	45,1004	51,5978	0,85	0,23	0,36	0,91	0,63
altküme_4	47,33	52,2116	0,85	0,27	0,37	0,9	0,64
altküme_3	47,4316	52,3902	0,86	0,27	0,37	0,9	0,66
altküme_2	62,5077	51,7775	0,78	0,62	0,44	0,63	0,65
altküme_1	52,7824	52,0507	0,81	0,4	0,39	0,8	0,61

Karar ağacı modelinde, bir özniteliğe göre yapılan modelin doğruluğu yüksek olmasına rağmen F ölçütü düşüktür. Model ağırlıklı olarak, 0 değeri ataması yapmak üzerine yoğunlaşmıştır. Ödeme ihtimali olan müşteriler için de ödemez diye işaretlenmiştir. Model en iyi sonuca 27 öznitelik ile ulaşmıştır. Ayrıca, yeni öznitelikler eklendikçe, modelin bilgi kazanımı artacağı için 8 öznitelik sayısına ulaşana kadar sürekli bir artış gözlenmiş, ancak bu noktadan sonra eklenen yeni öznitelikler modelin performansı üzerinde ciddi bir etkiye sahip olamamıştır. Ağaç temelli öznitelik seçimiyle belirlenmiş altkümelerin karar ağacı modeline uygulanması modelin doğruluğu üzerinde beklenen etkiyi göstermemiştir.

Çizelge 5.4 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – Karar ağaçları sınıflayıcısı.

Karar Ağaçları	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	67,6013	49,2315	0,76	0,76	0,49	0,49	0,63
altküme_37	67,6213	49,3618	0,76	0,76	0,49	0,49	0,63
altküme_36	67,7122	49,4223	0,76	0,76	0,49	0,49	0,63
altküme_35	67,646	49,3223	0,76	0,76	0,49	0,49	0,63
altküme_34	67,6675	49,2606	0,76	0,76	0,49	0,49	0,63

altküme_33	67,7738	49,3502	0,77	0,76	0,49	0,49	0,63
altküme_32	67,7815	49,5296	0,76	0,76	0,49	0,5	0,63
altküme_31	67,8307	49,4728	0,77	0,76	0,49	0,49	0,63
altküme_30	67,9093	49,7916	0,77	0,76	0,5	0,5	0,63
altküme_29	67,9201	49,6933	0,77	0,76	0,49	0,5	0,63
altküme_28	67,8692	49,6076	0,77	0,76	0,49	0,5	0,63
altküme_27	68,1341	50,0519	0,77	0,77	0,5	0,5	0,63
altküme_26	67,9524	49,721	0,77	0,76	0,49	0,5	0,63
altküme_25	67,84	49,5775	0,77	0,76	0,49	0,5	0,63
altküme_24	67,9278	49,6702	0,77	0,76	0,49	0,5	0,63
altküme_23	67,9293	49,703	0,77	0,76	0,49	0,5	0,63
altküme_22	68,1187	49,7464	0,77	0,77	0,49	0,5	0,63
altküme_21	67,9508	49,637	0,77	0,76	0,49	0,5	0,63
altküme_20	67,706	49,5174	0,76	0,76	0,49	0,5	0,63
altküme_19	67,6952	49,458	0,76	0,76	0,49	0,49	0,63
altküme_18	67,7399	49,529	0,76	0,76	0,49	0,5	0,63
altküme_17	67,6691	49,3597	0,76	0,76	0,49	0,49	0,63
altküme_16	67,5166	49,1221	0,76	0,76	0,49	0,49	0,63
altküme_15	67,4735	49,0693	0,76	0,76	0,49	0,49	0,63
altküme_14	67,3919	48,9526	0,76	0,76	0,49	0,49	0,62
altküme_13	67,2903	48,5081	0,76	0,76	0,48	0,49	0,62
altküme_12	67,6444	48,8722	0,77	0,76	0,48	0,49	0,63
altküme_11	67,2456	48,2987	0,76	0,76	0,48	0,48	0,62
altküme_10	67,0886	48,0533	0,76	0,76	0,48	0,48	0,62
altküme_9	67,2841	48,3582	0,76	0,76	0,48	0,48	0,62
altküme_8	66,9592	48,2241	0,76	0,76	0,48	0,48	0,62
altküme_7	66,505	47,6928	0,75	0,75	0,48	0,48	0,62
altküme_6	64,7096	44,7589	0,74	0,74	0,45	0,45	0,59
altküme_5	64,5818	44,6109	0,74	0,74	0,45	0,45	0,59
altküme_4	65,0006	44,8649	0,75	0,74	0,44	0,45	0,59
altküme_3	63,7734	43,1948	0,74	0,73	0,43	0,43	0,59
altküme_2	63,4731	43,2949	0,73	0,73	0,44	0,43	0,58
altküme_1	68,2311	23,2897	0,93	0,8	0,15	0,23	0,65

K en yakın komşu yöntemiyle, yüksek doğruluk oranlarına çok hızlı bir şekilde 8 öznitelik ile ulaşılmıştır. Sonrasında modele öznitelik eklenmeye devam ettikçe, noktaların uzaydaki dağılımını değiştirdiği için doğrulukta düşüş meydana gelmiştir.

Çizelge 5.5 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – K en yakın komşu sınıflayıcısı.

En Yakın Komşu	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	64,4678	37,0162	0,71	0,79	0,43	0,33	0,62
altküme_37	64,4678	37,0162	0,71	0,79	0,43	0,33	0,62
altküme_36	64,4678	37,0162	0,71	0,79	0,43	0,33	0,62
altküme_35	64,4678	37,0128	0,71	0,8	0,43	0,33	0,62
altküme_34	64,4709	37,0148	0,71	0,8	0,43	0,33	0,62
altküme_33	64,4555	36,9875	0,71	0,79	0,43	0,33	0,62

altküme_32	64,4263	36,9821	0,71	0,79	0,43	0,33	0,62
altküme_31	64,4278	36,9865	0,71	0,79	0,43	0,33	0,62
altküme_30	64,4063	36,9415	0,71	0,79	0,43	0,33	0,62
altküme_29	64,4786	37,0473	0,71	0,79	0,43	0,33	0,62
altküme_28	64,4848	37,0617	0,71	0,8	0,43	0,33	0,62
altküme_27	64,4678	37,0368	0,71	0,79	0,43	0,33	0,62
altküme_26	64,8805	37,6967	0,72	0,8	0,44	0,33	0,62
altküme_25	64,8805	37,6967	0,72	0,8	0,44	0,33	0,62
altküme_24	65,1115	38,2683	0,72	0,8	0,44	0,34	0,63
altküme_23	65,4595	39,3271	0,72	0,8	0,45	0,35	0,63
altküme_22	65,6519	40,1652	0,73	0,8	0,45	0,36	0,64
altküme_21	65,6519	40,1652	0,73	0,8	0,45	0,36	0,64
altküme_20	65,6519	40,1652	0,73	0,8	0,45	0,36	0,64
altküme_19	65,655	40,1706	0,73	0,8	0,45	0,36	0,64
altküme_18	65,7089	40,1473	0,73	0,8	0,46	0,36	0,64
altküme_17	65,926	40,5278	0,73	0,8	0,46	0,36	0,64
altküme_16	66,0169	40,5954	0,73	0,8	0,46	0,36	0,64
altküme_15	66,023	40,587	0,73	0,8	0,46	0,36	0,64
altküme_14	65,943	40,3763	0,73	0,8	0,46	0,36	0,64
altküme_13	66,6205	41,7227	0,73	0,8	0,47	0,37	0,65
altküme_12	66,6251	41,726	0,73	0,8	0,47	0,37	0,65
altküme_11	66,7406	42,0414	0,73	0,8	0,48	0,38	0,65
altküme_10	66,9608	42,3524	0,73	0,81	0,48	0,38	0,65
altküme_9	66,9577	42,3409	0,73	0,81	0,48	0,38	0,65
altküme_8	68,6822	45,1823	0,74	0,82	0,51	0,4	0,68
altküme_7	68,9702	44,6374	0,74	0,83	0,52	0,39	0,67
altküme_6	68,1464	43,091	0,74	0,83	0,5	0,38	0,66
altküme_5	68,1556	43,2573	0,74	0,82	0,5	0,38	0,66
altküme_4	68,1187	42,9724	0,74	0,83	0,5	0,37	0,66
altküme_3	68,1449	43,4352	0,74	0,82	0,5	0,38	0,66
altküme_2	66,9869	40,0548	0,73	0,82	0,48	0,34	0,64
altküme_1	64,9744	35,9041	0,71	0,81	0,43	0,31	0,61

Gradyen boosting yöntemi ile 25 öznitelik ile en iyi sonuçlara ulaşılmıştır. Gradyen boosting model, boosting olduğundan dolayı, sonrasında yeni öznitelikler ile doğruluk oranında ciddi bir düşüş gözlenmemiştir. İyileşme de gözlenebilirdi ancak bu altküme sırası için artış gözlenmemiştir. Değerler Çizelge 5.6’da verilmiştir.

Çizelge 5.6 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – Gradyen boosting sınıflayıcısı.

Gradyen Boosting	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	74,1192	50,372	0,76	0,9	0,65	0,41	0,78
altküme_37	74,1192	50,372	0,76	0,9	0,65	0,41	0,78
altküme_36	74,1192	50,372	0,76	0,9	0,65	0,41	0,78
altküme_35	74,1192	50,372	0,76	0,9	0,65	0,41	0,78
altküme_34	74,1192	50,372	0,76	0,9	0,65	0,41	0,78

altküme_33	74,1192	50,372	0,76	0,9	0,65	0,41	0,78
altküme_32	74,1008	50,3776	0,76	0,9	0,65	0,41	0,78
altküme_31	74,1962	50,5722	0,76	0,9	0,65	0,41	0,78
altküme_30	74,1731	50,4885	0,76	0,9	0,65	0,41	0,78
altküme_29	74,1701	50,4885	0,76	0,9	0,65	0,41	0,78
altküme_28	74,1393	50,447	0,76	0,9	0,65	0,41	0,78
altküme_27	74,1362	50,444	0,76	0,9	0,65	0,41	0,78
altküme_26	74,1531	50,6265	0,76	0,9	0,65	0,41	0,78
altküme_25	74,2132	50,5565	0,76	0,9	0,66	0,41	0,78
altküme_24	74,1577	50,4941	0,76	0,9	0,65	0,41	0,78
altküme_23	74,1855	50,5297	0,76	0,9	0,65	0,41	0,78
altküme_22	74,1023	50,4755	0,76	0,9	0,65	0,41	0,78
altküme_21	74,1855	49,9119	0,76	0,9	0,66	0,4	0,78
altküme_20	74,1685	50,0089	0,76	0,9	0,66	0,4	0,78
altküme_19	74,2162	49,8938	0,76	0,9	0,66	0,4	0,78
altküme_18	74,1531	49,9135	0,76	0,9	0,66	0,4	0,78
altküme_17	74,0746	49,9599	0,76	0,9	0,65	0,4	0,78
altküme_16	73,6789	48,9427	0,76	0,9	0,65	0,39	0,77
altküme_15	73,6234	48,7892	0,76	0,9	0,65	0,39	0,77
altküme_14	73,608	48,7133	0,76	0,9	0,65	0,39	0,77
altküme_13	73,605	48,8512	0,76	0,9	0,64	0,39	0,77
altküme_12	73,6034	48,6444	0,76	0,9	0,65	0,39	0,77
altküme_11	73,5341	48,6189	0,76	0,9	0,64	0,39	0,77
altküme_10	73,5326	48,6144	0,76	0,9	0,64	0,39	0,77
altküme_9	73,5572	48,7633	0,76	0,9	0,64	0,39	0,77
altküme_8	73,4879	48,8412	0,76	0,9	0,64	0,39	0,77
altküme_7	73,0676	45,3935	0,75	0,91	0,65	0,35	0,75
altküme_6	71,9512	45,0829	0,75	0,89	0,6	0,36	0,74
altküme_5	71,8573	45,0465	0,75	0,89	0,6	0,36	0,74
altküme_4	71,8404	44,9952	0,75	0,89	0,6	0,36	0,74
altküme_3	71,6633	43,9531	0,74	0,89	0,6	0,35	0,73
altküme_2	70,8303	40,1113	0,73	0,9	0,59	0,3	0,71
altküme_1	69,1781	22,6277	0,7	0,95	0,58	0,14	0,68

AdaBoost'da ise, 17 öznitelik ile en iyi sonuca ulaşıyor ve bu değerini biraz azalma gösterse de, ciddi bir düşüş olmadan yeni öznitelikler eklendikçe korumaya devam ediyor. Elde edilen sonuçlar Çizelge 5.7'de verilmiştir.

Çizelge 5.7 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – Adaboost sınıflandırıcısı.

AdaBoost Sınıflandırıcısı	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	73,3755	49,7866	0,76	0,89	0,63	0,41	0,77
altküme_37	73,3755	49,7866	0,76	0,89	0,63	0,41	0,77
altküme_36	73,3755	49,7866	0,76	0,89	0,63	0,41	0,77
altküme_35	73,3755	49,7866	0,76	0,89	0,63	0,41	0,77
altküme_34	73,3755	49,7866	0,76	0,89	0,63	0,41	0,77
altküme_33	73,3755	49,7866	0,76	0,89	0,63	0,41	0,77

altküme_32	73,3755	49,7866	0,76	0,89	0,63	0,41	0,77
altküme_31	73,2631	49,6345	0,76	0,88	0,63	0,41	0,77
altküme_30	73,2585	49,6156	0,76	0,88	0,63	0,41	0,77
altküme_29	73,2585	49,6156	0,76	0,88	0,63	0,41	0,77
altküme_28	73,3047	49,6997	0,76	0,88	0,63	0,41	0,77
altküme_27	73,2677	49,7322	0,76	0,88	0,63	0,41	0,77
altküme_26	73,3863	49,9566	0,76	0,88	0,63	0,41	0,77
altküme_25	73,2985	50,1194	0,76	0,88	0,62	0,42	0,77
altküme_24	73,2985	50,1194	0,76	0,88	0,62	0,42	0,77
altküme_23	73,2985	50,1194	0,76	0,88	0,62	0,42	0,77
altküme_22	73,4402	49,8736	0,76	0,89	0,63	0,41	0,77
altküme_21	73,3771	50,17	0,76	0,88	0,63	0,42	0,77
altküme_20	73,4109	49,9188	0,76	0,89	0,63	0,41	0,77
altküme_19	73,4109	49,9188	0,76	0,89	0,63	0,41	0,77
altküme_18	73,434	50,0072	0,76	0,89	0,63	0,41	0,77
altküme_17	73,5541	49,9257	0,76	0,89	0,63	0,41	0,77
altküme_16	72,8997	48,6371	0,76	0,88	0,62	0,4	0,76
altküme_15	72,8982	48,6357	0,76	0,88	0,62	0,4	0,76
altküme_14	72,8982	48,6357	0,76	0,88	0,62	0,4	0,76
altküme_13	72,8766	48,5348	0,76	0,88	0,62	0,4	0,76
altküme_12	72,9613	48,7479	0,76	0,88	0,62	0,4	0,76
altküme_11	72,9706	48,7803	0,76	0,88	0,62	0,4	0,76
altküme_10	72,9706	48,7803	0,76	0,88	0,62	0,4	0,76
altküme_9	72,9706	48,7803	0,76	0,88	0,62	0,4	0,76
altküme_8	72,7396	48,1612	0,76	0,88	0,62	0,4	0,76
altküme_7	71,9728	43,1436	0,74	0,9	0,62	0,33	0,73
altküme_6	70,7456	38,9943	0,73	0,9	0,59	0,29	0,72
altküme_5	70,6147	39,6496	0,73	0,9	0,58	0,3	0,72
altküme_4	71,0212	41,9172	0,74	0,89	0,59	0,33	0,72
altküme_3	70,6547	39,4523	0,73	0,9	0,58	0,3	0,72
altküme_2	70,6393	37,3505	0,73	0,91	0,59	0,27	0,71
altküme_1	69,1796	22,1954	0,7	0,95	0,58	0,14	0,68

Yerine koyma sınıflandırıcısı (bagging) ile 25 öznitelik ile en iyi doğruluk sonuçlarına ulaşmakta ve bu değerini, bazı özniteliklerin eklenmesiyle ufak düşüşler olsa da korumayı başardığı Çizelge 5.8 ile görülmektedir.

Çizelge 5.8 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – Yerine koyma sınıflandırıcısı

Bagging Sınıflandırıcısı	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	73,3925	49,8956	0,76	0,89	0,63	0,41	0,75
altküme_37	73,2908	49,4669	0,76	0,89	0,63	0,41	0,75
altküme_36	73,2677	49,6622	0,76	0,88	0,63	0,41	0,75
altküme_35	73,5819	50,1931	0,76	0,89	0,63	0,42	0,75
altküme_34	73,4741	50,0131	0,76	0,89	0,63	0,41	0,75
altküme_33	73,3894	49,8025	0,76	0,89	0,63	0,41	0,75
altküme_32	73,7174	50,1154	0,76	0,89	0,64	0,41	0,75

altküme_31	73,4202	49,6647	0,76	0,89	0,63	0,41	0,75
altküme_30	73,4048	49,7878	0,76	0,89	0,63	0,41	0,75
altküme_29	73,3247	49,6922	0,76	0,89	0,63	0,41	0,75
altküme_28	73,4633	49,717	0,76	0,89	0,63	0,41	0,75
altküme_27	73,4448	49,9855	0,76	0,89	0,63	0,41	0,75
altküme_26	73,3724	49,6257	0,76	0,89	0,63	0,41	0,75
altküme_25	73,6157	50,2685	0,76	0,89	0,63	0,42	0,75
altküme_24	73,5387	50,1233	0,76	0,89	0,63	0,41	0,75
altküme_23	73,163	49,547	0,76	0,88	0,62	0,41	0,75
altküme_22	73,2046	49,574	0,76	0,88	0,62	0,41	0,75
altküme_21	73,1061	49,4179	0,76	0,88	0,62	0,41	0,75
altküme_20	73,0907	49,3596	0,76	0,88	0,62	0,41	0,75
altküme_19	73,2554	49,6886	0,76	0,88	0,63	0,41	0,75
altküme_18	73,026	49,408	0,76	0,88	0,62	0,41	0,75
altküme_17	72,7165	48,8968	0,76	0,88	0,61	0,41	0,74
altküme_16	72,7227	48,609	0,76	0,88	0,61	0,4	0,74
altküme_15	72,7196	48,6062	0,76	0,88	0,61	0,4	0,74
altküme_14	72,8381	49,0409	0,76	0,88	0,61	0,41	0,74
altküme_13	72,1144	47,8369	0,75	0,87	0,6	0,4	0,74
altküme_12	72,3546	48,0528	0,76	0,88	0,6	0,4	0,74
altküme_11	72,2407	47,8899	0,76	0,88	0,6	0,4	0,73
altküme_10	72,1329	47,7208	0,75	0,87	0,6	0,4	0,73
altküme_9	72,4178	48,2986	0,76	0,88	0,6	0,4	0,73
altküme_8	71,6001	47,453	0,75	0,86	0,58	0,4	0,73
altküme_7	71,6294	46,8668	0,75	0,87	0,59	0,39	0,72
altküme_6	69,7786	43,3384	0,74	0,86	0,54	0,36	0,69
altküme_5	69,6523	43,3861	0,74	0,85	0,54	0,36	0,69
altküme_4	69,6184	43,4204	0,74	0,85	0,54	0,36	0,69
altküme_3	67,5105	42,7564	0,74	0,81	0,49	0,38	0,65
altküme_2	66,2802	42,3665	0,73	0,79	0,47	0,39	0,63
altküme_1	67,9709	25,4952	0,7	0,92	0,5	0,17	0,65

Rastgele orman sınıflandırma algoritması, 19 öznitelik ile en iyi doğruluk ve F ölçütüne ulaştığı görülse de 10 öznitelik ile ulaştığı doğruluk oranı oldukça başarılı bir model olduğu Çizelge 5.9 ile görülmektedir.

Çizelge 5.9 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – Rastgele orman sınıflayıcısı.

Rast. Orman Sınıflandırıcısı	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	71,9466	45,4439	0,75	0,89	0,6	0,36	0,73
altküme_37	72,4224	45,9631	0,75	0,89	0,62	0,37	0,74
altküme_36	72,447	46,4027	0,75	0,89	0,62	0,37	0,74
altküme_35	72,3546	45,7157	0,75	0,89	0,62	0,36	0,74
altküme_34	72,4332	46,4191	0,75	0,89	0,62	0,37	0,74
altküme_33	72,233	45,5739	0,75	0,89	0,61	0,36	0,74
altküme_32	72,3485	46,1206	0,75	0,89	0,61	0,37	0,74
altküme_31	72,2915	45,8357	0,75	0,89	0,61	0,37	0,74

altküme_30	72,4655	46,3905	0,75	0,89	0,62	0,37	0,74
altküme_29	72,3085	45,942	0,75	0,89	0,61	0,37	0,74
altküme_28	72,2253	45,6425	0,75	0,89	0,61	0,36	0,74
altküme_27	72,2823	45,8861	0,75	0,89	0,61	0,37	0,74
altküme_26	72,3254	45,98	0,75	0,89	0,61	0,37	0,74
altküme_25	72,3439	46,055	0,75	0,89	0,61	0,37	0,74
altküme_24	72,4424	46,453	0,75	0,89	0,62	0,37	0,74
altküme_23	72,4778	46,2371	0,75	0,89	0,62	0,37	0,74
altküme_22	72,1298	45,7954	0,75	0,89	0,61	0,37	0,74
altküme_21	72,4978	46,5799	0,75	0,89	0,62	0,37	0,74
altküme_20	72,3639	46,2763	0,75	0,89	0,61	0,37	0,74
altküme_19	72,6241	46,7615	0,75	0,89	0,62	0,38	0,74
altküme_18	72,2623	46,1529	0,75	0,89	0,61	0,37	0,74
altküme_17	72,4085	46,6554	0,75	0,89	0,61	0,38	0,74
altküme_16	72,0051	46,1479	0,75	0,88	0,6	0,37	0,73
altküme_15	71,9558	45,8253	0,75	0,88	0,6	0,37	0,73
altküme_14	71,7664	45,5161	0,75	0,88	0,6	0,37	0,73
altküme_13	71,8388	46,0103	0,75	0,88	0,6	0,37	0,73
altküme_12	71,8434	46,2524	0,75	0,88	0,6	0,38	0,73
altküme_11	71,7664	46,0738	0,75	0,88	0,59	0,38	0,73
altküme_10	72,2746	46,8034	0,75	0,88	0,61	0,38	0,74
altküme_9	72,0451	46,2631	0,75	0,88	0,6	0,38	0,73
altküme_8	71,6463	47,5474	0,75	0,87	0,58	0,4	0,73
altküme_7	71,9959	46,8139	0,75	0,88	0,6	0,38	0,72
altküme_6	69,8802	43,4702	0,74	0,86	0,55	0,36	0,69
altküme_5	69,951	43,4167	0,74	0,86	0,55	0,36	0,69
altküme_4	69,8956	43,5383	0,74	0,86	0,55	0,36	0,69
altküme_3	67,452	42,7992	0,74	0,81	0,49	0,38	0,65
altküme_2	66,7129	42,6365	0,73	0,8	0,48	0,39	0,64
altküme_1	68,0617	24,8424	0,7	0,92	0,5	0,16	0,65

Yapay sinir ağları, bagging veya boosting modeli olmamasına rağmen öznitelik sayısı arttıkça, gerçek dünyayı daha iyi modellediği için yeni parametreler ile doğruluk oranının artması beklenen bir modeldir. Elde edilen sonuçlarda bunu doğrulamaktadır. 8 öznitelik ile ağaç sıralamasında en iyi sonuca ulaşmayı başarmıştır. Bunu yeni öznitelikler eklendikçe korumayı başarmıştır. Sonuçlar Çizelge 5.10’da verilmiştir.

Çizelge 5.10 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – YSA sınıflayıcısı.

Yapay Sinir Ağları	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	72,4316	47,0014	0,75	0,89	0,61	0,38	0,75
altküme_37	72,096	48,6454	0,76	0,87	0,59	0,41	0,74
altküme_36	72,2638	48,8456	0,76	0,87	0,6	0,41	0,75
altküme_35	72,3439	44,7473	0,75	0,9	0,62	0,35	0,75
altküme_34	72,4132	46,9124	0,75	0,89	0,61	0,38	0,75
altküme_33	72,1375	50,8328	0,77	0,85	0,58	0,45	0,75

altküme_32	72,3993	46,1501	0,75	0,89	0,62	0,37	0,75
altküme_31	72,0744	49,5578	0,76	0,86	0,59	0,43	0,75
altküme_30	72,3762	50,6791	0,77	0,86	0,59	0,44	0,75
altküme_29	72,4132	45,3414	0,75	0,9	0,62	0,36	0,75
altküme_28	72,3177	48,1304	0,76	0,88	0,6	0,4	0,75
altküme_27	71,9327	51,4257	0,77	0,84	0,58	0,46	0,75
altküme_26	71,7018	53,334	0,78	0,82	0,57	0,5	0,75
altküme_25	72,3808	46,2884	0,75	0,89	0,61	0,37	0,75
altküme_24	72,4255	45,9626	0,75	0,89	0,62	0,37	0,75
altküme_23	72,213	48,5605	0,76	0,87	0,6	0,41	0,75
altküme_22	72,4085	50,6513	0,77	0,86	0,59	0,44	0,75
altküme_21	72,3916	50,2276	0,76	0,86	0,59	0,43	0,75
altküme_20	72,3085	50,4409	0,76	0,86	0,59	0,44	0,75
altküme_19	72,4286	45,6274	0,75	0,9	0,62	0,36	0,75
altküme_18	72,4301	44,6779	0,75	0,9	0,63	0,35	0,75
altküme_17	72,638	45,6608	0,75	0,9	0,63	0,36	0,75
altküme_16	72,4085	43,9523	0,74	0,91	0,63	0,34	0,75
altküme_15	72,37	49,0835	0,76	0,87	0,6	0,42	0,75
altküme_14	72,3685	49,1657	0,76	0,87	0,6	0,42	0,75
altküme_13	72,3346	48,6055	0,76	0,87	0,6	0,41	0,75
altküme_12	72,4024	49,1791	0,76	0,87	0,6	0,42	0,75
altküme_11	72,4917	48,5203	0,76	0,88	0,61	0,4	0,75
altküme_10	72,5025	47,3091	0,75	0,89	0,61	0,39	0,75
altküme_9	72,5456	50,8545	0,77	0,86	0,6	0,44	0,75
altküme_8	72,9967	49,6887	0,76	0,88	0,62	0,42	0,76
altküme_7	72,5025	45,8684	0,75	0,9	0,62	0,36	0,74
altküme_6	70,8626	36,6848	0,73	0,92	0,6	0,26	0,73
altküme_5	71,2629	44,8411	0,75	0,88	0,58	0,36	0,72
altküme_4	70,8672	44,6071	0,74	0,87	0,57	0,37	0,72
altküme_3	70,801	38,5575	0,73	0,91	0,59	0,29	0,71
altküme_2	70,219	36,3469	0,72	0,91	0,58	0,27	0,71
altküme_1	68,459	4,5569	0,68	1	0,75	0,02	0,68

Extra Gradyen Boosting modelinde, 24 öznitelik sayısı ile yüksek doğruluk oranlarına erişmiş ve bunu yeni öznitelikler eklendikçe korumayı ve artırmayı başarmıştır.

Çizelge 5.11 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – XGB sınıflayıcısı.

XGB	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	74,284	50,1478	0,76	0,9	0,66	0,4	0,78
altküme_37	74,284	50,1478	0,76	0,9	0,66	0,4	0,78
altküme_36	74,284	50,1478	0,76	0,9	0,66	0,4	0,78
altküme_35	74,284	50,1478	0,76	0,9	0,66	0,4	0,78
altküme_34	74,284	50,1478	0,76	0,9	0,66	0,4	0,78
altküme_33	74,284	50,1478	0,76	0,9	0,66	0,4	0,78
altküme_32	74,2578	50,2766	0,76	0,9	0,66	0,41	0,78
altküme_31	74,2378	50,2394	0,76	0,9	0,66	0,41	0,78

altküme_30	74,264	50,2382	0,76	0,9	0,66	0,41	0,78
altküme_29	74,264	50,2382	0,76	0,9	0,66	0,41	0,78
altküme_28	74,264	50,2382	0,76	0,9	0,66	0,41	0,78
altküme_27	74,264	50,2382	0,76	0,9	0,66	0,41	0,78
altküme_26	74,2547	50,2025	0,76	0,9	0,66	0,41	0,78
altküme_25	74,1839	49,7994	0,76	0,9	0,66	0,4	0,78
altküme_24	74,2732	50,1015	0,76	0,9	0,66	0,4	0,78
altküme_23	74,1855	49,8579	0,76	0,9	0,66	0,4	0,78
altküme_22	74,1131	49,8748	0,76	0,9	0,66	0,4	0,78
altküme_21	74,0592	49,419	0,76	0,9	0,66	0,4	0,78
altküme_20	74,0022	49,279	0,76	0,9	0,66	0,39	0,78
altküme_19	74,0161	49,2099	0,76	0,9	0,66	0,39	0,78
altküme_18	73,9683	49,3134	0,76	0,9	0,66	0,4	0,78
altküme_17	73,9452	49,2365	0,76	0,9	0,66	0,39	0,78
altküme_16	73,5387	47,831	0,76	0,9	0,65	0,38	0,77
altküme_15	73,5434	47,7783	0,75	0,9	0,65	0,38	0,77
altküme_14	73,5372	47,8991	0,76	0,9	0,65	0,38	0,77
altküme_13	73,4925	47,9611	0,76	0,9	0,65	0,38	0,77
altküme_12	73,5449	47,9222	0,76	0,9	0,65	0,38	0,77
altküme_11	73,5865	48,2471	0,76	0,9	0,65	0,38	0,77
altküme_10	73,5341	48,0348	0,76	0,9	0,65	0,38	0,77
altküme_9	73,5357	48,0394	0,76	0,9	0,65	0,38	0,77
altküme_8	73,4387	47,9324	0,76	0,9	0,64	0,38	0,77
altküme_7	72,8782	44,6901	0,75	0,91	0,64	0,34	0,75
altküme_6	71,8573	44,2927	0,74	0,89	0,61	0,35	0,74
altküme_5	71,7942	44,3357	0,74	0,89	0,6	0,35	0,74
altküme_4	71,8989	44,9638	0,75	0,89	0,6	0,36	0,74
altküme_3	71,6848	43,7525	0,74	0,89	0,6	0,34	0,73
altküme_2	70,8149	39,9277	0,73	0,9	0,59	0,3	0,71
altküme_1	69,2027	21,8558	0,7	0,95	0,58	0,13	0,68

Extra trees modelinde ise, 7 öznitelik ile en iyi sonuca ulaşılmış olduğu görülmüştür.

Yeni öznitelik eklendikçe, doğruluk oranında çok az da olsa bir düşüş gözlenmiş, model ulaştığı doğruluk oranını korumakta çok etkili olamamış.

Çizelge 5.12 : Ağaç temelli yöntemlerle oluşturulmuş öznitelik önemleri – Extra Trees sınıflayıcısı.

Extra Trees	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	69,4876	39,864	0,73	0,87	0,54	0,32	0,69
altküme_37	69,332	39,5557	0,73	0,87	0,54	0,31	0,69
altküme_36	69,3659	39,6774	0,73	0,87	0,54	0,31	0,69
altküme_35	69,6215	39,9915	0,73	0,88	0,54	0,32	0,69
altküme_34	69,7555	40,3631	0,73	0,88	0,55	0,32	0,69
altküme_33	69,5184	40,0702	0,73	0,87	0,54	0,32	0,69
altküme_32	69,6246	40,0195	0,73	0,88	0,54	0,32	0,69
altküme_31	69,4614	39,7521	0,73	0,87	0,54	0,31	0,69
altküme_30	69,5692	39,9793	0,73	0,87	0,54	0,32	0,69

altküme_29	70,1481	40,6054	0,73	0,88	0,56	0,32	0,70
altküme_28	69,6431	40,0085	0,73	0,88	0,55	0,32	0,69
altküme_27	69,951	40,6063	0,73	0,88	0,55	0,32	0,70
altküme_26	70,1897	41,1085	0,73	0,88	0,56	0,32	0,70
altküme_25	69,9726	40,8433	0,73	0,88	0,55	0,32	0,69
altküme_24	69,814	40,9127	0,73	0,87	0,55	0,33	0,69
altküme_23	69,6693	41,0275	0,73	0,87	0,54	0,33	0,69
altküme_22	69,6092	41,1591	0,73	0,87	0,54	0,33	0,69
altküme_21	69,7632	41,4677	0,73	0,87	0,55	0,33	0,69
altküme_20	69,6847	41,464	0,73	0,87	0,54	0,34	0,69
altküme_19	69,7586	41,1024	0,73	0,87	0,55	0,33	0,69
altküme_18	69,4891	41,509	0,73	0,86	0,54	0,34	0,69
altküme_17	69,3397	41,9069	0,74	0,86	0,53	0,35	0,69
altküme_16	68,6576	41,2028	0,73	0,85	0,52	0,34	0,68
altküme_15	68,79	41,414	0,73	0,85	0,52	0,34	0,69
altküme_14	68,9163	41,5361	0,73	0,85	0,52	0,34	0,69
altküme_13	69,0395	42,0598	0,74	0,85	0,53	0,35	0,69
altküme_12	69,4629	42,8341	0,74	0,85	0,54	0,36	0,70
altküme_11	69,757	43,4026	0,74	0,86	0,54	0,36	0,70
altküme_10	69,5245	42,8472	0,74	0,85	0,54	0,36	0,70
altküme_9	69,4999	42,867	0,74	0,85	0,54	0,36	0,70
altküme_8	70,4869	45,2481	0,75	0,86	0,56	0,38	0,71
altküme_7	71,1644	45,3019	0,75	0,87	0,58	0,37	0,71
altküme_6	69,4244	42,4318	0,74	0,86	0,53	0,35	0,68
altküme_5	69,2073	42,5312	0,74	0,85	0,53	0,36	0,68
altküme_4	69,3983	42,8842	0,74	0,85	0,53	0,36	0,68
altküme_3	66,8576	42,1709	0,73	0,81	0,48	0,38	0,64
altküme_2	65,7459	42,2752	0,73	0,78	0,46	0,39	0,62
altküme_1	68,2373	23,2875	0,7	0,93	0,51	0,15	0,65

5.3.2 Temel bileşen analizi (PCA) yöntemiyle oluşturulmuş alt kümeler

Lojistik regresyon modeli için, öznitelik sayısı arttıkça iyileşme göstermiştir. Model, eklenen her yeni öznitelik ile lineer fonksiyona bir yeni bilinmeyen daha eklediğinden, eklenen her öznitelik ile iyileşme gösterdiği görülmüştür. En iyi sonuca en son eklenen öznitelik ile ulaşılmıştır. Değerler Çizelge 5.13'te verilmiştir.

Çizelge 5.13 : Temel bileşen yöntemlerle oluşturulmuş öznitelik önemleri – Lojistik Regresyon sınıflayıcısı.

Lojistik Reg.	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_33	69,2951	20,2775	0,7	0,96	0,6	0,12	0,7
altküme_32	69,2843	20,0545	0,7	0,96	0,6	0,12	0,7
altküme_31	69,2504	19,7992	0,7	0,96	0,6	0,12	0,7
altküme_30	69,2566	19,8024	0,7	0,96	0,6	0,12	0,7
altküme_29	69,2474	19,7396	0,7	0,96	0,6	0,12	0,7
altküme_28	69,2397	19,4703	0,7	0,96	0,6	0,12	0,7

altküme_27	69,252	19,6321	0,7	0,96	0,6	0,12	0,7
altküme_26	68,9948	17,8055	0,7	0,97	0,59	0,1	0,7
altküme_25	68,9964	17,8063	0,7	0,97	0,59	0,1	0,7
altküme_24	68,9302	17,1267	0,7	0,97	0,59	0,1	0,7
altküme_23	68,9409	17,2879	0,7	0,97	0,59	0,1	0,7
altküme_22	68,9132	17,1597	0,7	0,97	0,59	0,1	0,7
altküme_21	68,9101	17,0222	0,69	0,97	0,59	0,1	0,7
altküme_20	68,9163	17,0454	0,69	0,97	0,59	0,1	0,7
altküme_19	68,9209	16,9929	0,69	0,97	0,59	0,1	0,7
altküme_18	68,9225	16,9936	0,69	0,97	0,59	0,1	0,7
altküme_17	68,0525	14,7506	0,69	0,96	0,51	0,09	0,67
altküme_16	68,5375	18,232	0,7	0,96	0,54	0,11	0,66
altküme_15	68,5852	17,9654	0,69	0,96	0,55	0,11	0,66
altküme_14	68,6468	17,0726	0,69	0,96	0,56	0,1	0,66
altküme_13	68,3681	7,5764	0,69	0,99	0,59	0,04	0,64
altküme_12	68,3466	7,38	0,69	0,99	0,59	0,04	0,64
altküme_11	68,1079	1,6804	0,68	1	0,69	0,01	0,64
altküme_10	68,1125	1,5966	0,68	1	0,71	0,01	0,64
altküme_9	68,0833	1,2576	0,68	1	0,72	0,01	0,64
altküme_8	68,0817	1,2482	0,68	1	0,72	0,01	0,64
altküme_7	68,0833	1,2576	0,68	1	0,72	0,01	0,64
altküme_6	68,0787	1,1067	0,68	1	0,75	0,01	0,63
altküme_5	68,0201	0,3837	0,68	1	1	0	0,56
altküme_4	68,0201	0,3837	0,68	1	1	0	0,54
altküme_3	68,0201	0,3837	0,68	1	1	0	0,55
altküme_2	68,0201	0,3837	0,68	1	1	0	0,51
altküme_1	68,0201	0,3837	0,68	1	1	0	0,51

Naive bayes yönteminde, öznitelik sayısı 7 doğruluk ve f ölçütü oldukça başarılı bir performans göstermiş iken, yeni eklenen özniteliklerle bu performansı 17 öznitelik sayısına ulaşana kadar koruyabilmiştir. Ancak ‘transfer_ay’ parametresiyle doğrulukta düşüş meydana gelmiş iken, F ölçütü ve ROC değeri artışa geçmiştir.

Çizelge 5.14 : Temel bileşen yöntemlerle oluşturulmuş öznitelik önemleri – Naive Bayes sınıflayıcısı.

Naive Bayes	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_33	41,4311	51,9583	0,96	0,14	0,35	0,99	0,61
altküme_32	41,4326	51,9602	0,96	0,14	0,35	0,99	0,61
altküme_31	41,4326	51,9602	0,96	0,14	0,35	0,99	0,61
altküme_30	41,4049	51,9496	0,96	0,14	0,35	0,99	0,61
altküme_29	41,3679	51,9448	0,96	0,14	0,35	0,99	0,62
altküme_28	41,3064	51,9271	0,97	0,14	0,35	0,99	0,62
altküme_27	41,2956	51,9237	0,97	0,14	0,35	0,99	0,62
altküme_26	41,2294	51,9174	0,97	0,14	0,35	0,99	0,61
altküme_25	41,8607	51,941	0,94	0,15	0,35	0,98	0,64
altküme_24	41,8191	51,9415	0,95	0,15	0,35	0,98	0,64

altküme_23	41,7975	51,9384	0,95	0,15	0,35	0,98	0,64
altküme_22	41,6898	51,8922	0,95	0,15	0,35	0,98	0,64
altküme_21	41,6944	51,9101	0,95	0,15	0,35	0,98	0,64
altküme_20	41,5881	51,9013	0,95	0,15	0,35	0,98	0,63
altküme_19	45,1358	51,5837	0,85	0,23	0,36	0,91	0,63
altküme_18	44,8248	51,524	0,85	0,23	0,36	0,92	0,63
altküme_17	54,1282	49,2565	0,77	0,47	0,38	0,69	0,61
altküme_16	53,2212	49,3852	0,77	0,45	0,38	0,71	0,61
altküme_15	53,2474	49,4329	0,77	0,45	0,38	0,71	0,61
altküme_14	53,2151	49,4072	0,77	0,45	0,38	0,71	0,61
altküme_13	52,9364	49,0846	0,76	0,45	0,38	0,71	0,6
altküme_12	52,5391	48,9406	0,76	0,44	0,37	0,71	0,6
altküme_11	52,3559	48,8545	0,76	0,44	0,37	0,71	0,6
altküme_10	52,3405	48,8312	0,76	0,44	0,37	0,71	0,6
altküme_9	44,9664	50,9113	0,82	0,24	0,36	0,89	0,6
altküme_8	53,5277	49,669	0,77	0,45	0,38	0,72	0,6
altküme_7	53,6832	49,9234	0,77	0,45	0,38	0,72	0,59
altküme_6	60,5968	35,0903	0,7	0,73	0,37	0,33	0,59
altküme_5	67,6398	4,1328	0,68	0,99	0,41	0,02	0,59
altküme_4	67,6752	3,7769	0,68	0,99	0,41	0,02	0,56
altküme_3	67,6752	3,7769	0,68	0,99	0,41	0,02	0,52
altküme_2	67,6752	3,7769	0,68	0,99	0,41	0,02	0,51
altküme_1	67,6752	3,7769	0,68	0,99	0,41	0,02	0,51

Karar ağaçlarında ise 13 öznitelikle yüksek doğruluk oranlarına ulaşmış ve bu oranı yeni özniteliklerle artırarak devam etmiştir.

Çizelge 5.15 : Temel bileşen yöntemlerle oluşturulmuş öznitelik önemleri – Karar ağaçları sınıflayıcısı.

Karar Ağaçları	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_33	65,0299	44,9684	0,75	0,74	0,45	0,45	0,6
altküme_32	65,1777	45,3387	0,75	0,74	0,45	0,45	0,6
altküme_31	65,0699	45,058	0,75	0,74	0,45	0,45	0,6
altküme_30	64,9868	44,9832	0,75	0,74	0,45	0,45	0,6
altküme_29	65,3332	45,5263	0,75	0,75	0,45	0,46	0,6
altküme_28	64,9929	44,9902	0,75	0,74	0,45	0,45	0,6
altküme_27	65,401	45,6459	0,75	0,75	0,45	0,46	0,6
altküme_26	65,3532	45,1235	0,75	0,75	0,44	0,45	0,6
altküme_25	65,3286	45,3245	0,75	0,75	0,45	0,45	0,6
altküme_24	65,3563	45,4054	0,75	0,75	0,45	0,45	0,6
altküme_23	65,2793	45,5851	0,75	0,75	0,45	0,46	0,6
altküme_22	65,2593	45,4866	0,75	0,75	0,45	0,45	0,6
altküme_21	65,4194	45,5669	0,75	0,75	0,45	0,46	0,6
altküme_20	65,4764	45,2866	0,75	0,75	0,45	0,45	0,6
altküme_19	65,4564	45,2482	0,75	0,75	0,45	0,45	0,6
altküme_18	65,267	44,785	0,75	0,75	0,44	0,45	0,6
altküme_17	65,4764	45,2973	0,75	0,75	0,45	0,45	0,6

altküme_16	65,2839	45,1168	0,75	0,75	0,45	0,45	0,6
altküme_15	65,4964	45,3597	0,75	0,75	0,45	0,45	0,6
altküme_14	65,441	45,1113	0,75	0,75	0,44	0,45	0,6
altküme_13	65,3671	45,2749	0,75	0,75	0,45	0,45	0,6
altküme_12	64,6126	44,2563	0,74	0,74	0,44	0,44	0,59
altküme_11	64,1891	43,5496	0,74	0,74	0,43	0,44	0,59
altküme_10	63,5424	42,5357	0,74	0,73	0,42	0,43	0,58
altküme_9	63,4408	42,1453	0,74	0,73	0,42	0,42	0,57
altküme_8	63,5332	42,2774	0,74	0,73	0,42	0,42	0,57
altküme_7	63,4454	42,3814	0,74	0,73	0,42	0,42	0,57
altküme_6	63,3115	42,0845	0,74	0,73	0,42	0,42	0,57
altküme_5	63,0974	41,6175	0,73	0,73	0,41	0,42	0,57
altküme_4	61,6608	28,1103	0,8	0,74	0,23	0,28	0,54
altküme_3	63,3792	18,2462	0,87	0,76	0,13	0,18	0,53
altküme_2	65,0201	3,4247	1	0,81	0	0	0,51
altküme_1	68,0201	0,3837	1	0,81	0	0	0,5

K en yakın komşu, az sayıdaki öznitelik ile yüksek doğruluk oranlarına ulaşmış iken, bu öznitelik sıralaması ile 5 öznitelik ile en yüksek değerine ulaşmıştır. Yeni özniteliklerin eklenmesiyle uzaydaki noktaların dağılımı doğrulukta ve F ölçütünde düşüş göstermiştir. Elde edilen sonuçlar Çizelge 5.16’da verilmiştir.

Çizelge 5.16 : Temel bileşen yöntemlerle oluşturulmuş öznitelik önemleri – K-en yakın komşu sınıflayıcısı.

K En Yakın Komşu	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_33	63,37	34,5035	0,71	0,79	0,4	0,3	0,6
altküme_32	63,5101	35,3397	0,71	0,79	0,41	0,31	0,61
altküme_31	63,6518	35,7765	0,71	0,79	0,41	0,32	0,61
altküme_30	63,6518	35,7765	0,71	0,79	0,41	0,32	0,61
altküme_29	63,6625	35,8042	0,71	0,79	0,41	0,32	0,61
altküme_28	63,6595	35,7988	0,71	0,79	0,41	0,32	0,61
altküme_27	63,9767	36,1508	0,71	0,79	0,42	0,32	0,61
altküme_26	64,1337	36,5987	0,71	0,79	0,42	0,32	0,61
altküme_25	64,146	36,6067	0,71	0,79	0,42	0,32	0,61
altküme_24	64,6341	38,0048	0,72	0,79	0,43	0,34	0,62
altküme_23	64,6141	37,9412	0,72	0,79	0,43	0,34	0,62
altküme_22	64,6034	37,9508	0,72	0,79	0,43	0,34	0,62
altküme_21	64,7111	38,1597	0,72	0,79	0,44	0,34	0,62
altküme_20	64,7111	38,1597	0,72	0,79	0,44	0,34	0,62
altküme_19	64,7096	38,1587	0,72	0,79	0,44	0,34	0,62
altküme_18	64,7389	38,0244	0,72	0,79	0,44	0,34	0,62
altküme_17	64,7327	38,0705	0,72	0,79	0,44	0,34	0,62
altküme_16	64,705	38,0252	0,72	0,79	0,43	0,34	0,62
altküme_15	64,7219	38,1168	0,72	0,79	0,44	0,34	0,62
altküme_14	64,7296	38,1287	0,72	0,79	0,44	0,34	0,62
altküme_13	64,959	38,3329	0,72	0,8	0,44	0,34	0,63

altküme_12	64,6911	37,3145	0,72	0,8	0,43	0,33	0,62
altküme_11	64,899	38,0947	0,72	0,8	0,44	0,34	0,63
altküme_10	64,9267	36,4311	0,71	0,81	0,43	0,31	0,61
altküme_9	66,4388	38,6788	0,72	0,82	0,47	0,33	0,64
altküme_8	66,4388	38,6788	0,72	0,82	0,47	0,33	0,64
altküme_7	66,5789	38,9881	0,72	0,82	0,47	0,33	0,64
altküme_6	66,6867	39,0134	0,72	0,82	0,47	0,33	0,64
altküme_5	67,3242	40,8342	0,73	0,82	0,49	0,35	0,65
altküme_4	62,6986	29,7235	0,69	0,81	0,37	0,25	0,56
altküme_3	62,1551	25,9431	0,69	0,82	0,35	0,21	0,52
altküme_2	67,7322	1,9281	0,68	0,99	0,37	0,01	0,5
altküme_1	69,4523	0,7335	0,68	0,99	0,37	0,01	0,47

Gradyen Boosting model ile modele yeni öznitelik eklendikçe, bu özniteliklerden de hata öğrenerek ilerlediğinden, eklenmemesi gereken bir öznitelik bile modelin doğruluğunu iyileştirmektedir. En iyi sonuca, tüm öznitelik altkütmesiyle ulaşılmıştır.

Çizelge 5.17 : Temel bileşen yöntemlerle oluşturulmuş öznitelik önemleri – Gradyen Boosting sınıflayıcısı.

Gradyen Boosting	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_33	71,035	38,9432	0,73	0,91	0,6	0,29	0,74
altküme_32	70,8764	38,3788	0,73	0,91	0,6	0,28	0,74
altküme_31	70,8102	38,565	0,73	0,91	0,59	0,29	0,74
altküme_30	70,8595	38,5093	0,73	0,91	0,59	0,28	0,74
altküme_29	70,8888	38,6089	0,73	0,91	0,6	0,29	0,74
altküme_28	70,8826	38,58	0,73	0,91	0,6	0,29	0,74
altküme_27	70,9365	38,0477	0,73	0,91	0,6	0,28	0,74
altküme_26	70,7363	37,8942	0,73	0,91	0,59	0,28	0,74
altküme_25	70,7363	37,8942	0,73	0,91	0,59	0,28	0,74
altküme_24	70,6978	38,767	0,73	0,9	0,59	0,29	0,74
altküme_23	70,6424	38,5562	0,73	0,9	0,59	0,29	0,74
altküme_22	70,6424	38,5562	0,73	0,9	0,59	0,29	0,74
altküme_21	70,6578	38,0534	0,73	0,91	0,59	0,28	0,74
altküme_20	70,587	36,5719	0,72	0,91	0,59	0,26	0,74
altküme_19	70,587	36,5719	0,72	0,91	0,59	0,26	0,74
altküme_18	70,5793	36,6332	0,73	0,91	0,59	0,27	0,74
altküme_17	70,4853	34,7672	0,72	0,92	0,6	0,25	0,74
altküme_16	70,4853	34,7672	0,72	0,92	0,6	0,25	0,74
altküme_15	70,4853	34,7672	0,72	0,92	0,6	0,25	0,74
altküme_14	70,4853	34,7672	0,72	0,92	0,6	0,25	0,74
altküme_13	70,2913	33,1323	0,72	0,93	0,59	0,23	0,73
altküme_12	70,1543	31,4725	0,72	0,93	0,6	0,21	0,73
altküme_11	70,0896	29,5506	0,71	0,94	0,6	0,2	0,72
altküme_10	69,6169	24,0376	0,7	0,95	0,6	0,15	0,7
altküme_9	69,446	23,9295	0,7	0,95	0,59	0,15	0,69
altküme_8	69,446	23,9295	0,7	0,95	0,59	0,15	0,69
altküme_7	69,4675	24,1169	0,7	0,95	0,59	0,15	0,69

altküme_6	69,4337	23,928	0,7	0,95	0,59	0,15	0,69
altküme_5	69,4891	24,8093	0,7	0,95	0,59	0,16	0,69
altküme_4	68,0325	0,8122	0,68	1	0,7	0	0,62
altküme_3	68,0248	0,7741	0,68	1	0,68	0	0,57
altküme_2	68,0201	0,3837	0,68	1	1	0	0,51
altküme_1	68,0201	0,3837	0,68	1	1	0	0,51

Adaboost algoritmasıyla, öznitelik eklendikçe iyileşmenin devam ettiği gözlenmiştir.

En iyi sonuca en son eklenen öznitelik ile ulaşılmıştır.

Çizelge 5.18 : Temel bileşen yöntemlerle oluşturulmuş öznitelik önemleri – AdaBoost sınıflayıcısı.

AdaBoost	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_33	70,7687	41,1093	0,73	0,89	0,58	0,32	0,74
altküme_32	70,7656	41,1615	0,74	0,89	0,58	0,32	0,74
altküme_31	70,647	41,0453	0,73	0,89	0,58	0,32	0,74
altküme_30	70,5823	40,7505	0,73	0,89	0,57	0,32	0,74
altküme_29	70,5823	40,7505	0,73	0,89	0,57	0,32	0,74
altküme_28	70,5254	40,656	0,73	0,89	0,57	0,32	0,74
altküme_27	70,5377	40,8202	0,73	0,89	0,57	0,32	0,74
altküme_26	70,4684	39,7171	0,73	0,89	0,57	0,3	0,74
altküme_25	70,4684	39,7171	0,73	0,89	0,57	0,3	0,74
altküme_24	70,4391	39,4576	0,73	0,89	0,57	0,3	0,74
altküme_23	70,4391	39,4576	0,73	0,89	0,57	0,3	0,74
altküme_22	70,4391	39,4576	0,73	0,89	0,57	0,3	0,74
altküme_21	70,3622	39,0153	0,73	0,9	0,57	0,3	0,74
altküme_20	70,3021	37,5482	0,73	0,9	0,58	0,28	0,74
altküme_19	70,2836	37,6063	0,73	0,9	0,57	0,28	0,74
altküme_18	70,3514	39,5315	0,73	0,89	0,57	0,3	0,74
altküme_17	69,7262	32,4712	0,72	0,92	0,57	0,23	0,72
altküme_16	69,7016	34,087	0,72	0,91	0,56	0,24	0,72
altküme_15	69,7016	34,087	0,72	0,91	0,56	0,24	0,72
altküme_14	69,7016	34,087	0,72	0,91	0,56	0,24	0,72
altküme_13	69,8232	30,2711	0,71	0,93	0,58	0,2	0,72
altküme_12	69,4814	29,2244	0,71	0,93	0,57	0,2	0,71
altküme_11	69,5599	27,8425	0,71	0,94	0,58	0,18	0,71
altküme_10	69,5184	25,9464	0,71	0,94	0,59	0,17	0,69
altküme_9	69,3521	25,6814	0,71	0,94	0,58	0,17	0,69
altküme_8	69,3521	25,6814	0,71	0,94	0,58	0,17	0,69
altküme_7	69,3875	22,8671	0,7	0,95	0,59	0,14	0,69
altküme_6	69,3367	25,0565	0,7	0,94	0,58	0,16	0,68
altküme_5	69,2982	25,2801	0,7	0,94	0,57	0,16	0,69
altküme_4	68,0402	0,727	0,68	1	0,77	0	0,62
altküme_3	68,0325	0,7363	0,68	1	0,73	0	0,57
altküme_2	68,0201	0,3837	0,68	1	1	0	0,51
altküme_1	68,0201	0,3837	0,68	1	1	0	0,51

Yerine koyma (bagging) yöntemi ile öznitelik eklendikçe iyileşmenin devam ettiği gözlenmiştir. En iyi sonuca 32 öznitelik alt küme öznitelik grubuyla ulaşılmıştır. Değerler Çizelge 5.19’da verilmiştir.

Çizelge 5.19 : Temel bileşen yöntemlerle oluşturulmuş öznitelik önemleri – Yerine koyma (Bagging) sınıflayıcısı.

Bagging	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_33	69,6369	42,066	0,74	0,86	0,54	0,34	0,71
altküme_32	69,7093	42,5367	0,74	0,86	0,54	0,35	0,71
altküme_31	69,5507	42,4057	0,74	0,86	0,54	0,35	0,71
altküme_30	69,5461	42,5092	0,74	0,86	0,54	0,35	0,7
altküme_29	69,4275	42,3138	0,74	0,86	0,54	0,35	0,7
altküme_28	69,6092	42,5599	0,74	0,86	0,54	0,35	0,71
altküme_27	69,7847	42,5371	0,74	0,86	0,54	0,35	0,71
altküme_26	69,5122	42,3379	0,74	0,86	0,54	0,35	0,7
altküme_25	69,4121	42,1234	0,74	0,86	0,53	0,35	0,7
altküme_24	69,5245	42,5453	0,74	0,86	0,54	0,35	0,7
altküme_23	69,4953	42,7808	0,74	0,85	0,54	0,36	0,7
altküme_22	69,4876	42,7051	0,74	0,86	0,54	0,35	0,7
altküme_21	69,349	42,591	0,74	0,85	0,53	0,35	0,7
altküme_20	69,369	42,6368	0,74	0,85	0,53	0,36	0,7
altküme_19	69,4337	42,4425	0,74	0,86	0,54	0,35	0,7
altküme_18	69,3028	42,4214	0,74	0,85	0,53	0,35	0,7
altküme_17	69,4383	42,913	0,74	0,85	0,53	0,36	0,7
altküme_16	69,5014	42,6333	0,74	0,86	0,54	0,35	0,7
altküme_15	69,5461	42,5759	0,74	0,86	0,54	0,35	0,7
altküme_14	69,1349	42,0212	0,74	0,85	0,53	0,35	0,7
altküme_13	69,1673	42,3039	0,74	0,85	0,53	0,35	0,7
altküme_12	68,633	40,6906	0,73	0,85	0,52	0,34	0,69
altküme_11	68,6407	40,4259	0,73	0,85	0,52	0,33	0,68
altküme_10	67,8723	38,5619	0,72	0,85	0,5	0,31	0,66
altküme_9	67,0562	39,4886	0,73	0,83	0,48	0,34	0,65
altküme_8	66,7775	38,9474	0,72	0,83	0,47	0,33	0,65
altküme_7	67,1948	39,5243	0,73	0,83	0,48	0,33	0,65
altküme_6	66,7514	39,2482	0,72	0,82	0,47	0,34	0,65
altküme_5	66,7236	39,5446	0,73	0,82	0,47	0,34	0,64
altküme_4	61,1573	30,7168	0,69	0,77	0,36	0,27	0,55
altküme_3	62,8557	21,231	0,68	0,85	0,33	0,16	0,53
altküme_2	68,0201	0,3837	0,68	1	1	0	0,51
altküme_1	68,0201	0,3837	0,68	1	1	0	0,51

Rastgele ormanlar yöntemiyle, ilk başta doğruluk oranı yüksek çıksa da bu değerler için F ölçütü oldukça düşüktür. Bu model, 5 öznitelik ile yüksek doğruluk oranlarına erişmiştir. Bilgi birikerek ilerlemiş ve en doğru sonuca 33. altküme ile ulaşılmıştır. Modelden elde edilen sonuçlar Çizelge 5.20’de verilmiştir.

Çizelge 5.20 : Temel bileşen yöntemiyle oluşturulmuş öznitelik önemleri – Rastgele orman sınıflayıcısı.

Rast. Orman Sınıflayıcısı	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_33	69,2243	40,2606	0,73	0,87	0,53	0,32	0,69
altküme_32	68,9271	40,0262	0,73	0,86	0,52	0,32	0,69
altküme_31	68,9594	40,3668	0,73	0,86	0,53	0,33	0,69
altküme_30	69,1088	40,6309	0,73	0,86	0,53	0,33	0,69
altküme_29	68,8793	40,5366	0,73	0,86	0,52	0,33	0,69
altküme_28	68,6037	40,1632	0,73	0,85	0,52	0,33	0,69
altküme_27	69,0241	40,6596	0,73	0,86	0,53	0,33	0,69
altküme_26	68,7854	40,8255	0,73	0,85	0,52	0,34	0,69
altküme_25	68,8639	40,7339	0,73	0,86	0,52	0,33	0,69
altküme_24	68,5298	41,0465	0,73	0,85	0,51	0,34	0,69
altküme_23	68,536	40,795	0,73	0,85	0,51	0,34	0,69
altküme_22	68,4728	40,5367	0,73	0,85	0,51	0,34	0,69
altküme_21	68,6145	40,6867	0,73	0,85	0,52	0,34	0,69
altküme_20	68,924	41,345	0,73	0,85	0,52	0,34	0,69
altküme_19	69,0102	41,3031	0,73	0,86	0,53	0,34	0,69
altküme_18	68,8978	40,9472	0,73	0,86	0,52	0,34	0,69
altküme_17	69,1873	41,253	0,73	0,86	0,53	0,34	0,69
altküme_16	68,8655	40,5609	0,73	0,86	0,52	0,33	0,69
altküme_15	68,9748	40,9172	0,73	0,86	0,52	0,34	0,69
altküme_14	69,1288	40,9924	0,73	0,86	0,53	0,33	0,7
altküme_13	68,8963	40,272	0,73	0,86	0,52	0,33	0,69
altküme_12	68,5021	39,1335	0,73	0,86	0,51	0,32	0,69
altküme_11	68,2034	38,7531	0,73	0,86	0,51	0,31	0,68
altküme_10	67,104	37,0833	0,72	0,84	0,48	0,3	0,65
altküme_9	66,7606	38,6437	0,72	0,83	0,47	0,33	0,64
altküme_8	66,8037	38,8623	0,72	0,83	0,47	0,33	0,65
altküme_7	66,9207	39,0432	0,72	0,83	0,48	0,33	0,64
altküme_6	66,9454	39,3913	0,73	0,83	0,48	0,34	0,64
altküme_5	66,6097	39,4291	0,72	0,82	0,47	0,34	0,64
altküme_4	61,0049	29,6352	0,69	0,78	0,35	0,26	0,55
altküme_3	62,5493	21,2166	0,68	0,85	0,33	0,16	0,53
altküme_2	68,0201	0,3837	0,68	1	1	0	0,51
altküme_1	68,0201	0,3837	0,68	1	1	0	0,51

Yapay sinir ağları, 5 öznitelik ile yüksek doğruluk değerlerine ulaşmış, sonrasında eklenen öznitelikler ile en iyi değerine 12. öznitelikte ulaşmış, sonrasında modelin performans artışı yaklaşık bu değerde sabit kaldığı Çizelge 5.21 ile görülmektedir.

Çizelge 5.21 : Temel bileşen yöntemiyle oluşturulmuş öznitelik önemleri – YSA sınıflayıcısı.

YSA Sınıflayıcısı	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
-------------------	----------	----------	------------	-------------	------------	-------------	-----

altküme_33	70,0357	35,7968	0,72	0,91	0,57	0,26	0,71
altküme_32	69,4537	44,3409	0,74	0,84	0,53	0,38	0,71
altküme_31	69,3474	38,0789	0,73	0,88	0,54	0,29	0,71
altküme_30	69,369	39,5772	0,73	0,87	0,54	0,31	0,71
altküme_29	69,5615	43,3419	0,74	0,85	0,54	0,36	0,71
altküme_28	69,1057	46,2581	0,75	0,82	0,52	0,41	0,71
altküme_27	69,5445	40,3726	0,73	0,87	0,54	0,32	0,71
altküme_26	69,175	46,4174	0,75	0,82	0,52	0,42	0,72
altküme_25	69,8756	35,5812	0,72	0,91	0,57	0,26	0,72
altküme_24	69,9926	37,6264	0,73	0,9	0,56	0,28	0,72
altküme_23	69,9849	38,1469	0,73	0,89	0,56	0,29	0,72
altküme_22	70,0234	41,365	0,73	0,87	0,55	0,33	0,72
altküme_21	70,0573	40,8577	0,73	0,88	0,56	0,32	0,72
altküme_20	70,1373	37,6679	0,73	0,9	0,57	0,28	0,72
altküme_19	69,7308	28,9915	0,71	0,94	0,58	0,19	0,72
altküme_18	70,1882	39,6948	0,73	0,89	0,56	0,31	0,72
altküme_17	70,2775	37,1094	0,73	0,91	0,58	0,27	0,73
altküme_16	70,1758	41,7666	0,74	0,88	0,56	0,33	0,73
altküme_15	69,9541	29,9882	0,71	0,93	0,59	0,2	0,72
altküme_14	70,2436	37,7677	0,73	0,9	0,57	0,28	0,73
altküme_13	69,774	39,9951	0,73	0,88	0,55	0,31	0,71
altküme_12	69,5645	41,9944	0,74	0,86	0,54	0,34	0,72
altküme_11	69,6231	28,0315	0,71	0,94	0,58	0,18	0,71
altküme_10	69,1488	28,6213	0,71	0,93	0,55	0,19	0,69
altküme_9	68,8609	24,3066	0,7	0,94	0,55	0,16	0,69
altküme_8	68,9563	27,8599	0,71	0,93	0,55	0,19	0,69
altküme_7	69,0333	22,3543	0,7	0,95	0,57	0,14	0,68
altküme_6	69,1319	21,301	0,7	0,96	0,58	0,13	0,69
altküme_5	68,6545	32,1637	0,71	0,9	0,52	0,23	0,69
altküme_4	68,0386	0,8597	0,68	1	0,7	0	0,62
altküme_3	68,034	0,7363	0,68	1	0,73	0	0,57
altküme_2	68,0201	0,3837	0,68	1	1	0	0,51
altküme_1	68,0201	0,3837	0,68	1	1	0	0,51

XGB sınıflayıcı, öznitelik sayısı arttıkça performansında artış göstermeye devam etmiştir. En yüksek değerine 33. altküme ile ulaşmıştır.

Çizelge 5.22 : Temel bileşen yöntemiyle oluşturulmuş öznitelik önemleri – XGB sınıflayıcısı.

XGB Sınıflayıcısı	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_33	71,0905	38,2584	0,73	0,91	0,61	0,28	0,74
altküme_32	71,0135	37,9757	0,73	0,91	0,6	0,28	0,74
altküme_31	70,9211	38,1043	0,73	0,91	0,6	0,28	0,74
altküme_30	70,9565	37,4104	0,73	0,92	0,6	0,27	0,74
altküme_29	70,918	37,8247	0,73	0,91	0,6	0,28	0,74
altküme_28	70,8888	37,7601	0,73	0,91	0,6	0,28	0,74
altküme_27	70,8287	37,4773	0,73	0,91	0,6	0,27	0,74

altküme_26	70,7132	36,6886	0,73	0,92	0,6	0,26	0,74
altküme_25	70,6747	36,7675	0,73	0,91	0,59	0,27	0,74
altküme_24	70,6686	37,3985	0,73	0,91	0,59	0,27	0,74
altküme_23	70,6732	37,468	0,73	0,91	0,59	0,27	0,74
altküme_22	70,6732	37,468	0,73	0,91	0,59	0,27	0,74
altküme_21	70,6208	37,2203	0,73	0,91	0,59	0,27	0,74
altküme_20	70,7286	35,664	0,72	0,92	0,6	0,25	0,74
altküme_19	70,6717	36,2401	0,72	0,92	0,6	0,26	0,74
altküme_18	70,7009	35,8852	0,72	0,92	0,6	0,26	0,74
altküme_17	70,5469	36,1122	0,72	0,92	0,59	0,26	0,74
altküme_16	70,5469	36,1122	0,72	0,92	0,59	0,26	0,74
altküme_15	70,5423	36,1385	0,72	0,92	0,59	0,26	0,74
altküme_14	70,5439	36,1439	0,72	0,92	0,59	0,26	0,74
altküme_13	70,3683	32,5907	0,72	0,93	0,6	0,22	0,73
altküme_12	70,2251	31,4752	0,72	0,93	0,6	0,21	0,73
altküme_11	70,0126	29,0554	0,71	0,94	0,6	0,19	0,72
altküme_10	69,6	23,781	0,7	0,95	0,6	0,15	0,7
altküme_9	69,3952	22,9134	0,7	0,95	0,59	0,14	0,69
altküme_8	69,3952	22,9134	0,7	0,95	0,59	0,14	0,69
altküme_7	69,4275	21,6487	0,7	0,96	0,61	0,13	0,69
altküme_6	69,4213	23,8564	0,7	0,95	0,59	0,15	0,69
altküme_5	69,446	24,0227	0,7	0,95	0,59	0,15	0,69
altküme_4	68,014	0,9536	0,68	1	0,61	0	0,62
altküme_3	68,0248	0,7741	0,68	1	0,68	0	0,57
altküme_2	68,0201	0,3837	0,68	1	1	0	0,51
altküme_1	68,0201	0,3837	0,68	1	1	0	0,51

Extra trees sınıflayıcı, 5 öznitelik ile yüksek doğruluk sonuçlarına ulaşmış ve yeni öznitelikler geldikçe bunu artırarak devam etmiştir. En yüksek değerine, en son alt küme ile ulaşmıştır. Sonuçlar Çizelge 5.23'te verilmiştir.

Çizelge 5.23 : Temel bileşen yöntemiyle oluşturulmuş öznitelik önemleri – Extra Trees sınıflayıcısı.

Extra Trees Sınıflayıcısı	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_33	67,1794	38,4831	0,72	0,84	0,48	0,32	0,66
altküme_32	67,0901	39,3106	0,73	0,83	0,48	0,33	0,65
altküme_31	67,0239	40,028	0,73	0,82	0,48	0,34	0,65
altküme_30	66,8268	39,8313	0,73	0,82	0,48	0,34	0,65
altküme_29	66,6528	39,7329	0,73	0,82	0,47	0,34	0,65
altküme_28	66,5158	39,9514	0,73	0,81	0,47	0,35	0,65
altküme_27	66,8314	40,0256	0,73	0,82	0,48	0,35	0,65
altküme_26	66,6159	40,6277	0,73	0,81	0,47	0,36	0,65
altküme_25	66,6282	40,7177	0,73	0,81	0,47	0,36	0,65
altküme_24	66,5466	41,2493	0,73	0,81	0,47	0,37	0,65
altküme_23	66,3218	41,0585	0,73	0,8	0,47	0,37	0,65
altküme_22	66,3649	40,9302	0,73	0,81	0,47	0,36	0,65
altküme_21	66,7421	41,0556	0,73	0,81	0,48	0,36	0,66

altküme_20	66,6374	40,9795	0,73	0,81	0,47	0,36	0,65
altküme_19	66,4834	40,8971	0,73	0,81	0,47	0,36	0,65
altküme_18	66,7113	41,2479	0,73	0,81	0,47	0,36	0,66
altküme_17	66,8853	41,2276	0,73	0,81	0,48	0,36	0,66
altküme_16	66,7837	40,9827	0,73	0,81	0,48	0,36	0,66
altküme_15	66,8668	41,2269	0,73	0,81	0,48	0,36	0,66
altküme_14	67,1517	41,2492	0,73	0,82	0,48	0,36	0,66
altküme_13	66,9577	40,2872	0,73	0,82	0,48	0,35	0,66
altküme_12	66,5758	39,6105	0,73	0,82	0,47	0,34	0,65
altküme_11	66,1955	39,5173	0,72	0,81	0,46	0,34	0,64
altküme_10	64,8821	37,5065	0,72	0,8	0,44	0,33	0,62
altküme_9	65,5519	39,0442	0,72	0,8	0,45	0,34	0,63
altküme_8	65,575	38,917	0,72	0,8	0,45	0,34	0,63
altküme_7	65,461	38,8684	0,72	0,8	0,45	0,34	0,62
altküme_6	65,6612	38,9666	0,72	0,8	0,45	0,34	0,62
altküme_5	65,3117	38,1167	0,72	0,8	0,44	0,33	0,62
altküme_4	61,3898	28,3469	0,69	0,79	0,35	0,24	0,54
altküme_3	61,1973	25,3643	0,68	0,8	0,33	0,21	0,52
altküme_2	68,0201	0,3837	0,68	1	1	0	0,51
altküme_1	68,0201	0,3837	0,68	1	1	0	0,51

5.3.3 Lojistik regresyon ile özyinelemeli yöntem ile oluşturulmuş alt kümeler

Özyinelemeli yöntemler için kullanılan ilk alt küme, model eğitilirken başlangıç parametresi olarak 5 verildiğinden ilk alt küme, en önemli 5 öz niteliği içeriyor. Diğerleri bu 5 niteliğe ilave edilerek ekleniyor. Aşağıdaki bütün yöntemlerde dikkat çekmiştir ki, ilk 10 öz nitelik ile hemen hemen bütün yöntemler için iyi sonuçlar elde edilebilmiştir.

Lojistik regresyon yönteminde, ilk alt küme ile yüksek sonuçlara ulaşmış ve yeni öz niteliklerle doğruluk oranını artırmaya devam etmiştir. En iyi sonuca tüm öz niteliklerin eklendiği alt küme 34 ile ulaşıldığı Çizelge 5.24'te görülmektedir.

Çizelge 5.24 : Özyinelemeli yöntemle oluşturulmuş öz nitelik önemleri – Lojistik regresyon sınıflayıcısı.

Lojistik Reg.	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	69,7586	25,4649	0,71	0,95	0,61	0,16	0,71
altküme_33	69,5137	23,7474	0,7	0,95	0,6	0,15	0,71
altküme_32	69,2951	21,2534	0,7	0,96	0,6	0,13	0,71
altküme_31	69,0749	19,3996	0,7	0,96	0,59	0,12	0,71
altküme_30	69,078	19,3947	0,7	0,96	0,59	0,12	0,71
altküme_29	69,098	19,3465	0,7	0,96	0,59	0,12	0,71
altküme_28	69,0241	19,056	0,7	0,96	0,59	0,11	0,71
altküme_27	69,0318	19,164	0,7	0,96	0,59	0,11	0,71
altküme_26	69,0287	19,1494	0,7	0,96	0,59	0,11	0,71

altküme_25	69,0256	19,1089	0,7	0,96	0,59	0,11	0,71
altküme_24	68,5467	15,4896	0,69	0,97	0,56	0,09	0,70
altküme_23	68,5467	15,4896	0,69	0,97	0,56	0,09	0,70
altküme_22	68,5822	15,7973	0,69	0,97	0,56	0,09	0,70
altküme_21	68,5883	15,8277	0,69	0,97	0,56	0,09	0,70
altküme_20	68,5067	15,3681	0,69	0,97	0,55	0,09	0,70
altküme_19	68,5113	15,349	0,69	0,97	0,55	0,09	0,70
altküme_18	68,4836	15,1621	0,69	0,97	0,55	0,09	0,70
altküme_17	68,5021	15,3451	0,69	0,97	0,55	0,09	0,70
altküme_16	68,4328	13,9301	0,69	0,97	0,55	0,08	0,70
altküme_15	68,2065	11,283	0,69	0,97	0,53	0,06	0,70
altküme_14	68,2049	11,2444	0,69	0,97	0,53	0,06	0,70
altküme_13	68,2111	11,2616	0,69	0,97	0,53	0,06	0,70
altküme_12	68,1988	11,2272	0,69	0,97	0,53	0,06	0,70
altküme_11	68,1988	11,2195	0,69	0,97	0,53	0,06	0,70
altküme_10	68,1957	11,2033	0,69	0,97	0,53	0,06	0,70
altküme_9	68,1834	11,1689	0,69	0,97	0,53	0,06	0,70
altküme_8	68,1895	11,1555	0,69	0,97	0,53	0,06	0,70
altküme_7	68,1895	11,1555	0,69	0,97	0,53	0,06	0,70
altküme_6	67,8277	6,145	0,68	0,98	0,47	0,03	0,66
altküme_5	67,8323	6,0868	0,68	0,98	0,47	0,03	0,66
altküme_4	67,8	3,7201	0,68	0,99	0,44	0,02	0,65
altküme_3	67,9632	0,048	0,68	1	0,71	0	0,64
altküme_2	67,9678	0,0769	0,68	1	0,8	0	0,64
altküme_1	67,9662	0,0672	0,68	1	0,78	0	0,64

Naive bayes yöntemiyle ilk alt küme ile en doğru sonuçlara ulaşmıştır. Bu durum üzerinde alt kümenin 5 öznitelik içermesi etkilidir. Diğer öznitelik seçim yöntemlerinde, ilk alt küme yüksek doğruluk değerine sahip olsa bile, kabul edilemez F ölçütü değerlerine sahip oluyordu. Yeni özniteliklerin gelmesiyle doğruluk oranında azalma gözlenmiştir. Sonuçlar Çizelge 5.25’te verilmiştir.

Çizelge 5.25 : Özyinelemeli yöntemle oluşturulmuş öznitelik önemleri – Naive Bayes sınıflayıcısı.

Naive Bayes	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	41,5281	51,9973	0,96	0,15	0,35	0,99	0,61
altküme_33	40,9091	51,9013	0,98	0,13	0,35	1	0,59
altküme_32	40,8475	51,8898	0,98	0,13	0,35	1	0,59
altküme_31	40,8475	51,8898	0,98	0,13	0,35	1	0,59
altküme_30	40,8352	51,8906	0,99	0,13	0,35	1	0,59
altküme_29	40,8244	51,8873	0,99	0,13	0,35	1	0,59
altküme_28	40,8244	51,8873	0,99	0,13	0,35	1	0,59
altküme_27	40,8198	51,8914	0,99	0,13	0,35	1	0,59
altküme_26	40,8814	51,9017	0,98	0,13	0,35	1	0,60
altküme_25	40,8814	51,9017	0,98	0,13	0,35	1	0,60
altküme_24	40,8583	51,8943	0,98	0,13	0,35	1	0,59

altküme_23	40,8537	51,8924	0,98	0,13	0,35	1	0,59
altküme_22	40,8552	51,893	0,98	0,13	0,35	1	0,59
altküme_21	40,9276	51,9128	0,98	0,13	0,35	1	0,61
altküme_20	40,926	51,9145	0,98	0,13	0,35	1	0,61
altküme_19	41,3248	51,9082	0,96	0,14	0,35	0,99	0,64
altküme_18	42,0532	51,965	0,94	0,16	0,35	0,98	0,64
altküme_17	41,9808	51,9498	0,94	0,16	0,35	0,98	0,65
altküme_16	41,9223	51,943	0,94	0,16	0,35	0,98	0,65
altküme_15	41,9346	51,9434	0,94	0,16	0,35	0,98	0,64
altküme_14	41,9361	51,9759	0,94	0,15	0,35	0,98	0,65
altküme_13	41,9361	51,9735	0,94	0,15	0,35	0,98	0,65
altküme_12	41,893	51,9635	0,94	0,15	0,35	0,98	0,65
altküme_11	41,8976	51,9655	0,94	0,15	0,35	0,98	0,65
altküme_10	41,8376	51,9458	0,95	0,15	0,35	0,98	0,65
altküme_9	41,7236	51,936	0,95	0,15	0,35	0,98	0,65
altküme_8	41,6513	51,9368	0,95	0,15	0,35	0,98	0,65
altküme_7	41,6359	51,9314	0,95	0,15	0,35	0,98	0,65
altküme_6	44,195	51,3752	0,85	0,22	0,36	0,92	0,63
altküme_5	44,0995	51,4042	0,85	0,21	0,36	0,92	0,63
altküme_4	44,0718	51,3801	0,85	0,21	0,36	0,92	0,62
altküme_3	44,0395	51,3643	0,85	0,21	0,36	0,92	0,60
altküme_2	44,0426	51,3774	0,85	0,21	0,36	0,92	0,60
altküme_1	44,5137	51,3481	0,85	0,22	0,36	0,91	0,60

Karar ağaçlarında, ilk alt küme ile de kabul edilebilir doğruluk oranlarına erişmiş olsa da, en iyi sonucu 5. alt kümesinde almaya başlayıp, bu değer yeni özniteliklerle, bilgi kazanımının artırılmasıyla, artırarak devam etmektedir. En iyi sonuca, en çok bilgi edilebilmiş tüm özniteliklere sahip alt küme 34’te ulaşılmıştır.

Çizelge 5.26 : Özyinelemeli yöntemle oluşturulmuş öznitelik önemleri – Karar ağaçları sınıflayıcısı.

Karar Ağaçları	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	67,5289	49,2076	0,76	0,76	0,49	0,49	0,63
altküme_33	67,4997	49,0034	0,76	0,76	0,49	0,49	0,63
altküme_32	66,9515	48,7255	0,75	0,76	0,49	0,49	0,62
altküme_31	67,0162	48,6245	0,76	0,76	0,49	0,49	0,62
altküme_30	67,1502	48,8761	0,76	0,76	0,49	0,49	0,62
altküme_29	66,8129	48,4292	0,75	0,76	0,49	0,48	0,62
altküme_28	67,0947	48,6471	0,76	0,76	0,49	0,49	0,62
altküme_27	67,0008	48,4249	0,76	0,76	0,48	0,48	0,62
altküme_26	66,913	48,3685	0,76	0,76	0,48	0,48	0,62
altküme_25	66,9977	48,6229	0,76	0,76	0,49	0,49	0,62
altküme_24	66,9069	48,4035	0,76	0,76	0,48	0,48	0,62
altküme_23	66,893	48,4351	0,76	0,76	0,49	0,48	0,62
altküme_22	66,8607	48,4033	0,76	0,76	0,49	0,48	0,62
altküme_21	66,7498	48,1736	0,75	0,76	0,48	0,48	0,62

altküme_20	66,913	48,4527	0,76	0,76	0,49	0,48	0,62
altküme_19	66,9115	48,4243	0,76	0,76	0,48	0,48	0,62
altküme_18	66,9977	48,581	0,76	0,76	0,49	0,49	0,62
altküme_17	66,7822	48,2897	0,75	0,76	0,48	0,48	0,62
altküme_16	67,0147	48,5617	0,76	0,76	0,49	0,49	0,62
altküme_15	66,8807	48,4283	0,76	0,76	0,49	0,48	0,62
altküme_14	66,6805	47,8541	0,76	0,76	0,48	0,48	0,62
altküme_13	66,5712	47,82	0,75	0,75	0,48	0,48	0,62
altküme_12	66,7591	48,0858	0,76	0,76	0,48	0,48	0,62
altküme_11	66,7683	48,2322	0,75	0,76	0,48	0,48	0,62
altküme_10	66,7991	48,2404	0,76	0,76	0,48	0,48	0,62
altküme_9	66,9608	48,516	0,76	0,76	0,49	0,49	0,62
altküme_8	66,9869	48,4665	0,76	0,76	0,48	0,48	0,62
altküme_7	66,8591	48,4343	0,75	0,76	0,49	0,48	0,62
altküme_6	66,4034	47,4987	0,75	0,75	0,47	0,47	0,61
altküme_5	66,4619	47,6129	0,75	0,75	0,48	0,48	0,61
altküme_4	64,8035	45,161	0,74	0,74	0,45	0,45	0,60
altküme_3	62,9573	41,7633	0,73	0,73	0,41	0,42	0,57
altküme_2	63,3931	41,9069	0,74	0,73	0,41	0,42	0,58
altküme_1	63,3053	41,9224	0,74	0,73	0,41	0,42	0,58

K en yakın komşu algoritmasıyla, yine ilk alt küme ile de düzgün sonuçlar alınmış olsa da alt küme 7 için en yüksek değerine ulaşmış ve bu değeri altküme 28'e kadar koruyabilmiştir. Normalde çok kırılgan olup ulaştığı en yüksek doğruluktan, yeni öznitelikler ile hızlıca uzaklaşan bu model, bu öznitelik sıralaması ile en yüksek doğruluğunu uzun süre devam ettirebilmiştir. Değerler Çizelge 5.27'de verilmiştir.

Çizelge 5.27 : Özyinelemeli yöntemle oluşturulmuş öznitelik önemleri – K en yakın komşu sınıflayıcısı.

K En Yakın Komşu	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	64,2569	36,5088	0,71	0,79	0,42	0,32	0,61
altküme_33	64,3308	36,7716	0,71	0,79	0,43	0,32	0,62
altküme_32	65,7921	39,7745	0,72	0,8	0,46	0,35	0,64
altküme_31	66,8715	41,8529	0,73	0,81	0,48	0,37	0,65
altküme_30	66,9669	41,9609	0,73	0,81	0,48	0,37	0,66
altküme_29	67,6013	43,2352	0,74	0,81	0,49	0,39	0,66
altküme_28	67,9493	43,8661	0,74	0,82	0,5	0,39	0,67
altküme_27	68,1141	44,1652	0,74	0,82	0,5	0,39	0,67
altküme_26	68,1141	44,1652	0,74	0,82	0,5	0,39	0,67
altküme_25	68,1834	44,3571	0,74	0,82	0,5	0,4	0,67
altküme_24	68,3912	44,8765	0,74	0,82	0,51	0,4	0,68
altküme_23	68,3897	44,8723	0,74	0,82	0,51	0,4	0,68
altküme_22	68,4082	44,8986	0,74	0,82	0,51	0,4	0,68
altküme_21	68,4097	44,8998	0,74	0,82	0,51	0,4	0,68
altküme_20	68,3928	44,8273	0,74	0,82	0,51	0,4	0,68
altküme_19	68,3928	44,8273	0,74	0,82	0,51	0,4	0,68

altküme_18	68,402	44,8612	0,74	0,82	0,51	0,4	0,68
altküme_17	68,422	44,8057	0,74	0,82	0,51	0,4	0,68
altküme_16	68,5421	45,0659	0,74	0,82	0,51	0,4	0,68
altküme_15	68,4482	45,0629	0,74	0,82	0,51	0,4	0,68
altküme_14	68,4374	45,0456	0,74	0,82	0,51	0,4	0,68
altküme_13	68,4374	45,0456	0,74	0,82	0,51	0,4	0,68
altküme_12	68,7731	45,8015	0,75	0,82	0,52	0,41	0,68
altküme_11	68,7731	45,8015	0,75	0,82	0,52	0,41	0,68
altküme_10	68,7731	45,7725	0,75	0,82	0,52	0,41	0,68
altküme_9	68,7762	45,772	0,75	0,82	0,52	0,41	0,68
altküme_8	68,7762	45,772	0,75	0,82	0,52	0,41	0,68
altküme_7	68,7762	45,772	0,75	0,82	0,52	0,41	0,68
altküme_6	68,5775	45,1999	0,74	0,82	0,51	0,4	0,68
altküme_5	68,5083	44,4028	0,74	0,82	0,51	0,39	0,68
altküme_4	67,8092	42,1623	0,73	0,83	0,5	0,37	0,65
altküme_3	66,6913	39,1231	0,72	0,82	0,47	0,33	0,63
altküme_2	67,0039	39,6417	0,73	0,83	0,48	0,34	0,64
altküme_1	66,9731	39,5888	0,73	0,83	0,48	0,34	0,64

Gradyen boosting yöntemde, altküme 5 ile ulaştığı doğruluk değeri ile altküme 34 ile ulaştığı doğruluk arasında ciddi bir fark bulunmadığı Çizelge 5.28 incelendiğinde görülmüştür. Önemli derecede bilgi edinilen özniteliklerin büyük bir kısmı ilk 10 öznitelik içerisinde olduğunu göstermektedir. (5 tanesi ilk alt kümeden gelen öznitelikler + 5 sırasıyla diğer alt kümelerden gelen öznitelikler)

Çizelge 5.28 : Özyinelemeli yöntemle oluşturulmuş öznitelik önemleri – Gradyen boosting sınıflayıcısı.

Gradyen Boosting	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	73,8929	49,9454	0,76	0,9	0,65	0,41	0,78
altküme_33	73,9283	50,0825	0,76	0,9	0,65	0,41	0,78
altküme_32	73,8929	49,9779	0,76	0,9	0,65	0,41	0,78
altküme_31	73,9499	50,0502	0,76	0,9	0,65	0,41	0,78
altküme_30	74,0145	50,1977	0,76	0,9	0,65	0,41	0,78
altküme_29	73,9083	50,0398	0,76	0,9	0,65	0,41	0,78
altküme_28	73,9129	50,0825	0,76	0,9	0,65	0,41	0,78
altküme_27	73,9576	50,1165	0,76	0,9	0,65	0,41	0,78
altküme_26	73,9576	50,1165	0,76	0,9	0,65	0,41	0,78
altküme_25	73,9206	50,175	0,76	0,89	0,65	0,41	0,78
altküme_24	73,8821	50,1059	0,76	0,89	0,65	0,41	0,78
altküme_23	73,8821	50,1059	0,76	0,89	0,65	0,41	0,78
altküme_22	73,8698	50,1117	0,76	0,89	0,65	0,41	0,78
altküme_21	73,8698	50,1117	0,76	0,89	0,65	0,41	0,78
altküme_20	73,8698	50,1117	0,76	0,89	0,65	0,41	0,78
altküme_19	73,8698	50,1117	0,76	0,89	0,65	0,41	0,78
altküme_18	73,8975	50,2407	0,76	0,89	0,65	0,41	0,78
altküme_17	73,8621	50,2156	0,76	0,89	0,64	0,41	0,78

altküme_16	73,722	50,0322	0,76	0,89	0,64	0,41	0,77
altküme_15	73,6404	49,855	0,76	0,89	0,64	0,41	0,77
altküme_14	73,7605	50,1069	0,76	0,89	0,64	0,41	0,77
altküme_13	73,7605	50,1069	0,76	0,89	0,64	0,41	0,77
altküme_12	73,7235	50,2754	0,76	0,89	0,64	0,41	0,77
altküme_11	73,7866	50,2746	0,76	0,89	0,64	0,41	0,77
altküme_10	73,7866	50,2746	0,76	0,89	0,64	0,41	0,77
altküme_9	73,7543	49,9221	0,76	0,89	0,64	0,41	0,77
altküme_8	73,588	49,6404	0,76	0,89	0,64	0,41	0,77
altküme_7	73,5849	48,9814	0,76	0,9	0,64	0,4	0,77
altküme_6	73,4233	47,7919	0,76	0,9	0,64	0,38	0,77
altküme_5	73,4325	47,9109	0,76	0,9	0,64	0,38	0,77
altküme_4	71,8958	44,1254	0,74	0,89	0,61	0,35	0,74
altküme_3	69,9326	31,3469	0,71	0,93	0,58	0,21	0,70
altküme_2	69,9695	29,2421	0,71	0,94	0,6	0,19	0,70
altküme_1	69,9618	29,3649	0,71	0,94	0,6	0,19	0,70

Adaboost algoritmasıyla da yine ilk alt kümeden itibaren makul sonuçlar elde edilmiş olsa da özellikle 5. alt küme den itibaren yüksek değerler elde edilmiş, doğruluk artarak devam etmiş ve en son alt küme ile en iyi doğruluk oranına ulaşmıştır.

Çizelge 5.29 : Özyinelemeli yöntemle oluşturulmuş öznelik önemleri – AdaBoost sınıflayıcısı.

AdaBoost	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	73,0121	49,0153	0,76	0,88	0,62	0,4	0,77
altküme_33	72,9736	49,0863	0,76	0,88	0,62	0,41	0,76
altküme_32	72,9505	49,1269	0,76	0,88	0,62	0,41	0,76
altküme_31	72,9552	49,1488	0,76	0,88	0,62	0,41	0,76
altküme_30	72,9736	49,0449	0,76	0,88	0,62	0,41	0,76
altküme_29	72,9736	49,0449	0,76	0,88	0,62	0,41	0,76
altküme_28	72,9567	48,8183	0,76	0,88	0,62	0,4	0,76
altküme_27	72,9567	48,8183	0,76	0,88	0,62	0,4	0,76
altküme_26	72,9567	48,8183	0,76	0,88	0,62	0,4	0,76
altküme_25	72,949	48,8321	0,76	0,88	0,62	0,4	0,76
altküme_24	72,9983	48,8776	0,76	0,88	0,62	0,4	0,76
altküme_23	72,9043	48,4971	0,76	0,89	0,62	0,4	0,76
altküme_22	72,9182	48,735	0,76	0,88	0,62	0,4	0,76
altküme_21	72,9182	48,735	0,76	0,88	0,62	0,4	0,76
altküme_20	72,9182	48,735	0,76	0,88	0,62	0,4	0,76
altküme_19	72,9182	48,735	0,76	0,88	0,62	0,4	0,76
altküme_18	72,8643	48,4572	0,76	0,88	0,62	0,4	0,76
altküme_17	72,8643	48,4572	0,76	0,88	0,62	0,4	0,76
altküme_16	72,815	48,1086	0,76	0,89	0,62	0,39	0,76
altküme_15	72,775	47,9864	0,76	0,89	0,62	0,39	0,76
altküme_14	72,775	47,9864	0,76	0,89	0,62	0,39	0,76
altküme_13	72,775	47,9864	0,76	0,89	0,62	0,39	0,76
altküme_12	72,7642	47,955	0,76	0,89	0,62	0,39	0,76

altküme_11	72,7642	47,955	0,76	0,89	0,62	0,39	0,76
altküme_10	72,7642	47,955	0,76	0,89	0,62	0,39	0,76
altküme_9	72,6765	47,7458	0,75	0,89	0,62	0,39	0,76
altküme_8	72,7088	47,8737	0,76	0,89	0,62	0,39	0,76
altküme_7	72,7919	47,7404	0,75	0,89	0,62	0,39	0,76
altküme_6	72,1622	44,2917	0,74	0,9	0,62	0,35	0,74
altküme_5	72,0806	44,312	0,74	0,9	0,61	0,35	0,74
altküme_4	71,2814	40,6605	0,73	0,9	0,6	0,31	0,72
altküme_3	69,9279	30,5526	0,71	0,93	0,59	0,21	0,70
altküme_2	69,8694	27,8785	0,71	0,94	0,6	0,18	0,70
altküme_1	69,7971	31,332	0,71	0,93	0,58	0,22	0,70

Yerine koyma yöntemiyle 5. alt küme’den itibaren yüksek değerler elde edilmiş, doğruluk artarak devam etmiş ve en son alt küme ile en iyi doğruluk oranına ulaşmıştır.

Çizelge 5.30 : Özyinelemeli yöntemle oluşturulmuş öznelik önemleri – Bagging sınıflayıcısı.

Bagging	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	73,3509	49,5731	0,76	0,89	0,63	0,41	0,75
altküme_33	73,2831	49,4332	0,76	0,89	0,63	0,41	0,75
altküme_32	72,6395	48,7349	0,76	0,88	0,61	0,41	0,74
altküme_31	72,4363	48,4492	0,76	0,88	0,6	0,4	0,74
altküme_30	72,8351	49,1497	0,76	0,88	0,61	0,41	0,75
altküme_29	72,4301	48,6507	0,76	0,87	0,6	0,41	0,74
altküme_28	72,3978	48,6125	0,76	0,87	0,6	0,41	0,74
altküme_27	72,4593	48,7066	0,76	0,87	0,6	0,41	0,74
altküme_26	72,4686	48,659	0,76	0,87	0,6	0,41	0,74
altküme_25	72,4147	48,5157	0,76	0,87	0,6	0,41	0,74
altküme_24	72,29	48,6181	0,76	0,87	0,6	0,41	0,74
altküme_23	72,2746	48,5043	0,76	0,87	0,6	0,41	0,74
altküme_22	72,3516	48,6942	0,76	0,87	0,6	0,41	0,74
altküme_21	72,31	48,3262	0,76	0,87	0,6	0,4	0,74
altküme_20	72,2623	48,4489	0,76	0,87	0,6	0,41	0,74
altküme_19	72,2746	48,3269	0,76	0,87	0,6	0,4	0,74
altküme_18	72,407	48,7678	0,76	0,87	0,6	0,41	0,74
altküme_17	72,3069	48,5069	0,76	0,87	0,6	0,41	0,74
altküme_16	72,1268	48,4567	0,76	0,87	0,59	0,41	0,74
altküme_15	72,0528	48,2021	0,76	0,87	0,59	0,41	0,73
altküme_14	72,0852	48,29	0,76	0,87	0,59	0,41	0,74
altküme_13	72,2114	48,3796	0,76	0,87	0,6	0,41	0,74
altküme_12	72,1021	48,3406	0,76	0,87	0,59	0,41	0,74
altküme_11	72,0097	48,3785	0,76	0,87	0,59	0,41	0,73
altküme_10	71,8865	47,9948	0,76	0,87	0,59	0,4	0,73
altküme_9	72,0975	48,242	0,76	0,87	0,59	0,41	0,74
altküme_8	72,0575	48,2623	0,76	0,87	0,59	0,41	0,74
altküme_7	71,9728	48,1543	0,76	0,87	0,59	0,41	0,73
altküme_6	71,6309	47,3148	0,75	0,87	0,58	0,4	0,73

altküme_5	71,4431	47,2616	0,75	0,86	0,58	0,4	0,72
altküme_4	69,9587	43,5703	0,74	0,86	0,55	0,36	0,69
altküme_3	67,1163	39,5459	0,73	0,83	0,48	0,34	0,65
altküme_2	66,8022	39,8337	0,73	0,82	0,48	0,34	0,65
altküme_1	66,9638	40,0849	0,73	0,82	0,48	0,34	0,65

Rastgele ormanlar algoritmasıyla 5. alt kümeden itibaren yüksek değerler elde edilmiş, doğruluk artarak devam etmiş ve en son alt küme ile en iyi doğruluk oranına ulaşmıştır.

Çizelge 5.31 : Özyinelemeli yöntemle oluşturulmuş öznelilik önemleri – Rastgele orman sınıflayıcısı.

Rast. Orman Sınıflayıcısı	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	72,1191	45,4888	0,75	0,89	0,61	0,36	0,74
altküme_33	71,9974	45,4398	0,75	0,89	0,6	0,36	0,73
altküme_32	71,9297	45,9948	0,75	0,88	0,6	0,37	0,73
altküme_31	71,5386	45,6257	0,75	0,88	0,59	0,37	0,73
altküme_30	71,5724	45,7638	0,75	0,88	0,59	0,37	0,73
altküme_29	71,6771	45,8395	0,75	0,88	0,59	0,37	0,73
altküme_28	71,7618	46,4945	0,75	0,88	0,59	0,38	0,73
altküme_27	72,0005	46,8553	0,75	0,88	0,6	0,39	0,73
altküme_26	71,7495	46,4649	0,75	0,88	0,59	0,38	0,73
altküme_25	71,5724	46,4279	0,75	0,87	0,59	0,38	0,73
altküme_24	71,5817	46,5539	0,75	0,87	0,59	0,39	0,73
altküme_23	71,6294	46,5338	0,75	0,87	0,59	0,39	0,73
altküme_22	71,6818	46,7468	0,75	0,87	0,59	0,39	0,73
altküme_21	71,6818	46,5487	0,75	0,87	0,59	0,38	0,73
altküme_20	71,3107	46,0067	0,75	0,87	0,58	0,38	0,73
altküme_19	71,6664	46,5972	0,75	0,87	0,59	0,39	0,73
altküme_18	71,9789	47,2797	0,75	0,87	0,6	0,39	0,73
altküme_17	71,7911	46,9078	0,75	0,87	0,59	0,39	0,73
altküme_16	71,8296	47,2569	0,75	0,87	0,59	0,39	0,73
altküme_15	71,8311	47,3402	0,75	0,87	0,59	0,4	0,73
altküme_14	71,731	47,1242	0,75	0,87	0,59	0,39	0,73
altküme_13	71,6494	46,915	0,75	0,87	0,59	0,39	0,73
altküme_12	71,6925	47,0537	0,75	0,87	0,59	0,39	0,73
altküme_11	71,6741	47,2924	0,75	0,87	0,59	0,4	0,73
altküme_10	71,8989	47,3942	0,75	0,87	0,59	0,4	0,73
altküme_9	72,0205	47,8459	0,75	0,87	0,59	0,4	0,73
altküme_8	72,0159	47,7067	0,75	0,87	0,59	0,4	0,73
altküme_7	71,8203	47,3519	0,75	0,87	0,59	0,4	0,73
altküme_6	71,6756	46,8552	0,75	0,87	0,59	0,39	0,73
altküme_5	71,7603	46,9298	0,75	0,87	0,59	0,39	0,72
altküme_4	70,3113	43,3695	0,74	0,87	0,56	0,35	0,69
altküme_3	67,1532	38,9398	0,72	0,83	0,48	0,33	0,65
altküme_2	66,8129	39,4017	0,72	0,82	0,47	0,34	0,65
altküme_1	66,8037	39,0765	0,72	0,83	0,47	0,33	0,65

YSA ile 5. alt küme’den itibaren yüksek değerler elde edilmiş, doğruluk artarak devam etmiş ve en iyi doğruluk oranına alt küme 29 ile ulaşılmıştır.

Çizelge 5.32 : Özyinelemeli yöntemle oluşturulmuş öznitelik önemleri – YSA sınıflayıcısı.

YSA Sınıflayıcısı	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	72,27	44,415	0,74	0,9	0,62	0,35	0,75
altküme_33	71,6617	52,4592	0,77	0,82	0,57	0,49	0,75
altküme_32	73,0722	49,8221	0,76	0,88	0,62	0,42	0,76
altküme_31	72,8335	53,1133	0,78	0,85	0,59	0,48	0,76
altküme_30	73,2939	47,7779	0,75	0,9	0,64	0,38	0,76
altküme_29	73,4602	51,3107	0,77	0,88	0,62	0,44	0,76
altküme_28	73,3309	47,7463	0,75	0,9	0,64	0,38	0,76
altküme_27	73,1907	44,2295	0,75	0,92	0,66	0,33	0,76
altküme_26	73,086	47,4805	0,75	0,9	0,63	0,38	0,76
altküme_25	73,2739	51,4937	0,77	0,87	0,62	0,44	0,76
altküme_24	73,3416	48,7917	0,76	0,89	0,63	0,4	0,76
altküme_23	72,8582	44,7897	0,75	0,91	0,64	0,34	0,76
altküme_22	73,2631	48,2475	0,76	0,89	0,64	0,39	0,76
altküme_21	73,3232	48,1023	0,76	0,9	0,64	0,39	0,76
altküme_20	73,2015	47,2925	0,75	0,9	0,64	0,38	0,76
altküme_19	73,3109	47,5441	0,75	0,9	0,64	0,38	0,76
altküme_18	73,2323	44,4849	0,75	0,92	0,66	0,33	0,77
altküme_17	73,1923	44,7968	0,75	0,92	0,66	0,34	0,76
altküme_16	73,1892	48,2648	0,76	0,89	0,63	0,39	0,77
altküme_15	72,8797	43,4302	0,74	0,92	0,65	0,32	0,76
altküme_14	73,0168	44,9727	0,75	0,91	0,65	0,34	0,76
altküme_13	72,6149	39,3604	0,73	0,94	0,68	0,28	0,76
altküme_12	73,3293	51,2675	0,77	0,87	0,62	0,44	0,76
altküme_11	73,0214	52,5703	0,77	0,85	0,6	0,47	0,76
altküme_10	73,4063	51,9141	0,77	0,87	0,62	0,45	0,77
altküme_9	73,1769	44,9814	0,75	0,92	0,66	0,34	0,77
altküme_8	73,2477	47,9103	0,76	0,9	0,64	0,38	0,77
altküme_7	73,1615	51,2584	0,77	0,87	0,61	0,44	0,76
altküme_6	73,0229	45,5088	0,75	0,91	0,64	0,35	0,76
altküme_5	73,1415	48,9179	0,76	0,89	0,63	0,4	0,76
altküme_4	70,7117	35,1947	0,72	0,92	0,6	0,25	0,72
altküme_3	69,4444	18,0068	0,7	0,97	0,64	0,1	0,70
altküme_2	69,6831	31,7752	0,71	0,92	0,57	0,22	0,70
altküme_1	69,4814	20,4591	0,7	0,96	0,62	0,12	0,69

XGB sınıflayıcı, 5. alt küme’den itibaren yüksek değerler elde edilmiş, doğruluk artarak devam etmiş ve en iyi doğruluk oranına tüm alt küme öznitelikleri ile ulaşılmıştır. Değerler Çizelge 5.33’te verilmiştir.

Çizelge 5.33 : Özyinelemeli yöntemle oluşturulmuş öznitelik önemleri – XGB sınıflayıcısı.

XGB Sınıflayıcısı	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	73,9745	49,7861	0,76	0,9	0,65	0,4	0,78
altküme_33	73,9529	49,8696	0,76	0,9	0,65	0,4	0,78
altküme_32	73,9699	50,0044	0,76	0,9	0,65	0,41	0,78
altküme_31	73,9591	49,7594	0,76	0,9	0,65	0,4	0,78
altküme_30	73,9221	49,6522	0,76	0,9	0,65	0,4	0,78
altküme_29	73,9068	49,6703	0,76	0,9	0,65	0,4	0,78
altküme_28	73,9021	49,645	0,76	0,9	0,65	0,4	0,78
altküme_27	73,9191	49,5863	0,76	0,9	0,65	0,4	0,78
altküme_26	73,9191	49,5863	0,76	0,9	0,65	0,4	0,78
altküme_25	73,9776	49,7831	0,76	0,9	0,65	0,4	0,78
altküme_24	73,8529	49,6158	0,76	0,9	0,65	0,4	0,77
altküme_23	73,8529	49,6158	0,76	0,9	0,65	0,4	0,77
altküme_22	73,8837	49,7616	0,76	0,9	0,65	0,4	0,78
altküme_21	73,8837	49,7616	0,76	0,9	0,65	0,4	0,78
altküme_20	73,8837	49,7616	0,76	0,9	0,65	0,4	0,78
altküme_19	73,8837	49,7556	0,76	0,9	0,65	0,4	0,78
altküme_18	73,8867	49,901	0,76	0,9	0,65	0,41	0,78
altküme_17	73,8821	49,9941	0,76	0,9	0,65	0,41	0,78
altküme_16	73,6111	49,6977	0,76	0,89	0,64	0,41	0,77
altküme_15	73,645	48,9471	0,76	0,9	0,65	0,39	0,77
altküme_14	73,702	49,5078	0,76	0,89	0,64	0,4	0,77
altküme_13	73,702	49,5078	0,76	0,89	0,64	0,4	0,77
altküme_12	73,6989	49,5883	0,76	0,89	0,64	0,4	0,77
altküme_11	73,7112	49,4088	0,76	0,9	0,64	0,4	0,77
altküme_10	73,6881	49,5423	0,76	0,89	0,64	0,4	0,77
altküme_9	73,6203	49,4989	0,76	0,89	0,64	0,4	0,77
altküme_8	73,5495	49,0055	0,76	0,9	0,64	0,4	0,77
altküme_7	73,4279	48,6078	0,76	0,9	0,64	0,39	0,77
altküme_6	73,3709	47,4985	0,75	0,9	0,64	0,38	0,77
altküme_5	73,3463	47,4818	0,75	0,9	0,64	0,38	0,76
altküme_4	71,8527	44,0773	0,74	0,89	0,61	0,35	0,74
altküme_3	70,0219	29,4831	0,71	0,94	0,6	0,2	0,70
altküme_2	69,9048	28,5975	0,71	0,94	0,6	0,19	0,70
altküme_1	69,8941	28,9277	0,71	0,94	0,59	0,19	0,70

Extra trees sınıflayıcı yönteminde de görülmüştür ki 5. alt küme'den itibaren iyi doğruluk oranına ulaşmış, en iyi sonuca 6. altküme ile ulaşmıştır. Ancak bagging yöntem olmasına rağmen 17. alt kümeden itibaren doğruluk oranı düşüşe geçmiştir. Modelin altkümeler ile çalışmasıyla ürettiği sonuçlar Çizelge 5.34'te verilmiştir.

Çizelge 5.34 : Özyinelemeli yöntemle oluşturulmuş öznitelik önemleri – Extra trees
sınıflayıcısı.

Extra Trees Sınıflayıcısı	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	69,4506	39,6484	0,73	0,87	0,54	0,31	0,69
altküme_33	69,252	39,7089	0,73	0,87	0,53	0,32	0,68
altküme_32	69,1996	40,8248	0,73	0,86	0,53	0,33	0,69
altküme_31	69,2012	41,6987	0,73	0,86	0,53	0,34	0,69
altküme_30	69,3012	41,7405	0,73	0,86	0,53	0,34	0,69
altküme_29	69,466	42,3412	0,74	0,86	0,54	0,35	0,69
altküme_28	69,5492	42,834	0,74	0,86	0,54	0,36	0,69
altküme_27	69,2304	42,3706	0,74	0,85	0,53	0,35	0,69
altküme_26	69,2766	42,3807	0,74	0,85	0,53	0,35	0,69
altküme_25	69,4568	43,1601	0,74	0,85	0,53	0,36	0,69
altküme_24	69,4029	43,0516	0,74	0,85	0,53	0,36	0,70
altküme_23	69,6508	43,466	0,74	0,85	0,54	0,36	0,70
altküme_22	69,6754	43,1795	0,74	0,86	0,54	0,36	0,70
altküme_21	69,8048	43,487	0,74	0,86	0,54	0,36	0,70
altküme_20	69,7555	43,4144	0,74	0,86	0,54	0,36	0,70
altküme_19	69,6338	43,3646	0,74	0,85	0,54	0,36	0,70
altküme_18	69,7108	43,492	0,74	0,85	0,54	0,36	0,70
altküme_17	70,0434	43,7245	0,74	0,86	0,55	0,36	0,70
altküme_16	70,3036	44,424	0,74	0,86	0,55	0,37	0,71
altküme_15	70,1358	43,9661	0,74	0,86	0,55	0,37	0,70
altküme_14	70,1404	44,186	0,74	0,86	0,55	0,37	0,71
altküme_13	70,2066	44,2215	0,74	0,86	0,55	0,37	0,71
altküme_12	70,2236	44,2355	0,74	0,86	0,55	0,37	0,71
altküme_11	70,3637	44,5156	0,74	0,86	0,56	0,37	0,71
altküme_10	70,4515	44,8912	0,74	0,86	0,56	0,38	0,71
altküme_9	70,4207	44,9065	0,74	0,86	0,56	0,38	0,71
altküme_8	70,6501	45,1466	0,75	0,86	0,56	0,38	0,71
altküme_7	70,5885	45,3898	0,75	0,86	0,56	0,38	0,71
altküme_6	70,7764	45,6827	0,75	0,86	0,56	0,38	0,71
altküme_5	70,7533	45,3001	0,75	0,86	0,57	0,38	0,71
altküme_4	69,0087	41,0681	0,73	0,86	0,53	0,34	0,68
altküme_3	66,0708	38,4583	0,72	0,82	0,46	0,33	0,63
altküme_2	66,1293	39,3437	0,72	0,81	0,46	0,34	0,63
altküme_1	66,0354	39,3778	0,72	0,81	0,46	0,34	0,63

5.3.4 Adaboost ile özyinelemeli yöntem ile oluşturulmuş alt kümeler

Lojistik regresyon, ilk alt küme ile yüksek sonuçlara ulaşmış ve yeni özniteliklerle doğruluk oranını artırmaya devam etmiştir. En iyi sonuca alt küme 30 ile ulaşılmıştır.

Çizelge 5.35 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznitelik önemleri –
Lojistik regresyon sınıflayıcısı.

Lojistik Reg. Sınıflayıcısı	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	70,1158	27,9049	0,71	0,95	0,61	0,18	0,72
altküme_33	70,1266	27,9122	0,71	0,95	0,62	0,18	0,72
altküme_32	70,1297	27,9142	0,71	0,95	0,62	0,18	0,72
altküme_31	70,1266	27,9122	0,71	0,95	0,62	0,18	0,72
altküme_30	70,1728	28,0397	0,71	0,95	0,62	0,18	0,72
altküme_29	70,1666	27,982	0,71	0,95	0,62	0,18	0,72
altküme_28	70,1712	27,9958	0,71	0,95	0,62	0,18	0,72
altküme_27	70,1712	27,9958	0,71	0,95	0,62	0,18	0,72
altküme_26	70,0927	27,7982	0,71	0,95	0,61	0,18	0,72
altküme_25	70,0804	27,7094	0,71	0,95	0,61	0,18	0,72
altküme_24	70,1481	27,8838	0,71	0,95	0,62	0,18	0,72
altküme_23	70,0989	27,6571	0,71	0,95	0,62	0,18	0,72
altküme_22	70,105	27,672	0,71	0,95	0,62	0,18	0,72
altküme_21	70,1019	27,6807	0,71	0,95	0,62	0,18	0,72
altküme_20	70,1019	27,6807	0,71	0,95	0,62	0,18	0,72
altküme_19	70,1404	27,8357	0,71	0,95	0,62	0,18	0,72
altküme_18	70,1481	27,8945	0,71	0,95	0,62	0,18	0,72
altküme_17	69,7478	24,2511	0,7	0,96	0,61	0,15	0,71
altküme_16	69,7539	24,284	0,7	0,96	0,61	0,15	0,71
altküme_15	69,7447	24,2609	0,7	0,95	0,61	0,15	0,71
altküme_14	69,737	24,2036	0,7	0,96	0,61	0,15	0,71
altküme_13	69,6785	23,6152	0,7	0,96	0,61	0,15	0,71
altküme_12	69,6277	23,1084	0,7	0,96	0,61	0,14	0,71
altküme_11	69,6277	23,0184	0,7	0,96	0,61	0,14	0,71
altküme_10	69,6	22,912	0,7	0,96	0,61	0,14	0,71
altküme_9	69,4183	20,5083	0,7	0,96	0,61	0,12	0,71
altküme_8	69,4183	20,6544	0,7	0,96	0,61	0,12	0,71
altküme_7	69,4152	20,6083	0,7	0,96	0,61	0,12	0,71
altküme_6	69,3598	20,1068	0,7	0,96	0,61	0,12	0,71
altküme_5	69,098	18,1024	0,7	0,97	0,6	0,11	0,70
altküme_4	68,8593	16,1804	0,69	0,97	0,59	0,09	0,70
altküme_3	68,8485	16,1687	0,69	0,97	0,59	0,09	0,70
altküme_2	68,8778	16,3272	0,69	0,97	0,59	0,09	0,70
altküme_1	68,0817	9,4527	0,69	0,98	0,52	0,05	0,70

Naive Bayes, en iyi doğruluk sonucunu 10 öznitelik içeren altküme ile vermiştir. Diğer özniteliklerin gelmesiyle, doğruluk oranında düşüş gözlenmiştir.

Çizelge 5.36 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznitelik önemleri –
Naive Bayes sınıflandırıcısı.

Naive Bayes	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
-------------	----------	-------------	------------	-------------	------------	-------------	-----

altküme_34	41,545	52,0045	0,96	0,15	0,35	0,99	0,61
altküme_33	43,4759	52,1713	0,91	0,19	0,36	0,96	0,64
altküme_32	43,0864	52,0709	0,92	0,18	0,36	0,96	0,64
altküme_31	42,9662	52,0307	0,92	0,18	0,36	0,97	0,64
altküme_30	42,9047	52,0298	0,92	0,18	0,36	0,97	0,64
altküme_29	43,0186	52,0436	0,92	0,18	0,36	0,96	0,64
altküme_28	44,3813	52,0121	0,88	0,21	0,36	0,94	0,64
altküme_27	44,3767	52,0087	0,88	0,21	0,36	0,94	0,64
altküme_26	44,2166	51,9879	0,88	0,21	0,36	0,94	0,64
altküme_25	44,2166	51,9879	0,88	0,21	0,36	0,94	0,64
altküme_24	44,2196	51,9931	0,88	0,21	0,36	0,94	0,64
altküme_23	48,7605	51,5738	0,82	0,32	0,37	0,85	0,63
altküme_22	48,3601	51,6556	0,82	0,31	0,37	0,86	0,63
altküme_21	48,3694	51,6698	0,82	0,31	0,37	0,86	0,63
altküme_20	48,3694	51,6698	0,82	0,31	0,37	0,86	0,63
altküme_19	48,3278	51,6455	0,82	0,31	0,37	0,86	0,63
altküme_18	48,337	51,6277	0,82	0,31	0,37	0,86	0,63
altküme_17	48,7636	51,6892	0,82	0,31	0,37	0,86	0,63
altküme_16	56,2808	50,596	0,78	0,5	0,4	0,7	0,63
altküme_15	56,8012	50,5787	0,78	0,51	0,4	0,69	0,64
altküme_14	58,281	49,4043	0,76	0,56	0,4	0,64	0,64
altküme_13	58,2548	49,3281	0,76	0,56	0,4	0,63	0,64
altküme_12	58,2625	49,3724	0,76	0,56	0,4	0,64	0,64
altküme_11	58,9585	48,6178	0,76	0,58	0,41	0,61	0,64
altküme_10	59,3111	48,2826	0,76	0,59	0,41	0,59	0,64
altküme_9	54,0589	52,5449	0,81	0,42	0,39	0,79	0,65
altküme_8	53,657	52,5845	0,82	0,41	0,39	0,8	0,65
altküme_7	53,1258	52,5293	0,82	0,4	0,39	0,81	0,65
altküme_6	52,4513	52,7511	0,82	0,38	0,39	0,83	0,65
altküme_5	41,9592	52,1638	0,96	0,15	0,35	0,99	0,66
altküme_4	41,7267	52,1246	0,97	0,15	0,35	0,99	0,66
altküme_3	41,6451	52,0982	0,97	0,15	0,35	0,99	0,66
altküme_2	41,6528	52,1003	0,97	0,15	0,35	0,99	0,66
altküme_1	41,4927	52,0585	0,97	0,14	0,35	0,99	0,66

Karar ağaçları, lojistik regresyona göre daha düşük doğruluk oranına ulaştığı Çizelge 5.37 incelendiğinde de görülmüştür. Karar ağacı yöntemiyle ulaşılabilen en iyi sonuç 30 öznitelik ile 67.91 değeri olduğu gözlenmiştir.

Çizelge 5.37 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznitelik önemleri –
Karar ağaçları sınıflandırıcısı.

Karar Ağaçları	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	67,7799	49,5187	0,76	0,76	0,49	0,5	0,63
altküme_33	67,6983	49,509	0,76	0,76	0,49	0,5	0,63
altküme_32	67,5967	49,2426	0,76	0,76	0,49	0,49	0,63
altküme_31	67,7122	49,51	0,76	0,76	0,49	0,5	0,63
altküme_30	67,9108	49,8605	0,76	0,76	0,5	0,5	0,63

altküme_29	67,6583	49,3098	0,76	0,76	0,49	0,49	0,63
altküme_28	67,7307	49,3633	0,77	0,76	0,49	0,49	0,63
altküme_27	67,8261	49,6494	0,76	0,76	0,5	0,5	0,63
altküme_26	67,686	49,3337	0,76	0,76	0,49	0,49	0,63
altküme_25	67,7276	49,4269	0,76	0,76	0,49	0,49	0,63
altküme_24	67,6906	49,3543	0,76	0,76	0,49	0,49	0,63
altküme_23	67,5628	49,0667	0,76	0,76	0,49	0,49	0,63
altküme_22	67,6275	49,1412	0,76	0,76	0,49	0,49	0,63
altküme_21	67,5782	49,1229	0,76	0,76	0,49	0,49	0,63
altküme_20	67,626	49,3044	0,76	0,76	0,49	0,49	0,63
altküme_19	67,5705	49,085	0,76	0,76	0,49	0,49	0,63
altküme_18	67,5567	49,1407	0,76	0,76	0,49	0,49	0,63
altküme_17	67,5751	49,177	0,76	0,76	0,49	0,49	0,63
altküme_16	67,5921	49,1041	0,76	0,76	0,49	0,49	0,63
altküme_15	67,7553	49,213	0,77	0,76	0,49	0,49	0,63
altküme_14	67,5289	49,0899	0,76	0,76	0,49	0,49	0,63
altküme_13	67,7291	49,3254	0,77	0,76	0,49	0,49	0,63
altküme_12	67,6398	49,1113	0,77	0,76	0,49	0,49	0,63
altküme_11	67,5536	49,0103	0,76	0,76	0,49	0,49	0,63
altküme_10	67,6367	49,2539	0,76	0,76	0,49	0,49	0,63
altküme_9	67,0131	48,7353	0,76	0,76	0,49	0,49	0,62
altküme_8	66,9715	48,6055	0,76	0,76	0,49	0,49	0,62
altküme_7	67,0193	48,5207	0,76	0,76	0,49	0,49	0,62
altküme_6	67,0424	48,5134	0,76	0,76	0,48	0,49	0,62
altküme_5	66,8684	48,5153	0,75	0,76	0,49	0,49	0,62
altküme_4	66,8622	48,5402	0,75	0,76	0,49	0,49	0,62
altküme_3	66,5573	48,1857	0,75	0,75	0,49	0,48	0,62
altküme_2	66,4172	47,8978	0,75	0,75	0,48	0,48	0,62
altküme_1	66,2956	47,7103	0,75	0,75	0,48	0,48	0,61

K en yakın komşu modeliyle, ilk öznitelik alt kümesi ile en yüksek sonuca ulaşılmış.

Bu alt küme 5 öznitelik içermektedir. Sonrasında düşüş yaşanmıştır. Doğruluğun düşüşü kritik olmasa da en iyi sonuç ilk öznitelik ile alınmıştır.

Çizelge 5.38 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznitelik önemleri –
K en yakın komşu sınıflayıcısı.

K En Yakın Komşu	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	64,4678	37,0162	0,71	0,79	0,43	0,33	0,62
altküme_33	64,474	37,0168	0,71	0,8	0,43	0,33	0,62
altküme_32	64,4725	37,0158	0,71	0,8	0,43	0,33	0,62
altküme_31	64,4725	37,0158	0,71	0,8	0,43	0,33	0,62
altküme_30	64,4663	36,9946	0,71	0,8	0,43	0,33	0,62
altküme_29	64,4663	36,9911	0,71	0,8	0,43	0,33	0,62
altküme_28	64,4663	36,9911	0,71	0,8	0,43	0,33	0,62
altküme_27	64,4648	36,9901	0,71	0,8	0,43	0,33	0,62
altküme_26	64,5479	37,1101	0,71	0,8	0,43	0,33	0,62
altküme_25	64,8235	37,9161	0,72	0,8	0,44	0,34	0,62

altküme_24	65,2039	38,7754	0,72	0,8	0,44	0,34	0,63
altküme_23	65,19	38,746	0,72	0,8	0,44	0,34	0,63
altküme_22	65,2039	38,7721	0,72	0,8	0,44	0,34	0,63
altküme_21	65,2223	38,7913	0,72	0,8	0,44	0,34	0,63
altküme_20	65,2223	38,7913	0,72	0,8	0,44	0,34	0,63
altküme_19	65,1962	38,8132	0,72	0,8	0,44	0,34	0,63
altküme_18	65,1869	38,8003	0,72	0,8	0,44	0,34	0,63
altküme_17	65,19	38,7693	0,72	0,8	0,44	0,34	0,63
altküme_16	65,19	38,7693	0,72	0,8	0,44	0,34	0,63
altküme_15	65,2932	38,9259	0,72	0,8	0,45	0,35	0,63
altküme_14	65,2855	38,914	0,72	0,8	0,45	0,35	0,63
altküme_13	65,5087	39,7461	0,72	0,8	0,45	0,36	0,63
altküme_12	65,5873	39,7617	0,72	0,8	0,45	0,35	0,63
altküme_11	65,5919	39,7585	0,72	0,8	0,45	0,35	0,63
altküme_10	65,5149	39,5943	0,72	0,8	0,45	0,35	0,63
altküme_9	66,9207	42,1987	0,73	0,81	0,48	0,38	0,65
altküme_8	66,9207	42,2018	0,73	0,81	0,48	0,38	0,65
altküme_7	66,9254	42,199	0,73	0,81	0,48	0,38	0,65
altküme_6	67,1717	42,5801	0,73	0,81	0,48	0,38	0,66
altküme_5	67,6675	43,2486	0,74	0,81	0,49	0,38	0,66
altküme_4	67,8939	43,7751	0,74	0,82	0,5	0,39	0,66
altküme_3	69,4183	46,128	0,75	0,83	0,53	0,41	0,69
altküme_2	69,3736	46,118	0,75	0,83	0,53	0,41	0,69
altküme_1	69,5907	46,5911	0,75	0,83	0,53	0,41	0,69

Gradyen boosting modeli uygulandığında, en iyi sonuca 26 öznitelik ile ulaşıldığı gözlenmiştir. İlk öznitelik ile ulaşılan doğruluk değeri 73.39 iken 34 öznitelik ile bu değer ancak 74.12 olabilmiştir. Değerler Çizelge 5.39’da verilmiştir.

Çizelge 5.39 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznitelik önemleri –
Gradyen boosting sınıflayıcısı.

Gradyen Boosting	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	74,1192	50,372	0,76	0,9	0,65	0,41	0,78
altküme_33	74,1192	50,372	0,76	0,9	0,65	0,41	0,78
altküme_32	74,1362	50,3473	0,76	0,9	0,65	0,41	0,78
altküme_31	74,2517	50,5705	0,76	0,9	0,66	0,41	0,78
altküme_30	74,2316	50,5043	0,76	0,9	0,66	0,41	0,78
altküme_29	74,247	50,5075	0,76	0,9	0,66	0,41	0,78
altküme_28	74,247	50,5075	0,76	0,9	0,66	0,41	0,78
altküme_27	74,247	50,5075	0,76	0,9	0,66	0,41	0,78
altküme_26	74,247	50,5075	0,76	0,9	0,66	0,41	0,78
altküme_25	74,1577	50,3123	0,76	0,9	0,66	0,41	0,78
altküme_24	74,1547	50,5902	0,76	0,9	0,65	0,41	0,78
altküme_23	74,1547	50,5902	0,76	0,9	0,65	0,41	0,78
altküme_22	74,1547	50,5902	0,76	0,9	0,65	0,41	0,78
altküme_21	74,1516	50,5756	0,76	0,9	0,65	0,41	0,78
altküme_20	74,1516	50,5756	0,76	0,9	0,65	0,41	0,78

altküme_19	74,15	50,5712	0,76	0,9	0,65	0,41	0,78
altküme_18	74,15	50,5712	0,76	0,9	0,65	0,41	0,78
altküme_17	74,1716	50,2816	0,76	0,9	0,66	0,41	0,78
altküme_16	74,1716	50,2816	0,76	0,9	0,66	0,41	0,78
altküme_15	74,1346	50,0565	0,76	0,9	0,66	0,4	0,78
altküme_14	74,2101	50,3248	0,76	0,9	0,66	0,41	0,78
altküme_13	74,0946	50,0208	0,76	0,9	0,66	0,4	0,78
altküme_12	74,0961	50,3116	0,76	0,9	0,65	0,41	0,78
altküme_11	74,0792	50,2277	0,76	0,9	0,65	0,41	0,78
altküme_10	74,0592	50,1583	0,76	0,9	0,65	0,41	0,78
altküme_9	74,1285	50,4453	0,76	0,9	0,65	0,41	0,78
altküme_8	74,11	50,4567	0,76	0,9	0,65	0,41	0,78
altküme_7	74,11	50,4567	0,76	0,9	0,65	0,41	0,78
altküme_6	74,1085	50,4757	0,76	0,9	0,65	0,41	0,78
altküme_5	74,0469	50,6341	0,76	0,89	0,65	0,42	0,78
altküme_4	73,9483	50,2806	0,76	0,89	0,65	0,41	0,78
altküme_3	74,053	50,6314	0,76	0,89	0,65	0,42	0,78
altküme_2	73,973	49,963	0,76	0,9	0,65	0,41	0,77
altküme_1	73,3971	48,0344	0,76	0,9	0,64	0,38	0,77

AdaBoost algoritmasıyla da yine ilk alt kümeden itibaren iyi sonuçlar elde edilmiş olsa da, doğruluk artarak devam etmiş ve en son alt küme ile en iyi doğruluk oranına ulaşmıştır. AdaBoost modeli ile elde edilen sonuçlar Çizelge 5.40’de verilmiştir.

Çizelge 5.40 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznelik önemleri – AdaBoost sınıflayıcısı.

AdaBoost	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	73,3755	49,7866	0,76	0,89	0,63	0,41	0,77
altküme_33	73,3755	49,7866	0,76	0,89	0,63	0,41	0,77
altküme_32	73,3755	49,7866	0,76	0,89	0,63	0,41	0,77
altküme_31	73,3755	49,7866	0,76	0,89	0,63	0,41	0,77
altküme_30	73,3678	49,7414	0,76	0,89	0,63	0,41	0,77
altküme_29	73,3678	49,7414	0,76	0,89	0,63	0,41	0,77
altküme_28	73,3678	49,7414	0,76	0,89	0,63	0,41	0,77
altküme_27	73,3678	49,7414	0,76	0,89	0,63	0,41	0,77
altküme_26	73,3678	49,7414	0,76	0,89	0,63	0,41	0,77
altküme_25	73,3678	49,7414	0,76	0,89	0,63	0,41	0,77
altküme_24	73,3678	49,7414	0,76	0,89	0,63	0,41	0,77
altküme_23	73,3155	50,0202	0,76	0,88	0,63	0,42	0,77
altküme_22	73,3155	50,0202	0,76	0,88	0,63	0,42	0,77
altküme_21	73,3155	50,0202	0,76	0,88	0,63	0,42	0,77
altküme_20	73,3155	50,0202	0,76	0,88	0,63	0,42	0,77
altküme_19	73,3155	50,0202	0,76	0,88	0,63	0,42	0,77
altküme_18	73,3155	50,0202	0,76	0,88	0,63	0,42	0,77
altküme_17	73,3155	50,0202	0,76	0,88	0,63	0,42	0,77
altküme_16	73,3155	50,0202	0,76	0,88	0,63	0,42	0,77
altküme_15	73,3386	50,1856	0,76	0,88	0,63	0,42	0,77

altküme_14	73,3032	49,9769	0,76	0,88	0,63	0,42	0,77
altküme_13	73,3339	50,0058	0,76	0,88	0,63	0,42	0,77
altküme_12	73,3124	49,8176	0,76	0,88	0,63	0,41	0,77
altküme_11	73,3278	50,0548	0,76	0,88	0,63	0,42	0,77
altküme_10	73,2477	49,1215	0,76	0,89	0,63	0,4	0,77
altküme_9	73,2169	49,5153	0,76	0,88	0,63	0,41	0,77
altküme_8	73,2169	49,5153	0,76	0,88	0,63	0,41	0,77
altküme_7	73,1523	49,2313	0,76	0,88	0,62	0,41	0,77
altküme_6	73,1523	49,2313	0,76	0,88	0,62	0,41	0,77
altküme_5	73,1569	49,0427	0,76	0,89	0,63	0,4	0,76
altküme_4	73,1723	49,459	0,76	0,88	0,62	0,41	0,76
altküme_3	73,2462	49,2982	0,76	0,89	0,63	0,41	0,76
altküme_2	73,3724	49,4903	0,76	0,89	0,63	0,41	0,76
altküme_1	72,7334	47,6621	0,75	0,89	0,62	0,39	0,76

Yerine koyma sınıflandırıcısı da yine ilk alt kümeden itibaren iyi sonuçlar elde edilmiş olsa da, doğruluk artarak devam etmiş ve en son alt küme ile en iyi doğruluk oranına ulaşmıştır.

Çizelge 5.41 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznelik önemleri – Bagging sınıflayıcısı.

Bagging	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	73,434	49,5718	0,76	0,89	0,63	0,41	0,75
altküme_33	73,4756	49,9069	0,76	0,89	0,63	0,41	0,75
altküme_32	73,4864	49,9084	0,76	0,89	0,63	0,41	0,75
altküme_31	73,4202	49,8955	0,76	0,89	0,63	0,41	0,75
altküme_30	73,4371	49,7831	0,76	0,89	0,63	0,41	0,75
altküme_29	73,4094	49,798	0,76	0,89	0,63	0,41	0,75
altküme_28	73,4417	49,6996	0,76	0,89	0,63	0,41	0,76
altküme_27	73,5064	49,9826	0,76	0,89	0,63	0,41	0,75
altküme_26	73,4895	49,9113	0,76	0,89	0,63	0,41	0,75
altküme_25	73,4063	49,587	0,76	0,89	0,63	0,41	0,75
altküme_24	73,1523	49,4404	0,76	0,88	0,62	0,41	0,75
altküme_23	73,4402	49,7539	0,76	0,89	0,63	0,41	0,75
altküme_22	73,3355	49,7023	0,76	0,89	0,63	0,41	0,75
altküme_21	73,4079	50,0376	0,76	0,88	0,63	0,42	0,75
altküme_20	73,4048	49,8577	0,76	0,89	0,63	0,41	0,75
altküme_19	73,5018	50,0566	0,76	0,89	0,63	0,41	0,75
altküme_18	73,4479	49,9245	0,76	0,89	0,63	0,41	0,75
altküme_17	73,5233	50,1117	0,76	0,89	0,63	0,42	0,75
altküme_16	73,3324	49,784	0,76	0,88	0,63	0,41	0,75
altküme_15	73,4464	49,9143	0,76	0,89	0,63	0,41	0,75
altküme_14	73,3817	49,8331	0,76	0,89	0,63	0,41	0,75
altküme_13	73,2908	49,7363	0,76	0,88	0,63	0,41	0,75
altküme_12	73,1384	49,1266	0,76	0,89	0,62	0,4	0,75
altküme_11	73,1769	49,4721	0,76	0,88	0,62	0,41	0,75
altküme_10	72,9937	49,0959	0,76	0,88	0,62	0,41	0,75

altküme_9	72,6934	48,7782	0,76	0,88	0,61	0,41	0,74
altküme_8	72,7519	49,0088	0,76	0,88	0,61	0,41	0,75
altküme_7	72,7288	48,8875	0,76	0,88	0,61	0,41	0,74
altküme_6	72,5856	48,8185	0,76	0,88	0,61	0,41	0,74
altküme_5	72,4563	48,6361	0,76	0,87	0,6	0,41	0,74
altküme_4	72,4085	48,6723	0,76	0,87	0,6	0,41	0,74
altküme_3	72,1529	47,9044	0,76	0,87	0,6	0,4	0,74
altküme_2	71,8696	47,4772	0,75	0,87	0,59	0,4	0,73
altküme_1	71,808	47,6332	0,75	0,87	0,59	0,4	0,73

Rastgele orman modelinde de aynı gradyen boosting, adaboost ve yerine koyma sınıflandırıcısında olduğu gibi, ilk alt kümeden itibaren iyi sonuçlar elde edilmiş olsa da, doğruluk artarak devam etmiş ve en son alt küme ile en iyi doğruluk oranına ulaşmıştır. Modelin çalıştırılmasıyla elde edilen sonuçlar Çizelge 5.42’de verilmiştir.

Çizelge 5.42 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznelilik önemleri – Rastgele orman sınıflayıcısı.

Rast. Orman Sınıflayıcısı	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	72,0282	45,3621	0,75	0,89	0,61	0,36	0,74
altküme_33	72,3931	46,1769	0,75	0,89	0,62	0,37	0,74
altküme_32	72,3054	45,6748	0,75	0,89	0,61	0,36	0,74
altküme_31	72,3346	45,724	0,75	0,89	0,62	0,36	0,74
altküme_30	72,156	45,5676	0,75	0,89	0,61	0,36	0,74
altküme_29	72,3654	46,0387	0,75	0,89	0,61	0,37	0,74
altküme_28	72,4424	46,3407	0,75	0,89	0,62	0,37	0,74
altküme_27	72,4316	45,8046	0,75	0,89	0,62	0,36	0,74
altküme_26	72,5456	46,3049	0,75	0,89	0,62	0,37	0,74
altküme_25	72,6903	46,8345	0,75	0,89	0,62	0,38	0,74
altküme_24	72,33	45,9321	0,75	0,89	0,61	0,37	0,74
altküme_23	72,2453	45,8009	0,75	0,89	0,61	0,37	0,74
altküme_22	72,4948	46,3267	0,75	0,89	0,62	0,37	0,74
altküme_21	72,3685	45,9797	0,75	0,89	0,62	0,37	0,74
altküme_20	72,4008	46,4091	0,75	0,89	0,61	0,37	0,74
altküme_19	72,6349	46,5118	0,75	0,89	0,62	0,37	0,74
altküme_18	72,6888	46,9889	0,75	0,89	0,62	0,38	0,74
altküme_17	72,6703	46,7557	0,75	0,89	0,62	0,37	0,74
altküme_16	72,5949	46,8336	0,75	0,89	0,62	0,38	0,74
altküme_15	72,4717	46,5019	0,75	0,89	0,62	0,37	0,74
altküme_14	72,4547	46,5473	0,75	0,89	0,62	0,37	0,74
altküme_13	72,6811	47,083	0,75	0,89	0,62	0,38	0,74
altküme_12	72,059	45,8393	0,75	0,89	0,6	0,37	0,74
altküme_11	72,2746	46,4298	0,75	0,89	0,61	0,37	0,74
altküme_10	72,3962	46,8499	0,75	0,89	0,61	0,38	0,74
altküme_9	72,2992	47,2712	0,75	0,88	0,61	0,39	0,74
altküme_8	72,2592	47,3278	0,75	0,88	0,6	0,39	0,74
altküme_7	72,3593	47,2386	0,75	0,88	0,61	0,39	0,74

altküme_6	72,2931	47,3213	0,75	0,88	0,61	0,39	0,74
altküme_5	71,8157	46,9664	0,75	0,87	0,59	0,39	0,73
altküme_4	71,8111	47,5008	0,75	0,87	0,59	0,4	0,73
altküme_3	71,7403	47,178	0,75	0,87	0,59	0,39	0,73
altküme_2	71,3353	46,6621	0,75	0,87	0,58	0,39	0,72
altküme_1	71,2614	46,7959	0,75	0,86	0,58	0,39	0,72

YSA ile öznitelik sayısı arttıkça doğruluk değeri diğer modellerde artsa da, en yüksek değere alt küme 4 ile yani 9 öznitelik ile ulaşılmıştır. Değerler 5.43'te verilmiştir.

Çizelge 5.43 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznitelik önemleri – YSA sınıflayıcısı.

YSA Sınıflayıcısı	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	72,0806	49,3859	0,76	0,86	0,59	0,43	0,74
altküme_33	72,2653	42,6771	0,74	0,91	0,63	0,32	0,75
altküme_32	71,4739	54,4614	0,78	0,8	0,56	0,53	0,75
altküme_31	71,9728	49,6459	0,76	0,86	0,58	0,43	0,75
altküme_30	72,1945	52,2805	0,77	0,84	0,58	0,48	0,75
altküme_29	72,0067	49,8427	0,76	0,85	0,59	0,43	0,74
altküme_28	72,3362	47,4525	0,75	0,88	0,61	0,39	0,75
altküme_27	72,3947	50,5843	0,76	0,86	0,59	0,44	0,75
altküme_26	72,0729	51,3166	0,77	0,84	0,58	0,46	0,75
altküme_25	72,233	46,6558	0,75	0,88	0,61	0,38	0,74
altküme_24	72,4809	47,3796	0,75	0,88	0,61	0,39	0,75
altküme_23	72,2546	51,5475	0,77	0,85	0,59	0,46	0,75
altküme_22	72,1576	52,6327	0,77	0,83	0,58	0,48	0,75
altküme_21	72,1822	48,7664	0,76	0,87	0,59	0,41	0,74
altküme_20	71,8912	51,6693	0,77	0,84	0,58	0,47	0,75
altküme_19	72,2361	49,4463	0,76	0,86	0,59	0,42	0,75
altküme_18	71,477	54,598	0,78	0,8	0,56	0,54	0,75
altküme_17	72,1729	53,8815	0,78	0,82	0,57	0,51	0,75
altküme_16	71,8896	51,2966	0,77	0,84	0,58	0,46	0,75
altküme_15	72,2345	48,2998	0,76	0,87	0,6	0,4	0,75
altküme_14	71,8573	52,7201	0,77	0,83	0,57	0,49	0,75
altküme_13	72,387	49,2256	0,76	0,87	0,6	0,42	0,75
altküme_12	72,4809	47,3951	0,75	0,88	0,61	0,39	0,75
altküme_11	72,2268	46,2368	0,75	0,89	0,61	0,37	0,74
altküme_10	71,6063	46,4763	0,75	0,87	0,59	0,38	0,74
altküme_9	73,3555	49,3591	0,76	0,89	0,63	0,41	0,76
altküme_8	73,3955	49,8491	0,76	0,89	0,63	0,41	0,76
altküme_7	73,3463	51,9327	0,77	0,87	0,62	0,45	0,76
altküme_6	73,3247	44,7929	0,75	0,92	0,66	0,34	0,77
altküme_5	73,4679	49,4944	0,76	0,89	0,63	0,41	0,76
altküme_4	73,608	50,4051	0,76	0,89	0,63	0,42	0,77
altküme_3	72,3362	43,1492	0,74	0,91	0,63	0,33	0,75
altküme_2	73,374	46,3781	0,75	0,91	0,65	0,36	0,76
altküme_1	73,046	46,2856	0,75	0,9	0,64	0,36	0,76

XGB sınıflayıcı, ilk öznitelik alt kümesinden itibaren oldukça yüksek sonuç üretmiş olsa da, doğruluk artarak devam etmiş ve en iyi doğruluk oranına tüm alt küme öznitelikleriyle ulaşılmıştır. Değerler Çizelge 5.44’te verilmiştir.

Çizelge 5.44 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznitelik önemleri – XGB sınıflayıcısı.

XGB Sınıflayıcısı	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	74,284	50,1478	0,76	0,9	0,66	0,4	0,78
altküme_33	74,284	50,1478	0,76	0,9	0,66	0,4	0,78
altküme_32	74,2132	49,9537	0,76	0,9	0,66	0,4	0,78
altküme_31	74,1947	49,7797	0,76	0,9	0,66	0,4	0,78
altküme_30	74,1993	49,724	0,76	0,9	0,66	0,4	0,78
altküme_29	74,1947	49,8008	0,76	0,9	0,66	0,4	0,78
altküme_28	74,1947	49,8008	0,76	0,9	0,66	0,4	0,78
altküme_27	74,1947	49,8008	0,76	0,9	0,66	0,4	0,78
altküme_26	74,1947	49,8008	0,76	0,9	0,66	0,4	0,78
altküme_25	74,1778	49,8145	0,76	0,9	0,66	0,4	0,78
altküme_24	74,244	49,9446	0,76	0,9	0,66	0,4	0,78
altküme_23	74,244	49,9446	0,76	0,9	0,66	0,4	0,78
altküme_22	74,244	49,9446	0,76	0,9	0,66	0,4	0,78
altküme_21	74,244	49,9446	0,76	0,9	0,66	0,4	0,78
altküme_20	74,244	49,9446	0,76	0,9	0,66	0,4	0,78
altküme_19	74,1932	49,8624	0,76	0,9	0,66	0,4	0,78
altküme_18	74,1932	49,8624	0,76	0,9	0,66	0,4	0,78
altküme_17	74,2162	50,0939	0,76	0,9	0,66	0,4	0,78
altküme_16	74,2162	50,0939	0,76	0,9	0,66	0,4	0,78
altküme_15	74,2162	50,0879	0,76	0,9	0,66	0,4	0,78
altküme_14	74,2116	50,0269	0,76	0,9	0,66	0,4	0,78
altküme_13	74,1824	49,858	0,76	0,9	0,66	0,4	0,78
altküme_12	74,2594	50,3844	0,76	0,9	0,66	0,41	0,78
altküme_11	74,2178	50,3086	0,76	0,9	0,66	0,41	0,78
altküme_10	74,1901	50,2257	0,76	0,9	0,66	0,41	0,78
altküme_9	74,1885	50,3657	0,76	0,9	0,66	0,41	0,78
altküme_8	74,1516	50,3123	0,76	0,9	0,65	0,41	0,78
altküme_7	74,1516	50,3093	0,76	0,9	0,66	0,41	0,78
altküme_6	74,1516	50,3093	0,76	0,9	0,66	0,41	0,78
altküme_5	74,1778	50,3905	0,76	0,9	0,66	0,41	0,78
altküme_4	74,0977	50,2984	0,76	0,9	0,65	0,41	0,78
altküme_3	73,9976	49,7993	0,76	0,9	0,65	0,4	0,77
altküme_2	73,8914	49,1666	0,76	0,9	0,65	0,39	0,77
altküme_1	73,3201	47,67	0,75	0,9	0,64	0,38	0,77

Extra trees sınıflayıcı modeli ile ilk alt kümeden (5 öznitelik) ulaşılmış olup, 34 öznitelik ile 1 puanlık düşüş gözlenmiş olup, öznitelik eklemenin avantajlı olmadığı görülmüştür. Değerler Çizelge 5.45’de verilmiştir.

Çizelge 5.45 : Özyinelemeli yöntemle (Adaboost) oluşturulmuş öznelik önemleri –

Extra trees sınıflayıcısı.

Extra Trees	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_34	69,6308	40,0972	0,73	0,88	0,54	0,32	0,69
altküme_33	69,6369	39,9342	0,73	0,88	0,55	0,32	0,69
altküme_32	69,6816	40,2283	0,73	0,88	0,55	0,32	0,69
altküme_31	69,7154	40,2769	0,73	0,88	0,55	0,32	0,69
altküme_30	69,7062	40,4035	0,73	0,87	0,55	0,32	0,69
altküme_29	69,8694	40,5264	0,73	0,88	0,55	0,32	0,69
altküme_28	69,9079	40,4231	0,73	0,88	0,55	0,32	0,70
altküme_27	69,6492	40,5202	0,73	0,87	0,54	0,32	0,69
altküme_26	69,9172	40,7772	0,73	0,88	0,55	0,32	0,70
altküme_25	69,7478	40,5555	0,73	0,87	0,55	0,32	0,69
altküme_24	69,4521	40,9114	0,73	0,87	0,54	0,33	0,69
altküme_23	69,7	41,1156	0,73	0,87	0,54	0,33	0,69
altküme_22	69,6816	40,8567	0,73	0,87	0,54	0,33	0,69
altküme_21	69,7955	41,1814	0,73	0,87	0,55	0,33	0,69
altküme_20	69,7632	41,0884	0,73	0,87	0,55	0,33	0,69
altküme_19	69,854	41,327	0,73	0,87	0,55	0,33	0,69
altküme_18	69,8771	41,2328	0,73	0,87	0,55	0,33	0,70
altküme_17	69,7077	41,0723	0,73	0,87	0,55	0,33	0,69
altküme_16	69,9372	41,6113	0,74	0,87	0,55	0,33	0,70
altküme_15	69,3105	40,5737	0,73	0,87	0,53	0,33	0,69
altküme_14	69,3644	40,7646	0,73	0,87	0,54	0,33	0,69
altküme_13	69,6462	41,3076	0,73	0,87	0,54	0,33	0,69
altküme_12	69,235	41,6676	0,73	0,86	0,53	0,34	0,69
altküme_11	69,0949	41,239	0,73	0,86	0,53	0,34	0,68
altküme_10	69,3906	42,0082	0,74	0,86	0,53	0,35	0,69
altküme_9	69,8663	44,0249	0,74	0,85	0,54	0,37	0,69
altküme_8	69,6138	43,6365	0,74	0,85	0,54	0,37	0,69
altküme_7	69,5014	43,478	0,74	0,85	0,54	0,37	0,69
altküme_6	69,8432	43,974	0,74	0,85	0,54	0,37	0,70
altküme_5	69,8525	44,24	0,74	0,85	0,54	0,37	0,70
altküme_4	70,3699	45,3526	0,75	0,85	0,55	0,38	0,70
altküme_3	70,3637	45,6591	0,75	0,85	0,55	0,39	0,70
altküme_2	69,9695	45,15	0,75	0,85	0,54	0,39	0,70
altküme_1	70,5023	45,6524	0,75	0,86	0,56	0,39	0,70

5.3.5 Ki-kare temelli yöntemler ile oluşturulmuş alt kümeler

Lojistik regresyon modeli için incelendiğinde, veri kümesine bilgi kazancını artıran ne kadar parametre eklenirse sonuçta o kadar iyileşme olduğu Çizelge 5.46’da görülmektedir.

Çizelge 5.46 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – Lojistik

Regresyon sınıflayıcısı.

Lojistik Reg.	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	70,1266	27,9122	0,71	0,95	0,62	0,18	0,72
altküme_37	70,1189	27,9016	0,71	0,95	0,61	0,18	0,72
altküme_36	69,834	24,8667	0,71	0,95	0,62	0,16	0,71
altküme_35	69,834	24,8667	0,71	0,95	0,62	0,16	0,71
altküme_34	69,8356	24,8677	0,71	0,95	0,62	0,16	0,71
altküme_33	69,1796	20,5966	0,7	0,96	0,59	0,12	0,69
altküme_32	69,1827	20,6298	0,7	0,96	0,59	0,12	0,69
altküme_31	69,0657	19,8516	0,7	0,96	0,58	0,12	0,69
altküme_30	69,0241	19,238	0,7	0,96	0,58	0,12	0,69
altküme_29	69,021	19,1456	0,7	0,96	0,58	0,11	0,69
altküme_28	68,8439	18,3323	0,7	0,96	0,57	0,11	0,69
altküme_27	68,8408	18,2978	0,7	0,96	0,57	0,11	0,69
altküme_26	68,6899	17,4823	0,69	0,96	0,56	0,1	0,68
altküme_25	68,556	17,1999	0,69	0,96	0,55	0,1	0,67
altküme_24	68,5498	17,2105	0,69	0,96	0,55	0,1	0,67
altküme_23	68,556	17,22	0,69	0,96	0,55	0,1	0,67
altküme_22	68,5544	17,2126	0,69	0,96	0,55	0,1	0,67
altküme_21	68,459	16,0079	0,69	0,96	0,55	0,09	0,67
altküme_20	68,4682	15,9842	0,69	0,96	0,55	0,09	0,67
altküme_19	68,442	15,7797	0,69	0,96	0,54	0,09	0,67
altküme_18	68,4328	15,7482	0,69	0,96	0,54	0,09	0,67
altküme_17	68,4205	15,8052	0,69	0,96	0,54	0,09	0,67
altküme_16	68,3743	15,2577	0,69	0,96	0,54	0,09	0,67
altküme_15	68,2804	10,2397	0,69	0,98	0,55	0,06	0,66
altküme_14	68,3004	9,9943	0,69	0,98	0,55	0,05	0,66
altküme_13	68,3019	10,0026	0,69	0,98	0,55	0,05	0,66
altküme_12	68,3035	10,0109	0,69	0,98	0,55	0,06	0,66
altküme_11	68,1957	7,5715	0,69	0,98	0,55	0,04	0,66
altküme_10	68,1957	7,5715	0,69	0,98	0,55	0,04	0,66
altküme_9	68,0602	5,4256	0,68	0,99	0,53	0,03	0,65
altküme_8	68,0602	5,4256	0,68	0,99	0,53	0,03	0,65
altküme_7	68,0648	5,4695	0,68	0,99	0,53	0,03	0,65
altküme_6	67,9585	0	0,68	1	0	0	0,64
altküme_5	67,9585	0	0,68	1	0	0	0,63
altküme_4	67,9585	0	0,68	1	0	0	0,63
altküme_3	67,9585	0	0,68	1	0	0	0,63
altküme_2	67,9585	0	0,68	1	0	0	0,63
altküme_1	67,9585	0	0,68	1	0	0	0,63

Naive Bayes algoritması için, 26 öznitelik ile en etkili sonuçları vermiştir. Ekleme devam ettikçe model performansında düşüş olduğu Çizelge 5.47 ile gözlenmektedir.

Çizelge 5.47 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – Naive Bayes sınıflayıcısı.

Naive Bayes	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	41,545	52,0045	0,96	0,15	0,35	0,99	0,61
altküme_37	41,545	52,0045	0,96	0,15	0,35	0,99	0,61
altküme_36	41,5543	52,0097	0,96	0,15	0,35	0,99	0,61
altküme_35	41,9207	52,0474	0,95	0,15	0,35	0,98	0,63
altküme_34	42,8015	52,1911	0,93	0,17	0,36	0,97	0,64
altküme_33	42,6968	52,1517	0,93	0,17	0,36	0,97	0,63
altküme_32	42,6952	52,151	0,93	0,17	0,36	0,97	0,63
altküme_31	42,8831	52,1874	0,93	0,17	0,36	0,97	0,64
altküme_30	42,8816	52,1855	0,93	0,17	0,36	0,97	0,64
altküme_29	42,8292	52,1835	0,93	0,17	0,36	0,97	0,64
altküme_28	42,9632	52,2051	0,93	0,17	0,36	0,97	0,64
altküme_27	42,9601	52,205	0,93	0,17	0,36	0,97	0,64
altküme_26	42,9724	52,2042	0,93	0,17	0,36	0,97	0,64
altküme_25	42,8215	52,1136	0,93	0,17	0,36	0,97	0,64
altküme_24	42,7399	52,0953	0,93	0,17	0,36	0,97	0,64
altküme_23	42,6922	52,0807	0,93	0,17	0,36	0,97	0,63
altküme_22	42,5844	52,0473	0,93	0,17	0,36	0,97	0,63
altküme_21	42,1224	51,9237	0,93	0,16	0,35	0,98	0,63
altküme_20	42,3642	51,9974	0,93	0,16	0,35	0,97	0,63
altküme_19	42,3288	51,9894	0,93	0,16	0,35	0,97	0,63
altküme_18	42,455	52,0368	0,93	0,17	0,35	0,97	0,63
altküme_17	42,4427	52,0326	0,93	0,17	0,35	0,97	0,63
altküme_16	42,3811	52,0121	0,93	0,16	0,35	0,97	0,63
altküme_15	42,3272	51,9937	0,93	0,16	0,35	0,97	0,63
altküme_14	41,9931	51,8888	0,93	0,16	0,35	0,98	0,63
altküme_13	41,8283	51,8647	0,94	0,15	0,35	0,98	0,63
altküme_12	41,7544	51,8759	0,94	0,15	0,35	0,98	0,63
altküme_11	40,0206	51,5769	0,99	0,12	0,35	1	0,62
altküme_10	40,0068	51,5723	0,99	0,12	0,35	1	0,62
altküme_9	40,0099	51,6061	0,99	0,12	0,35	1	0,62
altküme_8	40,0099	51,6097	0,99	0,12	0,35	1	0,62
altküme_7	40,0114	51,6163	0,99	0,12	0,35	1	0,62
altküme_6	39,8774	51,5931	1	0,12	0,35	1	0,61
altküme_5	39,8774	51,5931	1	0,12	0,35	1	0,60
altküme_4	39,8774	51,5931	1	0,12	0,35	1	0,60
altküme_3	39,8774	51,5931	1	0,12	0,35	1	0,60
altküme_2	39,8774	51,5931	1	0,12	0,35	1	0,63
altküme_1	39,8774	51,5931	1	0,12	0,35	1	0,63

Karar ağaçları için ise, görülmüştür ki öznitelik eklendikçe modelin bilgi kazancı artıyor ve modelin doğruluğunda iyileşme gözlemleniyor. Değerler Çizelge 5.48’de verilmiştir.

Çizelge 5.48 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – Karar ağaçları sınıflayıcısı.

Karar Ağaçları	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	67,6891	49,3409	0,76	0,76	0,49	0,49	0,63
altküme_37	67,5674	49,1833	0,76	0,76	0,49	0,49	0,63
altküme_36	67,8107	49,2536	0,77	0,76	0,49	0,49	0,63
altküme_35	67,7137	49,4357	0,76	0,76	0,49	0,49	0,63
altküme_34	67,8631	49,4441	0,77	0,76	0,49	0,49	0,63
altküme_33	67,2195	48,5213	0,76	0,76	0,48	0,49	0,62
altküme_32	67,1825	48,5479	0,76	0,76	0,48	0,49	0,62
altküme_31	67,2549	48,6279	0,76	0,76	0,48	0,49	0,62
altküme_30	67,1917	48,5947	0,76	0,76	0,48	0,49	0,62
altküme_29	67,2148	48,5974	0,76	0,76	0,48	0,49	0,62
altküme_28	66,7868	47,9564	0,76	0,76	0,48	0,48	0,62
altküme_27	66,8407	47,9215	0,76	0,76	0,48	0,48	0,62
altküme_26	66,3325	47,4689	0,75	0,75	0,47	0,47	0,61
altküme_25	64,5895	44,453	0,74	0,74	0,44	0,44	0,59
altküme_24	64,5941	44,4884	0,74	0,74	0,44	0,44	0,59
altküme_23	64,5803	44,4439	0,74	0,74	0,44	0,44	0,59
altküme_22	64,7435	44,7639	0,74	0,74	0,45	0,45	0,59
altküme_21	64,685	44,4794	0,74	0,74	0,44	0,44	0,59
altküme_20	64,608	44,4525	0,74	0,74	0,44	0,44	0,59
altküme_19	64,6665	44,6006	0,74	0,74	0,44	0,45	0,59
altküme_18	64,2554	43,7999	0,74	0,74	0,43	0,44	0,59
altküme_17	64,1229	43,4054	0,74	0,74	0,43	0,43	0,59
altküme_16	64,2215	43,7412	0,74	0,74	0,43	0,44	0,59
altküme_15	62,1474	40,2441	0,73	0,72	0,4	0,4	0,56
altküme_14	62,0627	40,416	0,72	0,72	0,4	0,4	0,56
altküme_13	62,0103	40,2846	0,72	0,72	0,4	0,4	0,56
altküme_12	62,0057	40,2123	0,72	0,72	0,4	0,4	0,56
altküme_11	61,8748	39,9787	0,72	0,72	0,4	0,4	0,56
altküme_10	61,841	39,9079	0,72	0,72	0,4	0,4	0,56
altküme_9	60,9402	37,9886	0,72	0,71	0,37	0,38	0,55
altküme_8	61,145	37,8013	0,73	0,72	0,37	0,38	0,55
altküme_7	67,7245	12,2273	0,96	0,8	0,07	0,12	0,64
altküme_6	67,9585	0	1	0,81	0	0	0,64
altküme_5	67,9585	0	1	0,81	0	0	0,63
altküme_4	67,9585	0	1	0,81	0	0	0,63
altküme_3	67,9585	0	1	0,81	0	0	0,63
altküme_2	67,9585	0	1	0,81	0	0	0,63
altküme_1	67,9585	0	1	0,81	0	0	0,63

K en yakın komşu modelinde; öznitelik sayısı ne kadar az ise o kadar doğru sonuçlar alınmıştır. Ancak f-puanına bakılınca modelin hataya karşı duyarlılığının çok arttığı görülmektedir. Ancak yine de en iyi sonuca 16 öznitelik ile ulaşabilmektedir.

Çizelge 5.49 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – K en yakın komşu sınıflayıcısı.

K En Yakın Komşu	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	64,4678	37,0162	0,71	0,79	0,43	0,33	0,62
altküme_37	64,7681	37,8082	0,72	0,8	0,44	0,33	0,62
altküme_36	64,7866	37,8442	0,72	0,8	0,44	0,33	0,62
altküme_35	64,7866	37,8442	0,72	0,8	0,44	0,33	0,62
altküme_34	64,7866	37,8442	0,72	0,8	0,44	0,33	0,62
altküme_33	64,7773	37,7873	0,72	0,8	0,44	0,33	0,62
altküme_32	64,7773	37,7873	0,72	0,8	0,44	0,33	0,62
altküme_31	64,782	37,7971	0,72	0,8	0,44	0,33	0,62
altküme_30	64,6865	37,5878	0,72	0,8	0,43	0,33	0,62
altküme_29	64,6957	37,5871	0,72	0,8	0,43	0,33	0,62
altküme_28	64,705	37,6034	0,72	0,8	0,43	0,33	0,62
altküme_27	64,7081	37,6055	0,72	0,8	0,43	0,33	0,62
altküme_26	64,6865	37,5606	0,72	0,8	0,43	0,33	0,62
altküme_25	64,685	37,5528	0,72	0,8	0,43	0,33	0,62
altküme_24	64,6603	37,533	0,72	0,8	0,43	0,33	0,62
altküme_23	64,6603	37,5228	0,72	0,8	0,43	0,33	0,62
altküme_22	64,6742	37,5456	0,72	0,8	0,43	0,33	0,62
altküme_21	64,9744	38,2563	0,72	0,8	0,44	0,34	0,63
altküme_20	64,9806	38,2671	0,72	0,8	0,44	0,34	0,63
altküme_19	64,9914	38,2878	0,72	0,8	0,44	0,34	0,63
altküme_18	65,1777	38,6046	0,72	0,8	0,44	0,34	0,63
altküme_17	65,421	39,3628	0,72	0,8	0,45	0,35	0,63
altküme_16	65,6165	39,9117	0,72	0,8	0,45	0,36	0,64
altküme_15	64,0382	35,9628	0,71	0,79	0,42	0,32	0,61
altküme_14	64,2138	36,0861	0,71	0,8	0,42	0,32	0,61
altküme_13	64,2153	36,0906	0,71	0,8	0,42	0,32	0,61
altküme_12	64,1537	35,889	0,71	0,8	0,42	0,31	0,61
altküme_11	64,0906	35,5079	0,71	0,8	0,42	0,31	0,61
altküme_10	64,0952	35,4965	0,71	0,8	0,42	0,31	0,61
altküme_9	63,9613	34,4012	0,71	0,8	0,41	0,29	0,60
altküme_8	63,9659	34,382	0,71	0,8	0,41	0,29	0,60
altküme_7	63,2622	32,0353	0,7	0,8	0,39	0,27	0,59
altküme_6	62,848	25,7737	0,69	0,83	0,36	0,2	0,58
altküme_5	67,6752	2,5711	0,68	0,99	0,38	0,01	0,56
altküme_4	67,4674	3,6571	0,68	0,98	0,36	0,02	0,52
altküme_3	67,4674	3,6571	0,68	0,98	0,36	0,02	0,52
altküme_2	67,9585	0	0,68	1	0	0	0,52
altküme_1	67,9585	0	0,68	1	0	0	0,52

Gradyen boosting model, bir önceki iterasyondaki hata oranlarını da modele ekleyerek ilerlediğinden, öznitelik sayısı arttıkça modelin doğruluğundaki artış Çizelge 5.50’de gözlenmiştir.

Çizelge 5.50 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – Gradyen boosting sınıflayıcısı.

Gradyen Boosting	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	74,1192	50,372	0,76	0,9	0,65	0,41	0,78
altküme_37	74,1608	50,5525	0,76	0,9	0,65	0,41	0,78
altküme_36	74,1316	50,4132	0,76	0,9	0,65	0,41	0,78
altküme_35	74,1316	50,4132	0,76	0,9	0,65	0,41	0,78
altküme_34	74,1316	50,4132	0,76	0,9	0,65	0,41	0,78
altküme_33	73,7112	46,2183	0,75	0,92	0,67	0,35	0,77
altküme_32	73,7112	46,2183	0,75	0,92	0,67	0,35	0,77
altküme_31	73,7312	46,3724	0,75	0,92	0,67	0,35	0,77
altküme_30	73,7066	46,18	0,75	0,92	0,67	0,35	0,77
altküme_29	73,685	46,3624	0,75	0,92	0,67	0,35	0,77
altküme_28	73,7666	46,1723	0,75	0,92	0,67	0,35	0,77
altküme_27	73,7666	46,1723	0,75	0,92	0,67	0,35	0,77
altküme_26	72,8705	42,0485	0,74	0,93	0,67	0,31	0,76
altküme_25	71,9066	37,9696	0,73	0,93	0,65	0,27	0,74
altküme_24	71,9066	37,9696	0,73	0,93	0,65	0,27	0,74
altküme_23	71,9019	37,7329	0,73	0,93	0,65	0,27	0,74
altküme_22	71,8896	37,8159	0,73	0,93	0,65	0,27	0,74
altküme_21	71,8665	37,8136	0,73	0,93	0,65	0,27	0,74
altküme_20	71,8989	38,0348	0,73	0,93	0,65	0,27	0,74
altküme_19	71,8989	38,0348	0,73	0,93	0,65	0,27	0,74
altküme_18	71,557	36,579	0,73	0,93	0,64	0,26	0,73
altküme_17	71,5724	36,4475	0,73	0,93	0,64	0,25	0,73
altküme_16	71,5293	36,1004	0,73	0,93	0,64	0,25	0,73
altküme_15	68,7839	14,4707	0,69	0,97	0,59	0,08	0,69
altküme_14	68,7808	14,0999	0,69	0,97	0,6	0,08	0,69
altküme_13	68,7931	14,3588	0,69	0,97	0,59	0,08	0,69
altküme_12	68,7823	14,1587	0,69	0,97	0,6	0,08	0,69
altküme_11	68,6268	13,2794	0,69	0,97	0,58	0,07	0,69
altküme_10	68,5976	13,3718	0,69	0,97	0,58	0,08	0,68
altküme_9	68,402	10,4395	0,69	0,98	0,57	0,06	0,68
altküme_8	68,3558	8,5199	0,69	0,98	0,58	0,05	0,68
altküme_7	67,9709	1,0466	0,68	1	0,52	0,01	0,65
altküme_6	67,9585	0	0,68	1	0	0	0,64
altküme_5	67,9585	0	0,68	1	0	0	0,63
altküme_4	67,9585	0	0,68	1	0	0	0,63
altküme_3	67,9585	0	0,68	1	0	0	0,63
altküme_2	67,9585	0	0,68	1	0	0	0,63
altküme_1	67,9585	0	0,68	1	0	0	0,63

Adaboost modeli ise, ağaç mantığı ile çalıştığından karar ağaçları gibi bilgi kazancı arttıkça modelin doğruluğunda iyileşme olduğu Çizelge 5.51’de gözlenmiştir.

Çizelge 5.51 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – AdaBoost sınıflayıcısı.

AdaBoost	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	73,3755	49,7866	0,76	0,89	0,63	0,41	0,77
altküme_37	73,3755	49,7866	0,76	0,89	0,63	0,41	0,77
altküme_36	73,3755	49,7866	0,76	0,89	0,63	0,41	0,77
altküme_35	73,3755	49,7866	0,76	0,89	0,63	0,41	0,77
altküme_34	73,3755	49,7866	0,76	0,89	0,63	0,41	0,77
altküme_33	72,3793	44,461	0,75	0,9	0,62	0,35	0,75
altküme_32	72,3793	44,461	0,75	0,9	0,62	0,35	0,75
altküme_31	72,3793	44,461	0,75	0,9	0,62	0,35	0,75
altküme_30	72,3223	44,7857	0,75	0,9	0,62	0,35	0,75
altküme_29	72,3531	44,9385	0,75	0,9	0,62	0,35	0,75
altküme_28	72,2268	44,0279	0,74	0,9	0,62	0,34	0,75
altküme_27	72,2268	44,0279	0,74	0,9	0,62	0,34	0,75
altküme_26	71,9974	42,5403	0,74	0,91	0,62	0,32	0,74
altküme_25	71,4015	39,2821	0,73	0,91	0,61	0,29	0,73
altküme_24	71,4015	39,2821	0,73	0,91	0,61	0,29	0,73
altküme_23	71,4015	39,2821	0,73	0,91	0,61	0,29	0,73
altküme_22	71,4015	39,2821	0,73	0,91	0,61	0,29	0,73
altküme_21	71,3291	39,1344	0,73	0,91	0,61	0,29	0,73
altküme_20	71,3984	39,4498	0,73	0,91	0,61	0,29	0,73
altküme_19	71,3984	39,4498	0,73	0,91	0,61	0,29	0,73
altküme_18	71,2013	37,8369	0,73	0,92	0,61	0,27	0,72
altküme_17	71,2013	37,8369	0,73	0,92	0,61	0,27	0,72
altküme_16	71,0581	37,1749	0,73	0,92	0,61	0,27	0,72
altküme_15	68,5514	21,0392	0,7	0,95	0,54	0,13	0,68
altküme_14	68,5514	21,0392	0,7	0,95	0,54	0,13	0,68
altküme_13	68,5514	21,0392	0,7	0,95	0,54	0,13	0,68
altküme_12	68,5514	21,0392	0,7	0,95	0,54	0,13	0,68
altküme_11	68,4805	18,9949	0,7	0,95	0,54	0,12	0,68
altküme_10	68,4805	18,9949	0,7	0,95	0,54	0,12	0,68
altküme_9	68,2988	17,5755	0,69	0,96	0,53	0,11	0,68
altküme_8	68,2573	17,7538	0,69	0,95	0,52	0,11	0,68
altküme_7	68,0278	4,1632	0,68	0,99	0,53	0,02	0,65
altküme_6	67,9585	0	0,68	1	0	0	0,64
altküme_5	67,9585	0	0,68	1	0	0	0,63
altküme_4	67,9585	0	0,68	1	0	0	0,63
altküme_3	67,9585	0	0,68	1	0	0	0,63
altküme_2	67,9585	0	0,68	1	0	0	0,63
altküme_1	67,9585	0	0,68	1	0	0	0,63

Yerine koyma sınıflandırıcı da ağaç kökenli olduğundan, bilgi kazancı arttıkça modelin doğruluğu arttığı Çizelge 5.52 incelendiğinde gözlenmektedir.

Çizelge 5.52 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – Bagging sınıflayıcısı.

Bagging	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	73,4063	49,5987	0,76	0,89	0,63	0,41	0,75
altküme_37	73,4541	49,9652	0,76	0,89	0,63	0,41	0,75
altküme_36	73,4895	50,071	0,76	0,89	0,63	0,41	0,75
altküme_35	73,6203	50,1861	0,76	0,89	0,64	0,41	0,75
altküme_34	73,4402	50,0014	0,76	0,89	0,63	0,41	0,75
altküme_33	72,6441	47,9705	0,76	0,88	0,61	0,39	0,74
altküme_32	72,7873	48,0252	0,76	0,89	0,62	0,39	0,74
altküme_31	72,7704	48,1712	0,76	0,88	0,62	0,39	0,74
altküme_30	72,6888	47,9808	0,76	0,88	0,62	0,39	0,74
altküme_29	72,6842	48,0953	0,76	0,88	0,61	0,39	0,74
altküme_28	72,6564	48,003	0,76	0,88	0,61	0,39	0,74
altküme_27	72,5194	47,857	0,76	0,88	0,61	0,39	0,74
altküme_26	71,8604	46,2358	0,75	0,88	0,6	0,38	0,73
altküme_25	70,105	41,8834	0,74	0,87	0,56	0,34	0,70
altküme_24	70,0711	41,5762	0,74	0,87	0,56	0,33	0,70
altküme_23	70,1866	42,0542	0,74	0,87	0,56	0,34	0,70
altküme_22	70,1697	41,7161	0,74	0,88	0,56	0,33	0,70
altküme_21	69,9664	41,2624	0,73	0,87	0,55	0,33	0,70
altküme_20	69,8448	41,0263	0,73	0,87	0,55	0,33	0,69
altküme_19	69,9649	41,3671	0,73	0,87	0,55	0,33	0,69
altküme_18	69,8202	41,1022	0,73	0,87	0,55	0,33	0,69
altküme_17	69,6	40,6422	0,73	0,87	0,54	0,32	0,69
altküme_16	69,5553	40,5925	0,73	0,87	0,54	0,32	0,68
altküme_15	66,4111	35,2816	0,71	0,84	0,46	0,29	0,64
altküme_14	66,1955	35,3495	0,71	0,84	0,46	0,29	0,64
altküme_13	66,0615	35,0532	0,71	0,84	0,45	0,29	0,64
altküme_12	65,9414	35,2014	0,71	0,83	0,45	0,29	0,64
altküme_11	65,4718	35,1908	0,71	0,83	0,44	0,29	0,63
altküme_10	65,3286	34,9577	0,71	0,82	0,44	0,29	0,63
altküme_9	63,6964	32,4267	0,7	0,81	0,4	0,27	0,60
altküme_8	63,9705	32,4139	0,7	0,81	0,41	0,27	0,60
altküme_7	67,5166	13,6189	0,69	0,96	0,46	0,08	0,64
altküme_6	67,9585	0	0,68	1	0	0	0,64
altküme_5	67,9585	0	0,68	1	0	0	0,63
altküme_4	67,9585	0	0,68	1	0	0	0,63
altküme_3	67,9585	0	0,68	1	0	0	0,63
altküme_2	67,9585	0	0,68	1	0	0	0,63
altküme_1	67,9585	0	0,68	1	0	0	0,63

Rastgele orman algoritması da aynı şekilde ağaç temelli olduğundan bilgi kazancı, yeni öznitelikler ile artmaktadır. Değerler Çizelge 5.53’te verilmiştir.

Çizelge 5.53 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – Rastgele orman sınıflayıcısı.

Rastgele Orman	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	72,2453	45,9115	0,75	0,89	0,61	0,37	0,74
altküme_37	72,1837	45,7002	0,75	0,89	0,61	0,37	0,74
altküme_36	72,2869	46,204	0,75	0,89	0,61	0,37	0,74
altküme_35	72,3362	45,9734	0,75	0,89	0,61	0,37	0,74
altküme_34	72,1375	45,6753	0,75	0,89	0,61	0,37	0,74
altküme_33	71,6248	43,9231	0,74	0,89	0,6	0,35	0,73
altküme_32	71,7433	43,8172	0,74	0,89	0,6	0,34	0,73
altküme_31	71,9589	44,5209	0,74	0,89	0,61	0,35	0,73
altküme_30	71,6448	43,8858	0,74	0,89	0,6	0,35	0,72
altküme_29	71,7203	43,7592	0,74	0,89	0,6	0,34	0,72
altküme_28	71,246	42,5804	0,74	0,89	0,59	0,33	0,71
altküme_27	71,5924	43,2111	0,74	0,89	0,6	0,34	0,72
altküme_26	70,4068	41,5356	0,74	0,88	0,57	0,33	0,70
altküme_25	69,0965	38,7288	0,73	0,87	0,53	0,3	0,68
altküme_24	69,3521	39,165	0,73	0,88	0,54	0,31	0,68
altküme_23	69,4044	39,2578	0,73	0,88	0,54	0,31	0,68
altküme_22	69,3567	39,0381	0,73	0,88	0,54	0,31	0,68
altküme_21	68,9917	38,8498	0,73	0,87	0,53	0,31	0,68
altküme_20	68,9163	38,8884	0,73	0,87	0,53	0,31	0,68
altküme_19	69,0071	38,8838	0,73	0,87	0,53	0,31	0,68
altküme_18	68,2773	38,0987	0,72	0,86	0,51	0,3	0,67
altküme_17	68,3943	38,8343	0,73	0,86	0,51	0,31	0,67
altküme_16	68,2911	39,155	0,73	0,85	0,51	0,32	0,67
altküme_15	64,3108	35,9051	0,71	0,8	0,42	0,31	0,61
altküme_14	64,2969	36,0658	0,71	0,8	0,42	0,31	0,61
altküme_13	64,4586	36,167	0,71	0,8	0,43	0,31	0,61
altküme_12	64,2323	36,0699	0,71	0,8	0,42	0,31	0,61
altküme_11	63,6163	36,2189	0,71	0,78	0,41	0,32	0,60
altküme_10	63,5671	36,4379	0,71	0,78	0,41	0,33	0,60
altküme_9	61,8902	36,3066	0,71	0,75	0,39	0,34	0,58
altküme_8	62,1351	36,4459	0,71	0,75	0,39	0,34	0,59
altküme_7	67,5443	13,9673	0,69	0,96	0,46	0,08	0,64
altküme_6	67,9585	0	0,68	1	0	0	0,64
altküme_5	67,9585	0	0,68	1	0	0	0,63
altküme_4	67,9585	0	0,68	1	0	0	0,63
altküme_3	67,9585	0	0,68	1	0	0	0,63
altküme_2	67,9585	0	0,68	1	0	0	0,63
altküme_1	67,9585	0	0,68	1	0	0	0,63

Yapay sinir ağları modeli, modele eklenen öznitelik sayısı arttıkça, gerçek hayat simülasyonu gibi olmaktadır. Gizli katmanda bu öznitelikler uygun ağırlıklarla modelin doğru tahmin yamasına katkısı artmaktadır.

Çizelge 5.54 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – YSA sınıflayıcısı.

YSA	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	72,3331	48,8412	0,76	0,87	0,6	0,41	0,75
altküme_37	72,3793	45,2342	0,75	0,9	0,62	0,36	0,75
altküme_36	71,9512	42,1163	0,74	0,91	0,62	0,32	0,74
altküme_35	72,2099	42,672	0,74	0,91	0,63	0,32	0,74
altküme_34	72,4578	46,1121	0,75	0,89	0,62	0,37	0,75
altküme_33	71,554	45,22	0,75	0,88	0,59	0,37	0,73
altküme_32	71,7341	45,2798	0,75	0,88	0,6	0,36	0,74
altküme_31	71,283	48,8424	0,76	0,85	0,57	0,43	0,73
altküme_30	71,7341	45,2569	0,75	0,88	0,6	0,36	0,74
altküme_29	71,0951	47,7714	0,75	0,85	0,57	0,41	0,73
altküme_28	70,9889	43,5408	0,74	0,88	0,58	0,35	0,72
altküme_27	70,781	43,1652	0,74	0,88	0,57	0,35	0,72
altküme_26	70,0881	42,0016	0,74	0,87	0,55	0,34	0,71
altküme_25	69,8001	41,6922	0,74	0,87	0,55	0,34	0,71
altküme_24	69,8586	41,1473	0,73	0,87	0,55	0,33	0,70
altküme_23	69,8078	43,0066	0,74	0,86	0,54	0,36	0,70
altküme_22	69,6261	41,9311	0,74	0,86	0,54	0,34	0,70
altküme_21	70,1635	36,7777	0,72	0,9	0,57	0,27	0,71
altküme_20	69,4937	43,8276	0,74	0,85	0,53	0,37	0,70
altküme_19	70,0819	39,6621	0,73	0,89	0,56	0,31	0,70
altküme_18	69,7817	39,7359	0,73	0,88	0,55	0,31	0,70
altküme_17	69,9726	38,9888	0,73	0,89	0,56	0,3	0,70
altküme_16	69,7663	40,3934	0,73	0,88	0,55	0,32	0,70
altküme_15	68,2172	23,1334	0,7	0,93	0,51	0,15	0,66
altküme_14	67,8692	26,2728	0,7	0,91	0,5	0,18	0,66
altküme_13	68,0864	25,5852	0,7	0,92	0,51	0,17	0,66
altküme_12	67,5875	30,1685	0,71	0,89	0,49	0,22	0,66
altküme_11	68,0787	22,9016	0,7	0,93	0,51	0,15	0,67
altküme_10	67,9693	25,4302	0,7	0,92	0,5	0,17	0,67
altküme_9	67,9493	15,2656	0,69	0,96	0,5	0,09	0,65
altküme_8	68,0094	18,3398	0,69	0,95	0,5	0,11	0,65
altküme_7	67,8523	11,233	0,69	0,97	0,49	0,06	0,64
altküme_6	67,9585	0	0,68	1	0	0	0,64
altküme_5	67,9585	0	0,68	1	0	0	0,63
altküme_4	67,9585	0	0,68	1	0	0	0,63
altküme_3	67,9585	0	0,68	1	0	0	0,63
altküme_2	67,9585	0	0,68	1	0	0	0,63
altküme_1	67,9585	0	0,68	1	0	0	0,63

XGB modeller, öznitelik sayısı eklendikçe iyileşme göstermektedir. Hem ağaç kökenli oluşu hem de hatayı her iterasyonda hesaba katması sayesinde modelin doğruluk oranı artmaktadır. Ancak F ölçütü değerinde doygunluk olabileceği, daha fazla öznitelik eklense de bir iyileşme gözlenmeme ihtimali olabilir.

Çizelge 5.55 : Ki-kare temelli yöntemle oluşturulmuş öznitelik önemleri – XGB
sınıflayıcısı.

XGB	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	74,284	50,1478	0,76	0,9	0,66	0,4	0,78
altküme_37	74,2671	50,2945	0,76	0,9	0,66	0,41	0,78
altküme_36	74,2378	50,1594	0,76	0,9	0,66	0,4	0,78
altküme_35	74,2378	50,1594	0,76	0,9	0,66	0,4	0,78
altküme_34	74,2378	50,1594	0,76	0,9	0,66	0,4	0,78
altküme_33	73,6635	45,7223	0,75	0,92	0,67	0,35	0,77
altküme_32	73,6635	45,7223	0,75	0,92	0,67	0,35	0,77
altküme_31	73,5557	45,3405	0,75	0,92	0,67	0,34	0,77
altküme_30	73,6327	45,6036	0,75	0,92	0,67	0,34	0,77
altküme_29	73,6388	45,6094	0,75	0,92	0,67	0,34	0,77
altküme_28	73,4956	44,9976	0,75	0,92	0,67	0,34	0,77
altküme_27	73,4956	44,9976	0,75	0,92	0,67	0,34	0,77
altküme_26	72,6949	40,9077	0,74	0,93	0,67	0,29	0,76
altküme_25	71,7726	36,7076	0,73	0,94	0,65	0,26	0,74
altküme_24	71,7726	36,7076	0,73	0,94	0,65	0,26	0,74
altküme_23	71,7726	36,7076	0,73	0,94	0,65	0,26	0,74
altküme_22	71,7741	36,7133	0,73	0,94	0,65	0,26	0,74
altküme_21	71,8019	36,8016	0,73	0,94	0,65	0,26	0,74
altküme_20	71,8496	37,0888	0,73	0,94	0,65	0,26	0,74
altküme_19	71,8496	37,0888	0,73	0,94	0,65	0,26	0,74
altküme_18	71,5555	35,5511	0,72	0,94	0,65	0,24	0,73
altküme_17	71,5278	35,3891	0,72	0,94	0,65	0,24	0,73
altküme_16	71,5771	35,5864	0,72	0,94	0,65	0,25	0,73
altküme_15	68,807	12,4092	0,69	0,98	0,62	0,07	0,69
altküme_14	68,7823	12,77	0,69	0,98	0,61	0,07	0,69
altküme_13	68,8024	12,7245	0,69	0,98	0,61	0,07	0,69
altküme_12	68,8362	13,3419	0,69	0,98	0,61	0,07	0,69
altküme_11	68,5668	11,8033	0,69	0,98	0,58	0,07	0,68
altküme_10	68,576	12,2538	0,69	0,98	0,58	0,07	0,68
altküme_9	68,4189	10,3897	0,69	0,98	0,57	0,06	0,68
altküme_8	68,3651	8,5385	0,69	0,98	0,58	0,05	0,68
altküme_7	67,9632	0,5449	0,68	1	0,51	0	0,65
altküme_6	67,9585	0	0,68	1	0	0	0,64
altküme_5	67,9585	0	0,68	1	0	0	0,63
altküme_4	67,9585	0	0,68	1	0	0	0,63
altküme_3	67,9585	0	0,68	1	0	0	0,63
altküme_2	67,9585	0	0,68	1	0	0	0,63
altküme_1	67,9585	0	0,68	1	0	0	0,63

Extra trees metodunda, model ağaç kökenli olduğundan yeni öznitelik eklendikçe iyileşme göstermesi beklenmiştir ve Çizelge 5.56’da gözlenmiştir. Ancak doygunluğa ulaşmış olabilir. 38 öznitelik arasından en iyi sonucu 37 öznitelik sayısı ile vermiştir.

Çizelge 5.56 : Ki-kare temelli yöntemle oluşturulmuş öznelilik önemleri – Extra trees sınıflayıcısı.

Extra Trees	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
altküme_38	69,389	39,743	0,73	0,87	0,54	0,32	0,69
altküme_37	69,4814	40,5412	0,73	0,87	0,54	0,32	0,69
altküme_36	69,4629	40,3226	0,73	0,87	0,54	0,32	0,69
altküme_35	69,4999	40,4808	0,73	0,87	0,54	0,32	0,69
altküme_34	69,62	40,4144	0,73	0,87	0,54	0,32	0,69
altküme_33	69,1719	39,0291	0,73	0,87	0,53	0,31	0,68
altküme_32	69,3505	39,1861	0,73	0,88	0,54	0,31	0,68
altküme_31	68,5637	38,2979	0,73	0,87	0,52	0,3	0,67
altküme_30	68,3435	38,5583	0,73	0,86	0,51	0,31	0,67
altküme_29	68,7284	39,0065	0,73	0,86	0,52	0,31	0,67
altküme_28	68,1202	37,7511	0,72	0,86	0,5	0,3	0,67
altküme_27	68,1972	38,3315	0,72	0,86	0,51	0,31	0,67
altküme_26	67,1856	37,8525	0,72	0,84	0,48	0,31	0,65
altküme_25	66,3864	37,625	0,72	0,83	0,46	0,32	0,64
altküme_24	66,6513	37,6784	0,72	0,83	0,47	0,31	0,64
altküme_23	66,6559	37,509	0,72	0,83	0,47	0,31	0,64
altküme_22	66,6082	37,3816	0,72	0,83	0,47	0,31	0,64
altküme_21	66,2833	37,5496	0,72	0,83	0,46	0,32	0,64
altküme_20	66,0815	37,6578	0,72	0,82	0,46	0,32	0,63
altküme_19	65,8783	37,2664	0,72	0,82	0,45	0,32	0,63
altküme_18	65,9507	37,7957	0,72	0,82	0,46	0,32	0,63
altküme_17	65,6073	38,2472	0,72	0,81	0,45	0,33	0,63
altküme_16	65,9337	38,5752	0,72	0,81	0,46	0,33	0,63
altküme_15	62,7587	37,4974	0,71	0,76	0,41	0,35	0,58
altküme_14	62,8772	37,8554	0,71	0,76	0,41	0,35	0,58
altküme_13	62,7402	37,7591	0,71	0,76	0,41	0,35	0,58
altküme_12	62,7186	37,768	0,71	0,76	0,41	0,35	0,58
altküme_11	62,5123	37,8231	0,71	0,75	0,4	0,36	0,58
altküme_10	62,3707	37,604	0,71	0,75	0,4	0,35	0,58
altküme_9	61,0218	36,8413	0,71	0,73	0,38	0,35	0,57
altküme_8	61,2497	36,921	0,71	0,73	0,39	0,35	0,57
altküme_7	67,726	12,2205	0,69	0,96	0,48	0,07	0,64
altküme_6	67,9585	0	0,68	1	0	0	0,64
altküme_5	67,9585	0	0,68	1	0	0	0,63
altküme_4	67,9585	0	0,68	1	0	0	0,63
altküme_3	67,9585	0	0,68	1	0	0	0,63
altküme_2	67,9585	0	0,68	1	0	0	0,63
altküme_1	67,9585	0	0,68	1	0	0	0,63

5.4 Öznitelik Altkümelerinin En İyi Sonuçlarının Karşılaştırılması

Bölüm 5.3'te her bir modele sırasıyla en önemliden başlayarak özniteliklerin birer birer eklendiğindeki doğruluk hesaplarının çizelgeleri verilmiştir. Bu çizelgelerden her model ve her öznitelik seçim yöntemi için en yüksek doğruluk değerini veren altkümeler bu bölümde karşılaştırılacaktır.

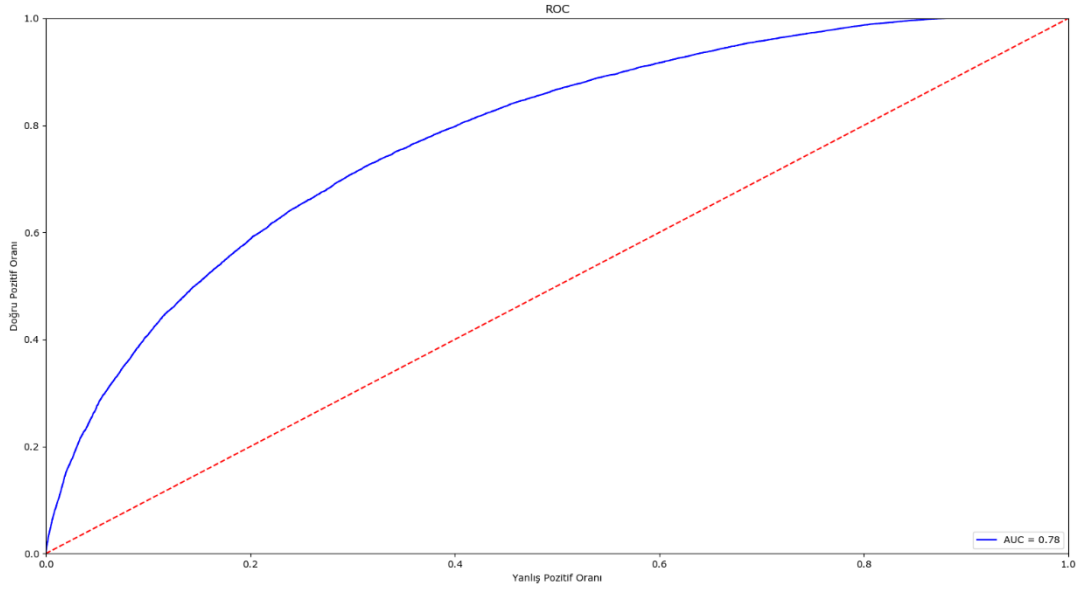
5.4.1 Ağaç modellerinin öznitelik veri kümesi en iyi alt kümesinin modeller üzerinde çalıştırılması

Bölüm 5.3.1'de verilmiş olan ağaç temelli yöntemler ile öznitelik önem listelerinin çalışmada kullanılan modellere sırasıyla verilmesiyle elde edilmiş olan sonuçlar arasından her model için en iyi sonuç Çizelge 5.57'de listelenmiştir. Bu çizelge incelendiğinde bu çizelgedeki en iyi sonucu veren modelin koyu renkli olarak işaretlendiği görülecektir. Bu değer ilk 33 öznitelik ile '*Extra gradyen boosting*' algoritmasından elde edilen 74,28 olmuştur.

Çizelge 5.57 : Ağaç modellerinin karşılaştırılması.

Ağaç temelli	Altküme no	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
Lojistik Regresyon	altküme_38	70,1266	27,9122	0,71	0,95	0,62	0,18	0,72
Naive Bayes	altküme_2	62,5077	51,7775	0,78	0,62	0,44	0,63	0,65
Karar Ağaçları	altküme_21	67,9508	49,637	0,77	0,76	0,49	0,5	0,63
K En yakın komşu	altküme_7	68,9702	44,6374	0,74	0,83	0,52	0,39	0,67
Gradyen Boosting	altküme_19	74,2162	49,8938	0,76	0,9	0,66	0,4	0,78
Adaboost	altküme_17	73,5541	49,9257	0,76	0,89	0,63	0,41	0,77
Bagging	altküme_32	73,7174	50,1154	0,76	0,89	0,64	0,41	0,75
Rastgele Orman	altküme_21	72,4978	46,5799	0,75	0,89	0,62	0,37	0,74
YSA	altküme_8	72,9967	49,6887	0,76	0,88	0,62	0,42	0,76
XGB	altküme_33	74,284	50,1478	0,76	0,9	0,66	0,4	0,8
Extra Trees	altküme_7	71,1644	45,3019	0,75	0,87	0,58	0,37	0,71

XGB için ağaç öznitelik önemine göre seçilmiş ilk 33 öznitelik ile çizdirilmiş ROC eğrisi Şekil 5.2'de gösterilmektedir.



Şekil 5.2 : Ağaç öznitelik sıralaması ile XGB modeli ROC grafiği.

Çizelge 5.57’de yer alan en yüksek değere sahip XGB modelinin çalıştırılmasıyla elde edilen karmaşıklık matrisinin değerleri Çizelge 5.58’de yer almaktadır.

Çizelge 5.58 : Ağaç öznitelik sıralaması ile karmaşıklık matrisi.

XGB karmaşıklık matrisi	Ödemez (Tahmin)	Öder (Tahmin)
Ödemez	39843	4292
Öder	12409	8400

5.4.2 Temel bileşen analizi yöntemiyle öznitelik veri kümesinin en iyi alt kümesinin modeller üzerinde çalıştırıldığı metrikler

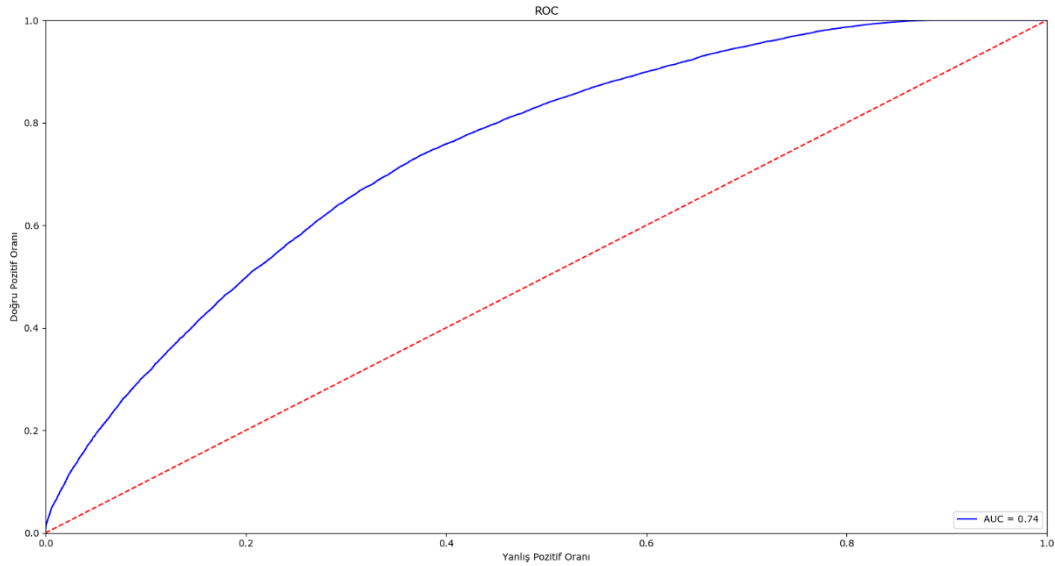
Bölüm 5.3.2’de verilmiş olan temel bileşen analizi ile öznitelik önem listelerinin çalışmada kullanılan modellere sırasıyla verilmesiyle elde edilmiş olan sonuçlar arasından her model için aşağıda Çizelge 5.59’de listelenmiştir. Bu çizelge incelendiğinde bu çizelgedeki en iyi sonucu veren modelin koyu renkli olarak işaretlendiği görülecektir. Bu değer ilk 33 öznitelik ile koyu renk ile belirtilen ‘gradyen boosting’ ve ‘extra gradyen boosting’ algoritmalarından elde edilen 71,04 ve 71,09 sonuçları olmuştur. F-ölçütü değerleri de dikkate alınarak karşılaştırma yapılmıştır.

Çizelge 5.59 : Temel bileşen analizi modellerinin karşılaştırılması.

Temel bileşen analizi	Altküme no	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
Lojistik Regresyon	altküme_33	69,2951	20,2775	0,7	0,96	0,6	0,12	0,7

Naive Bayes	altküme_6	60,5968	35,0903	0,7	0,73	0,37	0,33	0,59
Karar Ağaçları	altküme_15	65,4964	45,3597	0,75	0,75	0,45	0,45	0,6
K En yakın komşu	altküme_5	67,3242	40,8342	0,73	0,82	0,49	0,35	0,65
Gradyen Boosting	altküme_33	71,035	38,9432	0,73	0,91	0,6	0,29	0,7
Adaboost	altküme_33	70,7687	41,1093	0,73	0,89	0,58	0,32	0,74
Bagging	altküme_32	69,7093	42,5367	0,74	0,86	0,54	0,35	0,71
Rastgele Orman	altküme_33	69,2243	40,2606	0,73	0,87	0,53	0,32	0,69
YSA	altküme_17	70,2775	37,1094	0,73	0,91	0,58	0,27	0,73
XGB	altküme_33	71,0905	38,2584	0,73	0,91	0,61	0,28	0,7
Extra Trees	altküme_33	67,1794	38,4831	0,72	0,84	0,48	0,32	0,66

Bu iyi sonuçlar arasından gradyen boosting için 33 öznitelik ile ROC eğrisi çizdirilirse Şekil 5.3'teki gibi bir görsel elde edilmiş olacaktır.



Şekil 5.3 : TBA öznitelik sıralaması ile Gradyen Boosting modeli ROC grafiği.

Çizelge 5.60'da temel bileşen analizine göre en iyi sonucu veren gradyen boosting algoritmasının 33 öznitelik ile yaptığı tahminler sonucunda karmaşıklık matrisinin değerleri listelenmiştir.

Çizelge 5.60 : TBA öznitelik sıralaması ile karmaşıklık matrisi.

Gradyen Boost. Karmş. matrisi	Ödemez (Tahmin)	Öder (Tahmin)
Ödemez	40352	3783
Öder	14992	5817

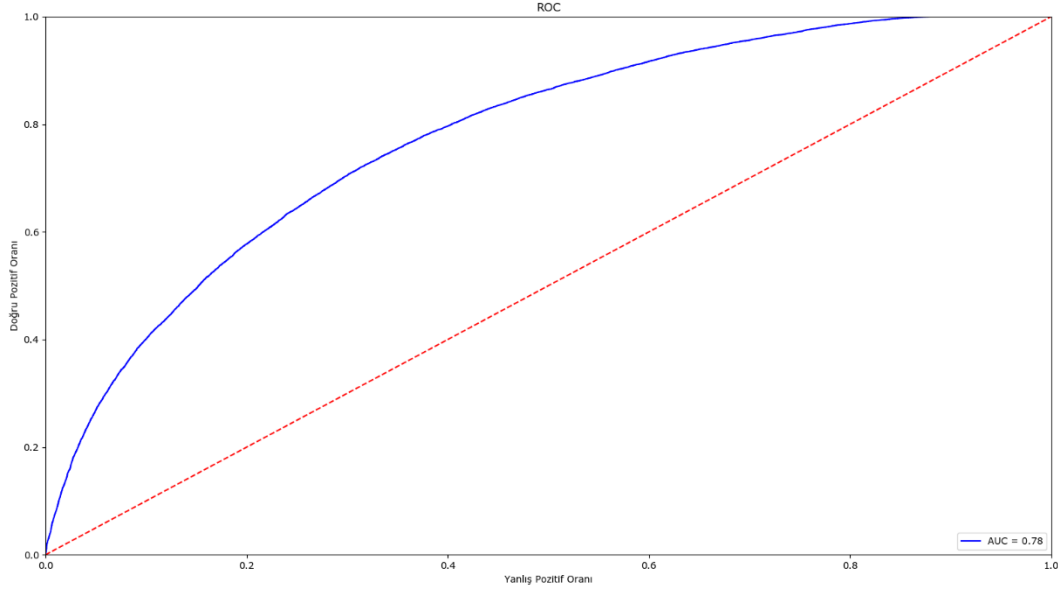
5.4.3 Lojistik regresyon ile özyinelemeli öznitelik veri kümesi en iyi alt kümenin modeller üzerinde çalıştırıldığında metrikler

Bölüm 5.3.3'te verilen özyinelemeli olarak lojistik regresyon uygulanarak listelenmiş olan öznitelik önem listelerinden elde edilen en iyi sonuçlar Çizelge 5.61'de listelenmiştir. Bu çalışma için özyinelemeli modellerde baz küme sayısı olarak 5 verilmiştir. Bu nedenle ilk altküme diğer yöntemlerde bir özniteliğe sahipken, özyinelemeli modellerde 5 özniteliğe sahiptir. Bu değerler arasından en iyi sonucu 30. altküme ile 34 özniteliğe sahip iken veren 'gradyen boosting' algoritması koyu renk ile belirtilmiştir. Ulaşılan doğruluk değeri 74,01'dir.

Çizelge 5.61 : Özyinelemeli lojistik regresyon modellerinin karşılaştırılması.

Özyinelemeli Lojistik	Altküme no	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
Lojistik Regresyon	altküme_34	69,7586	25,4649	0,71	0,95	0,61	0,16	0,71
Naive Bayes	altküme_1	44,5137	51,3481	0,85	0,22	0,36	0,91	0,6
Karar Ağaçları	altküme_34	67,5289	49,2076	0,76	0,76	0,49	0,49	0,63
K En yakın komşu	altküme_7	68,7762	45,772	0,75	0,82	0,52	0,41	0,68
Gradyen Boosting	altküme_30	74,0145	50,1977	0,76	0,9	0,65	0,41	0,8
Adaboost	altküme_34	73,0121	49,0153	0,76	0,88	0,62	0,4	0,77
Bagging	altküme_34	73,3509	49,5731	0,76	0,89	0,63	0,41	0,75
Rastgele Orman	altküme_34	72,1191	45,4888	0,75	0,89	0,61	0,36	0,74
YSA	altküme_29	73,4602	51,3107	0,77	0,88	0,62	0,44	0,76
XGB	altküme_34	73,9745	49,7861	0,76	0,9	0,65	0,4	0,78
Extra Trees	altküme_6	70,7764	45,6827	0,75	0,86	0,56	0,38	0,71

Gradyen Boosting modeli için 35 öznitelik ile elde edilen sonuçların ROC eğrisi Şekil 5.4'te verilmiştir.



Şekil 5.4 : Özyinelemeli lojistik reg. öznitelik sıralaması ile Gradyen Boosting modeli ROC grafiği.

Gradyen Boosting modeli için 35 öznitelik ile elde edilen sonuçların test veri kümesi ile olan karmaşıklık matrisi sonuçları Çizelge 5.62’de verilmiştir.

Çizelge 5.62 : Özyinelemeli lojistik reg. öznitelik sıralaması ile karmaşıklık matrisi.

Gradyen Boost. Karmş. matrisi	Ödemez (Tahmin)	Öder (Tahmin)
Ödemez	39563	4572
Öder	12304	8505

5.4.4 Adaboost ile özyinelemeli öznitelik veri kümesi en iyi alt kümenin modeller üzerinde çalıştırıldığında metrikler

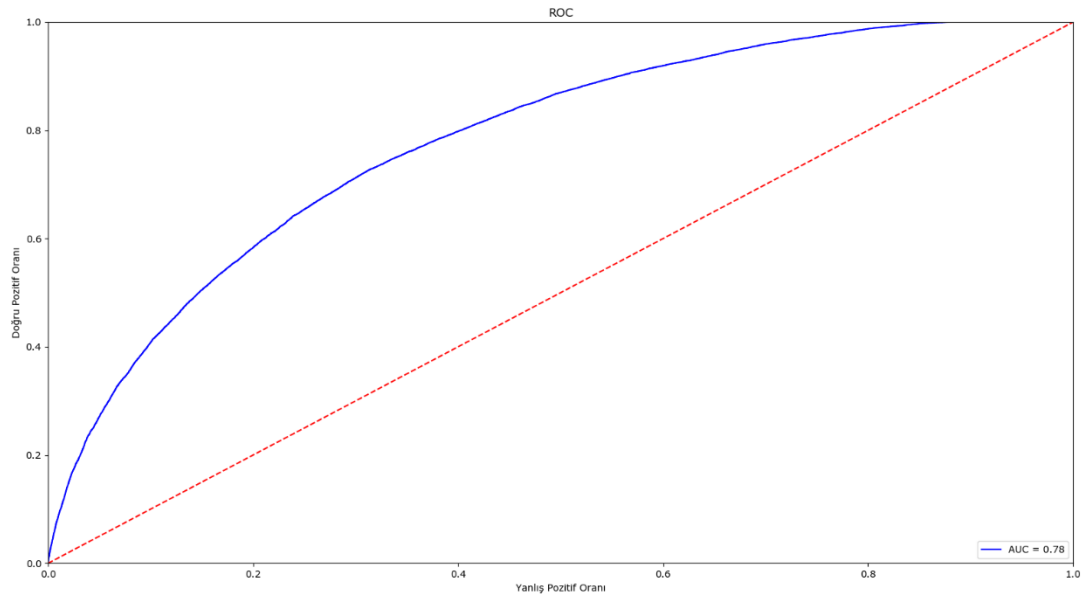
Bölüm 5.3.4’te verilen özyinelemeli olarak Adaboost uygulanarak listelenmiş olan öznitelik önem listelerinden elde edilen en iyi sonuçlar Çizelge 5.63’te listelenmiştir. Bu çalışma için özyinelemeli modellerde baz küme sayısı olarak 5 verilmiştir. Bu nedenle ilk altküme diğer yöntemlerde bir özniteliğe sahipken, özyinelemeli modellerde 5 özniteliğe sahiptir. Bu değerler arasından en iyi sonuçlar 31. altküme ile 35 özniteliğe sahip iken veren ‘gradyen boosting’ ve 34. altküme ile 38 öznitelik ile ‘extra gradyen boosting’ algoritmaları koyu renk ile belirtilmiştir. Ulaşılan doğruluk değeri sırasıyla 74,25 ve 74,28’dir. F-ölçütü de değerlendirilmede kullanılmıştır.

Çizelge 5.63 : Adaboost ile özyinelemeli modellerin karşılaştırılması.

Özyinelemeli Adaboost	Altküme no	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
-----------------------	------------	----------	----------	------------	-------------	------------	-------------	-----

Lojistik Regresyon	altküme_30	70,1728	28,0397	0,71	0,95	0,62	0,18	0,72
Naive Bayes	altküme_10	59,3111	48,2826	0,76	0,59	0,41	0,59	0,64
Karar Ağaçları	altküme_30	67,9108	49,8605	0,76	0,76	0,5	0,5	0,63
K En yakın komşu	altküme_1	69,5907	46,5911	0,75	0,83	0,53	0,41	0,69
Gradyen Boosting	altküme_31	74,2517	50,5705	0,76	0,9	0,66	0,41	0,8
Adaboost	altküme_31	73,3755	49,7866	0,76	0,89	0,63	0,41	0,77
Bagging	altküme_17	73,5233	50,1117	0,76	0,89	0,63	0,42	0,75
Rastgele Orman	altküme_13	72,6811	47,083	0,75	0,89	0,62	0,38	0,74
YSA	altküme_4	73,608	50,4051	0,76	0,89	0,63	0,42	0,77
XGB	altküme_34	74,284	50,1478	0,76	0,9	0,66	0,4	0,8
Extra Trees	altküme_1	70,5023	45,6524	0,75	0,86	0,56	0,39	0,7

Daha az öznitelik sayısı ile yüksek doğruluk değerine ulaşan Gradyen Boosting için 35 öznitelik (ilk altküme 5 öznitelik içeriyor) ile elde edilen sonuçlara ait ROC grafiği Şekil 5.5'te verilmiştir.



Şekil 5.5 : Özyinelemeli adaboost öznitelik sıralaması ile Gradyen Boosting modeli ROC grafiği.

Gradyen boosting modeli ile elde edilen sonuçların test veri kümesi üzerinde çalıştırıldığında elde edilen sonuçların karmaşıklık matrisine konumlandırılmış hali Çizelge 5.64'te verilmiştir.

Çizelge 5.64 : Özyinelemeli adaboost öznitelik sıralaması ile karmaşıklık matrisi.

Gradyen Boost. Karmş. matrisi	Ödemez (Tahmin)	Öder (Tahmin)
Ödemez	39668	4467
Öder	12255	8554

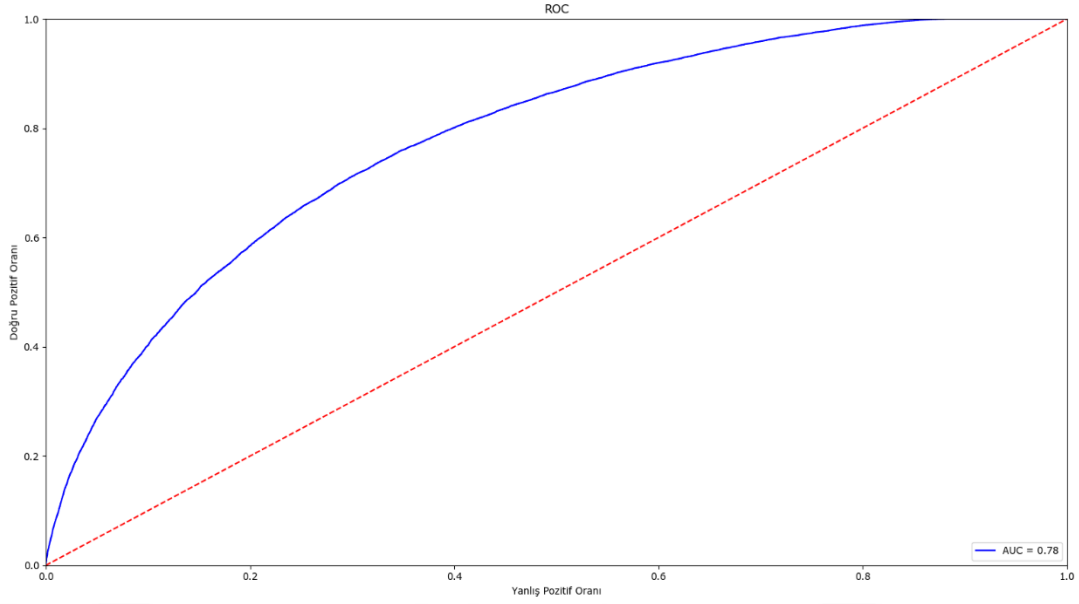
5.4.5 Ki-kare temelli öznitelik veri kümesi en iyi alt kümenin modeller üzerinde çalıştırıldığında metrikler

Bölüm 5.3.5’te verilen ki-kare değerine göre öznitelik önem derecelerine karar verilmiş olan listenin modeller üzerinde uygulanmasıyla elde edilmiş en iyi sonuçlar (her bir model için ayrı) Çizelge 5.65’te gösterilmiştir. En yüksek değere 37 öznitelik sayısı ile 74,16 doğruluk oranıyla ‘gradyen boosting’ algoritmasıyla erişilmiştir.

Çizelge 5.65 : Ki-kare modellerinin karşılaştırılması.

Ki-kare	Altküme no	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
Lojistik Regresyon	altküme_38	70,1266	27,9122	0,71	0,95	0,62	0,18	0,72
Naive Bayes	altküme_26	42,9724	52,2042	0,93	0,17	0,36	0,97	0,64
Karar Ağaçları	altküme_34	67,8631	49,4441	0,77	0,76	0,49	0,49	0,63
K en yakın komşu	altküme_16	65,6165	39,9117	0,72	0,8	0,45	0,36	0,64
Gradyen Boosting	altküme_37	74,1608	50,5525	0,76	0,9	0,65	0,41	0,8
Adaboost	altküme_34	73,3755	49,7866	0,76	0,89	0,63	0,41	0,77
Bagging	altküme_35	73,6203	50,1861	0,76	0,89	0,64	0,41	0,75
Rastgele Orman	altküme_35	72,3362	45,9734	0,75	0,89	0,61	0,37	0,74
YSA	altküme_34	72,4578	46,1121	0,75	0,89	0,62	0,37	0,75
XGB	altküme_38	74,284	50,1478	0,76	0,9	0,66	0,4	0,78
Extra Trees	altküme_34	69,62	40,4144	0,73	0,87	0,54	0,32	0,69

Gradyen Boosting için 37 öznitelik kullanılarak elde edilen sonuçlardan ROC eğrisi Şekil 5.6’daki gibi çizdirilmiştir.



Şekil 5.6 : Ki-kare öznitelik sıralaması ile Gradyen Boosting modeli ROC grafiği.

Modelin test veri kullanılarak çalıştırılmasıyla elde edilen sonuçlardan Çizelge 5.66'daki karmaşıklık matrisi elde edilmiştir.

Çizelge 5.66 : Ki-kare öznitelik sıralaması ile karmaşıklık matrisi.

Gradyen Boost. Karmş. matrisi	Ödemez (Tahmin)	Öder (Tahmin)
Ödemez	39585	4550
Öder	12231	8578

5.5 Tüm Öznitelik Seçim Yöntemlerinin En İyi Model Performanslarının Birbirleri ile Karşılaştırması

Yukarıdaki modellerde yer alan tablolardaki koyu renkle seçilmiş olan ve diğer altküme ve modellere oranla daha yüksek doğruluk gösteren modeller aşağıda Çizelge 5.67'de gösterilmiştir. Bunlar arasından da en iyi sonuç veren öznitelik kümesi seçilecek olursa, 33 öznitelik ile Ağaç temelli öznitelik sıralamasıyla Extra Gradyen Boosting %74,28 doğruluk ile modellenebilmiştir.

Çizelge 5.67 : Tüm özniteliklerin en iyi performansları.

Öznitelik Seçim Yöntemi	En iyi performanslı model	Altküme no	Doğruluk	F Ölçütü	Kesinlik-0	Hassaslık-0	Kesinlik-1	Hassaslık-1	ROC
Ağaç temelli	XGB	altküme_33	74,28	50,1478	0,76	0,9	0,66	0,4	0,78
Temel bileşen analizi	Gradyen Boosting	altküme_33	71,04	38,9432	0,73	0,91	0,6	0,29	0,74

Özyinelemeli Lojistik	Gradyen Boosting	altküme_30	74,02	50,1977	0,76	0,9	0,65	0,41	0,78
Özyinelemeli Adaboost	Gradyen Boosting	altküme_31	74,25	50,5705	0,76	0,9	0,66	0,41	0,78
Ki-kare	Gradyen Boosting	altküme_37	74,16	50,5525	0,76	0,9	0,65	0,41	0,78

“Kesinlik-0” kolonu, borcunu ödememiş müşterilerin model tarafından ne kadarlık bir başarıyla tespit edildiğini ölçer. “Hassaslık-0” kolonu ile modelin gerçekte ödeme yapmamış olan müşteriler arasından tahsilat gerçekleştirmenin olanları ayırma kabiliyetini ölçer. “Kesinlik-1” kolonu, borcunu ödememiş müşterilerin model tarafından ne kadarlık bir başarıyla tespit ettiğini belirtir. “Hassaslık-1” kolonu ise, modelin borcunu ödeyen müşterileri diğerlerinden ayrılma kabiliyetini ölçer.



6. SONUÇLAR VE ÖNERİLER

Bu çalışmada, bir telekomünikasyon şirketine borçlu müşteriler incelenmiştir. Borcunu 120 gün ödememiş ve yasal takibe girmiş müşteriler için 2018 yılına ait veri kümesi üzerinde bir tahmin modeli geliştirilmeye çalışılmıştır.

Genellikle, bankacılık ve finans sektörlerinde kredi başvuruları için müşteriler üzerine tahminler yapılıyor olsa da son yıllarda müşterilerinin kredibilitesini ölçmeye çalışan ve müşterisini daha iyi tanımak isteyen çoğu şirket tarafından kullanılmaktadır. Müşterinin borcunu ödeyip ödemeyeceği yönündeki tahminler için makine öğrenmesi yöntemlerinin kullanılmaya başlanmasındaki amaç; çok fazla miktarda, birbirinden farklı biçimlerde ve dağınık olarak depolanan veriden elde edilen bilgilerle uygulanabilir stratejik kararlar alabilmektir.

Müşterilerin davranışlarının yorumlanması, çoğunlukla finansal kuruluşlar için; kredi talep eden müşterilerinin finansal yükümlülüklerini yerine getirme kabiliyetleri veya müşteri sadakati açısından kullanılmasının yanı sıra, bu çalışma kapsamında telekomünikasyon müşterileri için müşterinin ödemediği borcunu ödeme ihtimalini hesaplamak için kullanılmıştır. Bu yöntemlerden elde edilecek sonuçlar kullanılarak şirkete uygun risk stratejileri hayata geçirilebilir. Makine öğrenmesi, istatistiksel analiz, yapay zeka, geliştirilebilir zeka ve veri görselleştirme tekniklerini kullanarak büyük veri yığınları arasında gizli kalan kalıplar ve ilişkiler bu yöntemler sayesinde ortaya çıkarılabilmektedir. Daha iyi ve nitelikli kararlar alabilmek ve daha etkin stratejiler geliştirebilmek için geleceği öngörmeyi ve hızlı aksiyon almayı sağlayacak araçlar kullanılmalıdır. Çok büyük veri yığınlarına sahip mobil telekomünikasyon sektöründe riskli müşterilerin tespit edilmesi de ancak güçlü bir uygulama metodolojisine sahip, doğru veri altyapısı üzerine kurulmuş güçlü yöntemler ve süreçler ile başarılabılır.

Çalışılan şirketin amacı, mümkün olduğunca objektif değerlendirme yapabilmek için şirketlere atanacak müşterilerin, eşit zorlukta olmasını temin edebilmektir. Yanlış atamalar ile hem birlikte çalışılan ortaklar açısından (paydaş) güven probleminin

yaşanmasına, hem de adil olmayan bir sistem ile borçlulardan temin edilebilecek olan geliri temin edememe ihtimali oluşabilir.

Uygulamada makine öğrenmesi modellerinde kullanılabilecek kadar yeterli gözlem ve öznitelik mevcuttur. Çalışmada, Lojistik Regresyon (LR), Karar Ağacı (KA), Naive Bayes, K-en yakın komşu, Adaboost, Bagging, Rastgele Ormanlar, Gradyen Boosting, XGB, Extra Ağaçlar ve Yapay Sinir Ağı (YSA) modelleri kullanılmıştır. Bunlara ilave öznitelik seçim yöntemleri olarak, ağaç temelli seçim yöntemlerinin ortalama değerleri, Temel Bileşen Analizi, Özyinelemeli Lojistik Regresyon, Özyinelemeli Adaboost, ki-kare testlerinden elde edilen öznitelik önem listeleri en etkiliden en az etkiliye doğru sırasıyla, öznitelik sayısı artırılarak modelde işlenmiştir.

Eğitim veri kümesiyle model eğitildikten sonra, test veri kümesine de uygulanması ile modelin tutarlılığı ve doğruluğunun gözlenebilmesi için ölçümler yapılmıştır. Bu ölçümler sonucunda karar ağaçları ve ağaç temelli boosting yöntemleri ile seçilmiş öznitelik alt kümesi karar ağaçları modelleri üzerinde yüksek oranlarda başarı sağlayamazken, boosting modeller üzerinde en iyi sonuçları veren alt küme olmuştur. En iyi sonuç ağaç temelli yöntemler kullanılarak elde edilen öznitelik sıralamasının extra gradyen boosting algoritmasıyla 74,28 doğruluk oranına erişilmiştir. 38 öznitelik kullanmak yerine 33 öznitelik ile 38 öznitelik ile alacağımızdan daha yüksek doğruluk değerine ulaşılmıştır. Hem modelin çalışma süresi kısalmıştır hem de doğruluğu artmıştır. Dahil edilmeyen öznitelikler; transfer_maas_yatma_araligi, transfer_gunu_araligi, bayram_var, yasal_faiz, yasal_faiz_bolu_toplam'dir.

Ayrıca literatürde görülmüştür ki, veri kümesi boyutunu düşürme ve özellik seçiminin model üzerindeki çalışmalarda sonuçları daha istikrarlı hale getirdiği gözlemlenmiştir. Modellerin doğruluğunu artırıp, modelin çalışma süresini kısaltmaktadır. Öznitelik seçme işlemi, çalışma süresinin uzamasına neden olsa da uzun vadede fayda sağlamaktadır.

Naive Bayes algoritmasında, kullanılan bütün özniteliklerin birbiri ile ilgili bilgi bulundurmadığı, birbirinden bağımsız ve aynı derecede öneme sahip olduğu kabul edilmektedir. Örüntü tanıma, makine öğrenmesi gibi çok sayıda alanda, kolay öğrenilebilmesi ve yüksek doğruluk oranı ile sınıflandırması nedeniyle kullanılan etkili bir sınıflandırıcıdır. Bu çalışmada, öznitelikler birbirinden bağımsız olmadığından ve aynı derece öneme sahip olmadığından sınıflandırma algoritmasının doğruluk oranı

düşmektedir. Öznitelikler arası korelasyon matrisi Şekil 4.21 numaralı şekilden incelenebilir. Ayrıca, NB parametrik olmayan sürekli özniteliklerle ilgilenmez Bu çalışmada da tutar miktarı, müşterilik süresi, toplam ödenmiş fatura sayısı vb. gibi öznitelikler sürekli değeri bulunan özniteliklerdir. Ancak öznitelik sıralamasında etkili olduğu görülen özniteliklerdir. Bu nedenle, Naive Bayes bu çalışmada kullanılan veri kümesi için yüksek doğruluk oranlarında sonuç üretmemiştir.

Tekli modellerde öznitelik sayısı en yüksek performansa ulaştıktan sonra eklenmeye devam ederse, performansta düşüş gözlenmeye başlanmaktadır. Ancak artırıcı modellerde bu durum tam tersidir. Ya doğruluk değeri sabit kalmakta veya öznitelik sayısı arttıkça modelin bilgi kazanımı arttığı için doğruluk değerinde artış gözlenmektedir. Tekli modellerden, lojistik regresyon’da öznitelik eklendikçe boosting metotlar gibi doğruluk oranında iyileşme görülmekte veya sabit kalmaktadır. Ancak doğruluk değeri, boosting modeller kadar yüksek olmasa da, davranış biçimi benzerlik göstermektedir.

Python programlama diline veriler aktarılırken, tarih alanları için bir veri tipi bulunmamaktadır. Bu alan gün, ay, yıl olarak ayrı ayrı kolonlar halinde verilmektedir. Ancak bu çalışmada firmanın yürüttüğü süreçler iyi incelenmiş olduğundan, tarih alanıyla ilgili olarak model ile daha açıklayıcı bilgiler paylaşılmıştır. Bu bilgilerde modele öznitelik olarak verildikten sonra modelin doğruluk değerinde ~%4’lük artış gözlenmiştir. Bu öznitelikler; atama gerçekleştikten sonra ay sonuna kadarki iş günü sayısı, atamanın yapıldığı mevsim, atamanın haftanın kaçınıcı günü yapıldığı, o aydaki toplam iş günü sayısı, atama yapıldığı ayda bayram olup olmadığı gibi.

Bölüm 3.6.2.1’de değinildiği üzere Breiman’a göre, Adaboost, boosting sınıflandırma metotları içerisinde en iyisi olarak kabul edilmektedir. Bu çalışma kapsamında, bu yöntem, karşılaştırılan diğer yöntemler arasında en iyi sonucu vermemiş olsa da, doğruluk oranı oldukça yüksek, en iyi sonuca yakın sonuçlar üretmiştir. Bu çalışmada Extra gradyen boosting algoritmasının en iyi sonucu ürettiği görülmüştür. Ona en yakın sonuçlara da ‘gradyen boosting’ algoritmasıyla erişilmiştir.

Yaş ve medeni durum kategorizasyonu için iki farklı model önerilmiştir. Bu modellerden Çizelge 6.1’de de görüldüğü gibi, medeni durum için çoğunlukla ikinci seçenek tercih edilmiş. Birinci seçeneğin tercih edildiği durumlarda da ikinci seçenek ile arasında çok küçük farklar bulunduğu görülmüştür. Yaş kategorizasyonu açısından

ise, tüm öznitelik seçme algoritmaları için sadece ilk kategorinin seçildiği görülmüştür.

Çizelge 6.1 : Yaş ve medeni durum çok seçmeli özniteliklerinin seçim tercihleri.

İki seçenekli öznitelikler	Ağaç	TBA	Özyineleme - Lojistik	Özyineleme - Adaboost	Ki-kare
medeni_durum	0,108678	0,000007	17	22	249,952342
medeni_durum2	0,177225	0,000004	18	3	1787,225341
yas_kategori	0,164427	0,000002	26	6	84,159674
yas_kategori2	0,073120	0,000001	31	27	73,243521

Müşteri bir finansal kuruluşa kredi başvurusunda bulunduğunda, kendisi hakkındaki bilgileri, gelir durumunu ve mal varlığını bu kuruluşlar ile paylaşmaktadır. Ancak telekomünikasyon sektörü için böyle bir durum söz konusu değildir. Başvuran müşterinin sadece kimlik bilgileri istendiğinden dolayı müşteriler hakkında edinilen bilgiler oldukça azdır. Ayrıca çalışılan şirketin güvenlik politikası nedeniyle, müşteri hakkında sadece bu çalışma ile ilgilenen birimin kendi alanıyla ilgili bilgilere erişilebilmesi, diğer birimlerin tablolarına erişilememesi müşteri hakkında bilgi edinme sürecini zorlaştırmıştır. Müşteri hakkında; ödediği fatura tutarı, yasal takibe düşmeden önceki dönemlerde fatura ödeme tarihlerindeki gecikme gün sayısı, bağlı olduğu tarife adı.. gibi bilgilere erişilebilmesi modelin doğruluk oranı üzerinde etkili olabilir.

Müşterilerin geri ödeme yapıp yapmayacağı bilgisi hemen hemen hizmet sağlayan tüm şirketlerin ihtiyacıdır. Bu alanda, geliştirilmesi ve incelenmesi gereken pek çok bilgi mevcuttur. Ancak ülkemizde bu sürecin daha bilimsel temellere oturabilmesi ve elde edilen faydanın artırılabilmesi için bu bilgilere erişimin kolaylaştırılması büyük önem taşımaktadır. Bu kapsamda yapılabilecekler:

- Endüstri 4.0'a geçilmesiyle birlikte, şirket içerisindeki bilgilerin daha kolay ulaşılabilir olması ve veri tabanı normalizasyonu ile müşteriye ait tüm bilgilere erişebilme imkanının çalışanlara sağlanması
- Aynı sektör içerisinde bilgi paylaşımının artırılması (kredi kuruluşları arasındaki işbirliği gibi)
- Kişisel verilerin korunması yasası kapsamında; müşteri borçlarının, diğer şahıslar ile paylaşılması alanındaki yasal engellerin düzenlenmesi
- MERNIS, UYAP, TARBS, EGM gibi kamuya ait veritabanlarından faydalanarak, müşterinin adresi, çalışma durumu bilgisi, üzerinde başka

dava/haciz olduđu bilgisi, mal varlığı bilgisi gibi bilgiler ile müşteri hakkında daha doğru ve detaylı bilgiye erişilme imkanının oluşturulması





KAYNAKLAR

- Anderson, R.** (2007). *The Credit Scoring Toolkit*, Oxford University Press
- Arunasakthi, K., KamatchiPriya, L., Askerunisa, A.** (2014). Fisher Score Dimensionality Reduction for Svm Classification. *International Journal of Innovative Research in Science, Engineering and Technology* 3, 1900
- Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J.** (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, 54, 627-635
- Bhatia, N. & Vandana, S.** (2010). Survey of nearest neighbor techniques. *International Journal of Computer Science and Information Security*, 8(2), 302-305.
- Breiman, L.** (1996). Bagging predictors. *Machine Learning*, 24, 123-140.
- Breiman, L.** (2001). Random forests. *Machine Learning*, 45, 5-32.
- Bordi, I. & Sutera, A.** (2001), Fifty years of precipitation: Some spatially remote teleconnections. *Water Resources Management*, 15(4), 247-280.
- Bugera V., Konno H. & Uryasev S.** (2002). Credit cards scoring with quadratic utility function. *Journal of Multi-Criteria Decision Analysis*, 11, 197-211.
- Chen, K. Y. & Wang, C. H.** (2007). A hybrid ARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. *Expert Systems with Applications*, 32, 254-264.
- Chen, W., Ma, C., & Ma, L.** (2009). Mining the customer credit using hybrid support vector machine technique. *Expert Systems with Applications*, 36, 7611-7616.
- Chen, W., Xiang, G., Liu, Y. & Wang, K.** (2012). Credit risk Evaluation by hybrid data mining technique. *Systems Engineering Procedia*, 3, 194-200.
- Chen, Y. W. & Lin, C. J.** (2006). Combining SVMs with various feature selection strategies. *Fuzziness and Soft Computing*, 207, 315-324.
- Fraiwani, L., Lweesy, K., Khasawneh, N., Wenz, H. & Dickhaus, H.** (2012). Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier. *Computer methods and programs in biomedicine*, 108(1), 10-19.
- Freund, Y. & Schapire, R. E.** (1997). A decision-theoretic generalization of online learning and an application to boosting. *Journal of computer and system sciences*, 55(1), 119-139.
- Friedman, J.** (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29(5), 1189-1232.

- Gao, J., Wu, Q., Burges, C., Svore, K., Su, Y., Khan, N., Shah, S. & Zhou, H.** (2009). Model adaptation via model interpolation and boosting for web search ranking. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2 - Volume 2, EMNLP '09*, 505–513, Morristown, NJ, USA.
- Geurts, P., Ernst D. & Wehenkel, L.** (2006). Extremely randomized trees. *Machine Learning*, 36(1), 3–42.
- Gurny, P. & Gurny, M.** (2010). Comparison Of Credit Scoring Models On Probability Of Default Estimation For Us Banks. *Economic and Business Review*, 14(4), 299–320.
- Han, J. & Kamber, M.** (2001). *Data Mining Concepts and Techniques*. Academic Press. New York.
- Hastie, T., Tibshirani, R., Friedman, J.** (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Second Edition*. Springer Series in Statistics. Springer.
- Hastie, T.J., Rosset, S., Zhu, J. & Zou, H.** (2009). Multi-class adaBoost. *Statistics and its Interface* 2(3), 349–360
- Jha, J., & Ragha, L.** (2013). Intrusion detection system using support vector machine. *International Journal of Applied Information Systems (HAIS)-ISSN*, 2249-0868.
- Keller, J. M., Gray, M. R. & Givens, J. A.** (1985) A fuzzy K-nearest neighbor algorithm. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-15, 580–585.
- Kim, H.S. & Sohn, S. Y.** (2010). Support vector machines for default prediction of SMEs based on technology credit. *European Journal of Operational Research*, 201(3), 838–846.
- Kleinbaum, D. G.** (1994). *Logistic Regression A Self Learning Text*. 1st ed., New York : Springer.
- Koh, H. C. & Low, C. K.** (2004). Going concern prediction using data mining techniques. *Managerial Auditing Journal*, 19 (3), 462–476.
- Kröse, B. & Van Der Smagt, P.** (1996). *An Introduction to Neural Networks*, 8th ed., The University of Amsterdam.
- Le, Chap T.** (1998). *Applied Categorical Data Analysis*. New York: John Wiley & Sons Ltd.
- Liberati C., Camillo F. & Saporta G.** (2017) Advances in credit scoring: Combining performance and interpretation in kernel discriminant analysis. *Advances in Data Analysis and Classification* 11(1), 121–138.
- Madeira, S. C., Oliveira, A. L. & Conceição, C. S.** (2003). A data-mining approach to credit risk evaluation and behaviour scoring. *Progress in Artificial Intelligence. EPIA 2003. Lecture Notes in Computer Science*, vol 2902. Springer, Berlin, Heidelberg.
- Martinez-Arroyo, M. & Sucar, L. E.** (2006). Learning an optimal naive bayes classifier. *Pattern Recognition, ICPR 18th International Conference on IEEE*, 3, 1236–1239.

- Masud, M., Thuraisingham, B. & Khan, L.** (2011). *Data Mining Tools for Malware Detection*. CRC Press. Taylor & Francis, London.
- Mayr, A., Binder, H., Gefeller, O. & Schmid, M.** (2014). The evolution of boosting algorithms - from machine learning to statistical modelling. *Methods of Information in Medicine*. 53(6), 419-427.
- Mihalovic, M.** (2016). Performance comparison of multiple discriminant analysis and logit models in bankruptcy prediction. *Economics and Sociology*. 9(4), 101-118.
- Minsky, M. L. & Papert, S. S.** (1969). *Perceptrons*. Cambridge, MA:MIT Press.
- Ogryczak W. & Ruszczyński, A.** (1999). From stochastic dominance to mean-risk models: semideviations as risk measures. *European Journal of Operational Research*, 116, 33-50.
- Olson, D., Delen, D. & Meng, Y.** (2012) Comparative analysis of data mining methods for bankruptcy prediction. *Decision Support Systems*, 52(2), 464-473.
- Rosenblatt, Frank** (1961). *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Spartan Books, Washington DC.
- Rumelhart, D. E., McClelland, J. L. & the PDP research group** (1986). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. MIT Press.
- Salem, M.** (2012). Mining techniques in network security to enhance intrusion detection systems. *International Journal of Network Security & Its Applications* 4(6), 51–66.
- Schapire, R. E.** (1999). Theoretical views Of boosting and applications. *Proceedings Of The Tenth International Conference On Algorithmic Learning Theory*, 13 – 25, New York: Springer.
- Schapire, R. E.** (1999). A brief introduction to boosting. *Proceedings Of The Sixteenth International Joint Conference On Artificial Intelligence*, 2, 1401-1406.
- Sousa, Marcos de Moraes & Figueiredo, Reginaldo Santana** (2014). Credit analysis using data mining: application in the case of a credit union. *JISTEM- Journal of Information Systems and Technology Management* 11(2), 379-396.
- Sustersic, M., Mramor, D., & Zupan, J.** (2009). Consumer credit scoring models with limited data. *Expert Systems with Applications*, 36, 4736-4744.
- Thomas, L .C., Edelman, D. B. & Crook, J. N.** (2002). *Credit Scoring and Its Applications*. Society for Industrial and Applied Mathematics, Philadelphia.
- Vicente-Serrano, S. M., Gonzalez-Hidalgo, J. S. & Luis, M. Ravento's, J.** (2004). Drought patterns in the Mediterranean area: The Valencia region (eastern Spain). *Climate Research* 26(1), 5–15.
- Wang, R.** (2012). AdaBoost for feature selection, classification and its relation with SVM, a review. *Physics Procedia*, 25, 800-807.

- Wang, W.** (2012). How the small and medium-sized enterprises' owners' credit features affect the enterprises' credit default behaviour? *E3 Journal of Business Management and Economics*, 3(2), 090–095.
- Wei, W., Knapskog, S. J. & Gombault, S.** (2009). Attribute normalization in network intrusion detection. *10th International Symposium on Pervasive Systems, Algorithms, and Networks IEEE*, 448-453.
- Wu, Q., Burges, C.J.C., Svore, K.M. and Gao, J.** (2008). Ranking, boosting, and model adaptation. *Technical Report MSR-TR-2008-109*, Microsoft Research.
- Ye, J., Chow, J.-H. & Chen, J., and Zheng, Z.** (2009). Stochastic gradient boosted distributed decision trees. *Proceeding of the 18th ACM conference on Information and knowledge management, CIKM '09*, 2061–2064, New York, NY, USA.
- Zhai, L., Sun, J., Sang, H., Yang, G. & Jia, Y.** (2012). *Large Area Land Cover Classification with Landsat ETM+ Images Based on Decision Tree*. Proceedings of the XXII ISPRS Congress held 25 August - 1 September. Melbourne: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences.
- Zhang, C. & Ma Y.** (2012). *Ensemble Machine Learning*, 1-37, New York: Springer.
- Zhou, Z. H.** (2012). *Ensemble Methods: Foundations and Algorithms*. CRC Machine Learning & Pattern Recognition Series. Boca Raton, FL, United States of America.
- Zhu, Y., Xie, C., Wang, G.J. & Yan, X.G.** (2016). Predicting China's SME credit risk in supply chain finance based on machine learning methods, *Entropy*, 18(5), 195.
- Url-1** <<https://www.turkcell.com.tr/tr/hakkimizda/genel-bakis>>, erişim tarihi 22.03.2019.

EKLER

EK A : Ağaç modellerinin öznitelik seçimi sonuçlarının ayrı ayrı değerleri

EK B : Özniteliklerin korelasyon matrisinin sayısal verileri



EK A

Çizelge A.1 : Ağaç modellerinin öznitelik seçme algoritması ile elde edilen sonuçları.

Öznitelikler	Karar Ağaçları	Rastgele Orman	Extra Gradyen Boosting	Gradyen Boosting	Ada Boost	Extra Trees
ofis_sehir	0.03403917	2.76851419e-02	0.01124451	5.20665883e-03	0.04	3.81346929e-02
ilk_tutar	0.02987486	5.71354556e-02	0.00087701	4.22935085e-04	0.02	5.69566736e-02
avukat_ucreti	0.02076849	4.17180958e-02	0.01467659	4.26372113e-03	0.	4.30390167e-02
masraf	0.15764108	1.07140126e-01	0.11191136	3.17472973e-01	0.16	7.64135763e-02
yasal_faiz	0.	3.64765742e-06	0.	0.	0.	1.35915736e-05
ilk_tutar_bolu_toplam	0.03259	6.40948705e-02	0.00357182	6.28677769e-04	0.	5.88426251e-02
avukat_bolu_toplam	0.04730836	5.95287143e-02	0.01478053	3.24420346e-03	0.02	6.21976115e-02
masraf_bolu_toplam	0.04514183	6.37806230e-02	0.0282724	2.54232515e-02	0.06	6.33596452e-02
yasal_faiz_bolu_toplam	0.	5.98014531e-06	0.	0.	0.	1.17246717e-05
abone_sehri	0.07709848	5.20576883e-02	0.01402355	3.00395154e-03	0.02	4.90397689e-02
tum_ay_isgunu_sayisi	0.00493894	5.62598095e-03	0.09498293	0.	0.	4.15495888e-03
transfer_ay_sonu_isgunu_sayisi	0.0103762	1.08124200e-02	0.00182121	0.	0.02	6.96274588e-03
transfer_ay_sonu_farki	0.00962643	1.14274236e-02	0.	3.53486911e-04	0.02	6.83172226e-03
mevsim	0.00181382	4.27179781e-03	0.08259013	4.68941247e-03	0.	5.15791443e-03
bayram_var	0.00124017	1.50600258e-03	0.	0.	0.	7.23838463e-04
transfer_maas_yatma_araligi	0.00308506	2.77542595e-03	0.	0.	0.	2.39106972e-03
transfer_gunu_araligi	0.00061631	1.09450045e-03	0.	0.	0.	1.70348683e-03
transfer_ay	0.0049525	1.93329382e-02	0.16145203	9.14072477e-03	0.	1.17492891e-02
transfer_haftanin_hangi_gun	0.01246015	1.00486073e-02	0.00121113	5.97043301e-04	0.02	7.04639760e-03
transfer_yilin_hangi_haftasi	0.017524	2.20173926e-02	0.07918213	6.29151024e-02	0.	5.52491661e-03
transfer_yilin_hangi_cerregi	0.00168947	1.01315148e-02	0.	2.64011985e-02	0.	1.71550080e-02
ayni_sehir	0.00770907	7.92293888e-03	0.00695167	4.70380913e-04	0.	1.05572580e-02
ayni_bolge	0.00844159	9.39789368e-03	0.00382614	1.49768049e-04	0.	1.37882057e-02
durum_tipi	0.00134463	1.83296734e-03	0.00401261	0.	0.	2.03209963e-03
medeni_durum	0.00718102	9.82299204e-03	0.03804494	2.64361250e-03	0.02	1.11581141e-02
medeni_durum2	0.01582816	1.45445889e-02	0.04376427	2.83800914e-02	0.04	1.65726522e-02
yeni_yil_transfer_fark_kategori	0.04501278	1.35095672e-02	0.0303639	1.85413046e-01	0.08	1.25247549e-02
yas_kategori	0.01731262	1.55240096e-02	0.01745951	7.56432007e-03	0.04	2.67452855e-02
yas_kategori2	0.0103352	1.00583601e-02	0.00908401	1.18493396e-03	0.	1.67673838e-02
musteri_bolge_kategori	0.03220362	2.83201069e-02	0.01993359	6.33987475e-03	0.04	3.97293415e-02
toplam_borc_tutari	0.0343181	6.18023687e-02	0.01385112	1.05303216e-03	0.04	5.91218003e-02
toplam_borc_dolar	0.07119113	6.28066677e-02	0.01030322	2.30869487e-04	0.	5.83319250e-02
dolar_kuru	0.01078988	6.25192081e-03	0.	3.35504221e-02	0.04	5.72241702e-03
gender	0.01120287	9.49702239e-03	0.00790749	1.02969497e-03	0.	1.67581290e-02
iletisim_kuruldu	0.01298814	1.16687074e-02	0.01172007	1.57788104e-03	0.	1.68838086e-02
yabanci_uyruk	0.00028349	2.15024521e-04	0.00691496	1.04560352e-03	0.02	3.53867397e-04
yasal_takip_sayi	0.02138305	1.95563446e-02	0.02149257	6.37643439e-03	0.02	3.55811246e-02
taksit_yapildi	0.00840499	7.38157336e-03	0.02374717	5.91961288e-03	0.02	9.00891398e-03
fatura_sayisi	0.04548963	4.31345006e-02	0.04841701	8.44708567e-02	0.1	5.64825865e-02
abonelik_sure	0.12579474	9.45580975e-02	0.06160836	1.68836224e-01	0.16	7.44700580e-02

Çizelge A.2 : Ağaç modellerinin normalize edilmiş ve ortalamalarının alınmış hali.

Öznitelikler	Karar Ağaçları	Rastgele Orman	Extra Gradyen Boosting	Gradyen Boosting	Ad a Bo ost	Extra Trees	Ortalama Değerler
ofis_sehir	0.2159283	2.58375995e-01	0.06964614	1.64003215e-02	0.25	4.98979637e-01	2.18221732e-01
ilk_tutar	0.18951192	5.33261953e-01	0.00543202	1.33219241e-03	0.13	7.45334671e-01	2.66645460e-01
avukat_ucreti	0.13174541	3.89358029e-01	0.09090372	1.34301862e-02	0.	5.63170802e-01	1.98101358e-01
masraf	1.	1.	0.6931555	1.	1.	1.	9.48859255e-01
yasal_faiz	0.	0.	0.	0.	0.	2.44352976e-05	4.07254960e-06
ilk_tutar_bolu_topla m	0.20673544	5.98220360e-01	0.02212311	1.98025603e-03	0. 0.12	7.70019301e-01	2.66513077e-01
avukat_bolu_toplam masraf_bolu_topla m	0.30010176	5.55600365e-01	0.09154749	1.02188335e-02	5 0.37	8.13931671e-01	3.16066687e-01
yasal_faiz_bolu_top lam	0.	2.17711830e-05	0.	0.	0.	0.	3.62853050e-06
abone_sehri	0.48907606	4.85866638e-01	0.08685892	9.46207014e-03	0.13	6.41712776e-01	3.06329410e-01
tum_ay_isgunu_say isi	0.03133029	5.24782348e-02	0.58830434	0.	0.	5.42294999e-02	1.21057062e-01
transfer_ay_sonu_is gunu_sayisi	0.06582166	1.00887881e-01	0.0112802	0.	0.13	9.09797480e-02	6.56615815e-02
transfer_ay_sonu_fa rki	0.06106548	1.06628257e-01	0.	1.11343938e-03	0.13	8.92648207e-02	6.38453329e-02
mevsim	0.011506	3.98384397e-02	0.5115459	1.47710604e-02	0.	6.73568722e-02	1.07503045e-01
bayram_var	0.00786704	1.40228141e-02	0.	0.	0.	9.32063525e-03	5.20174903e-03
transfer_maas_yatm a_araligi	0.01957017	2.58714709e-02	0.	0.	0.	3.11425051e-02	1.27640239e-02
transfer_gunu_arali gi	0.00390956	1.01818989e-02	0.	0.	0.	2.21429472e-02	6.03906699e-03
transfer_ay	0.03141629	1.80417453e-01	1.	2.87921352e-02	0.	1.53629319e-01	2.32375866e-01
transfer_haftanin_h angi_gun	0.07904124	9.37585380e-02	0.00750149	1.88061143e-03	0.13	9.20746392e-02	6.65427536e-02
transfer_yilin_hangi _haftasi	0.11116391	2.05473852e-01	0.49043754	1.98174672e-01	0.	7.21604492e-02	1.79568403e-01
transfer_yilin_hangi _ceyregi	0.01071717	9.45323872e-02	0.	8.31604601e-02	0.	2.24383087e-01	6.87988508e-02
ayni_sehir	0.04890265	7.39177852e-02	0.04305716	1.48164081e-03	0.	1.38027195e-01	5.08977390e-02
ayni_bolge	0.05354946	8.76848496e-02	0.02369829	4.71750547e-04	0.	1.80316062e-01	5.76200680e-02
durum_tipi	0.00852972	1.70746669e-02	0.02485327	0.	0.	2.64440575e-02	1.28169517e-02
medeni_durum	0.04555296	9.16526705e-02	0.23564236	8.32704741e-03	0.13	1.45891614e-01	1.08677775e-01
medeni_durum2	0.10040634	1.35723531e-01	0.27106673	8.93937241e-02	0.25	2.16760814e-01	1.77225189e-01
yeni_yil_transfer_fa rk_kategori	0.28553968	1.26062753e-01	0.18806766	5.84027813e-01	0.5	1.63779149e-01	3.07912843e-01
yas_kategori	0.10982304	1.44865335e-01	0.10814057	2.38266583e-02	0.25	3.49907237e-01	1.64427139e-01
yas_kategori2	0.06556162	9.38495700e-02	0.05626443	3.73239317e-03	0.	2.19309595e-01	7.31196003e-02
musteri_bolge_kate gori	0.20428444	2.64302688e-01	0.12346449	1.99698093e-02	0.25	5.19851495e-01	2.30312155e-01
toplam_borc_tutari	0.21769769	5.76822402e-01	0.08579096	3.31691908e-03	0.25	7.73673339e-01	3.17883552e-01
toplam_borc_dolar	0.45160266	5.86196418e-01	0.063816	7.27209893e-04	0.	7.63334907e-01	3.10946200e-01
dolar_kuru	0.06844585	5.83206881e-02	0.	1.05679617e-01	0.25	7.47454705e-02	9.28652708e-02
gender	0.07106566	8.86101060e-02	0.04897736	3.24340985e-03	0.	2.19188461e-01	7.18475003e-02
iletisim_kuruldu	0.08239055	1.08880373e-01	0.07259163	4.97012714e-03	0. 0.12	2.20833443e-01	8.16110199e-02
yabanci_uyruk	0.00179834	1.97296818e-03	0.04282982	3.29351978e-03	5	4.47819939e-03	2.98954737e-02
yasal_takip_sayi	0.13564388	1.82502703e-01	0.13312045	2.00849676e-02	0.13	4.65556778e-01	1.76984796e-01
taksit_yapildi	0.05331727	6.88647395e-02	0.14708501	1.86460372e-02	0.13	1.17761404e-01	8.84457431e-02
fatura_sayisi	0.2885646	4.02578594e-01	0.2998848	2.66072591e-01	0.63	7.39129492e-01	4.36871679e-01
abonelik_sure	0.79798197	8.82560739e-01	0.38158923	5.31812904e-01	1.	9.74561895e-01	7.61417790e-01

EK B

	ofis_sehir	ilk_tutar	avukat_ureti	masraf	yasal_faiz	ilk_tutar_bolu_toplam	avukat_bolu_toplam	masraf_bolu_toplam	yasal_faiz_bolu_toplam	abone_sehri	tum_ay_isgunu_sayisi	transfer_ay_sonu_isgunu_sayisi	transfer_ay_sonu_farki	mevsim	bayram_var	transfer_maas_yatma_araligi	transfer_gunu_araligi	transfer_ay	transfer_haftan_hangi_gun	transfer_yilin_hangi_haftasi	ayni_sehir	ayni_bolge	durum_tipi	medeni_durum	medeni_durum2	yeni_yil_transfer_fark_kategori	yas_kategori	yas_kategori2	abone_bolge_kategori	toplam_borc_tutari	toplam_borc_dolar	dolar_kuru	cinsiyet	iletisim_kuruldugu	yabanci_uyruk	yasal_takip_sayisi	taksit_yapildi	fatura_sayisi	abonelik_sure	durum_info											
ofis_sehir	1,00																																																		
ilk_tutar	0,00	1,00																																																	
avukat_ureti	-0,06	0,76	1,00																																																
masraf	-0,07	0,65	0,76	1,00																																															
yasal_faiz	0,00	0,11	0,09	0,06	1,00																																														
ilk_tutar_bolu_toplam	0,10	0,22	-0,16	-0,24	0,00	1,00																																													
avukat_bolu_toplam	-0,10	-0,20	0,25	0,17	-0,01	-0,95	1,00																																												
masraf_bolu_toplam	-0,09	-0,20	-0,08	0,30	-0,01	-0,81	0,60	1,00																																											
yasal_faiz_bolu_toplam	0,00	0,08	0,06	0,05	0,75	0,00	-0,01	-0,01	1,00																																										
abone_sehri	0,10	-0,01	-0,01	-0,01	0,00	-0,01	0,01	0,00	0,00	1,00																																									
tum_ay_isgunu_sayisi	-0,02	-0,01	-0,03	0,00	-0,02	0,01	0,04	0,00	0,00	1,00																																									
transfer_ay_sonu_isgunu_sayisi	0,00	0,03	-0,01	-0,05	0,00	0,11	-0,10	-0,12	-0,01	0,00	0,09	1,00																																							
transfer_ay_sonu_farki	0,00	0,03	-0,01	-0,04	0,00	0,11	-0,09	-0,11	-0,01	-0,01	-0,21	0,93	1,00																																						
mevsim	-0,06	0,00	0,15	0,10	-0,01	-0,28	0,30	0,16	-0,01	-0,01	-0,11	0,41	0,48	1,00																																					
bayram_var	-0,01	0,00	0,06	0,05	0,00	-0,11	0,11	0,07	0,00	0,00	-0,66	-0,15	0,11	0,23	1,00																																				
transfer_maas_yatma_araligi	-0,02	0,01	0,05	0,04	0,00	-0,07	0,08	0,04	0,00	0,00	0,20	0,28	0,27	0,10	-0,15	1,00																																			
transfer_gunu_araligi	-0,01	-0,03	-0,02	0,03	0,00	-0,10	0,08	0,12	0,00	0,00	0,38	-0,85	-0,91	-0,37	-0,12	-0,11	1,00																																		
transfer_ay	0,05	0,01	-0,18	-0,22	-0,01	0,38	-0,36	-0,30	-0,01	-0,01	-0,13	0,47	0,54	0,60	0,12	0,03	-0,45	1,00																																	
transfer_haftanin_hangi_gun	0,02	0,02	-0,03	-0,04	0,00	0,13	-0,13	-0,10	0,00	0,00	0,26	0,42	0,36	0,01	-0,43	0,21	-0,41	0,15	1,00																																
transfer_yilin_hangi_haftasi	0,05	0,01	-0,19	-0,22	-0,01	0,38	-0,36	-0,31	-0,01	-0,01	-0,12	0,43	0,50	0,60	0,12	0,01	-0,40	1,00	0,12	1,00																															
transfer_yilin_hangi_haftasi	0,04	0,01	-0,15	-0,19	-0,01	0,32	-0,30	-0,26	-0,01	-0,01	-0,07	0,42	0,50	0,63	0,06	0,09	-0,37	0,97	0,17	0,97	1,00																														
ayni_sehir	-0,11	-0,06	-0,08	-0,09	0,00	0,03	-0,03	-0,03	0,00	0,16	0,04	-0,08	-0,09	-0,09	-0,04	-0,02	0,09	-0,02	-0,01	-0,01	-0,02	1,00																													
ayni_bolge	-0,04	-0,08	-0,09	-0,09	-0,01	0,03	-0,02	-0,03	-0,01	0,08	0,04	-0,11	-0,13	-0,11	-0,05	-0,04	0,11	-0,04	-0,01	-0,03	-0,04	0,67	1,00																												
durum_tipi	0,03	0,29	0,19	0,17	0,02	0,12	-0,12	-0,09	0,03	-0,04	0,01	0,02	0,02	0,01	0,00	0,00	-0,02	0,01	0,03	0,01	0,01	-0,18	-0,24	1,00																											
medeni_durum	0,03	0,15	0,11	0,07	0,01	0,10	-0,08	-0,11	0,01	-0,05	-0,02	0,05	0,07	0,08	0,03	0,01	-0,05	0,10	0,01	0,10	0,10	-0,19	-0,23	0,46	1,00																										
medeni_durum2	0,03	0,17	0,12	0,08	0,01	0,11	-0,10	-0,11	0,01	-0,06	-0,02	0,06	0,07	0,09	0,03	0,02	-0,06	0,11	0,01	0,10	0,11	-0,24	-0,29	0,54	0,93	1,00																									
yeni_yil_transfer_fark_kategori	-0,05	-0,01	0,19	0,22	0,01	-0,38	0,36	0,31	0,01	0,01	0,16	-0,39	-0,47	-0,59	-0,15	0,01	0,40	-0,99	-0,12	-1,00	-0,96	0,01	0,03	-0,01	-0,10	-0,10	1,00																								
yas_kategori	0,00	0,00	-0,04	-0,02	0,00	0,02	-0,03	0,00	0,00	-0,02	-0,01	0,02	0,02	0,02	0,00	0,01	-0,02	0,02	0,01	0,02	0,02	-0,04	-0,05	0,02	-0,08	0,02	-0,02	1,00																							
yas_kategori2	0,00	0,01	-0,01	-0,01	0,00	0,04	-0,04	-0,03	0,00	-0,01	-0,02	0,06	0,06	0,04	0,01	0,01	-0,06	0,06	0,01	0,05	0,05	-0,02	-0,02	0,04	-0,08	-0,01	-0,05	0,56	1,00																						
abone_bolge_kategori	0,00	-0,09	-0,05	-0,05	-0,01	-0,06	0,06	0,04	-0,01	0,25	0,01	-0,03	-0,04	-0,04	-0,01	-0,02	0,03	-0,05	-0,01	-0,05	-0,06	0,34	0,56	-0,32	-0,33	-0,41	0,05	-0,06	-0,02	1,00																					
toplam_borc_tutari	-0,01	0,99	0,82	0,70	0,14	0,16	-0,13	-0,17	0,11	-0,01	-0,01	0,02	0,02	0,03	0,01	0,01	-0,03	-0,02	0,01	-0,03	-0,02	-0,06	-0,09	0,28	0,15	0,16	0,03	-0,01	0,01	-0,09	1,00																				
toplam_borc_dolar	-0,01	0,97	0,82	0,71	0,18	0,15	-0,13	-0,16	0,14	-0,01	0,01	-0,01	-0,02	-0,05	-0,02	0,01	0,00	-0,11	0,01	-0,11	-0,11	-0,06	-0,08	0,28	0,14	0,15	0,11	-0,01	0,00	-0,08	0,98	1,00																			
dolar_kuru	0,01	0,00	-0,07	-0,11	-0,01	0,14	-0,12	-0,13	-0,01	-0,01	-0,30	0,27	0,43	0,75	0,27	0,03	-0,34	0,87	0,05	0,87	0,85	-0,04	-0,06	0,01	0,09	0,10	-0,88	0,02	0,05	-0,05	-0,01	-0,11	1,00																		
cinsiyet	0,01	0,05	0,04	0,04	0,00	0,03	-0,04	-0,01	0,00	0,00	-0,01	0,01	0,01	0,02	0,01	0,01	-0,01	0,02	0,00	0,01	0,02	-0,04	-0,04	0,12	0,18	0,14	-0,02	0,00	-0,01	-0,05	0,05	0,04	0,02	1,00																	
iletisim_kuruldugu	-0,01	0,01	0,02	0,03	0,00	-0,02	0,02	0,02	0,00	0,00	0,00	0,01	0,01	-0,02	0,00	0,01	-0,01	-0,04	0,02	-0,04	0,00	0,00	0,03	-0,07	-0,06	0,04	0,00	-0,04	0,02	0,01	0,01	-0,04	0,04																		
yabanci_uyruk	0,01	0,00	0,00	-0,01	0,00	0,00	0,00	0,00	0,00	-0,01	0,01	0,00	0,00	0,00	0,00	-0,01	0,00	0,00	0,00	0,00	0,00	-0,03	-0,04	-0,01	0,05	0,06	0,00	0,00	0,00	-0,05	0,00	0,00	0,00	-0,01	-0,03	1,00															
yasal_takip_sayisi	0,00	0,02	0,04	0,05	0,00	-0,02	0,01	0,02	0,00	-0,01	0,00	0,01	0,01	0,00	-0,01	0,03	-0,01	-0,02	0,01	-0,03	-0,02	-0,06	-0,06	0,02	0,09	0,14	0,03	0,04	-0,01	-0,10	0,02	0,03	-0,01	0,12	0,16	-0,02	1,00														
taksit_yapildi	0,01	0,07	0,09	0,05	0,00	-0,06	-0,09	0,00	0,00	0,00	0,00	0,02	0,02	0,01	0,00	-0,01	-0,02	0,01	0,02	0,01	0,01	-0,01	0,00	0,04	0,03	-0,01	-0,04	-0,02	0,00	0,08	0,08	0,01	0,01	0,13	0,00	0,11	1,00														
fatura_sayisi	0,00	0,59	0,43	0,38	0,07	0,13	-0,12	-0,11	0,06	-0,01	0,00	0,01	0,																																						

ÖZGEÇMİŞ

Ad Soyad : Elif Ekiz

Doğum Tarihi ve Yeri : 1991 / İstanbul



ÖĞRENİM DURUMU:

- **Lisans** : 2014, İstanbul Teknik Üniversitesi, Bilgisayar ve Bilişim Fakültesi, Bilgisayar Mühendisliği

MESLEKİ DENEYİM:

Mesleki kariyerine bir danışmanlık şirketinde yazılımcı olarak başladıktan sonra, burada edindiği farklı tecrübeler ve farklı şirket deneyimlerine ilaveten, Türkiye'deki bir beyaz eşya üreticisi şirkette mobil uygulama ve Android TV ürünlerinin geliştirilmesine katkıda bulunmuştur. Şimdi ise Borsa İstanbul A.Ş.'de index hesaplama ekibinde yazılım geliştirmeye ve piyasaya destek vermeye devam etmektedir.