ZONGULDAK BÜLENT ECEVİT ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ İLE ÇEVRESEL TUTUMLARIN SINIFLANDIRILMASI: ÖZNİTELİK SEÇİMİ TEMELLİ MÜHENDİSLİK YAKLAŞIMI

ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

YÜKSEK LİSANS TEZİ

FUAT ALKAN

HAZİRAN 2025

ZONGULDAK BÜLENT ECEVİT ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ İLE ÇEVRESEL TUTUMLARIN SINIFLANDIRILMASI: ÖZNİTELİK SEÇİMİ TEMELLİ MÜHENDİSLİK YAKLAŞIMI

ELEKTRİK ELEKTRONİK MÜHENDİSLİĞİ ANABİLİM DALI

YÜKSEK LİSANS TEZİ

Fuat ALKAN

DANIŞMAN : Doç. Dr. Rukiye UZUN ARSLAN

İKİNCİ DANIŞMAN : Dr. Öğr. Üyesi İrem ŞENYER YAPICI

ZONGULDAK

Haziran 2025

KABUL:

Fuat ALKAN tarafından hazırlanan "Makine Öğrenmesi ile Çevresel Tutumların Sınıflandırılması: Öznitelik Seçimi Temelli Mühendislik Yaklaşımı" başlıklı bu çalışma jürimiz tarafından değerlendirilerek Zonguldak Bülent Ecevit Üniversitesi, Fen Bilimleri Enstitüsü, Elektrik Elektronik Mühendisliği Anabilim Dalında Yüksek Lisans Tezi olarak oybirliğiyle kabul edilmiştir. 03/06/2025

Danışman: Doç. Dr. Rukiye UZUN ARSLAN .....
Zonguldak Bülent Ecevit Üniversitesi, Mühendislik Fakültesi, Elektrik Elektronik Mühendisliği Bölümü

Üye: Doç. Dr. Tuğba Özge ONUR .....
Zonguldak Bülent Ecevit Üniversitesi, Mühendislik Fakültesi, Elektrik Elektronik Mühendisliği Bölümü

Üye: Dr. Öğr. Üyesi Didem ALTUN .....
Sivas Cumhuriyet Üniversitesi, Sivas Teknik Bilimler Meslek Yüksekokulu, Elektrik ve Enerji Bölümü

ONAY:

Yukarıdaki imzaların, adı geçen öğretim üyelerine ait olduğunu onaylarım. ..../..../20....

Prof. Dr. Fikret GÖLGELEYEN
Fen Bilimleri Enstitüsü Müdürü

“Bu tezdeki tüm bilgilerin akademik kurallara ve etik ilkelere uygun olarak elde edildiğini ve sunulduğunu; ayrıca bu kuralların ve ilkelerin gerektirdiği şekilde, bu çalışmadan kaynaklanmayan bütün atıfları yaptığımı beyan ederim.”

Fuat ALKAN

ÖZET

Yüksek Lisans Tezi

MAKİNE ÖĞRENMESİ İLE ÇEVRESEL TUTUMLARIN SINIFLANDIRILMASI: ÖZNİTELİK SEÇİMİ TEMELLİ MÜHENDİSLİK YAKLAŞIMI

Fuat ALKAN

Zonguldak Bülent Ecevit Üniversitesi
Fen Bilimleri Enstitüsü
Elektrik Elektronik Mühendisliği Anabilim Dalı

Tez Danışmanı: Doç. Dr. Rukiye UZUN ARSLAN
İkinci Danışman: Dr. Öğr. Üyesi İrem ŞENYER YAPICI
Haziran 2025, 97 sayfa

Günümüzde veriye dayalı karar alma süreçlerinde makine öğrenmesi (MÖ) modellerinin başarımı, büyük ölçüde kullanılan verinin kalitesi ile anlamlı özniteliklerin seçimine bağlıdır. Özellikle yüksek boyutlu veri kümelerinde, gereksiz veya gürültülü değişkenlerin varlığı öğrenme sürecini yavaşlatmakta, aşırı öğrenmeye yol açmakta ve model çıktılarının yorumlanabilirliğini azaltmaktadır. Bu nedenle, öznitelik seçimi yalnızca hesaplama karmaşıklığını azaltmakla kalmayıp, aynı zamanda doğruluk ve genellenebilirlik gibi temel performans ölçütlerini de iyileştiren kritik bir ön işleme adımı olarak değerlendirilmektedir. Literatürde yapılan çalışmalar, uygun öznitelik alt kümelerinin belirlenmesinin sınıflandırma başarımını artırdığını ve eğitim süresini kısalttığını ortaya koymaktadır.

Bu doğrultuda yürütülen tez çalışmasında, farklı özellik seçimi yöntemlerinin çeşitli MÖ algoritmalarının sınıflandırma performansı üzerindeki etkileri sistematik olarak incelenmiştir. Örnek uygulama olarak, çok değişkenli yapıya sahip bireylerin çevresel tutum düzeylerini

içeren açık erişimli bir anket veri seti kullanılmıştır. Bu veri seti üzerinde en anlamlı özellikleri belirlemek amacıyla Varyans Analizi, Ki-kare testi, Rasgele Orman (RO), LASSO (En Küçük Mutlak Daralma ve Seçim Operatörü), Temel Bileşenler Analizi (TBA), Oylama, Özyinelemeli Özellik Eliminasyonu, Karşılıklı Bilgi, varyans eşiği gibi on farklı yöntem uygulanmıştır. Seçilen özellik kümeleriyle Naive Bayes, Gradyan Artırma, k-En Yakın Komşu, Destek Vektör Makineleri, Lojistik Regresyon, Çok Katmanlı Algılayıcılar (ÇKA), Torbalama ve Doğrusal Diskriminant Analizi gibi on farklı sınıflandırıcı model eğitilmiştir. Modelleme sürecinde veri kümesi %80 eğitim ve %20 test olacak şekilde ayrılmış; performans değerlendirmesi ise 5 katlı çapraz doğrulama yöntemi ile gerçekleştirilmiştir. Elde edilen bulgular, özellikle TBA ve RO tabanlı özellik seçimi yöntemlerinin, ÇKA algoritmasıyla birlikte kullanıldığında %98,7 doğruluk oranına ulaşarak en yüksek performansı sağladığını ortaya koymuştur. Ayrıca LASSO ve Voting tabanlı yaklaşımlar da dikkat çekici başarımlar sergilemiştir. Genel olarak, özellik seçimi ile sınıflandırma algoritmalarının uygun biçimde eşleştirilmesinin model başarımı üzerinde istatistiksel olarak anlamlı etkiler yarattığı gözlemlenmiştir.

Sonuç olarak, bu tez çalışması, özellik seçimi tekniklerinin MÖ tabanlı sınıflandırma modelleri üzerindeki etkilerini çevresel tutum verileri özelinde kapsamlı biçimde incelemektedir. Elde edilen bulgular, yüksek boyutlu veri kümelerine yönelik sınıflandırma problemlerinde öznitelik seçiminin model başarımı açısından kritik bir rol oynadığını ortaya koymakta; bu doğrultuda çalışma, literatüre metodolojik ve uygulamalı düzeyde anlamlı katkılar sunmaktadır. Sayısal analizlerle desteklenen sonuçlar, model optimizasyon sürecinde öznitelik seçiminin önemini vurgulamakta ve farklı veri türleri ile problem alanlarında uygulanabilecek esnek bir çerçeve önermektedir.

Anahtar Kelimeler: Özellik seçimi, makine öğrenmesi, sınıflandırma, karar destek sistemi.

Bilim Kodu: 608.03.00

ABSTRACT

M. Sc. Thesis

CLASSIFICATION OF ENVIRONMENTAL ATTITUDES WITH MACHINE LEARNING ALGORITHMS: AN ENGINEERING APPROACH BASED ON FEATURE SELECTION

Fuat ALKAN

Zonguldak Bülent Ecevit University
Graduate School of Natural and Applied Sciences
Department of Electrical and Electronics Engineering

Thesis Advisor: Assoc. Prof. Dr. Rukiye UZUN ARSLAN
Co-Advisor: Assist. Prof. Dr. İrem ŞENYER YAPICI
June 2025, 97 pages

Nowadays, the performance of machine learning (ML) models in data-driven decision-making processes largely depends on the quality of the used data and the selection of meaningful features. Particularly in high-dimensional datasets, the presence of redundant or noisy feautres impedes the learning process, but also induces overfitting and diminishes the interpretability of model outputs. Therefore, feature selection is regarded as a critical preprocessing step that not only reduces computational complexity but also improves fundamental performance metrics, such as accuracy and generalizability. Studies in the literature show that identifying optimal feature subsets improves classification performance and reduces training duration.

In this thesis, the effects of different feature selection methods on the classification performance of various ML algorithms are systematically examined. As a case study, an open-access survey dataset containing individuals' environmental attitude levels with a multivariate structure has

been used. To identify the most significant features in this dataset, ten different methods have been applied, including Analysis of Variance, Chi-square test, Random Forest (RF), LASSO (Least Absolute Shrinkage and Selection Operator), Principal Component Analysis (PCA), Voting, Recursive Feature Elimination, Mutual Information, and Variance Threshold. Using the selected feature sets, ten different classification models have been trained, including Naive Bayes, Gradient Boosting, k-Nearest Neighbors, Support Vector Machines, Logistic Regression, Multilayer Perceptrons (MLP), Bagging, and Linear Discriminant Analysis. During the modelling process, the dataset has been divided into 80% training and 20% testing, and the performance evaluation has been conducted using 5-fold cross-validation method. The findings reveal that feature selection methods based on PCA and RF, when combined with MLP algorithm, has been achieved the highest performance with an accuracy rate of 98.7%. Additionally, LASSO- and Voting-based approaches have also demonstrated remarkable performance. Overall, it has been observed that the appropriate matching of feature selection methods with classification algorithms has statistically significant effects on model performance.

In conclusion, this thesis extensively investigates the effects of feature selection techniques on ML based classification models for environmental attitude data. The findings reveal that feature selection plays a critical role in terms of model performance in classification problems for high-dimensional datasets; accordingly, the study makes significant contributions to the literature at methodological and practical levels. The results, supported by numerical analyses, emphasise the importance of feature selection in the model optimisation process and suggest a flexible framework that can be applied to different data types and problem domains.

Keywords: Feature selection, machine learning, classification, decision support system.

Science Code: 608.03.00

TEŞEKKÜR

Yüksek lisans tezimi hazırlarken bana her aşamada rehberlik eden, bilgi ve deneyimiyle yolumu aydınlatan değerli tez danışmanım Doç. Dr. Rukiye UZUN ARSLAN'a en içten teşekkürlerimi sunuyorum. Tez sürecinin başından itibaren gösterdiği sabır, yapıcı yönlendirmeleri ve her zaman yanımda hissettirdiği desteği benim için çok kıymetlidir.

Ayrıca, çalışmalarım boyunca bilgi ve birikimiyle katkı sağlayan, destek ve motivasyonunu esirgemeyen ikinci danışmanım Dr. Öğr. Üyesi İrem ŞENYER YAPICI'ya da teşekkür ederim. Varlığı ve değerli katkıları, bu sürecin daha verimli ve anlamlı geçmesini sağlamıştır.

Bu sürecin her anında yanımda olan, sabrını ve sevgisini hiç eksik etmeyen aileme de sonsuz teşekkür ederim. Onların koşulsuz desteği, inancı ve manevi varlığı olmasaydı bu çalışmayı tamamlamak mümkün olmazdı.

Emek veren, destek olan ve yanımda yürüyen herkese gönülden teşekkür ederim.

İÇİNDEKİLER

İÇİNDEKİLER (devam ediyor)

Sayfa

İÇİNDEKİLER (devam ediyor)

Sayfa

İÇİNDEKİLER (devam ediyor)

Sayfa

ŞEKİLLER DİZİNİ

No Sayfa

ŞEKİLLER DİZİNİ (devam ediyor)

No Sayfa

ÇİZELGELER DİZİNİ

No Sayfa

SİMGELER VE KISALTMALAR DİZİNİ

SİMGELER

z	: Standartlaştırılmış değeri
x	: Orijinal gözlem değeri
μ	: İlgili değişkenin ortalaması
σ	: İlgili değişkenin standart sapması
Oi	: Gözlenen frekansı
Ei	: Beklenen frekansı
χ^2	: Ki-Kare test istatistiği
f	: Özelliğin önem skorunu
$T(f)$	: Özelliğin kullanıldığı tüm düğümler kümesi
Nt	: Düğüm t'ye ulaşan örnek sayısı
N	: Toplam örnek sayısı
$\Delta i(t)$	: Düğümde sağlanan saflık artışı
ni	: i. gruptaki gözlem sayısı
YMi	: i. grubun ortalaması
YM	: Veri setinin genel ortalaması
N	: Toplam gözlem sayısı
k	: Grup sayısı
y_i	: Bağımlı değişkeni
x_{ij}	: Bağımsız değişkeni
α	: Sabit terimi
β_j	: Regresyon katsayıları
p	: Bağımsız değişken sayısı

SİMGELER VE KISALTMALAR DİZİNİ (devam ediyor)

t : Cezalandırma parametresi

$p(x, y)$: Değişkenlerin ortak olasılık dağılımını

$p_{1(x)}$: Değişkeninin marjinal olasılığını

X : Veri matrisi

$\overline{X}$: Değişkenlerin ortalamaları

x_i : Değişkenin i. gözlemini

p_k : Düğüm v üzerindeki gözlemlerin k sınıfına ait olma oranı

ω_i : Bölünme sonucu oluşan alt düğümlerdeki örneklerin oransal ağırlığını

$Gini(Xj, vi$: Alt düğümlerin safsızlık düzeyini

$p(x)$: Bir gözlemin 1 sınıfına ait olma olasılığını

β_0 : Sabit terimini

$P(C \mid X)$: X gözlemleri verildiğinde C sınıfına ait olma olasılığı

$P(X \mid C)$: C sınıfı altında X gözlemlerinin olasılığı

$P(C)$: C sınıfının gerçekleşme olasılığı

$P(X)$: X gözlemlerinin gerçekleşme olasılığına

KISALTMALAR

ANOVA **:** Varyans Analizi

Bagging **:** Bootstrap Aggregating

ÇKA **:** Çok Katmanlı Algılayıcılar

DDA **:** Doğrusal Diskriminant Analizi

DFA **:** Doğrulayıcı Faktör Analizi

DN **:** Doğru Negatif

DP **:** Doğru Pozitif

SİMGELER VE KISALTMALAR DİZİNİ (devam ediyor)

DVM : Destek Vektör Makineleri

EA : Ekstra Ağaçlar

GA : Gradyan Artırma

KA : Karar Ağaçları

KDA : Kuadratik Diskriminant Analizi

KNN : K En Yakın Komşu

LASSO : Least Absolute Shrinkage and Selection Operator (En Küçük Mutlak Daralma ve Seçim Operatörü)

LR : Lojistik Regresyon

MÖ : Makine Öğrenmesi

NB : Naive Bayes

RFE : Özyinelemeli Özellik Eleme

RFE-LR : Yinelenen Özellik Eleme- Lojistik Regresyon

RFE-RO : Yinelenen Özellik Eleme - Rastgele Orman

RO : Rastgele Orman

SHAP : Shapley Additive Explanations

TBA : Temel Bileşenler Analizi

XAI : Açıklanabilir Yapay Zekâ

Voting : Oylama

YSA : Yapay Sinir Ağları

YN : Yanlış Negatif

YP : Yanlış Pozitif

BÖLÜM 1

GİRİŞ

1.1 ARAŞTIRMANIN GEREKÇESİ VE AMACI

Günümüzde veri bilimi ve makine öğrenmesi (MÖ) alanlarında yaşanan hızlı gelişmeler, özellikle yüksek boyutlu veri kümelerinde etkili ve anlamlı analizlerin gerçekleştirilmesini gerekli kılmaktadır. Büyük ölçekli veri setlerinde yer alan alakasız veya gürültülü özellikler, MÖ temelli karar destek sistemlerinin performansını olumsuz yönde etkileyebilmekte, hesaplama maliyetlerini artırmakta ve modelin yorumlanabilirliğini azaltmaktadır. Bu doğrultuda, özellik seçimi süreci, bu sistemlerde kritik bir ön işleme adımı olarak öne çıkmakta; hem model başarımını hem de işlem verimliliğini artırırken, aynı zamanda verideki en bilgilendirici ve ayırt edici değişkenlerin belirlenmesini sağlamaktadır.

Özellik seçim yöntemleri genel olarak üç temel kategori altında sınıflandırılmaktadır. Filtre tabanlı yöntemler**,** özellikleri veri setinin istatistiksel özelliklerine dayanarak bağımsız şekilde değerlendirmektedir. Korelasyon katsayısı, varyansa analizi (ANOVA) ve bilgi kazancı gibi istatistiksel ölçütler bu gruba örnek olarak verilebilir. Sarmal yöntemler**,** özellik alt kümelerini belirli bir sınıflandırıcının performansına göre değerlendirerek seçim yapmaktadır. Bu yöntemlerde genellikle yinelemeli stratejiler kullanılmakta olup, Özyinelemeli Özellik Eleme (Recursive Feature Elimination, RFE) bu gruba dâhildir. Gömülü yöntemler ise, özellik seçimini model eğitimiyle eşzamanlı olarak gerçekleştirmetktedir. LASSO (Least Absolute Shrinkage and Selection Operator, En Küçük Mutlak Daralma ve Seçim Operatörü) regresyonu ve Rasgele Orman (RO) gibi algoritmalar aracılığıyla özelliklerin önem dereceleri hesaplanmakta ve seçim süreci modelin öğrenme yapısına entegre biçimde yürütülmektedir.

Bu tez çalışmasının temel amacı, farklı öznitelik seçimi yöntemlerinin çeşitli MÖ algoritmalarıyla kombinasyonlarının, karar destek sistemlerinin sınıflandırma performansı üzerindeki etkilerini karşılaştırmalı olarak değerlendirmek ve bu yöntemlerin avantajları ile sınırlılıklarını ortaya koymaktır. Bireylerin çevreye yönelik tutumlarının sınıflandırılmasına

yönelik bir problem çerçevesinde yürütülen bu çalışma, yüksek boyutlu veri yapıları içerisinde öznitelik seçiminin model performansına olan katkısını kapsamlı biçimde analiz etmeyi hedeflemektedir.

Bu doğrultuda çalışma, aşağıdaki araştırma sorularına odaklanmaktadır:

- Hangi öznitelik seçimi yöntemi, hangi MÖ algoritmasıyla birlikte kullanıldığında, veri karakteristiklerine (boyut, korelasyon yapısı vb.) bağlı olarak daha yüksek sınıflandırma başarımı sunmaktadır?

- Öznitelik seçiminin uygulanması, MÖ tabanlı karar destek sistemlerinin performans metrikleri üzerinde anlamlı farklılıklara yol açmakta mıdır?

Bu gerekçeler doğrultusunda, çalışmanın hem MÖ tabanlı sınıflandırma sistemlerinde özellik seçimi süreçlerinin başarım üzerindeki etkisini ortaya koyması hem de çevresel tutum düzeylerinin belirlenmesine yönelik etkili ve sistematik bir yaklaşım sunması beklenmektedir. Ayrıca, bu bağlamda geliştirilen modellerin, mühendislik alanında veri odaklı karar destek sistemlerinin tasarımı ve optimizasyonuna yönelik uygulamalara katkı sağlayacağı öngörülmektedir.

1.2 ARAŞTIRMANIN KATKISI

Bu çalışma, özellik seçimi yöntemlerinin MÖ tabanlı karar destek sistemlerinin üzerindeki etkilerini sistematik bir yaklaşımla ele alarak, hem yöntemsel hem de uygulamaya yönelik önemli katkılar sunmayı amaçlamaktadır.

Yöntemsel olarak, farklı öznitelik seçimi yaklaşımlarının çeşitli MÖ algoritmalarıyla birlikte kullanımına dair kapsamlı bir analiz gerçekleştirilmekte ve bu doğrultuda araştırmacılar için modelleme sürecinde rehber niteliğinde bir çerçeve sunulmaktadır.

Uygulamalı katkı açısından ise, çevresel tutum düzeylerinin sınıflandırılması problemi üzerinden, öznitelik seçimi ile algoritma tercihleri arasındaki etkileşim incelenmekte; bu yöntemlerin gerçek dünya veri kümelerine uygulanabilirliği değerlendirilmektedir. Böylece,

benzer veri yapısına sahip problemler üzerinde çalışan araştırmalar için yönlendirici ve faydalı bir kaynak oluşturulması hedeflenmektedir.

1.3 ARAŞTIRMANIN KAPSAMI VE SINIRLILIKLARI

Yürütülen bu tez çalışmasında, özellik seçim yöntemlerinin MÖ tabanlı sınıflandırma modellerinin performansı üzerindeki etkilerini incelemeyi amaçlamaktadır. Çalışma, açık erişimli bir çevresel tutum veri seti kullanılarak gerçekleştirilmiştir. Bu veri seti, bireylerin çevresel davranış ve tutumlarını yansıtan çeşitli öznitelikleri içermektedir. çalışmad kapsamında, farklı istatistiksel ve MÖ temelli yaklaşımları temsil eden on özellik seçim yöntemi değerlendirilmiştir. Bu yöntemler arasında ANOVA, Ki-kare, Lasso, Karşılıklı Bilgi, Temel Bileşen Analizi (TBA), RO, Lojistik Regresyon (LR) ve RO ile Özyinelemeli Özellik Eleme Teknikleri (Recursive Feature Elimination, RFE) ile varyans eşiği tabanlı yöntem yer almaktadır. Sınıflandırma sürecinde ise Naive Bayes (NB), Gradyan Arttırma (GA), Destek Vektör Makineleri (DVM), K En Yakın Komşu (KNN), Lojistik Regresyon (LR), Oylama (Voting), Çok Katmanlı Algılayıcı (ÇKA) ve Torbalama gibi on bir farklı MÖ algoritma kullanılmıştır. Model performansları doğruluk, kesinlik, duyarlılık ve F1 skoru gibi çoklu değerlendirme ölçütleriyle karşılaştırılmıştır. Deneysel süreçte veri seti %80 eğitim ve %20 test olarak ayrılmış, 5 katlı çapraz doğrulama uygulanarak modellerin güvenilirliği artırılmış ve aşırı öğrenme riski en aza indirilmeye çalışılmıştır.

Bu çalışmanın mühendislik perspektifinden değerlendirildiğinde, çeşitli sınırlılıkları bulunmaktadır. Öncelikle, deneysel veri kısıtlamaları kapsamında kullanılan veri seti belirli bir çevresel tutum ölçeği ile sınırlı olup, gerçek mühendislik uygulamalarında karşılaşılan gürültülü, eksik ya da dinamik veri senaryolarını tam anlamıyla yansıtmamaktadır. Kullanılan özellik seçimi ve sınıflandırma yöntemleri, tüm olası teknik ve kombinasyonları kapsamamaktadır; özellikle derin öğrenme tabanlı özellik çıkarımı yöntemleri bu çalışmada değerlendirilmemiştir. Ayrıca, bazı karmaşık algoritmaların hiperparametre ayarlamaları, hesaplama kaynakları ve zaman kısıtlamaları nedeniyle tam olarak optimize edilememiş olabilir. Bu durum, bazı modellerin potansiyel performanslarının altında kalmasına neden olmuş olabilir. Son olarak, kapsamlı bir analiz gerçekleştirmek amacıyla daha fazla sayıda deney ve farklı veri setleri üzerinde testler yapılması mümkün olabilirdi; ancak zaman ve kaynak sınırlılıkları nedeniyle bu çalışmanın kapsamı belirli çerçevede tutulmuştur.

Tüm bu sınırlılıklara rağmen, çalışma, özellik seçim yöntemlerinin sınıflandırma başarımı üzerindeki etkilerini anlamaya yönelik önemli bulgular sunmakta ve sonraki araştırmalar için sağlam bir temel oluşturmaktadır. Elde edilen bulguların farklı veri kümeleri ve genişletilmiş algoritma yelpazesiyle test edilmesi, genellenebilirliği artırmak açısından gelecekte yapılacak çalışmalara katkı sağlayacaktır.

1.5 TEZİN YAPISI

Tezin birinci bölümünde, araştırmanın kuramsal temelleri sunulmuş; amacı, özgün katkısı, kapsamı ve sınırlılıkları ayrıntılı biçimde açıklanmıştır. İkinci bölümde, çevre, çevre bilinci, çevresel tutumlar ve bu tutumları etkileyen faktörler literatür temelli bir yaklaşımla kapsamlı biçimde ele alınacaktır. Üçüncü bölümde, araştırmada kullanılan veri seti, ön işleme süreci, öznitelik seçimi yöntemleri ve uygulanan sınıflandırma algoritmaları tanıtılacaktır. Dördüncü bölümde ise elde edilen bulgular sistematik olarak sunulacak ve ilgili analizler gerçekleştirilecektir. Beşinci ve son bölümde ise sonuçlar tartışılacak, önerilerde bulunulacak ve çalışmanın bilimsel katkıları değerlendirilecektir.

BÖLÜM 2

KURAMSAL TEMELLER

2.1 ÇEVRE KAVRAMI

Çevre, insanın içinde bulunduğu ve yaşamını sürdürdüğü fiziksel, biyolojik ve toplumsal koşulların bütünüdür (Çepel 1990). Herring ve Kaplan (2000), çevreyi bir organizmanın dışında kalan her şeyi kapsayan bir yapı olarak tanımlamaktadır. Bu yaklaşıma göre çevre; yaşamı destekleyen temel kaynakları (su, hava, toprak) ve bu kaynaklarla etkileşim hâlindeki tüm canlı-cansız varlıkları içeren, fiziksel, kimyasal, biyolojik ve toplumsal faktörlerin bir arada işlediği karmaşık bir sistemdir. Bu sistem, yalnızca doğal unsurları değil, aynı zamanda sosyal, kültürel ve ekonomik dinamikleri de içine alan çok katmanlı bir yapıya sahiptir. Atmosferden su kaynaklarına, topraktan biyolojik çeşitliliğe kadar uzanan bu yapı; bitki örtüsü, hayvanlar ve insan gibi tüm unsurların karşılıklı etkileşimleriyle şekillenmektedir. Bu nedenle çevre, ekoloji başta olmak üzere birçok disiplinin ilgi alanına girmekte ve farklı bilimsel bakış açılarıyla ele alınmaktadır.

Çevreye ilişkin sorunların günümüzde yalnızca bireysel ya da ulusal düzeyde değil, küresel ölçekte ele alınması gerekmektedir. Zira çevre, tüm insanlığın ortak değeri ve sorumluluğu olarak kabul edilmekte; bu da uluslararası iş birliklerini ve çok paydaşlı çözümleri zorunlu kılmaktadır (Downs vd. 2020). Doğal kaynakların korunması, biyolojik çeşitliliğin sürdürülmesi, iklim değişikliğiyle mücadele, hava, su ve toprak kirliliğinin önlenmesi, enerji yönetimi ve atık kontrolü gibi konular çevre kavramının günümüzde ulaştığı kapsamı yansıtmaktadır. Bu alanlara ek olarak, çevresel sorunların sosyal adalet, ekonomik kalkınma ve halk sağlığı gibi toplumsal boyutlarla da yakından ilişkili olduğu görülmektedir.

Küresel ısınma, ormansızlaşma, su kaynaklarının tükenmesi ve plastik kirliliği gibi çevresel tehditler, dünya genelinde ciddi riskler oluşturmakta; iklim değişikliği nedeniyle artan doğal afetler tarımsal üretimi ve gıda güvenliğini tehdit ederken, yükselen deniz seviyesi kıyı bölgelerindeki yaşam alanlarını riske sokmaktadır (Özdemir 2009, Sağsöz ve Doğanay 2019).

Bu bağlamda çevresel sorunların disiplinlerarası bir yaklaşımla değerlendirilmesi ve çözümlenmesi büyük önem taşımaktadır. Çevrenin korunmasına yönelik stratejiler yalnızca devlet politikalarıyla sınırlı olmayıp, bireylerden sivil toplum kuruluşlarına, ulusal otoritelerden uluslararası kuruluşlara kadar çok çeşitli aktörlerin iş birliğini gerektirmektedir. Sürdürülebilir kalkınma hedeflerine ulaşılmasında çevre bilincinin artırılması, çevre dostu teknolojilerin yaygınlaştırılması ve çevresel eğitim yoluyla toplumsal farkındalığın geliştirilmesi önem arz etmektedir (Arifin vd. 2023). Bu kapsamda, çevre kavramının çok boyutlu yapısını daha sistematik bir biçimde ele alabilmek amacıyla, farklı kategoriler altında sınıflandırılarak incelenmesi gerekmektedir.

2.2 ÇEVRENİN SINIFLANDIRILMASI

Çevre, içerdiği bileşenler ve etkileşim düzeylerine göre farklı şekillerde sınıflandırılmaktadır. Genel kabul gören bir yaklaşıma göre çevre; doğal çevre, yapay çevre ve sosyokültürel çevre olmak üzere üç ana başlık altında ele alınmaktadır (Şahin 2019). Doğal çevre, insan müdahalesi olmaksızın var olan ve kendiliğinden işleyen ekosistemleri kapsamaktadır. Bu kapsamda atmosferik koşullar, hidrolojik döngüler, toprak yapısı, bitki örtüsü, hayvan toplulukları ve jeolojik oluşumlar gibi unsurlar yer almaktadır. Doğal çevre, biyolojik çeşitliliğin devamlılığı ve ekosistem hizmetlerinin sürdürülebilirliği açısından hayati öneme sahiptir. Özellikle iklim düzenleme, su arıtma, toprak oluşumu ve tozlaşma gibi kritik ekolojik süreçler, sağlıklı işleyen doğal çevre sistemleri tarafından gerçekleştirilmektedir (Akkuş 2024, Keleş ve Hamamcı 1998). Yapay çevre ise insan faaliyetleri sonucunda oluşturulan ve dönüştürülen yapılı çevreyi ifade etmektedir. Kentsel yerleşim alanları, ulaşım ağları, tarımsal peyzajlar, endüstriyel bölgeler ve enerji üretim tesisleri bu kategoride değerlendirilmektedir. Sanayi Devrimi'nden bu yana hızla artan yapay çevre oluşumları, doğal çevre üzerinde önemli baskılar oluşturmaktadır. Özellikle kentleşme ve sanayileşme süreçleri, doğal kaynak tüketimini artırmakta ve ekosistemlerin dengesini bozmaktadır (Akkuş 2024). Sosyokültürel çevre, insan topluluklarının oluşturduğu soyut sistemleri içermektedir. Kültürel değerler, ekonomik modeller, yönetim yapıları, eğitim sistemleri ve sosyal ilişkiler ağı bu kapsamda değerlendirilmektedir. Sosyokültürel çevre, insanların doğal çevreyle kurduğu ilişkinin niteliğini belirleyen temel faktördür. Toplumların çevre algısı, tüketim alışkanlıkları ve doğal kaynak kullanım pratikleri, sosyo-kültürel yapı tarafından şekillendirilmektedir.

Bu üçlü sınıflandırma, çevresel sorunların çok boyutlu yapısını anlamada ve etkili çözüm stratejileri geliştirmede önemli bir analitik araç sunmaktadır. Günümüzde karşılaşılan çevresel sorunların çoğu, bu üç çevre türü arasındaki etkileşimlerin yönetilememesinden kaynaklanmaktadır. Sürdürülebilir bir gelecek için, doğal çevrenin korunması, yapay çevrenin ekolojik prensiplere uygun şekilde planlanması ve sosyokültürel çevrenin sürdürülebilirlik değerleriyle uyumlu hale getirilmesi gerekmektedir. Bu bağlamda, çevresel yönetim politikalarının başarısı, bu üç alan arasındaki dengenin sağlanabilmesine bağlıdır.

2.3 İNSAN VE ÇEVRE İLİŞKİSİ

İnsan ile çevre arasındaki ilişki, yalnızca fiziksel bir etkileşimle sınırlı kalmamış, tarihsel süreçte karşılıklı etkileşimle gelişerek çok boyutlu ve dinamik bir yapı kazanmıştır. Bu etkileşim sürecinde insan, yalnızca çevrenin bir ürünü olmanın ötesine geçmiş; onu dönüştürebilen, şekillendirebilen tek canlı türü olarak doğaya tarih boyunca çeşitli müdahalelerde bulunmuştur. Tarım devrimiyle başlayan bu müdahaleler, sanayi devrimiyle birlikte daha sistematik ve yaygın bir hâl almış; kentleşme, sanayileşme, nüfus artışı ve tüketim alışkanlıklarında köklü değişimleri beraberinde getirmiştir (Çepel 1990, Ponting 2007). Bu değişimler ise doğal kaynakların tükenmesine, ekosistem dengesinin bozulmasına ve çevresel sorunların küresel ölçekte yayılmasına neden olmuştur (Çetin vd. 2023).

Çevresel koşullarda meydana gelen bu dönüşümler, yalnızca doğayı değil; insan yaşamını, sağlığını, ekonomik faaliyetlerini ve toplumsal örgütlenme biçimlerini de doğrudan etkilemektedir. Bu etkilerin somut yansımaları ise iklim değişikliği, doğal afetlerin şiddetindeki artış, su kıtlığı ve toprak verimliliğindeki azalma gibi sorunlarda açıkça görülmektedir. Tüm bu çevresel tehditler, insan yerleşim düzeninden üretim biçimlerine kadar pek çok alanda belirleyici bir rol oynamaktadır (IPCC 2021).

Bu nedenle insan-çevre ilişkisi çift yönlü bir etkileşim barındırmakta ve sürdürülebilir bir temelde yeniden kurgulanması gerekmektedir. Doğaya zarar vermeyen, kaynakların bilinçli kullanıldığı ve ekolojik dengeyi gözeten yaşam biçimlerinin benimsenmesi bu sürecin temelini oluşturmaktadır. Bu doğrultuda geliştirilecek çevre dostu politikalar, çevre bilincini artırmayı hedefleyen eğitim programları ve toplumsal katılımı teşvik eden uygulamalar, dönüşümün kalıcı hâle gelmesine katkı sağlayacaktır. Böylece yalnızca günümüz toplumlarının değil,

gelecek kuşakların da sağlıklı, dengeli ve sürdürülebilir bir çevrede yaşamlarını sürdürebilmeleri mümkün olacaktır.

2.4 ÇEVRE SORUNLARI

İnsan-çevre etkileşiminin tarihsel süreçte geçirdiği dönüşüm, günümüzde karmaşık ve kapsamı giderek genişleyen çevresel sorunları beraberinde getirmiştir. Özellikle Sanayi Devrimi sonrasında hız kazanan üretim ve tüketim modelleri, doğal kaynaklar üzerinde geri dönüşü güç tahribatlara yol açmış; bu durum yalnızca ekolojik dengenin bozulmasıyla sınırlı kalmayıp, insan sağlığı, ekonomik istikrar ve toplumsal refah üzerinde de ciddi tehditler oluşturmuştur.

Günümüzde ise çevre sorunları, birbirine bağlı yapısıyla yalnızca yerel değil, aynı zamanda küresel ölçekte etkilerini hissettirmektedir. Doğal kaynakların aşırı tüketimi, hızlı nüfus artışı, plansız kentleşme, iklim değişikliği, çevresel kirlilik, biyolojik çeşitlilik kaybı ve atık yönetimi gibi temel problemler, bu sürecin başlıca göstergeleri arasında yer almaktadır. Bu sorunlar arasında özellikle doğal kaynakların tükenme süreci, hem kapsamı hem de sonuçları bakımından öncelikli bir tehdit olarak öne çıkmaktadır. Özellikle ormanlar, su kaynakları, madenler ve fosil yakıtlar gibi yaşam için kritik öneme sahip kaynakların hızla azalması, hem ekosistemlerin sürdürülebilirliğini hem de insan yaşamının devamlılığını riske sokmaktadır.

Doğal kaynaklar üzerindeki bu baskıyı artıran temel etkenlerden biri de hızla artan dünya nüfusudur. Nüfus artışı, enerji, su, gıda ve barınma gibi temel ihtiyaçlara olan talebi sürekli yükseltmekte; bu da doğal sistemlerin taşıma kapasitesini zorlamaktadır. Artan ihtiyaçları karşılamak amacıyla fosil yakıt kullanımı, endüstriyel üretim ve tarımsal faaliyetler hız kazanmakta; bu süreç, çevre kirliliğini ve sera gazı salınımını artırarak iklim değişikliğini daha da derinleştirmektedir. Ayrıca, kalabalık nüfusun sınırlı coğrafi alanlara yoğunlaşması, altyapı yetersizliklerine, atık yönetiminde aksaklıklara ve doğal yaşam alanlarının tahribine yol açmaktadır. Bu durum, yalnızca çevresel değil, aynı zamanda sosyal ve ekonomik sürdürülebilirliği de tehdit eden çok yönlü bir soruna dönüşmektedir (Yaylı 2017).

Nüfus artışıyla doğrudan ilişkili bir diğer sorun ise plansız kentleşmedir. Özellikle büyük kentlerde hızlı nüfus artışına bağlı olarak gelişen çarpık yapılaşma, doğal alanların tahribatına ve arazi kullanımında verimsizliklere neden olmaktadır. Tarım ve orman alanlarının yerini betonlaşmaya bırakması, hem gıda güvenliğini hem de ekosistem dengesini zayıflatmaktadır.

Ayrıca, altyapı eksiklikleri, ulaşım sorunları ve artan enerji talebi, hava kalitesini olumsuz etkilemekte ve yaşam kalitesini düşürmektedir. Plansız kentleşme, çevre sorunlarının kent ölçeğinde yoğunlaştığı, ancak etkilerinin bölgesel ve küresel düzeyde hissedildiği karmaşık bir olgudur (Çelikkıran 1997).

Çevresel sorunların en belirgin sonuçlarından biri olan iklim değişikliği, günümüzde küresel ölçekte karşı karşıya kalınan en ciddi tehditlerden biridir. Fosil yakıtların yoğun kullanımı ve sanayi kaynaklı sera gazı emisyonları, atmosferde ısı tutulumunu artırarak küresel sıcaklıkların yükselmesine neden olmaktadır. Bu durum, buzulların erimesi, deniz seviyelerinin yükselmesi, kuraklık, ani hava olayları ve mevsimsel kaymalar gibi pek çok çevresel riski beraberinde getirmektedir. Aynı zamanda tarım, su kaynakları ve doğal yaşam alanları üzerinde de yıkıcı etkiler oluşturmaktadır (Karaman ve Gökalp 2010).

İklim değişikliğiyle yakından ilişkili bir diğer sorun ise çevre kirliliğidir. Hava, su ve toprak kirliliği, endüstriyel faaliyetler, ulaşım, tarım ilaçları ve evsel atıklar gibi insan kaynaklı faktörlerden kaynaklanmaktadır. Hava kirliliği, özellikle kentsel bölgelerde solunum yolu hastalıklarını artırmakta; su kirliliği ise hem insan sağlığını tehdit etmekte hem de sucul yaşamı olumsuz etkilemektedir. Toprak kirliliği ise verimliliği azaltarak gıda güvenliği üzerinde risk oluşturmaktadır. Bu kirlilik türleri, yalnızca çevresel değil, aynı zamanda toplumsal sağlık açısından da ciddi sonuçlar doğurmaktadır (Güler ve Çobanoğlu 1994, Menteşe 2017, Özdemirkol 2023).

Çevresel kirliliğin ve habitat tahribatının doğrudan sonucu olarak biyolojik çeşitlilik kaybı da günümüzde önemli bir tehdit hâline gelmiştir. Ormanların yok edilmesi, sulak alanların kurutulması, tarım alanlarının genişletilmesi ve iklim değişikliği gibi faktörler; birçok hayvan ve bitki türünün yaşam alanlarını kaybetmesine neden olmaktadır. Ekosistemlerin bütünlüğünün bozulması, sadece türlerin yok olmasına değil, aynı zamanda doğal dengeyi sağlayan ekolojik hizmetlerin zayıflamasına da yol açmaktadır (Kurt 2017).

Son olarak, modern yaşamın önemli sorunlarından biri olan atık yönetimi de çevresel sürdürülebilirlik açısından ciddi bir risk oluşturmaktadır. Tüketim alışkanlıklarındaki artış, ambalaj kullanımının yaygınlaşması ve geri dönüşüm bilincinin yetersizliği; katı atıkların, plastiklerin ve tehlikeli atıkların çevreye kontrolsüz bir şekilde yayılmasına neden olmaktadır. Özellikle gelişmekte olan ülkelerde atıkların etkin bir şekilde toplanamaması hem çevre

sağlığını hem de kent estetiğini olumsuz etkilemektedir. Bu nedenle, atıkların azaltılması, geri dönüşüm uygulamalarının yaygınlaştırılması ve sürdürülebilir tüketim alışkanlıklarının teşvik edilmesi büyük önem taşımaktadır (Bulut ve Özer 2024).

2.5 ÇEVRE BİLİNCİ VE TUTUM KAVRAMI

Çevre bilinci, bireylerin çevresel sorunlara karşı farkındalık geliştirmesi, çevrenin korunması yönünde sorumluluk üstlenmesi ve bu doğrultuda tutum ve davranış geliştirmesi sürecini kapsayan çok boyutlu bir kavramdır (Zheng 2010). Bu bilinç, bireyin yalnızca çevresel bilgi sahibi olmasıyla sınırlı kalmayıp, çevreye yönelik değer yargılarını ve buna dayalı davranış eğilimlerini de kapsamaktadır (Türküm 1998). İnsanların doğal çevrenin korunması, kaynakların sürdürülebilir biçimde kullanılması ve çevresel etkilerin en aza indirilmesi konularında duyarlılık geliştirmeleri, çevre bilincinin temel göstergeleri arasında yer almaktadır. Bu bağlamda çevre bilinci; çevreye ilişkin düşünceler, bu düşüncelere bağlı duygular ve bunların yaşama yansıması olan davranışlar olmak üzere üç boyutta ele alınmaktadır (Yılmaz ve Zorlutuna 2024). Söz konusu bu boyutlar; bireyin çevresel sorunlara ilişkin bilgi ve farkındalık düzeyini yansıtan düşünsel boyut, çevreye yönelik değer yargıları ile duygusal tepkilerini içeren duyuşsal boyut ve çevreyle ilgili sergilenen somut eylemleri kapsayan davranışsal boyut olarak tanımlanmaktadır (Türküm 1998).

Çevresel bilincin çok boyutlu yapısı, bireylerin çevreye ilişkin bilgi sahibi olmalarının bu bilgiyi doğrudan çevre dostu davranışlara dönüştürmelerini garanti etmediğini göstermektedir. Nitekim bireyler, çevre kirliliği, iklim değişikliği ya da doğal kaynakların tükenmesi gibi sorunlar hakkında yeterli bilgiye sahip olsalar bile, bu bilgiyi davranışa dönüştürme konusunda güçlük yaşayabilmektedirler (Kollmuss ve Agyeman 2002, van Valkengoed vd. 2022). Türküm (1998), çevre bilincinin gelişiminin kişilik yapısı, sosyal çevre, değer sistemleri ve eğitim gibi birçok faktörle etkileşim hâlinde şekillendiğini belirtmektedir. Bu bağlamda çevresel tutum kavramı, çevre bilincinin davranışa dönüşmesinde belirleyici bir rol oynamaktadır. Genel anlamda tutum; bireyin belirli bir nesne, olgu ya da konuya karşı sergilediği bilişsel (inanç), duyuşsal (duygu) ve davranışsal (eylem eğilimi) boyutlardan oluşan eğilimler bütünüdür (De Groot 2019, Eagly ve Chaiken 1993). Çevresel tutum ise bu yapının çevre konularına yönelmiş biçimidir. Olumlu çevresel tutuma sahip bireyler, çevreyi korumaya yönelik duyarlılık sergilemekte ve geri dönüşüm, enerji tasarrufu ile su kaynaklarının etkin kullanımı gibi davranışları daha sık gerçekleştirmektedirler (Kollmuss ve Agyeman 2002). Bu doğrultuda

çevre bilinci, yalnızca bilişsel düzeydeki bilgi birikimiyle değil, aynı zamanda bu bilginin tutum ve davranışlara yansıma düzeyiyle birlikte ele alınmalıdır (Kapyla ve Wahlström 2000).

2.6 BİREYLERİN ÇEVREYE YÖNELİK TUTUMLARINI ŞEKİLLENDİREN FAKTÖRLER

Bireylerin çevresel tutumları, yalnızca bilgi düzeyleriyle sınırlı olmayan, çok sayıda içsel ve dışsal etmenin etkileşimiyle biçimlenen çok boyutlu bir yapıya sahiptir. Bu tutumların oluşumunda; bireysel değerler, kişilik özellikleri, inanç sistemleri, sosyal normlar, kültürel arka plan, eğitim düzeyi ve medyadan edinilen bilgiler önemli rol oynamaktadır (Kollmuss ve Agyeman 2002, Türküm 1998). Çevresel tutumlar, çoğunlukla çevresel davranışların öncülü olarak kabul edilmekte ve bireylerin çevreyle ilgili karar alma süreçlerini doğrudan etkilemektedir. Bu kapsamda çeşitli teorik yaklaşımlar, çevresel tutumların bireyler ve toplumlar arasındaki farklılıklarını açıklamaya çalışmakta; özellikle bireysel refah düzeyi, ulusal gelir, sosyal değerler ve yaşam deneyimleri gibi faktörlerin etkisi üzerinde durulmaktadır.

Çevresel tutumlar aynı zamanda bireyin çevresel davranışlarını belirleyen temel bir unsur olarak değerlendirilmekte ve çevreye duyarlılığın davranışsal düzeydeki yansımalarını şekillendirmektedir (Eagly ve Chaiken 1993).

Teorik yaklaşımlar, çevresel tutumlar arasındaki bireysel ve uluslararası farklılıkları açıklamaya çalışmaktadır. Örneğin, bireysel refah düzeyi ile çevresel kaygı arasındaki ilişki çevre sosyolojisi bağlamında hâlâ tartışmalı bir konudur (Gadenne vd. 2009). Bu kapsamda yapılan araştırmalar, sosyo-demografik değişkenlerin çevresel tutumlar üzerindeki etkisini ortaya koymaktadır. Özellikle kadınların, gelir ve eğitim düzeyleri sabit tutulduğunda dahi erkeklere kıyasla daha yüksek çevresel kaygı sergiledikleri rapor edilmiştir (Gadenne vd. 2009). Yaş faktörü de benzer biçimde tutumları etkileyen önemli bir değişkendir; genç bireylerin çevresel farkındalığa daha açık olduğu, özellikle çocuklukta kazanılan doğa deneyimlerinin bu farkındalık üzerinde belirleyici olduğu saptanmıştır (Fayiga vd. 2018). Çevresel tutumların gelişiminde çocukluk döneminde doğayla kurulan olumlu deneyimlerin belirleyici etkisi sıklıkla vurgulanmaktadır. Evde çevre konularının konuşulması, doğa temalı filmlerin izlenmesi ve çevre hakkında okuma alışkanlığı edinilmesi gibi deneyimlerin, bireylerin çevreye ilişkin olumlu tutumlar geliştirmesinde etkili olduğu bildirilmektedir. Ayrıca, çevresel

sorunlara karşı endişe duyan bireylerin doğa temelli etkinliklere katılma eğilimlerinin daha yüksek olduğu belirtilmiştir (Fayiga vd. 2018). Bazı çalışmalar ise değer sistemlerinin çevresel tutumlar üzerindeki etkisini incelemiştir. Dini inançların ve mezhep aidiyetinin çevresel kaygı ile doğrudan bir ilişkisi olmadığı, ancak özgecilik gibi evrensel insani değerlerin çevreyle ilgili duyarlılıkla pozitif yönde ilişkilendirilebileceği ileri sürülmektedir. Bu kapsamda Gadenne ve arkadaşları (2009) tarafından on dört ülkede üniversite öğrencileriyle gerçekleştirilen bir araştırmada, bireylerin değer öncelikleri ile çevresel tutumları arasında anlamlı ilişkiler tespit edilmiştir.

Sonuç olarak, çevresel tutumların oluşumu çok boyutlu bir süreçtir ve bireylerin bu tutumları benimsemesinde hem kişisel hem de toplumsal düzeyde çeşitli faktörlerin etkili olduğu görülmektedir. Bu faktörlerin anlaşılması, çevre dostu davranışların yaygınlaştırılmasına yönelik etkili eğitim ve politika yaklaşımlarının geliştirilmesine katkı sağlayacaktır.

2.7 BİREYLERİN ÇEVRESEL TUTUMLARINI BELİRLEMEDE KULLANILAN YÖNTEMLER

Bireylerin çevresel tutum ve davranışlarını değerlendirmek; çevre eğitimi politikalarının etkililiğini belirlemek, sürdürülebilir davranışları teşvik etmek ve çevresel farkındalık düzeyini artırmak açısından kritik öneme sahiptir. Bu amaçla geliştirilen yöntemler, genel olarak nicel ve nitel yaklaşımlar çerçevesinde sınıflandırılmakta ve her biri belirli avantajlar ve sınırlılıklarla birlikte değerlendirilmektedir.

Nicel temelli ölçüm araçları, çevresel tutumların değerlendirilmesinde en yaygın kullanılan yöntemler arasında yer almaktadır. Bu kapsamda, standartlaştırılmış anket formları ve psikometrik tutum ölçekleri, araştırmacılar tarafından sıklıkla tercih edilmektedir. Özellikle Dunlap ve Van Liere (1978) tarafından geliştirilen Yeni Çevre Paradigması Ölçeği, bireylerin çevreye ilişkin temel dünya görüşlerini ölçmede yaygın olarak kullanılan geçerli ve güvenilir bir araç olarak öne çıkmaktadır. Benzer şekilde, Likert tipi tutum ölçekleri, katılımcıların çevresel kaygı düzeyleri, sorumluluk algıları ve pro-çevresel eyleme geçme niyetleri gibi boyutları nicel verilerle analiz etme olanağı sunmaktadır (Milfont ve Duckitt 2010). Ancak, bu tür öz-bildirim tekniklerinin en önemli sınırlılıkları arasında sosyal beğenirlik yanlılığı (social desirability bias) ve katılımcıların verdikleri yanıtlardaki potansiyel yanıltıcı eğilimler yer almaktadır (Kollmuss ve Agyeman 2002).

Bu ölçeklerden elde edilen verilerin geçerlik ve güvenirlik analizleri, psikometrik değerlendirme teknikleri ile desteklenmektedir. Ayrıca, betimsel analizler, t-testi/ANOVA gibi gruplar arası karşılaştırmalar ve korelasyon analizleri, çevresel tutumların demografik, sosyokültürel ve bireysel değişkenlerle olan ilişkisini ortaya koymada yaygın olarak kullanılmaktadır (Schultz 2001).

Geleneksel ölçüm tekniklerinin sınırlılıklarını aşmak amacıyla, çevresel tutumların değerlendirilmesinde alternatif yaklaşımlar ön plana çıkmaktadır. Bu kapsamda, davranışsal gözlem yöntemleri ve teknoloji tabanlı izleme sistemleri, bireylerin çevre dostu davranışlarını gerçek yaşam bağlamında nesnel biçimde değerlendirme olanağı sunmaktadır. Geri dönüşüm ya da enerji tasarrufu gibi davranışların gözlemlenmesi, tutumların fiili davranışlara dönüşümünü anlamada etkili olurken (Kaiser vd. 2010), sensör tabanlı sistemler sayesinde enerji kullanımı ve atık ayrıştırma gibi etkileşimler gerçek zamanlı olarak izlenebilmektedir. Ayrıca, deneysel araştırma tasarımları bilgilendirme kampanyaları ve ekonomik teşviklerin çevresel tutumlara etkisini nedensel düzeyde analiz etmeye olanak tanımaktadır (Schultz vd. 2007).

Ancak, her yöntemin belirli sınırlılıkları bulunmaktadır. Öz-bildirim temelli tekniklerde yanıt yanlılıkları, gözleme dayalı yöntemlerde yüksek maliyet ve uygulama zorlukları, deneysel tasarımlarda dış geçerlilik sorunları ve büyük veri analizlerinde etik kaygılar araştırma sürecini sınırlayabilmektedir. Ayrıca, kültürel farkların çevresel tutumların ölçümündeki etkisi sıklıkla göz ardı edilmekte, bu durum da genellenebilirlik açısından sorun yaratmaktadır. Bu nedenle, karma yöntem yaklaşımları günümüzde daha fazla önem kazanmakta; nicel ve nitel verilerin birlikte kullanılması, çevresel tutumların çok boyutlu doğasını daha kapsamlı ve güvenilir bir şekilde analiz etmeye olanak tanımaktadır. Bu tür bütünleştirici yaklaşımlar, çevresel tutum ile davranış arasındaki ilişkiyi daha derinlemesine kavramamıza katkı sağlamaktadır.

2.8 MAKİNE ÖĞRENMESİ YAKLAŞIMIYLA ÇEVRESEL TUTUM ANALİZİ

Veri temelli karar alma süreçlerinin önem kazandığı günümüzde, çevresel tutumların değerlendirilmesinde daha güçlü ve esnek analiz yöntemlerine ihtiyaç duyulmaktadır. Bu doğrultuda, MÖ yaklaşımları, sundukları yüksek öngörü gücü ve çok boyutlu veri işleme kapasitesi sayesinde çevresel tutum analizine yeni bir perspektif kazandırmaktadır (Gholami

vd. 2021, Ren 2025). Bu yaklaşımlar, klasik istatistiksel sınırlamaların ötesine geçerek, bireylere ait karmaşık veri yapılarını modelleme ve anlamlı sonuçlar üretme imkânı sunar.

MÖ, sistemlerin geçmiş verilere dayanarak örüntüleri öğrenmesini ve bu öğrenme süreci aracılığıyla gelecekteki veriler için tahminler yapabilmesini sağlayan algoritmalardan oluşmaktadır. Bu çerçevede, özellikle sınıflandırma problemleri için geliştirilen yöntemler, bireylerin çevresel tutum düzeylerinin veriye dayalı olarak belirlenmesini mümkün kılmaktadır. Modelleme süreci, etiketli veriler üzerinde öğrenme gerçekleştirerek, benzer özelliklere sahip bireylerin hangi tutum kategorisine ait olabileceğine ilişkin kestirimlerde bulunmaktadır.

Bu tür veri odaklı yaklaşımlar, çevresel tutumları etkileyen değişkenler arasındaki doğrusal olmayan ve çok değişkenli ilişkileri daha etkili biçimde yakalayabilmektedir (Lee ve Kim 2023). Ayrıca, bazı algoritmalar tarafından sağlanan değişken önem düzeyleri, tutumların belirlenmesinde rol oynayan temel faktörlerin ortaya konmasına yardımcı olmakta; bu da elde edilen bulguların yalnızca sınıflandırma amacıyla değil, aynı zamanda açıklayıcı analizler için de kullanılmasını sağlamaktadır.

MÖ yöntemleri, yüksek doğruluk oranları ve genellenebilirlik kapasitesiyle birlikte, çevresel tutumların bireysel, bölgesel veya toplumsal düzeyde öngörülmesine olanak tanımaktadır. Böylece eğitim politikalarının hedefe yönelik olarak tasarlanması, toplumsal farkındalık düzeylerinin haritalanması ya da davranış değişikliği programlarının daha etkili biçimde planlanması mümkün hale gelmektedir (Wang vd. 2023).

Sonuç olarak, MÖ temelli yaklaşımlar, çevresel tutum analizinde yalnızca verimliliği artırmakla kalmamakta, aynı zamanda nesnellik, ölçeklenebilirlik ve çok boyutlu analiz kapasitesi ile bu alana önemli metodolojik katkılar sunmaktadır. Ancak bu yöntemlerin başarılı biçimde uygulanabilmesi, kaliteli veri setlerinin temini, model seçimi ve doğrulama süreçlerinin titizlikle yürütülmesi ile doğrudan ilişkilidir. Gelecekte, MÖ ile geleneksel yöntemlerin bir arada kullanıldığı hibrit modellerin, çevresel tutumların değerlendirilmesinde çok daha güçlü bir analitik çerçeve sunması beklenmektedir.

2.9 LİTERATÜR TARAMASI

Bu bölümde, bireylerin çevresel tutum düzeylerinin modellenmesine yönelik geliştirilen yaklaşımlar sistematik biçimde ele alınmakta; özellikle MÖ temelli sınıflandırma yöntemlerinin açıklanabilirlik odaklı analiz teknikleriyle entegrasyonuna dayanan güncel çalışmalar değerlendirilmektedir. Literatürde öne çıkan yöntemsel farklılıklar, kullanılan değişken setleri, modelleme stratejileri ve performans metrikleri üzerinden karşılaştırmalı bir çerçeve sunulmaktadır.

Beiser-McGrath ve Huber (2018), bireylerin iklim değişikliği ve çevresel sorunlara yönelik tutumlarını öngörmede etkili olan bireysel özellikleri kapsamlı bir biçimde ele almıştır. Çin, İsviçre ve Amerika Birleşik Devletleri'nde yürütülen üç anket çalışmasından elde edilen veriler doğrultusunda, psikolojik ve sosyodemografik değişkenlerin öngörü gücü karşılaştırılmıştır. Bu amaçla, bireylerin yaş, cinsiyet, eğitim düzeyi, gelir, siyasi görüş gibi klasik demografik özelliklerinin yanı sıra çeşitli psikolojik eğilimler RO algoritmasıyla analiz edilmiştir. Elde edilen bulgular, geleneksel demografik değişkenlerin çevresel tutumları öngörmede sınırlı kaldığını; buna karşılık zaman perspektifi gibi psikolojik özelliklerin daha yüksek bir öngörü gücü sunduğunu göstermektedir. Ayrıca, siyasi görüşler ile iklim değişikliğine ilişkin inançların bazı ülkelerde tutumları belirlemede etkili olduğu, ancak bu etkinin ülkelere göre farklılık gösterdiği ifade edilmiştir. Bu sonuçlar, çevresel tutumların değerlendirilmesinde psikolojik değişkenlerin önemine işaret etmekte; tahmine dayalı modellerin çevresel politika geliştirme süreçlerinde kullanılabilirliğini ortaya koymaktadır. MÖ tabanlı bu yaklaşım, sosyal bilimlerdeki klasik analiz yöntemlerine güçlü bir alternatif sunmakta ve çevresel tutumların sınıflandırılmasında çok boyutlu veri yapılarının dikkate alınmasının model başarımını artırabileceğini göstermektedir (Beiser-McGrath ve Huber 2018).

Jiang et al. (2022), çevresel tutumların sürdürülebilir tüketim davranışları üzerindeki etkisini, Ahlaki Norm Aktivasyon Teorisi çerçevesinde incelemiştir. Çalışmada, çevresel kaygı, kişisel normlar ve algılanan davranışsal kontrol gibi psikolojik ve tutumsal değişkenlerin, bireylerin sürdürülebilir tüketim eğilimleri üzerindeki rolü yapısal eşitlik modellemesi ile analiz edilmiştir. Anket yoluyla toplanan veriler üzerinden yürütülen bu modelleme, teorik yapıların doğruluğunu sınamak açısından değerli bulgular ortaya koymuştur. Bununla birlikte, çalışmada kullanılan geleneksel modelleme yöntemi, değişkenler arası doğrusal ilişkileri temel almakta; bu da çok boyutlu ve karmaşık veri yapılarını açıklamada sınırlılıklar barındırabilmektedir.

Oysa MÖ algoritmaları, özellikle çok sayıda değişken içeren ve doğrusal olmayan ilişkilere sahip veri yapılarında daha esnek ve yüksek öngörü kapasitesine sahip modeller sunabilmektedir. Bu açıdan bakıldığında, Jiang ve arkadaşlarının (2022) çalışması, çevresel tutumların davranış üzerindeki etkisini anlamaya yönelik önemli bir teorik çerçeve sunsa da, benzer problem alanlarında MÖ tabanlı sınıflandırma ve tahmin modellerinin kullanılması, bu ilişkinin daha derinlemesine ve veri odaklı şekilde değerlendirilmesine imkân tanıyabilmektedir.

Zhao ve arkadaşları (2022), bireylerin çevresel davranışlarını etkileyen temel belirleyicileri ortaya koymak amacıyla, sosyo-demografik ve psikolojik değişkenleri içeren kapsamlı bir analiz gerçekleştirmiştir. Çalışma, Çin genelinde uygulanan geniş ölçekli bir anket verisine dayanmaktadır. Katılımcıların çevresel davranış düzeylerini etkileyebilecek faktörler; yaş, cinsiyet, eğitim düzeyi, gelir seviyesi, çevre bilgisi, çevreye yönelik endişe düzeyi ve kişisel sorumluluk algısı gibi değişkenler üzerinden değerlendirilmiştir. Bu kapsamda, elde edilen veriler üzerinde çeşitli denetimli MÖ algoritmaları (örneğin RO ve GA) uygulanarak, her bir değişkenin çevresel davranış üzerindeki görece etkisi analiz edilmiştir. Modelleme sürecinde, algoritmaların doğruluk düzeyleri çapraz doğrulama yöntemleriyle test edilmiş; ayrıca değişken önem sıralaması aracılığıyla, hangi faktörlerin karar sürecinde belirleyici rol oynadığı görselleştirilmiştir. Bulgular, bireylerin çevresel davranışlarında en etkili değişkenin çevre bilgisi düzeyi olduğunu, bunu ise kişisel sorumluluk algısının izlediğini ortaya koymuştur. Ayrıca, eğitim düzeyi ve çevreye yönelik endişe gibi psikolojik değişkenlerin de çevresel davranışları anlamlı biçimde etkilediği belirlenmiştir. Bu yönüyle çalışma, çevresel tutum ve davranışların çok boyutlu yapısını MÖ teknikleriyle analiz ederek, klasik istatistiksel yöntemlerin ötesinde daha esnek ve güçlü bir öngörüsel çerçeve sunmaktadır.

Wang ve arkadaşları (2023), üniversite öğrencilerinin çevresel olarak sorumlu davranışlarını öngörmek amacıyla karar ağacı (KA) temelli bir MÖ yaklaşımı geliştirmiştir. Araştırmada, Çin'in Guangdong bölgesindeki 334 üniversite öğrencisine uygulanan ölçekler aracılığıyla; algılanan davranışsal kontrol, sosyal kimlik, yenilikçi davranış, mekânsal aidiyet, öznel normlar, çevresel aktivizm ve çevresel sorumluluk bilinci gibi değişkenlerin, çevresel davranış üzerindeki etkileri analiz edilmiştir. Bu çok boyutlu veri seti, C5.0 algoritması kullanılarak oluşturulan KA modeli ile değerlendirilmiş ve çevreci davranış düzeylerini yüksek ve düşük olarak sınıflandıran öngörücü bir yapı kurulmuştur. Modelin doğruluk oranı test verisi üzerinde %72,89 olarak hesaplanırken, özellikle çevresel olarak sorumlu davranma istekliliği (önem

katsayısı: 0,1562), yenilikçi davranış (0,1404) ve algılanan davranışsal kontrol (0,1322) en güçlü belirleyiciler olarak öne çıkmıştır. Bu çalışma, hem KA temelli MÖ yaklaşımının çevreci davranışların modellenmesinde etkili bir araç olabileceğini göstermesi hem de özellikle genç bireylerde bu davranışları etkileyen psikososyal değişkenleri önceliklendirerek literatüre önemli bir katkı sunmaktadır. Ayrıca, çevreci davranışın sadece bireysel eğilimler değil, sosyal kimlik, normatif baskılar ve yenilik kapasitesi gibi çok boyutlu unsurlarla şekillendiğine işaret ederek, ileride geliştirilecek eğitim politikaları ve farkındalık stratejileri açısından da yol gösterici niteliktedir.

Değirmenci (2024), bireylerin çevresel tutumlarını öngörmek amacıyla KNN, NB ve DVM olmak üzere üç farklı MÖ yöntemini karşılaştırmalı olarak değerlendirmiştir. Çalışmada Türkiye'de 384 katılımcıdan elde edilen 37 değişkenden oluşan açık kaynaklı bir çevresel tutum veri seti kullanılmıştır. Modelleme sürecinde, hiperparametre optimizasyonu için ızgara arama (grid search) yöntemi kullanılmış ve 10 katlı çapraz doğrulama (k=10) ile modellerin genellenebilirliği test edilmiştir. Elde edilen bulgular, tüm performans metriklerinde (doğruluk, F1 skoru, kesinlik ve geri çağırma) en yüksek başarıyı DVM yönteminin sağladığını ortaya koymuştur (doğruluk = 0,9576, F1 skoru = 0,9615, kesinlik = 0,9576, geri çağırma = 0,9670). KNN ve NB yöntemlerinin ise birbirine yakın ve nispeten daha düşük performans gösterdiği belirlenmiştir. Ayrıca, çalışmada parametre ayarlamalarının model başarısını önemli ölçüde etkilediği vurgulanmış, bu bağlamda hiperparametre optimizasyonunun sınıflandırma başarısındaki kritik rolüne dikkat çekilmiştir. Çevresel tutumların MÖ algoritmaları ile etkili biçimde tahmin edilebileceğini ortaya koyan bu çalışma, ilgili alanda yapılan literatürün yöntemsel çeşitliliğine önemli bir katkı sunmaktadır.

Koklu ve Sulak (2024), bireylerin çevresel tutum düzeylerinin tahmin edilmesine yönelik MÖ temelli sınıflandırma yöntemlerinin uygulanabilirliğini incelemiştir. Çalışmanın temel amacı, çevreye karşı duyarlılığı etkileyen faktörleri istatistiksel olarak modellemek ve sınıflandırma algoritmaları aracılığıyla bireylerin çevresel tutumlarını belirleyebilmektir. Bu kapsamda, araştırmacılar 18–45 yaş aralığında yer alan 384 bireyden veri toplamış ve çevresel olarak sorumlu davranışları ölçmeye yönelik 37 değişkenden oluşan bir ölçek kullanılmıştır. Veri setinde katılımcıların demografik özelliklerine ek olarak çevreye yönelik davranış ve tutumlarına ilişkin bilgiler de yer almıştır. Çalışmada, LR, DVM ve KA gibi farklı MÖ algoritmaları uygulanarak sınıflandırma işlemleri gerçekleştirilmiştir. Modellerin başarımı; doğruluk, kesinlik, duyarlılık ve F1 skoru gibi performans ölçütleri kullanılarak

değerlendirilmiş, ayrıca 10 katlı çapraz doğrulama yöntemiyle test edilmiştir. Elde edilen bulgular, LR algoritmasının %94,53 doğruluk oranı ile en yüksek sınıflandırma başarısını gösterdiğini, DVM algoritmasının %92,96 ve KA algoritmasının %82,55 doğruluk oranı ile takip ettiğini ortaya koymuştur. Elde edilen sonuçlar, çevresel tutumların sınıflandırılmasında MÖ temelli yaklaşımların etkili bir yöntem olarak kullanılabileceğini göstermektedir. Bu bağlamda, çevre eğitimi ve farkındalık çalışmalarında hedef kitlenin belirlenmesi, davranış değişikliği odaklı programların planlanması ve sürdürülebilirlik stratejilerinin veri temelli olarak desteklenmesi açısından önemli bir uygulama alanı sunmaktadır.

Choudhury ve arkadaşları (2024), MÖ yöntemlerinin bireylerin yeşil satın alma niyetlerini tahmin etmedeki başarısını incelemiş ve bu bağlamda Hindistan'daki tüketicilerden toplanan anket verilerini analiz etmiştir. Çalışmada çevresel kaygı, kişisel tutum, algılanan tüketici etkinliği ve sosyal etki gibi psikososyal faktörler dikkate alınarak, lojistik regresyon, karar ağaçları, destek vektör makineleri ve rastgele orman gibi çeşitli MÖ algoritmaları karşılaştırılmıştır. Elde edilen bulgular, özellikle topluluk öğrenmesine dayalı algoritmaların (örneğin RO) yüksek sınıflandırma performansı gösterdiğini ve geleneksel istatistiksel modellere kıyasla daha isabetli öngörülerde bulunabildiğini ortaya koymuştur. Çalışma, yalnızca tüketici davranışlarının modellenmesinde değil, aynı zamanda sürdürülebilir pazarlama stratejilerinin geliştirilmesinde de MÖ yöntemlerinin etkili bir karar destek aracı olarak kullanılabileceğini göstermektedir.

Agarwal ve arkadaşları (2024), çevresel tutumları sınıflandırmak ve bu sınıflandırma sürecinin açıklanabilirliğini artırmak amacıyla Açıklanabilir Yapay Zeka (Explainable Artificial Intelligence, XAI) yöntemleri ile desteklenen çeşitli makine öğrenimi modellerini karşılaştırmalı olarak incelemiştir. Türkiye'de 384 katılımcıdan toplanan çevresel tutum anket verileri kullanılarak gerçekleştirilen çalışmada, bireylerin demografik bilgileri, çevre eğitimi geçmişleri ve çevresel konulara yönelik tutumları analiz edilmiştir. En yüksek doğruluk oranı %94 ile lojistik regresyon modeli tarafından elde edilmiştir. Modelin öngörü süreci SHapley Additive exPlanations (SHAP) yöntemiyle açıklanarak, çevresel tutumları belirlemede en etkili değişkenlerin ortaya konması sağlanmıştır. Özellikle, "Satın aldığım ürünlerin ambalajında geri dönüşüm sembolü olup olmadığına dikkat etmem" ifadesine verilen yanıtın sınıflandırmada en belirleyici faktör olduğu saptanmıştır. Bu sonuçlar, geri dönüşüm farkındalığı ve çevresel tüketim bilincinin tutumları doğrudan etkilediğini göstermektedir.

İlgili literatürde yapılan çalışmalar, MÖ algoritmalarının çevresel tutumların sınıflandırılmasında yüksek öngörü gücüne sahip olduğunu ve geleneksel yöntemlerin ötesine geçerek çok boyutlu veri yapılarının analizine olanak tanıdığını ortaya koymaktadır. Bu tür yaklaşımlar, yalnızca sınıflandırma doğruluğunu artırmakla kalmamakta, aynı zamanda sürdürülebilirlik odaklı politika geliştirme süreçlerinde hedef kitlenin daha doğru biçimde tanımlanmasına ve çevre eğitimi stratejilerinin daha etkili şekilde yapılandırılmasına katkı sunmaktadır. Bu bağlamda, çevresel tutumların nesnel, sistematik ve veri odaklı biçimde değerlendirilmesine yönelik bir model geliştirmeyi amaçlayan bu tez çalışması, literatürdeki mevcut boşluğu doldurmayı hedeflemekte ve ilgili alandaki çalışmalara önemli bir katkı sağlamaktadır. Çalışmada, farklı öznitelik seçimi teknikleriyle çevresel tutumları en iyi temsil eden değişkenler belirlenmiş; bu değişkenler üzerinden çeşitli MÖ algoritmaları ile sınıflandırma modelleri oluşturulmuştur. Bu modellerin başarımı, çoklu performans metrikleriyle değerlendirilmiş ve algoritmalar arasında karşılaştırmalı bir analiz gerçekleştirilmiştir. Böylece yalnızca en yüksek doğruluk oranına ulaşan modellerin belirlenmesi değil, aynı zamanda sınıflandırma başarımına etki eden temel faktörlerin de anlaşılması sağlanmıştır. Bu yönüyle tez, çevre eğitimi ve farkındalık politikalarının şekillendirilmesinde kullanılabilecek güçlü bir karar destek altyapısı sunmakta ve MÖ yöntemlerinin sosyal bilimler bağlamındaki uygulama potansiyelini ortaya koyan disiplinlerarası bir katkı niteliği taşımaktadır.

BÖLÜM 3

YÖNTEM

Bu bölümde, tez çalışması kapsamında kullanılan veri seti, ön işleme adımları, özellik seçimi yöntemleri ve sınıflandırma algoritmaları ayrıntılı bir şekilde açıklanmaktadır. Ek olarak modelleme süreci, performans değerlendirme ölçütleri ve sonuçların analizine ilişkin uygulamalar da sunulmuştur. Yöntemsel yaklaşım, araştırmanın amacına uygun sonuçlara ulaşılmasını sağlamak üzere sistematik bir yapı içerisinde ele alınmıştır.

3.1 VERİ SETİ

Tez çalışmasında, Koklu ve Sulak (2024) tarafından açık erişime sunulan "Environmental Attitude Dataset" adlı veri seti kullanılmıştır (URL-1). Veri seti, Türkiye'de yaşayan ve yaşları 18 ile 45 arasında değişen 177 erkek ve 207 kadından oluşan toplam 384 katılımcıdan çevrim içi anket yöntemiyle toplanmıştır. Katılımcılara uygulanan anket, bireylerin çevresel sorumluluk bilinci, sürdürülebilirlik algısı, çevre dostu davranış eğilimleri ve çevresel farkındalık düzeylerini değerlendirmek amacıyla yapılandırılmıştır (Koklu ve Sulak 2024).

Bu doğrultuda elde edilen veri seti, katılımcıların demografik bilgilerini (cinsiyet, yaş, eğitim düzeyi, gelir durumu, okul öncesi eğitim alma durumu ve yaşanılan yerleşim birimi gibi) içeren değişkenlerin yanı sıra, çevresel tutum ve farkındalıklarını ölçmeye yönelik 27 anket sorusundan oluşmaktadır. Anket sorularına verilen yanıtlar beşli Likert ölçeği (1: Kesinlikle katılmıyorum – 5: Kesinlikle katılıyorum) formatında kodlanmıştır. Ayrıca, katılımcıların çevre eğitimi, küresel ısınma ve iklim değişikliği gibi konularda eğitim alıp almadıklarına ilişkin değişkenler de veri setine dahil edilmiştir. Bu doğrultuda, veri seti Çizgelde 3.1'de verilen sorulara katılımcıların verdikleri yanıtlar temel alınarak oluşturulmuştur.

Çizelge 3.1 Veri Setinde Kullanılan Demografik ve Tutum Değişkenleri

Nitelikler	**Değerler**
Cinsiyet	1. Erkek
	2. Kadın
Yaş	Tam sayılardaki değerler
Seviye	1, 2, 3, 4
Çevre eğitimi dersi aldınız mı?	1-Evet, 2-Hayır
	2. Hayır
Küresel ısınma ve iklim dersi aldınız mı?	1. Evet
	2. Hayır
Çevre okuryazarlığı dersi aldınız mı?	1. Evet
	2. Hayır
Gelir durumu	1. 0-7500₺
	2. 7501₺-10000₺
	3. 10001₺-15000₺
	4. 15001₺-daha
Okul öncesi eğitime gittiniz mi?	1. Evet
	2. Hayır
Nerede yaşıyorsun	1. Kent
	2. Semt
	3. Şehir
Kardeş sayınız	0, 1, 2, 3, 4, 5 daha
S1- Havayı en az düzeyde kirletecek araçlar icat etme fikri beni heyecanlandırıyor.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S2- Doğaya salınan zararlı gazların çevrenin taşıma kapasitesini aşabileceği düşüncesi beni korkutuyor.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S3- Atmosferdeki artan kirliliğin küresel iklim değişikliğine sebep olduğunu bilmek beni korkutuyor.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S4- Gelecekteki su kıtlığının nedenlerinden birinin insan nüfusunun artması olmasından endişe ediyorum.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S5- Gelecek nesiller için suyun devamlılığını sağlamak amacıyla daha az kirletici tarım kimyasalları, endüstriyel ürünler ve ev temizleyicileri kullanmayı tercih ediyorum.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S6- Ürünlerde biriken kimyasalların gıda zincirindeki diğer halkalara olumsuz etkisi beni rahatsız ediyor.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
	1) Kesinlikle katılmıyorum

S7- Dünyanın diğer bölgelerinde meydana gelen toprak erozyonu beni ilgilendirmiyor.	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S8- Gelecek için yenilenebilir enerji kaynaklarına yatırım yapmak gereksizdir.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S9- Enerji kaynaklarının sürdürülebilirliğini sağlamak için bu kaynakları dikkatli kullanma düşüncesi gereksizdir.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S10- Fosil enerji kaynaklarının bir gün tükenebileceği düşüncesiyle dikkatli kullanılması gereksizdir.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S11- Hızla tükettiğimiz kaynakların doğa tarafından yenilenemeyeceği düşüncesi beni endişelendiriyor.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S12- Sürdürülebilir bir çevre için geri dönüşüm reklamları görmek beni mutlu ediyor.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S13- Aldığım ürünlerin ambalajında geri dönüşüm sembolü olup olmadığına dikkat etmiyorum.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S14- Okullarda geri dönüşüm konusunda eğitim verilmesinin gerekli olduğunu düşünüyorum.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S15- Ben depozitolu şişelerde gelen ürünleri kullanmayı tercih ediyorum.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S16- Plastik poşet yerine bez çanta, file çanta veya kağıt torba kullanmayı tercih etmiyorum.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S17- Aldığım ürünlerin tek kullanımlık değil, tekrar kullanılabilir olmasına dikkat ediyorum.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum

	5) Kesinlikle katılıyorum
S18- Çevremizde yeterli sayıda geri dönüşüm kutusu görememek üzücü.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S19- Tüketimin hızla artmasının çevresel sürdürülebilirliğin önünde önemli bir engel teşkil etmesi beni korkutuyor.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S20- Doğanın bize sağlayabileceğinden fazlasını tükettiğimizde geleceğin etkileneceğini düşünmek gereksizdir.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S21- Sürdürülebilirlik için tüketim alışkanlıkları konulu seminerlere katılmaktan mutluluk duyarım.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S22- İnsan nüfusu arttıkça kaynakların tükeneceğini düşünmek gereksizdir.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S23- İnsan nüfusunun artışının doğal dengenin sürdürülebilirliğine engel teşkil edeceği kaygısını taşımıyorum.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S24- Sürdürülebilirlik hakkında öğrendiklerimi aileme ve yakın arkadaşlarıma anlatmak zaman kaybıdır.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S25- Çocuklarımıza güzel bir çevre bırakmak için sürdürülebilirliğin bir yaşam biçimi haline gelmesi fikrini beğeniyorum.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S26- İnsanların hammadde ihtiyaçlarını geri dönüşüm uygulamalarıyla karşılamaları ve doğaya olan baskılarını azaltmaları beni mutlu ediyor.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
S27- Doğal kaynakların sonsuz olmadığı konusunda insanları geri dönüşüm kampanyaları aracılığıyla bilgilendirmenin önemli olduğunu düşünüyorum.	1) Kesinlikle katılmıyorum
	2) Katılmıyorum
	3) Hiçbir fikrim yok
	4) Katılıyorum
	5) Kesinlikle katılıyorum
Sınıf	0. Hayır
	1. Evet

3.2 VERİ ÖN İŞLEME

Veri ön işleme, ham verinin analiz ve modelleme süreçlerine uygun hale getirilmesi için gerçekleştirilen temel bir adımdır. Bu aşama, veri setinde yer alan eksikliklerin, tutarsızlıkların ve ölçüm farklılıklarının giderilmesini sağlayarak modelin güvenilirliğini artırmakta ve daha doğru, tutarlı sonuçlar elde edilmesine katkıda bulunmaktadır. Çalışmada, veri setinde eksik veya hatalı verilerin bulunup bulunmadığı kontrol edilmiş ve herhangi bir eksik ya da tutarsız veri tespit edilmemiştir. Veri kalitesinin sağlanmasının ardından, değişkenler arasındaki ölçüm birimi farklılıklarını ortadan kaldırmak amacıyla ölçeklendirme işlemi gerçekleştirilmiştir. Bu kapsamda, verilerin ortak bir ölçeğe getirilmesi ve modelleme sürecinde algoritmaların daha sağlıklı çalışmasını sağlamak için StandardScaler yöntemi uygulanmıştır. StandardScaler yöntemi, her bir değişkenin ortalamasını sıfıra ve standart sapmasını bire eşitleyerek verileri standart bir dağılıma dönüştürmektedir. Bu işlem aşağıdaki formül ile ifade edilmektedir:

$$z = \frac{x-\mu}{\sigma} \tag{3.1}$$

Bu formülde z standartlaştırılmış değeri, x orijinal gözlem değerini, μ ilgili değişkenin ortalamasını ve σ ise ilgili değişkenin standart sapmasını ifade etmektedir. Her bir gözlem değerinden değişkenin ortalaması çıkarılmakta ve elde edilen fark, değişkenin standart sapmasına bölünerek standartlaştırılmış değer hesaplanmaktadır. Bu işlem sonucunda, veri setindeki tüm değişkenler ortalaması sıfır ve standart sapması bir olacak şekilde yeniden ölçeklendirilmekte, böylece farklı ölçeklere sahip değişkenler arasında denge sağlanmakta ve modelleme sürecinin daha sağlıklı ilerlemesi sağlanmaktadır (Marques vd. 2025).

3.3 ÖZELLİK SEÇİMİ

Özellik seçimi, bir veri seti içerisinden problemin çözümüne en fazla katkı sağlayacak alt kümenin belirlenmesi sürecidir. Bu süreç, sınıflandırma, regresyon veya diğer MÖ problemlerinde modelin başarımını artırmak amacıyla, gereksiz veya etkisiz özelliklerin elenmesini ve dolayısıyla özellik sayısının azaltılmasını hedeflemektedir. Özellik seçimi sayesinde hem modelin öğrenme süreci hızlanmakta hem de aşırı öğrenme (overfitting) riski azaltılarak daha genellenebilir sonuçlar elde edilebilmektedir (Efeoğlu 2022). Özellik seçimi için kullanılan yöntemler genel olarak üç ana grupta sınıflandırılmaktadır: filtreleme yöntemleri

(filter methods), sarmal yöntemler (wrapper methods) ve gömülü yöntemler (embedded methods).

Filtreleme yöntemleri, veri madenciliği ve MÖ uygulamalarında kullanılan en eski ve temel özellik seçim yaklaşımlarındandır. Bu yöntemlerde herhangi bir sınıflandırıcı model kullanılmadan, yalnızca istatistiksel ölçütlere dayalı olarak özelliklerin önem derecesi hesaplanmaktadır. Özellikler, model eğitimi öncesinde bağımsız olarak değerlendirilip seçildiği için, büyük veri setlerinde hızlı ve düşük hesaplama maliyetine sahip çözümler sunmaktadır. Ancak, filtreleme yöntemleri seçilen özelliklerin modelin nihai başarımı üzerindeki etkisini doğrudan dikkate almamaktadır.

Sarmal yöntemlerde, bir sınıflandırıcı veya regresyon algoritması, özellik seçimi sürecinin doğrudan bir parçası olarak kullanılmaktadır. Bu yöntemde farklı özellik alt kümeleri oluşturularak her bir alt küme üzerinde model eğitilmekte ve performans ölçülmetkedir. En iyi tahmin performansını sağlayan özellikler belirlenmektedir. Ancak, özellikle büyük veri setlerinde hesaplama maliyeti yüksek olabilmektedir.

Gömülü yöntemler, sınıflandırma veya regresyon algoritmalarının eğitim süreci ile özellik seçimini eşzamanlı olarak gerçekleştiren bir yaklaşımdır (Budak 2018). Model eğitimi sırasında önemli özellikler doğrudan belirlenirken, gereksiz olanlar elenmektedir. Bu yöntemler, filtreleme ve sarmal yöntemlerin avantajlarını birleştirerek hem daha düşük hesaplama maliyeti sağlar hem de model başarımı üzerinde doğrudan etkili olmaktadır.

3.3.1 Ki Kare ile Özellik Seçimi

Ki Kare yöntemi, gözlenen ve beklenen frekanslar arasındaki farkı istatistiksel olarak değerlendirerek bir değişkenin hedef değişkenle ilişkili olup olmadığını belirlemeye yarayan bir özellik seçim tekniğidir. Özellikle kategorik veri içeren çalışmalarda etkili sonuçlar sunan bu yöntem, her bir bağımsız değişkenin hedef değişken üzerindeki etkisini ölçmektedir. Hesaplanan Ki Kare değerinin sıfıra yakın olması, ilgili değişken ile hedef değişken arasında anlamlı bir ilişki bulunmadığını gösterirken, yüksek bir Ki Kare değeri değişkenin hedef değişken üzerinde etkili olduğunu ve modelde tutulması gerektiğini ifade etmektedir. Ki Kare değeri aşağıdaki formül ile hesaplanmaktadır:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \quad (3.2)$$

Burada Oi gözlenen frekansı, Ei beklenen frekansı ve $\chi 2$ ise Ki Kare test istatistiğini temsil etmektedir. Bu sayede, veri setinde yer alan anlamsız değişkenler elenerek modelin daha verimli çalışması sağlanmaktadır. Ki Kare yöntemi ile gerçekleştirilen özellik seçimi, modelin doğruluğunu artırmakla kalmayıp, eğitim süresinin kısalmasına ve aşırı öğrenme riskinin azaltılmasına da katkı sağlamaktadır (Aksu 2023, Budak 2018, Emekli 2024).

3.3.2 Ekstra Ağaçlar ile Özellik Seçimi

Ekstra Ağaçlar (EA) algoritması, veri kümesindeki önemli değişkenleri belirlemek amacıyla kullanılan etkili bir özellik seçim yöntemidir. Bu yöntem, çok sayıda rastgele KA oluşturarak çalışmakta ve her ağacın eğitim sürecinde veri kümesindeki özellikler düğümlerle, belirli sınıflara ait örnekler ise yapraklarla temsil edilmektedir. Ağaçların oluşturulmasında bilgi kazancı veya Gini saflığı gibi ölçütler kullanılarak hedef değişkenle en yüksek ilişkiye sahip olan özellikler öncelikli olarak bölünmektedir. Her bir KA tamamlandıktan sonra, iç düğümler kökten yapraklara doğru sıralanarak, hangi özelliklerin karar verme sürecinde ne ölçüde etkili olduğu hesaplanmaktadır. Özelliklerin önemi, düğümde sağlanan saflık artışı ve o düğümün veri kümesindeki örnekler üzerindeki etkisi dikkate alınarak belirlenmektedir. Bu kapsamda, bir özelliğin önem skoru aşağıdaki formül ile hesaplanmaktadır:

$$\text{Özellik Önemi } (f) = \sum_{t \in T(f)} \frac{N_T}{N} * \Delta i(t) \quad (3.3)$$

Bu formülde, $\text{Özellik Önemi } (f)$ ilgili özelliğin önem skorunu, $T(f)$ özelliğin kullanıldığı tüm düğümler kümesini, Nt düğüm t'ye ulaşan örnek sayısını, N toplam örnek sayısını ve $\Delta i(t)$ düğümde sağlanan saflık artışını ifade etmektedir. Böylece, veri setindeki her bir özelliğin hedef değişken üzerindeki katkısı sistematik bir şekilde sıralanmakta ve en etkili özellikler öne çıkarılmaktadır. EA algoritması, özellikle çok boyutlu veri setlerinde yüksek verimlilik ve istikrarlı sonuçlar sağlaması nedeniyle özellik seçimi sürecinde sıklıkla tercih edilen yöntemlerden biridir (Yılmaz ve Kuvat 2023).

3.3.3 ANOVA F-Testi ile Özellik Seçimi

ANOVA, grup ortalamaları arasındaki farkları istatistiksel olarak analiz etmek için kullanılan bir yöntem olup, özellik seçimi sürecinde çıktı değişkeni ile girdi değişkeni arasındaki ilişkiyi inceleyen filtreleme yöntemlerinden biri olarak değerlendirilmektedir. Özellikle girdi veya çıktı değişkenlerinden birinin kategorik veri olduğu durumlarda tercih edilen ANOVA, farklı kategorilere ait grupların ortalamalarının birbirinden istatistiksel olarak anlamlı şekilde farklılaşıp farklılaşmadığını belirlemeyi hedeflemektedir. Yöntem, bağımsız değişkenin oluşturduğu grupların bağımlı değişken üzerindeki etkisini değerlendirerek, gruplar arasında anlamlı bir fark bulunmaması durumunda ilgili değişkenin modelden çıkarılmasına olanak tanımaktadır. Böylece yalnızca anlamlı katkı sağlayan değişkenlerin modele dahil edilmesi sağlanarak, modelin doğruluğu ve verimliliği artırılmaktadır. ANOVA, gruplar arasındaki varyansı inceleyerek hangi değişkenlerin çıktı değişkeni üzerinde daha etkili olduğunu belirlemekte ve değişkenlerin önem derecelerine göre sıralanmasına yardımcı olmaktadır. Bu yönüyle, çoklu kategorik veri içeren veri setlerinde ANOVA tabanlı özellik seçimi etkili ve güvenilir bir yöntem olarak öne çıkmaktadır (Yıldırım 2023). Bu yöntemde, gruplar arasındaki farklılığı değerlendirmek için önce gruplar arası varyans ve gruplar içi varyans hesaplanmakta, ardından bu iki varyansın oranı alınarak F-istatistiği elde edilmektedir. İlgili hesaplamalar aşağıdaki formüllerle ifade edilmektedir:

$$Gruplar\ Arası\ Varyans = \frac{\sum_{i=1}^{k} n_i (YM_i - YM)^2}{k-1} \tag{3.4}$$

$$Gruplar\ \text{İçi}\ Varyans = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - YM_i)^2}{N-k} \tag{3.5}$$

$$F = \frac{Gruplar\ Arası\ Varyans}{Gruplar\ \text{İçi}\ Varyans} \tag{3.6}$$

Burada $ni\ i.$ gruptaki gözlem sayısını, $YMi\ i.$ grubun ortalamasını, YM veri setinin genel ortalamasını, N toplam gözlem sayısını ve k grup sayısını temsil etmektedir. Hesaplanan F-istatistiğinin yüksek olması, gruplar arasında anlamlı bir fark bulunduğunu, düşük olması ise gruplar arasında anlamlı bir fark bulunmadığını göstermektedir. Bu değerlendirme sonucunda yalnızca hedef değişken ile anlamlı ilişkiye sahip özellikler modelde tutulmakta ve bu sayede modelin başarım düzeyi artırılmaktadır (Al-Shamaa 2020).

3.3.4 LASSO ile Özellik Seçimi

LASSO yöntemi, doğruluk oranını artırmak ve modelin daha sade bir yapıya kavuşmasını sağlamak amacıyla regresyon analizinde kullanılan etkili bir özellik seçim yöntemidir. Bu yöntem, regresyon modelinde yer alan katsayılar üzerine L1 normuna dayalı bir ceza uygulayarak, bazı katsayıların sıfırlanmasına ve dolayısıyla ilgili değişkenlerin modelden çıkarılmasına imkân tanımaktadır. Böylece, yalnızca hedef değişken ile anlamlı ilişkisi bulunan değişkenler modelde tutulmaktadır. LASSO, tüm regresyon katsayılarının mutlak değerlerinin toplamının belirli bir sınır değeri aşmamasını şart koşarak, tahmin hatasını minimize etmeyi amaçlamaktadır. Bu yapı aşağıdaki optimizasyon formülü ile ifade edilmektedir:

$$(\alpha, \beta) = argmin\left(\sum_{i=1}^{N}\left(y_i - \alpha - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2\right) \qquad s.t. \quad \sum_{j=1}^{p}|\beta_j| \leq t \tag{3.7}$$

Bu formülde y_i bağımlı değişkeni, x_{ij} bağımsız değişkenleri, α sabit terimi, β_j regresyon katsayılarını, N toplam gözlem sayısını, p bağımsız değişken sayısını ve t cezalandırma parametresini ifade etmektedir. Cezalandırma parametresi t değeri küçüldükçe, daha fazla regresyon katsayısının sıfırlanması sağlanmakta ve böylece modelde yalnızca anlamlı değişkenler kalmaktadır. Bu mekanizma, modelin hem aşırı öğrenmeye karşı daha dayanıklı hale gelmesini hem de yorumlanabilirliğinin artmasını sağlamaktadır. Özellikle yüksek boyutlu veri setlerinde, LASSO yöntemi etkili bir özellik seçimi aracı olarak öne çıkmakta ve model başarımını önemli ölçüde artırmaktadır (Chen vd. 2024).

3.3.5 Karşılıklı Bilgi ile Özellik Seçimi

Karşılıklı Bilgi, (Mutual Information, MI) yöntemi, iki değişken arasındaki bağımlılık düzeyini hem doğrusal hem de doğrusal olmayan ilişkileri dikkate alarak ölçen bir istatistiksel tekniktir. Özellik seçimi sürecinde, bir değişkenin diğer bir değişken hakkında ne kadar bilgi sağladığını nicel olarak değerlendirmek amacıyla kullanılmaktadır. Eğer iki değişken tamamen bağımsızsa, karşılıklı bilgi değeri sıfır olurken; karşılıklı bilginin yüksek olması, değişkenler arasında güçlü bir bağımlılık ilişkisinin bulunduğunu göstermektedir. Bu yöntem, yalnızca doğrusal ilişkileri değil, karmaşık ve doğrusal olmayan bağımlılıkları da tespit etme yeteneğine sahiptir (Emekli 2024). Özellik seçimi sürecinde, çıktı değişkeni ile yüksek karşılıklı bilgi değerine sahip öznitelikler belirlenerek modele dahil edilmekte; bilgi taşıma kapasitesi düşük olan veya benzer

özellikler ise elenerek modelin daha verimli ve daha sade bir yapıya ulaşması sağlanmaktadır. İki değişken arasındaki karşılıklı bilgi değeri aşağıdaki formül ile hesaplanmaktadır:

$$I(X,Y) = \sum_{y\in Y}\sum_{x\in X} p(x,y) log\left(\frac{p(x,y)}{p_1(x)p_2(y)}\right) \quad (3.8)$$

Bu formülde, $p(x,y)$ değişkenlerin ortak olasılık dağılımını, $p_{1(x)}$ X değişkeninin marjinal olasılığını ve $p_{2(y)}$ Y değişkeninin marjinal olasılığını ifade etmektedir. Karşılıklı bilgi ölçümü sayesinde, hedef değişkenle anlamlı bilgi paylaşımı yapan öznitelikler tespit edilerek modelin başarımının artırılması sağlanmaktadır (Biricik 2011).

3.3.6 Temel Bileşen Analizi ile Özellik Seçimi

TBA yüksek korelasyonlu çok değişkenli verilerin varyansı maksimum düzeyde yansıtan yeni bir koordinat sistemine doğrusal olarak dönüştürülmesini sağlayan bir boyut indirgeme yöntemidir. Bu yöntem, veri setinin boyutunu azaltırken, verideki en fazla bilgi taşıyan yönlerin korunmasını amaçlamaktadır. TBA sürecinde, orijinal değişkenler arasındaki korelasyonlar dikkate alınarak, verinin varyansını en yüksek düzeye çıkaracak yeni bileşenler oluşturulmaktadır. Bu yeni bileşenler, orijinal değişkenlerin doğrusal kombinasyonları ile elde edilmekte olup, veri setinin daha sade ve anlaşılır bir yapıya kavuşturulmasına katkı sağlamaktadır. Analiz süreci, veri setinin kovaryans matrisinin hesaplanmasıyla başlamakta ve ardından bu matrise ait özvektörler ile özdeğerler elde edilmektedir. TBA sürecinde, veri setinin kovaryans matrisi hesaplanarak değişkenler arasındaki ilişki yapısı ortaya konmakta, ardından bu matris üzerinden elde edilen özvektörler ve özdeğerler yardımıyla verinin en fazla varyans taşıyan yönleri belirlenerek boyut indirgeme işlemi gerçekleştirilmektedir. TBA'da, özellik seçimi amacıyla kullanılan temel adım, kovaryans matrisinin özvektörlerinin ve özdeğerlerinin elde edilmesiyle gerçekleştirilir. Veri setinin kovaryans matrisi aşağıdaki şekilde hesaplanmaktadır:

$$Cov\ (X) = \frac{1}{n-1}(X-\bar{X})^T(X-\bar{X}) \quad (3.9)$$

Bu formülde X veri matrisini, $\bar{X}$ ise değişkenlerin ortalamalarını ifade etmektedir. Özellik seçimi sürecinde, elde edilen özdeğerler sıralanarak en yüksek özdeğere sahip temel bileşenler seçilmekte, böylece veri setinin taşıdığı bilgi yoğunluğu korunmakta ve bilgi katkısı düşük olan

bileşenler elenmektedir. Bu yaklaşım sayesinde, verinin en önemli yönleri temsil edilmekte ve modelin daha az gürültü ile daha etkili çalışması sağlanmaktadır (Biricik 2011).

3.3.7 Varyans Eşiği ile Özellik Seçimi

Yüksek boyutlu veri setlerinde bilgi içeriği düşük değişkenlerin elenmesi ve modelin daha yalın bir yapıya kavuşturulması amacıyla varyans eşikleme yöntemi uygulanmaktadır. Bu yöntem, her bir değişkenin varyans değeri temel alınarak, varyansı düşük olan değişkenlerin veri setinden çıkarılmasına dayanmaktadır. Özellikle tüm gözlemlerde sabit kalan veya çok sınırlı değişkenlik gösteren değişkenler, sınıflar arasında ayırt edici bilgi taşımadığı için modelin öğrenme sürecine anlamlı bir katkı sunmamaktadır. Bu tür değişkenlerin veri setinden çıkarılması, hem modelin hesaplama yükünü azaltmakta hem de öğrenme performansını artırmaktadır (Yıldırım 2023). Bir değişkenin varyansı, aşağıdaki formül ile hesaplanmaktadır:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 \tag{3.10}$$

Burada x_i değişkenin i. gözlemini, μ değişkenin ortalamasını ve N toplam gözlem sayısını temsil etmektedir. Sürekli değişkenler için kullanılan bu standart varyans hesaplamasına ek olarak, yalnızca 0 ve 1 gibi ikili (binary) değerlere sahip değişkenler için varyans aşağıdaki formül ile ifade edilmektedir:

$$T_h = \alpha(1 - \alpha) \tag{3.11}$$

Burada α, ilgili değişkenin 1 değerini alma olasılığını göstermektedir. Bu tür ikili değişkenlerde varyans değeri, bu olasılığa bağlı olarak $\alpha(1 - \alpha)$ ifadesi üzerinden hesaplanmakta ve değişkenin dağılımı bu doğrultuda değerlendirilmektedir. Varyans eşikleme yöntemi, sınıf etiketlerini dikkate almayan denetimsiz bir özellik seçimi tekniği olup, model eğitimi öncesinde ön eleme işlemi olarak uygulanmaktadır. Bu yöntem sayesinde yalnızca yeterli varyansa ve dolayısıyla bilgi taşıma potansiyeline sahip değişkenler veri setinde tutulmakta, gereksiz değişkenlerin elenmesi ile modelin genelleme performansı artırılmaktadır (Walpole vd. 2012).

3.3.8 Rastgele Orman ile Özellik Seçimi

RO algoritması, çok sayıda KA'dan oluşan bir topluluk yöntemine dayanmakta olup, her bir ağaç, rastgele seçilen veri örnekleri ve değişkenler üzerinden bağımsız şekilde inşa edilmektedir. Bu yapı sayesinde model, hem değişkenler arasındaki karmaşık ilişkileri öğrenebilmekte hem de aşırı öğrenme riskini azaltarak genelleme performansını artırabilmektedir.

Özellik seçim sürecinde, değişkenlerin sınıflandırma başarısına katkı düzeyleri, RO algoritması tarafından hesaplanan değişken önem puanları aracılığıyla değerlendirilmektedir. Bu önem puanları, KA'da bir değişkenin kullanıldığı bölünme noktalarında sağladığı safsızlık (impurity) azalımı üzerinden hesaplanmaktadır. Safsızlık ölçütü olarak Gini indeksine dayalı safsızlık değeri kullanılmakta ve her bir düğüm için aşağıdaki formül ile hesaplanmaktadır:

$$Gini(v) = \sum_{k=1}^{K} p_k(1 - p_k) \tag{3.12}$$

Burada p_k, düğüm v üzerindeki gözlemlerin k sınıfına ait olma oranını ifade etmektedir. Bir değişkenin önem düzeyi, bu safsızlık değerinde sağladığı azalma ile belirlenmekte olup, söz konusu kazanç değeri aşağıdaki şekilde hesaplanmaktadır:

$$Gini\big(X_j, v\big) = Gini\big(X_j, v\big) - \sum_{i\in\{L,R\}} \omega^i \, Gini(X_j, v^i) \tag{3.13}$$

Bu ifadede ω_i, bölünme sonucu oluşan alt düğümlerdeki örneklerin oransal ağırlığını; $Gini(Xj, vi)$ ise bu alt düğümlerin safsızlık düzeyini göstermektedir. Böylece, her bir değişkenin tüm ağaçlar boyunca sağladığı toplam Gini kazancı toplanarak o değişkene ait önem puanı elde edilmektedir. Elde edilen önem skorları doğrultusunda, belirlenen eşik değerinin altında kalan değişkenler analiz dışı bırakılmakta ve yalnızca sınıflar arası ayrımı güçlendiren öznitelikler veri setinde tutulmaktadır. Bu yöntem, modelin hem daha düşük hesaplama maliyetiyle eğitilmesini hem de sınıflandırma performansının artırılmasını sağlamaktadır (Iranzad ve Liu 2024, Liu vd. 2024, Şahin 2017).

3.3.9 Rastgele Orman Modelini Kullanarak Özyinelemeli Özellik Eliminasyon İle Özellik Seçimi

RFE bir MÖ modelinin oluşturduğu özellik önem skorlarına dayanarak çalışan, denetimli bir özellik seçimi yöntemidir. Yöntem, tüm özelliklerin kullanıldığı bir başlangıç modeli ile süreci başlatmakta ve her yinelemede modelin karar mekanizmasında en düşük öneme sahip özelliği sistematik olarak elemektedir. RO modeli, RFE sürecinde temel öğrenici olarak kullanıldığında, KA'ları üzerinden değişken önem derecelerini hesaplayarak her adımda hangi özelliklerin çıkarılması gerektiğini belirlemektedir. Özelliklerin çıkarılması işlemi sonrası model yeniden eğitilmekte ve kalan özelliklerin önem dereceleri güncellenmektedir. Bu iteratif süreç, belirlenen son özellik sayısına ulaşılana kadar devam etmektedir. RFE ile özellikler arasındaki ilişkiler dikkate alınarak, modelin genelleme kapasitesini artıracak en bilgilendirici özellik alt kümesi elde edilmektedir. Böylece modelin aşırı öğrenme riski azaltılırken sınıflandırma performansı da optimize edilmektedir. Bu yöntem, özellikle çok boyutlu veri kümelerinde daha güvenilir ve daha dengeli bir model performansı elde etmek amacıyla yaygın olarak tercih edilmektedir. Ayrıca, RO algoritmasının değişken önemini güvenilir bir şekilde ölçebilme kapasitesi, RFE sürecinde daha kararlı ve doğru bir özellik seçimi yapılmasına olanak sağlamaktadır (Auliyani vd. 2024, Bilgin 2022).

3.3.10 Lojistik Regresyon Modelini Kullanarak Özyinelemeli Özellik Eliminasyon İle Özellik Seçimi

Bu yöntemde RFE süreci, temel model olarak LR kullanılarak yürütülmüştür. Başlangıçta tüm özellikler modele dahil edilmekte; ardından, modelin sınıflandırma performansına en az katkı sağlayan değişkenler yinelemeli olarak elenmektedir. Her çıkarım sonrası model yeniden eğitilmekte ve önem skorları güncellenmektedir. Bu süreç, belirlenen hedef özellik sayısına ulaşıncaya kadar sürdürülmüştür. Sonuçta elde edilen alt özellik kümesiyle oluşturulan LR modeli, daha sade, dengeli ve optimize bir yapı sunarak sınıflandırma başarımını artırmıştır (Lazrek vd. 2024, Zhang vd. 2024).

3.4 VERİ SETİNİN AYRILMASI ve ÇAPRAZ DOĞRULAMA YÖNTEMİ İLE MODEL PERFORMANSININ DEĞERLENDİRİLMESİ

Gerçekleştirilen çalışmada özellik seçimi sürecinin tamamlanmasının ardından, modelleme çalışmalarında kullanılacak veri seti eğitim ve test alt kümelerine ayrılmıştır. Bu kapsamda, veri setinin %80'i eğitim verisi, %20'si ise test verisi olarak belirlenmiştir. Eğitim verisi, modellerin oluşturulması ve değerlendirilmesi sürecinde kullanılırken, test verisi ise eğitim aşamasında kullanılmadan ayrılmış ve yalnızca nihai modelin performansını bağımsız olarak ölçmek amacıyla kullanılmıştır. Bu ayrım, modelin daha önce görmediği veri üzerinde gösterdiği başarının objektif şekilde değerlendirilmesine olanak sağlamıştır.

Eğitim verisi üzerinde, modelin farklı veri alt kümeleri üzerindeki başarımını değerlendirmek ve genellenebilirliğini artırmak amacıyla 5 katlı çapraz doğrulama yöntemi uygulanmıştır. Bu yöntemde, eğitim verisi eşit büyüklükte beş alt kümeye bölünmüş; her yinelemede dört alt küme modelin eğitilmesi için, kalan bir alt küme ise doğrulama amacıyla kullanılmıştır. Her bir alt küme, bir kez doğrulama verisi olarak kullanılacak şekilde işlem tekrarlanmıştır. Böylece modelin tüm eğitim verisi üzerinde dengeli bir şekilde test edilmesi sağlanmıştır.

Çapraz doğrulama süreci sonunda her kat için elde edilen doğrulama sonuçları birleştirilmiş ve modelin eğitim verisi üzerindeki genel başarımı hesaplanmıştır. Ayrıca, çapraz doğrulama süreci tamamlandıktan sonra, en iyi performansı gösteren model daha önce ayrılan %20'lik test verisi üzerinde de değerlendirilmiştir. Böylece hem eğitim verisindeki çapraz doğrulama sonuçları hem de bağımsız test verisi üzerindeki sonuçlar karşılaştırmalı olarak analiz edilmiş ve modelin genel sınıflandırma performansı iki aşamada doğrulanmıştır.

3.5 MAKİNE ÖĞRENME ALGORİTMALARI

Bireylerin çevresel tutumlarının tahmini üzerine gerçekleştirilen bu çalışmada farklı MÖ algoritmalarından yararlanılmıştır. Kullanılan algoritmalar alt başlıklarda detaylandırılmıştır.

3.5.1 Lojistik Regresyon

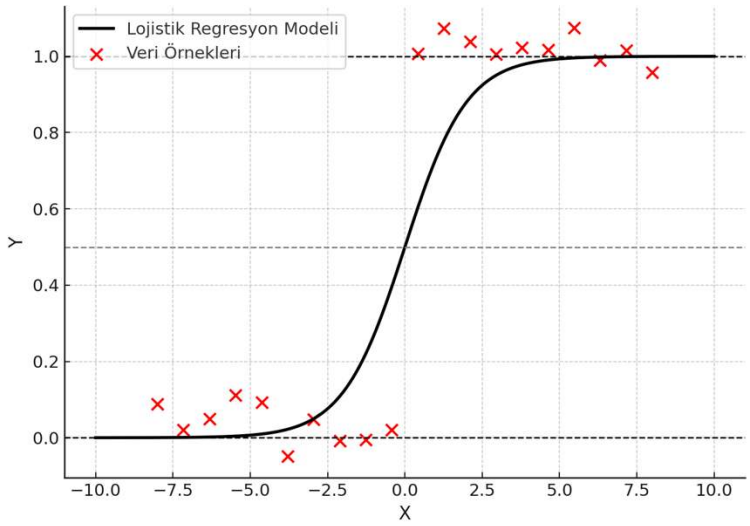
LR, istatistiksel MÖ alanında, özellikle ikili sınıflandırma problemlerinde yaygın olarak kullanılan bir veri madenciliği yöntemidir. Bu yöntemde, bağımsız değişkenlerin doğrusal bir

kombinasyonu elde edilmekte ve bu kombinasyon, sigmoid (lojistik) fonksiyon aracılığıyla normalize edilerek model çıktıları üretilmektedir (Gürgen ve Serttaş 2023, Uzun vd. 2019). Lojistik regresyonun matematiksel ifadesi aşağıda verilmiştir:

$$p(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x_1+\cdots\ldots+\beta_n x_n)}} \tag{3.14}$$

Burada $p(x)$, bir gözlemin 1 sınıfına ait olma olasılığını; β_0 sabit terimini ve $\beta_1, \beta_2 \ldots, \beta_n$ regresyon katsayıları ile $x_1, x_2, \ldots, x_n$ bağımsız değişkenlerini temsil etmektedir. Model parametreleri, en büyük olabilirlik (maximum likelihood) yöntemi kullanılarak tahmin edilmekte olup, model çıktıları daima 0 ile 1 aralığında kalmaktadır. Bu yapı, LR'nun sınıflandırma problemlerinde olasılık temelli yorum yapılmasına olanak tanımaktadır.

LR'da tahmin edilen olasılıkların sınıflandırılması için bir eşik değeri kullanılmakta olup, genellikle bu değer 0.5 olarak belirlenmektedir. Tahmin edilen olasılıklar eşik değerinin üzerinde ise 1, altında ise 0 olarak etiketlenmektedir (Gürgen ve Serttaş 2023). Bu yaklaşım, doğrusal karar sınırları üretirken modelin çıktılarının olasılık temelli yorumlanmasına olanak sağlamaktadır. Modelin karakteristik sigmoid eğrisi Şekil 3.1'de gösterilmektedir.



Şekil 3.1. LR Modeli ve Veri Örnekleri.

LR, basitliği, yorumlanabilirliği ve düşük hesaplama maliyeti gibi avantajlarıyla öne çıkmakta; tıbbi tanı, finansal risk analizi ve sosyal bilimler gibi alanlarda yaygın uygulama alanı

bulmaktadır. Ancak doğrusal sınırları nedeniyle karmaşık ilişkileri modellemede sınırlı kalabilmekte ve aykırı değerlere karşı duyarlılık gösterebilmektedir.

3.5.2 Rastgele Orman

RO, birden çok KA'nın bir araya gelmesiyle oluşturulan bir topluluk (ensemble) öğrenme yöntemidir. Bu algoritma, 2001 yılında Leo Breiman tarafından geliştirilmiş olup hem sınıflandırma hem de regresyon problemlerinde etkili sonuçlar vermektedir (Breiman 2001). Algoritmanın temel mantığı, çok sayıda zayıf öğreniciyi (KA'larını) birleştirerek güçlü bir tahmin modeli oluşturmaktır (Dietterich 2000). RO algoritması iki temel rastgelelik prensibine dayanmaktadır. Birincisi, her bir KA'nın eğitimi için orijinal veri setinden bootstrap yöntemiyle rastgele örneklemlerin seçilmesidir. İkincisi ise her düğüm bölünmesinde rastgele öznitelik alt kümelerinin kullanılmasıdır (Breiman 2001). Bu iki mekanizma, modelin genelleme yeteneğini artırırken aşırı öğrenme riskini de minimize etmektedir (Genuer vd. 2010).

Algoritmanın eğitim sürecinde, öncelikle veri setinden yerine koymalı örnekleme yöntemiyle çok sayıda alt örneklem oluşturulmaktadır. Daha sonra her bir alt örneklem için bir KA eğitilmektedir. KA'nın oluşturulması sırasında, her düğüm bölünmesinde rastgele seçilen öznitelikler arasından en iyi bölünme noktası belirlenmektedir. Bu işlem, ağaçların çeşitliliğini artırarak modelin kararlılığını sağlamaktadır. Tahmin aşamasında, sınıflandırma problemleri için tüm ağaçların tahminlerinin çoğunluk oyu alınırken, regresyon problemlerinde ağaç çıktılarının ortalaması kullanılmaktadır (Hastie vd. 2009). Bu yaklaşım, tek bir karar ağacının aksine daha dengeli ve güvenilir sonuçlar üretmektedir (Yapıcı ve Arslan 2024). RO algoritmasının önemli avantajları arasında yüksek tahmin doğruluğu, aşırı öğrenmeye karşı direnç, eksik verilerle başa çıkabilme yeteneği ve öznitelik önem sıralaması yapabilme özelliği yer almaktadır. Ancak algoritma, çok sayıda ağaç içerdiğinden yüksek hesaplama maliyeti gerektirebilir ve sonuçların yorumlanması nispeten daha zordur (Akgün ve Barlık 2023).

3.5.3 Naive Bayes

NB, olasılık temelli bir MÖ algoritmasıdır ve özellikle sınıflandırma problemlerinde yaygın olarak kullanılmaktadır. Algoritmanın temelini, Bayes Teoremi ile birlikte özelliklerin birbirinden bağımsız olduğu varsayımı oluşturmaktadır. Bu bağımsızlık varsayımı nedeniyle

"Naive" (saf, aşırı basitleştirilmiş) olarak adlandırılmaktadır (Rish 2001, Russell 2016, Yudhana vd. 2023, Zhang 2004).

Bayes teoremi, bir olayın gerçekleşme olasılığını, gözlemlenen verilere dayanarak güncellemeye olanak tanımakta ve matematiksel olarak aşağıdaki şekilde ifade edilmektedir:

$$P(C|X) = \frac{P(X|C)*P(C)}{P(X)} \quad (3.15)$$

Burada $P(C \mid X)$, X gözlemleri verildiğinde C sınıfına ait olma olasılığına (posterior olasılık), $P(X \mid C)$ C sınıfı altında X gözlemlerinin olasılığına (likelihood), $P(C)$, C sınıfının gerçekleşme olasılığına (prior); $P(X)$ ise X gözlemlerinin gerçekleşme olasılığına (evidence) karşılık gelmektedir (Ayan ve Bilgin 2024, Mitchell ve Mitchell 1997).

NB algoritması, özellikle büyük boyutlu veri kümelerinde ve gerçek zamanlı sınıflandırma uygulamalarında tercih edilmektedir. Çünkü modelin eğitilmesi ve çıkarım yapılması hızlıdır. Ayrıca, veri setinde özellikler arasındaki bağımlılıkların minimum olduğu durumlarda oldukça yüksek doğruluk oranları elde edilebilmektedir (Zhang 2004).

NB algoritmasının farklı varyantları bulunmaktadır. Bunlar arasında Gaussian NB, Multinomial NB ve Bernoulli NB en sık kullanılan modellerdir. Gaussian NB, özelliklerin normal dağıldığı varsayımıyla çalışırken, Multinomial NB genellikle belge sınıflandırma gibi kelime frekanslarına dayalı uygulamalarda kullanılmaktadır. Bernoulli NB ise ikili (binary) özelliklere sahip veriler için uygundur (McCallum ve Nigam 1998). Ancak NB'in temel sınırlaması, tüm özelliklerin birbirinden tamamen bağımsız olduğu varsayımıdır. Gerçek dünya verilerinde bu varsayım çoğu zaman geçerli olmayabilir. Buna rağmen, özellikler arasında bağımlılıklar olsa bile, algoritma çoğu uygulamada başarılı sonuçlar vermektedir.

3.5.4 K En Yakın Komşu

KNN algoritması, hem sınıflandırma hem de regresyon problemlerinde kullanılan temel bir MÖ yöntemidir. Algoritmanın temel prensibi, benzer özelliklere sahip veri noktalarının öznitelik uzayında birbirine yakın bulunacağı varsayımına dayanmaktadır (Cover ve Hart 1967, Deveci ve Özarı 2020).

Algoritmanın temel adımları şu şekildedir: Öncelikle, sınıflandırılacak veya değeri tahmin edilecek yeni bir veri noktası için, eğitim veri setindeki tüm noktalarla mesafe hesaplanmaktadır. Mesafe ölçümü için genellikle Öklid, Manhattan veya Minkowski metrikleri kullanılmaktadır. Hesaplamalar sonucunda, belirlenen K sayısı kadar en yakın komşu seçilmektedir. Sınıflandırma problemlerinde, bu komşuların çoğunluk sınıfı yeni veri noktasına atanırken; regresyon problemlerinde, komşuların hedef değerlerinin ortalaması alınarak tahmin yapılmaktadır (Duda ve Hart 2006, Görmez vd. 2024, Han vd. 2011).

KNN algoritmasının başlıca avantajları arasında, model eğitimi gerektirmemesi ve veri setinde meydana gelen değişikliklere kolaylıkla uyum sağlayabilmesi bulunmaktadır. Ancak, dezavantajları da bulunmaktadır. Özellikle büyük veri setlerinde tüm mesafelerin hesaplanması ciddi bir hesaplama yükü oluşturabilmektedir. Ayrıca, öznitelikler arasındaki ölçek farklılıkları mesafe hesaplamalarını olumsuz etkileyebileceği için, veri ön işleme aşamasında standartlaştırma veya normalizasyon yapılması gerekmektedir (Zhang 2016). Ayrıca, K değerinin uygun şekilde belirlenmesi algoritmanın doğruluğu üzerinde doğrudan etkili olmaktadır.

3.5.5 Destek Vektör Makineleri

DVM, veri noktalarını sınıflandırmak amacıyla en uygun ayırıcı hiperdüzlemi bulmaya dayanan güçlü bir denetimli MÖ algoritmasıdır. Temel amacı, iki sınıf arasındaki mesafeyi maksimum yapan bir karar sınırı oluşturarak sınıflar arasında mümkün olan en geniş ayrımı sağlamaktır (Öklü ve Canbay 2023). Bu algoritma özellikle yüksek boyutlu veri uzaylarında etkili performans sergilemektedir (Vapnik 1999).

DVM'nin temel özelliklerinden biri, yalnızca sınıf sınırına yakın konumlanan ve "destek vektörleri" olarak adlandırılan veri noktalarını kullanmasıdır. Bu yaklaşım, algoritmanın hem verimliliğini artırmakta hem de modelin genelleme kapasitesine katkı sağlamaktadır (Cristianini ve Shawe-Taylor 2000).

Doğrusal olarak ayrılabilir olmayan veriler için, DVM çekirdek fonksiyonları (kernel trick) kullanarak verileri daha yüksek boyutlu bir uzaya taşımaktadır. Böylece, karmaşık sınırların daha basit hiperdüzlemlerle ayrılması mümkün hale gelmektedir. Polinom ve sigmoid

fonksiyonlar en yaygın kullanılan çekirdek fonksiyonları arasında yer almaktadır (Scholkopf ve Smola 2002).

DVM algoritması, özellikle küçük ve orta ölçekli veri setlerinde, metin sınıflandırma, biyomedikal veri analizi ve görüntü tanıma gibi alanlarda yüksek başarı sağlamaktadır. Bununla birlikte, büyük veri setlerinde hesaplama maliyetinin artması ve model parametrelerinin doğru şekilde ayarlanmasının zorluk oluşturması gibi bazı sınırlamaları da bulunmaktadır (Hsu vd. 2010).

3.5.6 Gradyan Arttırma

GA, MÖ'inde tahmin performansını artırmak amacıyla zayıf öğrenicilerin ardışık olarak birleştirildiği güçlü bir topluluk yöntemidir. Friedman tarafından 1999 yılında temelleri atılan bu yaklaşım, gradyan iniş prensiplerine dayanarak adım adım hataları minimize eden bir optimizasyon süreci izlemektedir (Arslan ve Yapıcı 2024, Friedman 1999).

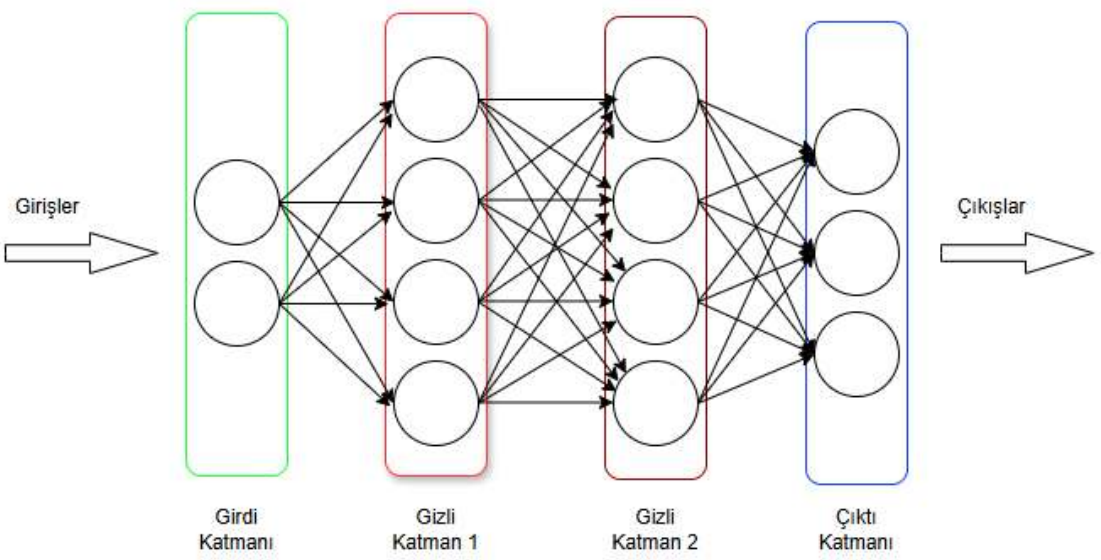
Algoritma, başlangıçta basit bir model oluşturarak sürece başlamakta ve her yinelemede (iterasyonda), önceki modelin tahmin hatalarını azaltmaya yönelik yeni bir zayıf öğrenici ekleyerek modeli kademeli olarak geliştirmektedir. Bu süreçte, regresyon problemlerinde hata terimlerinin (residuals) minimize edilmesi, sınıflandırma problemlerinde ise logaritmik kayıp gibi uygun kayıp fonksiyonlarının optimizasyonu hedeflenmektedir (Hastie vd. 2005).

GA algoritmasının temel avantajı, her yeni modelin, önceki modellerin hatalı tahmin ettiği örneklere odaklanarak aşamalı bir iyileştirme sağlamasıdır. Öğrenme oranı (learning rate) parametresi ise her eklenen modelin katkı düzeyini belirleyerek, genel modelin esneklik ve genelleme kapasitesi üzerinde doğrudan belirleyici bir rol oynamaktadır (Chen ve Guestrin 2016). Bununla birlikte, büyük veri setlerinde eğitim süresinin uzaması ve hiperparametre ayarlarının hassasiyet gerektirmesi gibi sınırlamalar, yöntemin uygulamadaki etkinliğini doğrudan etkileyebilmektedir.

3.5.7 Çok Katmanlı Algılayıcılar

ÇKA denetimli öğrenmeye dayalı YSA modelleri arasında yer almakta olup, özellikle sınıflandırma ve regresyon problemlerinde sıklıkla kullanılmaktadır (Kılınç 2022). Temel

olarak bir giriş katmanı, bir veya birden fazla gizli katman ve bir çıkış katmanından oluşmaktadır (Şekil 3.2). Her bir katmandaki nöronlar, kendisinden önceki katmandaki nöronlarla tam bağlantılıdır ve bu bağlantılar, öğrenme süreci boyunca güncellenen ağırlık değeriyle temsil edilmektedir (Somuncu ve Oral 2024).



Şekil 3.2 ÇKA mimari yapısı.

ÇKA'ların çok katmanlı mimarisi, giriş verileri ile hedef değişken arasındaki karmaşık ve doğrusal olmayan ilişkilerin öğrenilebilmesine olanak sağlamaktadır. Bu süreçte, giriş verileri katmanlar arasında aktarılırken ağırlık çarpımı ve bias eklenmesi gibi doğrusal işlemler gerçekleştirilmekte, ardından ReLU, tanh veya sigmoid gibi doğrusal olmayan aktivasyon fonksiyonları uygulanarak modelin karar sınırları karmaşıklaştırılmaktadır. Böylelikle model, veri içerisindeki karmaşık örüntüleri yakalayarak yüksek düzeyde temsil gücü elde edebilmektedir (Bishop 2006, Haykin 2009). Bu kapsamda, modelin başarıyla öğrenim gerçekleştirebilmesi, ağırlık parametrelerinin uygun şekilde optimize edilmesini gerekli kılmaktadır.

ÇKA modellerinde ağırlıkların optimize edilmesi süreci, genellikle geri yayılım (backpropagation) algoritması ile birlikte kullanılan gradyan iniş temelli optimizasyon yöntemleri (örneğin Stochastic Gradient Descent veya Adam) aracılığıyla yürütülmektedir. Bu yöntemler aracılığıyla modelin hata fonksiyonundaki değerler minimize edilmekte ve böylece genel öğrenme performansı iyileştirilmektedir. Bununla birlikte, çok katmanlı yapının sahip

olduğu yüksek temsil kapasitesi, aşırı öğrenme riskini de beraberinde getirmektedir. Bu durumu engelleyebilmek amacıyla, eğitim sürecinde L2 normu gibi düzenlileştirme (regularization) teknikleri veya dropout gibi yöntemler uygulanmakta, böylece modelin hem eğitim hem de doğrulama verileri üzerinde daha yüksek bir genelleme başarımı göstermesi sağlanmaktadır.

Genel performans açısından değerlendirildiğinde, ÇKA'lar, doğrusal olmayan ve karmaşık ilişkileri modelleyebilme kapasiteleri sayesinde sınıflandırma ve regresyon problemlerinde yüksek doğruluk oranlarına ulaşabilmektedir. Özellikle büyük, çeşitli ve yapısal karmaşıklık içeren veri kümeleri üzerinde, esnek mimarileri sayesinde güçlü bir genelleme performansı sergilemektedirler. Bununla birlikte, çok katmanlı yapılar, yüksek parametre sayıları nedeniyle eğitim süresinin uzamasına ve hesaplama maliyetlerinin artmasına yol açabilmektedir. Ayrıca, yapının karmaşıklığına bağlı olarak aşırı öğrenme riski de artış gösterebilmektedir. Bu nedenle, ÇKA tabanlı modellerde yapılandırma sürecinde model karmaşıklığının probleme uygun şekilde dengelenmesi ve düzenlileştirme stratejilerinin etkin biçimde uygulanması, performansın sürdürülebilirliği açısından kritik bir gereklilik oluşturmaktadır.

3.5.8 Doğrusal Diskriminant Analizi

Doğrusal Diskriminant Analizi (DDA), sınıflandırma problemlerinde yaygın olarak kullanılan, denetimli öğrenmeye dayalı bir boyut indirgeme yöntemidir. Temel amacı, veri setinde yer alan örnekleri ait oldukları sınıflara en iyi şekilde ayıracak doğrusal bir dönüşüm bulmaktır. DDA, sınıflar arasındaki ayrımı en üst düzeye çıkarırken, aynı sınıf içerisindeki örnekler arasındaki varyansı minimize etmeyi hedeflemektedir (Fisher 1936). Bu yöntemde, her sınıfın ortalama vektörleri ile toplam veri kümesinin ortalama vektörü hesaplanarak, sınıflar arası (between-class) ve sınıf içi (within-class) varyans matrisleri oluşturulmaktadır. DDA, bu iki matris arasındaki optimizasyonu sağlayan projeksiyon doğrultularını bularak, yüksek boyutlu verileri daha düşük boyutlu bir alt uzaya indirgemektedir. Böylece, veri setindeki örneklerin sınıflar bazında daha net ayrışması sağlanmaktadır (McLachlan 2004, Yavuz ve Yavuz 2021).

DDA, özellikle sınıflar arası dağılımın doğrusal ayrılabilir olduğu durumlarda yüksek sınıflandırma performansı sunmaktadır. Ancak, sınıflar arasındaki sınırların doğrusal olmadığı karmaşık veri setlerinde performansı sınırlı kalabilmektedir (Duda vd. 2001). DDA, yalnızca sınıflandırma amacıyla değil, aynı zamanda yüksek boyutlu veri kümelerinin boyutunu azaltarak görselleştirme ve veri ön işleme süreçlerinde de etkin biçimde kullanılmaktadır.

3.5.9 Kuadratik Diskriminant Analizi

Kuadratik Diskriminant Analizi (KDA), sınıflandırma problemlerinde kullanılan ve sınıflar arasındaki sınırların doğrusal olmadığı durumlarda DDA'ya göre daha esnek bir yaklaşım sunan bir yöntemdir. Temel amacı, veri setindeki örnekleri ait oldukları sınıflara ayırmak için her sınıfa özgü karar sınırları oluşturmaktır. KDA, her bir sınıf için ayrı ayrı kovaryans matrisleri hesaplayarak sınıf içi varyans farklılıklarını dikkate almakta ve böylece sınıflar arasındaki ayrımı doğrusal değil, kuadratik karar sınırları ile gerçekleştirmektedir (McLachlan 2004).

KDA'nın çalışma prensibi, sınıflara ait olasılık dağılımlarının çok değişkenli normal (Gauss) dağılım gösterdiği varsayımına dayanmaktadır. Ancak DDA'nın aksine, her sınıfın kovaryans matrisinin farklı olmasına izin verilmektedir. Bu durum, sınıflar arasında daha karmaşık ve kıvrımlı ayırıcı sınırların öğrenilebilmesini sağlamaktadır. Özellikle sınıfların varyans-yapılarının önemli ölçüde farklılık gösterdiği veri kümelerinde KDA, DDA'ya kıyasla daha yüksek sınıflandırma performansı sunabilmektedir (Hastie vd. 2009).

Bununla birlikte, KDA'nın esnek yapısı daha fazla parametre öğrenilmesini gerektirdiği için, özellikle küçük veri kümelerinde aşırı öğrenme riskini artırabilmektedir. Ayrıca, her sınıf için ayrı kovaryans matrisi hesaplanması hesaplama maliyetlerini artırmakta ve modelin daha fazla veri gerektirmesine neden olmaktadır. Bu nedenle, veri setinin boyutu ve sınıflar arasındaki varyans yapısı dikkate alınarak DDA ve KDA yöntemleri arasında seçim yapılması önerilmektedir (Duda vd. 2001, Nejad 2017).

3.5.9 Torbalama

Torbalama (Bagging, Bootstrap Aggregating) algoritması, istatistiksel yeniden örnekleme yöntemine dayalı bir topluluk öğrenmesi tekniğidir. Bu yöntemin temel amacı, farklı modellerin çıktılarının birleştirilmesi yoluyla bireysel modellerin varyansını azaltmak ve genel model performansını artırmaktır (Ürün ve Sönmez 2024). Bagging, özellikle yüksek varyans gösteren ve aşırı öğrenmeye eğilimli sınıflandırıcılar üzerinde etkili sonuçlar sağlamaktadır (Onan 2018).

Bagging algoritmasında, orijinal eğitim verisinden rastgele ve tekrarlı seçimlerle birçok alt küme (bootstrap örneklemleri) oluşturulmaktadir. Her bir alt küme üzerinde bağımsız bir model eğitilmekte ve tahmin aşamasında bu modellerin çıktıları birleştirilmektedir. Sınıflandırma problemlerinde genellikle çoğunluk oyu (majority voting) yöntemi, regresyon problemlerinde ise ortalama alma (averaging) yöntemi kullanılarak nihai tahmin üretilmektedir. Bu yaklaşım, tek bir modelin eğitim verisine aşırı uyum göstermesini önleyerek modelin genelleme yeteneğini artırmaktadır (Friedman 1999).

Bagging algoritmasının en önemli avantajlarından biri, temel öğrenicilerin her biri yüksek varyanslı olsa bile, bir araya geldiklerinde toplam model varyansının önemli ölçüde düşmesidir. Özellikle KA gibi varyansı yüksek algoritmalarla birlikte kullanıldığında, Bagging algoritması sınıflandırma doğruluğunu önemli ölçüde artırabilmektedir. Bununla birlikte, Bagging yalnızca varyansı azaltmakta etkili olup, modelin önyargısını (bias) azaltmada sınırlı bir etkiye sahiptir. Ayrıca, birden fazla model eğitilmesi nedeniyle eğitim süresi ve hesaplama maliyeti, tek model yaklaşımlarına kıyasla daha yüksek olabilmektedir (Rokach 2010).

3.5.10 Oylama

Voting algoritması, topluluk öğrenme yaklaşımları arasında yer alan ve birden fazla sınıflandırıcının çıktılarının birleştirilerek nihai kararın oluşturulmasını amaçlayan bir yöntemdir. Temel prensibi, bireysel sınıflandırıcıların tahmin sonuçlarının bir araya getirilerek daha güvenilir ve kararlı bir karar mekanizması oluşturulmasıdır. Böylece, tek bir sınıflandırıcının hatalarından kaynaklanan belirsizlikler azaltılarak genel sınıflandırma performansı artırılabilmektedir (Kuncheva 2004).

Voting algoritmasında, her bir temel sınıflandırıcı bağımsız olarak eğitilmekte ve tahmin aşamasında her bir model kendi kararını üretmektedir. Nihai kararın verilmesi sürecinde ise iki temel strateji uygulanmaktadır: çoğunluk oyu ve ağırlıklı oylama (weighted voting). Çoğunluk oyu yaklaşımında, sınıflandırıcıların tahmin ettiği sınıflar arasında en fazla oyu alan sınıf tercih edilirken; ağırlıklı oylama stratejisinde, her sınıflandırıcının doğruluk oranı gibi belirli bir kriter doğrultusunda farklı ağırlıklar atanmakta ve bu ağırlıklar doğrultusunda nihai karar belirlenmektedir (Kuncheva 2004, Ürün ve Sönmez 2024).

Voting algoritmasının en önemli avantajlarından biri, temel sınıflandırıcılar arasında çeşitliliğin sağlanması durumunda, genel modelin aşırı öğrenme eğilimini azaltması ve veri seti üzerinde daha iyi bir genelleme başarımı sunabilmesidir. Bununla birlikte, birden fazla sınıflandırıcının eşzamanlı olarak eğitilmesi, eğitim süresinin ve hesaplama maliyetinin artmasına neden olabilmektedir. Ayrıca, topluluk yapısında yer alan sınıflandırıcıların yeterli çeşitliliğe sahip olmaması durumunda elde edilen performans kazancı sınırlı kalabilmektedir (Opitz ve Maclin 1999).

3.6 DEĞERLENDİRME ÖLÇÜTLERİ

MÖ tabanlı sınıflandırma modellerinin etkinliğinin doğru bir şekilde ölçülebilmesi, yalnızca modelin genel doğruluk oranına bakmakla sınırlı değildir. Bir modelin performansını kapsamlı bir şekilde değerlendirebilmek için farklı yönlerden ölçüm yapmayı sağlayan çeşitli performans metriklerine ihtiyaç duyulmaktadır. Özellikle sınıflar arasında dengesizlik bulunan veri kümelerinde, doğruluk oranı tek başına yeterli bir gösterge oluşturmamakta; bu nedenle kesinlik, duyarlılık ve F1 skoru gibi tamamlayıcı ölçütlerin de dikkate alınması gerekmektedir.

Bu çalışmada, geliştirilen modellerin sınıflandırma başarımını daha ayrıntılı bir şekilde analiz edebilmek amacıyla doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1 skoru gibi yaygın sınıflandırma metriklerinden yararlanılmıştır. Söz konusu metriklerin hesaplanmasında temel alınan yapı, modelin tahmin sonuçlarını gerçek sınıf etiketleriyle karşılaştırarak doğru ve yanlış sınıflandırmaları gösteren karmaşıklık matrisidir. Karmaşıklık matrisi, modelin sınıflandırma performansını dört temel kategori üzerinden değerlendirir: Doğru Pozitif (DP), Doğru Negatif (DN), Yanlış Pozitif (YP) ve Yanlış Negatif (YN). Doğru Pozitif (DP), gerçek sınıfı pozitif olan ve model tarafından doğru şekilde pozitif olarak tahmin edilen örnekleri ifade etmektedir. Doğru Negatif (DN), gerçek sınıfı negatif olan ve model tarafından doğru şekilde negatif olarak tahmin edilen örneklerdir. Yanlış Pozitif (YP) ise gerçek sınıfı negatif olmasına rağmen model tarafından pozitif olarak tahmin edilen örneklerdir. Yanlış Negatif (YN) ise gerçek sınıfı pozitif olmasına rağmen model tarafından negatif olarak tahmin edilen örnekleri temsil etmektedir.

Karmaşıklık matrisi sayesinde, modelin yalnızca genel doğruluğu değil, aynı zamanda pozitif ve negatif sınıflar üzerindeki ayırt etme yeteneği de ayrıntılı bir şekilde analiz edilebilmektedir. Böylece modelin güvenilirliği, hata yapma eğilimleri ve sınıflar arasındaki performans dengesi

daha sağlıklı bir biçimde değerlendirilmektedir. Bu doğrultuda, çalışmada kullanılan performans ölçütlerinin matematiksel tanımları aşağıda sunulmaktadır.

Modelin genel doğruluk oranı, DP ve DN tahminlerin tüm tahminler içerisindeki oranı olarak ifade edilir ve aşağıdaki şekilde hesaplanır:

$$Doğruluk\ (Accuracy) = \frac{DP+DN}{DP+DN+YP+YN} \tag{3.16}$$

Kesinlik, modelin pozitif olarak tahmin ettiği örneklerden kaç tanesinin gerçekten pozitif olduğunu gösterir ve şu şekilde tanımlanır:

$$Kesinlik\ (Precision) = \frac{DP}{DP+YP} \tag{3.17}$$

Duyarlılık, gerçek pozitif örneklerin model tarafından doğru şekilde tahmin edilme oranını ifade eder ve aşağıdaki formülle hesaplanır:

$$Duyarlılık\ (Recall) = \frac{DP}{DP+YN} \tag{3.18}$$

F1 skoru, kesinlik ve duyarlılık değerlerinin harmonik ortalaması alınarak elde edilir ve modelin pozitif sınıfı tanıma performansını daha dengeli bir şekilde değerlendirmek için kullanılır. F1 skoru şu şekilde ifade edilmektedir:

$$F1\ Skoru = 2 * \frac{Kesinlik*Duyarlılık}{Kesinlik+Duyarlılık} \tag{3.19}$$

Bu ölçütler kullanılarak, modelin sınıflandırma başarımı hem eğitim sürecinde çapraz doğrulama sonuçları üzerinden hem de ayrılan bağımsız test verisi üzerinde değerlendirilmiş ve analiz edilmiştir (Yapıcı ve Arslan 2024).

BÖLÜM 4

SONUÇLAR

Bu tez çalışmasında, bireylerin çevresel tutum düzeylerinin sınıflandırılması amacıyla veri odaklı sistematik bir MÖ yaklaşımı sunulmuştur. Türkiye genelinden 388 üniversite öğrencisinden toplanan ve 34 değişkenden oluşan anket verileri kullanılarak gerçekleştirilen bu çalışmada, çevreye yönelik duyusal eğilimlerin varlığına ilişkin ikili sınıflandırma (0: eğilim yok, 1: eğilim var) gerçekleştirilmiştir. Analiz süreci üç temel aşamadan oluşmaktadır:

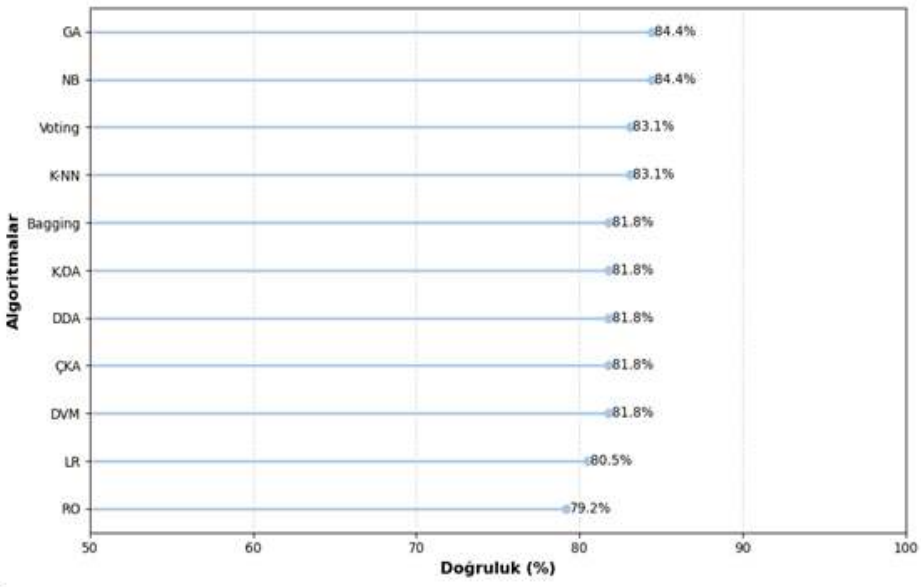
İlk aşamada, modelin tahmin gücünü artırmak amacıyla 10 farklı öznitelik seçim yöntemi (ANOVA, Ki-Kare, EA, LASSO, Karşılıklı Bilgi, TBA, RFE-LR, RFE-RF, RO ve Varyans Eşiği) sistematik olarak uygulanmış ve çevresel tutumları en iyi şekilde ayırt eden kritik özellikler belirlenmiştir. Bu çok yöntemli yaklaşım sayesinde, değişkenlerin hedef sınıfla ilişkisi hem istatistiksel hem de algoritmik düzeyde kapsamlı şekilde değerlendirilmiş, böylece modelin güvenilirliği artırılmıştır. İkinci aşamada, belirlenen bu önemli özellikler kullanılarak 11 farklı MÖ algoritması (LR, RO, DVM, KNN, NB, GA, ÇKA, DDA, KDA, Bagging ve Voting) ile sınıflandırma modelleri geliştirilmiştir.

Son aşamada, tüm modellerin performansı %80 eğitim ve %20 test şeklinde yapılan veri bölümlemesi ile 5 katlı çapraz doğrulama yöntemi kullanılarak sistematik biçimde değerlendirilmiştir. Performans analizinde doğruluk, kesinlik, duyarlılık ve F1 skoru gibi çoklu metrikler birlikte ele alınmış, özellikle en yüksek doğruluk değerlerine ulaşan algoritmaların sınıflandırma başarısı karmaşıklık matrisleriyle derinlemesine analiz edilmiştir. Bu kapsamlı değerlendirme yaklaşımı, çevresel tutumların sınıflandırılmasında en optimal algoritmanın nesnel kriterlere dayalı olarak belirlenmesini sağlamıştır.

Bu sistematik ve çok aşamalı analiz süreci sonucunda elde edilen bulgular aşağıda ayrıntılı şekilde sunulmakta; çevresel tutumların belirleyici özellikleri ve en başarılı sınıflandırma modelinin performansı kapsamlı biçimde tartışılmaktadır.

4.1 ANOVA TABANLI ÖZELLİK SEÇİMİ SONRASI SINIFLANDIRICILARIN PERFORMANS ANALİZİ

Bu alt bölümde, çevresel tutumların sınıflandırılmasında etkili olan belirleyici (anlamlı) özelliklerin tespiti amacıyla ANOVA yöntemi uygulanmıştır. İstatistiksel olarak anlamlı bulunan bu özellikle, sınıflandırma modellerinin eğitiminde bağımsız değişken olarak kullanılmış ve farklı sınıflandırma algoritmaları üzerindeki etkileri sistematik biçimde incelenmiştir. Bu kapsamda, Şekil 4.1'de ilgili modellerin doğruluk metriği açısından performansları karşılaştırmalı olarak sunulmuştur.



Şekil 4.1 ANOVA Tabanlı Özellik Seçimiyle Oluşturulan MÖ Modellerinin Doğruluk Değerleri.

Şekil 4.1'de görüldüğü gibi, ANOVA ile seçilen özellikler kullanılarak oluşturulan modeller arasında NB ve GA tabanlı yaklaşımlar, %84,4 doğruluk oranı ile en yüksek sınıflandırma performansını göstermiştir. Bu modelleri %83,1 doğruluk oranı ile KNN ve Voting tabanlı modeller takip etmiştir. DVM, ÇKA, DDA, KDA ve Bagging temelli modeller ise %81,8 doğrulukla tutarlı ve birbirine yakın sonuçlar vermiştir. Öte yandan, RO modeli %79,2 doğrulukla nispeten daha düşük bir performans sergilerken, LR modeli %80,5 doğruluk oranıyla orta düzeyde bir başarı elde etmiştir.

Doğruluk metriğinin yanı sıra, modellerin diğer performans ölçütlerine ilişkin ayrıntılı sonuçlar Çizelge 4.1'de sunulmuştur.

Çizelge 4.1 ANOVA İle Seçilen Özelliklerle Eğitilen Modellerin Kesinlik, Duyarlılık Ve F1 Skoru Performansları.

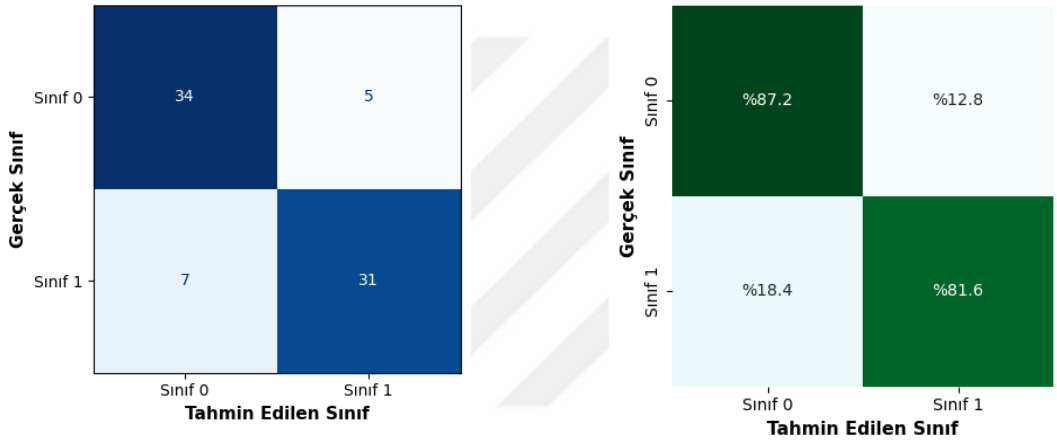
Modeller	**Kesinlik (%)**	**Duyarlılık (%)**	**F1 skor (%)**
LR	81,1	78,9	80,0
RO	82,4	73,7	77,8
DVM	83,3	78,9	81,1
KNN	82,1	84,2	83,1
NB	86,1	81,6	83,8
GA	86,1	81,6	83,8
ÇKA	85,3	76,3	80,6
DDA	81,6	81,6	81,6
KDA	83,3	78,9	81,1
Bagging	87,5	73,7	80,0
Voting	85,7	78,9	82,2

Çizelge 4.1'deki kesinlik metriği sonuçlarına göre, Bagging modeli %87,5'lik performansıyla en yüksek başarıyı sergilemiştir. İkinci sırada %86,1'lik kesinlik oranlarıyla NB ve GA modelleri yer almıştır. Voting (%85,7) ve ÇKA (%85,3) modelleri orta-üst performans grubunu oluştururken, KDA ve DVM modelleri %83,3'lük birbirine yakın değerler kaydetmiştir. Performans skalasının alt kısmında ise RO (%82,4), KNN (%82,1), DDA (%81,6) ve LR (%81,1) modelleri yer almıştır.

Duyarlılık metriği açısından değerlendirildiğinde, en yüksek başarı %84,2 ile KNN modeli tarafından elde edilmiştir. Bunu, %81,6'lık duyarlılık oranlarıyla NB, GA ve DDA modelleri takip etmiştir. Voting, KDA, DVM ve LR modelleri ise %78,9'luk benzer performans düzeyleriyle orta sıralarda yer almıştır. ÇKA modeli %76,3 ile bu gruptan daha düşük bir performans sergilerken, en düşük değerler (%73,7) RO ve Bagging modellerinde gözlemlenmiştir.

F1 skor metriği açısından en yüksek performans, %83,8 ile NB ve GA modelleri tarafından sergilenmiştir. Bu modelleri sırasıyla %83,1 ile KNN ve %82,2 ile Voting modeli takip etmiştir. DDA modeli %81,6'lık bir başarı elde ederken, KDA ve DVM modelleri %81,1, ÇKA modeli %80,6, LR ve Bagging modelleri ise %80 F1 skoruyla göreli olarak daha düşük performans göstermiştir.

Genel bir değerlendirme yapıldığında, NB ve GA modellerinin tüm performans metrikleri bakımından tutarlı ve yüksek düzeyde performans sergilediği tespit edilmiştir. Bu modellerin doğruluk, kesinlik, duyarlılık ve F1 skor metrikleri açısından da elde ettiği güçlü sonuçlar, sınıflandırma sürecinde güvenilir ve etkili bir yaklaşım sunduklarını ortaya koymaktadır. Söz konusu metriklerdeki bu başarının altında yatan sınıflandırma dinamiklerini incelemek amacıyla, modellere ait karmaşıklık matrisleri Şekil 4.2'de hem mutlak hem de normalize edilmiş yüzdelik değerler içeren karmaşıklık matrisleri sunulmuştur. Bu matris modelin sınıflandırma performansının nicel değerlendirilmesine ilişkin önemli bulgular ortaya koymaktadır.



(a) (b)

Şekil 4.2 NB ve GA modellerinin karmaşıklık matrisi: (a) Mutlak Değerler (b) Normalize Değerler (%)

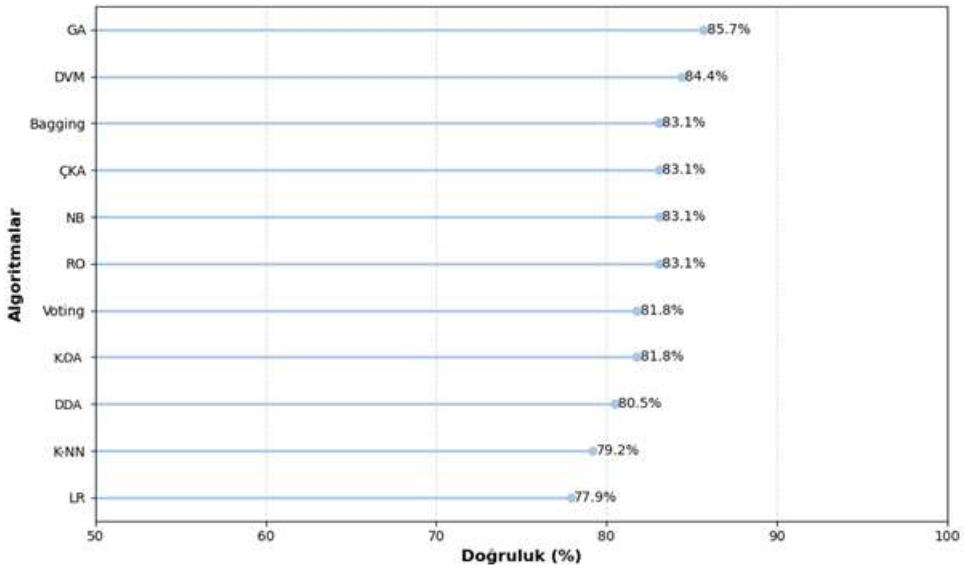
Şekil 4.2 (a)'da mutlak değerlerle sunulan karmaşıklık matrisine göre, gerçek sınıfı 0 olan 39 örneğin 34'ü doğru bir şekilde 'Sınıf 0' olarak tahmin edilirken, 5 örnek 'Sınıf 1' şeklinde yanlış sınıflandırılmıştır. Benzer şekilde, gerçek sınıfı 1 olan 38 örneğin 31'i doğru olarak tanımlanmış, ancak 7 örnek 'Sınıf 0' olarak hatalı etiketlenmiştir. Bu bulgular, NB ve GA modellerinin her iki sınıf için de yüksek bir sınıflandırma doğruluğu sergilediğini, ancak 'Sınıf 1' örneklerinde nispi olarak daha yüksek bir hata oranı gösterdiğini ortaya koymaktadır.

Şekil 4.2 (b)'de normalize edilmiş değerlerle sunulan karmaşıklık matrisi ise bu durumu yüzdesel oranlarla desteklemektedir. Modeller, 'Sınıf 0' örneklerini %87,2 doğrulukla sınıflandırırken, 'Sınıf 1' örneklerinde bu oran %81,6'ya düşmektedir. Yanlış sınıflandırma

oranlarının karşılaştırılması ('Sınıf 1'de %18,4'e karşılık 'Sınıf 0'da %12,8), modellerin 'Sınıf 1' örneklerini ayırt etmede görece daha fazla zorlandığını göstermektedir.

4.2 Kİ KARE TABANLI ÖZELLİK SEÇİMİ SONRASI SINIFLANDIRICILARIN PERFORMANS ANALİZİ

Bu alt bölümde, çevresel tutumların sınıflandırılmasında etkili olabilecek özellikleri tespit etmek amacıyla Ki kare bağımsızlık testi uygulanmıştır. Analiz neticesinde, hedef değişken ile istatistiksel olarak anlamlı ilişki gösteren bağımsız değişkenler belirlenmiş ve bu değişkenler sınıflandırma modellerinin eğitim sürecinde kullanılmıştır. Ardından, geliştirilen sınıflandırma modellerinin performansları çeşitli performans değerlendirme metrikleri açısından karşılaştırmalı olarak analiz edilmiştir. Bu bağlamda, Şekil 4.3'de söz konusu modellerin doğruluk metriğine dayalı sınıflandırma performansları sunularak algoritmaların başarı düzeyleri nicel olarak ortaya konulmuştur.



Şekil 4.3 Ki-Kare Tabanlı Özellik Seçimiyle Oluşturulan MÖ Modellerinin Doğruluk Değerleri

Şekil 4.3'de sunulan sınıflandırma modellerinin performans analizine ilişkin bulgular değerlendirildiğinde, en yüksek doğruluk değerinin %85,7 ile GA tabanlı model tarafından elde edildiği gözlemlenmiştir. Performans açısından GA modelini, %84,4 doğruluk oranıyla DVM modelinin izlediği ve istatistiksel olarak anlamlı bir fark olmaksızın benzer bir başarım düzeyi sergilediği tespit edilmiştir. Bagging, ÇKA, NB ve RO modellerinin ise %83,1 doğruluk

oranıyla istatistiksel açıdan birbirine yakın sonuçlar ürettiği ve homojen bir performans grubu oluşturduğu belirlenmiştir. Voting ve KDA modellerinin %81,8'lik doğruluk değerleriyle bu grubun hemen ardından geldiği ve anlamlı bir performans farkı göstermediği kaydedilmiştir. DDA modelinin %80,5'lik oranla bir miktar geride kaldığı, KNN modelinin ise %79,2 doğruluk değeriyle nispeten düşük bir performans sergilediği gözlenmiştir. Son olarak, en düşük sınıflandırma başarımının %77,9 ile LR) modelinde kaydedildiği tespit edilmiştir. Doğruluk metriğinin yanı sıra, modellere ait diğer performans ölçütlerinin ayrıntılı sonuçları Çizelge 4.2'de sunulmuştur.

Çizelge 4.2 Ki-Kare Testi İle Seçilen Özelliklerle Eğitilen Modellerin Kesinlik, Duyarlılık ve F1 Skoru Performansları.

Modeller	Kesinlik (%)	Duyarlılık (%)	F1 skor (%)
LR	74,4	84,2	79,0
RO	82,1	84,2	83,1
DVM	86,1	81,6	83,8
KNN	80,6	76,3	78,4
NB	80,5	86,8	83,5
GA	86,5	84,2	85,3
ÇKA	82,1	84,2	83,1
DDA	75,6	89,5	81,9
KDA	78,6	86,8	82,5
Bagging	82,1	84,2	83,1
Voting	83,3	78,9	81,1

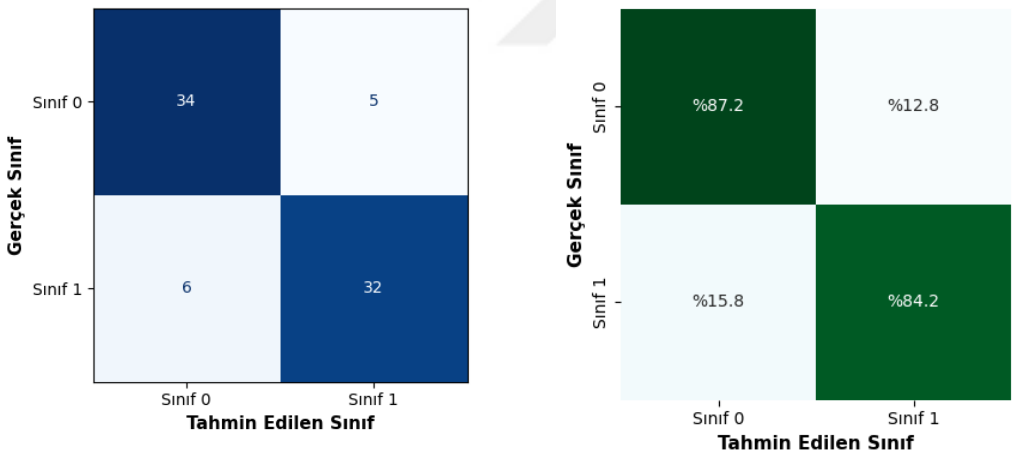
Çizelge 4.2 incelendiğinde, kesinlik metriği açısından en yüksek performans %86,5 değeri ile GA tabanlı model tarafından elde edilmiştir. Bu modeli sırasıyla %86,1 ile DVM ve %83,3 ile Voting modelleri izlemiştir. RO, Bagging ve ÇKA modelleri ise %82,1 kesinlik oranı ile birbirine yakın sonuçlar üretmiştir. Buna karşılık KNN, NB ve KDA modelleri sırasıyla %80,6, %80,5 ve %78,6 kesinlik değerleri ile orta düzeyde bir başarı göstermiştir. Öte yandan, DDA (%75,6) ve LR (%74,4) modelleri, kesinlik ölçütü bazında diğer modellere kıyasla anlamlı derecede düşük performans sergilemiştir.

Duyarlılık metriği açısından yapılan değerlendirmede, en yüksek başarı oranı %89,5 ile DDA modeli tarafından elde edilmiştir. Bu modeli, %86,8 duyarlılık değerleri ile NB ve KDA modelleri takip etmiştir. LR, GA, RO, ÇKA ve Bagging modelleri ise %84,2 duyarlılık oranı ile benzer bir performans düzeyi sergilemiştir. DVM modeli %81,6 duyarlılık değeri ile bu grup

içinde nispeten daha düşük bir performans gösterirken, Voting modeli %78,9 ile daha sınırlı bir başarı elde etmiştir. En düşük duyarlılık değeri ise %76,3 ile KNN modelinde gözlemlenmiştir.

F1 skoru metriği açısından ise, GA modeli %85,3 değeri ile en yüksek başarıyı göstermiştir. Bu modeli sırasıyla %83,8 ile DVM ve %83,5 ile NB modelleri izlemiştir. RO, ÇKA ve Bagging modelleri ise %83,1 değeriyle benzer düzeyde performans sergilemiştir. KDA, DDA ve Voting modelleri ise sırasıyla %82,5, %81,9 ve %81,1'lik F1 skoru değerleriyle nispeten daha düşük sonuçlar üretmiştir. Buna karşılık, en düşük F1 skoru değerleri %79,0 ile LR ve %78,4 ile KNN modellerinde gözlemlenmiştir.

Yapılan kapsamlı analizler sonucunda, GA tabanlı modelin, özellikle doğruluk (%85,7) kesinlik (%86,5) ve F1 skoru (%85,3) metriklerinde diğer modellere kıyasla belirgin şekilde üstün performans sergilediği tespit edilmiştir. Bu bulgular, GA modelinin sınıflandırma görevlerinde hem doğruluk hem de dengeli bir hassasiyet-duyarlılık performansı sunduğunu ortaya koymaktadır. Modelin sınıf bazlı performansını daha kapsamlı analiz edebilmek amacıyla, Şekil 4.4'de mutlak ve normalize edilmiş değerleri içeren karmaşıklık matrisleri sunulmuştur.



(a) (b)

Şekil 4.4 GA Modelinin Karmaşıklık Matrisi: (a) Mutlak Değerler, (b) Normalize Değerler (%)

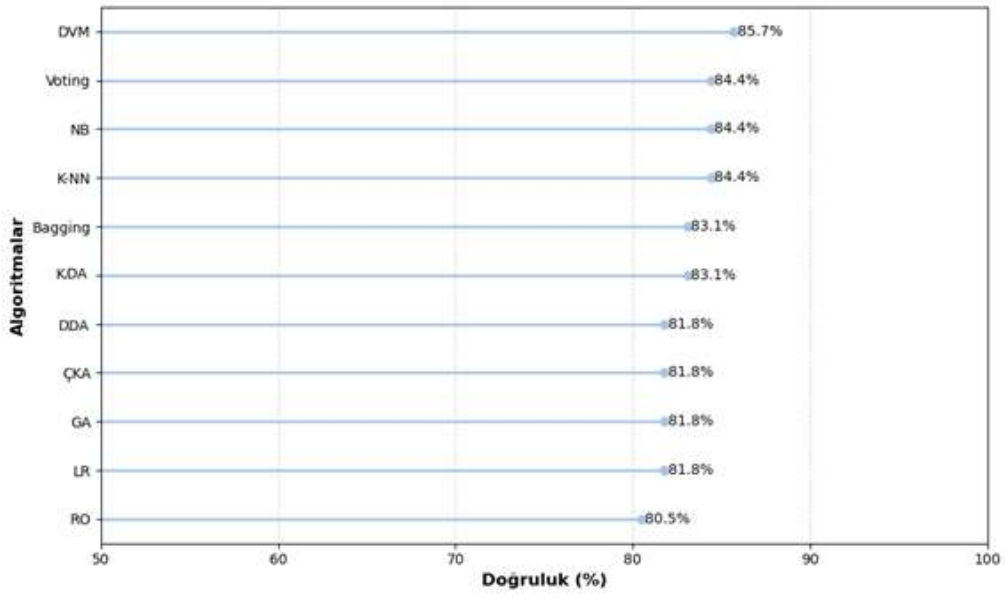
Şekil 4.4 (a)'da mutlak değerlerle sunulan karmaşıklık matrisine göre, gerçek sınıf etiketi 0 olan toplam 39 örneğin 34'ü doğru bir şekilde 'Sınıf 0' olarak sınıflandırılırken, 5 örnek yanlışlıkla 'Sınıf 1' olarak tahmin edilmiştir. Öte yandan, gerçek sınıfı 1 olan 38 örneğin 32'si başarıyla 'Sınıf 1' olarak tanımlanmış, ancak 6 örnek hatalı bir şekilde 'Sınıf 0' olarak etiketlenmiştir. Bu sonuçlar, GA tabanlı modelin her iki sınıf için de yüksek genel doğruluk sergilediğini, ancak

özellikle 'Sınıf 1' örneklerinde nispeten daha yüksek bir yanlış sınıflandırma eğilimi gösterdiğini ortaya koymaktadır.

Şekil 4.4 (b)'de normalize edilmiş değerlerle sunulan karmaşıklık matrisi bu durumu yüzdesel oranlarla desteklemektedir. Model, 'Sınıf 0' örneklerini %87,2 doğruluk oranıyla sınıflandırırken, 'Sınıf 1' için bu performans %84,2 seviyesinde kalmıştır. Yanlış sınıflandırma oranlarının karşılaştırılması ('Sınıf 0'da %12,8 iken 'Sınıf 1'de %15,8), modelin 'Sınıf 1' örneklerini ayırt etmede görece daha fazla güçlük çektiğini göstermektedir.

4.3 EKSTRA AĞAÇLAR TABANLI ÖZELLİK SEÇİMİ SONRASI SINIFLANDIRICILARIN PERFORMANS ANALİZİ

Bu alt bölümde, çevresel tutumların sınıflandırılmasında etkili olan belirleyici özelliklerin tespiti amacıyla EA yöntemi kullanılarak özellik seçimi gerçekleştirilmiştir. Modelin KA üzerinden hesapladığı önem düzeyleri doğrultusunda, hedef değişkenle istatistiksel açıdan anlamlı ilişkiler sergileyen özellikler belirlenmiştir. Seçilen bu özellikler, sınıflandırma algoritmalarının eğitim sürecinde bağımsız değişken olarak kullanılmış ve farklı MÖ modelleri üzerindeki etkileri sistematik biçimde analiz edilmiştir. Bu doğrultuda, Şekil 4.5'de söz konusu modellerin doğruluk değerleri karşılaştırmalı olarak sunularak, EA yöntemiyle yapılan özellik seçiminin model performansına katkısı ortaya konulmuştur.



Şekil 4.5 EA Tabanlı Özellik Seçimiyle Oluşturulan Mö Modellerinin Doğruluk Değerleri.

Şekil 4.5'de yer alan sınıflandırma modellerinin doğruluk metriği bazında karşılaştırmalı analizi incelendiğinde, en yüksek başarım değerinin %85,7 ile DVM modelinde gözlendiği tespit edilmiştir. İkinci en yüksek performans grubunu ise %84,4 doğruluk oranıyla Voting, NB ve KNN modelleri göstermiştir. Bagging ve KDA modellerinin ise %83,1'lik doğruluk oranıyla birbirine yakın performans sergiledikleri belirlenmiştir. Bununla birlikte, DDA, ÇKA, GA ve LR modellerinin %81,8 doğruluk değeriyle istatistiksel olarak benzer bir başarım düzeyi gösterdiği gözlemlenmiştir. Öte yandan, söz konusu modeller arasında en düşük sınıflandırma performansı %80,5 ile RO modeli tarafından elde edilmiştir. Doğruluk metriğine ek olarak, modellerin duyarlılık, özgüllük ve F1 skoru gibi diğer performans ölçütlerine ilişkin ayrıntılı sonuçlar ise Çizelge 4.3'de sunulmuştur.

Çizelge 4.3 EA Yöntemiyle ile Seçilen Özelliklerle Eğitilen Modellerin Kesinlik, Duyarlılık ve F1 Skoru Performansları

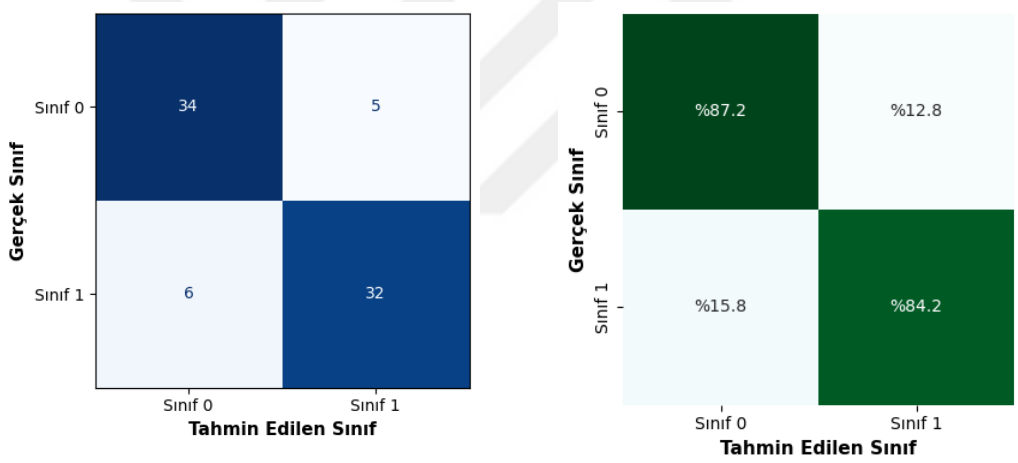
Modeller	Kesinlik (%)	Duyarlılık (%)	F1 skor (%)
LR	83,3	78,9	81,1
RO	79,5	81,6	80,5
DVM	86,5	84,2	85,3
KNN	88,2	78,9	83,3
NB	88,2	78,9	83,3
GA	80	84,2	82,1
ÇKA	83,3	78,9	81,1
DDA	83,3	78,9	81,1
KDA	83,8	81,6	82,7
Bagging	82,1	84,2	83,1
Voting	88,2	78,9	83,3

Çizelge 4.3'de göre kesinlik metriği açısından en yüksek performans %88,2 değeriyle KNN, NB ve Voting modelleri tarafından elde edilmiştir. Bu modelleri sırasıyla %86,5 kesinlik oranı ile DVM ve %83,8 ile KDA modelleri takip etmiştir. LR, ÇKA ve DDA modelleri %83,3 kesinlik değeri ile istatistiksel olarak anlamlı bir farklılık göstermezken, Bagging modeli %82,1 ve GA modeli ise %80 kesinlik oranı kaydetmiştir. Kesinlik metriğinde en düşük performans %79,5 ile RO modelinde gözlemlenmiştir.

Duyarlılık metriği açısından incelendiğimde, en yüksek başarı oranı %84,2 ile DVM, GA ve Bagging modelleri tarafından sağlanmıştır. RO ve KDA modelleri %81,6 duyarlılık değeri sergilerken, diğer tüm modellerin bu ölçütteki performansı %78,9 seviyesinde kalmıştır.

F1 skor metriği kapsamında yapılan değerlendirmede ise en yüksek performans değeri %85,3 ile DVM modeli tarafından elde edilmiştir. Bu modeli sırasıyla %83,3 F1 skor değeriyle KNN, NB ve Voting modelleri izlemiş olup, bunları %83,1 ile Bagging modeli; %82,7 ile KDA modeli ve %82,1 ile GA modeli takip etmiştir. LR, ÇKA ve DDA modelleri %81,1 F1skor değeri ile nispeten daha düşük performans sergilemişlerdir. Analizler sonucunda, en düşük F1 skor değerinin ise %80,5 ile RO modelinde kaydedildiği tespit edilmiştir.

Sonuçlar genel olarak değerlendirildiğinde DVM modeli, hem duyarlılık hem de F1 skor açısından en iyi performansı göstererek dengeli bir sınıflandırma sağlamıştır. Bunun yanı sıra doğruluk metriği açısından da en yüksek sınıflandırma başarımı DVM modeliyle elde edilmiştir. Bu kapsamda bu başarının altında yatan sınıflandırma dinamiklerini incelemek amacıyla, modellere ait karmaşıklık matrisleri Şekil 4.6'da hem mutlak hem de normalize edilmiş yüzdelik değerler içeren karmaşıklık matrisleri sunulmuştur.



(a) (b)

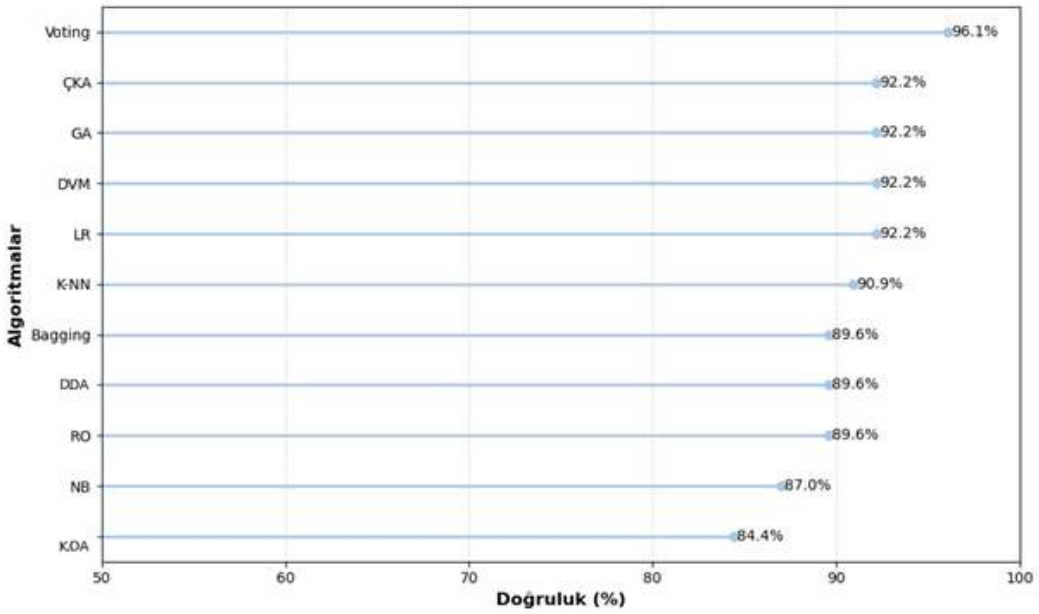
Şekil 4.6 DVM Modelinin Karmaşıklık Matrisi: (a) Mutlak Değerler (b) Normalize Değerler (%)

Şekil 4.6 (a)'da mutlak değerlerle sunulan karmaşıklık matrisine göre, gerçek sınıf etiketi 0 olan toplam 39 örneğin 34'ü doğru bir şekilde 'Sınıf 0' olarak sınıflandırılırken, 5 örnek yanlışlıkla 'Sınıf 1' olarak tahmin edilmiştir. Öte yandan, gerçek sınıfı 1 olan 38 örneğin 32'si başarıyla 'Sınıf 1' olarak tanımlanmış, ancak 6 örnek hatalı bir şekilde 'Sınıf 0' olarak etiketlenmiştir. Bu sonuçlar, DVM tabanlı modelin her iki sınıf için de yüksek genel doğruluk sergilediğini, ancak özellikle 'Sınıf 1' örneklerinde nispeten daha yüksek bir yanlış sınıflandırma eğilimi gösterdiğini ortaya koymaktadır.

Şekil 4.6 (b)'de normalize edilmiş değerlerle sunulan karmaşıklık matrisi bu durumu yüzdesel oranlarla desteklemektedir. Model, 'Sınıf 0' örneklerini %87,2 doğruluk oranıyla sınıflandırırken, 'Sınıf 1' için bu performans %84,2 seviyesinde kalmıştır. Yanlış sınıflandırma oranlarının karşılaştırılması ('Sınıf 0'da %12,8 iken 'Sınıf 1'de %15,8), modelin 'Sınıf 1' örneklerini ayırt etmede görece daha fazla güçlük çektiğini göstermektedir.

4.3 LASSO TABANLI ÖZELLİK SEÇİMİ SONRASI SINIFLANDIRICILARIN PERFORMANS ANALİZİ

Bu alt bölümde, çevresel tutumların sınıflandırılmasında en belirleyici özelliklerin seçilmesi amacıyla LASSO temelli bir özellik seçimi yöntemi uygulanmıştır. LASSO yöntemi, model eğitimi sırasında önemsiz değişkenlerin katsayılarını sıfıra yakınsatarak, hedef değişkenle anlamlı ilişki gösteren özellikleri otomatik olarak seçme avantajı sunmaktadır. Bu sayede, boyutsallığın azaltılması (dimensionality reduction) ve model karmaşıklığının kontrol altına alınması sağlanmış, aynı zamanda aşırı öğrenme riski minimize edilmiştir. LASSO yöntemiyle belirlenen özellikler, sınıflandırma algoritmalarının eğitim sürecinde girdi olarak kullanılmış ve farklı MÖ modelleri üzerindeki etkileri sistematik biçimde analiz edilmiştir. Bu doğrultuda, Şekil 4.7'de ilgili modellerin doğruluk değerleri karşılaştırmalı olarak sunularak, LASSO yöntemiyle yapılan özellik seçiminin model performansına katkısı ortaya konulmuştur.



Şekil 4.7 Lasso Tabanlı Özellik Seçimiyle Oluşturulan MÖ Modellerinin Doğruluk Değerleri.

Şekil 4.7'ye göre, %96,1 doğruluk oranıyla en yüksek performansı Voting modeli sergilemiştir. Bu modeli sırasıyla %92,2 doğruluk oranına sahip ÇKA, GA, DVM ve LR modelleri takip etmiştir. KNN modeli ise %90,9 doğruluk oranıyla ortalama düzeyde bir performans göstermiştir. Diğer taraftan, Bagging, DDA ve RO modelleri %89,6 doğruluk oranıyla benzer sonuçlar üretirken, NB modelinde bu oran %87'ye gerilemiştir. En düşük doğruluk oranı ise %84,4 ile KDA modeline ait olup, değerlendirmeye alınan modeller arasında en düşük başarıyı göstermiştir. Doğruluk metriğine ek olarak, modellerin duyarlılık, özgüllük ve F1 skoru gibi diğer performans ölçütlerine ilişkin ayrıntılı sonuçlar ise Çizelge 4.4'de sunulmuştur.

Çizelge 4.4 LASSO Yöntemiyle ile Seçilen Özelliklerle Eğitilen Modellerin Kesinlik, Duyarlılık ve F1 Skoru Performansları.

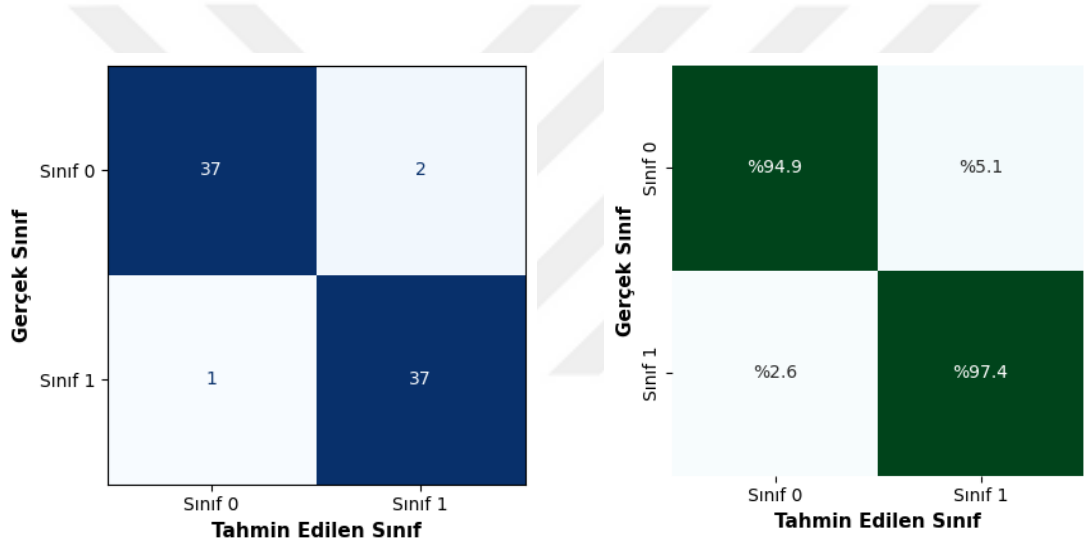
Modeller	Kesinlik (%)	Duyarlılık (%)	F1 skor (%)
LR	92,1	92,1	92,1
RO	87,5	92,1	89,7
DVM	92,1	92,1	92,1
KNN	91,9	89,5	90,7
NB	88,9	84,2	86,5
GA	90,0	94,7	92,3
ÇKA	92,1	92,1	92,1
DDA	89,5	89,5	89,5
KDA	81,0	89,5	85,0
Bagging	87,5	92,1	89,7
Voting	94,9	97,4	96,1

Çizelge 4.4'te yer alan kesinlik ölçütü değerlendirildiğinde, en yüksek performansın %94,9 ile Voting modeli tarafından sergilendiği gözlemlenmiştir. Bu modeli, %92,1 kesinlik oranıyla LR, DVM ve ÇKA modellerinin izlediği belirlenmiştir. KNN modeli %91,9, GA modeli %90,0 ve DDA modeli ise %89,5 kesinlik değeri elde etmiştir. Öte yandan, NB (%88,9), RO (%87,5), Bagging (%87,5) ve KDA (%81,0) modellerinin bu ölçüt kapsamında nispeten daha düşük performans sergilediği tespit edilmiştir.

Duyarlılık ölçütü açısından incelendiğinde, Voting modeli %97,4 ile en yüksek başarıyı gösterirken, bu modeli %94,7 duyarlılık oranıyla GA modelinin takip ettiği görülmüştür. LR, DVM, ÇKA, RO ve Bagging modellerinin benzer şekilde %92,1 duyarlılık değeri kaydettiği, DDA, KDA ve KNN modellerinin ise %89,5 seviyesinde kaldığı belirlenmiştir. NB modeli, %84,2 ile duyarlılık ölçütünde en düşük performansı sergileyen model olarak öne çıkmıştır.

F1 skoru metriği dikkate alındığında, Voting modeli %96,1 ile yine en yüksek performansı gösterirken, bu modeli %92,3 ile GA modeli izlemiştir. LR, DVM ve ÇKA modelleri %92,1 F1 skor değeriyle birbirine yakın sonuçlar üretmiştir. KNN modeli %90,7 F1 skoru ile dikkat çekerken, RO ve Bagging modelleri sırasıyla %89,7; DDA modeli ise %89,5 F1 skoru elde etmiştir. NB modeli %86,5 ile daha düşük bir performans sergilemiş, en düşük F1 skoru ise %85,0 ile KDA modeline ait olmuştur.

Bu bulgular, Voting modelinin diğer modellere kıyasla sınıflandırma performansı açısından istatistiksel olarak daha üstün olduğunu göstermektedir. En yüksek doğruluk oranına sahip olan bu modele ilişkin ayrıntılı sınıflama sonuçları ise Şekil 4.8'de sunulan karmaşıklık matrisi üzerinden görselleştirilmiştir.



(a) (b)

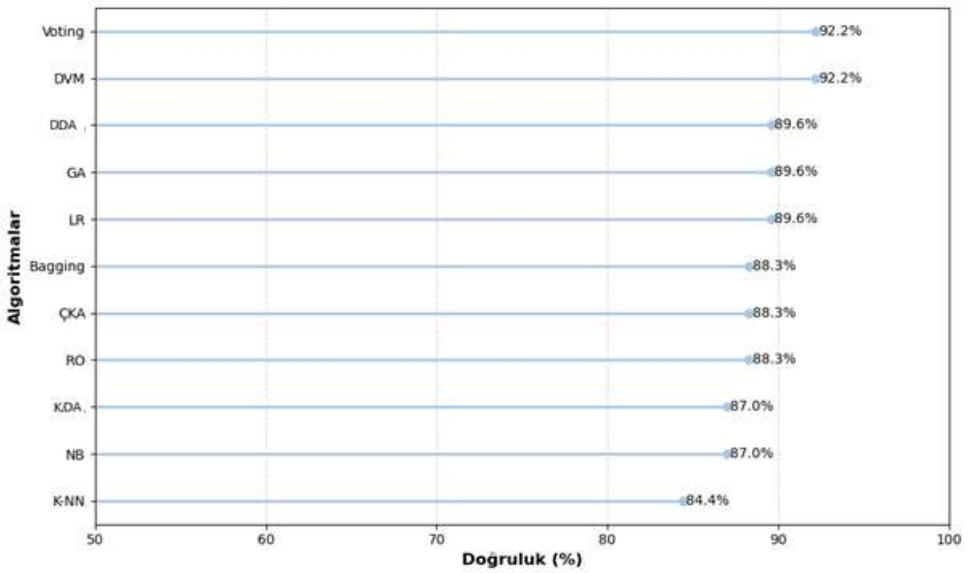
Şekil 4.8 Voting Modelinin Karmaşıklık Matrisi: (a) Mutlak Değerler (b) Normalize Değerler (%)

Şekil 4.8 (a)'da mutlak değerlerle sunulan karmaşıklık matrisine göre, gerçek sınıf etiketi 0 olan toplam 39 örneğin 37'sini doğru bir şekilde 'Sınıf 0' olarak sınıflandırılırken, 2 örnek yanlışlıkla 'Sınıf 1' olarak tahmin edilmiştir. Öte yandan, gerçek sınıfı 1 olan 38 örneğin 37'sini başarıyla 'Sınıf 1' olarak tanımlanmış, ancak 1 örnek hatalı bir şekilde 'Sınıf 0' olarak etiketlenmiştir. Bu sonuçlar, Voting tabanlı modelin her iki sınıf için de yüksek genel doğruluk sergilediğini, ancak özellikle 'Sınıf 0' örneklerinde nispeten daha yüksek bir yanlış sınıflandırma eğilimi gösterdiğini ortaya koymaktadır.

Şekil 4.8 (b)'de normalize edilmiş değerlerle sunulan karmaşıklık matrisi bu durumu yüzdesel oranlarla desteklemektedir. Model, 'Sınıf 0' örneklerini %94,9 doğruluk oranıyla sınıflandırırken, 'Sınıf 1' için bu performans %97,4 seviyesinde kalmıştır. Yanlış sınıflandırma oranlarının karşılaştırılması ('Sınıf 0'da %5,1 iken 'Sınıf 1'de %2,6), modelin 'Sınıf 0' örneklerini ayırt etmede görece daha fazla güçlük çektiğini göstermektedir.

4.4 KARŞILIKLI BİLGİ TABANLI ÖZELLİK SEÇİMİ SONRASI SINIFLANDIRICILARIN PERFORMANS ANALİZİ

Bu alt bölümde, çevresel tutumların sınıflandırılmasında en baskın özelliklerin belirlenmesi amacıyla karşılıklı bilgi temelli bir özellik seçimi yöntemi uygulanmıştır. Karşılıklı bilgi yönteminde; özellikler ile hedef değişken arasındaki istatistiksel bağımlılık ölçülerek, en yüksek bilgi kazancı sağlayan özellikler tespit edilmiştir. Daha sonra bu özellikler, sınıflandırma algoritmalarının eğitim sürecinde girdi olarak kullanılmış ve farklı MÖ modelleri üzerindeki etkileri sistematik biçimde analiz edilmiştir. Bu doğrultuda, Şekil 4.9'da ilgili modellerin doğruluk değerleri karşılaştırmalı olarak sunularak, karşılıklı bilgi yöntemiyle yapılan özellik seçiminin sınıflandırıcıların performansına katkısı ortaya konulmuştur.



Şekil 4.9 Karşılıklı Bilgi Tabanlı Özellik Seçimiyle Oluşturulan MÖ Modellerinin Doğruluk Değerleri.

Şekil 4.9'a göre, en yüksek doğruluk değeri %92,2 ile Voting ve DVM modellerinden elde edilmiştir. Bu yüksek başarımı %89,6 doğruluk oranıyla DDA, GA ve LR modelleri takip etmiştir. Bagging, ÇKA ve RO modelleri ise ortalama bir performans sergilemiş olup, doğruluk değerleri %88,3 olarak hesaplanmıştır. Yanı sıra KDA ve NB modelleri %87 doğruluk değeriyle ortalamanın altında bir başarım ortaya koymuşlardır. En düşük başarım ise %84,4 ile KNN modelinde gözlenmiştir. Doğruluk metriğine ek olarak, modellerin duyarlılık, özgüllük ve F1 skoru gibi diğer performans ölçütlerine ilişkin ayrıntılı sonuçlar ise Çizelge 4.5'de sunulmuştur.

Çizelge 4.5 Karşılıklı Bilgi Yöntemiyle Seçilen Özelliklerle Eğitilen Modellerin Kesinlik, Duyarlılık ve F1 Skoru Performansları

Modeller	Kesinlik (%)	Duyarlılık (%)	F1 skor (%)
LR	96,9	81,6	88,6
RO	89,2	86,8	88,0
DVM	92,1	92,1	92,1
KNN	82,5	86,8	84,6
NB	91,2	81,6	86,1
GA	91,7	86,8	89,2
ÇKA	87,2	89,5	88,3
DDA	94,1	84,2	88,9
KDA	85,0	89,5	87,2
Bagging	89,2	86,8	88,0
Voting	92,1	92,1	92,1

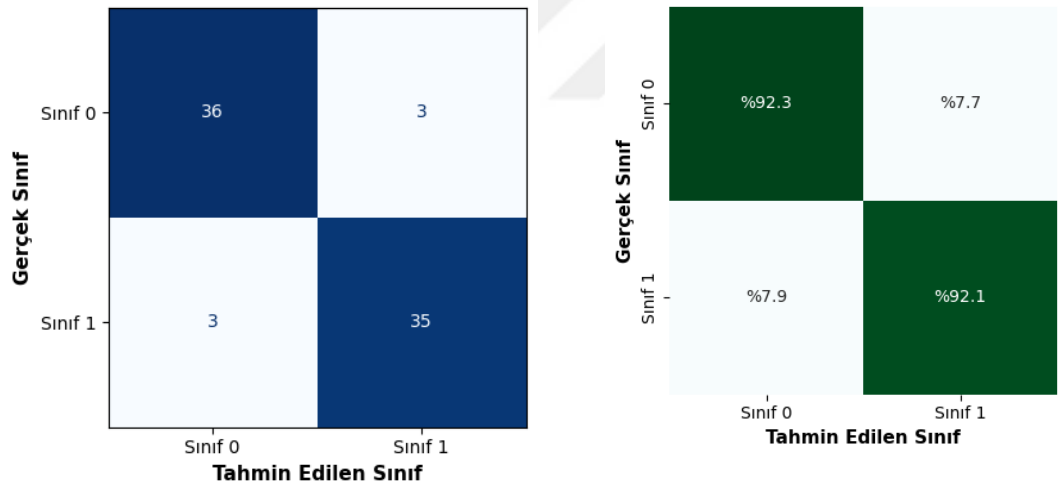
Çizelge 4.5'de sunulan model performans metrikleri incelendiğinde, kesinlik (precision) metriği açısından, en yüksek değer %96,9 ile LR modeli tarafından elde edildiği gözlemlenmiştir. Bu modeli sırasıyla %94,1 ile DDA, %92,1 ile DVM ve Voting modelleri takip etmiştir. GA (%91,7) ve NB (%91,2) modelleri de yüksek düzeyde kesinlik değerleri sergilemiştir. Öte yandan, RO ve Bagging modelleri %89,2 kesinlik değerine sahipken, ÇKA %87,2 ve KDA %85,0 değerleriyle daha düşük performans göstermiştir. En düşük kesinlik performansı ise %82,5 ile KNN modelinde gözlemlenmiştir.

Duyarlılık metriği açısından değerlendirildiğinde, en yüksek başarım %92,1 ile DVM ve Voting modelleri tarafından elde edilmiştir. Bu modelleri, %89,5 duyarlılık değeri ile ÇKA ve KDA modelleri izlemiştir. RO, GA, Bagging ve KNN modelleri %86,8 oranında benzer bir

performans sergilerken, DDA modelinin %84,2 ile daha düşük bir performans gösterdiği görülmüştür. En düşük duyarlılık değerleri ise %81,6 ile LR ve NB modellerinde gözlemlenmiştir.

F1 skoru ölçütüne göre yapılan değerlendirmede, en yüksek performansın yine %92,1 ile DVM ve Voting modelleri tarafından elde edildiği belirlenmiştir. Bu modelleri sırasıyla GA (%89,2), DDA (%88,9), LR (%88,6), ÇKA (%88,3) ve RO ile Bagging (%88,0) modelleri takip etmiştir. KDA (%87,2) ve NB (%86,1) modelleri bu ölçütte nispeten daha sınırlı bir başarı sergilerken, en düşük F1 skoru değeri %84,6 ile KNN modelinde gözlemlenmiştir.

Sonuçlar genel olarak değerlendirildiğinde, DVM ve Voting modellerinin tüm performans metrikleri açısından daha dengeli ve tutarlı bir performans sergilediği tespit edilmiştir. Söz konusu modellerin sınıflandırma performanslarını nicel ve nitel olarak doğrulamak amacıyla Şekil 4.10'da karmaşıklık matrisi hem mutlak hem de normalize edilmiş değerler üzerinden görselleştirilmiştir.



(a) (b)

Şekil 4.10 DVM ve Voting Modellerinin Karmaşıklık Matrisi: (a) Mutlak Değerler(b) Normalize Değerler (%)

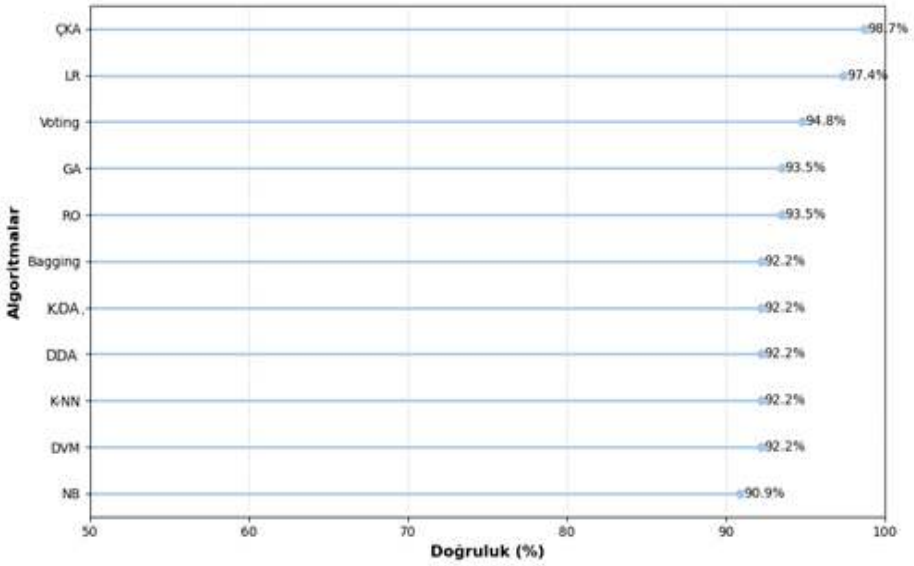
Şekil 4.10 (a)'da mutlak değerlerle sunulan karmaşıklık matrisine göre, gerçek sınıf etiketi 0 olan toplam 39 örneğin 36'sını doğru bir şekilde 'Sınıf 0' olarak sınıflandırılırken, 3 örnek yanlışlıkla 'Sınıf 1' olarak tahmin edilmiştir. Öte yandan, gerçek sınıfı 1 olan 38 örneğin 35'ini başarıyla 'Sınıf 1' olarak tanımlanmış, ancak 3 örnek hatalı bir şekilde 'Sınıf 0' olarak etiketlenmiştir. Bu sonuçlar, DVM ve Voting tabanlı modellerin her iki sınıf için de yüksek

genel doğruluk sergilediğini ve yanlış sınıflandırmaların her iki sınıfta da benzer düzeyde gerçekleştiğini göstermektedir.

Şekil 4.10 (b)'de normalize edilmiş değerlerle sunulan karmaşıklık matrisi bu durumu yüzdesel oranlarla desteklemektedir. Model, 'Sınıf 0' örneklerini %92,3, 'Sınıf 1' örneklerini ise %92,1 doğruluk oranıyla sınıflandırmıştır. Her iki sınıfa ait yanlış sınıflandırma oranları birbirine oldukça yakın olup, bu durum modelin sınıflar arasında dengeli bir performans sergilediğini göstermektedir.

4.5 TBA TABANLI ÖZELLİK SEÇİMİ SONRASI SINIFLANDIRICILARIN PERFORMANS ANALİZİ

Bu alt bölümde, çevresel tutumların sınıflandırılmasında en anlamlı özelliklerin belirlenmesi amacıyla TBA temelli bir özellik seçimi yöntemi uygulanmıştır. TBA yöntemiyle verideki boyutluluk azaltılarak temel özellikler tespit edilmiştir. Daha sonra seçilen bu özellikler, sınıflandırma algoritmalarının eğitim sürecinde girdi olarak kullanılmış ve farklı MÖ modelleri üzerindeki etkileri sistematik biçimde analiz edilmiştir. Bu doğrultuda, Şekil 4.11'de ilgili modellerin doğruluk değerleri karşılaştırmalı olarak sunularak, TBA tabanlı özellik seçiminin sınıflandırıcıların performansına katkısı ortaya konulmuştur.



Şekil 4.11 TBA Tabanlı Özellik Seçimiyle Oluşturulan MÖ Modellerinin Doğruluk Değerleri

Şekil 4.11'e göre, en yüksek doğruluk değeri %98,7 ile ÇKA modelinde elde edilmiştir. Bu yüksek başarımı %97,4 doğruluk oranıyla LR modeli takip etmiştir. Voting modeliyle %94,8'lik bir doğruluk değerine ulaşılırken, GA ve RO modellerinde bu oran %93,5 seviyelerinde kalmıştır. Bagging, KDA, DDA, KNN ve DVM modelleri için doğruluk oranı %92,5 olarak hesaplanmıştır. En düşük doğruluk oranı (%90,9) ise NB modelinde elde edilmiştir. Doğruluk metriğine ek olarak, modellerin duyarlılık, özgüllük ve F1 skoru gibi diğer performans ölçütlerine ilişkin ayrıntılı sonuçlar ise Çizelge 4.6'da sunulmuştur.

Çizelge 4.6 TBA Yöntemiyle Seçilen Özelliklerle Eğitilen Modellerin Kesinlik, Duyarlılık Ve F1 Skoru Performansları

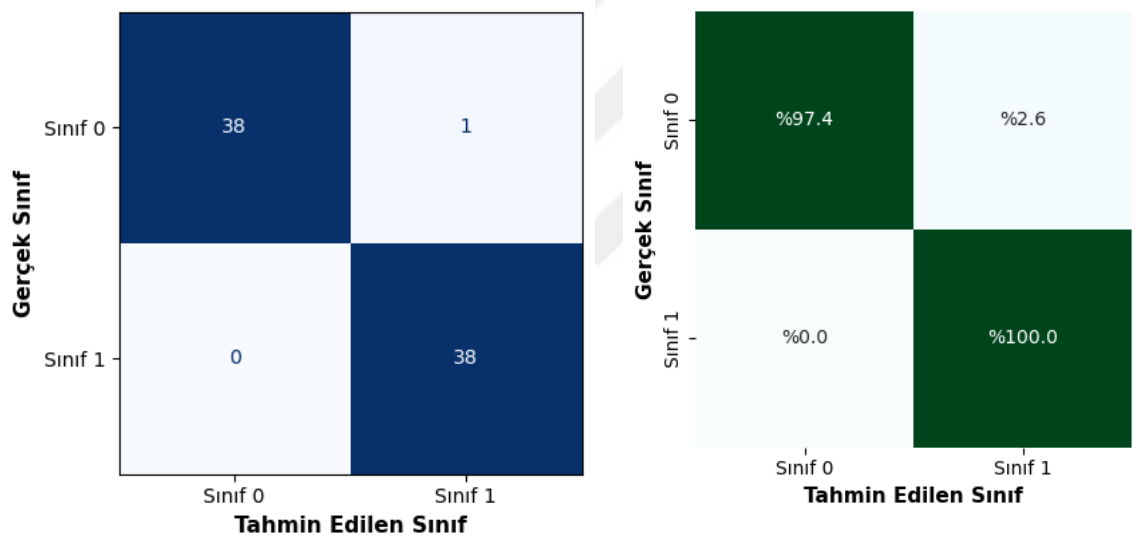
Modeller	Kesinlik (%)	Duyarlılık (%)	F1 skor (%)
LR	95,0	100,0	97,4
RO	90,2	97,4	93,7
DVM	88,1	97,4	92,5
KNN	92,1	92,1	92,1
NB	86,0	97,4	91,4
GA	92,3	94,7	93,5
ÇKA	97,4	100,0	98,7
DDA	90,0	94,7	92,3
KDA	92,1	92,1	92,1
Bagging	92,1	92,1	92,1
Voting	92,5	97,4	94,9

Çizelge 4.6'da sunulan bulgulara göre, kesinlik ölçütü açısından en yüksek değer %97,4 ile ÇKA modeli tarafından elde edilmiştir. Bu modeli sırasıyla %95,0 ile LR ve %92,5 ile Voting modelleri izlemiştir. GA, KNN, KDA ve Bagging modelleri %92,1 kesinlik değeriyle benzer düzeyde performans sergilemiştir. RO %90,2, DDA %90,0, DVM %88,1 ve NB %86,0 kesinlik oranı ile bu ölçüt açısından daha düşük sonuçlar sunmuştur.

Duyarlılık ölçütüne açısından değerlendirildğinde, en yüksek oran %100 ile LR ve ÇKA modellerinde elde edilmiştir. Bu modelleri %97,4 oranındaki duyarlılık değerleri ile RO, DVM, NB ve Voting modelleri takip etmiştir. GA ve DDA modelleri %94,7 ile yüksek düzeyde duyarlılık gösterirken, KNN, KDA ve Bagging modelleri %92,1 oranında benzer performanslar sergilemiştir.

F1 skoru ölçütüne göre yapılan değerlendirmede, en yüksek performansın yine %98,7 ile ÇKA modeli tarafından elde edildiği belirlenmiştir. Bu modeli sırasıyla %97,4 ile LR ve %94,9 ile Voting modelleri izlemiştir. Ayrıca, RO (%93,7) ve GA (%93,5) modelleri de yüksek performans göstermiştir. DVM (%92,5), DDA (%92,3) ile KNN, KDA ve Bagging modelleri (%92,1) ise birbirine yakın ve dengeli sonuçlar üretmiştir. Bu ölçütte en düşük performans %91,4 ile NB modelinde gözlemlenmiştir.

Sonuçlar genel olarak değerlendirildiğinde, ÇKA modelinin tüm performans metrikleri açısından daha dengeli ve tutarlı bir performans sergilediği tespit edilmiştir. Söz konusu modelin sınıflandırma performansını nicel ve nitel olarak doğrulamak amacıyla Şekil 4.12'de karmaşıklık matrisi hem mutlak hem de normalize edilmiş değerler üzerinden görselleştirilmiştir.



(a) (b)

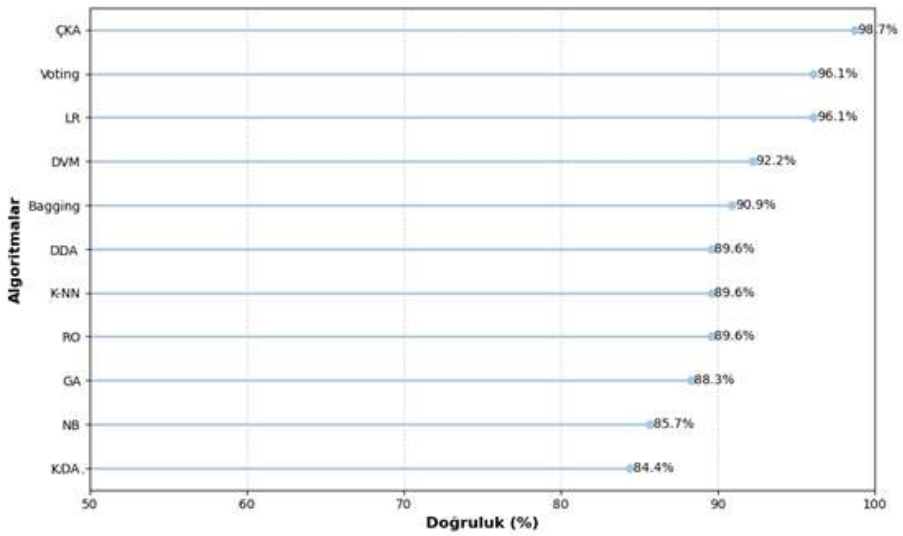
Şekil 4.12 ÇKA Modelinin Karmaşıklık Matrisi: (a) Mutlak Değerler (b) Normalize Değerler (%)

Şekil 4.12 (a)'da mutlak değerlerle sunulan karmaşıklık matrisine göre, gerçek sınıf etiketi 0 olan toplam 39 örneğin 38'sini doğru bir şekilde 'Sınıf 0' olarak sınıflandırılırken, 1 örnek yanlışlıkla 'Sınıf 1' olarak tahmin edilmiştir. Öte yandan, gerçek sınıfı 1 olan 38 örneğin 38'ini başarıyla 'Sınıf 1' olarak tanımlanmıştır. Bu sonuçlar, ÇKA modelinin her iki sınıf için de yüksek genel doğruluk sergilediğini ve yalnızca Sınıf 0'da düşük düzeyde bir yanlış sınıflandırma gerçekleştiğini göstermektedir.

Şekil 4.12 (b)'de normalize edilmiş değerlerle sunulan karmaşıklık matrisi bu durumu yüzdesel oranlarla desteklemektedir. Model, 'Sınıf 0' örneklerini %97,4, 'Sınıf 1' örneklerini ise %100,0 doğruluk oranıyla sınıflandırmıştır. Bu sonuçlar, ÇKA modelinin özellikle 'Sınıf 1' örneklerini tamamen doğru sınıflandırabildiğini ve genel olarak sınıflar arasında oldukça dengeli bir performans ortaya koyduğunu göstermektedir.

4.6 RASTGELE ORMAN TABANLI ÖZELLİK SEÇİMİ SONRASI SINIFLANDIRICILARIN PERFORMANS ANALİZİ

Bu alt bölümde, çevresel tutumların sınıflandırılmasında en anlamlı değişkenleri belirlemek amacıyla RO tabanlı bir özellik seçimi yöntemi uygulanmıştır. Bu yöntemde, değişkenlerin sınıflandırma üzerindeki etkileri karar ağacı yapıları aracılığıyla değerlendirilmiş ve en yüksek bilgi katkısına sahip olanlar seçilmiştir. Daha sonra seçilen bu özellikler, sınıflandırma algoritmalarının eğitim aşamasında kullanılmış ve farklı MÖ modelleri üzerindeki etkileri sistematik olarak incelenmiştir. Bu kapsamda, Şekil 4.13'de modellerin doğruluk değerleri karşılaştırmalı olarak sunularak, RO temelli özellik seçiminin sınıflandırıcıların performansına katkısı ortaya konulmuştur.



Şekil 4.13 RO Tabanlı Özellik Seçimiyle Oluşturulan Mö Modellerinin Doğruluk Değerleri

Şekil 4.13'de yer alan doğruluk değerlerine göre, en yüksek doğruluk değeri %98,7 ile ÇKA modeli tarafından elde edilmiştir. Bu yüksek başarım değerini %96,1 doğruluk oranıyla Voting ve LR modelleri takip etmiştir. Ayrıca DVM modeli %92,2, Bagging modeli %90,9 ve DDA,

KNN ile RO modelleri %89,6 ile ortalama bir doğruluk değeri sergilemişlerdir. GA modelinde doğruluk %88,3'e gerilerken, NB modelinde bu oran %85,7 olarak hesaplanmıştır. Doğruluk değerinin en düşük olduğu model ise %84.4 oranıyla KDA'dır. Doğruluk metriğine ek olarak, modellerin duyarlılık, özgüllük ve F1 skoru gibi diğer performans ölçütlerine ilişkin ayrıntılı sonuçlar ise Çizelge 4.7'de sunulmuştur.

Çizelge 4.7 RO Yöntemiyle Seçilen Özelliklerle Eğitilen Modellerin Kesinlik, Duyarlılık ve F1 Skoru Performansları

Modeller	Kesinlik (%)	Duyarlılık (%)	F1 skor (%)
LR	94,9	97,4	96,1
RO	87,5	92,1	89,7
DVM	88,1	97,4	92,5
KNN	94,1	84,2	88,9
NB	88,6	81,6	84,9
GA	87,2	89,5	88,3
ÇKA	97,4	100,0	98,7
DDA	89,5	89,5	89,5
KDA	81,0	89,5	85,0
Bagging	89,7	92,1	90,9
Voting	92,7	100,0	96,2

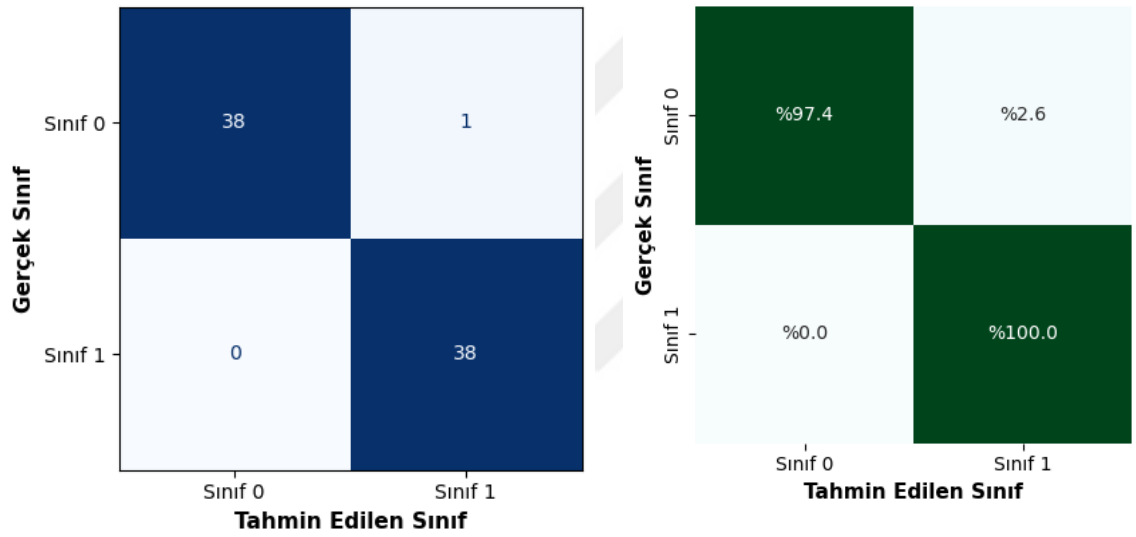
Çizelge 4.7'de sunulan bulgulara göre, kesinlik ölçütü açısından en yüksek başarı %97,4 ile ÇKA modelinde elde edilmiştir. Bu modeli sırasıyla %94,9 ile LR, %94,1 ile KNN ve %92,7 ile Voting modelleri izlemiştir. Bagging ve DDA ise sırasıyla %89,7 ve %89,5 kesinlik oranlarıyla üst sıralarda konumlanmıştır. DVM, NB, RO ve GA modelleri ise sırasıyla %88,1, %88,6, %87,5 ve %87,2 oranlarıyla daha sınırlı düzeyde performans sergilemiştir. Bu ölçüt bakımından en düşük değer, %81 kesinlik oranı ile KDA modelinde gözlemlenmiştir.

Duyarlılık ölçütü açısından yapılan değerlendirmede, en yüksek başarı %100 ile ÇKA ve Voting modellerinde elde edilmiştir. Bu modelleri %97,4 duyarlılık değeriyle LR ve DVM takip etmiştir. Bagging ve RO modelleri %92,1 oranında benzer düzeyde performans sunmuştur. DDA, KDA ve GA modelleri ise %89,5 oranında duyarlılık göstermiştir. KNN %84,2, NB ise %81,6 ile bu ölçüt açısından en düşük değeri elde etmiştir.

F1 skoru ölçütü bakımından yapılan değerlendirmede, en yüksek performans %98,7 ile ÇKA modelinde elde edilmiştir. Bu modeli sırasıyla %96,2 ile Voting ve %96,1 ile LR modelleri

takip etmiştir. DVM %92,5, Bagging %90,9, RO %89,7 ve DDA %89,5 oranında başarılı sonuçlar sunmuştur. KNN ve GA modelleri ise sırasıyla %88,9 ve %88,3 ile orta düzeyde performans sergilemiştir. Bu ölçüt açısından en düşük değerler %85 ile KDA ve %84,9 ile NB modellerinde gözlemlenmiştir.

Sonuçlar genel olarak değerlendirildiğinde, ÇKA modelinin tüm performans metrikleri açısından daha dengeli ve tutarlı bir performans sergilediği tespit edilmiştir. Söz konusu modelin sınıflandırma performansını nicel ve nitel olarak doğrulamak amacıyla Şekil 4.14'de karmaşıklık matrisi hem mutlak hem de normalize edilmiş değerler üzerinden görselleştirilmiştir.



(a) (b)

Şekil 4.14 ÇKA Modelinin Karmaşıklık Matrisi: (a) Mutlak Değerler (b) Normalize Değerler (%)

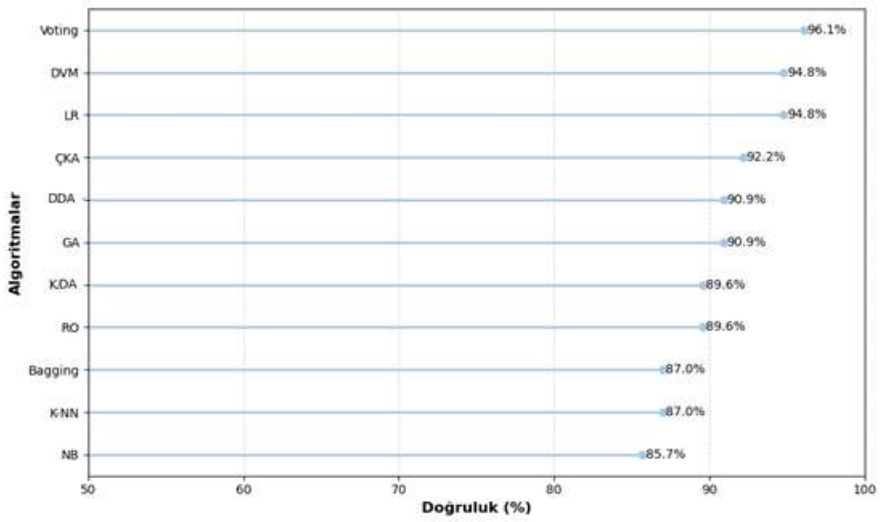
Şekil 4.14 (a)'da mutlak değerlerle sunulan karmaşıklık matrisine göre, gerçek sınıf etiketi 0 olan toplam 39 örneğin 38'sini doğru bir şekilde 'Sınıf 0' olarak sınıflandırılırken, 1 örnek yanlışlıkla 'Sınıf 1' olarak tahmin edilmiştir. Öte yandan, gerçek sınıfı 1 olan 38 örneğin 38'ini başarıyla 'Sınıf 1' olarak tanımlanmıştır. Bu sonuçlar, ÇKA modelinin her iki sınıf için de yüksek genel doğruluk sergilediğini ve yalnızca Sınıf 0'da düşük düzeyde bir yanlış sınıflandırma gerçekleştiğini göstermektedir.

Şekil 4.14 (b)'de normalize edilmiş değerlerle sunulan karmaşıklık matrisi bu durumu yüzdesel oranlarla desteklemektedir. Model, 'Sınıf 0' örneklerini %97,4, 'Sınıf 1' örneklerini ise %100,0

doğruluk oranıyla sınıflandırmıştır. Bu sonuçlar, ÇKA modelinin özellikle 'Sınıf 1' örneklerini tamamen doğru sınıflandırabildiğini ve genel olarak sınıflar arasında oldukça dengeli bir performans ortaya koyduğunu göstermektedir.

4.7 ÖZYİNELEMELİ ÖZELLİK SEÇME-LOJİSTİK REGRESYON TABANLI ÖZELLİK SEÇİMİ SONRASI SINIFLANDIRICILARIN PERFORMANS ANALİZİ

Bu alt bölümde, çevresel tutumların sınıflandırılmasında en anlamlı değişkenleri belirlemek amacıyla RFE-LR tabanlı bir özellik seçimi yöntemi uygulanmıştır. Bu yöntemde, değişkenlerin sınıflandırma üzerindeki katkısı LR modeli aracılığıyla değerlendirilmiş ve en etkili olanlar seçilmiştir. Daha sonra seçilen bu özellikler, sınıflandırma algoritmalarının eğitim aşamasında kullanılmış ve farklı MÖ modelleri üzerindeki etkileri sistematik olarak incelenmiştir. Bu kapsamda, Şekil 4.15'de modellerin doğruluk değerleri karşılaştırmalı olarak sunularak, RFE-LR temelli temelli özellik seçiminin sınıflandırıcıların performansına katkısı ortaya konulmuştur.



Şekil 4.15 RFE-LR Tabanlı Özellik Seçimiyle Oluşturulan Mö Modellerinin Doğruluk Değerleri.

Şekil 4.15'de verilen değerler incelendiğinde, Voting modeli %96,1 doğruluk oranıyla en iyi performansı sergilemiştir. Bunu %94,8'lik doğruluk oranıyla DVM ve LR modelleri takip etmiştir. ÇKA modeli %92,2 doğruluk oranıyla bir sonraki sırada yer alırken, DDA ve GA modelleri %90,9 ile benzer düzeyde performans göstermiştir. KDA ve RO modelleri ise %89,6 doğruluk oranıyla daha düşük sonuçlar sunarken, Bagging ve KNN modelleri %87 ile nispten

sınırlı başarı elde etmiştir. En düşük doğruluk oranı ise %85,7 ile NB modelinde gözlemlenmiştir. Doğruluk metriğine ek olarak, modellerin duyarlılık, özgüllük ve F1 skoru gibi diğer performans ölçütlerine ilişkin ayrıntılı sonuçlar ise Çizelge 4.8.'de sunulmuştur.

Çizelge 4.8 RFE-LR Yöntemiyle Seçilen Özelliklerle Eğitilen Modellerin Kesinlik, Duyarlılık Ve F1 Skoru Performansları.

Modeller	**Kesinlik (%)**	**Duyarlılık (%)**	**F1 skor (%)**
LR	97,2	92,1	94,6
RO	89,5	89,5	89,5
DVM	94,7	94,7	94,7
KNN	86,8	86,8	86,8
NB	88,6	81,6	84,9
GA	87,8	94,7	91,1
ÇKA	92,1	92,1	92,1
DDA	94,3	86,8	90,4
KDA	85,7	94,7	90,0
Bagging	83,3	92,1	87,5
Voting	94,9	97,4	96,1

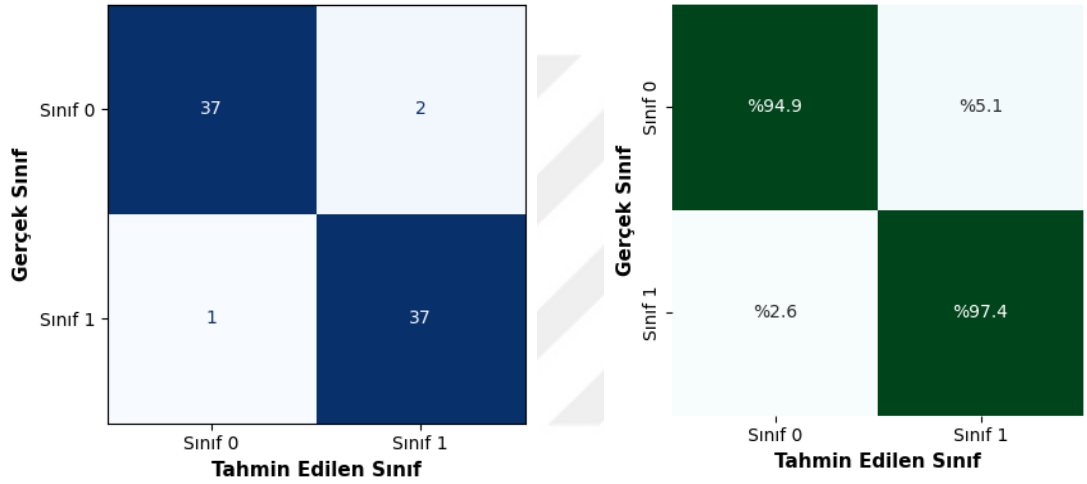
Çizelge 4.8.'de sunulan bulgulara göre, kesinlik ölçütü bakımından en yüksek değer %97,2 ile LR modeli tarafından elde edilmiştir. Bu modeli sırasıyla %94,9 ile Voting, %94,7 ile DVM ve %94,3 ile DDA modelleri takip etmiştir. ÇKA modeli %92,1 oranında kesinlik değeriyle üst sıralarda yer alırken; RO (%89,5), NB (%88,6), GA (%87,8), KNN (%86,8) ve KDA (%85,7) modelleri bu ölçüt açısından nispeten daha düşük performans sergilemiştir. En düşük kesinlik değeri ise %83,3 ile Bagging modelinde gözlemlenmiştir.

Duyarlılık ölçütü açısından değerlendirildiğinde, en yüksek oran %97,4 ile Voting modelinde elde edilmiştir. Bu modeli %94,7 ile DVM, GA ve KDA modelleri takip etmiştir. LR, ÇKA ve Bagging modelleri %92,1 oranında benzer bir performans sergilemiştir. RO modeli oranında duyarlılık sunarken, KNN ve DDA %86,8 ile daha düşük bir başarı elde etmiştir. En düşük değer ise %81,6 ile NB modelinde kaydedilmiştir.

F1 skoru ölçütüne göre, en yüksek değer %96,1 ile yine Voting modeli tarafından elde edilmiştir. Bu modeli sırasıyla %94,7 ile DVM ve %94,6 ile LR modelleri takip etmiştir. ÇKA modeli %92,1 ile yüksek düzeyde bir F1 skoru sunarken, GA (%91,1), DDA (%90,4) ve KDA

(%90,0) onu izlemiştir. RO ve Bagging modelleri sırasıyla %89,5 ve %87,5 değerlerine ulaşırken, KNN (%86,8) ve NB (%84,9) bu ölçüt açısından en düşük performansı sergileyen modeller olmuştur.

Sonuçlar genel olarak değerlendirildiğinde, Voting modelinin tüm performans metrikleri açısından daha dengeli ve tutarlı bir performans sergilediği tespit edilmiştir. Söz konusu modelin sınıflandırma performansını nicel ve nitel olarak doğrulamak amacıyla Şekil 4.16'da karmaşıklık matrisi hem mutlak hem de normalize edilmiş değerler üzerinden görselleştirilmiştir.



(a) (b)

Şekil 4.16 Voting Modelinin Karmaşıklık Matrisi: (a) Mutlak Değerler (b) Normalize Değerler (%)

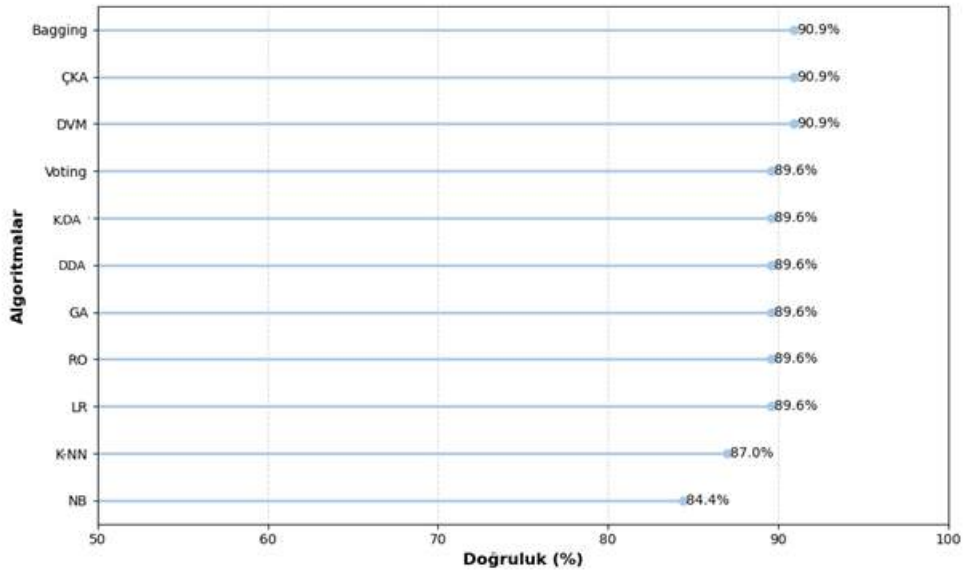
Şekil 4.16 (a)'da mutlak değerlerle sunulan karmaşıklık matrisine göre, gerçek sınıf etiketi 0 olan toplam 39 örneğin 37'sini doğru bir şekilde 'Sınıf 0' olarak sınıflandırılırken, 2 örnek yanlışlıkla 'Sınıf 1' olarak tahmin edilmiştir. Öte yandan, gerçek sınıfı 1 olan 38 örneğin 37'sini başarıyla 'Sınıf 1' olarak tanımlarken, 1 örnek hatalı sınıflandırılmıştır. Bu sonuçlar, Voting modelinin her iki sınıf için de yüksek düzeyde genel doğruluk sergilediğini ve yanlış sınıflandırmaların sınıflar arasında dengeli bir şekilde dağıldığını ortaya koymaktadır.

Şekil 4.16 (b)'de normalize edilmiş değerlerle sunulan karmaşıklık matrisi bu durumu yüzdesel oranlarla desteklemektedir. Model, 'Sınıf 0' örneklerini %94,9, 'Sınıf 1' örneklerini ise %97,4 doğruluk oranıyla sınıflandırmıştır. Bu oranlar, Voting modelinin özellikle 'Sınıf 1' sınıfında

daha yüksek doğruluk sunduğunu ve genel olarak sınıflar arasında tutarlı bir performans sergilediğini göstermektedir.

4.8 ÖZYİNELEMELİ ÖZELLİK SEÇME- RASTGELE ORMAN TABANLI ÖZELLİK SEÇİMİ SONRASI SINIFLANDIRICILARIN PERFORMANS ANALİZİ

Bu alt bölümde, çevresel tutumların sınıflandırılmasında en anlamlı değişkenleri belirlemek amacıyla RFE-RO tabanlı bir özellik seçimi yöntemi uygulanmıştır. Bu yöntemde, değişkenlerin önem düzeyleri RO algoritması aracılığıyla değerlendirilmiş ve sınıflandırma sürecine en fazla katkı sağlayanlar seçilmiştir. Daha sonra seçilen bu özellikler, sınıflandırma algoritmalarının eğitim aşamasında kullanılmış ve farklı MÖ modelleri üzerindeki etkileri sistematik olarak incelenmiştir. Bu kapsamda, Şekil 4.17'de modellerin doğruluk değerleri karşılaştırmalı olarak sunularak, RFE-RO temelli özellik seçiminin sınıflandırıcıların performansına katkısı ortaya konulmuştur.



Şekil 4.17 RFE-RO Tabanlı Özellik Seçimiyle Oluşturulan MÖ Modellerinin Doğruluk Değerleri

Şekil 4.17'de yer alan sonuçlar irdelendiğinde, en yüksek doğruluk değeri %90,9 ile Bagging, ÇKA ve DVM modelleri tarafından elde edilmiştir. Bu modelleri %89,6 ile Voting, KDA, DDA, GA, RO ve LR modelleri takip etmiştir. En düşük başarımlar ise KNN (%87) v NB (%84,4) modellerinde kaydedilmiştir. Doğruluk metriğine ek olarak, modellerin duyarlılık, özgüllük ve F1 skoru gibi diğer performans ölçütlerine ilişkin ayrıntılı sonuçlar ise Çizelge 4.9.'da sunulmuştur.

Çizelge 4.9 RFE-RO Yöntemiyle Seçilen Özelliklerle Eğitilen Modellerin Kesinlik, Duyarlılık Ve F1 Skoru Performansları

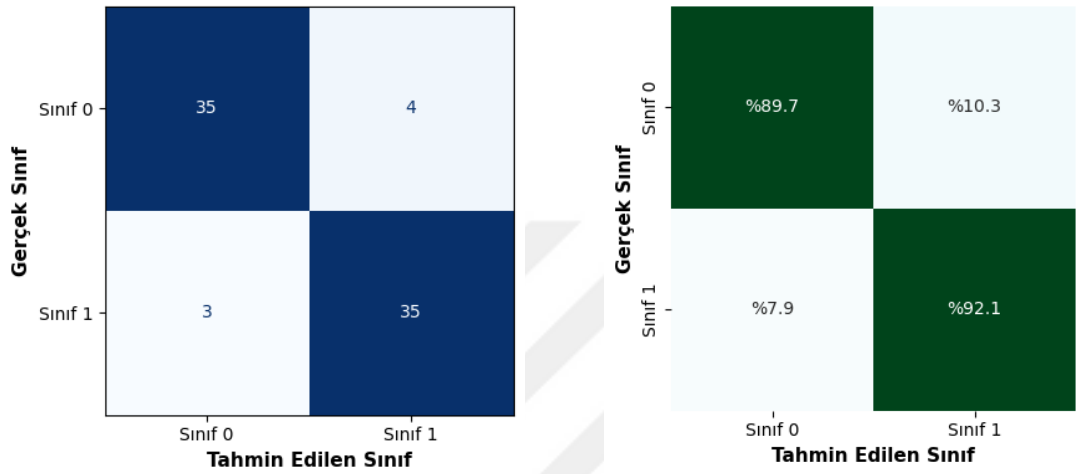
Algoritmalar	Kesinlik (%)	Duyarlılık (%)	F1 skor (%)
LR	89,5	89,5	89,5
RO	87,5	92,1	89,7
DVM	89,7	92,1	90,9
KNN	88,9	84,2	86,5
NB	90,6	76,3	82,9
GA	87,5	92,1	89,7
ÇKA	91,9	89,5	90,7
DDA	89,5	89,5	89,5
KDA	89,5	89,5	89,5
Bagging	89,7	92,1	90,9
Voting	91,7	86,8	89,2

Çizelge 4.9'da sunulan bulgulara göre, kesinlik ölçütü bakımından en yüksek değer %91,9 ile ÇKA modeli tarafından elde edilmiştir. Bu modeli sırasıyla %91,7 ile Voting ve %90,6 ile NB modelleri takip etmiştir. DVM ve Bagging modelleri %89,7 oranında kesinlik sergilerken; KDA, DDA ve LR modelleri %89,5 ile benzer düzeyde performans göstermiştir. KNN modeli %88,9 oranında bir kesinlik değerine sahipken, GA ve RO modelleri %87,5 ile bu ölçüt açısından en düşük sonuçları vermiştir.

Duyarlılık ölçütü açısından değerlendirildiğinde, en yüksek oran %92,1 ile RO, DVM, GA ve Bagging modellerinde elde edilmiştir. Bu modelleri %89,5 ile ÇKA, DDA, KDA ve LR takip etmiştir. Voting modeli %86,8 oranında duyarlılık sergilerken, KNN modeli %84,2 ile daha düşük bir performans göstermiştir. En düşük duyarlılık değeri ise %76,3 ile NB modelinde kaydedilmiştir.

F1 skoru ölçütüne göre, en yüksek performans %90,9 ile DVM ve Bagging modellerinde elde edilmiştir. Bu modelleri %90,7 ile ÇKA modeli takip etmiştir. RO ve GA modelleri %89,7, LR, DDA ve KDA ise %89,5 F1 skoru ile benzer düzeyde sonuçlar sunmuştur. Voting modeli %89,2 ile bu modellerin bir miktar gerisinde kalmıştır. En düşük F1 skor değerleri ise %86,5 ile KNN ve %82,9 ile NB modellerinde kaydedilmiştir.

Sonuçlar genel olarak değerlendirildiğinde, DVM, Bagging ve ÇKA modellerinin tüm performans metrikleri açısından daha dengeli ve tutarlı bir performans sergilediği tespit edilmiştir. Söz konusu modellerin sınıflandırma performansını nicel ve nitel olarak doğrulamak amacıyla Şekil 4.18'de DVM ve Bagging modellerine, Şekil 4.19'da ÇKA modeline ait karmaşıklık matrisleri hem mutlak hem de normalize edilmiş değerler üzerinden görselleştirilmiştir.

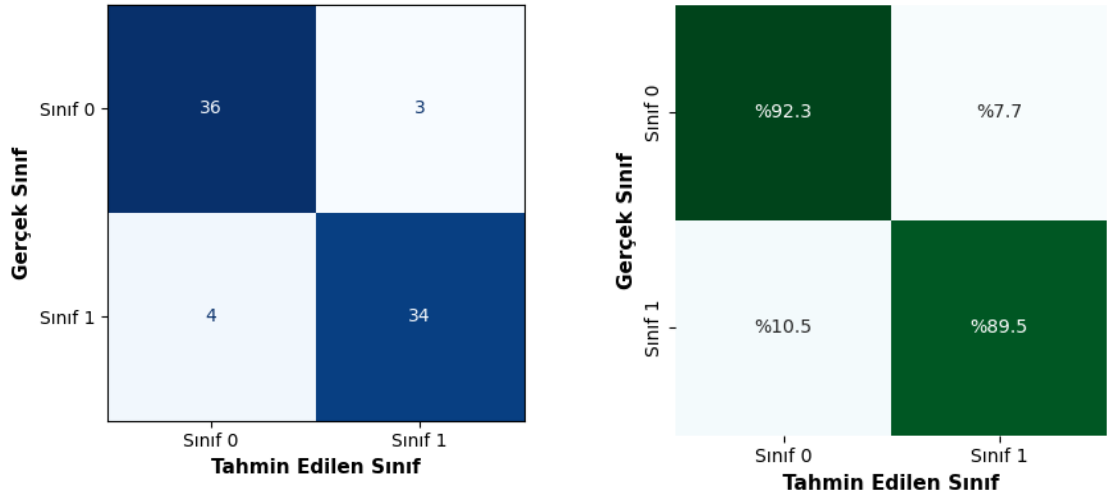


(a) (b)

Şekil 4.18 DVM ve Bagging Modellerinin Karmaşıklık Matrisi: (a) Mutlak Değerler (b) Normalize Değerler (%)

Şekil 4.18 (a)'da mutlak değerlerle sunulan karmaşıklık matrisine göre, gerçek sınıf etiketi 0 olan toplam 39 örneğin 35'ini doğru bir şekilde 'Sınıf 0' olarak sınıflandırılırken, 4 örnek yanlışlıkla 'Sınıf 1' olarak tahmin edilmiştir. Öte yandan, gerçek sınıfı 1 olan 38 örneğin 35'ini başarıyla 'Sınıf 1' olarak tanımlarken, 3 örnek hatalı sınıflandırılmıştır. Bu sonuçlar, DVM ve Bagging modellerinin her iki sınıf için de yüksek düzeyde genel doğruluk sergilediğini ve yanlış sınıflandırmaların sınıflar arasında nispeten dengeli bir şekilde dağıldığını ortaya koymaktadır.

Şekil 4.18 (b)'de normalize edilmiş değerlerle sunulan karmaşıklık matrisi bu durumu yüzdesel oranlarla desteklemektedir. Model, 'Sınıf 0' örneklerini %89,7, 'Sınıf 1' örneklerini ise %92,1 doğruluk oranıyla sınıflandırmıştır. Bu oranlar, DVM ve Bagging modellerinin özellikle 'Sınıf 1' sınıfında daha yüksek doğruluk sunduğunu ve genel olarak sınıflar arasında tutarlı bir performans sergilediğini göstermektedir.



(a) (b)

Şekil 4.19 ÇKA Modelinin Karmaşıklık Matrisi: (a) Mutlak Değerler (b) Normalize Değerler (%)

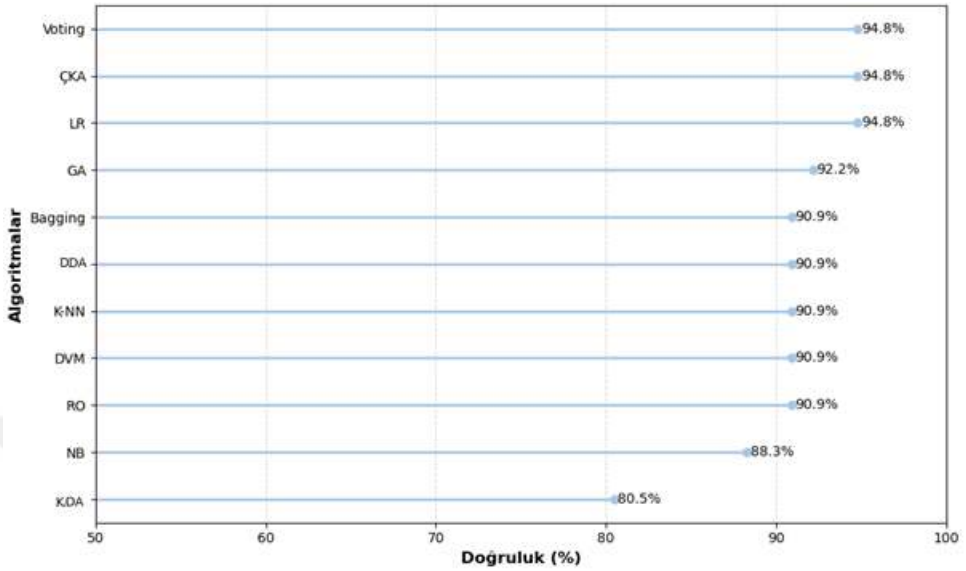
Şekil 4.19 (a)'da mutlak değerlerle sunulan karmaşıklık matrisine göre, gerçek sınıf etiketi 0 olan toplam 39 örneğin 36'sını doğru bir şekilde 'Sınıf 0' olarak sınıflandırılırken, 3 örnek yanlışlıkla 'Sınıf 1' olarak tahmin edilmiştir. Öte yandan, gerçek sınıfı 1 olan 38 örneğin 34'ünü başarıyla 'Sınıf 1' olarak tanımlarken, 4 örnek hatalı sınıflandırılmıştır. Bu sonuçlar, ÇKA modelinin her iki sınıf için de yüksek düzeyde genel doğruluk sergilediğini ve yanlış sınıflandırmaların sınıflar arasında nispeten dengeli bir şekilde dağıldığını ortaya koymaktadır.

Şekil 4.19 (b)'de normalize edilmiş değerlerle sunulan karmaşıklık matrisi bu durumu yüzdesel oranlarla desteklemektedir. Model, 'Sınıf 0' örneklerini %92,3, 'Sınıf 1' örneklerini ise %89,5 doğruluk oranıyla sınıflandırmıştır. Bu oranlar, ÇKA modelinin özellikle 'Sınıf 0' sınıfında daha yüksek doğruluk sunduğunu ve genel anlamda sınıflar arasında tutarlı bir performans sergilediğini göstermektedir.

4.9 VARYANS EŞİĞİ TABANLI ÖZELLİK SEÇİMİ SONRASI SINIFLANDIRICILARIN PERFORMANS ANALİZİ

Bu alt bölümde, çevresel tutumların sınıflandırılmasında en anlamlı değişkenleri belirlemek amacıyla varyans eşiğine dayalı bir filtreleme tabanlı özellik seçimi yöntemi uygulanmıştır. U yöntemde eşik değerin altında varyansa sahip özellikler veri setinden çıkarılarak, bilgi değeri yüksek özellikler belirlenmiştir. Daha sonra belirlene bu özellikler, sınıflandırma algoritmalarının eğitim aşamasında kullanılmış ve farklı MÖ modelleri üzerindeki etkileri

sistematik olarak incelenmiştir. Bu kapsamda, Şekil 4.20'de modellerin doğruluk değerleri karşılaştırmalı olarak sunularak, varyans eşiği tabanlı özellik seçiminin sınıflandırıcıların performansına katkısı ortaya konulmuştur.



Şekil 4.20 Varyans Eşiği Tabanlı Özellik Seçimiyle Oluşturulan MÖ Modellerinin Doğruluk Değerleri.

Şekil 4.20'de verilen doğruluk değerleri göz önüne alındığında, Voting, ÇKA ve LR modellerinde %94,8 ile en yüksek doğruluk değerine ulaşılmıştır. Bu doğruluk değerini %92,2 ile GA modeli takip etmiştir. Bunu %90,9 ile Bagging, DDA, KNN, DVM ve RO modelleri izlemiştir. En düşük doğruluk değerleri ise NB (%88,3) ve KDA (%80,5) modellerinde kaydedilmiştir. Doğruluk metriğine ek olarak, modellerin duyarlılık, özgüllük ve F1 skoru gibi diğer performans ölçütlerine ilişkin ayrıntılı sonuçlar ise Çizelge 4.10'da sunulmuştur.

Çizelge 4.10 Varyans Eşiği Yöntemiyle Seçilen Özelliklerle Eğitilen Modellerin Kesinlik, Duyarlılık Ve F1 Skoru Performansları.

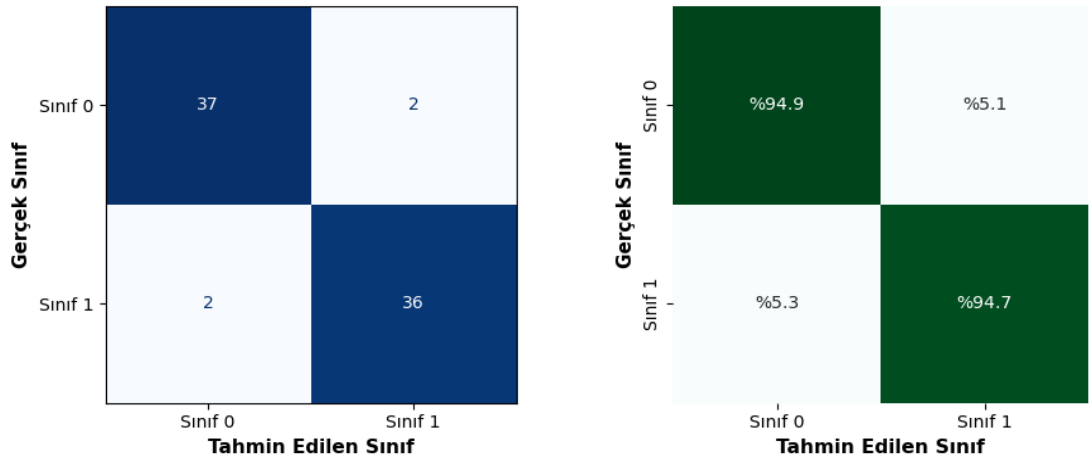
Algoritmalar	Kesinlik (%)	Duyarlılık (%)	F1 skor (%)
LR	94,7	94,7	94,7
RO	87,8	94,7	91,1
DVM	87,8	94,7	91,1
KNN	94,3	86,8	90,4
NB	89,2	86,8	88,0
GA	90,0	94,7	92,3
ÇKA	94,7	94,7	94,7
DDA	91,9	89,5	90,7
KDA	78,0	84,2	81,0
Bagging	89,7	92,1	90,9
Voting	92,5	97,4	94,9

Çizelge 4.10'da sunulan bulgular doğrultusunda, kesinlik ölçütü açısından en yüksek değer %94,7 ile LR ve ÇKA modelleri tarafından elde edilmiştir. Bu modelleri sırasıyla %94,3 ile KNN, %92,5 ile Voting ve %91,9 ile DDA modelleri takip etmiştir. GA (%90,0), Bagging (%89,7), NB (%89,2), RO ve DVM (%87,8) modelleri ise nispeten daha düşük kesinlik değerleri ortaya koymuştur. En düşük kesinlik performansı ise %78,0 ile KDA modelinde gözlemlenmiştir.

Duyarlılık ölçütü açısından en yüksek değer %97,4 ile Voting modeli tarafından elde edilmiştir. Bu modeli %94,7 oranıyla LR, ÇKA, GA, RO ve DVM modelleri takip etmiştir. Bagging modeli %92,1 duyarlılık göstermiştir. DDA modeli %89,5 oranında bir performans sergilerken, KNN ve NB modelleri %86,8 ile yakın bir performans sergilemiştir. En düşük duyarlılık değeri ise %84,2 ile KDA modelinde kaydedilmiştir.

F1 skoru ölçütüne göre, en yüksek değer %94,9 ile Voting modelinde elde edilmiştir. Bu modeli %94,7 ile LR ve ÇKA modelleri izlemiştir. Bu modelleri GA modeli %92,3, RO ve DVM modelleri ise %91,1 F1 skor değeriyle takip etmiştir. Bagging modeli %90,9, DDA modeli %90,7 ve KNN modeli %90,4 F1 skoru değerleri ile birbirine yakın performans sergilemişlerdir. NB modeli %88,0 ile bu modellerin gerisinde kalırken, en düşük F1 skoru %81,0 ile KDA modelinde elde edilmiştir.

Sonuçlar genel olarak değerlendirildiğinde, ÇKA, LR ve Voting modellerinin tüm performans metrikleri açısından daha dengeli ve tutarlı bir performans sergilediği tespit edilmiştir. Söz konusu modellerin sınıflandırma performansını nicel ve nitel olarak doğrulamak amacıyla Şekil 4.21'de ÇKA ve LR modellerine, Şekil 4.22'de Voting modeline ait karmaşıklık matrisleri hem mutlak hem de normalize edilmiş değerler üzerinden görselleştirilmiştir.

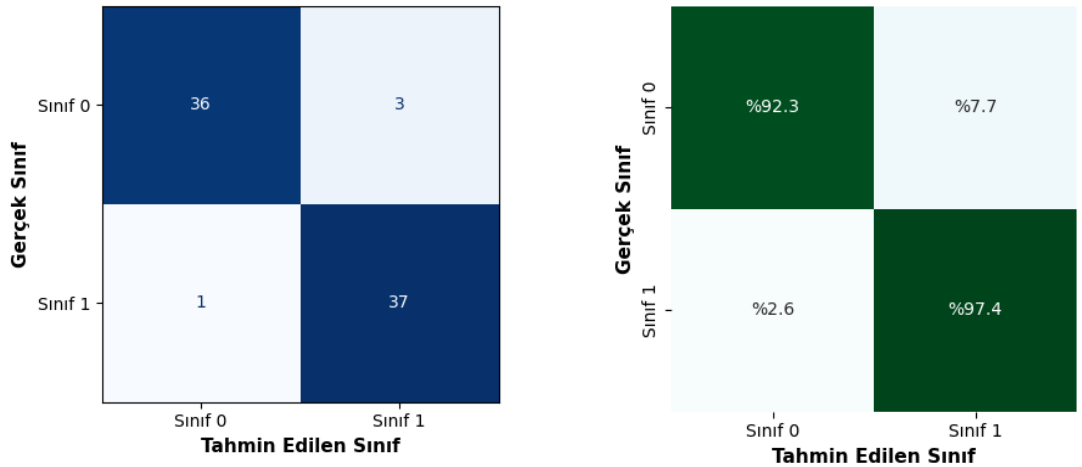


(a) (b)

Şekil 4.21 ÇKA ve LR Modelleri İçin Karmaşıklık Matrisi: (A) Mutlak Değerler, (B) Normalize Değerler (%)

Şekil 4.21 (a)'da mutlak değerlerle sunulan karmaşıklık matrisine göre, gerçek sınıf etiketi 0 olan toplam 39 örneğin 37'sini doğru bir şekilde 'Sınıf 0' olarak sınıflandırılırken, 2 örnek yanlışlıkla 'Sınıf 1' olarak tahmin edilmiştir. Öte yandan, gerçek sınıfı 1 olan 38 örneğin 36'sını başarıyla 'Sınıf 1' olarak tanımlarken, 2 örnek hatalı sınıflandırılmıştır. Bu sonuçlar, ÇKA ve LR modellerinin her iki sınıfta da yüksek doğruluk sağladığını göstermektedir.

Şekil 4.21 (b)'de normalize edilmiş değerlerle sunulan karmaşıklık matrisi bu durumu yüzdesel oranlarla desteklemektedir. Model, 'Sınıf 0' örneklerini %94,9, 'Sınıf 1' örneklerini ise %94,7 doğruluk oranıyla sınıflandırmıştır. Bu oranlar, ÇKA ve LR modellerinin her iki sınıfı da benzer doğruluk düzeylerinde sınıflandırdığını ve özellikle 'Sınıf 0' sınıfında hafifçe daha yüksek doğruluk sağladığını göstermektedir.



(a) (b)

Şekil 4.22 Voting Modeli İçin Karmaşıklık Matrisi: (a) Mutlak değerler, (b) Normalize Değerler (%)

Şekil 4.22 (a)'da mutlak değerlerle sunulan karmaşıklık matrisine göre, gerçek sınıf etiketi 0 olan toplam 39 örneğin 36'sını doğru bir şekilde 'Sınıf 0' olarak sınıflandırılırken, 3 örnek yanlışlıkla 'Sınıf 1' olarak tahmin edilmiştir. Öte yandan, gerçek sınıfı 1 olan 38 örneğin 37'sini başarıyla 'Sınıf 1' olarak tanımlarken, 1 örnek hatalı sınıflandırılmıştır. Bu sonuçlar, Voting modelinin her iki sınıfta da yüksek doğruluk sağladığını; ancak hataların, özellikle Sınıf 0'ta daha belirgin olduğunu göstermektedir.

Şekil 4.22 (b)'de normalize edilmiş değerlerle sunulan karmaşıklık matrisi bu durumu yüzdesel oranlarla desteklemektedir. Model, 'Sınıf 0' örneklerini %92,3, 'Sınıf 1' örneklerini ise %97,4 doğruluk oranıyla sınıflandırmıştır. Bu oranlar, Voting modelinin özellikle 'Sınıf 1' sınıfında daha yüksek doğruluk sunduğunu göstermektedir.

BÖLÜM 5

TARTIŞMA

Gerçekleştirilen bu tez çalışması, özellik seçim yöntemlerinin MÖ temelli karar destek sistemlerinin performansları üzerindeki etkilerini kapsamlı bir şekilde analiz etmeyi amaçlamaktadır. Özellikle yüksek boyutlu veri kümelerinde, öznitelik seçimi; model karmaşıklığını azaltma, hesaplama verimliliğini artırma ve genelleme yeteneğini geliştirme açısından kritik bir ön işleme adımı olarak kabul edilmektedir. Bu doğrultuda gerçekleştirilen çalışmada, açık erişimli bir çevresel tutum veri seti kullanılarak, mühendislik temelli, veri odaklı ve çok aşamalı bir MÖ yaklaşımı önerilmiştir. Özellik seçimi süreci, bu sistemin temel bileşenlerinden biri olarak sistematik biçimde değerlendirilmiştir.

Bu çerçevede, ANOVA, Ki kare, EA, RO gibi farklı istatistiksel ve algoritmik temellere dayanan on farklı özellik seçim yöntemi kullanılmıştır. Bu yöntemlerle belirlenen optimal özellik kümeleri, LR, DVM, Voting gibi çeşitli sınıflandırma algoritmaları ile edilmiştir. Modelleme sürecinde, istatistiksel genellenebilirliği sağlamak üzere veri kümesi %80 eğitim ve %20 test olacak şekilde alt kümelere ayrılmıştır. Eğitim aşamasında ise 5-kat çapraz doğrulama yöntemi uygulanarak modellerin kararlılığı sayısal metriklerle değerlendirilmiştir. Bu yaklaşım, aşırı uyum riskinin minimize edilmesine ve performans metriklerinin güvenilirliğinin artırılmasına olanak tanımıştır. Modellerin başarımı; doğruluk, kesinlik, duyarlılık ve F1 skoru gibi çok boyutlu ölçütler çerçevesinde kapsamlı olarak incelenmiştir. Elde edilen bulgular, çevresel tutum sınıflandırmasında optimal algoritma-özellik seçim kombinasyonlarını bütüncül bir şekilde ortaya koymuştur.

Yapılan analizlerde, ANOVA ile seçilen özelliklerle eğitilen NB ve GA modelleri, %84,4 doğruluk oranıyla en yüksek performansı göstermiştir. Bu modeller ayrıca %81,6 kesinlik ve duyarlılık düzeylerine ulaşarak dengeli bir sınıflandırma başarımı ortaya koymuştur. Bununla birlilte, her iki modelin de %83,8 F1 skoru elde etmesi, sınıflandırma performanslarının hem duyarlılık hem de kesinlik açısından tutarlı olduğunu doğrulamaktadır. Öte yandan, duyarlılık

metriği özelinde değerlendirildiğinde, en yüksek değer %84,2 ile KNN modeli tarafından elde edilmiştir. Ancak çoklu performans ölçütlerinin harmonik ortalaması olan F1 skoru dikkate alındığında NB ve GA modellerinin genel anlamda daha tutarlı sonuçlar verdiği tespit edilmiştir. Elde edilen bulgular genel olarak değerlendirildiğinde, ANOVA'nın özellikler arasındaki farklara dayalı seçme yapısının, belirli algoritmalarla bir araya geldiğinde sınıflandırma başarımını anlamlı ölçüde artırabileceğini göstermektedir.

Ki kare testi uygulandığında, genel sınıflandırma performansı açısından en başarılı sonuçların GA modeli tarafından elde edildiği gözlemlenmiştir. GA modeli, %85,7 doğruluk, %86,5 kesinlik, %84,2 duyarlılık ve %85,3 F1 skoru ile tüm metriklerde dengeli ve tutarlı bir performans sergilemiştir. Özellikle %86,5'lik kesinlik değeriyle bu metrikte en başarılı model olmuştur. Karşılaştırmalı analizler, duyarlılık metriği açısından en yüksek değerin (%89,5) DDA modeli tarafından sağlandığını göstermekle birlikte, çoklu performans ölçütlerinin harmonik ortalaması olan F1 skoru dikkate alındığında GA modelinin diğer algoritmalara göre daha istikrarlı ve bütünsel bir sınıflandırma performansı sunduğu tespit edilmiştir. Bu sonuçlar Ki kare testi ile seçilen özelliklerin GA gibi topluluk tabanlı algoritmalarla kombinasyon halinde kullanılması durumunda hem yüksek sınıflandırma doğruluğu hem de metrikler arası denge sağlanabileceğini ortaya koymaktadır.

EA yöntemiyle yapılan özellik seçimi sonrası DVM modeli, Ki kare tabanlı özellik seçimiyle eğitilen GA modeliyle birebir aynı performansı (%85,7 doğruluk, %86,5 kesinlik, %84,2 duyarlılık ve %85,3 F1 skoru) göstermiştir. Bu bulgu, farklı özellik seçim yöntemlerinin benzer sınıflandırıcı kombinasyonlarıyla aynı başarıyı üretebileceğini ortaya koymaktadır. Bununla birlikte, kesinlik metriği açısından DVM modelinden daha yüksek performans sergileyen modeller bulunsa da, genel performansı yansıtan F1 skoru metriğine göre en yüksek başarımın DVM modelinde olduğu belirlenmiştir. Bu bulgular, EA yönteminin özellikle DVM gibi sınır tabanlı algoritmalarla birlikte kullanıldığında sınıflandırma başarımını artırabildiğini göstermektedir.

LASSO yöntemiyle seçilen özelliklerle Voting modeli %96,1 doğruluk ve %96,1 F1 skoru ile en yüksek performansı göstermiştir. %94,9 kesinlik ve %97,4 duyarlılık değerleri de dikkate alındığında, bu modelin hem pozitif hem de negatif sınıflar arasında güçlü bir denge sağladığı anlaşılmaktadır. Bu durum, LASSO yönteminin düzenlileştirme temelli yapısının topluluk öğrenme algoritmalarıyla birlikte başarılı sonuçlar doğurduğunu göstermektedir.

Karşılıklı bilgi tabanlı özellik seçimi uygulandığında, DVM ve Voting modelleri de benzer şekilde yüksek performans göstermiştir. Her iki model için %92,2 doğruluk, %92,1 kesinlik, %92,1 duyarlılık ve %92,1 F1 skoru değerleri elde edilmiştir. Bu sonuçlar, karşılıklı bilginin özellikler arasındaki doğrusal olmayan ilişkileri yakalama kapasitesinin, hem karar sınırlarını etkili şekilde tanımlayabilen DVM gibi algoritmalar hem de topluluk öğrenme temelli modeller açısından başarılı bir özellik seçimi sağladığını göstermektedir. Bununla birlikte, kesinlik metriği açısından en yüksek değer %96,9 ile LR modeli tarafından elde edilmesine rağmen, genel performans değerlendirildiğinde DVM ve Voting modelleri bu özellik seçim yöntemiyle daha tutarlı ve dengeli sonuçlar vermiştir. Elde edilen bulgular genel olarak değerlendirildiğinde, karşılıklı bilgiye dayalı özellik seçiminin, farklı yapısal özelliklere sahip sınıflandırma algoritmalarıyla birlikte uygulandığında, sınıflandırıcı performansını istikrarlı ve dengeli bir biçimde optimize edebileceğini göstermektedir.

TBA tabanlı özellik seçimi uygulanan sınıflandırma modelleri arasında en yüksek performans ÇKA modeli tarafından elde edilmiştir. Model, %98,7 doğruluk, %97,4 kesinlik, %100,0 duyarlılık ve %98,7 F1 skoru elde ederek tüm değerlendirme metrikleri açısından üstün bir performans sergilemiştir. Elde edilen sonuçlar, özellikle TBA yöntemiyle boyut indirgeme gerçekleştirildiğinde ÇKA algoritmasının öne çıktığını ve diğer özellik seçimi yöntemleriyle elde edilenlere kıyasla daha yüksek bir performans sunduğunu göstermektedir. Bu durum, TBA'nın veri yapısındaki anlamlı varyansı koruyarak boyut indirgeme sağlaması ve ÇKA gibi topluluk temelli algoritmalarla birlikte kullanıldığında sınıflandırma başarımını önemli ölçüde artırabildiğini ortaya koymaktadır.

Benzer şekilde, RO tabanlı özellik seçimi uygulandığında da en yüksek performans ÇKA modeli ile elde edilmiştir. TBA ile ulaşılan doğruluk ve sınıflandırma metriklerinin RO yöntemiyle de aynı model (ÇKA) tarafından tekrarlanması, RO'nun da etkili özellik seçimi sağlayarak yüksek ve tutarlı bir sınıflandırma performansı elde edilmesine katkı sunduğunu göstermektedir.

RFE-LR tabanlı özellik seçimi uygulanması durumunda en yüksek sınıflandırma performansı, %96,1 doğruluk oranı ile Voting modeli tarafından elde edilmiştir. Model, %97,4 duyarlılık ve %96,1 F1 skoru değerleriyle dengeli bir performans göstermesine rağmen, kesinlik metriğinde nispeten daha düşük bir başarım elde etmiştir. Kesinlik metriği açısından en yüksek oran %97,2 ile LR modeline ait olsa da, tüm değerlendirme ölçütleri birlikte ele alındığında Voting modeli,

RFE-LR yöntemiyle en tutarlı ve etkili sınıflandırma başarımını sunan model olarak öne çıkmıştır. Bu bulgular, RFE-LR yönteminin anlamlı öznitelikleri başarıyla belirleyerek özellikle topluluk temelli modellerin genel sınıflandırma başarımını optimize etmede katkı sağladığını ortaya koymaktadır.

RFE-RO tabanlı özellik seçimi uygulanması durumunda ise Bagging ve DVM modelleri %90,9 doğruluk, %91,9 kesinlik, %92,1 duyarlılık ve %90,9 F1 skoru değerleriyle dikkat çekici bir performans sergilemiştir. Kesinlik metriği açısından en yüksek değer %90,7 ile ÇKA modeli tarafından elde edilmiş, duyarlılık metriği bakımından ise benzer düzeyde performans sergileyen başka modellerin de bulunduğu gözlemlenmiştir. Ancak genel model başarımını yansıtan F1 skoru dikkate alındığında, DVM ve Bagging modelleri ön plana çıkmaktadır. Bu bulgular, RFE-RO yönteminin özellikle sınıflar arası dengeyi koruyarak etkili öznitelikler belirleyebildiğini ve topluluk temelli algoritmalarla bir araya geldiğinde yüksek sınıflandırma başarımı sağlayabildiğini ortaya koymaktadır.

Varyans eşiği tabanlı özellik seçimi uygulanması durumunda, tüm performans metrikleri göz önünde tutularak yapılan analizlerde Voting algoritması en yüksek başarımı sergilemiştir. Model, %94,8 doğruluk, %92,5 kesinlik, %97,4 duyarlılık ve %94,9 F1 skoru elde etmiştir. Özellikle duyarlılık ve F1 skoru metriklerinde elde edilen değerler, modelin sınıflar arası dengeyi korumadaki başarısını ortaya koymaktadır. Bununla birlikte, ÇKA ve LR modelleri doğruluk metriği açısından Voting modeliyle aynı performansı sergilemiştir. Ancak Voting algoritmasının tüm performans ölçütlerindeki dengeli dağılımı ve varyans temelli özellik seçimiyle uyumlu yapısı, onu diğer modeller arasında öne çıkarmıştır. Bu durum, topluluk öğrenme yöntemlerinin, özellikle de Voting algoritmasının, varyans analizine dayalı özellik seçimi süreçlerinde daha kararlı ve genellenebilir sonuçlar üretebileceğini göstermektedir.

Sonuçlar genel olarak değerlendirildiğinde, TBA ve RO tabanlı özellik seçim yöntemleriyle eğitilen ÇKA modelinin, doğruluk metriği açısından %98,7 gibi oldukça yüksek bir sınıflandırma başarımı sunduğu görülmüştür. Bu bulgu, sınıflandırma sistemlerinin performansında yalnızca güçlü algoritmaların değil, aynı zamanda uygun özellik seçim stratejilerinin de kritik bir rol oynadığını açıkça ortaya koymaktadır. Özellikle mühendislik uygulamalarında; veri yoğunluğu, yapısal karmaşıklık ve hesaplama kaynaklarının etkin kullanımı gibi etkenler göz önüne alındığında, özellik seçimi süreci, modelleme başarımını belirleyen temel adımlardan biri olarak öne çıkmaktadır.

Bu çalışmada kullanılan TBA yöntemi, veri kümesindeki anlamlı varyansı koruyarak boyut indirgeme sağlarken; RO yöntemi, değişkenler arası etkileşimleri dikkate alarak önemli özniteliklerin önceliklendirilmesine olanak tanımıştır. Her iki yöntem de bilgi kaybını en aza indirerek modelin genellenebilirliğini artırmış; özellikle topluluk temelli güçlü sınıflayıcılarla (örneğin ÇKA) bir araya geldiğinde sınıflandırma başarımını kayda değer biçimde iyileştirmiştir. Bu durum, mühendislik alanında karar destek sistemlerinin tasarımında, algoritma seçimi kadar probleme özgü uygun özellik seçim yönteminin belirlenmesinin de sistem başarımı açısından vazgeçilmez bir bileşen olduğunu göstermektedir.

Sonuç olarak, bu çalışma, farklı yapısal temellere sahip özellik seçim yöntemlerinin, uygun sınıflayıcı algoritmalarla etkili biçimde bütünleştirilerek sınıflandırma performansının nasıl optimize edilebileceğini ortaya koymakta; mühendislik temelli karar destek sistemlerinde özellik seçiminin stratejik önemini güçlü biçimde vurgulamaktadır.

KAYNAKLAR

Agarwal A K, Bhatt C, Tiwari R G ve Gupta N (2024) Classifying Environmental Attitudes Using Explainable Machine Learning Models: A Study on Interpretability in Environmental Psychology. *13th International Conference on System Modeling & Advancement in Research Trends (SMART),* Aralık 2024, *Proceedings of SMART 2024*, IEEE, 267–273.

Akgün M ve Barlık N (2023) Makine Öğrenmesi Algoritmaları Kullanılarak Hava Kalitesi İndeksinin Tahmini. *Avrupa Bilim ve Teknoloji Dergisi*, (51): 97-107.

Akkuş M M (2024) Çevre Kirliliğinin Önlenmesinde Çevre Vergilerinin Etkinliği: Hesaplanabilir Genel Denge Modeli. *Doktora Tezi*, Kocaeli Üniversitesi, Sosyal Bilimleri Enstitüsü, İktisat Anabilim Dalı, Kocaeli, 226 s.

Aksu K (2023) Makine öğrenmesi yöntemleri kullanılarak üç boyutlu nokta bulutlarının sınıflandırılması. *Yüksek Lisans Tezi*, İstanbul Teknik Üniversitesi, Lisansüstü Eğitim Enstitüsü, Geomatik Mühendisliği Anabilim Dalı, İstanbul, 51 s.

Al-Shamaa Z Z R (2020) Dengesiz veri kümesinde nadir sınıfın sınıflandırılmasını geliştirmek için alt örnekleme modelleri. *Doktora Tezi*, Altınbaş Üniversitesi, Lisansüstü Çalışmalar Enstitüsü, Elektrik ve Bilgisayar Mühendisliği Ana Bilim Dalı, İstanbul, 73 s.

Arifin H H, Kawamura K, Ong H K R, Biber B,Chimplee N ve Pavalkis S (2023) Model-Based FMEA & FTA with Case-Based Reasoning. *INCOSE International Symposium,* 33(1): 123-134.

Arslan R U ve Yapıcı İ Ş (2024) Farklı Örnekleme Tekniklerine ve Farklı Sınıflandırıcılara Dayanarak Kalp Yetmezliği Tanılı Hastaların Sağkalımlarının İncelenmesi. *EMO Bilimsel Dergi*, 14(2): 35-47.

Auliyani D, Setiawan O, Nugroho H Y S H, Wahyuningrum N, Hardjo K S, Videllisa G A ve Ardiyanti N (2024) Micro Hydro Power Site Characterization in Indonesia: Variable Optimization for Site Selection Using GeoDetector and RFE-Random Forest. In *IOP Conference Series: Earth and Environmental Science*, 1357(1): 012025.

Ayan S ve Bilgin T T (2024) Uyku Sağlığı ile Yaşam Tarzı Arasındaki İlişkinin PCA, Naive Bayes ve Rastgele Orman Ağaçları Yöntemleri ile İncelenmesi ve Karşılaştırılması. *Uluslararası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi*, 8(1): 41-56.

Beiser-McGrath L F ve Huber R A (2018) Assessing the relative importance of psychological and demographic factors for predicting climate and environmental attitudes. *Climatic Change*, 149(3–4): 335–347.

KAYNAKLAR (devam ediyor)

Biegelmeyer U H, Torcheto M, Craco T, Moreira L F ve Pozzo D N (2024) Explorando horizontes sustentáveis: um olhar sobre as energias renováveis. *Journal on Innovation and Sustainability RISUS*, 15(4): 22-41.

Bilgin A A (2022) Vücut Yağ Yüzdesi Tahmini İçin Özellik Seçim Yöntemlerinin Karşılaştırılması. *Yüksek Lisans Tezi*, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Elektrik Elektronik Mühendisliği Anabilim Dalı, Sakarya, 80 s.

Biricik G (2011) Metin sınıflama için yeni bir özellik çıkarım yöntemi. *Doktora Tezi*, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Ana Bilim Dalı, İstanbul, 168 s.

Bishop C M (2006) *Pattern Recognition and Machine Learning*, Bishop C M (Ed.), 1. Baskı, ISBN: 978-0-387-31073-2, Springer, New York,

Bradshaw C J ve Brook B W (2014) Human population reduction is not a quick fix for environmental problems. *Proceedings of the National Academy of Sciences*, 111(46): 16610–16615.

Breiman L (2001) Random forests. *Machine learning*, 45(1): 5-32.

Budak H (2018) Özellik Seçim Yöntemleri ve Yeni Bir Yaklaşım. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, **(**): **-**.

Bulut G ve Özer M A (2024) Ekolojik Modernleşmeye İklim Değişikliği ve "Çevresel Ayak İzleri" Üzerinden Bakmak. *Çankırı Karatekin Üniversitesi İktisadi Ve İdari Bilimler Fakültesi Dergisi*, 14(3), 752-778.

Cover T M and Hart P E (1967) Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1): 21-27.

Chen T ve Guestrin C (2016) XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2016)*, San Francisco, ABD, 785–794.

Chen X, Cao L, Cao Z ve Zhang H (2024) A multi-feature stock price prediction model based on multi-feature calculation, LASSO feature selection, and Ca-LSTM network. *Connection Science*, 36(1): 2286188.

Choudhury N, Mukherjee R, Yadav R, Liu Y ve Wang W (2024) Can machine learning approaches predict green purchase intention?-A study from Indian consumer perspective. *Journal of Cleaner Production*, 456: 142218.

Cristianini N ve Shawe-Taylor J (2000) *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, 1. Baskı, Cambridge University Press, Cambridge,

KAYNAKLAR (devam ediyor)

Çelikkıran A (1997) Çevre Sorunları ve Eğitimi (Çevre Konusunda Formatör Öğretmen Eğitimi Kursu Uygulama Örneği). *Yüksek Lisans Tezi*, Ankara Üniversitesi, Sosyal Bilimler Enstitüsü, Sosyal Bilimler ve Çevre Anabilim Dalı, Ankara, 144 s.

Çepel N (1990) Ekoloji Terimleri Sözlüğü. İstanbul, 258.

Çetin M, Sarıgül S S, Topcu B A, Alvarado R, ve Karataser B (2023) Does globalization mitigate environmental degradation in selected emerging economies? Assessment of the role of financial development, economic growth, renewable energy consumption and urbanization. *Environmental Science and Pollution Research*, 30(45): 100340-100359.

De Groot J I (2019) Environmental Psychology: An Introduction. Steg L (ed.), ISBN: 978-1-119-24108-9, Wiley-Blackwell, 448 s.

Değirmenci A (2024) Comparison of K-Nearest Neighbors, Decision Tree and Support Vector Machines Methods in Predicting Environmental.Attitudes.*Emerging Trends in Electrical and Electronics Engineering*, Reşat M A(ed.), ISBN: 978-625-5530-00-4, Duvar Yayınları, İzmir, (4): 4-21.

Deveci E ve Özarı Ç (2020) K-en Yakın Komşu Algoritması ile Oecd Ülkelerinin Ekonomik Özgürlük Kategorilerinin Tahmin Edilmesi. *Sosyal Bilimler Dergisi*, (44): 577-590.

Dietterich T G (2000) Ensemble methods in machine learning. *In International workshop on multiple classifier systems*, Berlin, Heidelberg: Springer, 1-15.

Downs S M, Ahmed S, Fanzo J ve Herforth A (2020) Food Environment Typology: Advancing an Expanded Definition, Framework, and Methodological Approach for Improved Characterization of Wild, Cultivated, and Built Food Environments toward Sustainable Diets. *Foods*, 9(4): 532.

Duda R O, Hart P E ve Stork D G (2001) *Pattern Classification,* 2. Baskı, Wiley-Interscience, Hoboken,

Duda R O ve Hart P E (2006) *Pattern classification*, 2. Baskı, John Wiley & Sons, Hoboken,

Dunlap R E ve Van Liere K D (1978) The "new environmental paradigm". *The Journal of Environmental Education*, 9(4): 10-19.

Eagly A H ve Chaiken S (1993) *The psychology of attitudes*. Harcourt Brace Jovanovich College Publishers, Fort Worth, Teksas,

Efeoğlu E (2022) Covid-19 Hastalarının Ölüm Oranlarının ve Yüksek Ölüm Riskine Sahip Hastaların Belirlenmesi için Temel Bileşen Analizinin Kullanılması. *Zeki Sistemler Teori ve Uygulamaları Dergisi*, 5(2): 127-136.

Emekli B (2024) Yazılım Tanımlı Ağlarda Makine Öğrenme Temelli Saldırı Tespit Sistemi. *Yüksek Lisans Tezi*, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Bilişim Sistemleri Mühendisliği Anabilim Dalı, Sakarya, 57 s.

KAYNAKLAR (devam ediyor)

Eroğlu E, Aydemir Dev M ve Dalgın K (2025) Exploring the moderating role of university environmental performance in shaping students' environmental awareness and behavior. *International Journal of Sustainability in Higher Education,*

Fayiga A O, Ipinmoroti M O ve Chirenje T (2018) Environmental pollution in Africa. *Environment, Development and Sustainability*, 20(1): 41-73.

Fisher R A (1936) The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2): 179-188.

Friedman J H (1999) Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5): 1189–1232.

Frietsch M, Loos J, Löhr K, Sieber S ve Fischer J (2023) Future-proofing ecosystem restoration through enhancing adaptive capacity. *Communications Biology,* 6:377

Gadenne D L, Kennedy J ve McKeiver C (2009) An empirical study of environmental awareness and practices in SMEs. *Journal of Business Ethics*, 84(1): 45-63.

Genuer R, Poggi J M ve Tuleau-Malot C (2010) Variable selection using random forests. *Pattern Recognition Letters*, 31(14): 2225-2236.

Gholami M, Fattahi Y ve Alesheikh A A (2021) Predicting pro-environmental behavior using machine learning techniques. *Environmental Monitoring and Assessment*, 193(6): 1–15.

Görmez Y, Arslan H, Işık Y E ve Tomaç S (2024) Nesnelerin internetinde iç mekân lokalizasyonu için makine öğrenimini kullanan bluetooth düşük enerji tabanlı özgün hibrit model. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 29(1): 36-43.

Güler Ç ve Çobanoğlu Z (1994) *Su Kirliliği*. ISBN-10(13) 975-7572-60-8, Aydoğdu Ofset, Ankara, 112.

Gürgen G ve Serttaş S (2023) Kalp Yetmezliği Hastalığının Erken Teşhisinde Makine Öğrenimi Algoritmalarının Performans Karşılaştırması. *Euroasia Journal of Mathematics, Engineering, Natural & Medical Sciences,* 10(28): 165-174.

Han J, Kamber M ve Pei J (2011) *Data Mining: Concepts and Techniques,* 3. Baskı, Morgan Kaufman Publishers, USA, 740s.

Hastie T, Tibshirani R, Friedman J ve Franklin J (2005) The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2): 83-85.

Hastie T, Tibshirani R ve Friedman J (2009) *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2. Baskı, Springer, New York

Haykin S (2009) *Neural networks and learning machines,* 3. Baskı, Pearson, New Jersey,

KAYNAKLAR (devam ediyor)

Herring C ve Kaplan S (2000) Component-based software systems for smart environments. *IEEE Personal Communications*, 7(5): 60–61.

Hsu C W, Chang C C ve Lin C J (2010) *A practical guide to support vector classification*, Technical report, National Taiwan University, Department of Computer Science,

IPCC (2021) *Climate Change 2021: The Physical Science Basis*. Intergovernmental Panel on Climate Change. Adres: https://www.ipcc.ch

Iranzad R ve Liu X (2024) A review of random forest-based feature selection methods for data science education and applications. *International Journal of Data Science and Analytics*, 1-15.

Jiang A, Lehnert K, You L ve Snell R G (2022) ICARUS, an interactive web server for single cell RNA-seq analysis. *Nucleic Acids Research* 50(1): 427-433.

Kaiser F G, Byrka K ve Hartig T (2010) Reviving Campbell's paradigm for attitude research. *Personality and Social Psychology Review*, 14(4): 351-367.

Kapyla M ve Wahlström R (2000) Environmental education in teacher training: A case study. *International Research in Geographical and Environmental Education*, 9(3): 229–241.

Karaman S ve Gökalp Z (2010) Küresel Isınma ve İklim Değişikliğinin Su Kaynakları Üzerine Etkileri. *Tarım Bilimleri Araştırma Dergisi.* 3(1): 59-66.

Kaypak Ş (2011) Küreselleşme sürecinde sürdürülebilir bir kalkınma için sürdürülebilir bir çevre. *Karamanoğlu Mehmetbey Üniversitesi Sosyal ve Ekonomik Araştırmalar Dergisi*, 2011(1): 19–33.

Keleş R ve Hamamcı C (1998) *Çevrebilim*. 1. Baskı, ISBN:9789755330242, İmge Kitabevi, Ankara, 368.

Kılınç E E (2022) Yapay Sinir Ağı Yöntemleri Kullanılarak Kalp Krizinin Tahmin Edilmesi. *Yüksek Lisans Tezi*, Burdur Mehmet Akif Ersoy Üniversitesi, Sosyal Bilimler Enstitüsü, Yönetim Bilişim Sistemleri Anabilim Dalı, Burdur , 95 s.

Koklu N ve Sulak S A (2024) Classification of Environmental Attitudes with Artificial Intelligence Algorithms. *Intelligent Methods In Engineering Sciences*, 3(2): 54-62.

Kollmuss A ve Agyeman J (2002) Mind the Gap: Why do people act environmentally and what are the barriers to pro-environmental behavior?. *Environmental Education Research*, 8(3): 239–260.

Kuncheva L I (2004) *Combining Pattern Classifiers: Methods and Algorithms*, Kuncheva L I (Ed.), 1. Baskı, ISBN: 978-0-471-21078-8, Wiley-Interscience, Hoboken, New Jersey,

Kurt H (2017) Çevre Sorunlarının Kavşağında Biyolojik Çeşitlilik. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 22(3): 825-837.

KAYNAKLAR (devam ediyor)

Küçükbesleme H F (2024) Aile Hekimliği Asistanlarının Gezegen Sağlığı ve Çevre-sağlık İlişkisi Konularında Bilgi ve Davranışlarının Değerlendirilmesi. *Tıpta Uzmanlık Tezi*, Sağlık Bilimleri Üniversitesi, Adana Tıp Fakültesi, Aile Hekimliği Kliniği, Adana, 80 s.

Lazrek G, Chetioui K, Balboul Y, Mazer, S ve El bekkali M (2024) An RFE/Ridge-ML/DL based anomaly intrusion detection approach for securing IoMT system. *Results in Engineering*, 23: 102659.

Lee J ve Kim H (2023) Machine learning-based prediction of environmental awareness in urban populations. *Sustainable Cities and Society*, 92: 104497.

Liu S, Huang Z, Zhu J, Liu B ve Zhou P (2024) Continuous blood pressure monitoring using photoplethysmography and electrocardiogram signals by random forest feature selection and GWO-GBRT prediction model. *Biomedical Signal Processing and Control*, 88: 105354.

Marques C S, Dufourq E, Pereira S, Santos V F ve Malafaia E (2025) Enhancing the classification of isolated theropod teeth using machine learning: a comparative study. *PeerJ*, 13(3): 19116.

McCallum A ve Nigam K (1998) A comparison of event models for naive Bayes text classification. *In AAAI-98 workshop on learning for text categorization,* AAAI Press, 41–48.

McLachlan G J (2004) *Discriminant analysis and statistical pattern recognition*, 1.Baskı, John Wiley & Sons, Hoboken,

Menteşe S (2017) Çevresel Sürdürülebilirlik Açısından Toprak, Su ve Hava Kirliliği: Teorik Bir İnceleme. *Uluslararası Sosyal Araştırmalar Dergisi*. 10(1): 53.

Milfont T L ve Duckitt J (2010) The environmental attitudes inventory: A valid and reliable measure to assess the structure of environmental attitudes. *Journal of Environmental Psychology*, 30(1): 80-94.

Mitchell T M ve Mitchell T M (1997) *Machine learning,* 1. Baskı, McGraw-hill, New York,

Nejad A S (2017) Sıvılaşma Riskini Tahmin Etmek için Çeşitli Makine Öğrenmesi Yaklaşımlarının Uygulanması. *Yüksek Lisans Tezi*, Boğaziçi Üniversitesi, Fen Bilimleri Enstitüsü, İnşaat Mühendisliği Anabilim Dalı, İstanbul, 62 s.

Onan A (2018) Parçacık Sürüsü Eniyilemesine Dayalı Yığılmış Genelleme Yöntemi ve Metin Sınıflandırma Üzerinde Uygulanması. *Academic Platform Journal of Engineering and Science*, 6(2): 134-141.

Opitz D ve Maclin R (1999) Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, 11: 169–198.

KAYNAKLAR (devam ediyor)

Öklü M ve Canbay P (2023) Makine Öğrenmesi Yöntemleri ile Şehirlerin Hava Kalitesi Tahmini. *International Journal of Advances in Engineering and Pure Sciences*, 35(1): 39-53.

Özdemir B (2009) Küresel kirlenme sürdürülebilir ekonomik büyüme ve çevre vergileri. *Maliye Dergisi*, 156: 1-36.

Özdemirkol M (2023) Bir dayanışma hakkı olarak çevre hakkı: Ilısu barajı ve hidroelektrik enerji santralinin çevre hakkı çerçevesinde incelenmesi. *Doktora Tezi*, Hatay Mustafa Kemal Üniversitesi, Sosyal Bilimler Enstitüsü, Siyaset Bilimi ve Kamu Yönetimi Anabilim Dalı, Hatay, 411 s.

Ponting C (2007) *A new green history of the world: the environment and the collapse of great civilizations*, Random House, New York,

Ren X (2025) A hybrid model combining environmental analysis and machine learning for predicting AI education quality. *Sci Rep*, 15: 12577.

Rish I (2001) *An empirical study of the naive Bayes classifier*. *IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence,* Seattle, Washington, USA, 41–46.

Rokach L (2010) Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1-2): 1-39.

Russell S (2016) *Artificial Intelligence: A Modern Approach*, Norvig P (Ed.), 3. Baskı, ISBN: 978-1-292-15396-4, Pearson Education, Upper Saddle River, New Jersey,

Sağsöz G ve Doğanay G (2019) İlkokul öğrencilerinin çevre ve çevre sorunlarına ilişkin görüşlerinin incelenmesi (Giresun İli Örneği). *Anadolu University Journal of Education Faculty*, 3(1): 1-20.

Scholkopf B ve Smola A J (2002) *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, 1. Baskı, MIT Press, Cambridge,

Schultz P W (2001) The structure of environmental concern: Concern for self, other people, and the biosphere. *Journal of Environmental Psychology*, 21(4): 327-339.

Schultz P W, Nolan J M, Cialdini R B, Goldstein N J ve Griskevicius V (2007) The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18(5): 429-434.

Singh R L ve Singh P K (2017) Global environmental problems. *Principles and Applications of Environmental Biotechnology for a Sustainable Future*, Singh R L (Ed.), 1. Baskı, ISBN:978-981-10-9465-1, Springer, 13–41.

Somuncu S ve Oral C (2024) Yapay Sinir Ağı ve ANFIS Kullanılarak Meteorolojik Verilere Bağlı Güneş Enerjisi Tahmini. *OKU Fen Bilimleri Enstitüsü Dergisi*, 7(4):1685-1701.

Şahin C M (2019) Korku, Kaygı ve Kaygı (Anksiyete) Bozuklukları. *Avrasya Sosyal ve Ekonomi Araştırmaları Dergisi*, 6(10): 117–135.

KAYNAKLAR (devam ediyor)

Şahin E K (2017) Özellik Seçimi Algoritmaları Kullanılarak Heyelanda Etkili Faktörlerin Belirlenmesi ve Heyelan Duyarlılık Haritalarının Üretilmesi. *Doktora Tezi*, İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Geomatik Mühendisliği Anabilim Dalı, İstanbul, 223 s.

Türküm A S (1998) Çağdaş Toplumda Çevre Sorunları ve Çevre Bilinci. *Çağdaş Yaşam Çağdaş İnsan*, G. Can (Ed.), Anadolu Üniversitesi Açık Öğretim Fakültesi İlköğretim Öğretmenliği Lisans Tamamlama Programı, Eskişehir, 165-182.

Uzşen H, Büyük E T ve Koyun M (2024) Çevrimiçi Eğitimin Çocuk Hemşireliği Yeterliliğine Etkisi: Yarı Deneysel Bir Araştırma. *İzmir Kâtip Çelebi Üniversitesi Sağlık Bilimleri Fakültesi Dergisi*, 9(1): 49-56.

Uzun R, İşler Y ve Toksan M (2019) WEKA yazılım paketinin siğil tedavi yöntemlerinin başarısının tahmininde kullanımı. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, *7*(1): 699-708.

Ürün M B ve Sönmez Y (2024) Nelder-Mead Optimized Weighted Voting Ensemble Learning for Network Intrusion Detection. *Düzce University Journal of Science & Technology*, 12(4): 2139-2158.

van Valkengoed A M, Abrahamse W ve Steg L (2022) To select effective interventions for pro-environmental behaviour change, we need to consider determinants of behaviour. *Nature Human Behaviour*, 6(11): 1482-1492.

Vapnik V (1999) *The nature of statistical learning theory*, 2. Baskı, Springer science & business media, New York,

Vaughan C, Gack J, Solorazano H ve Ray R (2003) The effect of environmental education on schoolchildren, their parents, and community members: A study of intergenerational and intercommunity learning. *The Journal of Environmental Education*, 34(3): 12-21.

Walpole R E, Myers R H, Myers S L ve Ye K (2012) Varyans. *Probability & Statistics for Engineers & Scientists*, Myers RH (Ed.), 9. Baskı, ISBN: 978-0-321-62911-1, Prentice Hall. Upper Saddle River, New Jersey, 10-15

Wang Y, Li X ve Zhang M (2023) Integrating machine learning and behavioral data for environmental attitude classification. *Environmental Science & Policy*, 142: 97–106.

Yavuz A A ve Yavuz H S (2021) Denetimli Makine Öğrenme Yöntemleri ile Yüzey Su Kalitesinin Sınıflandırılması. *Biyoloji Bilimleri Araştırma Dergisi*, 14(2): 142-155.

Yapıcı İ Ş ve Arslan R U (2024) Gebelikte Anne Sağlığı Risk Gruplarının Tahminine Yönelik Makine Öğrenmesi Tabanlı Bir Karar Destek Sistem Tasarımı. *Black Sea Journal of Engineering and Science*, *7*(3): 509-520.

Yaylı H (2017) Çevre Etiği Bağlamında Kalkınma, Çevre ve Nüfus. *Süleyman Demirel Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 1(15): 151 – 169.

KAYNAKLAR (devam ediyor)

Yıldırım O (2023) Makine Öğrenmesi Yöntemleriyle Orman Yangını Tahmini. *Yüksek Lisans Tezi*, Atatürk Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı, Erzurum, 65 s.

Yılmaz S ve Zorlutuna Ş (2024) Öğretmen Adaylarının Çevresel Duyarlılık, Çevre Bilinci ve Çevre Sorunlarına Yönelik Davranışlarının İncelenmesi. *Sinop Üniversitesi Sosyal Bilimler Dergisi*, 8(Eğitim Bilimleri Özel Sayısı): 25-49.

Yılmaz Ü ve Kuvat Ö (2023) İnvestigating The Effect of Feature Selection Methods on The Success of Overall Equipment Effectiveness Prediction. *Uludağ University Journal of The Faculty of Engineering*, 28(2): **-**.

Yudhana A, Cahyo A D, Sabila L Y, Subrata A C ve Mufandi I (2023) Spatial distribution of soil nutrient content for sustainable rice agriculture using geographic information system and Naïve Bayes classifier. *International Journal on Smart Sensing and İntelligent Systems*, 16(1): **-**.

Zhang C, Cui H, Li Y ve Chang X (2024) Predicting CD27 expression and clinical prognosis in serous ovarian cancer using CT-based radiomics. *Journal of Ovarian Research*, 17(1): 131.

Zhang H (2004) *The Optimality of Naive Bayes*. *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference (FLAIRS 2004)*. Miami, Florida, USA, 562–567.

Zhang Z (2016) Introduction to machine learning: k-nearest neighbors. *Annals of translational medicine*, 4(11): 218.

Zhao X, Ma X, Chen B, Shang Y ve Song M (2022) Challenges toward carbon neutrality in China: Strategies and countermeasures. *Resources, Conservation and Recycling*, 176: 105959.

Zheng Y (2010) Association analysis on pro-environmental behaviors and environmental consciousness in main cities of East Asia. *Behaviormetrika*, 37(1): 55-69.

(URL-1) < https://www.kaggle.com/datasets/suleymansulak/environmental-attitude-dataset>, Ziyaret tarihi: 01.01. 2025

ÖZGEÇMİŞ

Fuat ALKAN, İlkokul, ortaokul ve lise öğrenimimi Zonguldak'ta tamamladım. Lisans eğitimimi Abant İzzet Baysal Üniversitesi Düzce Teknik Eğitim Fakültesi Elektrik-Elektronik Teknolojisi Alan Öğretmenliği bölümünde tamamladım. 2013 yılında Bülent Ecevit Üniversitesi Elektrik-Elektronik Mühendisliği Tamamlama Programı'na kabul edildim ve 2015 yılında mühendislik diploması almaya hak kazandım.

Meslek hayatıma 2014 yılında Yozgat iline Elektrik-Elektronik öğretmeni olarak atanarak başladım. Yedi yıl süren doğu görevimin ardından, 2021 yılında memleketim olan Zonguldak'a tayin oldum ve halen Devrek Mesleki ve Teknik Anadolu Lisesi'nde Elektrik-Elektronik Öğretmeni (Uzman Öğretmen) olarak görev yapmaktayım.

Alanımda edindiğim teorik ve pratik bilgi birikimini, mesleki deneyimimle birleştirerek genç nesillere aktarma hedefiyle Bülent Ecevit Üniversitesi Fen Bilimleri Enstitüsü Elektrik-Elektronik Mühendisliği Anabilim Dalında Yüksek Lisans eğitim hayatıma devam etmekteyim.