

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENME YÖNTEMLERİ İLE YANGIN
VERİLERİNİN ANALİZİ VE SINIFLANDIRILMASI

YÜKSEK LİSANS TEZİ

Zeynep Nazlı ASLAN

Afet Yönetimi Anabilim Dalı

HAZİRAN 2025

T.C.
SAKARYA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE YANGIN
VERİLERİNİN ANALİZİ VE SINIFLANDIRILMASI

YÜKSEK LİSANS TEZİ

Zeynep Nazlı ASLAN

Afet Yönetimi Anabilim Dalı

Tez Danışmanı: Doç. Dr. Beytullah EREN

HAZİRAN 2025

Zeynep Nazlı ASLAN tarafından hazırlanan “Makine Öğrenmesi Yöntemleri İle Yangın Verilerinin Analizi ve Sınıflandırılması“ adlı tez çalışması 23.06.2025 tarihinde aşağıdaki jüri tarafından oy birliği ile Sakarya Üniversitesi Fen Bilimleri Enstitüsü Afet Yönetimi Anabilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

Tez Jürisi

Jüri Başkanı : **Doç. Dr. Beytullah EREN (Danışman)**
Sakarya Üniversitesi

Jüri Üyesi : **Prof. Dr. Mehmet İŞLEYEN**
Sakarya Üniversitesi

Jüri Üyesi : **Dr. Öğr. Üyesi Ahmet İyad CEYHUNLU**
Sakarya Uygulamalı Bilimler Üniversitesi



ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Sakarya Üniversitesi Fen Bilimleri Enstitüsü Lisansüstü Eğitim-Öğretim Yönetmeliğine ve Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesine uygun olarak hazırlamış olduğum “MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE YANGIN VERİLERİNİN ANALİZİ VE SINIFLANDIRILMASI” başlıklı tezin bana ait, özgün bir çalışma olduğunu; çalışmamın tüm aşamalarında yukarıda belirtilen yönetmelik ve yönergeye uygun davrandığımı, tezin içerdiği yenilik ve sonuçları başka bir yerden almadığımı, tezde kullandığım eserleri usulüne göre kaynak olarak gösterdiğimi, bu tezi başka bir bilim kuruluna akademik amaç ve unvan almak amacıyla vermediğimi ve 20.04.2016 tarihli Resmi Gazete’de yayımlanan Lisansüstü Eğitim ve Öğretim Yönetmeliğinin 9/2 ve 22/2 maddeleri gereğince Sakarya Üniversitesi’nin abonesi olduğu intihal yazılım programı kullanılarak Enstitü tarafından belirlenmiş ölçütlere uygun rapor alındığını, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun ortaya çıkması halinde doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi beyan ederim.

(23/06/2025)

Zeynep Nazlı ASLAN



TEŞEKKÜR

Yüksek lisansımın ve araştırma sürecimde tecrübelerini, bilgilerini ve zamanını esirgemeyip her fırsatta yardımcı olan değerli danışman hocam Doç. Dr. Beytullah EREN'e, araştırmam boyunca bütün zorluklarda bana destek ve yardımcı olan değerli hocam Öğr. Gör. Dr. Muhammed Fatih PEKŞEN ve iş arkadaşlarıma teşekkürlerimi sunarım.

Hayatımın her alanında olduğu gibi tez sürecimde de elini hep omzumda hissettiğim sevgileri ve destekleriyle beni daima güçlü kılan canım annem Nimet KOCA ve babam Ercan KOCA'ya emekleri, sabırları ve maddi manevi katkıları için sonsuz teşekkür ederim.

Hayatıma girdiği andan itibaren her zaman yanımda olan, karşılaştığım her türlü zorlukta beni motive eden, eğitimimde büyük desteği olan değerli eşim Mikail ASLAN'a...

Sizlerin katkıları olmadan bu çalışmayı tamamlamak mümkün olmazdı.

Tüm kalbimle sonsuz teşekkürlerimi sunuyorum.

Zeynep Nazlı ASLAN



İÇİNDEKİLER

Sayfa

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ	v
TEŞEKKÜR	vii
İÇİNDEKİLER	ix
KISALTMALAR	xi
TABLO LİSTESİ	xiii
ŞEKİL LİSTESİ	xv
ÖZET	xvii
SUMMARY	xix
1. GİRİŞ	1
1.1. Çalışmanın Amacı ve Önemi	2
1.2. Problem Tanımı ve Araştırma Soruları	4
1.3. Çalışmanın Kapsamı ve Sınırları	5
2. KAVRAMSAL ÇERÇEVE VE LİTERATÜR TARAMASI	7
2.1. Yangın Analizleri ve Risk Değerlendirme Yaklaşımları	7
2.1.1. Yangın risk faktörleri	8
2.1.1.1. Meteorolojik faktörler	8
2.1.1.2. Topoğrafik faktörler	8
2.1.1.3. İnsan kaynaklı faktörler	9
2.1.2. Yangın risk değerlendirme yaklaşımları	9
2.1.2.1. Geleneksel yangın risk analizi yöntemleri	9
2.1.2.2. Veri tabanlı ve makine öğrenmesi yaklaşımları	10
2.2. Yangın Tahmininde Kullanılan Makine Öğrenmesi Teknikleri	10
2.3. Yangın Yönetiminde Veri Madenciliği ve Özellik Seçimi	14
2.4. Literatür Çalışmaları	16
2.5. Çalışmanın Literatüre Katkısı	18
3. MAKİNE ÖĞRENMESİ	21
3.1. Makine Öğrenmesinin Tanımı ve Gelişimi	21
3.2. Makine Öğrenmesi Algoritmaları	23
3.2.1. Denetimli öğrenme yöntemleri	25
3.2.1.1. Karar ağacı (fine tree)	28
3.2.1.2. Destek Vektör Makineleri (Linear Support Vector Machine (SVM))	29
3.2.1.3. Gaussian Naive Bayes (GNB)	30
3.2.2. Denetimsiz öğrenme yöntemleri	31
3.2.2.1. Kümeleme (clustering) yöntemleri	32
3.3. Model Doğrulama Yöntemleri	32
3.3.1. Sınama seti yöntemi (holdout method)	33
3.3.2. Bootstrap doğrulama	33
3.3.3. K-Katlı çapraz doğrulama (k-fold cross validation)	34
3.3.3.1. Boyut azaltma (dimensionality reduction) yöntemleri	36
3.3.4. Denetimsiz öğrenme yöntemlerinin yangın analizindeki rolü	36

4. BÖLÜM YÖNTEM.....	39
4.1. Kullanılan Veri Seti	39
4.1.1. Veri kaynağı ve özellikleri	39
4.1.2. Veri ön işleme	40
4.2. Analiz Yöntemleri	40
4.2.1. Yangın verilerinin istatistiksel analizi.....	40
4.2.2. Makine öğrenmesi modelleri.....	41
4.2.2.1. Karar ağacı (fine tree) modeli	41
4.2.2.2. Gaussian Naive Bayes (GNB) modeli.....	42
4.2.2.3. Destek Vektör Makineleri (SVM) modeli.....	43
4.2.3. Makine öğrenmesi model eğitimi ve doğrulama süreci	44
4.2.4. Eğitim ve test veri kümesinin oluşturulması	44
4.2.5. Çapraz doğrulama yönteminin uygulanması.....	45
4.2.6. Makine öğrenmesi modellerinin performans değerlendirme ölçütleri	45
4.2.6.1. Doğruluk (accuracy) ölçütü.....	47
4.2.6.2. Karışıklık matrisi (confusion matrix) ile değerlendirme.....	47
4.2.6.3. ROC eğrileri ve AUC değeri	47
4.2.6.4. F1-Skoru, duyarlılık (recall) ve kesinlik (precision) ölçütleri.....	48
5. BULGULAR	49
5.1. Yangın Verilerinin Tanımlayıcı İstatistikleri ve Veri Analizi.....	49
5.1.1. Yıllara göre yangın sayısı trendi	49
5.1.2. Yangınların bölgesel dağılımı	50
5.1.3. Mahallelere göre yangın dağılımı	50
5.1.4. Yangınların gün ve saat içerisindeki dağılımı.....	51
5.1.5. Yangınların mevsimlere göre dağılımı.....	52
5.1.6. Yangın türleri ve mevsimsel dağılımı	53
5.1.7. Yangın çıkış nedenleri.....	55
5.1.8. Yangınların mesafeye göre dağılımı	56
5.1.9. Yangınların insan üzerindeki etkileri	56
5.1.10. Hasar durumu	57
5.2. Makine Öğrenmesi Modelleme Bulguları	58
5.2.1. Veri seti özellikleri ve ön işleme süreci	58
5.2.2. Özellik seçimi ve etkili değişkenlerin belirlenmesi	60
5.2.3. Doğrulama yöntemi ve modelleme yaklaşımı.....	62
5.2.4. Eğitim aşamasında elde edilen performans bulguları	63
5.2.5. Test aşaması bulguları ve gerçek performans değerlendirmesi	65
5.2.6. ROC Eğrileri ve AUC değerleriyle ayırt etme gücünün analizi.....	68
5.2.7. Model performans ölçümlerinin karşılaştırmalı analizi	72
5.2.8. Bulguların yorumlanması ve kritik tespitler.....	73
6. SONUÇ VE TARTIŞMA.....	75
KAYNAKLAR.....	79
ÖZGEÇMİŞ.....	89

KISALTMALAR

AFAD	: Afet ve Acil Durum Yönetimi Başkanlığı
ANOVA	: Varyans Analizi (Analysis of Variance)
ANN	: Yapay Sinir Ağı (Artificial Neural Network)
AUC	: Eğri Altındaki Alan (Area Under the Curve)
FN	: Yanlış Negatif (False Negative)
GNB	: Gaussian Naive Bayes
IL	: Yanıcı Sıvı
IRDR	: Afet Riskine İlişkin Entegre Araştırma (Integrated Research on Disaster Risk)
K-NN	: En Yakın Komşu
LST	: Arazi Yüzey Sıcaklığı
MAMA	: Minimum Artıklık Maksimum Alaka Düzeyi
ML	: Makine Öğrenmesi (Machine Learning)
MLR	: Çoklu Doğrusal Regresyon (Multiple Linear Regression)
NVDI	: Normalize Edilmiş Fark Bitki Örtüsü İndeksi (Normalized Difference Vegetation Index)
PCA	: Temel Bileşenler Analizi (Principal Component Analysis)
RBFNN	: Radyal Taban Fonksiyonlu Sinir Ağı (Radial Basis Function Neural Network)
SVM	: Destek Vektör Makinesi (Support Vector Machine)
TA	: Termal Anomali
TP	: Doğru Pozitif (True Positive)
TS EN	: Türk Standardı
UNISDR	: Birleşmiş Milletler Afet Riskini Azaltma Stratejisi (United Nations International Strategy for Disaster Reduction)



TABLO LİSTESİ

Sayfa

Tablo 5.1. Ki-kare testine göre özelliklerin önem sıralaması.....	61
Tablo 5.2. Eğitim aşamasında modellerin doğruluk oranları.	63
Tablo 5.3. Test aşamasında modellerin doğruluk oranları.	66
Tablo 5.4. Modellerin precision, recall ve f1-score performans ölçümleri.	72





ŞEKİL LİSTESİ

Sayfa

Şekil 3.1. Makine öğrenimi şeması (Akin & Şahin, 2023).....	22
Şekil 3.2. Makine öğrenmesi algoritmaları hiyerarşisi (Rashidi, Tran, Betts, Howell, & Green, 2019).....	23
Şekil 3.3. Çapraz doğrulama modeli (Bulut, 2023).	35
Şekil 5.1. Yıllara göre yangın sayısı trendi.	49
Şekil 5.2. En riskli ilçeler (en çok yangın çıkan ilk 5).....	50
Şekil 5.3. İlçelere göre yangın dağılımı.	50
Şekil 5.4. En riskli mahalleler (en çok yangın çıkan ilk 5).....	51
Şekil 5.5. Yangınların gün içindeki dağılımı.	52
Şekil 5.6. Gece ve gündüz yangın dağılımı.	52
Şekil 5.7. Mevsimlere göre yangın dağılımı.	53
Şekil 5.8. Yangın türlerinin dağılımı.	54
Şekil 5.9. Yangın türlerine göre mevsimsel dağılım.	55
Şekil 5.10. Yangın çıkış nedenlerinin dağılımı.....	55
Şekil 5.11. Yangın mesafesi dağılımı.	56
Şekil 5.12. Yangınlardan kaynaklanan can kaybı ve yaralanmalar.	56
Şekil 5.13. Yangınlarda hasar durumu.	57
Şekil 5.14. Doğrulama tekniği seçim ekranı.	59
Şekil 5.15. Analiz yönetimi seçim ekranı.	60
Şekil 5.16. Ki-kare özellik sıralama ekranı.....	61
Şekil 5.17. Karar ağacı modeline ait eğitim karışıklık matrisi.	64
Şekil 5.18. Gaussian naive bayes modeline ait eğitim karışıklık matrisi.....	64
Şekil 5.19. Linear svm modeline ait eğitim karışıklık matrisi.	65
Şekil 5.20. Karar ağacı modeline ait test karışıklık matrisi.	66
Şekil 5.21. Gaussian naive bayes modeline ait test karışıklık matrisi.	67
Şekil 5.22. Linear SVM modeline ait test karışıklık matrisi.	67
Şekil 5.23. Karar ağacı modeline ait roc eğrisi.	69
Şekil 5.24. Linear svm modeline ait roc eğrisi.	70
Şekil 5.25. Gaussian naive bayes modeline ait roc eğrisi.	71



MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE YANGIN VERİLERİNİN ANALİZİ VE SINIFLANDIRILMASI

ÖZET

Yangınlar, hem doğal süreçlerin hem de insan kaynaklı faaliyetlerin sonucu olarak ortaya çıkabilen ve yerleşim yerlerinden sanayi alanlarına kadar geniş bir coğrafyada ciddi can ve mal kayıplarına neden olan afetler arasında yer almaktadır. Artan kentleşme, iklim değişikliği ve sanayileşme yangın riskini artırmakta; bu durum, geleneksel yöntemlerin ötesine geçerek veri odaklı karar destek sistemlerine olan ihtiyacı ortaya koymaktadır. Bu çalışmanın temel amacı, yangın verileri üzerinde makine öğrenmesi algoritmalarını uygulayarak yangın türlerinin sınıflandırılması ve elde edilen modellerin performanslarının karşılaştırılmasıdır.

Araştırma kapsamında, 2015–2023 yılları arasında bir ilin yerel itfaiye kayıtlarından elde edilen 29.135 adet yangın verisi kullanılmıştır. Veriler ön işleme tabi tutulmuş, ardından Ki-Kare testi ile en etkili öznitelikler belirlenmiştir. Sınıflandırma sürecinde üç farklı makine öğrenmesi algoritması olan Karar Ağacı (Fine Tree), Gaussian Naive Bayes (GNB) ve Lineer Destek Vektör Makineleri (Linear SVM) kullanılmıştır. Tüm modeller, 5 katlı çapraz doğrulama yöntemiyle test edilmiş; doğruluk oranı, karışıklık matrisi, F1-Skoru ve ROC eğrisi (AUC) gibi performans ölçütleriyle değerlendirilmiştir.

Analizler sonucunda, en yüksek doğruluk oranı %61,4 ile Linear SVM modelinde elde edilmiş; onu %61,2 ile Karar Ağacı ve %58,5 ile GNB modeli takip etmiştir. AUC değerleri açısından da Linear SVM modeli 0,7511 ile en yüksek ayırt etme gücüne sahip model olarak öne çıkmıştır. Karar Ağacı modeli, yorumlanabilirliği sayesinde karar vericiler için avantaj sağlamış; GNB ise düşük hesaplama maliyeti ile dikkat çekmiştir. Modellerin “Tamamen” yangın sınıfını sınıflandırmada zorlandığı gözlemlenmiş ve bu durum, örnek dengesizliği ile açıklanmıştır.

Çalışma, literatüre; kentsel ve kırsal yangınları birlikte ele alan, gerçek saha verilerine dayalı geniş kapsamlı bir sınıflandırma sunması; farklı makine öğrenmesi algoritmalarının aynı veri seti üzerinde karşılaştırmalı olarak analiz edilerek en uygun modelin belirlenmesi; ve özellik seçiminin model başarısına etkisinin somut olarak ortaya konması açısından katkı sağlamaktadır. Sonuç olarak bu tez, yangın tahmin süreçlerine yönelik makine öğrenmesi tabanlı yaklaşımıyla yerel yönetimlerin müdahale stratejilerine bilimsel bir temel sunmakta; özellikle kaynak planlaması, risk yönetimi ve erken uyarı sistemlerinin geliştirilmesine destek olmaktadır. İleride yapılacak çalışmalarda, farklı coğrafi bölgelerin verileriyle desteklenmiş ve sınıf dengesi sağlanmış veri kümeleriyle modelin genellenebilirliğinin artırılması önerilmektedir.



ANALYSIS AND CLASSIFICATION OF FIRE DATA USING MACHINE LEARNING METHODS

SUMMARY

Fires represent one of the most catastrophic types of disasters that can have profound and far-reaching consequences for human life, infrastructure, public health, and the natural environment. As societies continue to evolve under the pressures of climate change, increasing population density, urban sprawl, and accelerating industrialization, fire-related incidents have not only become more frequent, but also more destructive, unpredictable, and complex in nature. The interconnection between anthropogenic activities and environmental changes has exacerbated fire risks, thereby demanding a paradigm shift from conventional fire response mechanisms to more proactive, data-driven, and intelligent systems for prevention, early detection, and rapid mitigation.

In traditional fire management systems, decision-making is predominantly reactive, relying heavily on human expertise, past experiences, and manual assessments of risk. While such approaches have served well in the past, they are no longer sufficient in today's data-rich and dynamically evolving environments. The sheer volume, variety, and velocity of fire-related data necessitate more sophisticated analytical tools capable of processing complex datasets, extracting hidden patterns, and making timely, reliable predictions. In this context, machine learning (ML)—a subfield of artificial intelligence that focuses on algorithms capable of learning from data and improving over time—has emerged as a powerful solution for enhancing fire risk analysis and emergency response planning.

This study seeks to develop and evaluate a machine learning-based classification framework that can accurately analyze and predict the severity of fire incidents using real-world fire occurrence data. Specifically, the research is centered on building supervised classification models that categorize fire events into three severity-based levels: “Initial,” “Partial,” and “Complete.” The dataset employed in this research consists of 29,135 individual fire incident records obtained from a municipal fire department in a province of Turkey, covering a comprehensive nine-year period from 2015 to 2023. The dataset is rich in features and includes spatial variables (e.g., district, neighborhood), temporal variables (e.g., date, time of day, season), causal attributes (e.g., electrical faults, human negligence, unknown causes), and impact-related metrics (e.g., damage extent, number of casualties, distance to nearest fire station).

The methodological approach adopted in this study followed a structured, multi-phase process, beginning with extensive data preprocessing. This stage involved identifying and addressing missing or anomalous values, converting categorical variables into numerical representations through encoding techniques, and applying normalization where appropriate to standardize feature scales. Following preprocessing, a feature selection phase was conducted using the Chi-Square statistical test. This method was employed to determine the degree of association between independent variables and the categorical target variable (fire severity level), thereby enabling the exclusion of

less relevant or redundant features and improving both model interpretability and computational efficiency.

Subsequently, the study implemented three widely recognized supervised learning algorithms to perform the classification task: Decision Tree (specifically, the Fine Tree variant), Gaussian Naive Bayes (GNB), and Linear Support Vector Machines (SVM). Each of these models offers distinct advantages and challenges in terms of computational complexity, interpretability, and classification performance. To ensure the robustness and generalizability of the results, a 5-fold cross-validation approach was employed during model training. This method involves dividing the dataset into five equally sized subsets and iteratively using four subsets for training and one for validation, thus minimizing the risk of overfitting and providing a more reliable assessment of each model's performance.

The classification performance of the models was evaluated using a comprehensive set of metrics, including overall accuracy, precision, recall, F1-score, confusion matrix analysis, and the Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) curve. Among the models tested, the Linear SVM classifier demonstrated the highest classification accuracy, achieving a rate of 61,4%, closely followed by the Decision Tree model at 61,2%. The GNB model, although slightly less accurate with a score of 58,5%, exhibited superior computational efficiency and fast inference times. The AUC analysis further reinforced the superior performance of the Linear SVM model, which attained an AUC value of 0,7511. Notably, the model was particularly effective at correctly identifying fire incidents within the "Partial" severity class, demonstrating strong discriminative power for this category.

However, the performance of all models was less satisfactory when classifying "Complete" severity fires. This shortfall is likely attributed to an imbalanced class distribution in the dataset, where the number of "Complete" fire cases was significantly lower compared to the other categories. Additionally, overlapping feature characteristics between severity levels may have contributed to classification ambiguity. These limitations highlight the importance of addressing class imbalance through advanced sampling methods such as Synthetic Minority Over-sampling Technique (SMOTE) or by collecting more granular and representative data for underrepresented fire types.

Beyond numerical performance, the interpretability of the models was also a critical consideration in this study. The Decision Tree model, for instance, offered easily understandable, rule-based decision pathways that can be directly interpreted and utilized by emergency planners, fire department personnel, and municipal decision-makers. On the other hand, while the GNB model did not achieve the highest accuracy, its simplicity and computational speed make it a favorable choice for deployment in environments with limited processing capabilities or real-time operational requirements.

This research offers several original contributions to the existing body of literature on fire risk prediction and machine learning applications in disaster management. First, the study introduces a large-scale, real-world fire classification framework that integrates both urban and rural fire data—an important yet often overlooked dimension in current fire-related studies, which typically focus on forest fires or specific building types. Second, by systematically comparing three machine learning algorithms under uniform conditions, the research provides valuable insights into model selection for similar use cases. Third, the application of the Chi-Square feature selection method

effectively demonstrates how careful attribute selection can enhance model performance and reduce unnecessary computational overhead.

In addition to model development and evaluation, the study delves into exploratory data analysis to uncover meaningful temporal and spatial patterns within the fire dataset. The findings indicate that fire incidents tend to peak during the summer and autumn seasons, with industrial areas and densely populated urban neighborhoods showing higher fire frequencies. Furthermore, human-related causes such as electrical failures, careless behavior, and unsafe disposal of flammable materials emerged as predominant factors in fire ignition. These insights align with national and international fire statistics and reinforce the need for more targeted fire prevention strategies that consider both behavioral and environmental risk factors.

One of the key lessons drawn from this research is the importance of high-quality, comprehensive datasets in machine learning-driven fire management. The challenges encountered due to class imbalance and feature overlap underscore the necessity for future studies to incorporate more diverse and balanced data sources. Enhancing the dataset with external variables—such as meteorological conditions (e.g., temperature, humidity, wind speed), land use data, topographical information, and remote sensing imagery—could significantly improve the granularity and predictive accuracy of classification models.

Ultimately, this thesis illustrates that machine learning offers a powerful and scalable approach for enhancing fire risk assessment and supporting data-informed decision-making in emergency management. By integrating these models into operational workflows, local governments, fire brigades, and public safety agencies can proactively allocate resources, anticipate periods and regions of elevated fire risk, and implement timely interventions to reduce casualties and property loss. The findings of this study not only contribute to the academic discourse on artificial intelligence applications in public safety but also offer practical pathways toward the development of intelligent, adaptive fire management systems tailored to the evolving needs of modern societies.



1. GİRİŞ

Afetler, tarih boyunca insan toplumlarının güvenliğini tehdit eden ve büyük ölçekli fiziksel, ekonomik ve sosyal kayıplara neden olan olaylar arasında yer almaktadır. Küresel iklim değişikliği, hızlı kentleşme ve sanayileşme gibi faktörler, afet risklerini artırarak bu olayların daha sık ve yıkıcı sonuçlarla ortaya çıkmasına yol açmaktadır (Altıparmak & Sezgin, 2023). Yangınlar hem doğal hem de insan kaynaklı faktörlerden kaynaklanabilen ve kontrol altına alınmadığında ciddi can ve mal kayıplarına sebep olabilen afetlerden biridir. Özellikle yerleşim alanlarının genişlemesi, sanayi faaliyetleri (Şahin F. , 2022) ve tarımsal uygulamalar, yangınların sıklığını ve şiddetini artıran başlıca unsurlar arasında yer almaktadır (Gökkaya & Kaya, 2021).

Yangınların önlenmesi ve etkilerinin azaltılması için erken tahmin sistemlerinin geliştirilmesi büyük önem taşımaktadır. Geleneksel yangın yönetimi yöntemleri genellikle saha tecrübelerine ve manuel süreçlere dayanmakta olup, bu yöntemler geniş veri kümelerinin analiz edilmesi ve karmaşık yangın dinamiklerinin modellenmesi konusunda yetersiz kalabilmektedir (Rao, Eisenberg, & Schmitt, 2007). Günümüzde makine öğrenmesi ve büyük veri analitiği, afet yönetimi süreçlerini daha etkin hale getiren yeni nesil çözümler sunmaktadır (Yavuz & Bozatay, 2015). Bu teknolojiler, geçmiş yangın verilerini kullanarak riskli bölgeleri önceden belirleme, kaynak yönetimini iyileştirme ve tahminleme modelleri oluşturarak müdahale süreçlerini optimize etme potansiyeline sahiptir (Aksoy, Korucu, Çalışkan, Osmanbey, & Halis, 2021).

Türkiye'nin yangın riski taşıyan bölgeleri, geniş ormanlık alanları ve sanayi bölgeleri nedeniyle hem doğal hem de kentsel yangınlarla karşı karşıya kalmaktadır. Bölgedeki yangınlar, yalnızca doğal alanları tehdit etmekle kalmayıp, yerleşim bölgelerine yayılma riski taşıdığı için proaktif yangın yönetimi stratejilerinin geliştirilmesi gerekliliğini ortaya koymaktadır. Özellikle yaz aylarında artan sıcaklıklar, düşük nem oranları ve rüzgar gibi meteorolojik faktörler, yangın riskini yükseltmektedir (Pekşen, Uyaroglu, & Soyhan, 2020). Buna ek olarak, tarımsal faaliyetler, kırsal yapılaşma ve

insan ihmalleri yangın çıkış sebepleri arasında önemli bir yer tutmaktadır (Kaya, 2018). Bu durum, yangın risk haritalarının çıkarılması ve mekânsal-zamansal analizlerin yapılması gerekliliğini ortaya koymaktadır.

Makine öğrenmesi teknikleri, afet yönetimi süreçlerinde giderek daha fazla kullanılan yöntemler arasında yer almakta olup, yangın tahmini, yangın dinamiklerinin modellenmesi (Smith, 2011) ve risk analizine dayalı müdahale stratejilerinin oluşturulması gibi kritik süreçlerde etkili bir şekilde kullanılmaktadır (Kavlak, Kurtipek, & Çabuk, 2020). Günümüzde birçok araştırma, makine öğrenmesi tekniklerinin yangın yönetiminde kullanılabileceğini göstermekte, ancak bu çalışmaların büyük bir kısmı orman yangınları üzerine yoğunlaşmaktadır. Kentsel yangınların tahmin edilmesi konusunda ise literatürde sınırlı sayıda araştırma bulunmaktadır (Aksoy, Korucu, Çalışkan, Osmanbey & Halis, 2021). Bu bağlamda, yangın tahmin modelinin geliştirilmesi, yerel yönetimler ve itfaiye teşkilatları için bilimsel ve uygulamalı bir katkı sunacaktır.

Bu tez çalışması, yangın verilerini kullanarak makine öğrenmesi tabanlı bir tahminleme modeli geliştirmeyi amaçlamaktadır. Çalışmada Karar Ağacı (Fine Tree), Destek Vektör Makineleri (SVM) ve Gaussian Naive Bayes (GNB) algoritmaları kullanılarak yangın türlerinin sınıflandırılması yapılacak ve modellerin doğruluk oranları karşılaştırılacaktır. Bu analizler sayesinde, yangın çıkma olasılığı yüksek bölgeler tespit edilerek, karar alıcılara yönelik önleyici stratejiler geliştirilmeye yönelik bilimsel bir temel sunulacaktır.

Sonuç olarak, bu çalışma yangın yönetiminde makine öğrenmesi tekniklerinin uygulanabilirliğini inceleyerek, veri odaklı bir yangın tahmin modeli geliştirmeyi amaçlamaktadır. Geliştirilecek modelin, yangın risklerinin erken belirlenmesine, yangın önleme stratejilerinin güçlendirilmesine ve afet yönetiminde yapay zekâ tabanlı çözümler geliştirilmesine katkı sağlaması beklenmektedir.

1.1. Çalışmanın Amacı ve Önemi

Yangınlar, insan hayatını tehdit eden ve büyük maddi kayıplara yol açan doğal ve insan kaynaklı afetler arasında yer almaktadır. Yangınların kontrol altına alınması ve can ile mal kaybının en aza indirilmesi, etkili müdahale ve önceden alınan önlemlerle mümkün olmaktadır. Bu bağlamda, yangın verilerinin analiz edilmesi ve tahmin

edilmesi, itfaiye ekiplerinin operasyonel süreçlerini optimize etmek ve yangın önleme stratejilerini geliştirmek açısından kritik bir öneme sahiptir.

Son yıllarda, büyük veri analitiği ve makine öğrenmesi teknikleri, yangın olaylarının tahmin edilmesi ve erken uyarı sistemlerinin geliştirilmesi konusunda önemli fırsatlar sunmaktadır. Geleneksel istatistiksel yöntemler yangın tahmini konusunda belirli ölçüde başarılı olsa da yangın dinamiklerinin karmaşıklığı ve çok boyutlu veri yapıları nedeniyle daha esnek ve güçlü modelleme yöntemlerine ihtiyaç duyulmaktadır. Bu çalışmada, makine öğrenmesi algoritmalarının yangın verileri üzerinde uygulanarak, yangınların türlerini ve etkilerini tahmin eden bir model geliştirilmesi amaçlanmaktadır.

Bu çalışmanın temel hedefleri şu şekilde özetlenebilir:

- Yangın verilerini analiz ederek önemli değişkenleri belirlemek,
- Farklı makine öğrenmesi modelleri (Karar Ağacı, Gaussian Naive Bayes, Destek Vektör Makineleri) kullanarak yangın türlerinin sınıflandırmasını yapmak,
- Modellerin doğruluk, hata oranları, karışıklık matrisi ve ROC eğrisi gibi performans ölçütleri ile karşılaştırılmasını gerçekleştirmek,
- Elde edilen bulgular doğrultusunda yangın tahmini için en uygun modelin belirlenmesi ve gelecekteki yangın yönetim stratejilerine katkı sağlamak.

Bu çalışmanın önemi birkaç temel noktada değerlendirilebilir. İlk olarak, yangın olaylarının veri odaklı analiz edilmesi, geleneksel yöntemlere kıyasla daha hassas ve öngörülebilir bir tahmin süreci sunmaktadır. İtfaiye teşkilatları ve afet yönetimi birimleri için geliştirilecek tahminleme modelleri, müdahale sürelerini optimize etmek, yangın şiddetine bağlı riskleri belirlemek ve kaynak yönetimini iyileştirmek açısından büyük bir katkı sağlayacaktır.

İkinci olarak, makine öğrenmesi algoritmalarının yangın verilerine uygulanması, Türkiye’de bu alandaki akademik çalışmalara katkı sunacaktır. Literatürde yangın tahmini ile ilgili yapılan çalışmaların çoğu orman yangınlarına odaklanmakta olup, kentsel yangınların tahmini konusunda sınırlı sayıda araştırma bulunmaktadır. Bu bağlamda, bu çalışmanın sonuçları, kentsel alanlarda yangın risk analizine yönelik yeni yaklaşımlar geliştirmek için önemli bir referans olacaktır.

Son olarak, bu araştırma, veri odaklı karar destek sistemlerinin geliştirilmesine katkıda bulunarak, yerel yönetimlerin ve itfaiye birimlerinin yangın önleme politikalarını güçlendirilmesine yardımcı olacaktır. Gelecekte daha büyük ölçekli yangın analizleri ve tahminleme modelleri oluşturmak için bir temel sağlayarak, afet yönetimi konusunda veri bilimi ve yapay zekâ tekniklerinin kullanımını teşvik edecektir.

Bu çalışma, yangın önleme ve müdahale süreçlerinin daha etkin hale getirilmesi amacıyla makine öğrenmesi tabanlı modellerin potansiyelini ortaya koyarak, bilimsel ve uygulamalı bir katkı sunmayı hedeflemektedir.

1.2. Problem Tanımı ve Araştırma Soruları

Yangınlar, dünya genelinde can ve mal kayıplarına neden olan, ekosistemler üzerinde uzun vadeli etkiler bırakan afetler arasında yer almaktadır. Özellikle kentleşme, sanayileşme ve iklim değişikliğinin etkisiyle, yangınların sıklığı ve şiddeti artmakta, bu durum yangın yönetimi stratejilerinin daha etkin hale getirilmesini zorunlu kılmaktadır (Altıparmak & Sezgin, 2023). Geleneksel yangın yönetimi yöntemleri genellikle geçmiş verilere dayalı reaktif müdahale stratejilerine odaklanırken, modern teknolojiler sayesinde proaktif yangın tahmini ve risk değerlendirme yöntemleri geliştirilmesi mümkün hale gelmiştir (Rao, Eisenberg, & Schmitt, 2007).

Ülkemiz, geniş ormanlık alanları, sanayi bölgeleri ve kırsal yerleşimleri ile yangın riski açısından kritik bölgelerden biri olarak öne çıkmaktadır. Bölgede meydana gelen yangınlar farklı nedenlere bağlı olarak ortaya çıkmakta ve yangınların mekânsal ve zamansal dağılımı tahmin edilmesi zor bir dinamik sergilemektedir (Pekşen, Uyaroğlu, & Soyhan, 2020). Sanayi yangınları, orman yangınları ve kentsel yangınlar farklı karakteristiklere sahip olduğundan, bu yangınların türlerini ve nedenlerini doğru bir şekilde modelleyerek tahminleme yapmak büyük bir önem taşımaktadır.

Makine öğrenmesi teknikleri, yangın tahmini süreçlerinde büyük veri setlerini analiz ederek olası yangınların meydana gelme ihtimalini belirlemede güçlü bir araç olarak kullanılmaktadır (Yavuz & Bozatay, 2015). Bu teknikler, geçmiş yangın verilerinden öğrenerek risk haritaları oluşturabilir, müdahale süreçlerini optimize edebilir ve karar alıcılara erken uyarı sistemleri sunabilir (Aksoy, Korucu, Çalışkan, Osmanbey, & Halis, 2021). Ancak yangınların tahmin edilmesi sürecinde hangi makine öğrenmesi modelinin daha yüksek doğruluk sunduğu ve hangi değişkenlerin yangın türlerini belirlemede daha etkili olduğu gibi konular literatürde yeterince detaylandırılmamıştır.

Bu çalışma, yangın verilerini kullanarak farklı makine öğrenmesi algoritmalarının yangın tahmin performanslarını karşılaştırmayı ve yangın yönetimine katkı sağlayacak bir model geliştirmeyi amaçlamaktadır. Çalışmada Karar Ağacı (Fine Tree), Gaussian Naive Bayes (GNB) ve Destek Vektör Makineleri (SVM) algoritmaları kullanılarak yangınların sınıflandırılması yapılacak ve modellerin doğruluk oranları ile hata payları analiz edilecektir. Bu sayede, itfaiye ve yerel yönetimlerin erken müdahale kapasitesini artıracak karar destek sistemlerinin geliştirilmesine bilimsel bir katkı sunulması hedeflenmektedir.

Bu bağlamda, çalışma aşağıdaki araştırma sorularına yanıt aramaktadır:

- Makine öğrenmesi modelleri, yangın verileri üzerinde nasıl bir performans göstermektedir ve hangi model en yüksek doğruluk oranına sahiptir?
- Yangın türleri arasındaki sınıflandırma hataları hangi faktörlerden kaynaklanmaktadır ve model doğruluğunu artırmak için hangi optimizasyon teknikleri uygulanabilir?
- Karar Ağacı (Fine Tree), Gaussian Naive Bayes (GNB) ve Destek Vektör Makineleri (SVM) modellerinin doğruluk oranları nasıl karşılaştırılmaktadır ve hangi model yangın tahmininde daha güçlüdür?
- Yangın tahmin modellerinin geliştirilmesinde karşılaşılan temel zorluklar nelerdir ve bu zorlukları aşmak için hangi veri işleme teknikleri uygulanabilir?

Bu araştırma soruları, yangın tahmin modellerinin daha verimli hale getirilmesi, veri odaklı tahminleme sistemlerinin güçlendirilmesi ve makine öğrenmesi tekniklerinin afet yönetimi süreçlerine entegrasyonu gibi konulara ışık tutarak, yangın yönetiminde yeni bir perspektif sunmayı hedeflemektedir.

1.3. Çalışmanın Kapsamı ve Sınırları

Bu çalışma, gerçekleşen yangın olaylarına ilişkin verilerin incelenmesi ve bu verilerden hareketle sınıflandırma temelli tahminleme modellerinin geliştirilmesini kapsamaktadır. Kullanılan veri seti, ilgili yerel kurumların itfaiye kayıtlarından elde edilmiş; yangınların türü, çıkış nedeni, konumu, zamanı ve bazı etkileri gibi nitelikleri içermektedir. Çalışma, veri madenciliği ve makine öğrenmesi yöntemlerinin afet yönetimi alanında uygulanabilirliğini araştırmayı amaçlamaktadır.

Çalışmanın kapsamı aşağıdaki şekilde özetlenebilir:

- Coğrafi kapsam: Bir ilin idari sınırları içinde meydana gelen yangın vakaları ile sınırlıdır.
- Zamansal kapsam: Yangın verileri belirli bir tarih aralığını kapsamakta olup, analizler bu döneme özgü yapılmıştır.
- Yöntemsel kapsam: Makine öğrenmesi yöntemlerinden Karar Ağacı (Fine Tree), Gaussian Naive Bayes (GNB) ve Linear Support Vector Machine (SVM) algoritmaları uygulanmıştır.
- Araçlar: Modelleme ve analiz süreçlerinde MATLAB yazılımı ve içindeki Makine Öğrenmesi modülleri kullanılmıştır.
- Değerlendirme ölçütleri: Sınıflandırma algoritmalarının başarısı doğruluk oranı, karışıklık matrisi, ROC eğrisi, AUC değeri ve benzeri metrikler aracılığıyla değerlendirilmiştir.

Çalışmanın sınırları ise şu şekildedir:

- Veri sınırlılığı: Kullanılan veri seti yalnızca bir ile ait yangın kayıtlarını içermektedir. Bu nedenle farklı bölgelere genellenebilirlik sınırlı olabilir.
- Model kapsamı: Yalnızca belirli sayıda klasik makine öğrenmesi algoritması incelenmiş; derin öğrenme, zaman serisi modelleme ya da birleşik yöntemler bu çalışmanın dışında tutulmuştur.
- Öznitelik içeriği: Yangına etki eden dışsal değişkenlerden bazıları (örneğin iklimsel koşullar, bina yoğunluğu, trafik durumu) veri setinde yer almadığından, modelleme sürecine dahil edilememiştir.

Bu kapsam ve sınırlar çerçevesinde, çalışma yalnızca eldeki veri seti ve seçili yöntemlerle yürütülmüş olup, bulgular bu bağlamda değerlendirilmelidir. Çalışma, afet verilerinin sınıflandırılması ve bu alanda yapay zekâ tekniklerinin kullanımı açısından örnek bir uygulama sunmayı hedeflemektedir.

2. KAVRAMSAL ÇERÇEVE VE LİTERATÜR TARAMASI

Yangınlar, doğal ve insan kaynaklı afetler arasında yer almakta olup, meydana geldikleri bölgelerde önemli can ve mal kayıplarına yol açmaktadır. Yangın yönetimi ve tahmini süreçlerinde geleneksel yöntemler uzun yıllar boyunca kullanılmış olsa da veri odaklı analiz tekniklerinin gelişmesiyle birlikte makine öğrenmesi tabanlı yaklaşımlar yangın tahmini (Kaya H. , 2018) ve risk değerlendirme alanında önemli bir yere sahip olmuştur (Rao, Eisenberg, & Schmitt, 2007).

Bu bölümde, öncelikle yangın analizleri ve geleneksel risk değerlendirme yöntemleri ele alınacak, ardından makine öğrenmesi tekniklerinin yangın tahminine uygulanabilirliği incelenecektir. Makine öğrenmesi tabanlı tahmin modellerinin nasıl çalıştığı, hangi faktörlerin model doğruluğunu artırdığı ve bu tekniklerin başarısının nasıl değerlendirildiği detaylandırılacaktır. Son olarak, literatürde yapılan benzer çalışmalar incelenerek bu çalışmanın akademik alana sunduğu katkılar açıklanacaktır.

2.1. Yangın Analizleri ve Risk Değerlendirme Yaklaşımları

Yangınlar, çıkış nedenleri, yayılma dinamikleri ve etkiledikleri alanlara göre farklı analiz teknikleri gerektiren afetler arasında yer almaktadır (Altıparmak & Sezgin, 2023). Yangınlar, doğrudan can ve mal kayıplarına neden olmanın yanı sıra, ekosistemler üzerinde uzun vadeli etkiler yaratmakta ve ekonomik kayıplara yol açmaktadır. Yangınların neden olduğu zararları en aza indirmek ve yangın yönetimi stratejilerini optimize etmek için risk değerlendirme ve tahminleme süreçlerinin doğru bir şekilde yönetilmesi gerekmektedir (Rao, Eisenberg, & Schmitt, 2007).

Yangın riski, yangının meydana gelme olasılığını ve olası etkilerini değerlendirmek için kullanılan risk analizi ve modelleme yöntemleri ile belirlenmektedir. Risk analizi, yangının oluşmasını tetikleyen faktörleri belirlemek, yangınların mekânsal ve zamansal dağılımını analiz etmek ve olası etkilerini modellemek amacıyla gerçekleştirilir (Yavuz & Bozatay, 2015). Geleneksel yangın analiz yöntemleri, genellikle istatistiksel modelleme, mekânsal analizler, uzman görüşlerine dayalı

değerlendirmeler ve Coğrafi Bilgi Sistemleri (GIS) tabanlı yaklaşımlar gibi teknikleri içermektedir (Smith, 2011).

2.1.1. Yangın risk faktörleri

Yangınların oluşumunu etkileyen risk faktörleri, meteorolojik, topoğrafik ve insan kaynaklı olmak üzere üç temel gruba ayrılmaktadır.

2.1.1.1. Meteorolojik faktörler

Meteorolojik koşullar, yangının meydana gelme olasılığını doğrudan etkileyen en önemli faktörler arasındadır. İklim değişikliği ve artan sıcaklıklar, yangın riskini önemli ölçüde artırmaktadır (Pekşen, Uyaroğlu, & Soyhan, 2020). Yangın riskini belirleyen başlıca meteorolojik faktörler şunlardır:

- Sıcaklık: Yüksek sıcaklıklar, yangınların başlamasına neden olan koşulları güçlendirir ve yangının ilerleme hızını artırır.
- Bağıl Nem: Düşük nem oranı, bitki örtüsünün kurumasına ve yanmaya daha elverişli hale gelmesine neden olur.
- Rüzgâr Hızı ve Yönü: Rüzgar, yangının yayılma hızını ve yönünü belirleyen en kritik faktörlerden biridir. Güçlü rüzgarlar, yangının kontrol altına alınmasını zorlaştırmaktadır.
- Yağış Miktarı: Uzun süreli kuraklıklar, bitki örtüsünü daha yanıcı hale getirmekte ve yangın ihtimalini artırmaktadır.

Meteorolojik faktörlerin etkileri, özellikle mevsimsel değişkenlik ile bağlantılı olup, yaz aylarında yangın riskinin belirgin şekilde arttığı bilinmektedir (Aksoy, Korucu, Çalışkan, Osmanbey, & Halis, 2021).

2.1.1.2. Topoğrafik faktörler

Yangınların yayılma dinamikleri, arazi yapısına ve mekânsal değişkenlere bağlı olarak farklılık göstermektedir. Topoğrafik faktörler, yangının ilerleyişini ve büyümesini etkileyen fiziksel unsurlardır. Bunlar:

- Eğim ve Yükseklik: Yangınlar, eğimli arazilerde yukarı doğru daha hızlı ilerleyebilir. Yüksek rakımlı bölgelerde oksijen miktarının azalması, yangının yayılma hızını etkileyebilir.

- Bitki Örtüsü Yoğunluğu: Yangının hızla yayılmasını sağlayan yanıcı madde miktarıdır. Ormanlık alanlar, çalılıklar ve kuru otlar yangınların başlangıç ve yayılım süreçlerinde kritik rol oynar.
- Arazi Kullanımı: Tarım arazileri, yerleşim bölgeleri ve sanayi alanları gibi farklı arazi türleri, yangın riskini farklı şekillerde etkileyebilir (Gökkaya & Kaya, 2021).

2.1.1.3. İnsan kaynaklı faktörler

Yangınların büyük bir kısmı insan faaliyetlerinden kaynaklanmaktadır. Bu faktörler, doğrudan veya dolaylı olarak yangınların çıkmasına ve yayılmasına neden olabilir. Özellikle kentsel ve sanayi alanlarında yangın riski, insan etkisiyle büyük ölçüde artmaktadır. Başlıca insan kaynaklı yangın risk faktörleri şunlardır:

- Tarımsal Faaliyetler: Tarım alanlarında anız yakma gibi uygulamalar, kontrolsüz yangınlara neden olabilmektedir.
- Sanayi Kazaları: Kimyasal fabrikalar, petrol rafinerileri ve enerji santralleri, yangın riski yüksek alanlar arasındadır.
- Elektrik İletim Hatları: Eski ve bakımsız elektrik hatları, kısa devre nedeniyle kıvılcımlara ve büyük çaplı yangınlara yol açabilir.
- İnsan İhmali: Sigara izmaritleri, dikkatsizce yakılan ateşler ve atılan cam şişeler gibi insan ihmalleri, birçok yangının temel sebebidir (Kaftandjieva, 2010).

Bu faktörlerin modellenmesi ve analizi, yangın tahmini sistemlerinin daha doğru çalışmasını sağlamak ve risk haritalarının oluşturulmasına katkıda bulunmaktadır (Smith, 2011).

2.1.2. Yangın risk değerlendirme yaklaşımları

Yangınların olasılığını belirlemek ve risk haritalarını oluşturmak için farklı analiz yöntemleri geliştirilmiştir. Bu yaklaşımlar, yangının çıkış nedenlerini ve yayılım dinamiklerini daha iyi anlamaya yöneliktir.

2.1.2.1. Geleneksel yangın risk analizi yöntemleri

Geleneksel yöntemler genellikle:

- İstatistiksel Modeller: Geçmiş yangın verilerine dayalı olarak yangın olasılığını tahmin eden istatistiksel yaklaşımlar kullanılır (Yavuz & Bozatay, 2015).
- Coğrafi Bilgi Sistemleri (GIS): Mekânsal analizler yaparak yangın risk haritalarını oluşturan güçlü bir yöntemdir (Smith, 2011).
- Uzaktan Algılama Teknikleri: Uydu görüntüleri ve hava fotoğrafları kullanılarak orman yangınlarının erken tespit edilmesini sağlar (Kaftandjieva, 2010).

2.1.2.2. Veri tabanlı ve makine öğrenmesi yaklaşımları

Geleneksel yöntemlerin sınırlamaları nedeniyle, makine öğrenmesi ve büyük veri analitiği tabanlı yaklaşımlar ön plana çıkmaktadır. Bu yaklaşımlar, çok büyük veri kümelerini analiz ederek karmaşık desenleri ve ilişkileri keşfetmek için daha etkili çözümler sunmaktadır (Keçeoğlu, Gelbal, & Doğan, 2016).

- Makine öğrenmesi tabanlı yangın tahmini modelleri:
- Veri madenciliği teknikleri kullanarak yangın örüntülerini analiz eder.
- Farklı faktörler arasındaki ilişkileri ortaya koyarak daha hassas tahminler yapar.
- Gerçek zamanlı veri işleme yeteneği sayesinde yangın risk haritalarını dinamik hale getirir.

Bu bağlamda, geleneksel yöntemlerin eksik kaldığı noktalarda makine öğrenmesi teknikleri devreye girmekte ve yangın tahmini süreçlerini daha güvenilir hale getirmektedir (Urul, 2017).

Yangın risk analizleri, meteorolojik, topoğrafik ve insan kaynaklı faktörlerin modellenmesi ile gerçekleştirilmekte ve geleneksel yöntemlerin yetersiz kaldığı durumlarda makine öğrenmesi teknikleri önemli katkılar sunmaktadır. Bu nedenle, gelişmiş veri analitiği teknikleri ve makine öğrenmesi tabanlı modellerin kullanılması, yangın tahmini ve yönetimi süreçlerinde önemli bir ilerleme sağlamaktadır.

2.2. Yangın Tahmininde Kullanılan Makine Öğrenmesi Teknikleri

Geleneksel yangın tahmini yöntemleri genellikle istatistiksel analizlere, geçmiş verilere dayalı modellemelere ve uzman görüşlerine dayanırken, makine öğrenmesi

tabanlı yaklaşımlar, büyük veri analitiğini kullanarak daha yüksek doğruluk oranları ile yangın riskini tahmin edebilme potansiyeline sahiptir (Keçeoğlu, Gelbal, & Doğan, 2016). Makine öğrenmesi, karmaşık veri setlerinden anlamlı desenler çıkarmayı, belirli risk faktörleri ile yangın olasılığı arasındaki ilişkileri öğrenmeyi ve gelecekte oluşabilecek yangınları tahmin etmeyi sağlayan bir yapay zekâ disiplindir.

Makine öğrenmesi teknikleri, denetimli öğrenme (supervised learning), denetimsiz öğrenme (unsupervised learning) ve pekiştirmeli öğrenme (reinforcement learning) olmak üzere üç ana kategoriye ayrılmaktadır (Urul, 2017). Yangın tahmini bağlamında en yaygın kullanılan yöntemler, geçmiş yangın verilerine dayalı olarak tahminler yapan denetimli öğrenme algoritmalarıdır. Bu algoritmalar, yangınların oluşma olasılığını belirlemek ve mekânsal risk haritaları oluşturmak için kullanılmaktadır.

Yangın tahmini ve risk değerlendirme süreçlerinde farklı makine öğrenmesi algoritmaları kullanılmaktadır. Bu algoritmaların etkinliği, veri setinin yapısına, model parametrelerine ve tahmin edilen değişkenlerin doğruluğuna bağlı olarak değişmektedir. En yaygın kullanılan makine öğrenmesi teknikleri aşağıda detaylandırılmıştır.

Denetimli öğrenme, makine öğrenmesi modellerinin en yaygın kullanılan türlerinden biridir ve özellikle yangın tahmini süreçlerinde büyük öneme sahiptir. Bu yöntem, giriş verileri ile etiketlenmiş çıktılar arasındaki ilişkileri öğrenerek yeni verilere dayalı tahminler yapmayı amaçlar (Gökkaya & Kaya, 2021). Yangın tahmini çalışmalarında kullanılan denetimli öğrenme algoritmaları, yangının çıkma olasılığını belirleme, yangın türünü sınıflandırma ve yangın yayılma hızını modelleme gibi kritik görevlerde etkili bir şekilde uygulanmaktadır. Bu bağlamda en yaygın kullanılan denetimli öğrenme algoritmaları arasında karar ağaçları, rastgele ormanlar, destek vektör makineleri ve yapay sinir ağları yer almaktadır.

Karar ağaçları, yangın riskini etkileyen faktörleri (örneğin sıcaklık, nem, rüzgar hızı, bitki örtüsü, yerleşim yoğunluğu) belirleyerek tahmin yapan hiyerarşik bir modeldir (Keçeoğlu, Gelbal, & Doğan, 2016). Yangın tahmini süreçlerinde karar ağaçları, verilerin mantıksal olarak ayrılmasına olanak tanıyan yapısıyla kullanışlıdır. Bununla birlikte, rastgele ormanlar, birden fazla karar ağacının birleşiminden oluşarak daha sağlam tahminler üretir ve aşırı öğrenme (overfitting) sorununu minimize eder. Bu

nedenle, yangın riski belirleme süreçlerinde rastgele ormanlar sıklıkla tercih edilmektedir (Urul, 2017).

Destek vektör makineleri (SVM), özellikle büyük veri kümelerinde yüksek doğruluk oranı sağlayan güçlü bir sınıflandırma algoritmasıdır. Yangın tahmini süreçlerinde, mekânsal ve çevresel değişkenlerin etkili bir şekilde sınıflandırılmasını sağlayarak yangın risk bölgelerini belirleme açısından avantaj sağlamaktadır (Kaftandjieva, 2010). Bunun yanı sıra, yapay sinir ağları (ANN), insan beyninin çalışma mekanizmasını taklit ederek büyük veri setlerinde karmaşık ilişkileri öğrenme yeteneğine sahiptir. Derin öğrenme modelleri ile entegre edildiğinde, mekânsal ve zamansal yangın dinamiklerini analiz etmede üstün başarı gösterir (Gökkaya & Kaya, 2021).

Denetimsiz öğrenme yöntemleri ise, etiketlenmemiş veri kümeleri üzerinde çalışarak veri içindeki örüntüleri ve yapısal ilişkileri keşfetmeye yönelik bir yaklaşımdır (Urul, 2017). Yangın tahmini süreçlerinde denetimsiz öğrenme algoritmaları, bilinmeyen yangın örüntülerini belirleme, yangın risk bölgelerini haritalama ve yangınların mekânsal yayılımını analiz etme amacıyla kullanılmaktadır. Özellikle kümeleme (clustering) teknikleri, yangın riski taşıyan bölgeleri sınıflandırarak mekânsal analizler yapmaya imkân tanımaktadır.

K-Means kümeleme algoritması, yangın riskine sahip bölgeleri belirlemek için kullanılan popüler bir yöntemdir. Yangın olasılığına bağlı olarak verileri belirli gruplara ayırarak, benzer risk seviyelerine sahip alanları tespit edebilir (Keçeoğlu, Gelbal, & Doğan, 2016). Bunun yanı sıra, hiyerarşik kümeleme, yangın risk faktörlerini detaylı şekilde gruplayarak yangın olaylarının mekânsal analizine katkıda bulunur. Bu yöntem, yangın bölgelerinin farklı risk seviyelerine göre katmanlandırılmasını sağlayarak karar destek sistemlerinde kullanılmaktadır.

Makine öğrenmesi teknikleri içinde önemli bir yere sahip olan pekiştirmeli öğrenme, bir ajanın çevresiyle etkileşim kurarak ödül mekanizmasına dayalı öğrenmesini sağlayan bir yaklaşımdır (Urul, 2017). Yangın tahmini ve müdahale süreçlerinde optimal stratejilerin belirlenmesi için kullanılan pekiştirmeli öğrenme algoritmaları, özellikle yangın söndürme ekiplerinin kaynak yönetimini optimize etmeye ve yangın müdahale süreçlerini iyileştirmeye yönelik uygulamalarda yaygın olarak tercih edilmektedir.

Q-öğrenme (Q-Learning), büyük ölçekli yangın söndürme operasyonlarında kaynakların en iyi şekilde yönetilmesini sağlamak için kullanılan bir pekiştirmeli öğrenme tekniğidir. Bu model, itfaiye ekiplerinin en verimli müdahale rotalarını belirleyerek yangınlarla mücadelede karar destek sistemlerini güçlendirmektedir. Özellikle karmaşık afet yönetim süreçlerinde, Q-öğrenme gibi algoritmalar, değişken koşullara uyum sağlayarak en iyi müdahale stratejilerini oluşturabilir.

Derin öğrenme yöntemleri, özellikle büyük veri setleri ile çalışırken yüksek doğruluk oranlarına ulaşabilen güçlü yapay zeka modelleridir (Gökkaya & Kaya, 2021). Yangın tahmini süreçlerinde derin öğrenme teknikleri, hem mekânsal hem de zamansal verilerin işlenmesini sağlayarak yangın olasılığına dair daha detaylı tahminler üretebilmektedir. Evrişimli sinir ağları (CNN), özellikle uydu görüntüleri ve hava fotoğrafları gibi mekânsal verileri analiz ederek yangın risk haritalarının oluşturulmasında büyük avantaj sağlamaktadır. Yangınların erken tespiti ve yayılma tahminleri açısından bu model oldukça etkilidir.

Tekrarlayan sinir ağları (RNN) ve uzun kısa süreli bellek (LSTM) modelleri ise, zaman serisi verilerini analiz ederek yangın riskinin zaman içindeki değişimini tahmin etmek için kullanılmaktadır. Yangın mevsimselliğini ve trendlerini değerlendiren bu modeller, uzun vadeli tahminler yaparak yangın yönetim planlarının oluşturulmasına yardımcı olmaktadır. Ayrıca, generatif adversaryal ağlar (GAN) gibi yapay zeka teknikleri, anormal yangın örüntülerini belirleyerek risk analizlerinde kullanılabilir.

Son olarak, ensemble öğrenme yöntemleri, birden fazla modelin bir araya getirilerek daha güçlü tahminler oluşturmasını sağlayan tekniklerdir. Yangın tahmini süreçlerinde, özellikle hata oranlarını düşürmek ve tahmin doğruluğunu artırmak amacıyla yaygın olarak kullanılmaktadır. Gradient Boosting Machines (GBM), birden fazla zayıf öğreniciyi birleştirerek yüksek doğruluklu tahminler üreten güçlü bir modeldir. Bunun yanı sıra, XGBoost ve LightGBM gibi optimizasyon teknikleri, büyük veri setlerinde hızlı çalışarak yangın risk analizi uygulamalarında önemli avantajlar sunmaktadır.

Makine öğrenmesi tabanlı yangın tahmini modelleri, yangın yönetiminde veri odaklı karar alma süreçlerini destekleyerek risk değerlendirme ve erken müdahale sistemlerinin geliştirilmesine katkıda bulunmaktadır.

2.3. Yangın Yönetiminde Veri Madenciliği ve Özellik Seçimi

Yangın yönetimi, yangın riskinin belirlenmesi, yangınların önlenmesi, erken müdahale stratejilerinin geliştirilmesi ve kaynak tahsisi süreçlerini kapsayan çok boyutlu bir süreçtir. Bu süreçlerde doğru kararlar alınabilmesi için büyük miktarda verinin toplanması, işlenmesi ve analiz edilmesi gerekmektedir. Günümüzde geleneksel analiz yöntemleri yangın tahmini ve (Balcı & Altundağ, 2023) yönetimi için yetersiz kalmakta, bu nedenle veri madenciliği ve makine öğrenmesi teknikleri yangın yönetiminde giderek daha fazla kullanılmaktadır (Yıldırım, Günay, & Yağanoğlu, 2023). Veri madenciliği, büyük veri setlerinden anlamlı desenleri çıkarmak, bilinmeyen ilişkileri keşfetmek ve karar destek sistemlerine katkı sağlamak için kullanılan bir süreçtir. Özellikle yangın yönetiminde, mekânsal, zamansal ve meteorolojik değişkenleri analiz etmek için güçlü bir araç olarak görülmektedir (Pang, ve diğerleri, 2022). Bu bağlamda, veri madenciliği teknikleri yangın tahmini, yangın risk haritalarının oluşturulması ve (Pham, ve diğerleri, 2020) yangın müdahale süreçlerinin optimize edilmesi gibi kritik alanlarda uygulanmaktadır; (Mishra, ve diğerleri, 2024).

Yangın yönetiminde veri madenciliği teknikleri, yangın tahmini ve erken uyarı sistemlerinin geliştirilmesinde önemli bir rol oynamaktadır. Mevcut hava koşulları, bitki örtüsü, coğrafi yapı ve geçmiş yangın verileri kullanılarak yangın riskinin belirlenmesi mümkündür (Shahzad, ve diğerleri, 2024). Veri madenciliği, özellikle makine öğrenmesi modelleri ile birleştirilerek erken uyarı sistemlerinin oluşturulmasına katkı sağlar. Yangın risk haritalarının oluşturulması, yangın olasılığı yüksek bölgeleri belirlemek ve yangın yönetim süreçlerinde önceliklendirme yapmak için kullanılan bir diğer veri madenciliği uygulamasıdır (Balcı & Altundağ, 2023). Bu haritalar, mekânsal analizler, Coğrafi Bilgi Sistemleri (GIS) ve veri madenciliği teknikleriyle oluşturularak yangın müdahale ekiplerine daha etkin planlama imkânı sunmaktadır (Pham et al., 2020). Bunun yanı sıra, yangın söndürme ekiplerinin yönlendirilmesi, kaynak yönetimi ve lojistik planlamalar, veri madenciliği ile optimize edilerek müdahale süreleri kısaltılabilir ve yangınların büyümesi önlenabilir (Pang, ve diğerleri, 2022). Ayrıca, yangın sonrası analizler ve hasar değerlendirme süreçlerinde, uydu görüntüleri ve mekânsal analizlerle yangının çevresel etkileri incelenebilir (Mishra, ve diğerleri, 2024).

Veri madenciliği tekniklerinin etkinliği, doğru özellik seçiminden büyük ölçüde etkilenmektedir. Özellik seçimi, analiz edilen veri setinde en fazla bilgi içeren değişkenlerin belirlenmesini ve gereksiz veya düşük etkili değişkenlerin çıkarılmasını sağlayarak modelin performansını artırmaktadır (Peng, Long, & Ding, 2005). Yangın tahmini modellerinde kullanılan özellikler genellikle meteorolojik, coğrafi, çevresel ve insani faktörlerden oluşmaktadır. Meteorolojik faktörler arasında sıcaklık, nem oranı, rüzgar hızı ve yağış miktarı yer almaktadır. Yüksek sıcaklık, düşük nem ve kuvvetli rüzgar, yangın riskini artıran en önemli etkenlerdir (Shahzad, ve diğerleri, 2024). Coğrafi ve çevresel faktörler, yangınların mekânsal dağılımı üzerinde büyük etkiye sahiptir. Bitki örtüsü türü, rakım, eğim ve toprak yapısı gibi değişkenler, yangınların çıkış noktaları ve yayılma hızları üzerinde belirleyici rol oynamaktadır (Mishra, ve diğerleri, 2024). İnsan kaynaklı faktörler ise yangınların büyük bir kısmının sebebi olduğu için yangın tahmini süreçlerinde dikkate alınmaktadır. Yerleşim yoğunluğu, trafik durumu, sanayi bölgeleri ve enerji iletim hatları gibi değişkenler, yangın çıkma olasılığını etkileyen temel unsurlar arasında yer almaktadır (Balcı & Altundağ, 2023).

Özellik seçimi, veri madenciliği süreçlerinde model performansını artırmak ve analiz sürecini optimize etmek için kullanılan önemli bir tekniktir. Yangın tahmini modellerinde yaygın olarak kullanılan özellik seçimi yöntemleri, istatistiksel testler ve makine öğrenmesi tabanlı yöntemlerden oluşmaktadır. Ki-Kare testi ve ANOVA gibi istatistiksel yöntemler, değişkenler arasındaki ilişkiyi analiz ederek hangi değişkenlerin modelde daha büyük bir etkiye sahip olduğunu belirler (Jamil & Khanam, 2023). Recursive Feature Elimination (RFE) ve Minimum Redundancy Maximum Relevance (MRMR) gibi makine öğrenmesi tabanlı özellik seçimi yöntemleri, gereksiz değişkenleri elemine ederek modelin daha az işlem gücü tüketmesini sağlar (Peng, Long, & Ding, 2005) ve doğruluk oranını artırır (Ramírez-Gallego, ve diğerleri, 2016). Bu teknikler, özellikle büyük veri setleri ile çalışırken veri boyutunun azaltılması ve modelin genel başarımının iyileştirilmesi açısından büyük avantaj sunmaktadır. MRMR yöntemi, özellikle yangın tahmini süreçlerinde gereksiz değişkenleri eleyerek optimum değişken setini belirlemek için kullanılmaktadır (Nagy, ve diğerleri, 2024).

Yangın yönetiminde veri madenciliği ve özellik seçimi, yangın riskinin daha doğru tahmin edilmesine ve etkili müdahale stratejilerinin geliştirilmesine katkı

sağlamaktadır. Büyük veri setlerinden anlamlı bilgi çıkarılması, yangın dinamiklerinin daha iyi anlaşılmasını sağlayarak yangın tahmini süreçlerini iyileştirmektedir. Özellik seçimi teknikleri ise modellerin doğruluğunu artırarak erken uyarı sistemlerinin daha etkin çalışmasına yardımcı olmaktadır (Pang, ve diğerleri, 2022).

2.4. Literatür Çalışmaları

Yangınların tahmini ve yönetimi, afet yönetimi disiplininin önemli bir alt alanı olarak öne çıkmakta ve özellikle makine öğrenmesi temelli yaklaşımlar, bu süreçlerde etkili çözümler sunmaktadır. Literatürde yapılan çalışmalar, yangınların çıkış nedenlerini ve yayılma risklerini modellemek, risk haritaları üretmek ve müdahale stratejilerini optimize etmek amacıyla farklı makine öğrenmesi tekniklerini kullanmaktadır.

Makine öğrenmesi, tarihsel olarak 1940'lı yıllarda sinir ağlarının temellerinin atılmasıyla gelişmeye başlamış; 1950'lerde Turing Testi, 1952'de Arthur Samuel'in öğrenen programı ve 1957'de Rosenblatt'ın algılayıcı modeliyle kuramsal temelleri atılmıştır. 1980'lerden itibaren veri madenciliği ile güçlenen bu alan, 2010'lu yıllarda derin öğrenmenin katkısıyla çevresel olayların tahmini gibi karmaşık problemlerde etkili hale gelmiştir. Foote (2021), makine öğrenmesinin günümüzde orman yangınları gibi doğal afetlerin analizinde yüksek başarı gösterdiğini belirtmiştir.

Türkiye'de Tarım ve Orman Bakanlığına bağlı Orman Genel Müdürlüğü verilerine göre, orman yangınlarının çıkış nedenlerinin %55'i ihmal ve dikkatsizlik, %10'u kasıt, %5'i kaza, %15'i doğal nedenler ve %15'i bilinmeyen nedenlere dayanmaktadır. Bu oranlar, veri temelli analizlerin yangın nedenlerini daha net sınıflandırmada önemli rol oynayabileceğini göstermektedir (OGM, 2020). Aydın (2018) tarafından yürütülen bir çalışmada, itfaiye istasyonlarının hizmet durumları beş farklı makine öğrenmesi algoritmasıyla sınıflandırılmış ve mekânsal analiz sonucunda %85,43 oranında başarı elde edilmiştir.

Yıldırım, Günay ve Yağanoğlu (2023) tarafından yapılan çalışmada, NASA'nın yangın veri seti üzerinde altı farklı makine öğrenmesi algoritması uygulanarak en başarılı modellerin Naive Bayes (GNB) ve RO algoritmaları olduğu belirlenmiştir. PCA ve ANOVA gibi öznitelik seçimi yöntemleri ile model eğitimi süresi azaltılmış ve aşırı öğrenme riski düşürülmüştür. Pakistan'da gerçekleştirilen bir araştırmada, MODIS verileri üzerinden Random Forest, SVM, XGBoost ve lojistik regresyon algoritmaları karşılaştırılmış; Random Forest algoritması tüm yangın türlerinde en yüksek doğruluk

ve AUC deęerlerini saęlayarak en bařarılı model olarak belirlenmiřtir (Shahzad, ve dięerleri, 2024). Benzer řekilde in’de yrtlen alıřmada, meteorolojik ve topoęrafik deęiřkenler kullanılarak RF modeli ile %89,25 doęruluk oranı elde edilmiřtir (Pang, ve dięerleri, 2022). Trkiye’de geliřtirilen OYAT (Orman Yangını Analiz Tahmin) modeli, derin renme ve Coęrafı Bilgi Sistemleri (CBS) ile entegre alıřmakta ve %98 doęrulukla yangın risk haritaları retmektedir. Bu modelin zellikle veri gncellenebilirlięi aısından avantaj saęladıęı belirtilmiřtir (Fidanboy, Adar, & Okyay, 2022). Elektrik pano yangınları zelinde yapılan alıřmalarda ise, Gaussian SVM ve KNN algoritmaları %99,3 doęruluk ile en bařarılı modeller olarak ne ıkmıřtır. Bu modellerin, IoT tabanlı kestirimci bakım sistemleri ile entegrasyonu nerilmiřtir (Pekřen, 2024).

Yapay sinir aęları (ANN), lojistik regresyon, bulanık mantık ve CNN gibi yapılar da literatrde yaygın biimde kullanılmaktadır. ok katmanlı yapılar, kablosuz sensr aęları ile birlikte kullanıldığında, veri fzyonu ve enerji verimlilięi gibi avantajlar saęlamaktadır (Abid, 2021). Vietnam’da yrtlen bir alıřmada, Bayes Aęları (BN), Naive Bayes, Karar Aęaçları ve oklu Doęrusal Regresyon modelleri karřılařtırılmıř ve tm modellerin AUC deęerleri 0.90 zeri bulunmuřtur. BN modeli, en yksek doęruluęu saęlamıř; MLR ise karar vericiler aısından daha anlařılır olduęu iin ne ıkmıřtır (Pham et al., 2020). Sharma, Rani ve Memon (2020) tarafından yapılan alıřmada ise sekiz farklı model Azure ortamında test edilmiřtir. ROC, F1 ve hassasiyet gibi metriklerle yapılan karřılařtırmalarda, ensembel tabanlı modellerin doęruluk aısından daha stn olduęu gsterilmiřtir.

Meteorolojik parametrelere dayalı alıřmalar arasında Kanakaraddi ve arkadaşlarının (2021) arařtırmasında, Random Forest modeli ile dřk hata oranları elde edilmiř ve RandomizedSearchCV ile hiperparametre ayarlamalarının doęruluk zerinde etkili olduęu vurgulanmıřtır. Mishra ve dięerlerinin (2024) Hindistan’da yrttę alıřmada, SVM algoritması ile yangın duyarlılık blgeleri belirlenmiř; zellikle yksek biyolojik eřitlilięe sahip blgelerin risk altında olduęu tespit edilmiřtir. ABD Gney Karolina eyaletinde yapılan analizlerde, Karar Aęacı modeli %90,58 doęruluk ile en iyi sonucu vermiřtir. Bu alıřmada ayrıca karbon sıcak noktaları ile yangın alanlarının akıřtıęı ve nleyici stratejiler geliřtirilmesi gerektięi ortaya konmuřtur (Sharma & Khanal, 2024). Yangın tahmini dıřında, makine ęrenmesinin kimyasal bileřiklerin mutajenite tahmininde de etkili olduęu; Zhang ve arkadaşlarının (2017)

alışmasında Naive Bayes ile geliştirilen modellerin yüksek güvenilirlik sağladığı belirtilmiştir.

Bu alışmalar, makine öğrenmesinin farklı yangın türlerine, veri yapılarına ve bölgesel koşullara göre esnek biçimde uyarlanabileceğini ortaya koymaktadır. Tez kapsamında yürütölen bu alışma da gerçek yangın verileri kullanılarak üç farklı sınıflandırma algoritmasının karşılaştırıldığı uygulamalı bir yaklaşım sunmaktadır. Literatürde sıkça vurgulanan sınıf dengesizliği, özellik mühendisliği ve ROC tabanlı değerlendirme ölçütleri bu alışmada da dikkate alınarak, modelleme süreci bilimsel temele oturtulmuştur.

2.5. alışmanın Literatüre Katkısı

Bu alışma, makine öğrenmesi tabanlı yangın tahmini ve yönetimi konusunda literatüre çeşitli açılardan katkı sağlamaktadır. Yangın tahmini üzerine yapılan araştırmaların büyük bir kısmı orman yangınlarına odaklanırken, kentsel ve kırsal yangınların birlikte ele alındığı kapsamlı alışmalar sınırlıdır. Bu bağlamda, alışma, bir ilin yangın verileri kullanılarak yangın tahmini modellerinin performansını incelemekte ve kentsel yangın tahmini süreçlerinde veri odaklı bir yaklaşım sunmaktadır.

Öncelikli olarak, bu alışma farklı makine öğrenmesi algoritmalarının karşılaştırmalı analizini yaparak, yangın tahmini süreçlerinde hangi modellerin daha etkili olabileceğini ortaya koymaktadır. Literatürde, belirli bir model üzerine yoğunlaşan alışmalar yaygınken, farklı modelleme yaklaşımlarının yangın tahminine etkisini sistematik bir şekilde inceleyen araştırmalar daha sınırlıdır. Bu alışmada Karar Ağaçları (Fine Tree), Linear Support Vector Machine (SVM) ve Gaussian Naive Bayes (GNB) gibi farklı makine öğrenmesi algoritmaları test edilmiş ve doğruluk oranları karşılaştırılmıştır.

Ayrıca, alışma kapsamında özellik seçimi süreci ele alınmış ve Ki-Kare testi ile veri setindeki en etkili değişkenler belirlenmiştir. Literatürde, yangın tahmini üzerine yapılan alışmaların büyük bir kısmı geniş veri kümeleri üzerinden doğrudan modelleme yaparken, bu alışma özellik seçimi yöntemleri kullanılarak gereksiz değişkenlerin filtrelenmesini ve modelin genelleme başarısının artırılmasını hedeflemektedir. Elde edilen bulgular, yangın tahmininde kritik değişkenlerin

belirlenmesi ve veri ön işleme süreçlerinin model performansına etkisinin daha iyi anlaşılmasına katkı sağlamaktadır.

Çalışmanın bir diğer katkısı, makine öğrenmesi modellerinin performanslarını detaylı bir şekilde değerlendirerek, hangi modellerin hangi tür yangınları daha doğru tahmin ettiğini analiz etmesidir. Model doğruluk oranlarının yanı sıra, karışıklık matrisleri ve ROC eğrileri kullanılarak model hatalarının hangi sınıflarda yoğunlaştığı incelenmiş ve farklı modellerin güçlü ve zayıf yönleri karşılaştırılmıştır. Bu tür detaylı analizler, literatürde makine öğrenmesi tabanlı yangın tahmini modellerinin sınırlamaları ve avantajları konusunda önemli bilgiler sunmaktadır.

Bunun yanı sıra, çalışmada geliştirilen tahminleme modellerinin, yangın yönetimi süreçlerine nasıl entegre edilebileceğine dair öneriler sunulmaktadır. Yangın yönetimi ve itfaiye operasyonları genellikle saha deneyimlerine dayalı olarak yürütülmekte olup, veri odaklı karar destek sistemleri bu süreçlerin iyileştirilmesinde önemli bir rol oynamaktadır. Bu çalışma, makine öğrenmesi algoritmalarının yangın yönetim süreçlerine entegrasyonunu değerlendirerek, tahminleme modellerinin operasyonel kararlara nasıl katkı sağlayabileceğini göstermektedir.

Sonuç olarak, bu çalışma, yangın tahmini ve yönetimi alanında makine öğrenmesi tabanlı modelleme yöntemlerinin değerlendirilmesine yönelik kapsamlı bir analiz sunarak, literatüre katkıda bulunmaktadır. Farklı modellerin karşılaştırmalı analizi, özellik seçimi süreçlerinin model performansına etkisi ve yangın yönetiminde veri odaklı karar alma mekanizmalarının önemi gibi konular üzerine yaptığı değerlendirmeler ile, akademik ve uygulamalı araştırmalar için yeni bir perspektif sunmaktadır.



3. MAKİNE ÖĞRENMESİ

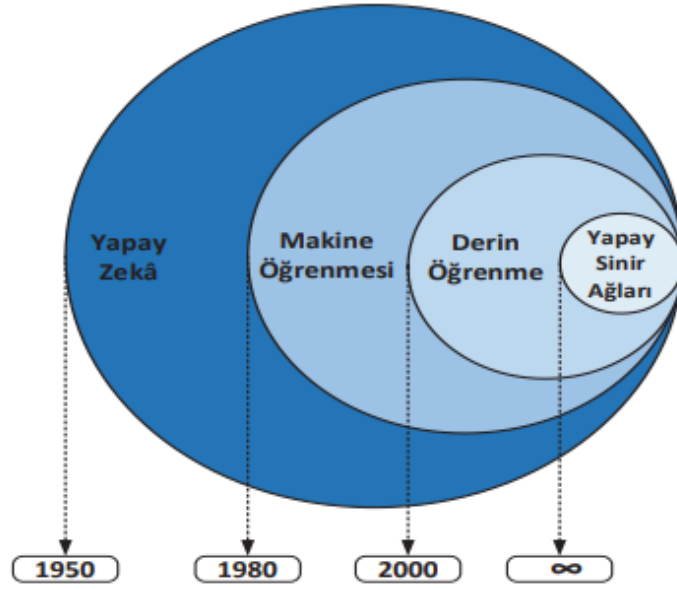
3.1. Makine Öğrenmesinin Tanımı ve Gelişimi

Makine öğrenmesi, ilk olarak 1959 yılında “Arthur Lee Samuel” tarafından IBM (International Business Machines; Uluslararası İş Makineleri) için geliştirilen bir dama oyunu tasarımı sırasında ortaya çıkmıştır (Samuel, 1959). Samuel’in geliştirdiği program, sınırlı bellek kapasitesine rağmen, tahtadaki taşların hamlelerini değerlendiren bir puanlama sistemi kullanarak her iki tarafın oyunu kazanma olasılıklarını ölçmek için tasarlanmıştır (Mockenhaupt, 2021). Veri setlerinden örüntüler öğrenerek olayları tahmin eden ve karar verme süreçlerini iyileştiren yapay zeka tekniklerinden biridir. 1940’lı yıllarda sinir ağlarının temellerinin atılmasıyla ortaya çıkan makine öğrenmesi, zaman içinde büyük veri ve derin öğrenme gibi teknolojilerle gelişimini sürdürmüş, çevresel olayların analizinde etkili bir araç haline gelmiştir. Orman yangınları gibi doğal afetlerin tahmini ve risk analizinde makine öğrenmesi, hızlı ve etkin müdahale sağlamak amacıyla yaygın olarak kullanılmaktadır (Foote, 2021).

Program, stratejik hamleleri belirlemek için “minimax algoritması” gibi teknikleri uygulamış ve zamanla “ezberci öğrenme” olarak adlandırılan bir mekanizma geliştirilmiştir (Gökalp, 2022). Bu mekanizma sayesinde program, yaptığı tüm hamleleri kaydedip hafızasında tutarak ilerleyen oyunlarda daha iyi performans göstermiştir (Foote, 2021). 1962 yılında IBM 7094 bilgisayarı, dama uzmanı Robert Nealey karşısında zafer elde ederek makine öğrenmesinin ilk önemli başarısını göstermiştir (IBM, 2020).

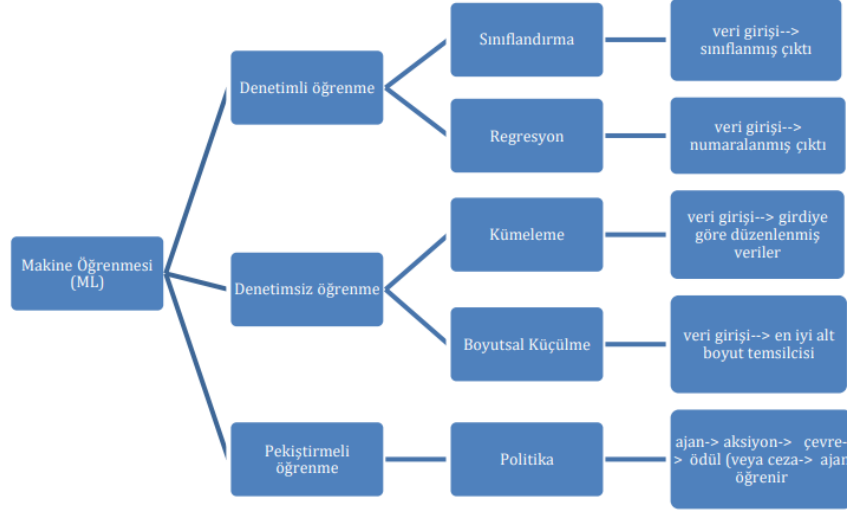
Makine öğrenmesi, günümüzde hayatın hemen her alanında kullanılmaktadır (Zhou, 2021). Biyoteknoloji, bankacılık, eğitim, pazarlama, oyun ve eğlence gibi sektörlerde geniş bir uygulama alanına sahiptir (Gökalp, 2022). Artan veri miktarları ve bu verilerden yeni çıkarımlar elde etme gerekliliği, makine öğrenmesi tekniklerinin önemini artırmıştır (Guerrero, Olivas, Romero, & Viedma, 2015). İnsanların büyük hacimli ve karmaşık verileri hızlı bir şekilde analiz etmesi mümkün olmadığından, makine öğrenmesi devreye girerek ileriye dönük doğru ve etkili analizler yapılmasını

sağlamaktadır (Türkmenoğlu & Tantuğ, 2014). Veri miktarının artmasıyla birlikte makine öğrenmesi algoritmalarının doğruluk payı da yükselmektedir (SAP, 2021).



Şekil 3.1. Makine öğrenimi şeması (Akın & Şahin, 2023).

Şekil 1 de görüldüğü üzere yapay zekanın oluşabilmesi için makine öğrenimi olmak zorundadır (Foote, 2021). Makine öğrenmesi, yapay zekâ sistemlerinin temelini oluşturur. Yapay zekânın oluşabilmesi için makine öğrenmesi yöntemlerine ihtiyaç duyulmaktadır (Gökalp, 2022). Bu yöntemler, algoritmaların istatistiksel yöntemlerle eğitilmesini ve sınıflandırma ya da tahmin yapmalarını sağlamaktadır (Dietterich, 1994). Veri büyüdükçe makine öğrenmesi işlevsiz hale gelmez; aksine, daha doğru tahminler yapar ve sonuçların kesinliğini artırır (IBM, 2020). Yapay zekâ ve makine öğrenmesi sıklıkla birbiriyle karıştırılsa da aslında farklı süreçleri ifade eder (Atalay & Çelik, 2017). Yapay zekâ, makinelerin insan benzeri kararlar almasını ve sorunları çözmesini amaçlarken, makine öğrenmesi verilerden öğrenmeye dayalı bir süreçtir (Balaban & Kartal, 2015). Bu süreçte algoritmalar, programlanmak yerine, kalıpları otomatik olarak tanıyacak şekilde eğitilir. Örneğin, bir algoritma elma ve armut gibi verileri doğru bir şekilde etiketleyebilir ve daha sonra bu bilgileri kullanarak yeni verileri sınıflandırabilir (Gökalp, 2022).



Şekil 3.2. Makine öğrenmesi algoritmaları hiyerarşisi (Rashidi, Tran, Betts, Howell, & Green, 2019).

3.2. Makine Öğrenmesi Algoritmaları

Makine öğrenmesi algoritmaları, denetimli, denetimsiz ve pekiştirmeli öğrenme olmak üzere üç ana kategoriye ayrılır (Özşahin, Yılmaz, & Akcan, 2022). Denetimli öğrenme, genellikle etiketli verilerle çalışır ve hedef değişken ile bağımsız değişkenler arasındaki ilişkiyi öğrenir (Eker, Alkiş, Uçar, & Aydın, 2023). Denetimsiz öğrenmede ise veriler üzerinde herhangi bir etiket bulunmaz ve algoritmanın kendisinin gruplama ya da kalıpları belirlemesi beklenir (Satılmış & Akleylek, 2021). Pekiştirmeli öğrenmede ise algoritma, belirli bir çevreyle etkileşime girerek en iyi sonucu elde etmek için ödül tabanlı bir mekanizma öğrenir (Ülker & Civalet, 2002).

Denetimli öğrenme, sınıflandırma ve regresyon olarak iki alt kategoriye ayrılmaktadır (Eker, Alkiş, Uçar, & Aydın, 2023). Sınıflandırma algoritmaları, nesneleri belirli gruplara ayırmayı amaçlarken, regresyon algoritmaları sayısal veriler arasındaki ilişkileri modellemeyi hedefler (Aylak, Oral, & Yazıcı, 2020). Örneğin, bir öğretmenin öğrencilerine meyve türlerini sınıflandırmayı öğrettiği bir senaryo, denetimli öğrenmenin tipik bir örneğidir. Öğretmen, her meyve türünü tanımlayan özellikleri açıklar ve öğrenciler bu özelliklere dayanarak meyveleri doğru gruplara ayırmayı öğrenir.

Makine öğrenmesi algoritmalarının doğru seçimi, başarıyı belirleyen en önemli faktörlerden biridir (Ülker & Civalet, 2002). Makine öğrenmesinde Doğrusal Regresyon (Linear Regression), Polinomsal Regresyon, Destek Vektör Regresyonu

(Support Vector Regression - SVM), K-En Yakın Komşu (K-Nearest Neighbors - KNN), Karar Ağacı Regresyonu (Decision Tree Regression), Rastgele Orman Regresyonu (Random Forest Regression), Naive Bayes, Yapay Sinir Ağları (Artificial Neural Networks - ANN) en çok bilinenlerdir (Bhavsar & Ganatra, 2012). Tez çalışması kapsamında, yangın verileri üzerinde Karar Ağacı (Fine Tree), Linear SVM ve Gaussian Naive Bayes algoritmaları kullanılmıştır. Bu algoritmalar, yangın risk faktörlerinin analiz edilmesi ve tahmin edilmesi amacıyla seçilmiş ve doğruluk, kesinlik, duyarlılık gibi performans metrikleri kullanılarak değerlendirilmiştir. Elde edilen sonuçlar, makine öğrenmesi modellerinin etkinliğini ve veri odaklı karar verme süreçlerindeki önemini bir kez daha ortaya koymuştur.

- Makine öğrenmesi yöntemleri, sınıflandırma, tahminleme ve öznitelik seçimi gibi adımlarda farklı algoritmalarla uygulanmaktadır. Literatürde yapılan çalışmalar, özellikle yangın risk haritalarının oluşturulmasında ve riskli bölgelerin belirlenmesinde bu yöntemlerin başarısını kanıtlamaktadır.
- Pakistan’da yapılan bir çalışmada, bitki örtüsü yangınlarının tahmini için lojistik regresyon, Destek Vektör Makineleri (SVM) ve XGBoost gibi algoritmalar kullanılmıştır. MODIS Global Fire Atlas verileri üzerinde yapılan analizlerde modellerin doğruluk oranları %66,6 ile %87,5 arasında değişmiştir. Bu çalışma, topografik ve iklimsel değişkenlerin etkili bir şekilde analiz edilerek risk bölgelerinin sınıflandırılmasını sağlamıştır (Shahzad, ve diğerleri, 2024).
- NASA’nın veri seti üzerinde gerçekleştirilen bir başka çalışmada, yangın risklerinin tahmini için 6 farklı makine öğrenmesi algoritması kullanılmıştır: Destek Vektör Makineleri (SVM), Karar Ağaçları (Fine Tree), Gaussian Naive Bayes (GNB), K-En Yakın Komşu (K-NN), Random Forest ve Naive Bayes (GNB). Çalışmada öznitelik seçimi için PCA ve ANOVA yöntemleri uygulanmış ve orijinal veri setindeki 35 öznitelik sırasıyla 15 ve 11 özniteliğe indirgenmiştir. Bu sayede eğitim süreleri kısılırken aşırı uyum riski azaltılmış ve doğruluk oranları korunmuştur. Sonuçlara göre, Gaussian Naive Bayes algoritması %96, Random Forest ise %97 doğruluk oranıyla en başarılı modeller olmuştur. Çalışma, yangın risk tahminlerinde öznitelik seçiminin performansı artırma potansiyeline dikkat çekmiştir (Yıldırım, Günay, & Yağanoğlu, 2023).

- Çin’de gerçekleştirilen çalışmalarda, orman yangınlarını etkileyen ana faktörler meteorolojik, topoğrafik, bitki örtüsü ve insan etkileri olarak belirlenmiştir. Yapay Sinir Ağları (ANN), Radial Basis Function Neural Network (RBFNN) ve Destek Vektör Makineleri (SVM) gibi algoritmalarla yapılan analizlerde, doğruluk oranları %75’in üzerinde bulunmuştur. Bu çalışmalarda özellikle yangın riski haritaları oluşturularak mevsimsel analizler gerçekleştirilmiş ve risk bölgeleri tespit edilmiştir (Pang, ve diğerleri, 2022).

Makine öğrenmesi yöntemleri, öznetelik seçimi ve model doğrulama gibi süreçlerle optimize edilerek, yangın risk sınıflandırması, tahminleme ve risk haritalarının oluşturulması gibi uygulamalarda güvenilir sonuçlar sunmaktadır. Destek Vektör Makineleri (SVM), Karar Ağaçları ve Naive Bayes (GNB) gibi algoritmalar, farklı veri setlerinde yüksek performans sergilemekte ve yangın yönetimi stratejilerine önemli katkılar sağlamaktadır. Gelecekte, yangın çıkış nedenleri, yangın noktalarının yerleşim yerlerine uzaklığı ve hava durumu gibi faktörlerin veri setlerine eklenmesiyle modellerin doğruluk oranlarının daha da artırılabilceği öngörülmektedir. Bu tür çalışmalar, yangın risklerinin daha kapsamlı bir şekilde analiz edilmesine olanak sağlayarak, proaktif önlemlerin alınmasını mümkün kılmaktadır (Mishra, ve diğerleri, 2024). Makine öğrenmesi modellerinin seçimi, veri analizi ve tahmin süreçlerinde başarıya ulaşmanın en kritik adımlarından biridir (Özdemir & Örsülü, 2019). Model seçiminde, veri setinin özellikleri, problem türü ve kullanılacak değerlendirme metrikleri dikkate alınır (Eryılmaz, Şahin, & Kılıç, 2020). Örneğin, sınıflandırma problemleri için Karar Ağacı, Destek Vektör Makineleri (SVM) veya Naive Bayes algoritmaları tercih edilirken, regresyon problemlerinde Doğrusal Regresyon veya Çok Katmanlı Sinir Ağları gibi yöntemler daha uygun olabilir. Ayrıca, model seçiminde hiperparametre ayarlarının doğru şekilde yapılandırılması ve modelin genelleme performansının değerlendirilmesi büyük önem taşır (Şahin & Kılıç, 2019). Bu süreçte, çapraz doğrulama gibi yöntemler kullanılarak aşırı öğrenme (overfitting) gibi sorunların önüne geçilebilir.

3.2.1. Denetimli öğrenme yöntemleri

Denetimli öğrenme yöntemleri, etiketli veri setleri kullanarak girdi ve çıktı arasındaki ilişkileri öğrenen algoritmalarlardır. Bu çalışmada, 2015-2023 yıllarına ait 29.136 yangın verisi kullanılarak yangın risk analizinde ve tahminlemelerinde üç temel denetimli öğrenme yöntemi uygulanmıştır: Karar Ağaçları (Fine Tree), Linear Support Vector

Machine (SVM) ve Gaussian Naive Bayes (GNB). Bu yöntemlerin seçiminde temel amaç, verilerin doğasına uygun algoritmalar kullanarak daha doğru ve güvenilir sonuçlar elde etmektir. Yöntemlerin detaylı açıklamaları şu şekildedir:

Karar ağaçları, verilerin dallara ayrılarak sınıflandırıldığı, anlaşılması kolay ve yorumlanabilir bir yöntemdir. Bu yöntem, özellikle yangın risk faktörlerinin belirlenmesi ve sınıflandırma problemlerinde etkili bir araç olarak kullanılmaktadır. Literatürde, orman yangınlarının tahmin edilmesi için meteorolojik ve uzaktan algılama verilerinin kullanıldığı çalışmalarda Karar Ağaçları başarılı sonuçlar vermiştir (Sun & Li, 2008) . Örneğin, bir çalışmada Normalize Edilmiş Fark Bitki Örtüsü İndeksi (NVDI), Arazi Yüzeyi Sıcaklığı (LST) ve Termal Anomali (TA) gibi uydudan elde edilen veriler kullanılarak orman yangınları tahmin edilmiştir. Veri setinin %70'i modelin eğitimi, %30'u ise test aşaması için ayrılmış ve 10 kat tekrarlı test yöntemi uygulanmıştır. Bu süreçte, Karar Ağaçları modeli, %98,62 duyarlılık oranı ve %93,11 doğruluk oranı ile orman yangınlarının tahmininde yüksek performans göstermiştir (Beşli & Tenekeci, 2020). Bu tür çalışmalar, Karar Ağaçları yönteminin yangın tahmini ve risk analizi açısından güçlü bir çözüm sunduğunu ortaya koymaktadır (Wu & Banzhaf, 2010). Modelin yorumlanabilir olması, yangına neden olan faktörlerin anlaşılmasına da katkı sağlamaktadır. Bu çalışmada, yangın risk faktörlerinin belirlenmesi ve sınıflandırılması amacıyla Karar Ağaçları yöntemi kullanılmıştır.

Destek Vektör Makineleri (SVM), sınıflandırma problemlerinde verilerin ayrılmasını sağlayan güçlü ve etkili bir algoritmadır. SVM, özellikle karmaşık ve yüksek boyutlu veri setlerinde yüksek doğruluk oranı sunmasıyla öne çıkmaktadır (Ayhan & Erdoğan, 2014). Literatürde SVM'nin başarısını ortaya koyan çeşitli çalışmalar bulunmaktadır. Örneğin, Muhammed Fatih Pekşen'in çalışmasında, elektrik pano yangınlarının tespiti için farklı makine öğrenmesi modelleri değerlendirilmiş ve Gaussian SVM modeli, %99,3 doğruluk oranı ile en başarılı modellerden biri olmuştur. Çalışmada 5-katlı çapraz doğrulama yöntemi kullanılarak modellerin performansı analiz edilmiş ve SVM'nin yüksek boyutlu veri setlerinde güçlü bir sınıflandırma performansı sergilediği belirtilmiştir. Bu başarı, yangın risklerinin erken tespiti ve önleyici bakım sistemlerinin geliştirilmesi açısından SVM algoritmasının önemini vurgulamaktadır (Pekşen M. , 2024). SVM, veri setlerinde başarılı sonuçlar vermesi nedeniyle yangın risk analizlerinde yaygın olarak tercih edilmektedir. Bu

çalışmada, yangın risk faktörlerinin sınıflandırılması amacıyla Linear Support Vector Machine (SVM) kullanılmıştır. SVM algoritmasının tercih edilmesindeki temel neden, sınıflandırma problemlerinde sunduğu yüksek doğruluk oranı ve genelleme kabiliyetidir. Modelin performans değerlendirmesi, K-Katlı Çapraz Doğrulama yöntemiyle gerçekleştirilmiş ve modelin doğruluk oranları ilerleyen bölümlerde detaylandırılacaktır.

Gaussian Naive Bayes (GNB), olasılık temelli sınıflandırma yöntemlerinden biri olup, özellikle büyük ve karmaşık veri setlerinde hızlı ve etkili sonuçlar vermektedir. GNB, sınıflar arasındaki olasılık dağılımlarını Gaussian (Normal) dağılım varsayımına dayanarak hesaplar ve bu özelliği sayesinde sınıflandırma problemlerinde yaygın olarak tercih edilmektedir (Şahinaslan, Dalyan, & Şahinaslan, 2022). Literatürde yapılan çalışmalar, GNB'nin performansını farklı alanlarda ortaya koymuştur. Örneğin, yangın enkazındaki yanıcı sıvı (IL) kalıntılarının tespiti üzerine yapılan bir çalışmada, 360 yangın enkazı numunesi kullanılarak makine öğrenmesi algoritmaları geliştirilmiştir. Bu çalışmada GNB, test örneklerinin sınıflandırılmasında “benzinli” ve “benzinsiz” olarak başarılı sonuçlar vermiş; ancak bazı durumlarda diğer IL türlerini sınıflandırmada performans sınırlılıkları göstermiştir. Buna rağmen, yöntemin hızlı bir tarama tekniği olarak hizmet etme potansiyeline sahip olduğu vurgulanmıştır. Bu çalışmada Gaussian Naive Bayes (GNB) yöntemi, yangın çıkış nedenlerinin sınıflandırılması için tercih edilmiştir. GNB'nin hızlı hesaplama kabiliyeti, küçük veri setlerinde etkin performansı ve basit uygulanabilirliği, yöntemin seçilmesinde temel faktörlerdir. Ayrıca, sınıflandırma modellerinin eğitiminde Ki Kare Testi ile öznitelik seçimi yapılmış ve K-Katlı Çapraz Doğrulama yöntemi kullanılarak modelin genelleme yeteneği artırılmıştır (Bogdal, Schellenberg, Höpli, Bovens, & Lory, 2022). Modelin performansına yönelik değerlendirme sonuçları ilerleyen bölümlerde detaylandırılacaktır. Gaussian Naive Bayes algoritması, yangın verilerinin sınıflandırılmasında hızlı ve güvenilir bir çözüm sunarak, tahminleme süreçlerine katkı sağlamaktadır.

Veri setinde öznitelik seçimi için Ki Kare Testi uygulanmış ve verilerin sınıflandırma performansını artırmak amacıyla gereksiz öznitelikler elenmiştir. Bu süreç, modelin daha az öznitelik kullanarak daha hızlı çalışmasını sağlamış ve aşırı uyum riskini azaltmıştır. Ayrıca, modellerin doğruluğunu artırmak ve genelleme hatasını en aza indirmek amacıyla K-Katlı Çapraz Doğrulama yöntemi kullanılmıştır. Bu çalışmada

performans değerlendirme metrikleri olarak doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1 skoru gibi kriterler dikkate alınmıştır. Doğruluk, modelin doğru sınıflandırdığı örneklerin oranını belirtirken; kesinlik, pozitif tahminlerin ne kadarının doğru olduğunu ifade eder. Duyarlılık, modelin pozitif sınıfları doğru tespit etme oranını ölçerken, F1 skoru ise kesinlik ve duyarlılık değerlerinin harmonik ortalaması olarak genel bir performans göstergesi sunmaktadır (Fidanboy, Adar, & Okyay, 2022). Denetimli öğrenme yöntemlerinin kullanımı, yangın risk faktörlerini belirlemek ve riskli bölgeleri sınıflandırmak açısından önemli bir katkı sağlamaktadır. Bu yöntemler sayesinde yangın olaylarının analiz edilmesi ve tahmin edilmesiyle, önleyici stratejilerin geliştirilmesi ve daha etkin bir yangın yönetim planının oluşturulması hedeflenmektedir.

3.2.1.1. Karar ağacı (fine tree)

Karar ağaçları, eğitim ve test süreçlerinin hızlı olması, güvenliklerinin daha iyi olması ve elde edilen sonuçların kolayca yorumlanabilmesi özellikleriyle sınıflandırma problemlerinde sıkça tercih edilen bir tekniktir (Wu & Banzhaf, 2010). Bu yöntem, basit karar verme adımları uygulanarak büyük veri kümelerini daha küçük ve anlamlı alt gruplara ayırma yeteneği sunar (Sun & Li, 2008). Karar ağaçları, sınıflandırma işlemlerini iki temel aşamada gerçekleştirir: ilk olarak, ağaç yapısı oluşturulur; ardından bu yapıdan sınıflandırma kuralları çıkarılarak tahminler yapılır (Kaya & Yıldız, 2014). Karar ağaçları, her bir karar noktasında bir özelliği veya değişkeni dikkate alarak dallara ayrılır ve bu süreç, her bir yaprak düğümünde nihai kararların alınmasıyla sonuçlanır (Kavzoğlu & Çölkesen, 2010). Bu yapı, özellikle büyük veri kümelerinde etkili ve hızlı sınıflandırmalar yapılmasını mümkün kılar (Beşli & Tenekeci, 2020). Ayrıca, sonuçların görselleştirilebilir bir formatta sunulması, modelin şeffaflığına ve kolay anlaşılabilir olmasına katkıda bulunur (Quinlan, 2014). Karar ağacı algoritmaları, verilerin yapısına ve hedef değişkenin türüne bağlı olarak farklı modeller kullanabilir, ancak temel mantık her zaman aynıdır: verileri anlamlı şekilde gruplamak ve bu grupları kullanarak tahminler yapmak (Padilla & Vega-Garcia, 2011).

Örnek verecek olursak; problemimiz bir bölgenin doğal afet riskine maruz kalıp kalmayacağını tahmin etmek olsun. Girdi veriler: yağış miktarı (mm), zemin türü (kayalık/toprak), eğim derecesi (%). Çıktısı da afet riski (yüksek/düşük). İlk karar (Kök Düğüm): İlk olarak bölgede ki yağış miktarına bakılır. Eğer yağış miktarı 100

mm' den fazlaysa karar zemin türüne yönelir. Yağış miktarı 100 mm' den az ise bu sefer eğim derecesi dikkate alınır. Şimdi yağışın 100 mm'den fazla olduğu durumda zemin toprak ise afet riski yüksek bölge olarak değerlendirilir. Zemin türü kayalık ise afet riski düşük bölge olarak değerlendirilebilir. Yağışın 100 mm'den az ve eğim derecesi %30'dan fazla ise afet riski yüksek bölge, eğim %30 dan azsa afet riski düşük bölge olarak kabul edilmektedir. Bu karar ağacı örneğinde yağış miktarı ve zeminin özelliklerine göre afet riski tahmini yapar.

Karar ağaçları, özellikle birden fazla değişkenin etkileşimini incelemek ve bu değişkenlere dayalı tahminler yapmak için oldukça etkili bir araçtır (Wu & Banzhaf, 2010). Bunun yanı sıra, modelin şeffaf ve açıklanabilir olması, kullanıcıların model çıktısını anlamasını ve bu çıktıları güven duymasını sağlar (Kaya & Yıldız, 2014). Tez çalışması kapsamında, yangın verilerinde karar ağacı algoritması, yangın risk faktörlerinin analiz edilmesi ve tahmin edilmesinde kullanılmıştır. Bu süreçte, karar ağacının sunduğu kolaylıklar sayesinde, yangın risk tahminlerinde yüksek doğruluk oranlarına ulaşılmıştır.

3.2.1.2. Destek Vektör Makineleri (Linear Support Vector Machine (SVM))

Destek Vektör Makineleri (SVM), makine öğreniminde sınıflandırma ve regresyon analizlerinde kullanılan güçlü ve etkili bir algoritmadır (Kavzoğlu, Çölkesen, & Şahin, 2012). SVM, bir modelin eğitim ve test verilerindeki hataları minimize etmek için matematiksel olarak sağlam bir yöntem olan dışbükey optimizasyonu kullanır (Suthaharan, 2016). Bu yaklaşım, yalnızca mevcut verilere değil, aynı zamanda gelecekteki verilere de iyi uyum sağlayabilen olarak genelleştirilebilir (Ghosh, Dasgupta, & Swetapadma, 2019) ve güvenilir sonuçlar üretmesini sağlar (Ayhan & Erdoğmuş, 2014). Bu özellik, SVM'yi geniş bir uygulama yelpazesinde popüler hale getirmiştir (Burges, 1998).

SVM, iki sınıfa ait verileri doğrusal olarak ayırabilmek için bu sınıflar arasındaki en iyi sınırı belirler (Kecman, 2005). Bu sınır, veriler arasındaki maksimum mesafeyi (marjı) koruyacak şekilde optimize edilir ve bu sayede sınıflandırma modeli (Hearst, Dumais, Osuna, Platt, & Schölkopf, 1998), hem doğruluk hem de genelleme performansı açısından etkili bir hale gelir (Forero, Cano, & Giannakis, 2010). Bu yöntem, istatistiksel öğrenme teorisine dayanır ve sınıflandırma problemlerinde sıklıkla kullanılır (Vapnik, 1996). Verilerin doğrusal olarak ayrılabilir olmadığı durumlarda ise, SVM algoritması çekirdek (kernel) fonksiyonları kullanarak verileri

daha yüksek boyutlu bir uzaya dönüştürerek ayrılabilir hale getirir (Ayhan & Erdoğan, 2014).

SVM'nin başarısı, yalnızca algoritmanın matematiksel gücüne değil, aynı zamanda veri işleme aşamalarındaki hassasiyete de bağlıdır. 2019 yılında Eyyup Ensar Başakın, Ömer Ekmekcioğlu ve Mehmet Özger tarafından yapılan “Makine Öğrenmesi Yöntemleri ile Kuraklık Analizi” başlıklı çalışmada, Kayseri iline ait 116 yıllık Palmer Kuraklık Şiddet İndeksi (PDSI) verileri, makine öğrenmesi yöntemleriyle analiz edilmiştir. Çalışmada bir, üç ve altı ay sonrası kuraklık tahminleri gerçekleştirilmiştir. Kullanılan yöntemler arasında Destek Vektör Makinesi (SVM) ve K-en Yakın Komşuluk (KNN) algoritmaları yer almıştır. Bunun yanı sıra, dalgacık dönüşümü ile hibrit modeller geliştirilmiştir. Kuraklık analizi sonuçları, SVM'nin KNN algoritmasına kıyasla daha yüksek bir tahmin başarı oranına sahip olduğunu göstermiştir. Karmaşık modeller (SVM gibi) genellikle daha iyi sonuçlar üretirken, çalışmada ön işlem aşamalarının basit modellerin performansını da artırabileceği ortaya konulmuştur. Bu durum, veri işleme süreçlerinin makine öğrenmesi uygulamalarındaki kritik önemini bir kez daha vurgulamaktadır. Özellikle kuraklık gibi belirsizlikler içeren doğal olaylarda uzun vadeli tahminler, su kaynakları yönetimi ve kriz planlamasında hayati bir rol oynamaktadır. Uzun dönemli tahminlerin doğruluğu, karar vericilerin alacağı önlemlerin etkinliğini artırarak kuraklık etkilerinin en aza indirilmesine katkıda bulunmaktadır (Başakın, Ekmekcioğlu, & Özger, 2019).

Tez çalışması kapsamında, Linear SVM algoritması, yangın verilerinin analizi ve yangın risk faktörlerinin sınıflandırılması için uygulanmıştır. SVM'nin genelleştirilebilir yapısı, elde edilen modelin güvenilir tahminler üretmesini sağlamış ve yangın riski değerlendirmelerinde yüksek doğruluk oranlarına ulaşılmıştır. Bu bağlamda, SVM algoritmasının hem teorik gücü hem de pratik avantajları, yangın risk yönetimi ve önleme stratejilerinin geliştirilmesinde önemli bir rol oynamıştır (Aydın, 2018).

3.2.1.3. Gaussian Naive Bayes (GNB)

Gaussian Naive Bayes sınıflandırıcısı, veri madenciliği, makine öğrenmesi ve duygu analizi gibi birçok alanda yaygın olarak kullanılan bir denetimli öğrenme algoritmasıdır (Şahinaslan, Dalyan, & Şahinaslan, 2022). Belirli bir özellik kümesinin belirli bir sınıfa ait olma ihtimalini tahmin etmek için kullanılmaktadır (Ontivero-Ortega, Lage-Castellanos, Valente, Goebel, & Valdes-Sosa, 2017). Sınıflandırmada

Naive Bayes teorimini esas alır (Alqaraleh, 2021). Olasılık hesaplamalarında kullanılarak en yüksek olasılığa sahip olan kararı seçmeyi hedeflemektedir (Tonkal & Polat, 2021).

Naive Bayes formülü şu şekilde (Arpacı & Kalıpsız, 2018);

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

GNB, verilerin normal dağılıma uygun olduğunu varsayar ve bu varsayım altında sürekli özelliklerin olasılıklarını hesaplar (Pérez, Larrañaga, & Inza, 2006). Formülden yola çıkarak bir örnek verecek olursak; yangın ihbarının belirli bir risk seviyesine (yüksek, orta, düşük) ait olma ihtimalini şu şekilde açıklayabiliriz:

$P(C_{yüksek}|X)$: Verilen X gözlemlerine (yoğun duman, yüksek sıcaklık gibi..) dayanarak yangının yüksek riskli olma olasılığıdır.

$P(C_{yüksek})$: Yangının yüksek riskli olma önsel olasılığıdır, yani geçmiş yangın ihbarlar verilerinde yüksek riskli olayların görülme oranıdır.

$P(X)$: Verilen X gözlemlerinin ortaya çıkma genel olasılığıdır. Örneğin; yoğun duman ve sıcaklık gibi belirtilerin herhangi bir yangında görülme sıklığı.

$P(X|C_{yüksek})$: Yüksek riskli yangınların belirtilen X özelliklerini (yoğun duman ve yüksek sıcaklık gibi) gösterme olasılığıdır.

Bu tür bir model yangın ihbarı alındığında ihbarın hangi risk sınıfına ait olma olasılığını daha önceden hesaplanan olasılık değerleri dikkate alınarak değerlendirilir. Bu yöntemle risk sınıfını doğru şekilde hesaplayarak itfaiyelerinde müdahale önceliklerini belirlemesinde yardımcı olmaktadır (Frank & Bouckaert, 2006).

Tez çalışmasında, Gaussian Naive Bayes algoritması, kullanılan yangın verilerinin sınıflandırılması ve risk analizi için uygulanmıştır. Bu yöntem, yangın ihbarlarının doğru bir şekilde analiz edilmesini sağlamış ve yangın yönetimi stratejilerinin geliştirilmesine önemli katkılar sunmuştur. Elde edilen sonuçlar, GNB algoritmasının hız ve doğruluk açısından sağladığı avantajları ortaya koymuştur.

3.2.2. Denetimsiz öğrenme yöntemleri

Denetimsiz öğrenme, makine öğrenmesinde etiketlenmemiş veriler üzerinde çalışan bir yöntemdir. Bu tür öğrenme, veriler arasında daha önceden belirlenmemiş örüntüleri

keşfetmek, veriyi gruplamak ve boyut azaltma gibi amaçlar için kullanılır. Denetimsiz öğrenmede model, sadece verilerin özelliklerini inceleyerek anlamlı sonuçlara ulaşmaya çalışır (Dike, Zhou, Deveerasetty, & Wu, 2018). Denetimsiz öğrenme yöntemleri, özellikle büyük ve karmaşık veri setlerinde veri sınıflandırma öncesinde kullanılan bir ön analiz tekniği olarak önemli rol oynamaktadır. Yangın analizi ve risk haritalarının oluşturulması gibi çalışmalar için veri setinin segmentlere ayrılması veya gruplar oluşturulması, denetimsiz öğrenme yöntemlerinin sunduğu avantajlardan biridir (Bilgin, 2017).

3.2.2.1. Kümeleme (clustering) yöntemleri

K-Means Kümeleme; K-Means algoritması, veriyi belirli sayıda küme (grup) içine ayırarak her kümenin merkezine en yakın verileri bir araya getirir. Örneğin, yangın verilerinde risk bölgelerinin kümelenmesi amacıyla kullanılabilir. Literatürde, mekânsal yangın verilerini analiz etmek için K-Means kullanılarak yüksek, orta ve düşük riskli bölgelerin gruplandırıldığı çalışmalar mevcuttur. Literatürde, “K-Means Algoritması ile Otomatik Kümeleme” başlıklı çalışmada, bu yöntemin otomatik kümeleme özelliğiyle farklı veri türlerinde başarılı olduğu gösterilmiştir (Çolak, Durdağ, & Erdoğmuş, 2016).

Hiyerarşik Kümeleme; Hiyerarşik kümeleme, verileri ağaç yapısında (dendrogram) organize ederek, verilerin birbirleriyle olan ilişkilerini hiyerarşik bir düzende gösterir. Bu yöntem, veri setindeki tüm verilerin birbirleriyle olan ilişkilerini gözlemleme imkanı sunar ve verilerdeki benzerlikleri keşfetmeye yardımcı olabilecek daha küçük kümeler oluşturmayı sağlar (Birant, 2017). Ersin Aytaç'ın “Havzaların Benzerliklerini Tanımlamada Alternatif Bir Yaklaşım: Hiyerarşik Kümeleme Yöntemi Uygulaması” başlıklı çalışmasında, hiyerarşik kümeleme yöntemi kullanılarak Türkiye'nin kuzeyindeki bazı havzaların hidrolojik müdahale birimlerine göre benzerlikleri incelenmiştir. Bu çalışma, hiyerarşik kümeleme yönteminin, veri setindeki tüm havzaların birbirleriyle olan ilişkilerini ortaya koymada ve havzaların benzerliklerini keşfetmede etkili bir yöntem olduğunu göstermektedir (Aytaç, 2021).

3.3. Model Doğrulama Yöntemleri

Makine öğrenmesi modellerinin güvenilirliğini ve genelleme kapasitesini değerlendirmek için doğrulama yöntemleri kritik bir öneme sahiptir (Dolu & Şen, 2024). Doğrulama süreci, modelin eğitim sırasında öğrendiği desenlerin ve ilişkilerin,

daha önce görmediği veri setleri üzerinde nasıl performans gösterdiğini ölçmeyi amaçlar (Bilgin, 2017). Genellikle eğitim ve test veri setlerinin farklı şekillerde bölünmesi esasına dayanan doğrulama yöntemleri, modelin aşırı öğrenme (overfitting) veya yetersiz öğrenme (underfitting) gibi problemler yaşayıp yaşamadığını anlamaya yardımcı olur (Dolu & Şen, 2024). Makine öğrenmesi modellerinde yaygın olarak kullanılan başlıca doğrulama yöntemleri Sınama Seti Yöntemi (Holdout Method), Bootstrap Doğrulama ve K-Katlı Çapraz Doğrulama (K-Fold Cross Validation) olarak üç ana başlık altında ele alınmaktadır.

3.3.1. Sınama seti yöntemi (holdout method)

Sınama Seti Yöntemi (Holdout Method), veri setini eğitim ve test seti olarak ikiye bölen basit ve yaygın bir model değerlendirme yöntemidir (Gümüştas & Pehlivanlı, 2021). Bu yöntemde, veri setinin genellikle %60-80'i eğitim seti, %20-40'ı ise test seti olarak ayrılır (MOE). Eğitim seti, modelin öğrenmesi için kullanılır ve bu süreçte model, özellikler (features) ile hedef değişkenler (labels) arasındaki ilişkileri öğrenir (Zhang, ve diğerleri, 2017). Eğitim işlemi tamamlandıktan sonra, modelin başarısı test seti üzerinde değerlendirilir. Test seti, modelin daha önce görmediği verilerden oluşur ve modelin gerçek hayattaki performansını anlamak için kullanılır (Gümüştas & Pehlivanlı, 2021).

Bu yöntemin avantajlarından biri, uygulama kolaylığı ve hesaplama açısından hızlı bir değerlendirme sağlamasıdır (Yadav & Shukla, 2016). Modelin yalnızca bir kez eğitilip test edilmesi, basit ve küçük veri setleri üzerinde etkili bir yöntem sunar (Dwork, ve diğerleri, 2015). Ancak, yöntemin doğruluğu veri setinin büyüklüğüne ve rastgele bölünme sürecine bağlıdır (Gümüştas & Pehlivanlı, 2021). Özellikle küçük veri setlerinde, eğitim ve test setlerinin dengeli bir şekilde bölünmemesi durumunda, modelin genelleme yeteneği yanıltıcı olabilir (Yadav & Shukla, 2016). Ayrıca, tek bir bölünme ile değerlendirme yapıldığı için sonuçlar, veri setinin rastgele seçimine bağlı olarak değişiklik gösterebilir (Güney & Öztoprak, 2018).

3.3.2. Bootstrap doğrulama

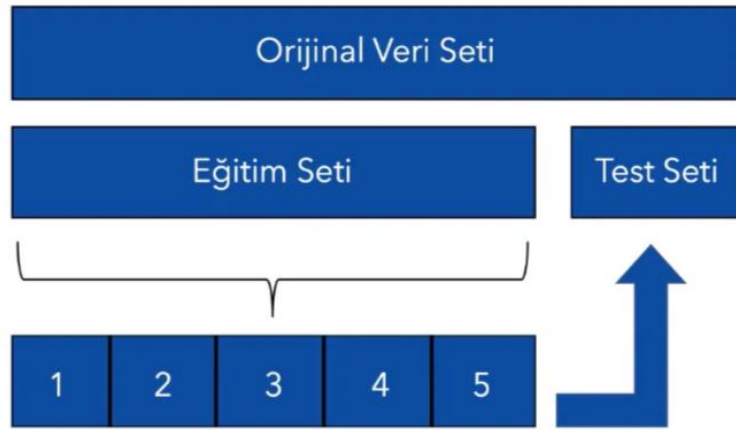
Bootstrap Doğrulama, veri setinden rastgele örnekler alınarak eğitim ve test veri setlerinin oluşturulmasını sağlayan bir yöntemdir (Özçelik, 2023). Bu yöntemde, veri setinden tekrar seçme (resampling) yöntemiyle eğitim verileri oluşturulur ve geri kalan veriler test seti olarak kullanılır (Biçer, Mirtagioğlu, Demir, Keskin, & Demir, 2024).

Bootstrap yöntemi, özellikle sınırlı sayıda veri ile çalışılan durumlarda model performansını değerlendirmek için etkili bir araçtır (Özçelik, 2023).

Bu yaklaşımın temel avantajlarından biri, orijinal veri setinden farklı ve çeşitli alt kümelerin elde edilmesini sağlamasıdır (Duman, 2006). Bu alt kümeler üzerinde model eğitilerek ve test edilerek, modelin genelleme yeteneği daha doğru bir şekilde ölçülebilir (Topuz, 2022). Ayrıca, Bootstrap yöntemi, veri setinin boyutunun küçük olduğu durumlarda, eğitim ve test setlerini dengeli bir şekilde oluşturmak için kullanışlıdır (Efron, 1992).

3.3.3. K-Katlı çapraz doğrulama (k-fold cross validation)

Makine öğrenimi modellerinin performansını değerlendirmek ve doğrulama süreçlerini daha güvenilir hale getirmek için kullanılan K-Katlı Çapraz Doğrulama, özellikle veri setlerinin sınırlı olduğu durumlarda etkili bir yöntem olarak öne çıkar (Polat & Güneş, 2006). Bu yöntem, modelin yeni verilere ne kadar iyi genelleme yapabildiğini anlamayı amaçlar ve veri setini farklı şekillerde kullanarak daha dengeli bir değerlendirme sağlar (Özkan & Ulaş, 2023). Bunu yaparken veri setinizi farklı şekillerde kullanır (Goh & Ubeynarayana, 2017). Eğitim ve test seti arasındaki dengesizliklerin model performansını olumsuz etkileyebileceği durumlarda, Holdout yöntemine alternatif olarak sıkça tercih edilmektedir (Temurtas, 2009). K-Katlı Çapraz Doğrulama, veri setini belirli bir sayıda alt gruba (kat) böler ve her bir alt grubun sırasıyla test seti olarak kullanıldığı bir döngü ile gerçekleştirilir (Bishop, 2006). Bu süreç, modelin tüm veri seti üzerinde dengeli bir şekilde değerlendirilmesini sağlar (Polat & Güneş, 2006). Yöntem, farklı veri yapıları ve problem türleri için uygulanabilir ve genelleme hatalarını en aza indirerek modelin performansını daha doğru bir şekilde ölçmeye olanak tanır (Özkan & Ulaş, 2023). Çalışmada bu yöntem iki farklı metotta uygulanmıştır (Çetin, Temurtaş, & Gülgönül, 2015):



Şekil 3.3. Çapraz doğrulama modeli (Bulut, 2023).

Birinci metot:

Orijinal veri seti 5 eşit parçaya bölünür. Bu yöntemle, her bir parça bir kez test verisi olarak kullanılırken kalan parçalar modelin eğitimi için kullanılır. Örneğin:

- İlk adımda, 1, 2, 3 ve 4 numaralı parçalar modelin eğitimi için kullanılır, 5 numaralı parça test edilir.
- Sonraki adımda, 1, 2, 3 ve 5 numaralı parçalar model eğitimi için kullanılır, 4 numaralı parça test edilir.
- Devamında, 1, 2, 4 ve 5 numaralı parçalar model kurmak için kullanılır, 3 numaralı parça test edilir.
- Daha sonra, 1, 3, 4 ve 5 numaralı parçalar model için kullanılırken 2 numaralı parça test edilir.
- Son olarak, 2, 3, 4 ve 5 numaralı parçalar model için kullanılır ve 1 numaralı parça test edilir.

Her bir adımda modelin başarı ve hata oranları kaydedilir. Bu süreç tamamlandığında, her adımda elde edilen sonuçların ortalaması alınarak çapraz doğrulama hatası (Cross Validation Error) hesaplanır (Bates, Hastie, & Tibshirani, 2023). Bu yöntem, modelin tüm veri seti üzerinde genelleme performansını ölçmek için kapsamlı bir değerlendirme sunar (Moore & Lee, 1994).

İkinci metot:

Zengin ve bol gözlem değerine sahip veri setlerinde, model performansını değerlendirmek için şu yöntem yaygın olarak kullanılır: Öncelikle, veri seti ikiye

ayrılır: bir kısmı Eğitim Seti, diğer kısmı ise Test Seti olarak belirlenir. Eğitim seti üzerinde, daha önce açıklanan Çapraz Doğrulama (Cross Validation) yöntemi uygulanır. Bu süreçte, eğitim seti tekrar tekrar bölünerek model eğitimi ve doğrulaması gerçekleştirilir. Böylece, modelin farklı veri alt kümeleri üzerinde performansı test edilerek güvenilir bir sonuç elde edilir. Son olarak, modelin tamamen bağımsız ve daha önce hiç görmediği Test Seti kullanılarak bir değerlendirme yapılır. Bu işlem, modelin gerçek dünyadaki yeni verilere nasıl performans göstereceğini anlamak için yapılır. Test seti, model değerlendirmesinde nihai bir doğrulama adımıdır ve modelin genelleştirme yeteneğini ölçer (Birba, 2020).

Makine öğrenmesi modellerinin doğruluğunu artırmak ve genelleme hatalarını en aza indirmek için doğrulama yöntemleri büyük önem taşımaktadır. Sınama Seti Yöntemi, büyük veri setleri için hızlı bir değerlendirme sunarken, Bootstrap Doğrulama, veri setinin çeşitliliğini artırarak modelin farklı senaryolar üzerindeki başarısını ölçmektedir. K-Katlı Çapraz Doğrulama ise, en güvenilir yöntemlerden biri olarak kabul edilmekte ve modelin tüm veri seti üzerinde kapsamlı bir değerlendirme yapmasına olanak tanımaktadır.

3.3.3.1. Boyut azaltma (dimensionality reduction) yöntemleri

Principal Component Analysis (PCA); PCA, yüksek boyutlu veri setlerinde fazla ve gereksiz öznitelikleri eleyerek verinin boyutunu azaltır. Bu süreç, modelin hesaplama süresini azaltırken performansını korur. Özellikle çok sayıda değişken içeren yangın verilerinde, en önemli faktörleri öne çıkararak analizleri kolaylaştırabilir. Örneğin, “Parameter Design in Friction Welding of Al/SiC/Al₂O₃ Composite Using Grey Theory Based Principal Component Analysis (GT-PCA)” başlıklı çalışmada, PCA, kaynak parametrelerini optimize etmek ve karmaşık veri setlerini sadeleştirmek için kullanılmıştır. Bu yöntem, kaynak kalitesini artırmak amacıyla etkili bir çözüm sunmuştur (Adalarasan & Sundaram, 2015).

3.3.4. Denetimsiz öğrenme yöntemlerinin yangın analizindeki rolü

Denetimsiz öğrenme yöntemleri, yangın verilerindeki örüntüleri keşfetme ve verilerin gruplandırılması gibi analiz süreçlerinde önemli bir araçtır. Etiketlenmemiş veriler üzerinde çalışarak, bu yöntemler yangın verileri içerisindeki gizli yapıları ortaya

çıkarabilir ve bu sayede daha anlamlı analizler yapılmasını sağlayabilir (Bröker, Holt, Roads, Dayan, & Love, 2024).

Yangın risk bölgelerinin belirlenmesinde denetimsiz öğrenme, mekânsal yangın verilerinde risk bölgelerinin belirlenmesinde sıklıkla kullanılmaktadır. Örneğin, K-Means kümeleme algoritması, yangın olaylarını yüksek, orta ve düşük risk seviyelerine göre sınıflandırabilir. Bu tür analizler, yangına müdahale süreçlerini iyileştirmek ve kaynakların daha etkin bir şekilde dağıtılmasını sağlamak için kritik bilgiler sunmaktadır (Hu, Liu, Zhang, & Fang, 2023). Yangınların çıkış nedenlerini belirlemek, yangın risk yönetimi stratejilerinin geliştirilmesinde temel bir adımdır. Denetimsiz öğrenme yöntemleri, verilerdeki belirli faktörlerin yangınlara olan etkisini gruplandırarak analiz edilmesine olanak tanır. Bu süreçte hiyerarşik kümeleme, yangın çıkış nedenlerini alt gruplara ayırarak daha detaylı analizler yapılmasını sağlar (Karabulut, Baran, & Eraslan, 2024). Principal Component Analysis (PCA) gibi boyut azaltma yöntemleri, yangın verilerindeki gereksiz veya az etkili değişkenleri eleyerek en önemli faktörleri öne çıkarır. Bu, veri setinin analiz süreçlerini hızlandırır ve model performansını artırır. Özellikle büyük veri setlerinde bu yöntemlerin kullanımı, analizlerin verimliliğini artırmak açısından oldukça önem arz etmektedir (Yıldırım, Günay, & Yağanoğlu, 2023).



4. BÖLÜM YÖNTEM

Bu bölümde, yangın verileri kullanılarak yapılan analiz süreci ayrıntılı olarak açıklanmaktadır. Çalışmada kullanılan veri setinin kaynağı, veri ön işleme adımları, makine öğrenmesi modelleri, model eğitimi ve doğrulama yöntemleri ile model performans değerlendirme kriterleri sistematik bir şekilde ele alınmıştır.

4.1. Kullanılan Veri Seti

4.1.1. Veri kaynağı ve özellikleri

Bu çalışmada kullanılan veri seti, bir ilin İtfaiye Daire Başkanlığı tarafından sağlanmıştır. 2015-2023 yılları arasındaki yangın olaylarına ait 29135 veri satırını içermektedir.

Veri seti, yangınların türü, çıkış nedenleri, mekânsal ve zamansal bilgileri, hasar durumu ve itfaiye müdahale süreleri gibi birçok değişkeni içermektedir. Bu özellikler, yangın yönetimi ve tahmin modelleri için kritik öneme sahiptir.

Veri setinde yer alan değişkenler şunlardır:

- Yangın Türü: Orman yangını, bina yangını, araç yangını vb.
- Hasar Durumu: Maddi hasar, can kaybı ve yaralanma bilgileri.
- Yangının Çıkış Nedeni: Elektrik arızası, ihmal, kasıt vb.
- Coğrafi Konum: İlçe, mahalle bilgileri ve yangın alanının mesafesi.
- Zaman Bilgisi: Yıl, ay, saat, gece/gündüz farkı.
- Müdahale Süresi: İtfaiyenin yangına ulaşma süresi.

Bu veriler, yangınların nedenlerini, dağılımlarını ve etkilerini kapsamlı bir şekilde analiz etmek ve makine öğrenmesi modeli geliştirmek için kullanılmıştır.

4.1.2. Veri ön işleme

Veri Bütünlüğü ve Etiketleme Sistemi:

Veri setinde eksik veri bulunmamaktadır. Ancak, bazı kategorilerde “Bilinmiyor” veya “Tespit Edilememiştir” etiketleri bulunmaktadır. Bu etiketler, ilgili yangın olayında kesin bir nedenin belirlenemediğini veya itfaiye tarafından bilinçli olarak bu şekilde kaydedildiğini ifade etmektedir.

- Yinelenen (Duplicate) Kayıtların Tespiti ve Temizlenmesi
- Veri setinde aynı yangın olayına ait yinelenen kayıtlar tespit edilerek temizlenmiştir.
- Tarih, saat ve mahalle bilgisi açısından birebir aynı olan kayıtlar analizden çıkarılmıştır.

Veri Formatlarının Dönüştürülmesi:

- Saat, tarih ve mevsim verileri uygun formatlara dönüştürülerek zamansal analizler için düzenlenmiştir.
- Müdahale süresi verisi standart hale getirilerek analiz edilmiştir.

Değişken Dönüştürme ve Gruplama İşlemleri:

Bazı değişkenler daha anlamlı analizler yapabilmek için kategorilere ayrılmıştır:

- Yangın Mesafesi: 0-5 km, 5-10 km, 10-20 km, 20-50 km, 50+ km
- Saat Verisi: Sabah (06:00-12:00), Öğle (12:00-18:00), Akşam (18:00-24:00), Gece (00:00-06:00)
- Mevsim Verisi: İlkbahar, Yaz, Sonbahar, Kış

4.2. Analiz Yöntemleri

4.2.1. Yangın verilerinin istatistiksel analizi

Veri seti üzerinde tanımlayıcı istatistikler ve veri görselleştirme yöntemleri kullanılarak detaylı analizler gerçekleştirilmiştir:

- Yangın Türlerinin Dağılımı: En sık karşılaşılan yangın türleri belirlenmiştir.
- Yangın Çıkış Nedenleri: En yaygın yangın nedenleri analiz edilmiştir.
- Mevsimsel Yangın Dağılımı: Hangi mevsimde hangi yangın türü daha fazla?

- Günün Saatine Göre Yangın Dağılımı: Yangınların en yoğun olduğu saat aralıkları incelenmiştir.
- Yangın Mesafesi Dağılımı: Yangınların olay yerine uzaklığı analiz edilmiştir.

4.2.2. Makine öğrenmesi modelleri

Bu çalışmada, yangın türlerini ve çıkış nedenlerini tahmin edebilmek amacıyla denetimli öğrenme teknikleri uygulanmıştır. Makine öğrenmesi algoritmaları, geçmiş yangın verilerinden öğrenerek gelecekteki yangın risklerini tahmin etmeyi ve bu doğrultuda önleyici tedbirler geliştirmeyi amaçlamaktadır. Çalışmada, Karar Ağaçları (Fine Tree), Gaussian Naive Bayes (GNB) ve Destek Vektör Makineleri (SVM) modelleri uygulanarak veriler üzerinde farklı sınıflandırma teknikleri test edilmiştir.

Seçilen algoritmalar, yangın verilerinin doğası ve tahminleme ihtiyacına uygun olarak belirlenmiştir. Karar Ağaçları modeli, verileri hiyerarşik olarak bölerek sınıflandırmayı sağlayan açıklanabilir bir model olup, değişkenler arasındaki ilişkileri anlamak açısından güçlüdür. Gaussian Naive Bayes modeli, olasılıklı sınıflandırma yöntemi kullanarak verinin istatistiksel dağılımına dayalı tahminler yapar ve büyük veri kümeleri üzerinde hızlı çalışmasıyla öne çıkar. Destek Vektör Makineleri modeli ise, veri noktalarını doğrusal veya doğrusal olmayan şekilde ayırarak sınıflandırma yapabilen ve yüksek boyutlu verilerde etkili sonuçlar veren bir algoritmadır.

Bu modellerin uygulanmasıyla, yangın türlerinin ve çıkış nedenlerinin özellik mühendisliği, çapraz doğrulama ve model optimizasyonu gibi süreçlerden geçirilerek analiz edilmesi hedeflenmiştir.

4.2.2.1. Karar ağacı (fine tree) modeli

Karar Ağaçları modeli, verileri dallara bölerek bir dizi karar kuralı ile sınıflandırma yapan bir makine öğrenmesi algoritmasıdır. Eğitim ve test süreçlerinin hızlı olması, güvenliklerinin daha iyi olması ve elde edilen sonuçların kolayca yorumlanabilmesi özellikleriyle sınıflandırma problemlerinde sıkça tercih edilen bir tekniktir (Wu & Banzhaf, 2010). Model, özellik seçiminde belirli eşik değerlerine göre karar düğümleri oluşturarak en uygun bölünmeleri yapar.

Bu modelde, önemli değişkenlerin belirlenmesi ve sınıflandırma sürecinde etkili olması için Ki-Kare testi gibi istatistiksel yöntemler kullanılmıştır (Sun & Li, 2008). Her düğümde, en iyi ayrımı sağlayan özellik seçilmiş ve veri kümeleri alt gruplara

ayrılmıştır. Bu sayede, yangın türleri ve çıkış nedenlerinin hangi faktörlerle daha güçlü ilişkili olduğu ortaya çıkarılmıştır (Padilla & Vega-Garcia, 2011).

Karar ağaçları, özellikle birden fazla değişkenin etkileşimini incelemek ve bu değişkenlere dayalı tahminler yapmak için oldukça etkili bir araçtır (Wu & Banzhaf, 2010). Bunun yanı sıra, modelin şeffaf ve açıklanabilir olması, kullanıcıların model çıktısını anlamasını ve bu çıktılara güven duymasını sağlar (Kaya & Yıldız, 2014). Karar Ağaçları modelinin avantajları arasında, kolay anlaşılabilir olması ve sonuçların görselleştirilebilir olması yer almaktadır (Sun & Li, 2008). Ancak, modelin aşırı öğrenmeye yatkın olabilmesi nedeniyle, budama (pruning) teknikleri kullanılarak modelin genelleme yeteneği artırılmıştır (Akçetin & Çelik, 2014).

4.2.2.2. Gaussian Naive Bayes (GNB) modeli

Gaussian Naive Bayes modeli, olasılıksal temellere dayanan bir sınıflandırma algoritmasıdır (Şahinaslan, Dalyan, & Şahinaslan, 2022). Model, her özelliğin bağımsız olduğu varsayımıyla çalışarak, her sınıf için olasılık dağılımlarını hesaplar ve en yüksek olasılığa sahip sınıfı tahmin eder (Tonkal & Polat, 2021).

Gaussian Naive Bayes modeli, büyük veri kümeleri üzerinde hızlı çalışması ve düşük hesaplama maliyetiyle öne çıkmaktadır (Ontivero-Ortega, Lage-Castellanos, Valente, Goebel, & Valdes-Sosa, 2017). Ancak, özelliklerin birbirinden bağımsız olması gerektiği varsayımı her zaman sağlanamayabilir (Pérez, Larrañaga, & Inza, 2006). Bu nedenle, modelin doğruluğunu artırmak için bazı özellik mühendisliği süreçleri ve veri temizleme işlemleri gerçekleştirilmiştir.

Naive Bayes formülü şu şekilde (Arpacı & Kalıpsız, 2018);

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)}$$

GNB, verilerin normal dağılıma uygun olduğunu varsayar ve bu varsayım altında sürekli özelliklerin olasılıklarını hesaplar (Pérez, Larrañaga, & Inza, 2006). Formülden yola çıkarak bir örnek verecek olursak; yangın ihbarının belirli bir risk seviyesine (yüksek, orta, düşük) ait olma ihtimalini şu şekilde açıklayabiliriz:

$P(C_{yüksek}|X)$: Verilen X gözlemlerine (yoğun duman, yüksek sıcaklık gibi..) dayanarak yangının yüksek riskli olma olasılığıdır.

$P(C_{yüksek})$: Yangının yüksek riskli olma önsel olasılığıdır, yani geçmiş yangın ihbarlar verilerinde yüksek riskli olayların görülme oranıdır.

$P(X)$: Verilen X gözlemlerinin ortaya çıkma genel olasılığıdır. Örneğin; yoğun duman ve sıcaklık gibi belirtilerin herhangi bir yangında görülme sıklığı.

$P(X|C_{yüksek})$: Yüksek riskli yangınların belirtilen X özelliklerini (yoğun duman ve yüksek sıcaklık gibi) gösterme olasılığıdır.

Bu tür bir model yangın ihbarı alındığında ihbarın hangi risk sınıfına ait olma olasılığını daha önceden hesaplanan olasılık değerlerini dikkate alınarak değerlendirilir. Bu yöntemle risk sınıfını doğru şekilde hesaplayarak itfaiyelerinde müdahale önceliklerini belirlemede yardımcı olmaktadır (Frank & Bouckaert, 2006).

Bu çalışmada, yangın çıkış nedenlerini tahmin etmek amacıyla Naive Bayes modelinin Gaussian versiyonu kullanılmıştır. Modelin performansını artırmak için, veri setindeki değişkenlerin dağılımı incelenmiş ve modelin varsayımlarına uygun hale getirilmesi için normalleştirme teknikleri uygulanmıştır.

4.2.2.3. Destek Vektör Makineleri (SVM) modeli

Destek Vektör Makineleri (SVM) modeli, sınıflandırma problemlerinde kullanılan güçlü bir algoritmadır (Kavzoğlu, Çölkesen, & Şahin, 2012) ve veri noktalarını doğrusal veya doğrusal olmayan bir hiper düzlem ile ayırarak sınıflandırma yapar (Suthaharan, 2016). Modelin amacı, iki sınıfı en iyi şekilde ayıran optimal sınır çizgisini belirlemektir.

SVM'nin başarısı, yalnızca algoritmanın matematiksel gücüne değil, aynı zamanda veri işleme aşamalarındaki hassasiyete de bağlıdır. 2019 yılında Eyyup Ensar Başakın, Ömer Ekmekcioğlu ve Mehmet Özger tarafından yapılan “Makine Öğrenmesi Yöntemleri ile Kuraklık Analizi” başlıklı çalışmada, Kayseri iline ait 116 yıllık Palmer Kuraklık Şiddet İndeksi (PDSI) verileri makine öğrenmesi yöntemleriyle analiz edilmiştir. Çalışmada bir, üç ve altı ay sonrası kuraklık tahminleri gerçekleştirilmiştir. Kullanılan yöntemler arasında Destek Vektör Makinesi (SVM) ve K-en Yakın Komşuluk (KNN) algoritmaları yer almıştır. Bunun yanı sıra, dalgacık dönüşümü ile hibrit modeller geliştirilmiştir. Kuraklık analizi sonuçları, SVM'nin KNN algoritmasına kıyasla daha yüksek bir tahmin başarı oranına sahip olduğunu

göstermiştir. Karmaşık modeller (SVM gibi) genellikle daha iyi sonuçlar üretirken, çalışmada ön işlem aşamalarının basit modellerin performansını da artırabileceği ortaya konulmuştur. Bu durum, veri işleme süreçlerinin makine öğrenmesi uygulamalarındaki kritik önemini bir kez daha vurgulamaktadır. Özellikle kuraklık gibi belirsizlikler içeren doğal olaylarda uzun vadeli tahminler, su kaynakları yönetimi ve kriz planlamasında hayati bir rol oynamaktadır. Uzun dönemli tahminlerin doğruluğu, karar vericilerin alacağı önlemlerin etkinliğini artırarak kuraklık etkilerinin en aza indirilmesine katkıda bulunmaktadır (Başakın, Ekmekçioğlu, & Özger, 2019). SVM modeli, düşük boyutlu veri kümelerinde etkili olmasının yanı sıra, gürültülü verilere karşı dayanıklı olması ile de avantaj sağlamaktadır. Ancak, modelin karmaşıklığını artırmamak için hiperparametre optimizasyonu yapılmış ve gereksiz özellikler çıkarılmıştır (Aydın, 2018).

Bu çalışmada, yangın türlerini sınıflandırmak için doğrusal (linear) SVM modeli kullanılmıştır. Modelin doğruluğunu artırmak için, çekirdek fonksiyonları (kernel trick) uygulanarak veri uzayı genişletilmiş ve daha iyi ayrıştırmalar yapılmıştır (Hearst, Dumais, Osuna, Platt, & Schölkopf, 1998).

4.2.3. Makine öğrenmesi model eğitimi ve doğrulama süreci

Bu çalışmada geliştirilen makine öğrenmesi modellerinin yalnızca eğitim verisi üzerinde değil, daha önce karşılaşmadığı örnekler üzerinde de tutarlı sonuçlar verebilmesi hedeflenmiştir. Bu doğrultuda, modelleme süreci veri kümesinin anlamlı biçimde bölünmesiyle başlamış; ardından modelin genelleme yeteneğini daha sağlıklı değerlendirebilmek için çapraz doğrulama yöntemlerine başvurulmuştur. Böylece, modellerin güvenilirliğini artırmak ve aşırı öğrenmeyi önlemek amacıyla yapılandırılmış bir eğitim-doğrulama süreci izlenmiştir.

4.2.4. Eğitim ve test veri kümesinin oluşturulması

Modelin hem eğitim sürecinde anlamlı bir öğrenme gerçekleştirmesi hem de test aşamasında genelleme yeteneğinin tarafsız biçimde değerlendirilebilmesi için veri kümesi, %80 eğitim ve %20 test oranında iki alt kümeye ayrılmıştır. Eğitim kümesi, algoritmaların yapısal örüntüleri öğrenmesini sağlarken; test kümesi, modelin daha önce karşılaşmadığı örnekler üzerinde nasıl performans gösterdiğini gözlemlemek amacıyla kullanılmıştır. Bu veri ayrımı, MATLAB ortamında rastgele ancak sınıf dengesi gözetilerek uygulanmıştır. Bu yaklaşım, sınıf dengesizliğinden

kaynaklanabilecek öğrenme sapmalarını azaltmak ve her sınıfın test verisinde yeterince temsil edilmesini sağlamak açısından önemlidir. Bu yöntem, modelin doğruluk, hata oranı ve diğer istatistiksel ölçütler temelinde nesnel olarak değerlendirilebilmesine olanak tanımaktadır (Kelleher, Namee, & D'Arcy, 2020).

4.2.5. Çapraz doğrulama yönteminin uygulanması

Modelin doğruluk oranlarının daha güvenilir ve genellenebilir biçimde hesaplanabilmesi amacıyla, bu çalışmada 5 katlı çapraz doğrulama (5-fold cross-validation) yöntemi tercih edilmiştir. Bu yöntemde, eğitim veri kümesi beş eşit alt gruba ayrılmış; her bir yinelemede dört grup modelin eğitimi için, kalan bir grup ise doğrulama amacıyla kullanılmıştır. Bu işlem, her alt kümenin bir kez doğrulama verisi olarak görev aldığı toplam beş döngüyle tamamlanmıştır. Bu yapılandırma, modelin farklı veri alt kümeleri üzerindeki performansını test etmeye imkân tanımış; böylece öğrenilen örüntülerin tüm veri setine genellenip genellenemeyeceği daha güvenilir şekilde değerlendirilmiştir. Ayrıca, çapraz doğrulama yöntemi, aşırı öğrenme (overfitting) riskini azaltması ve veri dengesizliklerinden kaynaklanabilecek sapmaları dengelemesi açısından da avantaj sağlamaktadır (Raschka, Liu, & Mirjalili, 2022).

4.2.6. Makine öğrenmesi modellerinin performans değerlendirme ölçütleri

Makine öğrenmesi uygulamalarında model başarımının değerlendirilmesi, yalnızca genel doğruluk oranı gibi tek boyutlu metriklerle sınırlandırılmaz. Özellikle çok sınıflı ve dengesiz dağılıma sahip veri kümelerinde, modelin farklı sınıflar üzerindeki tahmin başarısını ortaya koymak için çoklu değerlendirme ölçütlerinin birlikte kullanılması gerekmektedir. Bu kapsamda, bu çalışmada doğruluk (accuracy), karışıklık matrisi (confusion matrix), ROC eğrisi ve AUC (Area Under Curve) değeri ile birlikte kesinlik (precision), duyarlılık (recall) ve F1 skoru gibi detaylı metrikler analiz sürecine dahil edilmiştir.

Doğruluk, tüm sınıflar üzerindeki genel başarı oranını verirken, sınıf dengesizliği içeren veri setlerinde modelin başarısını yanıltıcı şekilde yansıtabileceği bilinmektedir. Bu nedenle kesinlik, duyarlılık ve F1 skoru gibi metrikler, azınlık sınıfların model tarafından ne ölçüde doğru tanındığını ve modelin pozitif sınıfları tahmin ederken ne kadar isabetli olduğunu ortaya koyar. ROC eğrisi ve AUC değeri ise modelin pozitif ve negatif sınıflar arasında ayırt edicilik gücünü ölçmektedir.

Tüm bu metriklerin birlikte değerlendirilmesi, sınıflandırma modelinin hem genel başarısını hem de sınıf düzeyinde performans farklılıklarını anlamak için gereklidir. Furkan Bulut'un da belirttiği gibi, makine öğrenmesi projelerinde her bir metriğin kendi kullanım amacı ve bağlamı bulunmaktadır; dolayısıyla hangi metriklerin kullanılacağı veri setinin yapısına ve çözülmek istenen probleme göre dikkatle seçilmelidir (Bulut, 2023).



4.2.6.1. Doğruluk (accuracy) ölçütü

Doğruluk, bir sınıflandırma modelinin genel başarımını değerlendirmede sıkça başvurulan temel metriklerden biridir. Bu ölçüt, modelin doğru şekilde sınıflandırdığı örnek sayısının toplam örnek sayısına oranı ile hesaplanır. Ancak bu metriğin, özellikle sınıf dengesizliği içeren veri kümelerinde yanıltıcı sonuçlar verebileceği literatürde sıkça vurgulanmaktadır. Örneğin, çoğunluk sınıfının baskın olduğu durumlarda, model yalnızca bu sınıfı tahmin ederek yüksek doğruluk değeri elde edebilir; ancak azınlık sınıflardaki başarımı göz ardı edilebilir düzeyde kalabilir (Bulut, 2023). Bu nedenle, bu çalışmada doğruluk metriği tek başına değil, tamamlayıcı diğer ölçütlerle birlikte kullanılarak yorumlanmıştır.

4.2.6.2. Karışıklık matrisi (confusion matrix) ile değerlendirme

Karışıklık matrisi, modelin sınıflandırma başarımını yalnızca genel bir yüzde ile değil, her sınıf özelinde hangi tahminlerin doğru veya yanlış olduğunu ayrıntılı biçimde sunan bir araçtır. Bu matris, özellikle çok sınıflı sınıflandırma problemlerinde modelin hangi sınıflar arasında ayırım yapmakta başarılı olduğunu, hangi sınıflarda ise hata eğiliminde olduğunu sayısal olarak gösterir. Sınıf bazlı doğru pozitif, yanlış pozitif, doğru negatif ve yanlış negatif değerleri üzerinden farklı metriklerin de hesaplanabilmesine olanak tanır (Bulut, 2023). Bu çalışmada üç sınıfa ayrılmış etiketler için oluşturulan karışıklık matrisleri, modellerin tahmin performansının daha derinlemesine değerlendirilmesini sağlamıştır.

4.2.6.3. ROC eğrileri ve AUC değeri

ROC (Receiver Operating Characteristic) eğrisi, sınıflandırma modelinin farklı eşik değerlerinde doğru pozitif oranı ile yanlış pozitif oranı arasındaki değişimi grafiksel olarak ortaya koyar. ROC eğrisi altında kalan alan, yani AUC (Area Under the Curve), modelin genel ayırt edicilik yeteneğini sayısal olarak temsil eder. AUC değeri 1'e yaklaştıkça modelin sınıflar arasında daha başarılı bir ayırım gerçekleştirdiği kabul edilir. Özellikle çok sınıflı yapılarda, her bir sınıf için ayrı ROC eğrileri üretilerek analiz daha kapsamlı hale getirilebilir (Bulut, 2023). Bu çalışmada AUC değeri, doğruluk metriğinden bağımsız olarak modelin sınıf ayırımı yapabilme kapasitesinin ölçülmesinde kritik rol oynamıştır.

4.2.6.4. F1-Skoru, duyarlılık (recall) ve kesinlik (precision) ölçütleri

F1-Skoru, bir modelin kesinlik (precision) ve duyarlılık (recall) metrikleri arasında kurduğu dengeyi yansıtan harmonik ortalama değeridir. Duyarlılık, modelin gerçek pozitif örnekleri ne oranda doğru sınıflandırabildiğini gösterirken; kesinlik, modelin pozitif tahminlerinin ne kadarının gerçekten doğru olduğunu ifade eder. Özellikle sınıf dengesizliği söz konusu olduğunda, F1-Skoru modeli daha adil şekilde değerlendirmeye olanak tanır. Bu nedenle, doğruluk oranı yüksek olsa bile modelin azınlık sınıflardaki başarısızlığı bu metrikler sayesinde daha net ortaya çıkabilmektedir (Bulut, 2023). Bu bağlamda, çalışmada tüm modellerin bu ölçütlere göre performansı analiz edilmiş ve sınıf bazlı güçlü/zayıf yönleri ortaya konulmuştur.



5. BULGULAR

5.1. Yangın Verilerinin Tanımlayıcı İstatistikleri ve Veri Analizi

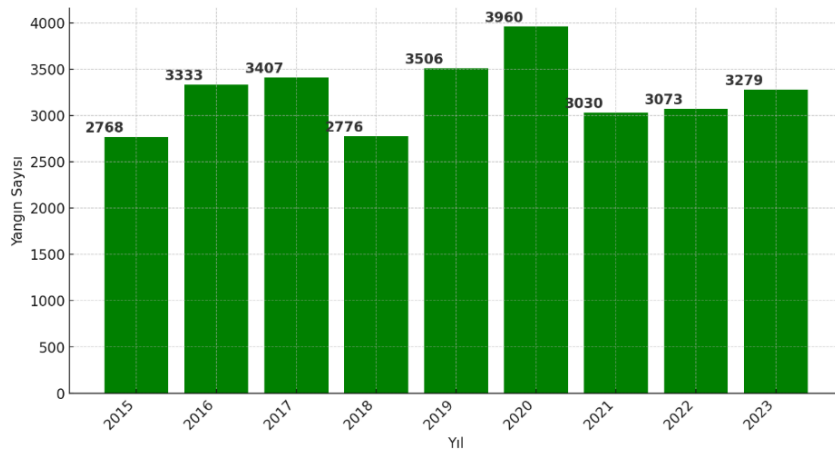
Çalışma kapsamında bir ilin 2015-2023 yılları arasında meydana gelen yangınların mekânsal, zamansal ve nedensel analizleri detaylandırılmıştır. Yangınların yıllara, ilçelere, mahallelere, günün saatlerine, mevsimlere ve çıkış nedenlerine göre dağılımı ele alınarak risk faktörleri belirlenmiştir. Ayrıca, yangınların neden olduğu can kayıpları, mesafeye göre dağılımı ve hasar durumları da incelenmiştir.

5.1.1. Yıllara göre yangın sayısı trendi

Yangınların yıllara göre değişimini gösteren Şekil 5.1 incelendiğinde, 2015-2023 yılları arasında dalgalı bir seyir izlendiği görülmektedir.

- 2020 yılı, 3960 yangın ile en fazla yangının yaşandığı yıl olmuştur.
- 2018 yılı ise 2776 yangın ile en düşük seviyededir.
- 2016, 2017 ve 2019 yıllarında yangın sayılarında artış gözlenirken, 2021 ve 2022 yıllarında göreceli bir düşüş yaşanmıştır.
- 2023 yılında ise 3279 yangın ile yeniden bir artış eğilimi görülmektedir.

Bu değişim, yangın önleme çalışmaları, hava koşulları ve insan faaliyetleri gibi faktörlerle ilişkilendirilebilir.



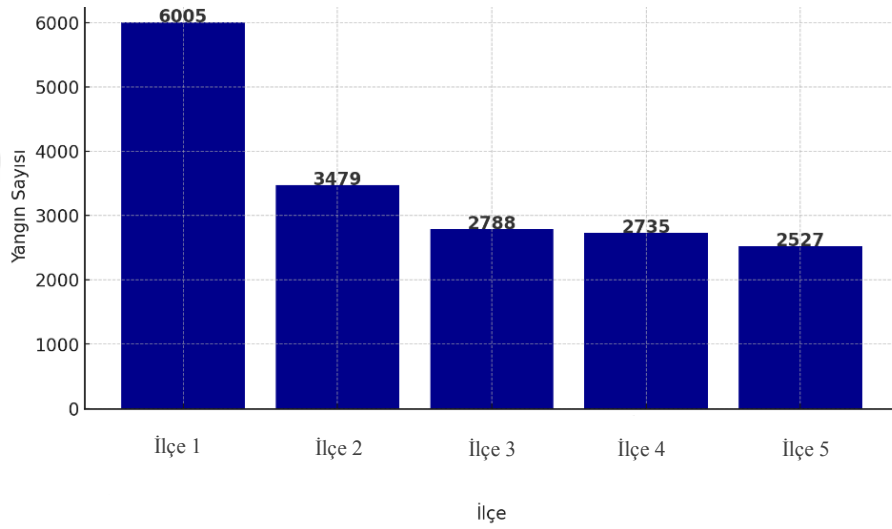
Şekil 5.1. Yıllara göre yangın sayısı trendi.

5.1.2. Yangınların bölgesel dağılımı

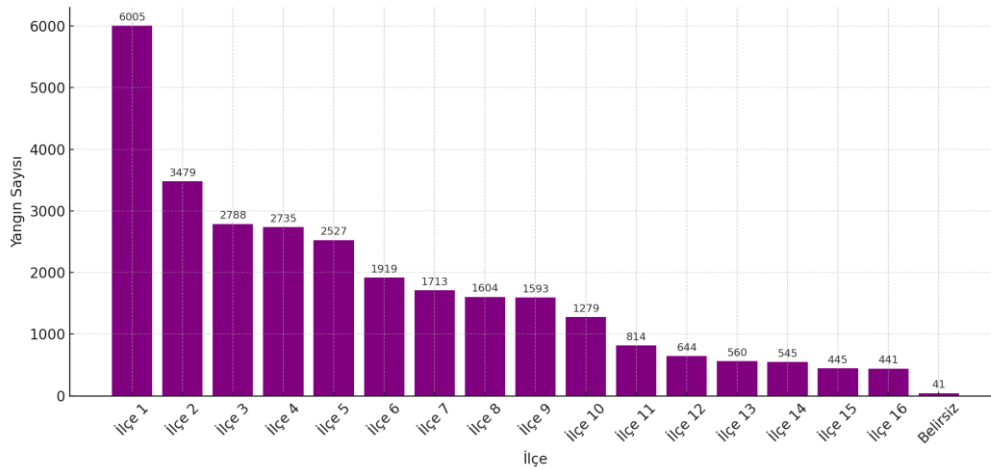
İlçelere Göre Yangın Dağılımı:

Yangın olaylarının ilçelere göre dağılımı, Şekil 5.2 ve Şekil 5.3'te gösterilmektedir.

- İlçe 1 (6005 yangın) en fazla yangın olayının yaşandığı ilçe olmuştur.
- İlçe 2 (3479), ilçe 3 (2788), ilçe 4 (2735) ve ilçe 5 (2527) yangın açısından en riskli diğer ilçelerdir.
- İlçe 14, ilçe 15 ve ilçe 16 da yangın sayıları oldukça düşüktür.



Şekil 5.2. En riskli ilçeler (en çok yangın çıkan ilk 5).



Şekil 5.3. İlçelere göre yangın dağılımı.

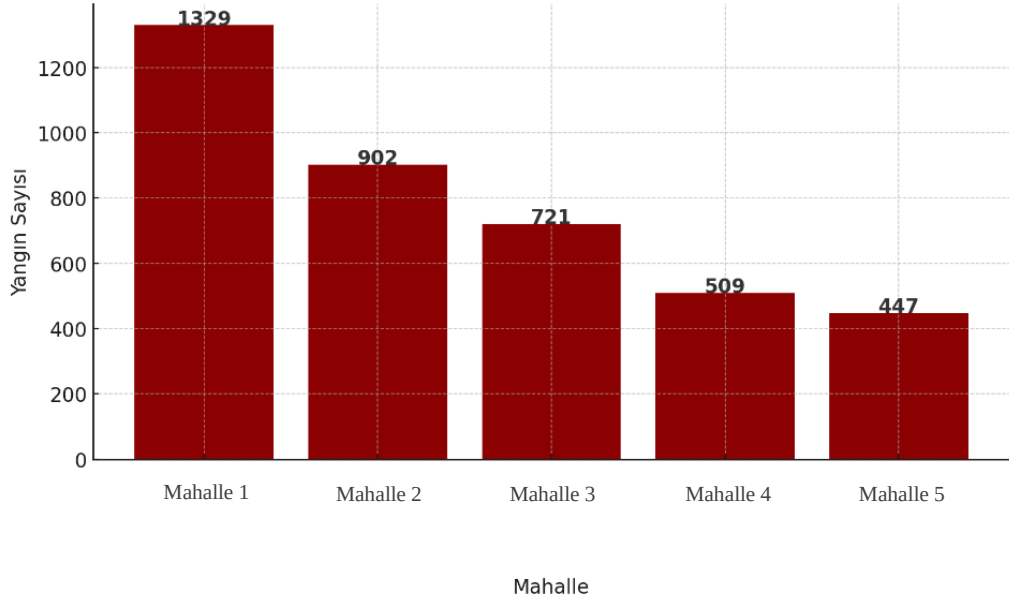
5.1.3. Mahallelere göre yangın dağılımı

Yangınların mahallelere göre dağılımı Şekil 5.4'te verilmiştir.

- Mahalle 1 (1329 yangın) en riskli mahalledir.

- Mahalle 2 (902), mahalle 3 (721), mahalle 4 (509) ve mahalle 5 (447) yangın açısından dikkat çeken diğer mahallelerdir.

Bu dağılım, şehirleşme, sanayi bölgeleri ve nüfus yoğunluğu gibi faktörlerle doğrudan ilişkili olabilir.



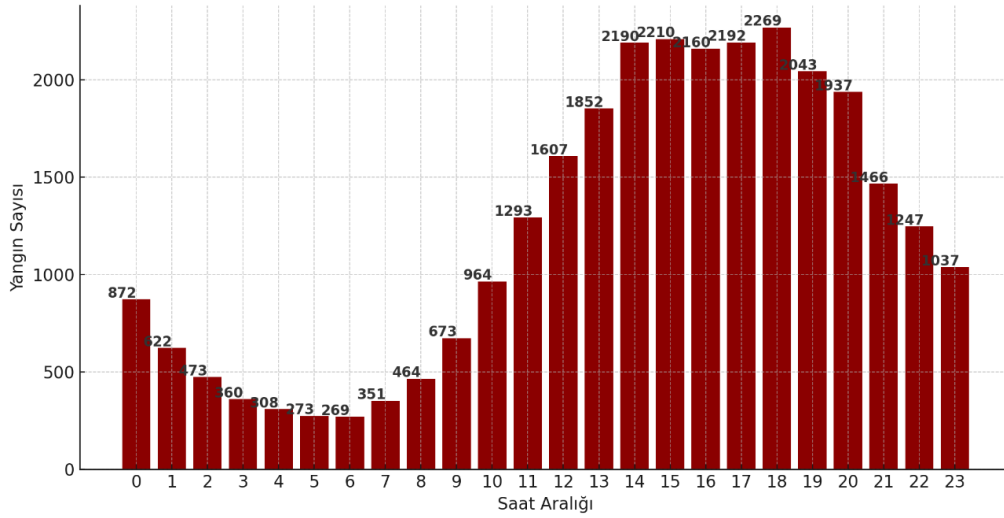
Şekil 5.4. En riskli mahalleler (en çok yangın çıkan ilk 5).

5.1.4. Yangınların gün ve saat içerisindeki dağılımı

Günün Saatlerine Göre Yangın Dağılımı:

Yangınların gün içindeki saatlere göre dağılımı Şekil 5.5'te gösterilmektedir.

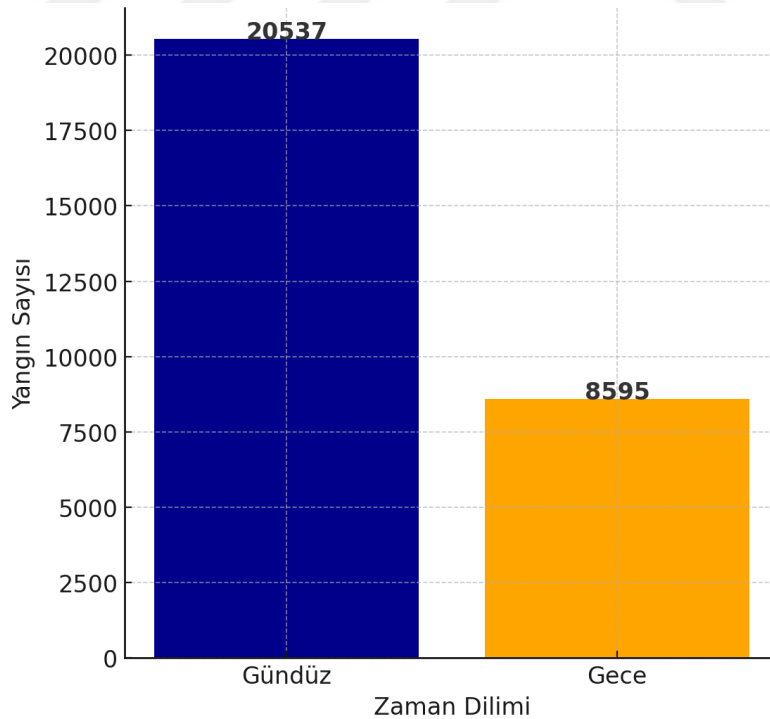
- En fazla yangın 17:00 saat diliminde (2269 yangın) meydana gelmiştir.
- Öğle ve akşam saatlerinde yangın sayıları belirgin şekilde artmaktadır.
- Gece saatlerinde yangın sayıları düşük olmakla birlikte, gece yarısından sonra 00:00-01:00 saatlerinde küçük bir artış gözlemlenmektedir.



Şekil 5.5. Yangınların gün içindeki dağılımı.

Gece ve Gündüz Yangın Dağılımı:

- Gündüz saatlerinde (06:00-18:00) yangın sayısı 20.537, gece saatlerinde (18:00-06:00) ise 8595'tir.
- Gündüz saatlerinde yangınların daha yoğun olması, insan faaliyetlerinin artışıyla doğrudan ilişkilendirilebilir (Şekil 5.6).



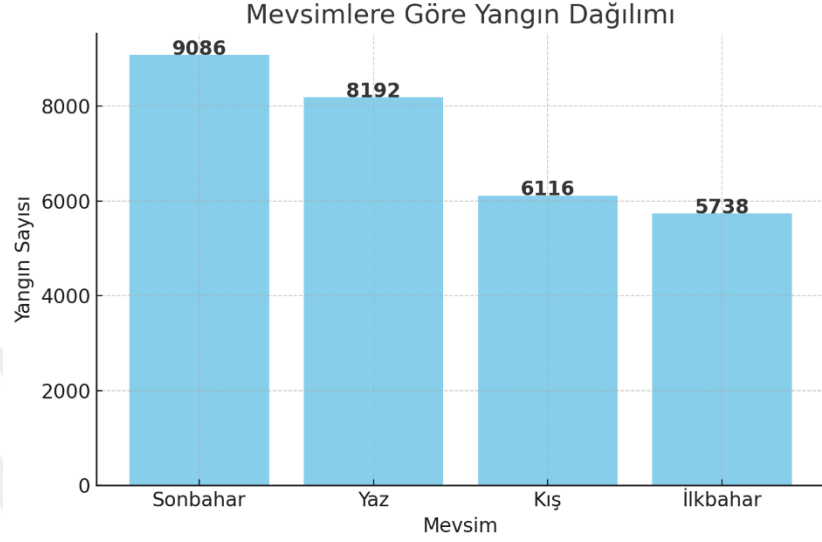
Şekil 5.6. Gece ve gündüz yangın dağılımı.

5.1.5. Yangınların mevsimlere göre dağılımı

Sonbahar (9086) ve yaz (8192) mevsimleri yangınların en yoğun olduğu dönemlerdir.

Kış (6116) ve ilkbahar (5738) mevsimlerinde yangın sayılarında belirgin bir azalma görülmektedir (Şekil 5.7).

Bu durum, hava sıcaklıklarının yangın riskini artırdığı ve özellikle sonbaharda kuru havanın etkili olduğu şeklinde yorumlanabilir.



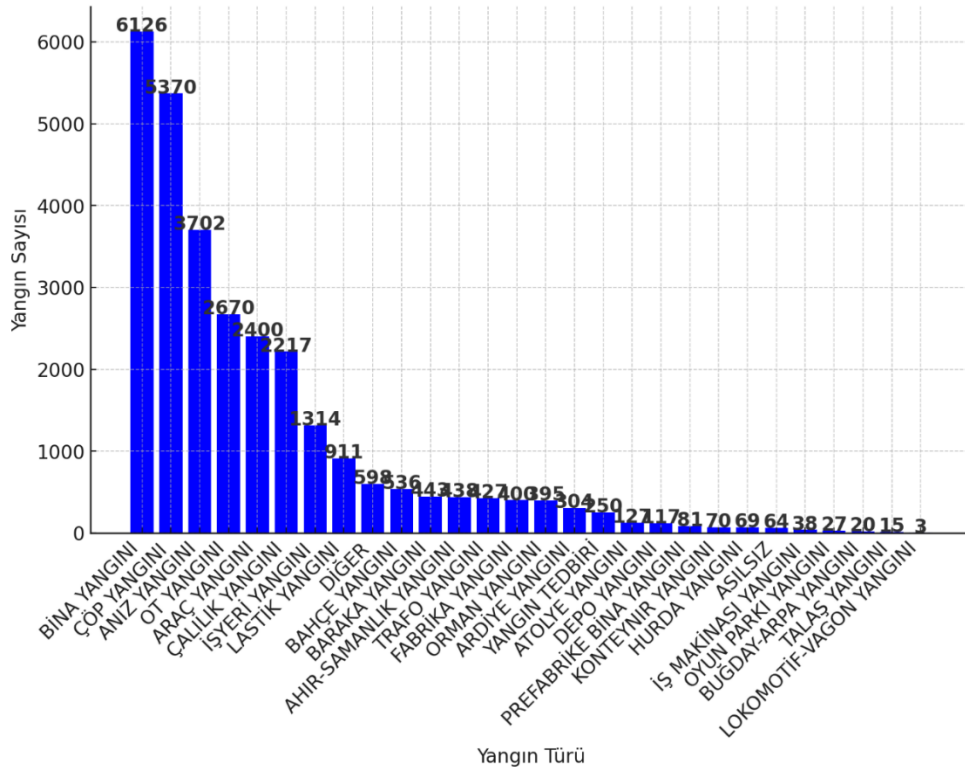
Şekil 5.7. Mevsimlere göre yangın dağılımı.

5.1.6. Yangın türleri ve mevsimsel dağılımı

Yangınların türlerine göre dağılımı ve mevsimsel etkileri, yangın önleme stratejileri açısından kritik bir öneme sahiptir. Şekil 5.8 ve Şekil 5.9, yangınların hangi türlerde yoğunlaştığını ve hangi mevsimlerde daha fazla meydana geldiğini göstermektedir.

Yangın Türlerine Göre Dağılım:

- Bina yangınları (6126) en yaygın yangın türü olup, toplam yangınlar içinde en büyük paya sahiptir.
- Çöp yangınları (5370) ve anız yangınları (3702) da sık karşılaşılan yangın türleri arasındadır.
- Araç yangınları (2670), çalılık yangınları (2400) ve iş yeri yangınları (2217) diğer önemli yangın türleridir.
- Daha düşük frekansta gerçekleşen yangın türleri arasında trafik kazası sonucu çıkan yangınlar (71), akaryakıt parlaması (69) ve lokomotif-vagon yangınları (3) yer almaktadır (Şekil 5.8).

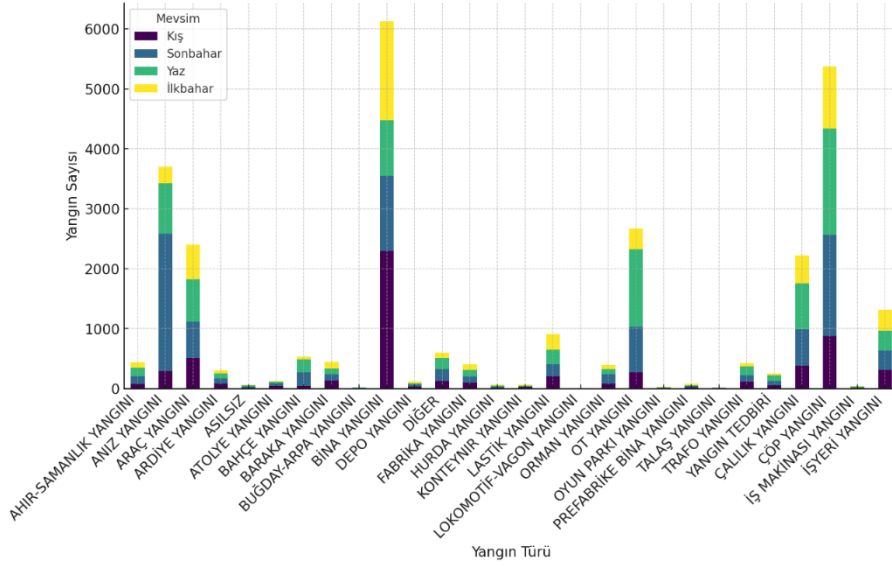


Şekil 5.8. Yangın türlerinin dağılımı.

Mevsimlere Göre Yangın Türleri:

Şekil 5.9, yangın türlerinin mevsimlere göre nasıl değiştiğini göstermektedir.

- Bina yangınları tüm mevsimlerde yüksek seyretmekle birlikte, özellikle kış aylarında artış göstermektedir.
- Anız, çöp ve çalılık yangınları genellikle yaz ve sonbahar aylarında daha sık görülmektedir.
- Depo yangınları ve fabrika yangınları yıl boyunca dengeli bir şekilde dağılmış görünmektedir, ancak sanayi bölgelerinde belirli mevsimlerde artış gösterebilir.
- Orman yangınları ve anız yangınları özellikle yaz aylarında yoğunlaşmaktadır, bu da sıcak hava ve kuraklık gibi çevresel faktörlerin etkisini göstermektedir.

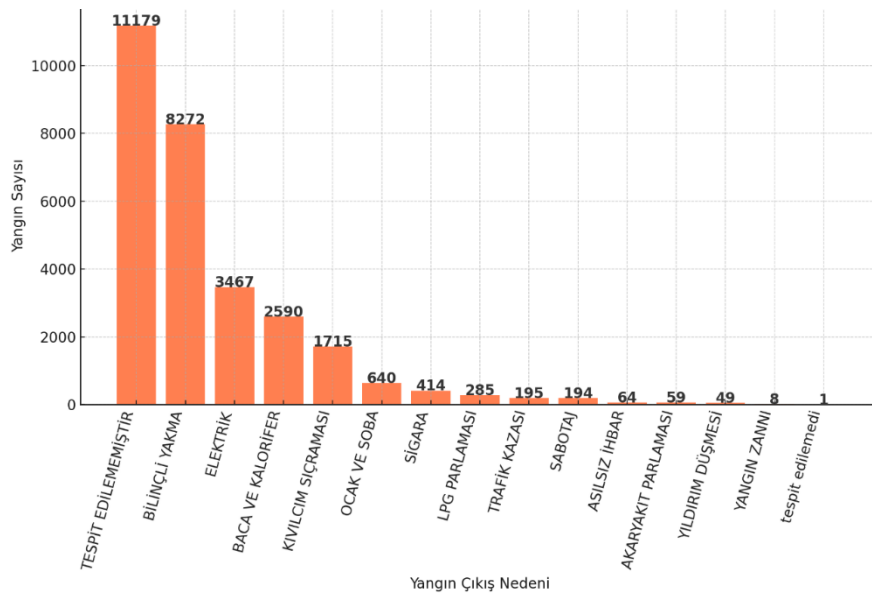


Şekil 5.9. Yangın türlerine göre mevsimsel dağılım.

5.1.7. Yangın çıkış nedenleri

Yangın çıkış nedenlerinin dağılımı Şekil 5.10'da gösterilmiştir.

- 11.179 yangının nedeni tespit edilememiştir, bu da kayıt sistemlerinin daha iyi hale getirilmesi gerektiğini göstermektedir.
- Bilerek çıkarılan yangınlar (8272), elektrik kaynaklı yangınlar (3467) ve baca/kalorifer sistemlerinden kaynaklanan yangınlar (2590) en yaygın nedenlerdir.
- Sigara, trafik kazası ve sabotaj gibi nedenlerin daha düşük oranlarda olduğu görülmektedir.

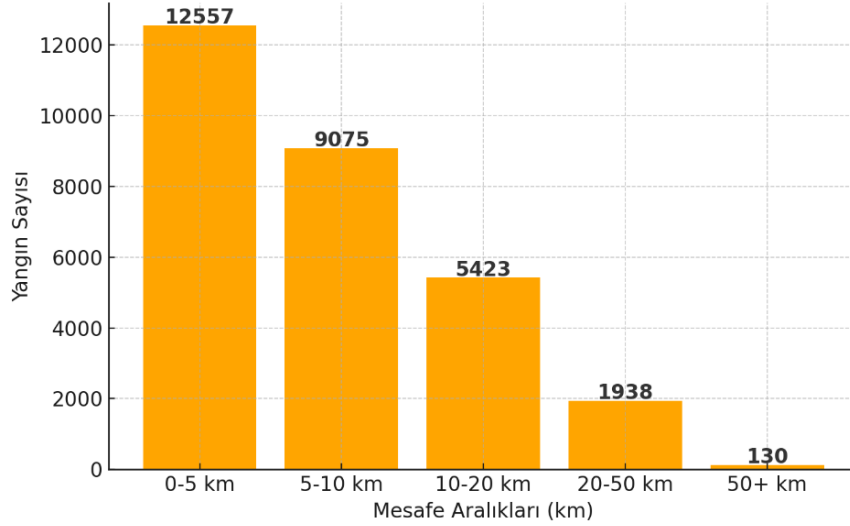


Şekil 5.10. Yangın çıkış nedenlerinin dağılımı.

5.1.8. Yangınların mesafeye göre dağılımı

Yangınların büyük çoğunluğu (12.557 yangın) 0-5 km mesafe içerisinde gerçekleşmiştir.

Mesafe arttıkça yangın sayılarında belirgin bir azalma görülmektedir (Şekil 5.11).



Şekil 5.11. Yangın mesafesi dağılımı.

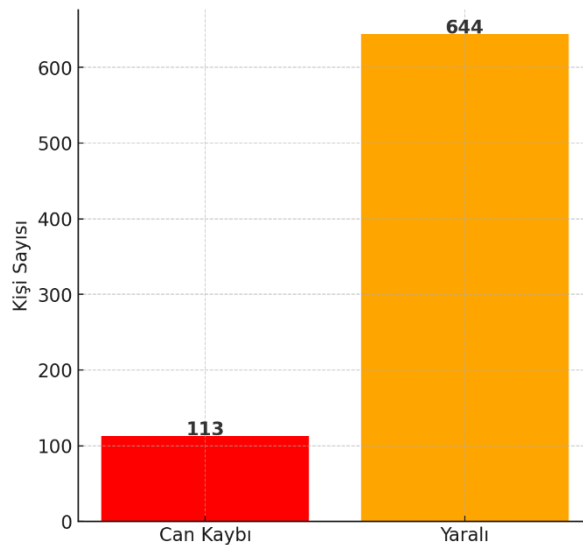
5.1.9. Yangınların insan üzerindeki etkileri

Can Kaybı ve Yaralanmalar:

113 kişi yangınlar nedeniyle hayatını kaybetmiştir.

644 kişi yangınlardan dolayı yaralanmıştır (Şekil 5.12).

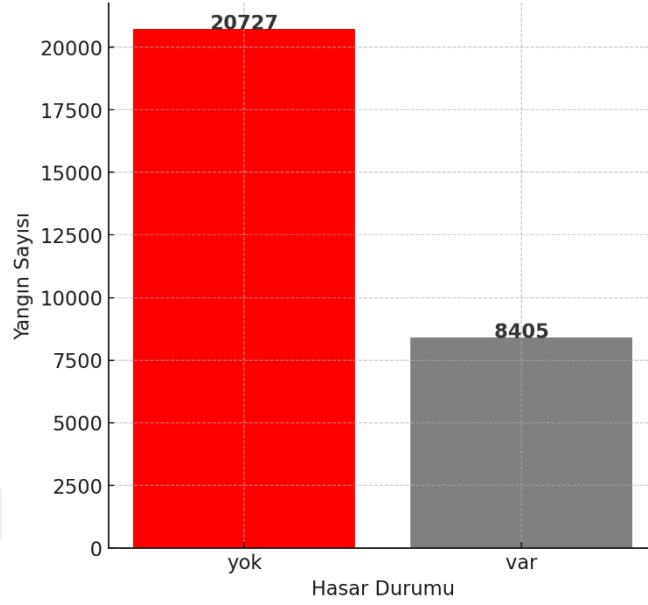
Bu veriler, yangın güvenliği önlemlerinin artırılması ve yangına karşı bilinçlendirme çalışmalarının önemini göstermektedir.



Şekil 5.12. Yangınlardan kaynaklanan can kaybı ve yaralanmalar.

5.1.10. Hasar durumu

20.727 yangında herhangi bir hasar oluşmazken, 8405 yangın maddi hasara yol açmıştır (Şekil 5.13).



Şekil 5.13. Yangınlarda hasar durumu.

Yapılan analizler, bir il de 2015-2023 yılları arasında gerçekleşen yangınların belirli dönemlerde, bölgelerde ve nedenlere bağlı olarak yoğunlaştığını göstermektedir. Yangın sayıları yıllara göre dalgalı bir seyir izlese de 2020 yılı en yüksek yangın sayısına sahip yıl olmuştur. İlçeler bazında değerlendirildiğinde, ilçe 1, ilçe 2 ve ilçe 3 en fazla yangın yaşanan bölgeler olarak öne çıkarken, mahalle düzeyinde mahalle 1, mahalle 2, mahalle 3 ve mahalle 4 en riskli alanlar arasında yer almaktadır. Yangınların zamansal dağılımı incelendiğinde, en fazla yangının öğleden sonra ve akşam saatlerinde meydana geldiği, özellikle 17:00 saat diliminde zirveye ulaştığı tespit edilmiştir. Gece saatlerinde yangın sayıları görece daha düşük seyrete de, gece yarısından sonra 00:00-01:00 saatlerinde küçük bir artış gözlemlenmiştir. Bu durum, yangınların büyük ölçüde insan faaliyetlerinden etkilendiğini ve özellikle gündüz saatlerinde artış gösterdiğini ortaya koymaktadır.

Mevsimsel olarak bakıldığında, yangınların en fazla sonbahar ve yaz aylarında meydana geldiği, buna karşılık kış ve ilkbahar dönemlerinde daha düşük seviyelerde seyrettiği görülmektedir. Bu durum, hava sıcaklıklarının artışı, kuraklık ve çevresel faktörlerle doğrudan ilişkilendirilebilir. Yangın türleri incelendiğinde, bina, çöp ve anız yangınlarının en yaygın türler olduğu anlaşılmaktadır. Özellikle bina yangınları kış aylarında, anız ve çöp yangınları ise yaz ve sonbaharda daha sık görülmektedir.

Yangın çıkış nedenleri açısından değerlendirildiğinde, elektrik arızaları, baca ve kalorifer sistemleri ile bilinçli yangın çıkarmanın en yaygın nedenler arasında olduğu tespit edilmiştir. Ancak 11.179 yangının nedeni tespit edilememiştir, bu da kayıt sistemlerinin iyileştirilmesi gerektiğini ortaya koymaktadır.

Yangınların mesafeye göre dağılımı incelendiğinde, çoğunluğunun 0-5 km mesafe içerisinde meydana geldiği ve şehir içi bölgelerde yangın riskinin daha yüksek olduğu görülmektedir. Yangınların insan hayatına olan etkileri değerlendirildiğinde, 113 kişinin hayatını kaybettiği, 644 kişinin ise yaralandığı tespit edilmiştir. Maddi hasar açısından bakıldığında ise 20.727 yangının herhangi bir hasara neden olmadığı, ancak 8405 yangının maddi kayıplara yol açtığı belirlenmiştir. Bu bulgular, yangınların yalnızca maddi zararlarla sınırlı kalmadığını, aynı zamanda insan hayatını da doğrudan tehdit ettiğini göstermektedir.

Genel olarak, yangınların yoğun olduğu ilçeler, mahalleler ve saat dilimlerinde yangın önleme ve müdahale stratejilerinin güçlendirilmesi gerekmektedir. Özellikle yüksek yangın riski taşıyan bölgelerde denetimlerin artırılması, elektrik tesisatlarının düzenli kontrol edilmesi ve baca temizliğinin teşvik edilmesi gibi önlemler hayata geçirilmelidir. Ayrıca, toplumun yangın riskleri konusunda bilinçlendirilmesi, yangın güvenlik önlemlerinin yaygınlaştırılması ve erken müdahale sistemlerinin geliştirilmesi, yangın sayılarının ve etkilerinin azaltılmasına katkı sağlayacaktır. Yangınlara ilişkin elde edilen veriler, gelecekte daha etkin yangın önleme politikalarının oluşturulması açısından önemli bir kaynak niteliğindedir.

5.2. Makine Öğrenmesi Modelleme Bulguları

5.2.1. Veri seti özellikleri ve ön işleme süreci

Bu çalışmada analiz edilen veri seti, bir ile ait 29.135 satırlık yangın vakası kaydını içermektedir. Her bir kayıt, bir yangın olayına ilişkin mekânsal (ilçe, mahalle, mesafe), zamansal (yıl, ay, saat, gece/gündüz) ve etkisel (yangının türü, çıkış nedeni, hasar durumu, yaralı, can kaybı) özellikleri kapsamaktadır. Bu özelliklere dayanarak yangınların şiddeti, üç ana sınıfta etiketlenmiştir: Başlangıç, Kısmen ve Tamamen. Sınıflar, yangının yayılma düzeyi ve kontrol altına alınma derecesine göre belirlenmiştir.

Veri seti, makine öğrenmesi modellemesine uygun hale getirilmeden önce aşağıdaki ön işleme adımlarından geçirilmiştir:

- Eksik veriler temizlenmiş, boş veya anlam ifade etmeyen değerler veri setinden çıkarılmıştır.
- Kategorik veriler sayısal değerlere dönüştürülmüş (örneğin: ilçe, mahalle, yangın türü gibi kategoriler etiket kodlamasıyla sayısallaştırılmıştır).
- Özellik ölçekleme işlemi, yalnızca SVM modelinde uygulanmıştır.
- Veri seti, %80 eğitim ve %20 test olacak şekilde ayrılmıştır.

Modelin genelleme kabiliyetini artırmak amacıyla 5 katlı çapraz doğrulama yöntemi uygulanmıştır.

Modelleme süreci başlamadan önce, uygun doğrulama stratejisi MATLAB Makine Öğrenimi Aracı üzerinden belirlenmiştir. Bu adım, modelin yalnızca eğitim verisi üzerinde değil, daha önce görmediği veriler üzerinde de tutarlı sonuçlar üretmesini sağlamak amacıyla uygulanmıştır.

Şekil 5.14’de doğrulama yöntemi seçiminin yapıldığı arayüz yer almaktadır.

Data set

Data Set Variable: sakarya yng tum veriler (29135x17 table)

Response: YanmaDerecesi (categorical, 3 unique)

Predictors

	Name	Type	Range
☒	x_le	categorical	17 unique
☒	Mahalle	string	586 unique
☒	Yang_nT_r_	categorical	28 unique
☐	YanmaDerecesi	categorical	3 unique
☒	Yang_n_k_Nedeni_K_sa	categorical	15 unique
☒	Yang_n_k_Nedeni	categorical	8026 unique

[How to prepare data](#)

Validation

Validation Scheme: Cross-Validation

Protects against overfitting. For data not set aside for testing, the app partitions the data into folds and estimates the accuracy on each fold.

Cross-validation folds: 5

[Read about validation](#)

Test

☒ Set aside a test data set

Percent set aside: 20

Use a test set to evaluate model performance after tuning and training models. To import a separate test set instead of partitioning the current data set, use the Test Data button after starting an app session.

[Read about test data](#)

Start Session Cancel

Şekil 5.14. Doğrulama tekniği seçim ekranı.

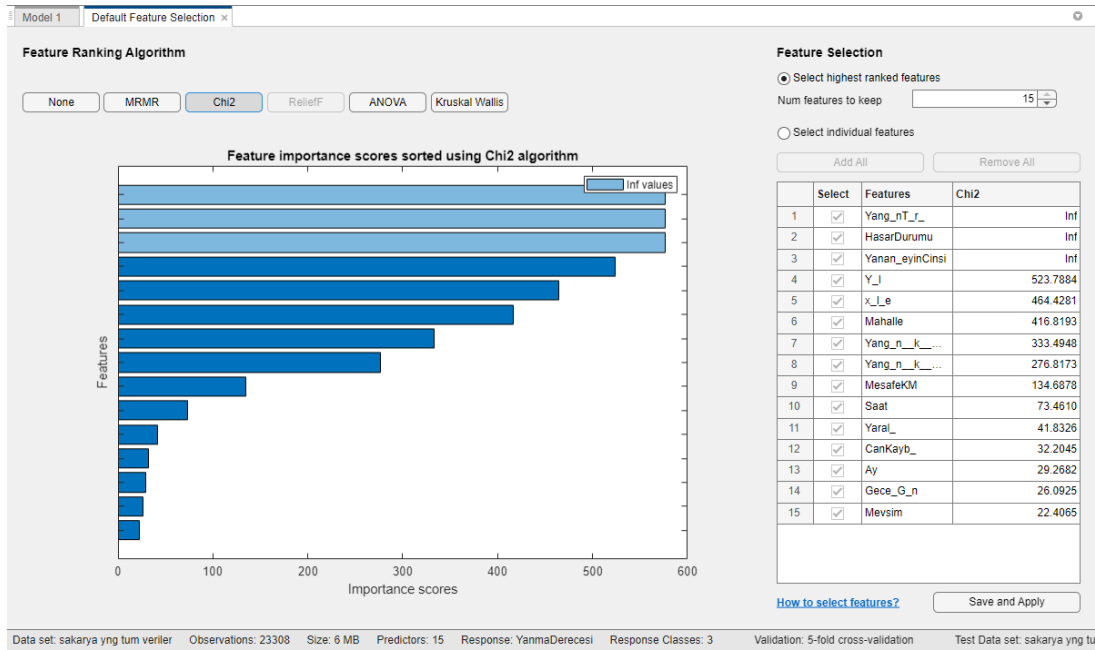
Çapraz doğrulama, veri setinin farklı alt kümelerinde dönüşümlü olarak eğitim ve test işlemlerine tabi tutulmasını sağlar. Böylece model, veri setine aşırı uyum sağlama (overfitting) eğilimi göstermeden genelleme yapabilir hale gelir. Bu yöntem sayesinde, sınıflar arasında dengesizlik olması durumunda bile her sınıfın test aşamasında temsil edilmesi garanti altına alınır.

5.2.2. Özellik seçimi ve etkili değişkenlerin belirlenmesi

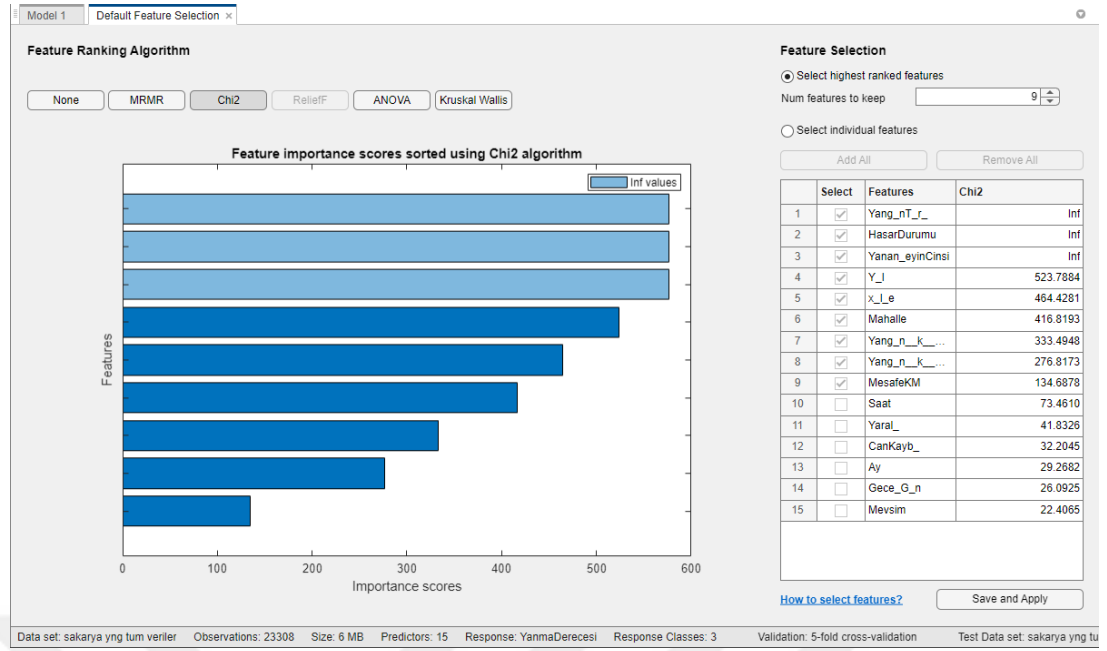
Makine öğrenmesi modellerinin başarısı, yalnızca algoritmaların gücüne değil, aynı zamanda modele sunulan girdilerin (özelliklerin) kalitesine de bağlıdır. Bu nedenle, model eğitimi öncesinde özellik seçimi süreci gerçekleştirilmiş ve veri setinde yer alan 15 değişkenin sınıflandırma performansına etkisi istatistiksel olarak değerlendirilmiştir.

Özellik seçimi sürecinde, MATLAB Makine Öğrenimi Aracı'nda yer alan "Feature Selection" ekranı üzerinden farklı istatistiksel yöntemler (None, MRMR, Anova, Kruskal-Wallis, Ki-Kare) karşılaştırılmış ve bu çalışmada Ki-Kare testi kullanılmıştır. Bu seçim, değişkenlerin hedef sınıf ile olan ilişkisini bağımsızlık varsayımı altında ölçmesi ve sınıflandırma performansına katkı sağlayan değişkenleri sıralaması açısından uygun bulunmuştur.

Şekil 5.15'de analiz yöntemi seçim ekranı; Şekil 5.14'te ise seçilen yönteme göre elde edilen özellik sıralaması sunulmuştur.



Şekil 5.15. Analiz yönetimi seçim ekranı.



Şekil 5.16. Ki-kare özellik sıralama ekranı.

Ki-Kare analizi sonucunda, 15 özelliğin sınıflandırma üzerindeki etkisi büyüktür ve ilk 9 özelliğin modelde tutulmasına karar verilmiştir. Diğer 6 özelliğin sınıflandırma sürecine anlamlı katkı sağlamadığı görülmüş ve modelden çıkarılmıştır. Bu durum, modelin öğrenme sürecini sadeleştirmiş, aynı zamanda eğitim süresini azaltarak daha kararlı tahminler üretilmesini sağlamıştır.

Aşağıda Tablo 5.1'de, özelliklerin etkisine göre sıralaması verilmiştir:

Tablo 5.1. Ki-kare testine göre özelliklerin önem sıralaması.

Sıra	Özellik Adı	Açıklama
1	Yangın Türü	Olayın türsel sınıflandırması
2	Hasar Durumu	Maddi hasarın varlığı ve seviyesi
3	Yanan Şeyin Cinsi	Yangına neden olan materyal
4	Yıl	Olayın gerçekleştiği yıl
5	İlçe	Coğrafi konum bilgisi
6	Mahalle	Daha spesifik yer bilgisi
7	Çıkış Nedeni (Kısa)	Kısa açıklamalı yangın nedeni
8	Çıkış Nedeni (Detaylı)	Uzun açıklamalı yangın nedeni
9	Mesafe (km)	İtfaiye müdahalesine olan mesafe
10	Saat	Yangının çıktığı saat
11	Yaralı	Yaralanma bilgisi
12	Can Kaybı	Hayat kaybı bilgisi
13	Ay	Olayın gerçekleştiği ay
14	Gece/Gündüz	Olayın zaman dilimi
15	Mevsim	Olayın gerçekleştiği mevsim

Seilen ilk 9 zellik model eęitimi iin kullanılmıř, geri kalan zellikler ıkarılmıřtır. zellik seim sreci, modelin doęruluk ve genel bařarımını artırmada nemli rol oynamıř ve ayrıca ařırı ęrenme (overfitting) riskini azaltmıřtır.

5.2.3. Doęrulama yntemi ve modelleme yaklařımı

Makine ęrenmesi alıřmalarında modelin yalnızca eęitim verisi zerinde iyi sonular vermesi yeterli deęildir; asıl nemli olan, modelin daha nce grmedięi veriler zerindeki tahmin bařarısıdır. Bu nedenle, alıřmada 5 katlı apraz doęrulama (5-fold cross-validation) yntemi tercih edilmiřtir. Bu yntem, modelin genelleme kabiliyetini objektif bir biimde deęerlendirmeye olanak saęlar.

apraz doęrulama teknięinde, eęitim veri kmesi eřit byklkte beř alt gruba ayrılır. Her bir adımda drt grup eęitim iin, kalan bir grup ise doęrulama (validation) iin kullanılır. Bu iřlem beř kez tekrarlanarak her alt kme bir kez doęrulama kmesi olarak kullanılır. Sonu olarak, modelin tm veri alt kmeleri zerinde denenmesi saęlanmış olur. Bu yapı, veriye ařırı uyum saęlama riskini azaltır ve modelin farklı veri alt gruplarındaki kararlılıęını test eder.

Daha nce řekil 5.14’de sunulan “Doęrulama Teknięi Seim Ekranı”, bu iřlemin MATLAB Makine ęrenimi Aracı’nda nasıl yapılandırıldığını gstermektedir. Bu ekran zerinden, 5 katlı apraz doęrulama seilerek modelleme iřlemi bařlatılmıřtır (řekil 5.14).

Bu alıřmada modelleme sreci ařaęıdaki ařamalardan oluřmuřtur:

- Veri n İřleme: Eksik deęerlerin temizlenmesi ve kategorik verilerin kodlanması.
- zellik Seimi: Ki-Kare testi ile en anlamlı zelliklerin belirlenmesi.
- Model Seimi: Karar Aęacı (Fine Tree), Linear Support Vector Machine (SVM) ve Gaussian Naive Bayes (GNB) algoritmaları kullanılmıřtır.
- apraz Doęrulama: Her model, 5 katlı apraz doęrulama yntemiyle deęerlendirilmiřtir.
- Performans lm: Doęruluk, karıřıklık matrisi ve ROC eęrileri zerinden analizler gerekleřtirilmiřtir.

Bu yntemsel yapı hem sınıf dengesizlięinin etkisini azaltmıř hem de her modelin genelleme kapasitesini daha gvenilir bir biimde ortaya koymuřtur. Ayrıca bu

yaklaşım sayesinde, düşük temsil oranına sahip sınıfların (özellikle “tamamen” sınıfının) model içinde yer bulması sağlanmıştır.

5.2.4. Eğitim aşamasında elde edilen performans bulguları

Makine öğrenmesi modellerinin eğitim süreci sonucunda, üç farklı algoritma ile elde edilen performans ölçümleri bu bölümde karşılaştırmalı olarak sunulmuştur. Kullanılan modeller; Karar Ağacı (Fine Tree), Gaussian Naive Bayes (GNB) ve Doğrusal Destek Vektör Makinesi (Linear SVM) olarak belirlenmiştir. Bu modeller, özellik seçimi yapılmış ve dengelenmiş veri seti üzerinde, 5 katlı çapraz doğrulama yöntemiyle eğitilmiştir. Eğitim süreci sonunda her modelin doğruluk oranı hesaplanarak Tablo 5.2’de verilmiştir.

Tablo 5.2. Eğitim aşamasında modellerin doğruluk oranları.

Model Adı	Doğruluk (%)
Karar Ağacı (Fine Tree)	60,1
Gaussian Naive Bayes (GNB)	60,0
Linear SVM	61,4

Elde edilen sonuçlara göre, doğruluk açısından en iyi performans %61,4 ile Linear SVM modeli tarafından gösterilmiştir. Bunu sırasıyla Karar Ağacı (%60,1) ve GNB (%60,0) modelleri takip etmiştir. Modellerin genel doğruluk oranları birbirine oldukça yakın olmakla birlikte, ayrıntılı performans değerlendirmesi için karışıklık matrislerine bakılması gerekmektedir.

Her model için oluşturulan karışıklık matrisleri; modelin sınıflar (Başlangıç, Kısmen, Tamamen) üzerindeki tahmin doğruluğunu daha detaylı bir biçimde ortaya koymaktadır. Şekil 5. 17, Karar Ağacı modeline ait karışıklık matrisini göstermektedir. Buna göre, model “Kısmen” sınıfında 9760 doğru tahminle yüksek bir başarı sergilemiştir. Ancak “Başlangıç” sınıfında yer alan 3636 örnek yanlışlıkla “Kısmen” sınıfında tahmin edilmiştir. Benzer şekilde, “Tamamen” sınıfı da büyük ölçüde karışmış ve çoğunlukla “Kısmen” olarak sınıflandırılmıştır.

Doğru Sınıf	Başlangıç	3740	3636	150
	Kısmen	2419	9760	569
	Tamamen	411	2109	514
		Başlangıç	Kısmen	Tamamen
		Tahmin Sınıfı		

Şekil 5.17. Karar ağacı modeline ait eğitim karışıklık matrisi.

Şekil 5.18’de yer alan GNB modeline ait karışıklık matrisi ise benzer bir eğilimi göstermektedir. GNB modeli, “Kısmen” sınıfı için 9255 doğru tahminle iyi bir performans sergilemiştir. Ancak, “Başlangıç” ve “Tamamen” sınıflarındaki örneklerin önemli bir bölümü yine “Kısmen” sınıfına kaymıştır. Özellikle “Tamamen” sınıfındaki örneklerin yalnızca 812’si doğru sınıflandırılırken, 1852’si “Kısmen” olarak hatalı etiketlenmiştir.

Doğru Sınıf	Başlangıç	3917	3418	191
	Kısmen	2693	9255	800
	Tamamen	370	1852	812
		Başlangıç	Kısmen	Tamamen
		Tahmin Sınıfı		

Şekil 5.18. Gaussian naive bayes modeline ait eğitim karışıklık matrisi.

Linear SVM modelinin eğitim aşamasındaki karışıklık matrisi Şekil 5.19’de verilmiştir. Bu model, özellikle “Kısmen” sınıfında 10775 doğru tahminle en başarılı sonuçlardan birini üretmiştir. Ancak “Başlangıç” sınıfına ait 4313 örnek ve

“Tamamen” sınıfına ait 2453 örnek yanlışlıkla yine “Kısmen” olarak sınıflandırılmıştır. Bu durum, Linear SVM'in “Kısmen” sınıfını baskın tahmin etme eğiliminde olduğunu göstermektedir.

Doğru Sınıf	Başlangıç	3163	4313	50
	Kısmen	1733	10775	240
	Tamamen	197	2453	384
		Başlangıç	Kısmen	Tamamen
		Tahmin Sınıfı		

Şekil 5.19. Linear svm modeline ait eğitim karışıklık matrisi.

Genel olarak tüm modellerin, “Kısmen” sınıfını oldukça yüksek başarıyla sınıflandırdığı, ancak “Tamamen” sınıfında belirgin bir başarısızlık yaşandığı görülmektedir. Bu durum, özellikle “Kısmen” ve “Tamamen” sınıflarının özellik uzayında birbirine oldukça yakın olmasından kaynaklanabilir. Karar Ağacı modeli, daha dengeli bir dağılım sergilese de doğruluk açısından geride kalmıştır. Linear SVM modeli ise en yüksek genel doğruluğa ulaşmasına rağmen sınıflar arasında ciddi karışmalar sergilemiştir. Gaussian Naive Bayes modeli, teorik olarak basit ve hızlı çalışmasına rağmen “Başlangıç” ve “Tamamen” sınıflarında yeterli ayrım sağlayamamıştır.

Bu sonuçlar, model başarısının yalnızca genel doğruluk oranı ile değil, aynı zamanda her bir sınıf için elde edilen tahmin doğruluğu ile birlikte değerlendirilmesi gerektiğini açıkça ortaya koymaktadır.

5.2.5. Test aşaması bulguları ve gerçek performans değerlendirmesi

Modelleme sürecinin test aşaması, geliştirilen algoritmaların daha önce görmedikleri veriler üzerindeki performanslarını değerlendirmek amacıyla gerçekleştirilmiştir. Bu aşamada, eğitim sırasında kullanılan %80’lik veri kümesi dışında kalan %20’lik test verisi kullanılmıştır. Her üç model (Karar Ağacı, Gaussian Naive Bayes ve Linear

SVM) bu test verisiyle çalıştırılmış ve elde edilen doğruluk oranları Tablo 5.6'da sunulmuştur.

Tablo 5.3. Test aşamasında modellerin doğruluk oranları.

Model Adı	Doğruluk (%)
Karar Ağacı (Fine Tree)	61,2
Gaussian Naive Bayes (GNB)	58,5
Linear SVM	61,0

Test aşamasında da en yüksek doğruluk oranı Karar Ağacı modeli ile elde edilmiştir (%61,2). Linear SVM modeli %61,0 doğruluk ile bu sonucu yakından takip etmiş, GNB modeli ise %58,5 doğruluk ile diğer modellerin gerisinde kalmıştır. Ancak doğruluk oranlarının birbirine oldukça yakın olması nedeniyle, sınıflar bazında analiz yapmak model tercihinde daha belirleyici olacaktır. Bu amaçla, her model için oluşturulan karışıklık matrisleri analiz edilmiştir.

Karar Ağacı (Fine Tree) modeline ait test karışıklık matrisi aşağıda Şekil 5. 20'de sunulmuştur. Bu model, “Kısmen” sınıfında 2516 doğru tahminle en yüksek başarıyı göstermiştir. “Başlangıç” sınıfında 873 örnek doğru tahmin edilirken, 953 örnek yanlışlıkla “Kısmen” olarak sınıflandırılmıştır. “Tamamen” sınıfındaki örneklerin ise büyük bir bölümü yine “Kısmen” olarak sınıflandırılmış ve yalnızca 176 örnek doğru şekilde tahmin edilmiştir.

Doğru Sınıf	Başlangıç	873	953	56
	Kısmen	504	2516	166
	Tamamen	64	519	176
		Başlangıç	Kısmen	Tamamen
		Tahmin Sınıfı		

Şekil 5.20. Karar ağacı modeline ait test karışıklık matrisi.

Gaussian Naive Bayes (GNB) modelinin test karışıklık matrisi, Şekil 5. 21.'de gösterilmektedir. GNB modeli, “Kısmen” sınıfında 2281 doğru tahmin yapmıştır. Ancak “Başlangıç” sınıfında 902 örnek yanlışlıkla “Kısmen” olarak, “Tamamen” sınıfında ise 481 örnek yine “Kısmen” olarak sınıflandırılmıştır. Bu tablo, modelin büyük ölçüde “Kısmen” sınıfına yöneldiğini göstermektedir.

Doğru Sınıf	Başlangıç	935	902	45
	Kısmen	702	2281	203
	Tamamen	84	481	194
		Başlangıç	Kısmen	Tamamen
		Tahmin Sınıfı		

Şekil 5.21. Gaussian naive bayes modeline ait test karışıklık matrisi.

Linear SVM modeline ait test karışıklık matrisi ise Şekil 5. 22’de verilmiştir. Bu model “Kısmen” sınıfında 2695 doğru tahminle en iyi sonucu vermiştir. Ancak “Tamamen” sınıfındaki örneklerin sadece 104’ü doğru sınıflandırılmış, 605’i yine “Kısmen” olarak etiketlenmiştir. “Başlangıç” sınıfında ise 758 doğru, 1112 yanlış sınıflama yapılmıştır.

Doğru Sınıf	Başlangıç	758	1112	12
	Kısmen	421	2695	70
	Tamamen	50	605	104
		Başlangıç	Kısmen	Tamamen
		Tahmin Sınıfı		

Şekil 5.22. Linear SVM modeline ait test karışıklık matrisi.

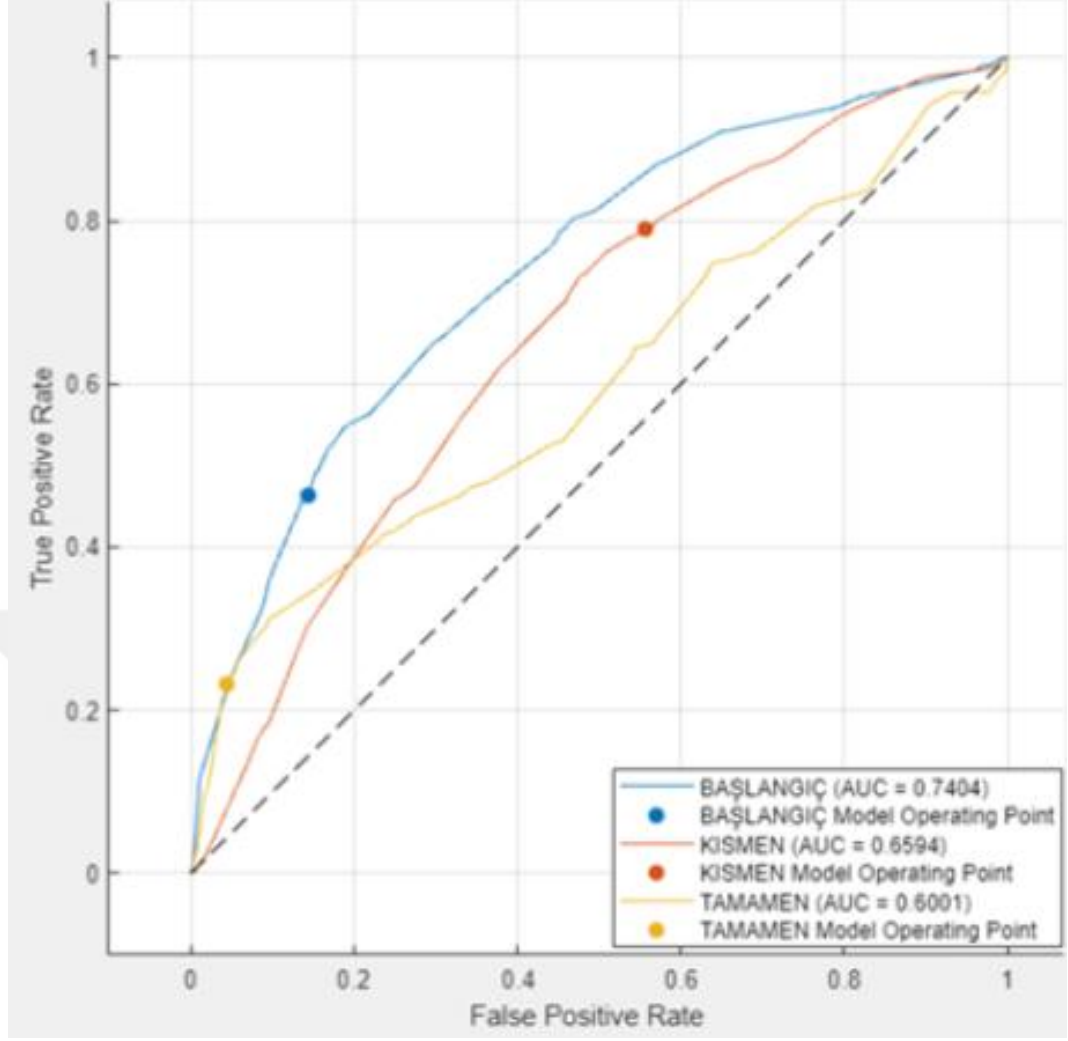
Test sonuçlarına genel olarak bakıldığında, tüm modellerin özellikle “Kısmen” sınıfında yüksek doğruluk oranlarına sahip olduğu, ancak “Tamamen” sınıfında sınıflandırma hatalarının yoğunlaştığı görülmektedir. “Tamamen” sınıfındaki düşük başarı, bu sınıfın özelliklerinin “Kısmen” sınıfına yakın olmasından kaynaklanıyor olabilir. Bu durum, hem eğitim hem de test verilerinde tutarlı bir şekilde tekrarlandığı için model yapılandırmasından ziyade veri dağılımının bir sonucu olarak değerlendirilmelidir.

Sonuç olarak, Karar Ağacı modeli sınıflar arası daha dengeli bir tahmin yapısı sergilerken, Linear SVM modeli “Kısmen” sınıfında daha baskın ve başarılı sonuçlar vermiştir. Gaussian Naive Bayes modeli ise hem doğruluk oranı hem de sınıf ayrımı açısından diğer modellere kıyasla daha zayıf bir performans ortaya koymuştur.

5.2.6. ROC Eğrileri ve AUC değerleriyle ayırt etme gücünün analizi

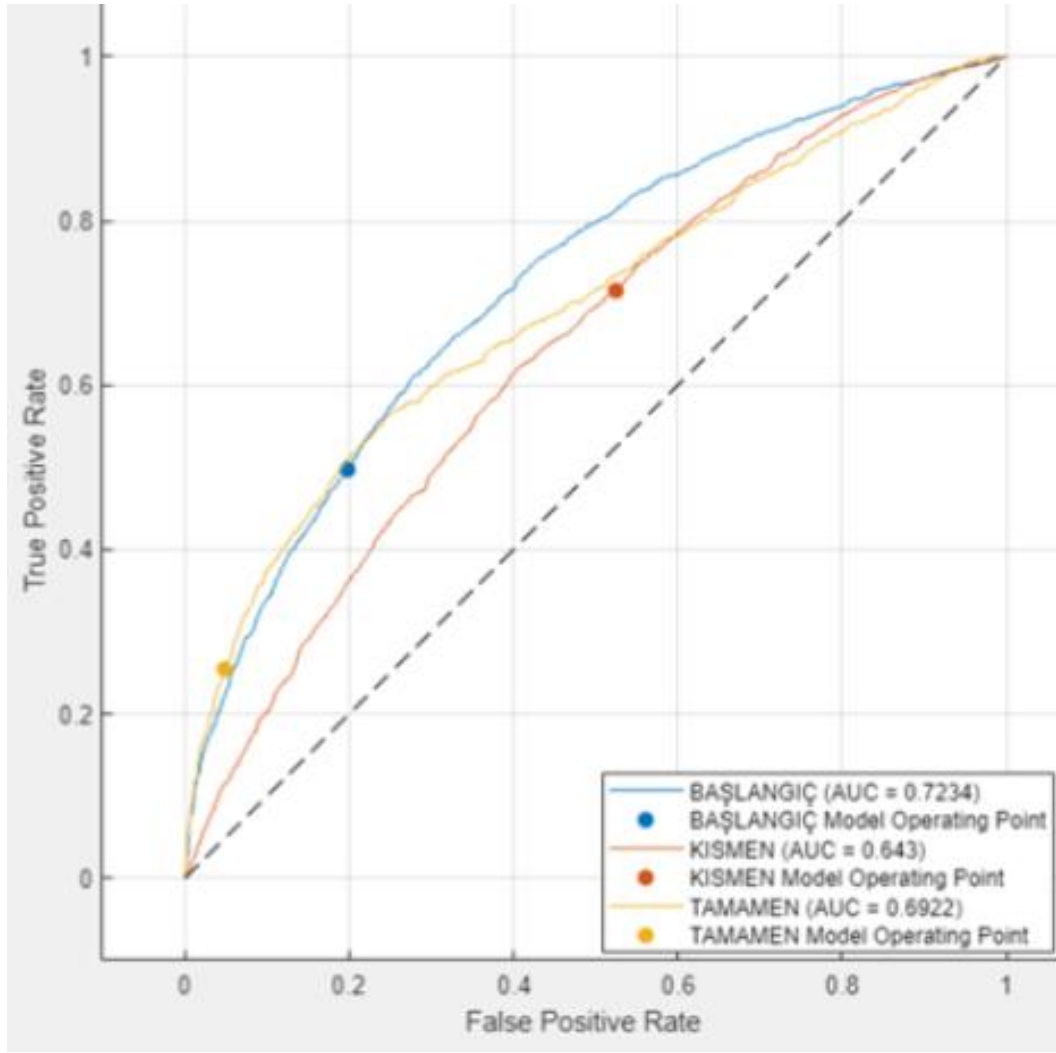
Model değerlendirmesinde yalnızca doğruluk oranları ile yetinmek, özellikle sınıf dengesizliği olan veri setlerinde yetersiz kalabilmektedir. Bu nedenle modellerin ayırt etme kabiliyetini ölçmek amacıyla, her bir sınıf için ROC (Receiver Operating Characteristic) eğrileri oluşturulmuş ve AUC (Area Under Curve) değerleri hesaplanmıştır. ROC eğrisi, bir sınıflandırma modelinin çeşitli eşik değerlerindeki doğru pozitif oranı (True Positive Rate) ile yanlış pozitif oranını (False Positive Rate) karşılaştırarak, modelin pozitif sınıfları ayırma yeteneğini grafiksel olarak gösterir.

Karar Ağacı (Fine Tree) modeli için oluşturulan ROC eğrisi Şekil 5.17’de gösterilmektedir. Bu modelin AUC değeri 0,7404 olarak hesaplanmıştır. Eğri, özellikle “Başlangıç” sınıfında belirgin şekilde üst bölgeye yakın seyretmekte, bu da modelin bu sınıfı diğerlerinden oldukça iyi ayırabildiğini göstermektedir. Elde edilen AUC değeri, modelin rastgele tahmin yapan bir sistemden belirgin biçimde üstün olduğunu ortaya koymaktadır.



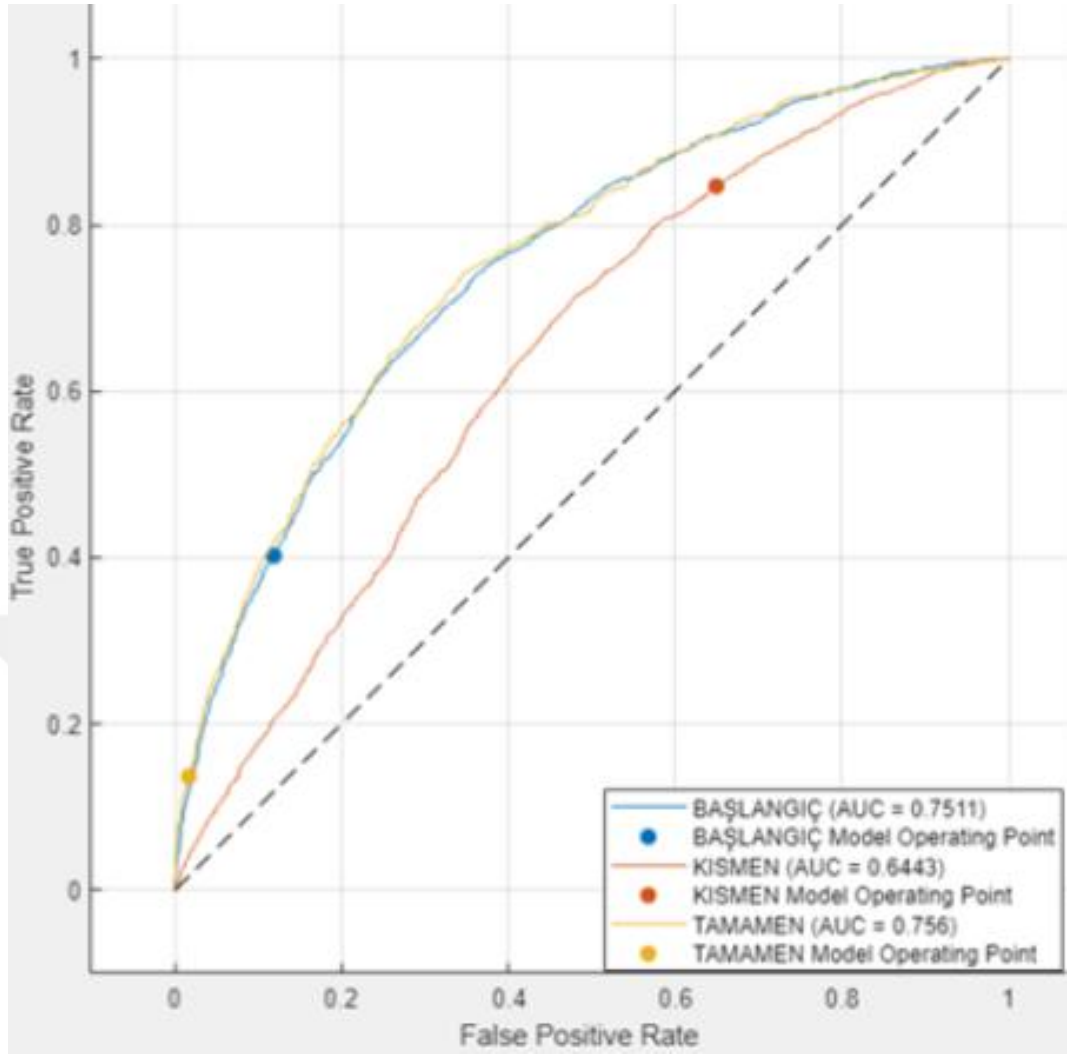
Şekil 5.23. Karar ağacı modeline ait roc eğrisi.

Linear SVM modelinin ROC eğrisi ise Şekil 5.18’de yer almaktadır. Bu modelin AUC değeri 0,7511 olarak hesaplanmış ve çalışmada en yüksek ayırt etme kabiliyetine ulaşmıştır. Eğrinin özellikle “Kısmen” ve “Tamamen” sınıfları için üst bölgeye yakın olması, bu sınıflarda daha güçlü pozitif tahminler üretilebildiğini göstermektedir. Ayrıca ROC eğrisinin ideal forma (yani sol üst köşeye yaklaşan eğri) oldukça yakın olması, modelin genellikle düşük yanlış pozitif oranı ile yüksek doğru pozitif oranı sağladığını ortaya koyar.



Şekil 5.24. Linear svm modeline ait roc eğrisi.

Gaussian Naive Bayes (GNB) modeline ait ROC eğrisi Şekil 5.19’da gösterilmektedir. Bu modelin AUC değeri 0.7234 olup, diğer modellere göre daha düşük bir ayırt etme performansı sergilemiştir. Özellikle “Tamamen” sınıfında ROC eğrisi, 45 dereceye yakın eğilim göstermiştir. Bu durum, modelin pozitif ve negatif örnekleri ayırt etmede zorlandığını ve tahmin başarısının sınırlı olduğunu ortaya koymaktadır.



Şekil 5.25. Gaussian naive bayes modeline ait roc eğrisi.

Tüm modellerde gözlemlenen ortak durum, “Kısmen” sınıfının ROC eğrilerinde diğer sınıflara kıyasla daha net ayrıştığı ve ayırt edicilik düzeyinin yüksek olduğudur. Buna karşın, “Tamamen” sınıfı için elde edilen ROC eğrilerinde eğriler 0,5 değerine (yani rasgele tahmin çizgisine) daha yakın seyretmiştir. Bu durum, bu sınıfın hem eğitim hem de test verilerinde yetersiz temsil edilmesinden veya bu sınıfa ait özelliklerin diğer sınıflardan net biçimde ayrışmamasından kaynaklanıyor olabilir.

Sonuç olarak ROC eğrileri, doğruluk oranlarından bağımsız olarak modellerin ayırt etme gücünü ortaya koymakta önemli bir araç olmuştur. Elde edilen AUC değerlerine göre, Linear SVM modeli hem genel doğruluk hem de ayırt edicilik açısından en güçlü model olarak öne çıkmıştır. Karar Ağacı modeli de güçlü performans sergilerken, Gaussian Naive Bayes modeli ROC açısından daha sınırlı bir başarı elde etmiştir.

5.2.7. Model performans ölçümlerinin karşılaştırmalı analizi

Modellerin genel doğruluk oranları, sınıflandırma başarısı hakkında önemli bir fikir verse de; özellikle çok sınıflı problemler için bu ölçüm tek başına yeterli değildir. Bu nedenle, her bir model için Precision (kesinlik), Recall (duyarlılık) ve F1-Score gibi detaylı sınıf bazlı metrikler hesaplanmış ve performans daha bütüncül bir yaklaşımla değerlendirilmiştir. Bu metrikler, modelin yalnızca doğru tahmin oranlarını değil, aynı zamanda yanlış pozitif ve yanlış negatif tahminleri de dikkate alarak dengeli bir başarı değerlendirmesi sunar.

Bu kapsamda, her bir model için Başlangıç, Kısmen ve Tamamen sınıflarına ait ölçümler aşağıda Tablo 5.10’da sunulmuştur.

Tablo 5.4. Modellerin precision, recall ve f1-score performans ölçümleri.

Model	Doğruluk	Başlangıç Precision	Başlangıç Recall	Başlangıç F1	Kısmen Precision	Kısmen Recall	Kısmen F1	Tamamen Precision	Tamamen Recall	Tamamen F1
1	0,6013	0,5693	0,4969	0,5306	0,6295	0,7656	0,6909	0,4169	0,1694	0,2409
2	0,6118	0,6058	0,4639	0,5254	0,6309	0,7897	0,7014	0,4504	0,2676	0,3357
3	0,6000	0,5612	0,5205	0,5401	0,6372	0,7260	0,6787	0,4504	0,2676	0,3357
4	0,5852	0,5433	0,4968	0,5190	0,6225	0,7159	0,6660	0,4389	0,2556	0,3231
5	0,6145	0,6210	0,4203	0,5013	0,6143	0,8452	0,7115	0,5697	0,1266	0,2071
6	0,6104	0,6168	0,4028	0,4873	0,6108	0,8459	0,7094	0,5591	0,1370	0,2201

Tablo 5.10 incelendiğinde, “Kısmen” sınıfında tüm modellerin oldukça başarılı olduğu, özellikle Model 2 (Fine Tree) ve Model 5 (SVM optimize)’in bu sınıfta yüksek duyarlılık (Recall > 0,84) ve tatmin edici F1 skorları (F1 > 0,70) sunduğu görülmektedir. Bu durum, verinin büyük çoğunluğunun “Kısmen” sınıfına ait olmasıyla da ilişkilendirilebilir.

“Başlangıç” sınıfı için modellerin başarıları birbirine yakın olmakla birlikte, duyarlılık oranlarının çoğunlukla %40–%50 aralığında kaldığı görülmektedir. Özellikle Model 2 ve Model 6, kesinlik açısından güçlü sonuçlar vermiştir; bu da modelin bu sınıfı tahmin ettiğinde çoğunlukla doğru tahminde bulunduğunu ancak bazı gerçek “Başlangıç” örneklerini gözden kaçırdığını göstermektedir.

“Tamamen” sınıfı, tüm modeller açısından en düşük performansın sergilendiği sınıf olmuştur. Bu sınıfa ait F1-skor değerleri genellikle 0,20–0,33 aralığında kalmıştır.

Düşük kesinlik ve duyarlılık değerleri, bu sınıfın hem sayıca az olması hem de diğer sınıflarla ortak özellikler taşıması nedeniyle modeller tarafından yeterince ayırt edilemediğini göstermektedir. Bu sonuçlar, önceki bölümlerde karışıklık matrislerinde gözlemlenen eğilimlerle de tutarlıdır.

Genel olarak, modellerin F1 skorları %60'ın biraz üzerinde seyretmiş; bu da çok sınıflı, dengesiz veri kümelerinde makul kabul edilebilecek bir performansa işaret etmektedir. En dengeli sonuçları veren modeller Linear SVM (özellikle Model 3 ve Model 5) ve Karar Ağacı (Model 2) olarak öne çıkmaktadır. Gaussian Naive Bayes modelleri ise hem genel doğruluk hem de sınıf bazlı ölçütlerde nispeten daha düşük skorlarla sınırlı kalmıştır.

5.2.8. Bulguların yorumlanması ve kritik tespitler

Bu çalışmada, bir ile ait yangın verileri üzerinde uygulanan makine öğrenmesi modelleri; eğitim ve test aşamalarında farklı doğruluk seviyeleri ve sınıflandırma başarıları ortaya koymuştur. Elde edilen bulgular, yalnızca doğruluk oranlarıyla değil, aynı zamanda sınıf bazlı başarı ölçütleriyle birlikte değerlendirilerek kapsamlı bir analiz sunmuştur.

En belirgin gözlemlerden biri, tüm modellerin “Kısmen” sınıfında yüksek performans göstermesidir. Bu sınıfa ait örneklerin hem sayıca fazla olması hem de belirgin ayırt edici özelliklere sahip olması, modellerin bu sınıfı doğru şekilde öğrenmesini kolaylaştırmıştır. ROC eğrileri ve F1-skorları da bu durumu doğrular niteliktedir. Örneğin, Linear SVM modeli bu sınıf için hem eğitim hem test aşamasında %84'e varan duyarlılık oranına ulaşmış; karar ağacı modeli de %79'a yakın değerlerle bu başarıyı desteklemiştir.

Öte yandan, “Tamamen” sınıfı tüm modeller için en zorlayıcı sınıf olmuştur. Karışıklık matrisleri ve ROC eğrileri, bu sınıfın sıklıkla “Kısmen” sınıfı ile karıştırıldığını ve ayırt edilmesinde ciddi güçlükler yaşandığını göstermiştir. Bu durum, “Tamamen” sınıfına ait örnek sayısının düşük olmasının yanı sıra, bu sınıfa ait olayların veri setindeki özelliklerle yeterince temsil edilememesinden de kaynaklanıyor olabilir. Nitekim bu sınıfa ait F1-skorlarının çoğunlukla 0,20–0,33 aralığında kalması, modelin hem bu sınıfı tespit etmede başarısız olduğunu hem de bu sınıfa ait örnekleri yanlış sınıflara tahmin ettiğini ortaya koymaktadır.

“Başlangıç” sınıfında ise modellerin daha tutarsız bir başarı grafiği sergilediği görülmektedir. Özellikle GNB ve SVM modelleri, başlangıç sınıfındaki birçok örneği yanlışlıkla “Kısmen” olarak sınıflandırmıştır. Bu da sınıflar arasında bazı örtüşen niteliklerin bulunduğunu ve modelin karar sınırlarını net şekilde çizemediğini göstermektedir. Başlangıç sınıfına ait başarı oranlarının genellikle %50 civarında kalması, modelin bu sınıfı tam anlamıyla öğrenemediğine işaret etmektedir.

Model karşılaştırmasına bakıldığında, Linear SVM ve Karar Ağacı modelleri öne çıkmaktadır. Linear SVM, özellikle “Kısmen” sınıfında sağladığı yüksek doğruluk ve ROC AUC değeri (0,7511) ile en güçlü model olarak dikkat çekerken; Karar Ağacı modeli ise sınıflar arasında daha dengeli bir dağılım sergileyerek yorumlanabilirliği yüksek bir alternatif sunmuştur. Gaussian Naive Bayes modeli ise hem eğitim hem test aşamasında daha düşük doğruluk ve ayırt edicilik oranları ile sınırlı kalmış; özellikle bağımsızlık varsayımının bu veri setine tam olarak uymadığı anlaşılmıştır.

Bulgular ayrıca, veri setinde sınıflar arası dengesizlik sorununun model performansını doğrudan etkilediğini açıkça göstermektedir. Bu nedenle, ilerleyen çalışmalarda özellikle “Tamamen” sınıfına ait örneklerin sayısal olarak artırılması, bu sınıfa özgü yeni nitelikler eklenmesi veya veri artırma (data augmentation) tekniklerinin uygulanması, modellerin genel başarısını artıracaktır.

Son olarak, elde edilen modellerin pratikte kullanımı açısından bakıldığında, yangın yönetimi, erken uyarı sistemleri ve müdahale stratejileri açısından anlamlı destekler sunduğu görülmektedir. Ancak bu modellerin gerçek zamanlı karar destek sistemlerine entegre edilebilmesi için hem çevresel faktörlerin (örneğin meteorolojik veriler) modele dâhil edilmesi hem de itfaiye müdahale sürelerini etkileyen coğrafi faktörlerin detaylı olarak modellenmesi gerekmektedir.

6. SONUÇ VE TARTIŞMA

Bu tez çalışmasında, bir ile ait yangın verileri kullanılarak, yangınların şiddet düzeylerini tahmin edebilecek makine öğrenmesi modelleri geliştirilmiştir. Çalışmanın temel amacı, yangın olaylarını mekânsal ve zamansal boyutlarıyla inceleyerek, denetimli öğrenme algoritmaları aracılığıyla yangının etkisinin sınıflandırılmasına olanak sağlayan bir tahmin sistemi oluşturmak olmuştur. Modelleme sürecinde, doğruluk oranı, karışıklık matrisi, ROC eğrisi, AUC değeri ve F1-skoru gibi çok boyutlu performans ölçütleri üzerinden üç farklı makine öğrenmesi algoritması (Karar Ağacı, Linear Support Vector Machine ve Gaussian Naive Bayes) karşılaştırmalı olarak değerlendirilmiştir.

Veri seti, 29135 yangın kaydından oluşmakta ve 15 özellik içermektedir. Bu özellikler arasında yangının türü, çıkış nedeni, konum bilgileri, hasar durumu ve zaman etiketleri gibi çeşitli parametreler yer almaktadır. Özellik seçimi sürecinde Ki-Kare testi uygulanarak, model başarısına en çok katkı sağlayan 9 değişken belirlenmiş, böylece modelin karmaşıklığı azaltılmış ve genelleme kapasitesi artırılmıştır. Eğitim ve test verisi oranı %80 - %20 olarak belirlenmiş, model değerlendirme süreçlerinde 5 katlı çapraz doğrulama yöntemi kullanılmıştır. Bu yapı sayesinde modellerin eğitim verisine aşırı uyum göstermesi engellenmiş, genellenebilirlikleri daha doğru şekilde ölçülmüştür.

Elde edilen bulgular, tüm modellerin özellikle “Kısmen” sınıfında yüksek doğruluk oranlarıyla başarılı sonuçlar ürettiğini ortaya koymuştur. Bu sınıf, hem sayıca veri setinde baskın olması hem de daha net ayırt edici özelliklere sahip olması nedeniyle modeller tarafından doğru sınıflandırılabilmiştir. En yüksek doğruluk oranı %61,4 ile Linear SVM modeli tarafından sağlanmış, Karar Ağacı modeli ise %61,2 oranıyla onu takip etmiştir. Gaussian Naive Bayes modeli, %58,5 doğruluk oranı ile diğer modellere kıyasla daha düşük bir başarı sergilemiştir. Ancak bu doğruluk değerlerinin birbirine oldukça yakın olması, daha ayrıntılı sınıf bazlı değerlendirmelerin yapılmasını gerekli kılmıştır.

Karışıklık matrisleri ve sınıf bazlı değerlendirmeler, modellerin “Başlangıç” sınıfında orta düzeyde başarı elde ettiğini, “Tamamen” sınıfında ise tüm modellerin sınıflandırma hatalarına açık olduğunu göstermiştir. Özellikle “Tamamen” sınıfına ait örneklerin büyük çoğunluğu yanlışlıkla “Kısmen” sınıfına tahmin edilmiştir. Bu durum, hem bu sınıfa ait örneklerin yetersizliğinden hem de özellik uzayında “Tamamen” sınıfının diğer sınıflarla örtüşmesinden kaynaklanmaktadır. ROC eğrileri üzerinden yapılan analizler de bu sonucu desteklemektedir. Linear SVM modelinin AUC değeri 0,7511 ile en yüksek ayırt etme kapasitesini gösterirken, Karar Ağacı modeli 0,7404, Gaussian Naive Bayes modeli ise 0,7234 AUC değeri üretmiştir. Özellikle “Kısmen” sınıfında modellerin ROC eğrileri ideal forma oldukça yakın çizilmiştir; buna karşın “Tamamen” sınıfında eğriler, 45 dereceye yakın eğim göstererek ayırım gücünün zayıf olduğunu ortaya koymuştur.

Çalışma bulguları, yalnızca teknik başarıya odaklanmakla kalmamış; aynı zamanda yangın müdahale süreçlerine yönelik uygulamalı çıkarımlar da sunmuştur. Geliştirilen modeller, yangının şiddet derecesini önceden tahmin edebilme yetenekleri sayesinde, risk bazlı müdahale planlarının oluşturulmasına, kaynakların öncelikli bölgelere dağıtılmasına ve müdahale sürelerinin kısaltılmasına katkı sağlayabilecek potansiyele sahiptir. Özellikle yüksek doğrulukla tahmin edilebilen “Kısmen” sınıfı, yangının erken evrelerinde alınacak kararların belirlenmesinde etkili olabilir. Karar Ağacı modelinin yorumlanabilirliği yüksek yapısı, sahada çalışan ekiplerin model çıktılarını daha kolay anlamasına ve güvenle kullanmasına olanak tanır. SVM modelinin doğrusal ayırım yeteneği ve yüksek ROC başarımı, teknik olarak güçlü bir sınıflandırma altyapısı sunduğunu göstermektedir.

Bununla birlikte, çalışmanın bazı sınırlılıkları mevcuttur. Veri setindeki “Tamamen” sınıfına ait örneklerin azlığı, bu sınıfın model tarafından yeterince öğrenilememesine yol açmıştır. Bu durum, hem sınıf dengesizliği hem de eğitim sırasında modelin bu sınıfla ilgili varyasyonları yeterince öğrenememesiyle ilgilidir. Ayrıca veri setinde meteorolojik değişkenlerin (sıcaklık, rüzgâr hızı, nem oranı vb.) yer almaması, yangının dinamik özelliklerinin modele tam anlamıyla yansıtılamamasına neden olmuştur. Bu eksiklikler, özellikle yangının hızlı yayılmasına neden olabilecek çevresel faktörlerin dışarıda bırakılması anlamına gelmekte ve model başarısını sınırlamaktadır.

Gelecekteki çalışmalarda, veri seti “Tamamen” sınıfı açısından zenginleştirilmeli, gerekirse veri üretme teknikleri (SMOTE, oversampling gibi) uygulanarak sınıf dengesi sağlanmalıdır. Ayrıca meteorolojik ve coğrafi bilgilerin (örneğin bitki örtüsü, eğim, hava durumu) modele entegre edilmesi, sınıflar arası ayırt ediciliği artıracaktır. Bunun yanı sıra, bu modellerin farklı şehirlerde uygulanması ve genellenebilirliklerinin test edilmesi, hem modelin doğruluğunu hem de operasyonel kullanım potansiyelini güçlendirecektir. Derin öğrenme algoritmaları, ensemble modeller ve zaman serisi tahmin sistemleri ile daha dinamik yapılar kurulabilir; yangın riski haritaları güncel ve anlık olarak üretilebilir.

Sonuç olarak, bu çalışma, kullanılan yangın verilerinden hareketle geliştirilen makine öğrenmesi tabanlı sınıflandırma modelleriyle yangın yönetimi süreçlerine stratejik destek sağlamayı hedeflemiştir. Modelleme süreçlerinde elde edilen doğruluk ve sınıf bazlı değerlendirme sonuçları, bu yaklaşımın hem bilimsel hem de uygulamalı değerini ortaya koymuştur. Elde edilen modeller, karar vericilerin yangınlara karşı daha hızlı ve doğru müdahale stratejileri geliştirmelerine olanak tanıyacak; böylece makine öğrenmesi yöntemlerinin afet yönetimi süreçlerinde kullanılabilirliği konusunda önemli bir katkı sağlayacaktır.



KAYNAKLAR

- Abid, F. (2021). A Survey of Machine Learning Algorithms Based Forest Fires Prediction and Detection Systems. *Fire Technology*, 559–590. doi:<https://doi.org/10.1007/s10694-020-01056-z>
- Adalarasan, R., & Sundaram, A. S. (2015). Parameter design in friction welding of Al/SiC/Al₂O₃ composite using grey theory based principal component analysis (GT-PCA). *Journal of the Brazilian Society of Mechanical Sciences and Engineering*, 1515-1528. doi: <https://doi.org/10.1016/j.jestch.2019.02.010>
- Akçetin, E., & Çelik, U. (2014). İstenmeyen Elektronik Posta (Spam) Tespitinde Karar Ağacı Algoritmalarının Performans Kıyaslaması. *İnternet Uygulamaları ve Yönetimi Dergisi*, 43-56. doi:<https://doi.org/10.5505/iuyd.2014.43531>
- Akın, E., & Şahin, M. E. (2023, Kasım 6). Derin Öğrenme ve Yapay Sinir Ağı Modelleri Üzerine Bir İnceleme. *Araştırma Makalesi*, 14(1). Rize: EMO Bilimsel Dergi.
- Aksoy, B., Korucu, K., Çalışkan, Ö., Osmanbey, Ş., & Halis, H. D. (2021). İnsansız Hava Aracı ile Görüntü İşleme ve Yapay Zekâ Teknikleri Kullanılarak Yangın Tespiti: Örnek Bir Uygulama. *Düzce University Journal of Science and Technology*, 112-122. doi:<https://doi.org/10.29130/dubited.1016195>
- Alqaraleh, S. (2021). Efficient Turkish Text Classification Approach for Crisis Management Systems. 718-731. doi:<https://doi.org/10.35378/gujs.715296>
- Altıparmak, Ö., & Sezgin, İ. (2023). *Birleşmiş Milletler Afet Risk Azaltma Ofisi Afet Terminolojisi*. Hukuk, Doğa ve Toplum Vakfı.
- Arpacı, S., & Kalıpsız, O. (2018). Yazılım Hata Sınıflandırmasında Farklı Naive Bayes Teknikleri Kıyaslama. *ÖHÜ Müh. Bilim. Derg.*, 1-13. doi:[doi:10.28948/ngumuh.383709](https://doi.org/10.28948/ngumuh.383709)
- Atalay, M., & Çelik, E. (2017). Büyük Veri Analizinde Yapay Zekâ ve Makine Öğrenmesi Uygulamaları. *Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 155-172. doi:<https://doi.org/10.20875/makusobed.309727>
- Aydın, C. (2018). Makine Öğrenmesi Algoritmaları Kullanılarak İtfaiye İstasyonu İhtiyacının Sınıflandırılması. *Avrupa Bilim ve Teknoloji Dergisi*, 169-175. doi:<https://doi.org/10.31590/ejosat.458613>
- Ayhan, S., & Erdoğan, Ş. (2014). Destek Vektör Makineleriyle Sınıflandırma Problemlerinin Çözümü İçin Çekirdek Fonksiyonu Seçimi. *Eskişehir Osmangazi Üniversitesi İİBF Dergisi*, 9(1), 175-198.

- Aylak, B., Oral, O., & Yazıcı, K. (2020). Yapay Zeka ve Makine Öğrenmesi Tekniklerinin Lojistik Sektöründe Kullanımı. *El-Cezerî Fen ve Mühendislik Dergisi*, 8, 74-93. doi:<https://doi.org/10.31202/ecjse.776314>
- Aytaç, E. (2021). Havzaların Benzerliklerini Tanımlamada Alternatif Bir Yaklaşım: Hiyerarşik Kümeleme Yöntemi Uygulaması. *Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi*, 958-970. doi:<https://doi.org/10.35414/akufemubid.870649>
- Balaban, M. E., & Kartal, E. (2015). *Veri Madenciliği ve Makine Öğrenmesi*. İstanbul: Çağlayan Kitabevi.
- Balcı, Y., & Altundağ, H. (2023). Büyükşehirlerde Yangın Risk Değerlendirmesi: Sakarya İli Örneği. *Dirençlilik Dergisi*, 7(1), 143-152. doi:<https://doi.org/10.32569/resilience.1248160>
- Başakın, E. E., Ekmekçioglu, Ö., & Özger, M. (2019). Makine Öğrenmesi Yöntemleri ile Kuraklık Analizi. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, 985 - 991. doi:[doi:10.5505/pajes.2019.34392](https://doi.org/10.5505/pajes.2019.34392)
- Bates, S., Hastie, T., & Tibshirani, R. (2023). Cross-Validation: What Does It Estimate and How Well Does It Do It? *Arxiv*, 1434-1445. doi:<https://doi.org/10.1080/01621459.2023.2197686>
- Beşli, N., & Tenekeci, E. (2020). Uydu Verilerinden Karar Ağaçları Kullanarak Orman Yangını Tahmini. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 899-906. doi:<https://doi.org/10.24012/dumf.661925>
- Bhavsar, H., & Ganatra, A. (2012). A Comparative Study of Training Algorithms for Supervised Machine Learning. *IJSCE*, 74-81.
- Biçer, E., Mirtagioğlu, H., Demir, C., Keskin, S., & Demir, Y. (2024). Jackknife ve Bootstrap Yöntemlerine İlişkin Bir Uygulama. *Journal of Statistics and Applied Sciences*, 45-56. doi:<https://doi.org/10.52693/jsas.1441998>
- Bilgin, M. (2017). Gerçek Veri Setlerinde Klasik Makine Öğrenmesi Yöntemlerinin Performans Analizi. *Akademik Bilişim Konferansı*. Aksaray.
- Birant, D. (2017). Hierarchical Clustering Ensemble with Different Linkage Methods. *International Journal of Computer Science and Information Security*. doi:[10.15317/Scitech.2019.189](https://doi.org/10.15317/Scitech.2019.189)
- Birba, D. (2020). A Comparative study of data splitting algorithms for machine learning model selection. *Degree project in computer science and engineering*. Digitala Vetenskapliga Arkivet. urn:nbn:se:kth:diva-287194 adresinden alındı
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.
- Bogdal, C., Schellenberg, R., Höpli, O., Bovens, M., & Lory, M. (2022). Recognition of gasoline in fire debris using machine learning: Part I, application of random forest, gradient boosting, support vector machine, and naïve bayes. *Forensic Science International*. doi:<https://doi.org/10.1016/j.forsciint.2021.111146>

- Bröker, F., Holt, L. L., Roads, B. D., Dayan, P., & Love, B. C. (2024). Demystifying Unsupervised Learning: How It Helps and Hurts. *Trends in Cognitive Sciences*, 974-986.
- Bulut, F. (2023). Makine Öğrenmesi Doğruluk Metrikleri: Modelinizin Performansını Anlamanın Anahtarı. *Topluluk Yayını*.
- Burges, C. J. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery 2*, 121-167.
- Çetin, O., Temurtaş, F., & Gülgönül, Ş. (2015). An application of multilayer neural network on hepatitis disease diagnosis using approximations of sigmoid activation function. *Dicle Tıp Dergisi*, 150-157., doi:https://doi.org/10.5798/diclemedj.0921.2015.02.0550
- Çolak, B., Durdağ, Z., & Erdoğan, P. (2016). K-Means Algoritması İle Otomatik Kümeleme. *El-Cezerî Fen ve Mühendislik Dergisi*, 315-323. doi:https://doi.org/10.31202/ecjse.264195
- Dietterich, T. G. (1994). Machine Learning. *Lecture Notes in Computer Science*. Department of Computer Science Oregon State University.
- Dike, H. U., Zhou, Y., Deveerasetty, K. K., & Wu, Q. (2018). Unsupervised Learning Based On Artificial Neural Network: A Review. *IEEE*. doi:10.1109/CBS.2018.8612259
- Dolu, D., & Şen, A. (2024). Coğrafi Referanslı Sosyal Medya Verilerinden Doğal Dil İşleme ve Derin Öğrenme Kullanılarak Duygu Analizi: Kahramanmaraş Depremleri Örneği. *Jeodezi ve Jeoinformasyon Dergisi*, 51-67. doi:https://doi.org/10.9733/JGG.2024R0005.T
- Duman, S. (2006). Markov Zincirlerinde Bootstrap. *Yüksek Lisans Tezi*. Ankara: Ankara Üniversitesi Fen Bilimleri Enstitüsü.
- Dwork, C., Feldman, V., Hardt, M., Pitassi, T., Reingold, O., & Roth, A. (2015). The reusable holdout: Preserving validity in adaptive data analysis. *Science*, 636-638. doi:10.1126/science.aaa937
- Efron, B. (1992). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 569-593.
- Eker, R., Alkış, K. C., Uçar, Z., & Aydın, A. (2023). Ormancılıkta Makine Öğrenmesi Kullanımı. *Türkiye Ormancılık Dergisi*, 150-177. doi:https://doi.org/10.18182/tjf.1282768
- Eryılmaz, E. E., Şahin, D. Ö., & Kılıç, E. (2020). Türkçe İstenmeyen E-postaların Farklı Öznetelik Seçim Yöntemleri Kullanılarak Makine Öğrenmesi Algoritmaları ile Tespit Edilmesi. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 13(2), 57-77.
- Fidanboy, M., Adar, N., & Okyay, S. (2022). Derin öğrenmeye dayalı orman yangını tahmin modeli geliştirilmesi ve Türkiye yangın risk haritasının oluşturulması. *Ormancılık Araştırma Dergisi*, 206-218. doi:https://doi.org/10.17568/ogmoad

- Foot, K. (2021). A Brief History of Machine Learning. *Dataversity*.
- Forero, P. A., Cano, A., & Giannakis, G. B. (2010). Consensus-based distributed linear support vector machines. *Journal of Machine Learning Research*, 35-46. doi:<https://doi.org/10.1145/1791212.179121>
- Frank, E., & Bouckaert, R. (2006). Naive Bayes for Text Classification With Unbalanced Classes. In *European Conference on Principles of Data Mining and Knowledge Discovery* (s. 503-510). Berlin: Springer.
- Ghosh, S., Dasgupta, A., & Swetapadma, A. (2019). A Study on Support Vector Machine based Linear and Non-Linear Pattern Classification. *IEEE*, 24-28. doi:10.1109/ISS1.2019.8908018
- Goh, Y. M., & Ubeynarayana, C. (2017). Construction accident narrative classification: An evaluation of text mining techniques. *Accident Analysis & Prevention*, 122-130. doi:<https://doi.org/10.1016/j.aap.2017.08.026>
- Gökalp, Ö. (2022). Makine Öğrenmesi. *Yüksek Lisans Tezi*. Gazi Üniversitesi, Gazi BilişimEnstitüsü, Adli Bilişim Bölümü.
- Gökkaya, E., & Kaya, A. (2021). Türkiye'de İtfaiyecilik Eğitimi İle Hava Kurtarma ve Yangın Mücadele Biriminin Yapısal Değerlendirmesi. *Hastane Öncesi Dergisi*, 6(1), 143-158.
- Guerrero, J. S., Olivas, J. A., Romero, F. P., & Viedma, E. H. (2015). Sentiment analysis: A review and comparative analysis of web. *Information Sciences*, 18-38.
- Gümüştas, E., & Pehlivanlı, A. (2021). In-Silico Mutajenisite Tahmininde İstatistiksel Öğrenme Modeli. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 365-370. doi:<https://doi.org/10.19113/sdufenbed.867067>
- Güney, H., & Öztoprak, H. (2018). The Impact of Under-sampling on the Performance of Bootstrap-based Ensemble Feature Selection. *IEEE*. doi:10.1109/SIU.2018.8404342
- Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Schölkopf, B. (1998). Support vector machines. *IEEE*, 18-28. doi:10.1109/5254.708428
- Hu, H., Liu, J., Zhang, X., & Fang, M. (2023). An Effective and Adaptable K-means Algorithm for Big Data Cluster Analysis. *Pattern Recognition*. doi:<https://doi.org/10.1016/j.patcog.2023.109404>
- IBM. (2020). *Makine Öğrenmesi*. Şubat 2025 tarihinde <https://www.ibm.com/>: <https://www.ibm.com/tr-tr/cloud/learn/machine-learning> adresinden alındı
- Jamil, M. A., & Khanam, S. (2023). Influence of One-Way ANOVA and Kruskal–Wallis Based Feature Ranking on the Performance of ML Classifiers for Bearing Fault Diagnosis. *Journal of Vibration Engineering & Technologies*, 3101-3132.

- Kaftandjieva, F. (2010). *Methods for Setting Cut Scores in Criterion-Referenced Achievement Tests: A Comparative Analysis of Six Recent Methods with an Application to Tests of Reading in EFL*. European Association for Language Testing and Assessment (EALTA).
- Karabulut, A., Baran, M., & Eraslan, E. (2024). Prevention of Occupational Accidents and Occupational Diseases with Supervised Machine Learning Algorithms: Different Sector Applications. *Journal of Turkish Operations Management*, 39-59. doi:<https://doi.org/10.56554/jtom.1245965>
- Kavlak, M. Ö., Kurtipek, A., & Çabuk, S. N. (2020). Coğrafi Bilgi Sistemleri İle Orman Yangını Risk Haritası Oluşturulması: Ören Örneği. *Resilience*, 4(1), 33-54. doi:<https://doi.org/10.32569/resilience.597887>
- Kavzoğlu, T., & Çölkesen, İ. (2010). Karar Ağaçları İle Uydu Görüntülerinin Sınıflandırılması: Kocaeli Örneği. *Harita Teknolojileri Elektronik Dergisi*, 36-45.
- Kavzoğlu, T., Çölkesen, İ., & Şahin, E. (2012). Heyelan Duyarlılık Haritasının Üretilmesinde Kullanılan Faktörlerin Etkilerinin Araştırılması: Düzköy Örneği. *Uzaktan Algılama ve Coğrafi Bilgi Sistemleri Sempozyumu*. Zonguldak.
- Kaya, Ç., & Yıldız, O. (2014). Makine Öğrenmesi Teknikleriyle Saldırı Tespiti: Karşılaştırmalı Analiz. *Marmara Fen Bilimleri Dergisi*, 89-104. doi:[10.7240/mufbed.24684](https://doi.org/10.7240/mufbed.24684)
- Kaya, H. (2018). Türkiye’de Yangınlar (1923-1960). *Uluslararası Tarih Araştırmaları Dergisi*, 28-29.
- Kecman, V. (2005). Support Vector Machines – An Introduction. L. Wang içinde, *Support Vector Machines: Theory and Applications* (s. 1-44). Berlin: Springer.
- Keçeoğlu, Ç. R., Gelbal, S., & Doğan, N. (2016). ROC Eğrisi Yöntemi ile Kesme Puanının Belirlenmesi. *International Journal of Social Science*, 553-562.
- Kelleher, J. D., Namee, B. M., & D'Arcy, A. (2020). *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*. The MIT Press.
- Mishra, M., Guria, R., Baraj, B., Nanda, A. P., Santos, C. A., Silva, R. M., & Laksono, F. A. (2024). Spatial Analysis and Machine Learning Prediction of Forest Fire Susceptibility: A Comprehensive Approach for Effective Management and Mitigation. *Science of the Total Environment*. doi:<https://doi.org/10.1016/j.scitotenv.2024.171713>
- Mockenhaupt, A. (2021). *Digitalisierung und Künstliche Intelligenz in der Produktion*. Wiesbaden: Springer Vieweg.
- MOE. (tarih yok). Molecular Operational Environment. Canada: Chemical Computing Group Inc.

- Moore, A. W., & Lee, M. S. (1994). Efficient Algorithms for Minimizing Cross Validation Error. *Proceedings of the Eleventh International Conference* (s. 190-198). San Francisco: Machine Learning. doi:https://doi.org/10.1016/B978-1-55860-335-6.50031-3
- Nagy, A., Szabó, A., Elbeltagi, A., Nxumalo, G. S., Bódi, E. B., & Tamás, J. (2024). Hyperspectral indices data fusion-based machine learning enhanced by MRMR algorithm for estimating maize chlorophyll content. *Frontiers in Plant Science*. doi: https://doi.org/10.3389/fpls.2024.1419316
- Ontivero-Ortega, M., Lage-Castellanos, A., Valente, G., Goebel, R., & Valdes-Sosa, M. (2017). Fast Gaussian Naïve Bayes for searchlight classification analysis. *NeuroImage*, 471-479. doi:https://doi.org/10.1016/j.neuroimage.2017.09.001
- Özçelik, O. (2023). Bootstrap Fourier AARDL Yöntemiyle J Eğrisi Hipotezinin Türkiye'nin Balkanlarla Dış Ticaretinde Geçerliliğinin Analizi. *İzmir İktisat Dergisi*, 969-991. doi:https://doi.org/10.24988/ije.1241877
- Özdemir, Ş., & Örsülü, S. (2019). Makine Öğrenmesinde Yeni Bir Bakış Açısı: Otomatik Makine Öğrenmesi (AutoML). *Bilişim Sistemleri ve Yönetim Araştırmaları Dergisi*, 23-29.
- Özkan, E. K., & Ulaş, H. B. (2023). Türkiye Metal Sektöründe Yaşanan İş Kazalarının Rassal Orman Algoritmasıyla Tahminlenmesi. *Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 1983-1997. doi:10.21597/jist.1285239
- Özşahin, M., Yılmaz, K. B., & Akcan, S. (2022). Makine Öğrenmesi Algoritmaları İle Ev Kirası Tahmini. C. Özalp, & N. Akçakale içinde, *Mühendislikte Araştırma ve Değerlendirmeler II* (s. 23-36). Gece Kitaplığı.
- Padilla, M., & Vega-Garcia, C. (2011). On the comparative importance of fire danger rating indices and their integration with spatial and temporal variables for predicting daily human-caused fire occurrences in Spain. *International Journal of Wildland Fire*, 46-58. doi:10.1071/WF09139
- Pang, Y., Li, Y., Feng, Z., Feng, Z., Zhao, Z., Chen, S., & Zhang, H. (2022). Forest Fire Occurrence Prediction in China Based on Machine Learning Methods. *Remote Sensing*. doi:https://doi.org/10.3390/rs14215546
- Pekşen, M. (2024, Mayıs). Elektrik Pano Yangınlarına Karşı Nesnelerin İnterneti Destekli Algılama Sisteminin Geliştirilmesi. *Yüksek Lisans Tezi*.
- Pekşen, M. F., Uyaroğlu, Y., & Soyhan, H. S. (2020). İstanbul, Ankara ve Sakarya İllerinin 2018 Yılı İtfaiye Olaylarının Karşılaştırılması. *Fuels, Fire and Combustion in Engineering Journal*, 1-19.
- Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1226-1238.
- Pérez, A., Larrañaga, P., & Inza, I. (2006). Supervised classification with conditional Gaussian networks: Increasing the structure complexity from naive Bayes.

- International Journal of Approximate Reasoning*, 1-25.
doi:<https://doi.org/10.1016/j.ijar.2006.01.002>
- Pham, B. T., Jaafari, A., Avand, M., Al-Ansari, N., Du, T. D., Yen, H. P., . . . Tuyen, T. T. (2020). Performance Evaluation of Machine Learning Methods for Forest Fire Modeling and Prediction. *Symmetry*, 1022.
doi:<https://doi.org/10.3390/sym12061022>
- Polat, K., & Güneş, S. (2006). Hepatitis Disease Diagnosis Using A New Hybrid System Based On Feature Selection (FS) and Artificial Immune Recognition System With Fuzzy Resource Allocation. *Elsevier*, 889-901.
doi:<https://doi.org/10.1016/j.dsp.2006.07.005>
- Quinlan, J. R. (2014). *C4.5: Programs for Machine Learning*. San Mateo, California, ABD: Morgan Kaufmann Publishers.
- Ramírez-Gallego, S., Lastra, I., Martínez-Rego, D., Bolón-Canedo, V., Benítez, J. M., Herrera, F., & Alonso-Betanzos, A. (2016). Fast-mRMR: Fast Minimum Redundancy Maximum Relevance Algorithm for High-Dimensional Big Data. *Wiley*, 134-152. doi: <https://doi.org/10.1002/int.21833>
- Rao, R. R., Eisenberg, J., & Schmitt, T. (2007). *Improving Disaster Management: The Pole of IT in Mitigation, Preparedness, Response and Recovery*. USA: The National Academies Press.
- Raschka, S., Liu, Y. (., & Mirjalili, V. (2022). *Machine Learning with PyTorch and Scikit-Learn: Develop Machine Learning and Deep Learning Models with Python*. Packt Publishing.
- Rashidi, H. H., Tran, N. K., Betts, E., Howell, L. P., & Green, R. (2019). Artificial Intelligence and Machine Learning in Pathology: The Present Landscape of Supervised Methods. *Academic pathology*.
doi:<https://doi.org/10.1177/2374289519873088>
- Samuel, A. (1959). Some Studies in Machine Learning Using the Game of Checkers. *BM Araştırma ve Geliştirme Dergisi*. doi:10.1147/rd.33.0210
- SAP. (2021). *Was ist Maschinelles Lernen?* Şubat 2024 tarihinde <https://www.sap.com/germany/insights/what-is-machine-learning.html>:
<https://www.sap.com/> adresinden alındı
- Satılmış, H., & Akleylek, S. (2021). IoT Güvenliği İçin Kullanılan Makine Öğrenimi ve Derin Öğrenme Modelleri Üzerine Bir Derleme. *Journal of Information Technologies*, 457-481.
- Shahzad, F., Mehmood, K., Hussain, K., Haidar, I., Anees, S. A., Muhammad, S., . . . Feng, Z. (2024). Comparing Machine Learning Algorithms to Predict Vegetation Fire Detections in Pakistan. *Fire Ecology*.
- Sharma, S., & Khanal, P. (2024). Forest Fire Prediction: A Spatial Machine Learning and Neural Network Approach. *Fire*. doi:<https://doi.org/10.3390/fire7060205>

- Smith, D. L. (2011). Firefighter fitness: improving performance and preventing injuries and fatalities. *Current Sports Medicine Reports*, 167-172. doi:10.1249/JSR.0b013e31821a9fec
- Sun, J., & Li, H. (2008). Data Mining Method for Listed Companies, Financial Distress Prediction. *Knowledge-Based Systems*, 1-5. doi:https://doi.org/10.1016/j.knosys.2006.11.003
- Suthaharan, S. (2016). Support Vector Machine. *Machine Learning Models and Algorithms for Big Data Classification*. içinde Boston: Integrated Series in Information Systems.
- Şahin, D. Ö., & Kılıç, E. (2019). Two new feature selection metrics for text classification. *Automatika*. doi:https://doi.org/10.1080/00051144.2019.1602293
- Şahin, F. (2022, Şubat). Yetişkinlerde yangın güvenliği bilincinin ve farkındalığının irdelenmesi (Bir fabrika çalışanları örneği). *Yüksek Lisans Tezi*.
- Şahinaslan, Ö., Dalyan, H., & Şahinaslan, E. (2022). Naive Bayes Sınıflandırıcısı Kullanılarak YouTube Verileri Üzerinden Çok Dilli Duygu Analizi. *Bilişim Teknolojileri Dergisi*, 15(2), 221-229. doi:https://doi.org/10.17671/gazibtd.999960
- Temurtas, F. (2009). A Comparative Study On Thyroid Disease Diagnosis Using Neural Networks. *Elsevier*, 944-949. doi:https://doi.org/10.1016/j.eswa.2007.10.010
- Tonkal, Ö., & Polat, H. (2021). Traffic Classification and Comparative Analysis with Machine Learning Algorithms in Software Defined Networks. *Gazi Üniversitesi Fen Bilimleri Dergisi Part C: Tasarım ve Teknoloji*, 71-83. doi:https://doi.org/10.29109/gujsc.869418
- Topuz, D. (2022). Regresyonda Yeniden Örneklem Yöntemlerinin Karşılaştırmalı Olarak İncelenmesi. *Yüksek Lisans Tezi*. Niğde: Niğde Üniversitesi Fen Bilimleri.
- Türkmenoğlu, C., & Tantuğ, A. C. (2014). Sentiment Analysis in Turkish Media. *In International Conference on Machine Learning*.
- Urul, A. K. (2017, Eylül). *Kaliteli Sunulması Zorunlu Olan Bir Hizmet: İtfaiye*. Ocak 08, 2025 tarihinde tepav: <https://www.tepav.org.tr/tr/yayin/s/1242> adresinden alındı
- Ülker, M., & Civalek, Ö. (2002). Yapay Sinir Ağları ile Eksenel Yüklü Kolonların Burkulma Analizi. *Turkish J. Eng. Env.*, 117-125.
- Vapnik, V. (1996). *The Nature of Statistical Learning Theory*. New York: Second Edition.
- Wu, S., & Banzhaf, W. (2010). The use of computational intelligence in intrusion detection systems: A review. *ScienceDirect*, 1-35. doi:https://doi.org/10.1016/j.asoc.2009.06.019

- Yadav, S., & Shukla, S. (2016). Analysis of k-Fold Cross-Validation over Hold-Out Validation on Colossal Datasets for Quality Classification. *IEEE*, 78-83. doi:10.1109/IACC.2016.25
- Yavuz, Ö., & Bozatay, Ş. A. (2015). Türkiye’de İtfaiye Hizmetlerinin Örgütlenmesi ve Etkinliği Sorunu. *Çankırı Karatekin Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 5(2), 785-804. doi:10.18074/cnuiibf.186
- Yıldırım, O., Günay, F., & Yağanoğlu, M. (2023). Makine Öğrenmesi Yöntemleriyle Orman Yangını Tahmini. *Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 1468-1481. doi:https://doi.org/10.21597/jist.1249908
- Zhang, H., Kang, Y.-L., Zhu, Y.-Y., Zhao, K.-X., Liang, J.-Y., Ding, L., . . . Zhang, J. (2017). Novel Naïve Bayes Classification Models For Predicting The Chemical Ames Mutagenicity. *Elsevier*, 56-63.
- Zhou, Z.-H. (2021). *Machine Learning*. Singapore: Springer Nature Singapore Pte Ltd.



ÖZGEÇMİŞ

Ad-Soyad : Zeynep Nazlı ASLAN

ÖĞRENİM DURUMU:

- **Lisans** : 2022, Sakarya Üniversitesi, Siyasal Bilgiler Fakültesi, Çalışma Ekonomisi
- **Yükseklisans** : 2025, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü, Afet Yönetimi Bölümü

TEZDEN TÜRETİLEN ESERLER:

- Zeynep Nazlı Aslan, Beytullah Eren, ‘Sakarya İlinde Yangın Vakalarının Mekânsal ve Zamansal Analizi’ Proceeding Book of 2nd International Conference on Recent and Innovative Results in Engineering and Technology ICRIRET 2025, 11-12 Şubat 2025, Konya, Turkey, s. 453-464

DİĞER ESERLER:

- Zeynep Nazlı Aslan, Beytullah Eren, ‘Yapay Zeka Destekli İşe Alım Süreçlerinin Cinsiyet Adaletini Sağlamadaki Etkilerinin Değerlendirilmesi’ Proceeding Book of 3rd International Conference on Engineering, Natural and Social Sciences ICENSOS 2024, 16 – 17 May 2024, Konya, Türkiye s. 859-877