

Atf İçin: Yıldırım, O., Günay, F. B. ve Yağanoğlu, M. (2023). Makine Öğrenmesi Yöntemleriyle Orman Yangını Tahmini. *İğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 13(3), 1468-1481.

To Cite: Yıldırım, O., Günay, F.B., & Yağanoğlu, M.. (2023). Forest Fire Prediction with Machine Learning Methods. *Journal of the Institute of Science and Technology*, 13(3), 1468-1481.

Makine Öğrenmesi Yöntemleriyle Orman Yangını Tahmini

Orhan YILDIRIM*, Faruk Baturalp GÜNAY, Mete YAĞANOĞLU

Öne Çıkanlar:

- Veri seti öznetelikleri PCA ve ANOVA yöntemleriyle seçilmiştir
- Orman yangınlarını sınıflandırmada rasgele orman iyi performans göstermiştir
- Bu çalışmada olası orman yangınına önceden tahmin ederek yangınlara hızlı müdahale edilebilecektir

Anahtar Kelimeler:

- Orman yangını
- Tahmin
- Makine öğrenmesi
- Öznetelik seçme
- Sınıflandırma

ÖZET:

Orman yangını, ormanda yaşama birliği içinde bulunan canlı ve cansız bütün varlıkları yakarak yok eden, ekonomik ve ekolojik zararları olan bir afettir. Son yıllarda küresel ısınma sebebi ile mevsim normalleri üzerinde seyreden sıcaklıklar ve kuraklıklar orman yangını riskini daha da artırmaktadır. Orman yangınları nedeniyle meydana gelen zararı en aza indirmek için yangınla mücadelede erken uyarı, hızlı ve etkin müdahale çok önemlidir. Makine öğrenmesi yöntemleri ise günümüzde erken uyarı sistemlerinde kullanılmaktadır. Bu çalışmada orman yangınlarıyla mücadele için olası orman yangınına önceden tahmin ederek yangınların kontrol edilmesi ve etkisinin azaltılması hedeflenmiştir. Orman yangını tahmin modeli için veri seti, NASA'nın Oak Ridge Ulusal Laboratuvarı (ORNL) Dağıtılmış Aktif Arşiv Merkezi'nin (DAAC) resmi web sitesinden alınarak geliştirilmiştir. Bu veriler makine öğrenmesi yöntemleriyle işlenerek orman yangını tahmin modeli oluşturulmuştur. Veri setine çeşitli ön işleme adımları uygulayarak sınıflandırma modeline uygun hale getirilmiştir. Öznetelik seçme teknikleri ile veri setinin tümü kullanılmadan en yüksek oranda veri bütünlüğü sağlanarak en az sayıda öznetelik alt kümesi seçilmiştir. Hedef değişkeni bulmada en önemli ve en faydalı öznetelikler seçilerek makine öğrenmesi algoritmalarından Destek Vektör Makinesi, Karar Ağacı, Rasgele Orman, Gradyan Artırma, K-En Yakın Komşu ve Naive Bayes olmak üzere 6 farklı sınıflandırma algoritmaları ile model oluşturulmuştur. Model performansını değerlendirmek için validasyon işlemi ve en iyi parametre seçimi için ise hiperparametre optimizasyonu yapılmıştır. Bu çalışmada kullanılan sınıflandırma algoritmaları arasında validasyon işlemi ile birlikte en başarılı iki algoritmadan Rasgele Orman ile %97 ve Naive Bayes ile %96 doğruluk oranı elde edilmiştir.

Forest Fire Prediction with Machine Learning Methods

ABSTRACT:

Forest fire is a disaster that destroys all living and non-living beings in the unity of life in the forest by burning and has economic and ecological damages. In recent years, temperatures and droughts that have been above the seasonal norms due to global warming have increased the risk of forest fires. In order to minimize the damage caused by forest fires, early warning, fast and effective intervention is very important in firefighting. Machine learning methods are used in early warning systems today. In this study, it is aimed to control and reduce the effects of fires by predicting possible forest fires in order to fight forest fires. The dataset for the wildfire prediction model was developed from the official website of NASA's Oak Ridge National Laboratory (ORNL) Center for Distributed Active Archives (DAAC). A forest fire prediction model was created by processing these data with machine learning methods. The data set was adapted to the classification model by applying various preprocessing steps. With the feature selection techniques, the least number of feature subsets were selected by providing the highest level of data integrity without using the entire data set. By choosing the most important and useful features in finding the target variable, a model was created with 6 different classification algorithms, namely Support Vector Machine, Decision Tree, Random Forest, Gradient Boosting, K-Nearest Neighbor and Naive Bayes. Validation process was performed to evaluate model performance and hyperparameter optimization was performed for best parameter selection. Among the classification algorithms used in this study, an accuracy rate of 97% was obtained with Random Forest and 96% with Naive Bayes, which is one of the two most successful algorithms with the validation process.

Highlights:

- Data set features were selected by PCA and ANOVA methods
- Random forest performed well in classifying forest fires
- In this study, it will be possible to respond quickly to fires by predicting a possible forest fire

Keywords:

- Forest fire
- Prediction
- Machine learning
- Feature selection
- Classification

¹Orhan YILDIRIM (Orcid ID: 0000-0003-3117-1147), Faruk Baturalp GÜNAY (Orcid ID: 0000-0001-5472-3608), Mete YAĞANOĞLU (Orcid ID: 0000-0003-3045-169X), Atatürk Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Erzurum, Türkiye

*Sorumlu Yazar/Corresponding Author: Orhan YILDIRIM, e-mail: orhanyildirim.com@gmail.com

GİRİŞ

Ormanlar, sadece ağaç ve ağaççık topluluklarından ibaret değildir. Ormanlar yaşamın her alanında canlıların en kıymetli ekolojik değerlerinden biridir. Yaşamımız için gerekli olan temiz hava, içilebilir su kaynakları, barınma ve yiyecek gibi temel yaşam koşullarının en önemli kaynağı ormanlardır. Doğal ve beşeri nedenlerle meydana gelen orman yangınları doğal afetlerin başında gelmekte olup birçok canlılığın bir arada yaşadığı ekosistemi yok ederek büyük bir çevresel tahribata yol açmaktadır. Bu afetler, diğer zararların yanı sıra en önemlisi insan hayatını tehdit etmektedir. Normalin üzerinde yüksek sıcaklıklar, düşük nispi nem ve şiddetli rüzgârlar ile birlikte orman yangınları zaman zaman bütün bir ülkenin hatta neredeyse tüm dünyanın birlikte mücadele ettiği afet boyutlarına ulaşmaktadır. Özellikle son yıllarda iklim değişikliğine bağlı küresel ısınmayla birlikte dünyada hızla artan orman yangınlarının yanı sıra artan nüfus, sanayileşme ve tarımsal gelişmeler ormanlık alanların hızla yok olmasına neden olmaktadır. Birleşmiş Milletler Gıda ve Tarım Örgütü'nün (FAO) "2020 Küresel Orman Kaynakları Değerlendirmesi Raporunda (FRA) dünyada 4,06 milyar hektar büyüklüğünde orman varlığı olduğunu ve bu ormanlık alanın dünya kara alanının üçte birine tekabül ettiğini açıklamıştır. Raporla başta orman yangınları olmak üzere çevre felaketleri, orman zararlıları gibi çeşitli nedenlerle dünya ormanlarının azaldığı ve 1990-2020 yılları arasında toplam 178 milyon hektar orman alanının tahrip edildiği belirtilmiştir (FAO,2020).

Tarım ve Orman Bakanlığına bağlı Orman Genel Müdürlüğünün resmi istatistik verilerinde Türkiye'de orman yangınlarının çıkış nedenlerine bakıldığında %29 ihmal ve dikkatsizlik, %4 kasıt, %6,4 kaza, %13 doğal ve %47,6 sebebi bilinmeyen nedenlerle çıkmaktadır (OGM, 2021). Orman ekosistemlerini tehdit eden faktörlerin ve orman yangınlarının nedenlerinin bilinmesi önleme faaliyetlerinin daha etkin bir şekilde yürütülmesini sağlayacaktır. Orman yangınlarını etkileyen ana faktörler arasında hava şartları (sıcaklık, nispi nem, yağış ve rüzgar hızı), zaman faktörü (mevsim, ay, günün belli saatleri), topografik yapı (bakı, yükseklik, arazi meyli, arazi şekli vb.) ve insan kaynaklı faktörler sayılmaktadır (Arif ve ark., 2021). Orman yangınlarının önlenmesi ve etkilerinin en aza indirilmesi için yangınla mücadelede erken uyarı, hızlı ve etkin müdahale çok önemlidir.

Literatüre bakıldığında orman yangınlarını tahmin etmek için birçok çalışma yapılmıştır. Sevinc ve ark. (2020), bayes ağ modelini kullanarak olası orman yangını nedenlerinin tahmin ve analizi yapılmaktadır. Bu model ile avcılık, anız yakma, piknik ve çoban ateşi gibi yangın çıkış nedenleri oranları bulunarak karşılaştırılmıştır. Diğer bir çalışmada (Arpaci ve ark., 2014), çevresel veri kümeleri ve geçmiş yangın verileri kullanılarak makine öğrenmesi algoritmalarından Maksimum Entropi (MaxEnt) ve Rastgele orman algoritması ile mekânsal yangın dağılımı tahmin edilmiştir. Çalışma sonucunda, mekânsal yangın dağılımını etkileyen iki faktörün nüfus yoğunluğu ve iklim olduğu görülmüştür. Fidanboy ve ark. (2022), Türkiye'nin çeşitli bölgelerinde 2013-2019 yılları arasında meydana gelen orman yangınlarından elde edilen veriler kullanarak derin öğrenme yöntemiyle Tensorflow ve sinir ağı olarak da Keras kütüphanesi ile orman yangını tahmin modeli geliştirmiş ve bölgesel yangın risk haritası oluşturmuşlardır. Liang ve ark. (2019) ise Kanada Alberta ormanlarındaki yangınların ölçeğini (yangın şiddetine göre 5 seviye) meteorolojik faktörleri girdi olarak tahmin etmek için geri yayılım sinir ağı (BPNN), tekrarlayan sinir ağı (RNN) ve uzun kısa süreli bellek (LSTM) yöntemlerini uygulamıştır. Çalışma sonucunda meteorolojik bilgileri kullanarak bir orman yangınının ölçeğini ve başlangıç aşamasında yangının tahminin mümkün olduğunu göstermiştir. Bir yangının ölçeği, yangının süresi ve yandığı alanın büyüklüğünün birleşimi ile belirlendiği ifade edilmiştir. Bayat ve Yıldız (2022), Portekiz'in bir milli parkında çıkan orman yangını sonucu elde edilen 512 orman

yangını verisinden sıcaklık, yağış, rüzgar ve bağıl nem gibi çeşitli öznelikleri kullanarak makine öğrenmesi algoritmalarıyla yanan alan boyutunu tahmin etmişlerdir.

Başka bir çalışmada ise Brezilya'nın Acre eyaletindeki hava ve coğrafi verileri dikkate alarak bulanık mantık ve makine öğrenmesi teknikleri ile orman yangını erken uyarı sistemi oluşturulmuştur. Çalışmada orman yangını uyarı haritası, orman yangını riski ve orman yangını tehlikesi olarak iki indeks bileştirilmiş, yangın risk indeksinin yangının belirli bir noktada çıkma olasılığını ölçtüğü, orman yangın riskinin ise bulanık KNN ile geçmiş verilerden elde edildiği belirtilmiştir (Silva ve ark., 2020). Orman yangınlarını tahmin ederek yangını önleme ve riskini ortadan kaldırmaya katkı sağlayacağı hususlarına değinen Shao ve ark. (2022), Çin için bir orman risk haritası oluşturmuşlardır. Makine öğrenmesi algoritmalarıyla kurulan modellerden elde edilen çeşitli metriklerin sonuçları karşılaştırılmış, Çin orman yangını risklerinde belirgin mevsimsel ve bölgesel farklılıklar olduğunu gösterilmiştir. Orman yangını tahmininin orman yangınlarını kontrol etmede en önemli bileşen olduğuna değinen Preeti ve ark. (2021), orman yangını oluşumunu tahmin etmek için sıcaklık, nem, rüzgar ve yağmur gibi meteorolojik parametreler kullanılarak makine öğrenmesi algoritmalarından rasgele orman regresyonu ile hiperparametre ayarlaması da yapılarak tahmin modeli geliştirmişlerdir. Spoorthy ve Kumar, (2022) yaptıkları çalışmada orman yangınlarının geniş bir alana yayılmadan hızlı önlem alabilmek için orman yangını alanını tespit etmişler ve makine öğrenmesi teknikleri karşılaştırılarak en iyi regresyon modeli oluşturmuşlardır. Mimboro ve ark. (2022), yaptıkları bir çalışmada veriler k-ortalama kümeleme ile kümelenecek verilere çok düşük risk, düşük risk, orta risk, yüksek risk, çok yüksek risk olmak üzere beş sınıf etiketiyle bir etiket yapılmıştır. Kümeleme sonucu elde edilen etiketlenmiş verileri kullanarak sınıflandırma işlemine tabi tutulmuş ve en iyi sınıflandırıcı rasgele orman algoritması olduğunu değerlendirmişlerdir. Rakshit ve ark. (2021) ise buna benzer bir çalışmalarında makine öğrenmesi sınıflandırma algoritmalarıyla yanan alan miktarına göre belirli bir alanın yangına eğilim düzeylerini oluşturmuşlar. Xie ve Peng (2018), topluluk öğrenme yaklaşımlarını kullanarak orman yangınlarının yanmış alanını ve büyük ölçekli orman yangınlarının oluşumu için tahmin modeli geliştirmişler ve ayarlanmış rastgele orman yaklaşımı yanmış alanı tahmin etmede diğer regresyon modellerine göre daha iyi performans gösterdiğini belirtmişlerdir. Castelli ve ark. (2015), meteorolojik ve orman ile ilgili verilerle yanmış alan miktarı arasındaki ilişkiler kullanılarak olası orman yangını esnasında yanacak alanı tahmin etmek için genetik programlamaya dayalı akıllı bir sistem geliştirmişlerdir. Yanan alanı tahmin etmek için yapılan diğer bir çalışmada ise yangını büyük veya küçük yangın olarak makine öğrenimi algoritmalarıyla sınıflandırılmış, yangın sırasında yangın yönetim ekibine uygun kaynak tahsisine yardımcı olması hedeflenmiştir (Niranjan ve ark., 2019). Orman yangınlarının çıkış nedenlerinden biri de yıldırım sebebiyle ortaya çıkan yangınlardır. Bu nedenle çıkan orman yangınlarını tahmini için Coughlan ve ark. 2020, yıldırım ile ateşleme arasındaki ilişkiyi analiz ederek makine öğrenimi yaklaşımıyla yıldırım tahminlerine ve çevresel koşullara dayalı orman yangını tahmin modeli oluşturmuşlardır. Pang ve ark. (2022), Çin'deki orman yangınlarının ana nedenlerini tahmin etmeye çalışmışlardır. Çalışmada makine öğrenmesi algoritmalarından yapay sinir ağı, radyal tabanlı fonksiyon ağı, destek vektör makinesi ve rastgele orman algoritmaları kullanılarak yangın çıkış nedenleri tahmin edilmiş ve yüksekten düşüğe doğru orman yangınına eğilimli alanların mekânsal dağılım haritaları çıkarılmıştır. Başka bir çalışmada yağışın, gelecekteki orman yangınlarının olasılığını etkileyen önemli bir faktör olduğunu ifade eden Chen ve ark. (2022), yürüttükleri bir çalışmada zamanla azalan yağış modeli ile Orta ve Kuzey Çin'de orman yangın riskini tahmin etmişlerdir. Çalışmada, kapsamlı yağış indeksini hesaplamak için zamanla azalan bir yağış algoritması kullanılmışlardır. Bu yöntem, orman yangınlarının oluşumunu tahmin etmede yağışın etkisinin daha iyi temsil edildiğini göstermiştir. Lai ve ark. (2022), büyük yangınların küçük yangınlardan daha az sayıda olduğu durumlarda dengesiz veri

dağılımı için yeni bir veri dengeleme prosedürü ve derin sinir ağı kullanarak orman yangını tahmini önermişlerdir. Önerilen yöntemde büyük ölçekli orman yangınlarının daha doğru şekilde tahmin edilebilir olduğuna, orman yangınlarının önceden yönetimine ve ciddi yangın kazalarının önlenmesine daha iyi fayda sağlayacağına değinmişlerdir.

Orman yangını tahmini, orman yangını yönetiminin önemli bir bileşenini oluşturmaktadır. Orman yangınlarının önceden tahmini ile yangına erken müdahale yapılarak ne kadar erken müdahale o kadar ormanın kurtarılması sağlanmaktadır. Olası orman yangınlarını önceden tahmin edebilmek, yangın öncesi önlemlerin alınabilmesi ve yangına müdahale edecek yangın yönetim ekibinin sağlıklı karar alabilmeleri gibi önemli faydalar sağlayabilmektedir. Böylece araç ve personel kaynağının yerinde ve zamanında kullanılması, yangına daha hızlı ve etkili müdahale edilmesi sağlanmaktadır. Bu çalışmanın amacı, orman yangınlarıyla mücadele için olası orman yangınına önceden tahmin ederek yangın çıkmadan önce veya henüz başlangıç aşamasında yangınların kontrol altına alınması ve etkilerinin en aza indirilmesidir.

Bu çalışmada önerilen yöntem diğer yöntemlerden farklı olarak yapay zekânın bir alt dalı olan makine öğrenmesi yöntemleriyle önceden sınıfı belli olan (yangın var/ yangın yok) veri seti işlenerek orman yangınları ikili sınıflandırma metodu ile tahmin edilmiştir. Literatürdeki diğer çalışmalarda genel olarak yanan alan büyüklüğünün tahmini, yangın risk haritalarının oluşturulması gibi regresyon modelleri oluşturulmuş veya veriler üzerinde belirli bir eşik değeri kullanılarak yeni bir hedef değişken oluşturularak veriler analiz edilmiştir. Önerilen yöntemin diğer çalışmalardan başka bir farkı ise boyut azaltma tekniklerinin kullanılmasıdır. Bu çalışmada orman yangını tahmin modeli için kullanılan veri seti, NASA'ya ait ORNL'nin DAAC resmi web sitesinden alınmıştır. Veriler, ön işleme adımlarından geçirilerek gereksiz verilerin silinmesi, eksik verilerin tamamlanması, normalizasyon gibi çeşitli işlemlere tabi tutulmuştur. Böylece model eğitilirken gereksiz girdi öznitelikleri eğitime girmediğinden işlem süresi ve işlem yükü önemli ölçüde azalmıştır. Ön işleme adımlarının tamamlanmasıyla birlikte veriler makine öğrenmesi yöntemleriyle işlenerek orman yangını tahmin modeli oluşturulmuştur. Veri setindeki sınıf etiketi ikili değerden (yangın var-1/ yangın yok-0) olduğundan ikili sınıflandırma işlemi yapılmıştır. Modele iyileştirme stratejileri uygulayarak performansı incelenmiştir. Öznitelik seçme teknikleri ile boyut azaltma yapılarak hedef değişkeni bulmada en önemli ve en faydalı öznitelikler belirlenerek sınıflandırma işlemi gerçekleştirilmiştir. Özellik seçme sonucu yeni veri kümesiyle model eğitilmiş ve doğruluk oranı tüm veri setiyle elde edilen orana yakın çıkmıştır. Bu çalışmada Makine öğrenmesi algoritmalarından DVM, KA, RO, GA, K-NN ve NB olmak üzere 6 farklı sınıflandırma algoritması kullanılarak model performansları karşılaştırılmıştır. İkili Sınıflandırma modellerini değerlendirmek için karışıklık matrisi, aşırı uyum (overfitting) sorununu gidermek için validasyon işlemi ve modele en iyi parametre seçimi için ise hiperparametre optimizasyonu uygulanmıştır.

MATERYAL VE METOT

Bu bölüm sınıflandırma işleminde kullanılan orman yangınları veri seti, veri ön işleme adımları ve sınıflandırma algoritmaları hakkında bilgiler içermektedir.

Veri Seti

Bu çalışmada kullanılan veri seti (veri kümesi), NASA'ya ait DAAC resmi web sitesinden alınmıştır (Walker ve ark., 2022). Bu veri seti, 1983-2016 yılları arasında Alaska ve Kanada'daki yanmış ve yanmamış alanlardaki arazilerden toplanan verilerden oluşmaktadır. 1171 örnek veriden oluşan bu veri seti, 1019 yanmış ve 152 de yanmamış olmak üzere iki sınıfa ayrılmıştır. Veri setinde her bir konuma ait bağıl nem, sıcaklık, rüzgar hızı, yağış, kuraklık kodları gibi yangın hava indekslerini de (FWI) içeren sistem bileşenleri, bakı, yükseklik, eğim gibi topoğrafik yapısı ve ağacın yaşı, ağaç

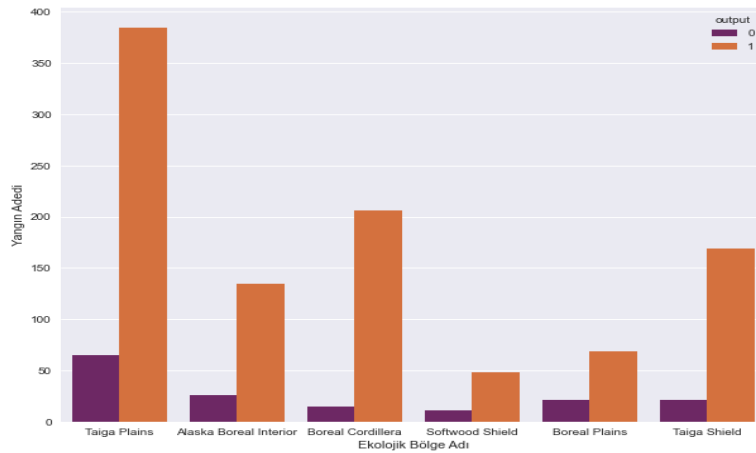
yoğunluğu gibi orman bilgisiyle birlikte 49 öznitelik bulunmaktadır. Veri kümesinde model eğitimi için gereksiz özellikler (proje id, ismi vb.) ve çok fazla boş değer bulunduran özellikler silinerek 35'e düşürülmüştür. Bu özniteliklerin 2'si kategorik ve 33'ü de sayısal veri içermektedir. Çalışmada kullanılan veri kümesinde sınıf değişkeni sıfır ile etiketlenmiş veri "yangın yok", bir ile etiketlenmiş veri ise "yangın var" sınıfını temsil etmektedir.

Aşağıdaki Şekil 1'de veri setinden örnek olarak 15 satır, 11 öznitelik ve bunlara ait değerler gösterilmiştir. Görüldüğü gibi veri kümesinde hem sayısal hem de kategorik veri bulunmaktadır. Makine öğrenmesi sayısal verilerle işlem yaptığından kategorik özniteliklere ait verilerin sayısal verilere dönüştürülmesi gerekmektedir.

	yükseklik	eğim	nem	yağış	sıcaklık	bağıl_nem	rüzgar_hızı	kuraklık_kodu	kuraklık_nem_kodu	yangın_havadurumu_indeksi	bölge
296	535.100213	2.000000	4.000000	1.910000e-11	20.22000	44.28000	8.64000	251.280000	49.340000	15.269000	Boreal Cordillera
716	214.775000	0.000000	3.000000	3.600000e-02	24.98000	37.99000	20.74500	315.510000	47.360000	32.195000	Taiga Plains
54	376.460085	3.208152	3.842397	6.985050e-01	22.25865	46.18538	11.58866	278.428543	46.910054	17.157185	Alaska Boreal Interior
400	535.100213	5.000000	3.000000	9.790000e-01	26.24000	40.88000	4.40000	296.780000	25.970000	5.059000	Boreal Cordillera
251	376.460085	3.208152	3.842397	8.240000e-03	30.18000	32.79000	5.13000	355.320000	88.992500	31.998000	Taiga Plains
144	376.460085	3.208152	3.842397	7.080000e-02	25.00000	47.11000	7.35000	203.930000	36.920000	10.053000	Boreal Cordillera
319	376.460085	3.208152	6.000000	2.760000e-03	13.54000	37.95000	9.62000	167.335000	22.120000	4.538000	Boreal Plains
111	458.500000	3.100000	3.842397	4.400000e-02	21.14000	71.09000	16.74000	304.460000	24.610000	10.262000	Boreal Cordillera
587	234.307000	0.000000	5.000000	8.790000e-01	24.90000	44.60000	4.93000	234.960000	82.330000	14.853000	Taiga Plains
687	394.506000	0.000000	2.000000	9.220000e-03	15.65000	37.53000	17.55000	306.500000	88.992500	28.859000	Taiga Shield
376	295.000000	2.000000	4.000000	2.190000e-02	17.62000	60.41000	12.58000	404.415000	21.430000	9.630000	Alaska Boreal Interior
1	376.460085	3.208152	6.000000	6.985050e-01	22.25865	46.18538	11.58866	278.428543	46.910054	17.157185	Taiga Plains
498	237.400000	0.000000	2.000000	6.985050e-01	22.25865	46.18538	11.58866	278.428543	46.910054	17.157185	Taiga Plains
539	196.800000	0.000000	5.000000	6.985050e-01	22.25865	46.18538	11.58866	278.428543	46.910054	17.157185	Taiga Plains
437	380.000000	3.208152	3.842397	6.985050e-01	22.25865	46.18538	11.58866	278.428543	46.910054	17.157185	Alaska Boreal Interior

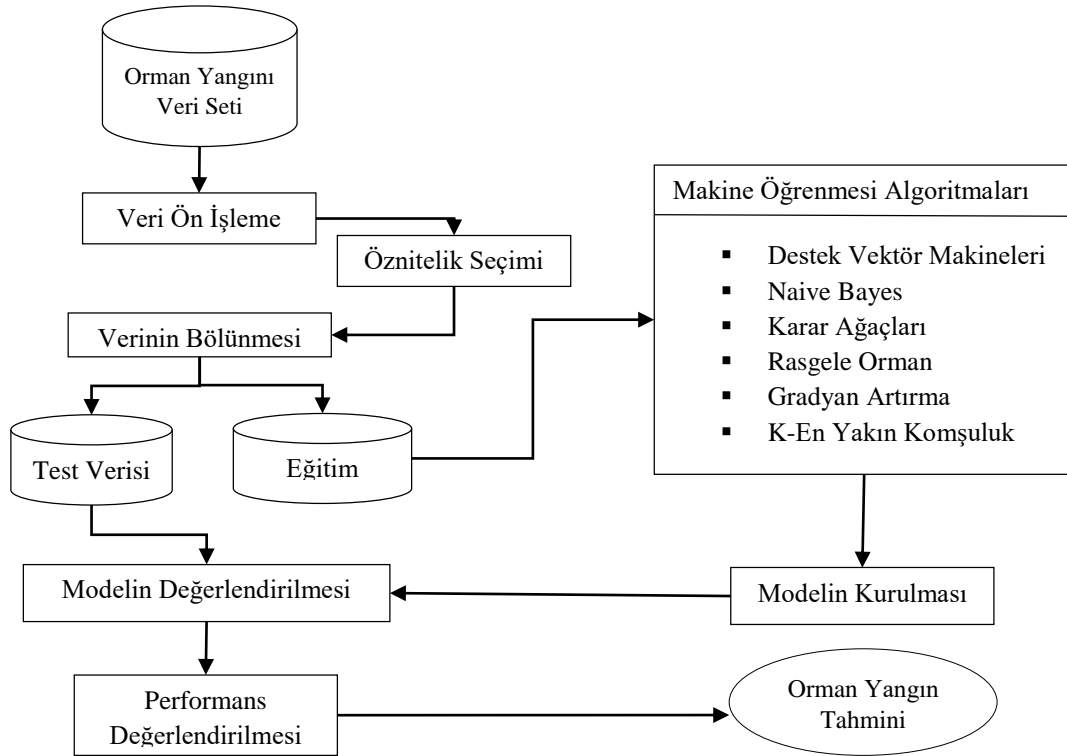
Şekil 1. Veri setinden rasgele alınan 15 örnek

Çalışmada kullanılan veri kümesinde bölgelere göre yanmış ve yanmamış orman yangın sayısının dağılımı Şekil 2'te gösterilmiş olup en çok yangın çıkan bölgenin Taiga Plains olduğu görülmüştür.



Şekil 2. Yangın durumunun bölgelere dağılımı

Yapılan çalışmada orman yangınları veri seti üzerinde 6 farklı makine öğrenme teknikleri kullanılarak orman yangınının varlığı tahmin edilmiştir. Sınıflandırma işlemi için önerilen modelin mimari yapısı Şekil 3'te verilmiştir.



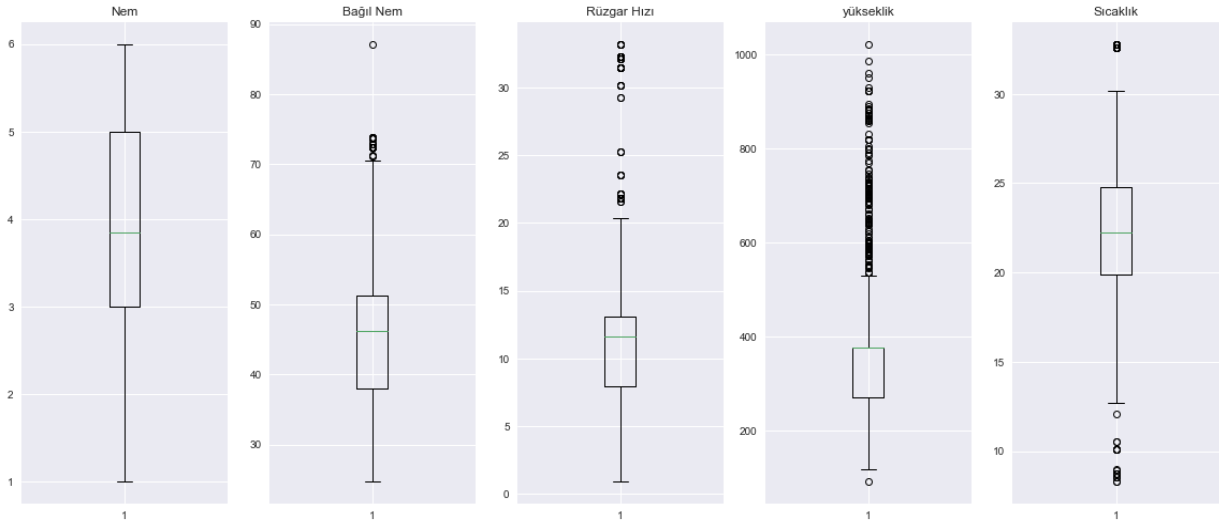
Şekil 3. Orman yangını tahmininde farklı modellerin karşılaştırılmasında kullanılan mimari

Veri Ön İşleme (Data Preprocessing)

Makine öğrenmesi modellerini kurma aşamasında en önemli ve gerekli adımlardan biri veriyi hazır hale getirmektir. Veri ön işleme; veri seti üzerinde eksik veriyi tamamlama, veri temizleme, dönüştürme, normalleştirme ve boyut indirgeme gibi yapılan bir takım işlemler ile verinin modele uygun hale getirilmesidir. Bu adımlar makine öğrenmesi modelinin güvenilir, doğru ve başarılı tahminde bulunması için uygulanmaktadır. Çalışmada kullanılan orman yangınları veri seti üzerinde modele katkısı olmayacak bazı öznitelikler (proje id, ismi vb.) silinmiştir. Veri setinde bazı özniteliklerde ise oldukça fazla boş değer bulunmaktadır. Söz konusu özniteliklerden %50'si boş olan öznitelikler veri setinden çıkarılmıştır. Kalan özniteliklerden eksik değeri bulunanlar ise ortalama değer ile doldurularak tamamlanmıştır. Bu işlem için Python'da sklearn impute kütüphanesinin SimpleImputer yöntemi uygulanmıştır. Aykırı verilerin (outlier) tespitinde kutu grafiği (boxplot-IQR) yöntemi kullanılmıştır. Aykırı değerler bulunduran sayısal öznitelikler, alt ve üst sınır değerine eşitlenmiştir. Şekil 4'te bazı özelliklere ait aykırı değerler gösterilmiştir.

Makine öğrenimi modelleri kurmak için hem giriş hem de çıkış değişkenlerinin sayısal biçimde olması gerekir. Bu çalışmada kategorik olan bağımlı değişken yangın durumu (yangın var/yangın yok) ikili sayısal forma dönüştürülmüştür. Giriş değişkenlerinde aynı şekilde kategorik olan öznitelikler sayısallaştırılmıştır. Bu işlem için Python'nun sklearn preprocessing kütüphanesinden LabelEncoder sınıfı kullanılmıştır.

Verinin normal dağılımda olmaması destek vektör makinaları ve k-en yakın komşu gibi uzaklığa dayalı makine öğrenmesi algoritmalarında doğruluk, hız ve performansı önemli ölçüde etkilemektedir.



Şekil 4. Aykırı değerlerin tespitinde kutu grafiği kullanımı

Bu nedenle veri setinin dağılımını değiştirmeden değerler arasındaki uzaklıkları ortak bir ölçüğe dönüştürmek gerekmektedir. Öznitelik ölçeklendirme işlemi olarak adlandırılan bu ön işleme adımında sadece sayısal özellikler üzerinde işlem yapılmaktadır. Bu çalışmada veriler arasındaki uzaklıkların giderilmesi için istatistiksel ve Min-Max normalizasyon işlemleri yapılmış olup tüm değerlerin kullanılan algoritmaların başarısına eşit şekilde katkıda bulunmaları amaçlanmıştır.

Model tasarımı oluştururken ilk ve en önemli adım ham veri seti üzerinde veri temizleme ve en iyi öznitelik seçme (özellik, nitelik veya değişken seçimi) işlemidir. Veri setinde girdi öznitelik sayısı çok fazla olduğu durumlarda çıktı özniteliğine olan etki azalmaktadır. Kurulan modelin karmaşıklığının artması, yorumlanabilirliğini ve uygulanabilirliğini azaltacaktır. Analiz edilecek veri seti, çeşitli nitelikler ve yüksek bir kardinalite içerebilir. Veri azaltma tekniği, model eğitimi sırasında performansı artırmak için veri kümesinin azaltılmış bir temsilini oluşturmayı amaçlar. Özniteliklerden bazıları model eğitiminde alakasız, hatta gereksiz olabilir (Moreira ve ark., 2018). Bu nedenle öznitelik seçiminin amacı veri seti içerisinde en önemli öznitelikleri model eğitime hazırlamaktır. Seçilen öznitelikler orijinal veri kümesinin yerini alabilecek bir veri alt kümesi olmalıdır. Öznitelik seçiminin başarılı bir uygulaması da modelin performansı azalmadan minimum bir alt küme ile modelin oluşturulmasıdır (García ve ark., 2016). Ayrıca öznitelik seçimi aşırı öğrenme (overfitting) olasılığını azaltır ve modelin eğitim aşamasında gereksiz kaynak tüketiminin önüne geçer. Bu çalışmada oluşturulacak model için kullanılacak öznitelikler, boyut azaltma (dimensionality reduction) tekniklerinde Temel Bileşenler Analizi (Principal Component Analysis-PCA) ve Öznitelik Seçme (Feature Selection) yöntemlerinden Varyansların Analizi (Analysis of Variance-ANOVA) yöntemleri uygulanarak belirlenmiştir. PCA ile veri seti 35 öznitelikten 15'e indirgenmiş, ANOVA ile ise 35 öznitelikten en iyi 11 öznitelik seçilerek model kurulmuştur. Sınıflandırma işlemleri veri setinin hem tüm öznitelikleriyle ve hem de boyut indirgeme sonrası belirlenen öznitelikler ile yapılmıştır.

ANOVA, girdi değişken ile çıktı değişkeni arasındaki ilişkiye göre seçim yapan filtreleme yöntemlerinden biridir. Girdi veya çıktı değişkenlerden birinin kategorik olması durumunda kullanılan bu yöntem kategorik bir değişken ile sayısal olan bir değişkenin arasındaki ilişkiyi istatistiksel olarak ölçmektedir. Bu yaklaşım, gruplar arasındaki varyans oranını hesaplayarak özellikleri sıralamaktadır (Lin ve Ding, 2011). İki grup arasındaki özelliğin önemli ölçüde farklı olup olmadığını tespit etmek için kullanılan ANOVA ile girdi değişkenindeki gruplar çıktı değişkenine göre istatistiksel olarak anlamlılık düzeyinde farklılık göstermiyorsa bu değişken modelden çıkarılmaktadır.

PCA ise, girdi vektörünün kovaryans matrisinin öz vektörleri ile öz değerlerini hesaplayarak elde edilmektedir. Doğrusal olarak yüksek boyutlu bir girdi vektörünü bileşenleri ilişkisiz olan düşük boyutlu bir girdiye dönüştürür (Qiu ve ark., 2012). Böylece veri kümesi, çok değişkenli verinin ana özelliklerini koruyarak daha az sayıda değişken ile temsil edilir. PCA uygulandığında verilerdeki varyasyonun çoğu ilk birkaç bileşende yoğunlaşarak yalnızca önemli farklılıkları olan bileşenler korunur ve geri kalanlar göz ardı edilir. Sonuç olarak boyut indirgeme ile mevcut değişkenlerin kombinasyonundan oluşan yeni değişkenler elde edilerek değişken sayıları azaltılmaktadır. Böylece hazırlanan modellerin optimum sürede ve en iyi performansla çalışması sağlanmaktadır. Bu çalışmada 35 özneliğe sahip veri kümesi PCA ile 15 özneliğe indirgenmiş ve verilerin toplam varyansının %99,9'unun korunabileceği görülmüştür.

Sınıflandırma Algoritmaları (Classification Algorithms)

Yapay zekânın bir alt dalı olan makine öğrenimi, kabul edilebilir bir aralıkta tahminler yapmak için giriş verilerini analiz ederek işlemlerini öğrenen ve optimize eden programlanmış algoritmaları kullanır. Yeni verilerin veri setine dahil edilmesiyle birlikte algoritmalar daha doğru tahminler yapma eğilimindedir (Uddin ve ark., 2019). Makine öğrenimi algoritmaları, geçmiş deneyimlerden öğrenmek ve büyük, yapılandırılmamış ve karmaşık veri kümelerinden faydalı kalıpları tespit etmek amaçlarıyla çeşitli istatistiksel, olasılık ve optimizasyon yöntemleri kullanır (Mitchell, 1997). Özetle makine öğrenmesi, istatistiksel ve matematiksel işlemler ile veriler üzerinden çıkarımlar ve tahminler yapan sistemlerin bilgisayarlar ile modellenmesidir. Sınıflandırma, makine öğrenmesinin gözetimli öğrenme problemlerinden biridir. Sınıflandırma işlemi bağımlı değişken sayısına göre ikiye ayrılmaktadır. Veri setindeki sınıf etiketi ikili değerden (kadın/erkek, pozitif/negatif vb.) oluşuyorsa İkili Sınıflandırma (Binary Classification) yöntemi, Sınıf sayısı üç veya daha fazla sayıdan (bitkilerin türlerinin belirlenmesi) oluşuyorsa çok sınıflı sınıflandırma (Multi-Class Classification) yöntemi kullanılmaktadır.

Bu çalışmada orman yangını veri seti kullanılarak DVM, KA, RO, GA, K-NN ve BN olmak üzere altı farklı denetimli makine öğrenmesi algoritmalarıyla modeller oluşturulmuştur. Veri setindeki sınıf sayısı iki değerden (yangın yok=0, yangın var=1) olduğundan ikili sınıflandırma yöntemi kullanılmıştır.

K-En Yakın Komşuluk Algoritması (K-NN)

K-NN Algoritması bir sınıflandırma algoritmasıdır. Sınıfı bilinmeyen verilerin, eğitim setindeki diğer veriler ile karşılaştırılıp bir uzaklık ölçümü sonucu hesaplanmaktadır. Uzaklığa göre eğitim setinden ilgili örneğe en yakın k tane örnek belirlenerek yeni örneğin sınıf etiketi, k adet en yakın komşunun sınıf etiketlerinin çoğunluğuna göre atanmaktadır. Nesnenin birçok etikete sahip olabileceği uygulamalar kadar çok modlu sınıflar için de uygun bir yöntemdir. Verimi düşüktür ve performansı iyi bir 'k' değerinin seçilmesine bağlıdır (Singh ve ark., 2016). Gürültüden olumsuz etkilenir ve alakasız özelliklere de duyarlıdır. Tüm verilerin tekrar ziyaret edilmesi gerektiğinden performans da boyuta göre değişmektedir (Islam ve ark., 2007).

Destek Vektör Makineleri (DVM)

DVM, sınıflandırma problemlerinde denetimli öğrenme yöntemi kapsamına girmektedir. Düzlem üzerindeki noktaların bir doğru veya hiper düzlem ile ayrıştırılarak sınıflandırılması esasına dayanır. İstatistiksel öğrenme teorisi ve yapısal riski en aza indirme ilkesine dayanan, sınıflandırma ve eğri uydurma problemlerinin çözümü amacıyla kullanılan bir öğrenme yöntemidir. Yüksek boyutlu verilere dayanıklıdır ve iyi bir genelleme yeteneğine sahiptir. Ancak eğitim hızı düşüktür ve performansı parametre seçimine bağlı olarak değişmektedir (Caruana ve Niculescu-Mizil, 2006).

Karar Ağaçları (KA)

Ağaç tabanlı öğrenme algoritmalarından olan KA, çok sayıda kayıt içeren bir veri kümesini, bir dizi karar kuralları uygulayarak daha küçük kümelere bölmek için kullanılan bir yapıdır. Kümenin tüm elemanları aynı sınıf etiketine sahip olana kadar kümeleme işlemi derinlemesine devam eder. Kararlar yapraklarda ve veriler düğümlerde bölünür. Sınıflandırma ağacında karar değişkeni kategoriktir ve Regresyon ağacında karar değişkeni sürekliktir. Karar Ağaçları, yorumlamada, kategorik ve nicel değerleri ele almada kolaylık gösterir, özniteliklerdeki eksik değerleri en olası değerle doldurabilme özelliğine sahiptir (Ray, 2019).

Rastgele Orman (RO)

RO, denetimli sınıflandırma algoritmalarından biridir. Bir dizi karar ağacını eğiterek ve topluluktaki tüm ağaçlar üzerinde çoğunluğun olduğu sınıfı döndürerek çalışan bir topluluk yöntemidir (Lorenna ve ark., 2011). Rastgele Orman, birbirinden bağımsız olarak çalışan birçok karar ağacının bir araya gelerek aralarından en yüksek puan alan değerini seçilmesi işlemidir. Karar ağaçları algoritmasından temel farkı, kök düğümü bulma ve düğümleri bölme işleminin rastgele gerçekleşmesidir. Verimliliği ve doğruluğu ile çok kullanışlı bir sınıflandırıcıdır (Gislason ve ark., 2006). Çoğu zaman, hiperparametre kullanılmadan da iyi sonuçlar verebilir. Karmaşık ve gürültülü veri setlerinde bile hızlı ve iyi sonuçlar verebilmektedir.

Gradyan Artırma (GA)

GA modeli, Friedman tarafından tanımlanan bir topluluk makine öğrenme modelidir (Friedman, 2001). Artırma terimi, daha güçlü bir tahmin üretmek için zayıf tahmin ağaçlarının bir araya getirildiği belirli bir algoritma türünü ifade eder (Xie ve ark., 2019). Karar ağaçlarına benzer tahmin modelleri oluşturan bir makine öğrenmesi tekniğidir. Gradyan artırma ve rastgele orman yöntemlerinin her ikisi de çekirdek karar ağacı algoritmalarını kullanmaktadır. Gradyan artırma modelinin ana fikri, nihai modelin doğruluğunu ve sağlamlığını geliştirmek için birden fazla zayıf öğrenciyi bir araya getirmektir. Aşırı uyum sorunlarını önlemek için gradyan artırma modeli, yeni ağaçtan gelen katkıyı ölçeklendirmek amacıyla bir öğrenme oranı kullanır. (Yoon, 2021). Rastgele ormandaki gibi veri setini çoklu sayıda alt veri setine bölme bu yaklaşımda bulunmamaktadır. Veri setini olduğu gibi kullanarak bir karar ağacı oluşturur ve bunun hatalarından yola çıkarak yeni bir karar ağacı daha oluşturmaktadır.

Naive Bayes (NB)

NB sınıflandırma algoritmasının temeli Bayes teoremine dayanmaktadır. Koşullu olasılığa dayanan bu teknikte eğitim verileri ile güncellenen bir olasılık tablosu bulunmaktadır. Bu tablo yeni bir gözlemi tahmin etmek için sınıf olasılıklarına bakılması gereken özellik değerlerine dayanmaktadır (Ray,2019). Algoritmanın çalışma yönteminde, bir eleman için her durumun olasılığı hesaplanıp olasılık değeri en yüksek olana göre sınıflandırma işlemi gerçekleştirilmektedir. Bu sınıflandırıcı, bir sınıftaki belirli bir özelliğin diğer herhangi bir özellik ile doğrudan ilişkili olmadığını varsaymakta ve ancak o sınıfa ait özelliklerin kendi aralarında karşılıklı bağımlılığı olabileceği esasına dayanmaktadır (Rish, 2001).

BULGULAR VE TARTIŞMA

Bu çalışmada oluşturulan ikili sınıflandırma modellerinin sınıflandırma performansları karışıklık matrisi (confusion matrix) ile değerlendirilmiştir. Karışıklık matrisinde tahmin edilen ve gerçek değerlerin kombinasyonunu içeren 4 farklı değer bulunmaktadır. Bunlardan; Doğru Pozitif (True Positive) değeri, gerçek durum pozitif iken test sonucu tahmininde pozitif olma durumu, Yanlış Pozitif (False Positive) değeri, gerçek durum negatif iken test sonucu tahmin değerinin pozitif olma durumu, Doğru Negatif (True Negative) değeri, gerçek durum negatif iken test sonucu tahmin değerinin de negatif

olma durumu ve Yanlış Negatif (False Negative) değeri ise gerçek durum pozitif iken test sonucu tahmin edilen değerin negatif olma durumudur. Doğru Pozitif (DP) ve Doğru Negatif (DN) modelin doğru tahminlerini, Yanlış Pozitif (YP) ve Yanlış Negatif (YN) ise modelin yanlış tahminlerini gösterir. Karışıklık matrisinde bu değerlerin gösterimi ve karışıklık matrisi içindeki değerler kullanılarak yapılan bazı ölçümler Şekil 5’ de verilmiştir.

		TAHMİN		
		Pozitif	Negatif	
GERÇEK	Pozitif	Doğru Pozitif	Yanlış Negatif	$Kesinlik = \frac{DP}{DP+YP}$
	Negatif	Yanlış Pozitif	Doğru Negatif	$Hassasiyet = \frac{DN}{DP+YN}$
				$Doğruluk = \frac{DP+DN}{DP+DN+YP+YN}$
				$F1-Skoru = 2x \frac{Kesinlik \times Hassasiyet}{Kesinlik + Hassasiyet}$

Şekil 5. İkili sınıflandırma için karışıklık matrisi ve diğer performans metrikleri

Farklı sınıflandırma modellerinin karşılaştırılması ve başarı değerlendirilmesi

Ön işleme adımları sonrası veri seti, %75’i eğitim verisi ve %25’i test verisi olacak şekilde ikiye ayrılmıştır. Bu veri seti üzerinde bağımlı değişken (sınıf etiketi) ikili değerden (1 yangın var, 0 yangın yok) oluştuğundan ikili sınıflandırma işlemi gerçekleştirilmiştir. Boyut indirgeme yapılmadan 35 özneliğin tamamıyla sınıflandırma işlemi yapılmıştır. Model doğrulama yöntemlerinden çapraz doğrulama (cross validation) katsayısı 5 seçilerek doğruluk oranları ve çalışma süreleri Çizelge 1’de gösterilmiştir. Söz konusu modellerin performansı analiz edilerek en hızlı çalışan algoritmanın KA, doğruluk oranı en yüksek olan algoritmanın ise GA algoritması olduğu görülmüştür.

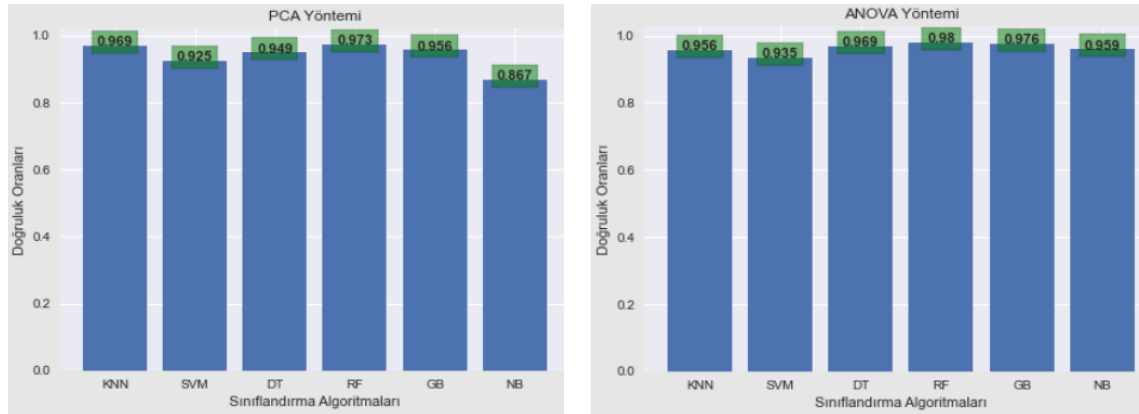
Çizelge 1. Boyut indirgemenen sınıflandırma modelleri için başarı değerlendirmesi

Sınıflandırma Algoritmaları	Doğruluk (Accuracy)	Cross-Validation Doğruluk	Kesinlik (Precision)	Hassasiyet (Sensitivity)	F1 Skor (F1 Score)	Çalışma Süresi (sn)
KNN	%98	%94	%99	%98	%99	0.046
DVM	%97	%89.5	%100	%97	%98	0.25
KA	%98	%93.5	%100	%98	%99	0.015
RO	%98	%96	%100	%98	%99	0.140
GA	%99	%93	%100	%100	%100	0.656
NB	%98	%98.5	%100	%98	%99	0.985

Bu çalışmada, veri kümesi üzerinde öznelik seçim yöntemleri uygulanarak elde edilen yeni öznelikler, çapraz doğrulama yardımıyla sınıflandırılmıştır. Bu yöntemlerden PCA ve ANOVA teknikleri ile boyut azaltma yapılarak hedef değişkeni bulmada en önemli öznelikler belirlenmiştir. Öğrenme algoritması için seçilen en yüksek doğruluk oranına sahip en küçük boyutlu özellik alt kümesi elde edilene kadar test işlemleri gerçekleştirilmiştir. Bu işlemler sonucu orijinal veri kümesini temsil eden en iyi alt küme seçimi için PCA ile 35 özellik arasından en iyi 15 adet özellik seçimi ve ANOVA ile ise 11 adet özellik seçimi yapılmıştır. Sınıflandırma için veri seti aynı şekilde %75’i eğitim ve %25’i test verisi olacak şekilde ayrılmıştır. Çizelge 2’de yer alan algoritmalar ve bunlara ait performans değerleri karşılaştırılmıştır.

Çizelge 2. Boyut azaltma sonrası sınıflandırma modelleri için başarı değerlendirilmesi

Sınıflandırma Algoritmaları	Doğruluk (Accuracy)	Cross-Validation Doğruluk	Keskinlik (Precision)	Hassasiyet (Sensitivity)	F1 Skor (F1 Score)	Çalışma Süresi (sn)
KNN (PCA)	%97	%94	%99	%97	%98	0.031
KNN (ANOVA)	%96	%84	%98	%97	%98	0.062
DVM (PCA)	%92.5	%89.5	%92	%100	%96	0.125
DVM (ANOVA)	%94	%87	%97	%95	%96	0.046
KA (PCA)	%96	%89	%99	%96	%98	0.015
KA (ANOVA)	%97	%95	%100	%97	%98	0.010
RO (PCA)	%97	%93	%100	%97	%98	0.218
RO (ANOVA)	%98	%97	%100	%98	%99	0.14
GA (PCA)	%96	%90	%98	%97	%98	1.140
GA (ANOVA)	%97.6	%95	%100	%98	%99	0.5
NB (PCA)	%90	%87.5	%98	%91	%94	0.015
NB (ANOVA)	%96	%96	%100	%95	%98	0.010



Şekil 6. Boyut indirgeme sonrası doğruluk oranları

Modellerin aşırı uyum (overfitting) sorununu gidermek için validasyon işleminde katsayı 5 olarak alınmış ve bu doğrultuda veri seti 5 eşit parçaya ayrılarak sınıflandırma işlemi gerçekleştirilmiştir. Buna göre her defasında veri setinin farklı bölümleri hem test hem de eğitim amaçlı kullanılmıştır. Validasyon işlemine ait ortalama doğruluk değerleri Çizelge 2’de gösterilmiştir. Öznitelik seçme işleminde PCA yöntemine göre ANOVA yönteminin hem doğruluk oranı hem de algoritmaların çalışma süreleri bakımından daha iyi sonuç verdiği görülmüştür. ANOVA ile özellik seçimi sonucu en yüksek doğruluk oranı %98 olarak RO algoritmasıyla elde edilmiş olup yine bu algoritmanın validasyon işlemi sonucu %97 başarı oranı ile diğer algoritmalara göre daha yüksek başarı gözlemlenmiştir (Şekil 6). Algoritmalar çalışma süreleri bakımından analiz edildiğinde KA ve NB algoritmaları 0.010 sn olarak en kısa sürede

model eğitimini tamamlamışlardır. Sonuçlar, Çizelge 1 ve Çizelge 2’de karşılaştırıldığında öznitelik sayısı önemli ölçüde azaldığından buna bağlı olarak modelin eğitim süresi de kısalmıştır. Doğruluk oranı ise yaklaşık %1.5 oranında azalmıştır. Veri setinin örnekleme sayısının çok fazla olması durumunda özellik seçimi daha belirgin şekilde modelin hızı ve performansını etkileyebilmektedir.

Bu çalışmada ayrıca öznitelik seçme yöntemlerinden Ki-Kare (Chi-Square), Özyinelemeli Özellik Eliminasyonu (Recursive Feature Elimination) ve Karar Ağaçları ile Değişken Seçimi (Feature Importances) yöntemleri test edilmiş olup model süresi ve başarı oranında önemli değişiklikler görülmemiştir.

SONUÇ

Bu çalışmada orman yangınlarıyla mücadele çalışmaları için olası orman yangınına önceden tahmin ederek yangınlara hızlı ve etkin müdahale edilmesi hedeflenmiştir. NASA’nın web sayfasından alınan veri seti üzerinde farklı sınıflandırma algoritmaları kullanılarak modeller geliştirilmiştir. Veri seti üzerinde birkaç ön işleme adımları uygulanarak veri seti sınıflandırmaya uygun hale getirilmiştir. Makine öğrenmesi algoritmalarından DVM, KA, RO, GA, K-NN ve NB olmak üzere 6 farklı sınıflandırma algoritması kullanılarak sınıflandırma performansları karşılaştırılmıştır. Sınıf etiketinde iki farklı değer bulunduğu için ikili sınıflandırma metodu uygulanmış ve model çıktıları karışıklık matrisi ile değerlendirilmiştir. Öznitelik seçimi ile orijinal veri setini temsil edebilecek en iyi altküme seçimi yapılarak 35 öznitelik PCA ile 15 ve ANOVA ile 11 özniteliğe indirgenmiştir. Model için en faydalı ve en önemli özellikler seçilerek yeni veri kümesiyle sınıflandırma işlemi yapılmıştır. Böylece daha az sayıda özniteliği işleme alarak orijinal veri kümesiyle ulaşılmak istenen doğruluk oranına en yakın olan tahmin değeri elde edilmiştir. Ayrıca öznitelik seçimi sonucu modelin eğitim süresi ve aşırı uyum riski azalmıştır. Çalışma sonunda validasyon işlemi ile birlikte en iyi iki algoritmadan NB ile %96 ve RO ile %97 doğruluk oranı elde edilmiştir. Gelecekteki çalışmalarda orman yangınlarının çıkış nedenleri ile yangın noktalarının yerleşim yerlerine uzaklığı gibi faktörlerin veri kümesine dâhil edilerek modelin güçlendirilmesi sağlanabilir. Ormanlık alanların yerleşim yerlerine yakın olması, yangınların daha çok insan kaynaklı nedenlerle meydana geldiğini göstermektedir. Yangınların çıkış nedenlerinin bilinmesi ise bu çıkış nedenine göre orman yangınlarına yönelik tedbirlerin alınması gibi önemli faydalar sağlanabilir. Gelecekte yine benzer bir problem derin öğrenme ve sezgi ötesi optimizasyon yöntemleriyle çözülebilir.

Çıkar Çatışması

Makale yazarları aralarında herhangi bir çıkar çatışması olmadığını beyan ederler.

Yazar Katkısı

Yazarlar makaleye eşit oranda katkı sağlamış olduklarını beyan eder.

KAYNAKLAR

- Arif, M., Alghamdi, K. K., Sahel, S. A., Alosaimi, S. O., Alsaft, M. E., Alharthi, M. A., & Arif, M. (2021). Role of machine learning algorithms in forest fire management: A literature review. *J. Robot. Autom.*, 5, 212-226.
- Arpacı, A., Malowerschnig, B., Sass, O., & Vacik, H. (2014). Using multi variate data mining techniques for estimating fire susceptibility of Tyrolean forests. *Applied Geography*, 53, 258-270.
- Bayat, G., & Yıldız, K. (2022). Comparison of the Machine Learning Methods to Predict Wildfire Areas. *Turkish Journal of Science and Technology*, 17(2), 241-250.

- Castelli, M., Vanneschi, L., & Popovič, A. (2015). Predicting burned areas of forest fires: an artificial intelligence approach. *Fire ecology*, 11(1), 106-118.
- Caruana, R., & Niculescu-Mizil, A. (2006, June). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning* (pp. 161-168).
- Chen, J., Wang, X., Yu, Y., Yuan, X., Quan, X., & Huang, H. (2022). Improved Prediction of Forest Fire Risk in Central and Northern China by a Time-Decaying Precipitation Model. *Forests*, 13(3), 480.
- Coughlan, R., Di Giuseppe, F., Vitolo, C., Barnard, C., Lopez, P., & Drusch, M. (2021). Using machine learning to predict fire-ignition occurrences from lightning forecasts. *Meteorological applications*, 28(1), e1973.
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.
- Fidanboy, M., Nihat, A., & Okyay, S. (2022). Derin öğrenmeye dayalı orman yangını tahmin modeli geliştirilmesi ve Türkiye yangın risk haritasının oluşturulması. *Ormanlık Araştırma Dergisi*, 9(2), 206-218.
- FAO. (2020). Global Forest Resources Assessment 2020 – Key findings. Rome., Erişim adresi: <https://www.fao.org/3/ca8753en/ca8753en.pdf> (Erişim Tarihi: 16.10.2022).
- Gislason, P. O., Benediktsson, J. A., & Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern recognition letters*, 27(4), 294-300.
- García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M., & Herrera, F. (2016). Big data preprocessing: methods and prospects. *Big Data Analytics*, 1(1), 1-22.
- Islam, M.J., Wu, Q.M., Ahmadi, M., & Sid-Ahmed, M.A. (2007). Investigating the Performance of Naive- Bayes Classifiers and K- Nearest Neighbor Classifiers. *2007 International Conference on Convergence Information Technology (ICCIT 2007)*, 1541-1546.
- Lorena, A. C., Jacintho, L. F., Siqueira, M. F., De Giovanni, R., Lohmann, L. G., De Carvalho, A. C., & Yamamoto, M. (2011). Comparing machine learning classifiers in potential distribution modelling. *Expert Systems with Applications*, 38(5), 5268-5275.
- Liang, H., Zhang, M., & Wang, H. (2019). A neural network model for wildfire scale prediction using meteorological factors. *IEEE Access*, 7, 176746-176755.
- Lai, C., Zeng, S., Guo, W., Liu, X., Li, Y., & Liao, B. (2022). Forest Fire Prediction with Imbalanced Data Using a Deep Neural Network Method. *Forests*, 13(7), 1129.
- Lin, H., & Ding, H. (2011). Predicting ion channels and their types by the dipeptide mode of pseudo amino acid composition. *Journal of theoretical biology*, 269(1), 64-69.
- Moreira, L., Dantas, C., Oliveira, L., Soares, J., & Ogasawara, E. (2018). On evaluating data preprocessing methods for machine learning models for flight delays. In *2018 International Joint Conference on Neural Networks (IJCNN)* (pp. 1-8). IEEE.
- Mimboro, P., Yanuargi, B., Surimbac, R., Kusri, K., & Khusnawi, K. (2022). Forest Fire Prediction Using K-Mean Clustering and Random Forest Classifier. *CSRID Journal*, 14(2): 157-165. DOI: <http://dx.doi.org/10.22303/csrid.14.2.2022.157-165>.
- Mitchell, T. M., & Mitchell, T. M. (1997). *Machine learning* (Vol. 1, No. 9). New York: McGraw-hill.
- Niranjan, T., Swetha, D., Charitha, V., & Stephen, A. J. (2019). Predicting Burned Area Of Forest Fires. *IRJCS:: International Research Journal of Computer Science*, 6, 132-136.
- OGM. (2021). Orman Genel Müdürlüğü, Ormanlık İstatistikleri. Erişim adresi: <https://www.ogm.gov.tr/tr/e-kutuphane/resmi-istatistikler> (Erişim Tarihi: 18.10.2022).

- 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*, 14(21), 5546.
- Preeti, T., Kanakaraddi, S., Beelagi, A., Malagi, S., & Sudi, A. (2021). Forest Fire Prediction Using Machine Learning Techniques, *2021 International Conference on Intelligent Technologies (CONIT)*, pp. 1-6, doi: 10.1109/CONIT51480.2021.9498448.
- Qiu, J., Wang, H., Lu, J., Zhang, B., & Du, K. L. (2012). Neural network implementations for PCA and its extensions. *International Scholarly Research Notices*, 2012.
- Rakshit, P., Sarkar, S., Khan, S., Saha, P., Bhattacharyya, S., Dey, N., Islam, S., & Pal, S., (2021). Prediction of Forest Fire Using Machine Learning Algorithms: The Search for the Better Algorithm, *2021 6th International Conference on Innovative Technology in Intelligent System and Industrial Applications (CITISIA)*, pp. 1-6, doi: 10.1109/CITISIA53721.2021.9719887.
- Ray, S. (2019). A quick review of machine learning algorithms. In *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)* (pp. 35-39). IEEE. doi: 10.1109/COMITCon.2019.8862451.
- Rish, I. (2001). An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence* (Vol. 3, No. 22, pp. 41-46).
- Silva, I. D. B., Valle, M. E., Barros, L. C., & Meyer, J. F. C. (2020). A wildfire warning system applied to the state of Acre in the Brazilian Amazon. *Applied Soft Computing*, 89, 106075.
- Sevinc, V., Kucuk, O., & Goltas, M. (2020). A Bayesian network model for prediction and analysis of possible forest fire causes. *Forest Ecology and Management*, 457, 117723.
- Shao, Y., Feng, Z., Sun, L., Yang, X., Li, Y., Xu, B., & Chen, Y. (2022). Mapping China's Forest Fire Risks with Machine Learning. *Forests*, 13(6), 856. <https://doi.org/10.3390/f13060856>.
- Spoorthy, M. R., & Kumar, H. 2022. Detection of Forest Fire Areas using Machine Learning. *Communication and Technology (IJARSCT)*, 2(2): DOI: 10.48175/IJARSCT-5623.
- Singh, A., Thakur, N., & Sharma, A. (2016). A review of supervised machine learning algorithms. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)* (pp. 1310-1315). Ieee.
- Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC medical informatics and decision making*, 19(1), 1-16.
- Walker, X.J., Baltzer, J.L., Bourgeau-Chavez, L.L., Day, N.J., De groot W.J., Dieleman, C., Hoy, E.E., Johnstone, J.F., Kane, E.S., Parisien, M.A., Potter, S., Rogers, B.M., Turetsky, M.R., Veraverbeke, S., Whitman, E., & Mack, M.C. (2020). ABoVE: Synthesis of Burned and Unburned Forest Site Data, AK and Canada, 1983-2016. ORNL DAAC, Oak Ridge, Tennessee, USA. <https://doi.org/10.3334/ORNLDAAC/1744>.
- Xie, Y., & Peng, M. (2019). Forest fire forecasting using ensemble learning approaches. *Neural Computing and Applications*, 31, 4541-4550. <https://doi.org/10.1007/s00521-018-3515-0>.
- Xie, Y., Jiang, B., Gong, E., Li, Y., Zhu, G., Michel, P., & Zaharchuk, G. (2019). Use of gradient boosting machine learning to predict patient outcome in acute ischemic stroke on the basis of imaging, demographic, and clinical information. *American Journal of Roentgenology*, 212(1), 44-51.
- Yoon, J. (2021). Forecasting of real GDP growth using machine learning models: Gradient boosting and random forest approach. *Computational Economics*, 57(1), 247-265.