



## Orman Yangınlarının Tahmini için Makine Öğrenimi Yaklaşımı: Türkiye Uygulaması

Senem BAYBAŞ<sup>ID\*</sup>, Özlem AKGÜN<sup>ID1</sup>, Emir Deniz Cem ADIGÜZEL<sup>ID1</sup>, Ehat Eser GÜRVARAR<sup>ID1</sup>, Esra GÖKPINAR<sup>ID1</sup>

<sup>1</sup>Gazi Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Ankara

### Özet

Son yıllarda küresel ısınmanın etkisiyle artan hava sıcaklıkları, orman yangınlarının dünya genelinde yaygınlaşmasına neden olmuştur. Türkiye’de bu durumdan ciddi şekilde etkilenmektedir. Bu sebeple, orman yangınlarına karşı önleyici tedbirler almak ve etkin mücadele stratejileri geliştirmek büyük önem taşımaktadır. Orman yangınlarının önceden tahmin edilmesi veya erken tespit edilmesi, hızlı müdahale ve zararları en aza indirmek için hayati öneme sahiptir. Literatürde, orman yangınlarını tahmin etmek amacıyla meteorolojik veriler ve uzaktan algılama verileri kullanılmaktadır. Bu çalışmada, T.C. Tarım ve Orman Bakanlığı’ndan alınan veriler kullanılarak, 2022 yılında Türkiye’de meydana gelen orman yangınlarının meteorolojik etkileri incelenmiş ve makine öğrenimi yöntemleriyle bir tahmin modeli geliştirilmiştir. Çalışmada, Lojistik Regresyon, K-En Yakın Komşu, Random Forest ve Karar Ağacı algoritmaları kullanılmıştır. Algoritmaların doğruluk oranları ise K-En Yakın Komşu %77.48, Lojistik Regresyon %77.98, Karar Ağacı %77.81 ve Random Forest %76.15 olarak hesaplanmıştır. Bu sonuçlar, Lojistik Regresyon modelinin en yüksek performansı gösterdiğini ortaya koymaktadır.

**Anahtar Kelimeler:** Orman Yangınları, Yanan Alan Tahmini, Makine Öğrenmesi Yöntemleri

### Machine Learning Approach For Predicting Forest Fires: The Case of Turkey

### Abstract

In recent years, increasing air temperatures due to global warming have caused forest fires to become widespread worldwide, with Türkiye also being seriously affected. Therefore, it is crucial to take preventive measures and develop effective combating strategies. Early prediction or detection of forest fires is vital for rapid response and minimizing damages. Meteorological and remote sensing data are used to predict forest fires. This study analyzes the meteorological impacts of forest fires in Türkiye in 2022 using data from the Turkish Ministry of Agriculture and Forestry and develops a prediction model using machine learning methods. Logistic Regression, K-Nearest Neighbor, Random Forest, and Decision Tree algorithms were used, with accuracy rates of 77.48% for K-Nearest Neighbor, 77.98% for Logistic Regression, 77.81% for Decision Tree, and 76.15% for Random Forest. The Logistic Regression model demonstrated the highest performance.

**Keywords:** Forest Fires, Burned Area Estimation, Machine Learning Methods

### Makale Bilgisi

Başvuru:  
16/07/2024

Kabul:  
06/10/2024

\* İletişim e-posta: senem.baybas@gazi.edu.tr

## 1 Giriş

Ormanlar; insanlığa gıda, yakıt, barınak, temiz hava ve su, ilaç, gelir kaynağı, istihdam, dinlenme, peyzaj gibi maddi-manevi birçok ekonomik, ekolojik, sosyokültürel faydalar sunan tabii bir kaynaktır. Bir ekosistem olarak orman, belirli bir kapalılıkta ağaçlar, diğer bitki ve hayvan topluluğu ile topraktaki gözle görülmeyen diğer organizmaların cansız çevreyle belli bir denge içinde karşılıklı olarak birbirleriyle etkileşimde bulunduğu canlı bir sistem ve topluluktur [1]. Orman varlıklarını en çok tehlikeye uğratan ve ülke ekonomisinde büyük kayıplara yol açan tahrip faktörlerinden biri de orman yangınlarıdır [2]. Dayananda [3] yaptığı çalışmada orman yangınlarının başlıca sebeplerini genel olarak iki sınıfa ayırmıştır. Bunlardan ilki yıldırım düşmesi, elektrik hatları, yanardağ patlaması ve aşırı derecede yükselen hava sıcaklıkları gibi doğal olaylardır. İkinci sebep olarak ise bilinçli ve bilinçsiz insan kaynaklı davranışlar olduğunu ifade etmiştir. Küçük ve Sağlam [4] yaptığı çalışmada ise yangın tehlikesinin, yangınların başlaması ve yangın davranışı üzerinde etkili olan en önemli hava koşullarının hava sıcaklığı, bağıl nem, rüzgâr hızı ve yağış miktarı olduğunu saptamıştır. Hava koşulları, özellikle yanıcı maddenin nem içeriğini değiştirdiğinden, yanıcı maddenin tutuşması, yanabilirliği ve yanmanın sürekliliği üzerinde doğrudan etkilidirler. Hava koşullarında meydana gelecek değişimler orman yangınlarının davranışı üzerinde önemli farklılıklara sebep olmaktadır. Bu nedenle hava koşullarının orman yangınları üzerine olan etkilerinin çok iyi bir şekilde ortaya konulması gerekmektedir.

Türkiye de bulunduğu Akdeniz iklim kuşağından dolayı çokça orman yangınları ile karşı karşıya kalmaktadır [5]. Türkiye Orman Genel Müdürlüğü verilerine göre; 2022 yılında 1 Ocak- 9 Ağustos tarihleri arasında 1203 orman yangını çıkarken toplam 8 bin 366 hektar alan zarar görmüştür. 2023 yılında ise orman yangın sayısı 1221'e yükselmiş ve 7 bin 908 hektar alan zarar görmüştür. 2022 yılında sadece 1-9 Ağustos tarihleri arasında ülkemizde 89 tane orman yangını tespit edilmiş ve toplam 207 hektar alan zarar görmüştür. Yine aynı dönemde 2023 yılında ise 148 tane orman yangını çıkmış ve toplam 175 hektar alan zarar görmüştür. İklim değişikliğinin etkisiyle ortalama sıcaklık değerindeki artış, orman yangınlarının çıkma olasılığını artırmaktadır. Bu durum doğal sebeplerle başlayabilecek yangınların izlenmesi ve potansiyel

risk bölgelerinin belirlenmesi açısından büyük önem taşımaktadır. Hızlı müdahale imkânı sağlayarak olası yangın alanlarının belirlenmesi ormanlık alanlara yönelik riskleri de minimize edecektir. Böylelikle ilgili kurumların harekete geçmesi ve gerekli kaynakların planlanması mümkün olacaktır. Bu çalışma Türkiye'de son zamanlarda meydana gelen orman yangınlarına odaklanmıştır. Bu amaç doğrultusunda hava sıcaklığı, rüzgâr şiddeti, nem gibi iklim parametreleri kullanılarak orman yangınlarının tahmin modeli geliştirilmiştir. Çalışmada kullanılan Makine Öğrenim (ML) teknikler ile son yıllarda ortaya çıkan orman yangınlarını modellemek oldukça bilgi verici olacaktır.

İlgili literatür incelendiğinde, yapılan çalışmaların çoğu çeşitli değişkenlere ve kullanılan yöntemlere göre orman yangınlarının erken tespitinde en iyi yöntemi bulmayı ortaya koymaya çalışmaktadır. Cortez ve Morais [6] yaptıkları çalışmada Kuzey Portekiz'deki milli parkta meydana gelen orman yangınının kayıtlarını kullanarak, orman yangınlarının yanan alanını tahmin etmek için Veri Madenciliği yaklaşımını kullanmışlardır. Bayat ve Yıldız [7], Portekiz'in kuzeyinde yer alan bir milli parkta meydana gelen 512 orman yangınının kayıtlarını kullanarak sıcaklık, rüzgâr, nem ve yağış gibi çeşitli özelliklere dayalı olarak yangın boyutunu tahmin etmek için farklı ML algoritmaları kullanmıştır. Chen ve ark. [8] insansız hava araçları (İHA) tarafından elde edilen görüntüler ile orman yangınların tespitini ML tekniklerinden biri olan CNN (Convolutional Neural Network) algoritmaları ile tahmin etmeye çalışarak orman yangınlarının erken tespitini gerçekleştirmeyi sağlamaya çalışmışlardır. Beşli ve Tenekeci [9] tarafından yapılan çalışmada ise Kanada'nın ormanlık alanlarında 2013 – 2014 yılları arasında meydana gelmiş olan orman yangınları ve uydulardan alınan veriler ile karar ağaçları yöntemi kullanılarak orman yangınları tahmin etmeye çalışmışlardır.

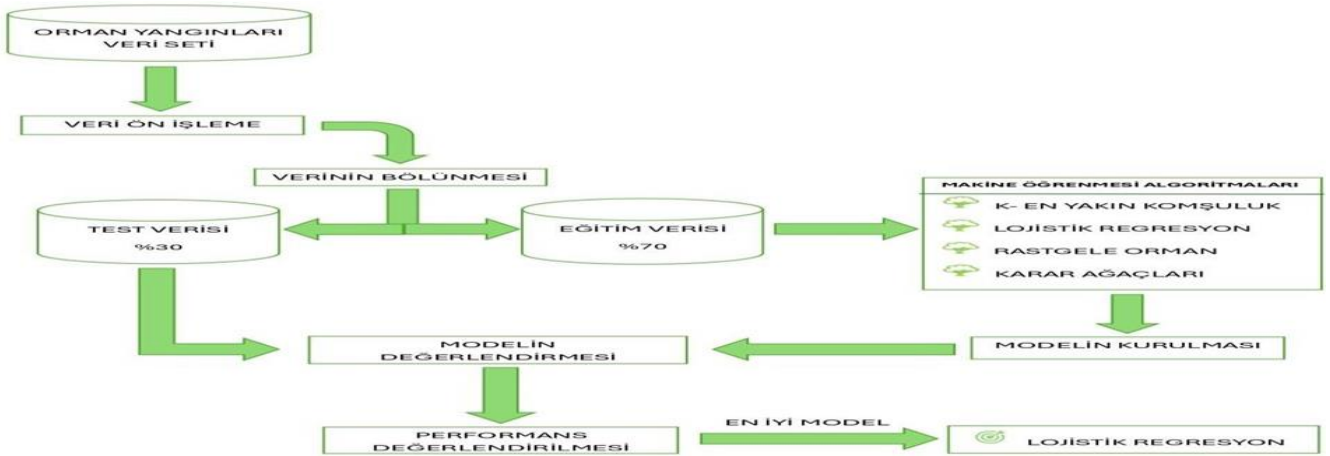
Sayad ve ark. [10] Kanada'nın merkezindeki (çoğunlukla Britanya Kolumbiyası ve Quebec) ormanlık alanlarında 2013-2014 yılları arasında meydana gelmiş olan orman yangınlarını karar ağaçları kullanarak tahmin etmiştir. Mohajane ve ark. [11] doğru stratejilerin geliştirilmesi olası etkileri önlemek ve felaket olaylarının meydana gelmesini mümkün olduğunca en aza indirmek için beş yeni hibrit ML algoritması ile şiddetli koşulları modellemeye ve tahmin etmeye çalışarak Akdeniz Bölgesi'nde orman yangını duyarlılığını

haritalandırmışlardır. Kukuk ve Kilimci [12] Google images, Shutterstock, Getty Images web sitelerinden veri kazıma yöntemi ile toplanan yangın görüntüsü verileri ile geleneksel ML algoritmaları, nesne algılama teknikleri, derin ve hibrit derin öğrenme modelleri kullanılarak orman yangını tespitinin analizini gerçekleştirmeye çalışmışlardır. Yıldırım ve ark. [13], 1983-2016 yılları arasında Alaska ve Kanada'daki yanmış ve yanmamış alanlardaki arazilerden toplanan veriler ile ML yöntemlerini kullanarak orman yangını tahmin modeli oluşturmuşlardır. Türkiye verileri kullanılarak yapılan çalışmaları incelediğimizde ise Akıncı ve Akıncı [14] yaptığı çalışmada farklı ML tekniklerini kullanarak Türkiye'nin Antalya ilinin Manavgat ilçesi için orman yangını duyarlılık haritaları üretmişlerdir. İban ve Sekertekin [15], Türkiye'nin Akdeniz Bölgesi'nde yer alan Adana ve Mersin illeri için ML tabanlı orman yangını duyarlılık haritaları oluşturmuşlardır. Kantarcioglu ve ark. [16] ML tekniği olan Yapay Sinir Ağları kullanılarak Türkiye'nin İstanbul İli ve Trakya Bölgesi'ndeki orman yangını duyarlılığı değerlendirilmiştir. Kantarcioglu ve ark. [17], Türkiye'nin kuzeydoğusunda Trabzon, Gümüşhane, Rize ve Bayburt illerini kapsayan bir bölgenin orman yangını duyarlılık haritalarını üretmek için Random Forest ve yapay sinir ağı sınıflandırıcılarının performansları, Trabzon Orman Bölge Müdürlüğü tarafından sağlanan serbestçe erişilebilen Dünya gözlem verileri ve orman envanteri kullanılarak değerlendirilmiştir.

Bu çalışmalarda görüldüğü gibi Türkiye'nin farklı bölgelerindeki orman yangınları ele alınmıştır. Türkiye Orman Genel Müdürlüğü'nün verilerine göre; ülkemizin özellikle Hatay'dan başlayıp Akdeniz ve Ege sahil bölgelerinden İstanbul'a kadar uzanan kıyı şeridi, yangınlar açısından en riskli bölgeyi oluşturmaktadır [18]. İncelenen literatürde bu bölgeleri de kapsayan çalışmalar bulunmakla birlikte, bu çalışmalardan elde edilen sonuçlar farklılık göstermektedir. Bu nedenle en önemli hava koşullarının, Türkiye'de 2022 yılında çıkan orman yangınları üzerindeki etkileri T.C. Tarım ve Orman Bakanlığından alınan veriler doğrultusunda irdelenmiştir. Bu doğrultuda yanan alan verileri meteorolojik bilgilere dayalı kullanılarak, farklı makine öğrenmesi algoritmalarının performansları karşılaştırılmıştır. Modellerin performanslarını güvenilir bir şekilde değerlendirmek amacıyla çapraz doğrulama yöntemi uygulanmıştır. Bu amaçla, araştırmanın ikinci bölümünde kullanılacak veri seti ve uygulanacak ML yöntemleri detaylandırılmıştır. Üçüncü bölümde, veri seti üzerinde ML algoritmaları uygulanarak modellerin başarı oranları belirlenmiş ve karşılaştırmaları yapılmıştır. Son bölümde ise çalışmanın katkıları ve gelecekte yapılabilecek araştırmalar hakkında bilgi verilmiştir.

## 2 Materyal ve metod

Bu bölüm sınıflandırma işleminde kullanılan orman yangınları veri seti, veri ön işleme adımları ve sınıflandırma algoritmaları hakkında bilgiler içermektedir. Çalışmada kullanılan modelin akış şeması, ana hatlarıyla Şekil 1'de gösterilmektedir.

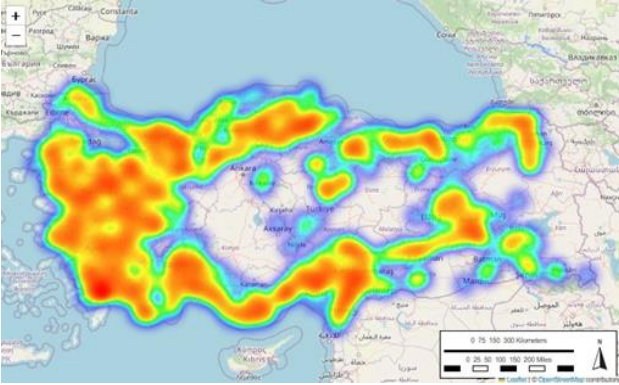


Şekil 1. Tahminde farklı modellerin karşılaştırılmasında kullanılan akış şeması

### 3 Veri seti

Çalışma kapsamında kullanılan veri seti T.C. Tarım ve Orman Bakanlığında alınan, Türkiye'deki 2022 yılında çıkan orman yangınları veri setidir. Bu veri seti 2011 adet veriden oluşmaktadır. Bunların 1555'i yangın var yani 1 ve 456'sı yangın yok yani 0 olmak üzere iki sınıftan meydana gelmiştir. Orman yangınlarının tahmin modeli için sıcaklık, nispi nem, rüzgar hızı (km/sa), rüzgar yönü, yükseklik (mt), yangının çıktığı bölge, yangının çıkış saati gibi farklı özelliklere dayalı değişkenler kullanılmıştır.

Şekil 2'de Türkiye'de 2022 yılında çıkan orman yangını verilerine göre yanan alan ısı haritası oluşturulmuştur. Buna göre, kırmızı alanlar riskli bölgeler, sarı alanlar orta riskli bölgeleri, yeşil ve mavi alanlar ise düşük riskli bölgeleri göstermektedir. Akdeniz bölgesinde özellikle Mersin ili taraflarında, Doğu Anadolu Bölgesinde Erzincan ili civarında, Ege Bölgesi'nde Muğla ilinde, özellikle de Marmaris çevresinde, Güneydoğu Anadolu Bölgesi'nde ise yangınların özellikle Mardin ve çevresinde yoğunlaştığı, Karadeniz Bölgesi'nde ise Çorum ilinde, Marmara Bölgesinde Çanakkale'de ve İç Anadolu Bölgesinde Eskişehir'de bölgedeki diğer illere göre gözle görülür bir yanan alan olduğu mevcuttur.

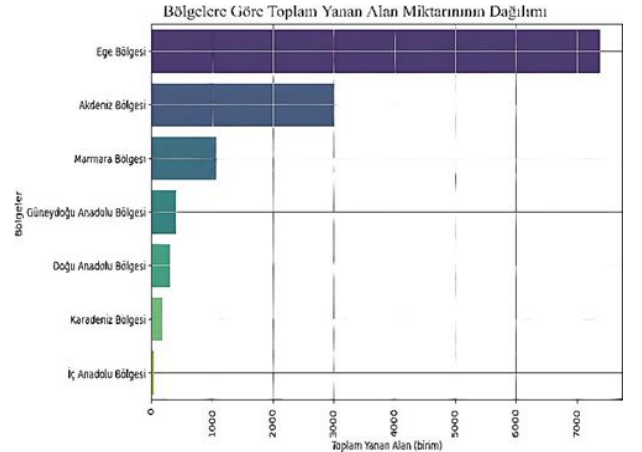


Şekil 1. Yanan alana göre ısı haritası

Ayrıca bölgelere göre yangınların dağılımı Şekil 3'te verilen grafikte de görülmektedir. Buna göre, T.C. Tarım ve Orman Bakanlığı verilerine göre 2022 yılında Akdeniz Bölgesi'nde 554 yangın, Doğu Anadolu Bölgesi'nde 79 yangın, Ege Bölgesi'nde 689 yangın, Güneydoğu Anadolu Bölgesi'nde 122 yangın, Karadeniz Bölgesi'nde 187 yangın, Marmara Bölgesi'nde 450 yangın ve İç Anadolu Bölgesi'nde 67 yangın meydana gelmiştir. Ege bölgesi yangın sayısında en yüksek bölgeye sahipken Doğu Anadolu, Güneydoğu Anadolu ve İç Anadolu bölgeleri, daha düşük yangın sayılarına sahiptir.



Şekil 3. Çıkan yangınların bölgesel dağılımı



Şekil 4. Bölgelere göre yanan alan miktarının dağılımı

Ormanlar; insanlığa gıda, yakıt, barınak, temiz hava ve su, ilaç, gelir kaynağı, istihdam, dinlenme, peyzaj gibi maddi-manevi birçok ekonomik, ekolojik, sosyokültürel faydalar sunan tabii bir kaynaktır. Bir ekosistem olarak orman, belirli bir kapalılıkta ağaçlar, diğer bitki ve hayvan topluluğu ile topraktaki gözle görülmeyen diğer organizmaların cansız çevreyle belli bir denge içinde karşılıklı olarak birbirleriyle Şekil 4'te görüldüğü üzere, çıkan yangınlar sonucunda en fazla yanan alan Ege Bölgesi'nde bulunmaktadır. Daha sonra Akdeniz, Marmara, Güneydoğu Anadolu, Doğu Anadolu, Karadeniz ve İç Anadolu bölgesi gelmektedir.

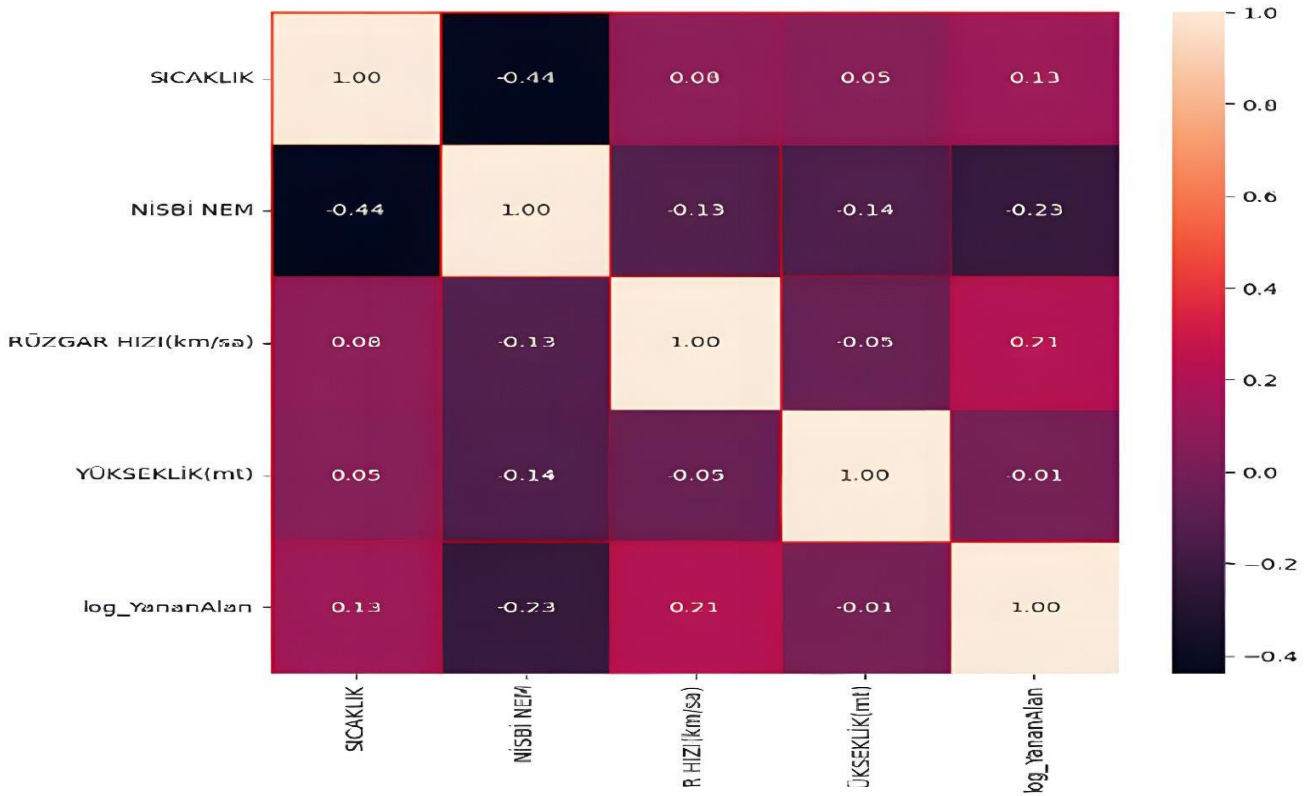
Çıkan orman yangınlarının aylık dağılımını incelediğimizde, Şekil 5'te görüldüğü üzere en çok yangın %13.6 ile Temmuz ayında çıkmakta, bunu %12.5 ile Eylül ayı izlemektedir. En az yangın ise %4.0 ile Aralık ayında meydana gelmiştir. Bu grafikten, en çok yangının yaz aylarında, en az yangının ise kış aylarında gerçekleştiği anlaşılmaktadır.



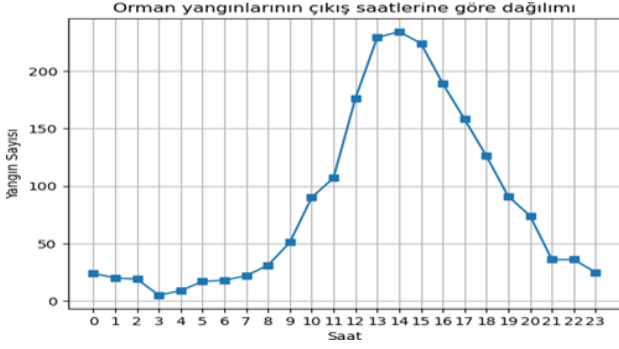
Şekil 5. Aylara göre yangınların dağılımı

Şekil 6 incelendiğinde, en yüksek ilişkinin sıcaklık ile nispi nem arasında olduğu görülmektedir; korelasyon katsayısı  $-0.44$ 'tür. Bu, sıcaklık arttıkça nispi nemin azaldığını gösterir. Sıcaklık ile yanan alan arasındaki korelasyon katsayısı ise  $0.13$  olup, sıcaklık arttıkça yanan alanın da artma eğiliminde olduğunu fakat bu ilişkinin zayıf olduğunu belirtir. Yanan alan ile nispi nem arasındaki korelasyon katsayısı  $-0.23$ 'tür, yani yanan alan arttıkça nispi nem azalma eğilimindedir. Ayrıca, yanan alan ile rüzgar hızı arasındaki korelasyon katsayısı  $0.21$ 'dir, bu da yanan alan ile rüzgar hızı arasında zayıf ve pozitif bir ilişki olduğunu göstermektedir.

Ayrıca yangınların çıkış saatlerini incelediğimizde, Şekil 7'de görüldüğü gibi, en çok yangının saat 13.00 ile 15.00 arasında, yani öğle vakitlerinde çıktığı görülmektedir.



Şekil 6. Değişkenlerin korelasyon matrisi



Şekil 7. Orman yangınlarının çıkış saatlerine göre dağılımı

### 3.1 Veri ön işleme

ML modellerini kurma aşamasında en önemli ve gerekli adımlardan biri veriyi hazır hale getirmektir. Veri ön işleme; veri seti üzerinde veri temizleme, dönüştürme, normalleştirme gibi yapılan birtakım işlemler ile verinin modele uygun hale getirilmesidir. Bu adımlar ML modelinin güvenilir, doğru ve başarılı tahminde bulunması için uygulanmaktadır [19]. Çalışmada kullanılan orman yangınları veri seti üzerinde modele katkısı olmayacak değişkenler silinmiştir. Veri setinde bazı değişkenlerin değerleri hatalı girilmiş olduğundan bu değerlerin bulunduğu satırlar silinmiştir. Aykırı verilerin (outlier) tespiti yapılarak bu veriler temizlenmiştir.

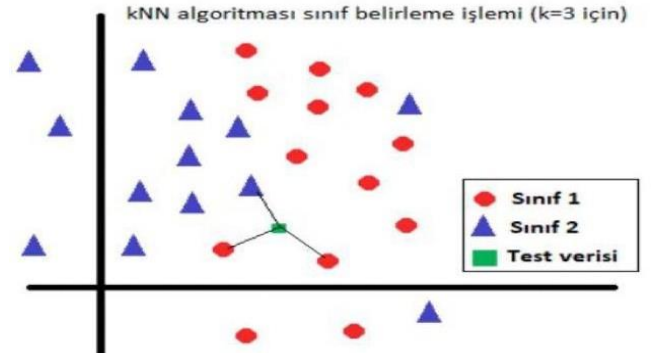
Yapılan çalışmada orman yangınları veri seti üzerinde 4 farklı ML teknikleri kullanılarak orman yangınının varlığı tahmin edilmiştir. Buna göre orman yangınının varlığının tahmini için kullanılan sınıflandırma algoritmaları aşağıda verilmiştir.

#### 3.1.1 Sınıflandırma algoritmaları

##### 3.1.1.1 KNN algoritması

K-En Yakın Komşuluk (KNN), denetimli bir öğrenme algoritması olup hem sınıflandırma hem de regresyon problemlerinde kullanılır. Temelde, bir veri noktasının tahmin edilmesi için KNN'lerini bulur ve bu komşuların değerlerini kullanarak tahmin yapar.

Sınıflandırma durumunda, veri noktası en yakın k komşusunun sınıfını alır ve çoğunluk oyu ile sınıflandırılır. Örneğin, bir nesne için en yakın k komşusunun sınıfları incelenir ve bu komşuların çoğunluğunun sınıfı, nesnenin sınıfı olarak tahmin edilir. Yani, "k" parametresi, hesaba katılacak en yakın nokta sayısını belirtir.



Şekil 8. KNN algoritması için örnek uygulama grafiği

Uzaklık hesaplamalarında genellikle Öklid, Manhattan veya Minkowski fonksiyonları kullanılır. Öklid fonksiyonu, en yaygın olarak kullanılan uzaklık ölçüsüdür ve veri noktaları arasındaki doğrudan mesafeyi hesaplar. Aşağıdaki Şekil 8'de, bu algoritmada kullanılan uzaklık ölçüleri ve k=3 olduğunda sınıflandırmanın nasıl yapıldığının bir grafik gösterimi verilmiştir [20, 21]

Lojistik regresyon sınıflandırma problemlerinde kullanılan ML modellerinin temellerinden birisidir. Lojistik regresyon, bağımsız değişkenlerden oluşan belirli bir veri kümesine dayalı olarak oy verme veya vermeme gibi bir olayın gerçekleşme olasılığını tahmin eder. Sonuç bir olasılık olduğundan, bağımlı değişken 0 ile 1 arasında sınırlandırılır. Lojistik regresyon, çoğunlukla ikili sınıflandırma problemleri için kullanılan bir denetimli öğrenme algoritmasıdır. "Regresyon" kelimesi ile "sınıflandırma" kelimesinin çeliştiği düşünülse de, burada dikkat çekilen kelime "lojistik"tir, yani lojistik fonksiyondur. Bu algoritma, sınıflandırma görevini bir fonksiyonla yapar. Lojistik regresyon, basit ancak çok etkili bir sınıflandırma algoritmasıdır, bu nedenle çok sayıda ikili sınıflandırma görevinde kullanılır. Lojistik regresyonda, olasılıklar üzerinde bir logit dönüşümü uygulanır; yani başarı olasılığı başarısızlık olasılığına bölünür. Bu, genellikle log odds veya odds'un doğal logaritması olarak da bilinir ve bu lojistik fonksiyon

$$p_i = \beta_0 + \beta_1 X_1 + \dots + \beta_m X_m \quad (1)$$

Olmak üzere,

$$\text{Logit}(p_i) = \frac{1}{1 + e^{-p_i}} \quad (2)$$

şekindedir. Burada  $p_i$ , olayın gerçekleşme olasılığını ifade ederken,  $\beta_0, \beta_1, \dots, \beta_m$  ise modeldeki bağımsız değişkenlerin olasılık üzerindeki etkilerini gösteren katsayılardır. Lojistik fonksiyon, modeldeki  $p_i$  değerini 0 ile 1 arasında bir olasılığa

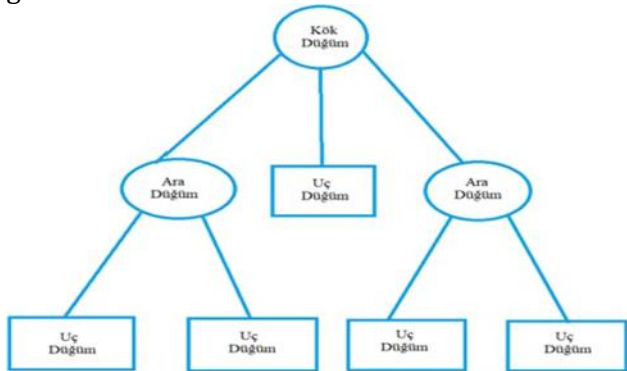
dönüştürür. Bu sayede lojistik fonksiyon, olayın gerçekleşme ihtimalini sınırlandırarak daha anlamlı bir olasılık elde edilmesini sağlar.

İkili sınıflandırma problemlerinde, 0.5'ten küçük olasılıkları 0 sınıfı, 0.5'ten büyük olasılıkları ise 1 sınıfı olarak tahmin eder. Bu yöntem, özellikle iki kategori arasındaki ayrımı yaparken oldukça etkili bir yaklaşım sunar ve lojistik regresyonun tahminleme kapasitesini artırır [23].

### 3.1.1.2 Karar ağaçları

Karar ağaçları, sınıfları bilinen örnek veriden tümevarım yöntemiyle öğrenilen ağaç şekilli bir karar yapısı çeşididir. Bir karar ağacı, basit karar verme adımları uygulanarak, büyük miktarlardaki kayıtları, çok küçük kayıt gruplarına bölerek kullanılan bir yapıdır. Her başarılı bölme işlemiyle, sonuç gruplarının üyeleri bir diğeriyle çok daha benzer hale gelmektedir [24].

Karar ağaçlarının temel bileşenleri olan kök, dallar ve yapraklar, veri setini işlemenin farklı aşamalarını temsil eder. Kök düğüm, bağımlı değişkeni temsil eder ve ağacın ilk karar noktasını oluşturur. Dallar, farklı öznelik değerlerine karşılık gelir ve veriyi bu öznelikler üzerinde bölerek ağacın derinleşmesini sağlar. Yapraklar ise sonuçları temsil eder ve nihai sınıflandırmayı veya tahmini sağlar [25]. Aşağıdaki Şekil 9'da karar ağaçlarının grafiğini göstermektedir.



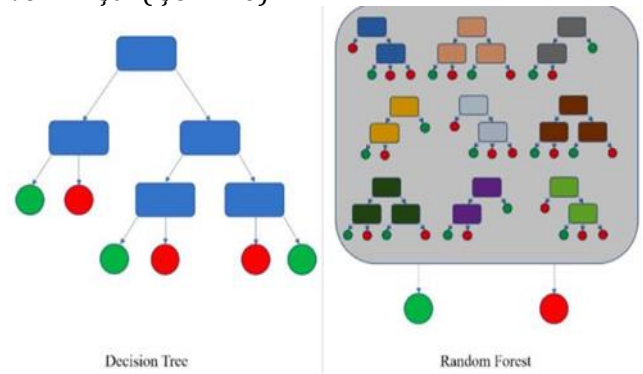
Şekil 9. Karar ağacı algoritması ile sınıf belirleme

Karar ağaçları, sınıflandırma problemlerini çözmek için kullanılan etkili bir araçtır. Karar ağaçları, genel olarak anlaşılması ve yorumlanması en kolay olan sınıflandırma algoritmalarından biridir. Bu özellikleri, karar ağaçlarını veri bilimi ve ML alanında popüler bir tercih haline getirir [26].

### 3.1.1.3 Rastgele orman

Rastgele orman modeli, sınıflama ve regresyon özelliklerini bir arada barındıran bir algoritmadır. Bu algoritma, karar ağaçlarının birleşiminden oluşur. Rastgele orman modeli, 2001 yılında Leo

Breiman tarafından geliştirilmiş ve rastgele seçilen verilerin alt uzaylarında büyüyen karar ağaçlarında bir tahmin kümesi oluşturmak ve bu ağaçlara dayanan bir ML metodudur. Bu yöntem sınıflandırma ve regresyon işlemlerinde oldukça sık tercih edilen bir veri madenciliği yöntemidir [27]. Rastgele ormanlar, ağaç tipi sınıflandırıcılar topluluğudur. Ormandaki her ağaç bağımsız olarak örneklenen bir rasgele vektörün değerlerine ve aynı dağılıma bağlıdır. Torbalama yönteminin gelişmiş bir şekli olarak kabul edilen rastgele ormanlar, ağaçların topluluğu üzerinden model oluşturur. Bu yöntemde torbalama ve rastgele değişken seçimi birlikte kullanılır. Her yeni eğitim seti, orijinal eğitim setinden iadeli olarak çekilir ve rastgele değişken seçimi kullanılarak yeni eğitim setinde bir ağaç yetiştirilir. Rastgele orman modelinin önemli avantajlarından biri, çok sayıda girdi değişkenini ele alabilmesidir. Ayrıca, sınıflandırma ve regresyon işlemlerinde yüksek doğruluk oranları sağlar. Ancak, dezavantaj olarak algoritmanın siyah kutu olması gösterilebilir, yani ağaç yapısını net bir şekilde görmek mümkün değildir. Aşağıda şekilde de Karar Ağacı ve Rastgele Orman modelinin birlikte şeması verilmiştir. Aşağıda şekilde de Karar Ağacı ve Rastgele Orman modelinin birlikte şeması verilmiştir( Şekil 10).



Şekil 10. Karar ağacı ve rastgele orman algoritması

## 4 Bulgular ve Tartışma

Bu çalışmada oluşturulan model doğrulama yöntemleri doğruluk oranı (accuracy rate) ve kesinlik (precision) değerleri kullanılarak hesaplanmıştır. Oluşturulan modelin değerlendirme kriterleri Tablo 1'de açıklamaları ve formülleri ile verilmiştir.

Tablo 1. Sınıflandırma modelleri değerlendirme kriterleri

DEĞERLENDİRME KRİTERLERİ	AÇIKLAMA	FORMÜL
DOĞRULUK ORANI (ACCURACY RATE):	Modeldeki hedef sınıfları doğru tahmin etme durumudur.	(Doğru Pozitif+Doğru Negatif)/ N
KESİNLİK (PRECISION):	Elde edilen modeldeki sonuçların ne kadar doğru olduğunu gösteren değerdir.	Doğru Pozitif/(Doğru Pozitif+Yanlış Pozitif)

**Doğru Pozitif :** Gerçek durum değeri pozitifdir (yangın var) ve kullanılan yöntemle test sonucu pozitif (yangın var) olarak tahmin edilmiştir. Yani yangın var olarak doğru tahmin edilmiştir.

**Doğru Negatif :** Gerçek durum değeri negatiftir (yangın yok) ve kullanılan yöntemle test sonucu negatif (yangın yok) olarak tahmin edilmiştir. Yani yangın yok olarak doğru tahmin edilmiştir.

**Yanlış Pozitif :** Gerçek durum değeri negatiftir (yangın yok) fakat kullanılan yöntemle test sonucu pozitif (yangın var) olarak tahmin edilmiştir. Yani yangın var olarak yanlış tahmin edilmiştir [28].

N: Toplam durum sayısı

Ön işleme adımları sonrası veri seti, %70'i eğitim verisi ve %30'i test verisi olacak şekilde ikiye ayrılmıştır. Bu veri seti üzerinde bağımlı değişken yangın durumu olmak üzere ikili değerden (1 yangın var, 0 yangın yok) oluştuğundan ikili sınıflandırma işlemi gerçekleştirilmiştir. 7 bağımsız değişken (sıcaklık, nispi nem, rüzgar hızı, yükseklik, bölge, rüzgar yönü, saat ) ile sınıflandırma işlemi yapılmıştır. Model doğrulama yöntemleri olarak doğruluk (accuracy) oranı ve kesinlik(precision) oranı elde edilmiştir. Sınıflandırma modellerinin başarı değerlendirmesi sonuçları Tablo 2'de gösterilmiştir.

Tablo 2. Sınıflandırma modelleri için başarı değerlendirmesi

Model Doğrulama Yöntemleri	KNN	Lojistik Regresyon	Karar Ağacı	Rastgele Orman
Doğruluk (Accuracy)	%77.48	%77.98	%77.81	%76.15
Kesinlik (Precision)	%78.00	%78.30	%78.07	%78.18

Tablo 2'den görüldüğü üzere, en başarılı sonuçlar Lojistik Regresyon modeli ile elde edilmiştir. Ayrıca modellerin doğruluk değerlerine bakıldığında, KNN ve Karar Ağacı modelleri de Lojistik Regresyon modeline yakın sonuçlar verirken, Rastgele Orman modeli diğer modellere göre en düşük performansı göstermiştir. Modellerin kesinlik değerleri incelendiğinde, tüm modellerin hemen hemen yakın sonuçlar verdiği gözlemlenmiş olmakla birlikte, %77.30'luk kesinlik oranı ile en yüksek performansı Lojistik Regresyon modelinin sağladığı tespit edilmiştir. Bu bulgu, Lojistik Regresyon modelinin orman yangınlarının tahminlenmesinde diğer yöntemlere göre biraz daha etkin olduğunu ortaya koymakta ve ileriki çalışmalarda bu modelin öncelikli olarak kullanılması gerektiğini göstermektedir.

Çalışmanın bulguları, literatürdeki diğer araştırmalarla karşılaştırıldığında, Türkiye'de kullanılan algoritmaların başarı oranlarının, farklı veri setleri ve koşullara göre değişiklik gösterdiği anlaşılmaktadır. Örneğin, Abdessemed ve ark. [29] çalışmasında Cezayir orman yangınları veri seti için Rastgele Orman modeli için %93.87 doğruluk oranı ve %100 kesinlik oranı; Karar Ağacı modeli için %95.91 doğruluk oranı ve %100 kesinlik oranını elde etmiştir. Stojanova ve arkd. [23] yaptıkları çalışmada, Slovenya'nın üç farklı bölgesindeki orman yangınları veri setini kullanarak orman yangınlarını bu bölgelere göre tahmin etmişlerdir. Çalışmaya göre, Kras bölgesi için Lojistik Regresyon modelinin doğruluk ve kesinlik oranları sırasıyla %77.2 ve %69.6; Rastgele Orman modeli için doğruluk ve kesinlik oranları sırasıyla %79.7 ve %75.1 olduğu görülmüştür. Primorska Bölgesi için Lojistik Regresyon modelinin doğruluk ve kesinlik oranları sırasıyla %83.4 ve %82.6; Rastgele Orman modelinin doğruluk ve kesinlik oranları sırasıyla %85.2 ve %82 olduğu görülmüştür. Kitasal Slovenya bölgesi için Lojistik Regresyon modelinin doğruluk ve kesinlik oranları sırasıyla %84 ve %83.1; Rastgele Orman modelinin doğruluk ve kesinlik oranları sırasıyla %84 ve %82.3 olduğu görülmektedir. Bu farklılıklar, çalışmalarda kullanılan veri setlerinde kullanılan özelliklere bağlı olarak coğrafi koşullardan kaynaklanabilir. Çalışma kapsamında veri setinde hava koşulları yer almakla birlikte, model performansını daha da artırmak amacıyla çeşitli iyileştirmeler yapılabilir. Örneğin, bitki örtüsü türü ve toprak nem içeriği gibi ek özellikler modele dahil edilebilir. Lojistik Regresyon modeli, basit yapısına rağmen, yangın tahmini gibi hızlı ve doğru tahmin gerektiren



uygulamalarda kullanışlı bir seçenek olmaktadır. Ancak, daha karmaşık veya doğrusal olmayan veri setleri için başka modelleme yöntemlerinin de denenmesi gerektiği unutulmamalıdır. Çalışmada oluşturulan makine öğrenmesi modeli ile yeni verileri sürekli işleyerek orman yangın risk tahminlerini güncelleyen otonom sistem ve alarm sistemleri oluşturulabilir. Otonom sistemlerin ve alarm sistemlerinin birlikte entegre edilmesiyle, orman yangınlarını önleyici tedbirler ve hızlı müdahale yöntemleri için kullanılacak bilginin güncel kalması sağlanabilir. Bu açıdan incelendiğinde makine öğrenmesi ile gelecekte orman yangınlarının tespitinde orman yangınlarının erken saptanması ve yangın dirençliliğini arttırması açısından daha sürdürülebilir bir ilerleme kaydedilebilir.

## 5 Sonuç

Bu çalışmada, Türkiye'deki orman yangınlarının tahmini için 2022 yılına ait veriler kullanılarak bir model oluşturulmuştur. Veri seti, ön işleme adımları uygulanarak veri sınıflandırma modellerine hazır hale getirilmiştir. KNN, Lojistik Regresyon, Karar Ağacı ve

Rastgele Orman olmak üzere dört farklı ML algoritmasının performansları karşılaştırılmıştır. Sonuçlar, en başarılı modelin %77.98 ile Lojistik Regresyon modeli olduğunu göstermiştir. Bununla birlikte, tahmin başarısını artırmak amacıyla meteorolojik veriler ve eski yangın kayıtları gibi ek parametreler modele dahil edilebilir. Bu yaklaşım, modelin daha kapsamlı hale gelmesini sağlayarak yangınların önceden belirlenmesinde daha etkili bir araç olmasını mümkün kılabilir. Bu çalışma, orman yangınlarının önlenmesi ve etkilerinin azaltılması ile ilgili gelecekte yapılacak araştırmalar için bir kaynak olarak literatüre katkı sağlayabilir. Ayrıca gelecek araştırmalarda, farklı makine öğrenimi teknikleri kullanılabilir ve yangınların yayılma hızını ve yönünü tahmin etmek için mekânsal ve zamansal veriler entegre edilebilir. Bu şekilde daha fazla parametre ile daha yüksek başarı oranına ulaşılabilir.

## 6 Teşekkür

Makalemizi gözden geçirip geliştirmek için değerli önerilerini bizimle paylaştığı için hakemlere çok teşekkür ederiz. Bu çalışma, Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından 2209-A numaralı proje kapsamında desteklenmiştir. TÜBİTAK'a sağladıkları destek için teşekkür ederiz. Ayrıca, araştırma sürecinde katkıda bulunan tüm kurum ve kişilere teşekkürlerimizi sunarız.

## 7 Kaynaklar

- [1] T.C. Orman ve Su İşleri Bakanlığı (2015). Türkiye Orman Varlığı. Orman Genel Müdürlüğü. [https://www.ogm.gov.tr/tr/ormanlarimizsitesi/TurkiyeOrmanVarligi/Yayinlar/2015%20T%C3%BCrkiye%20Orman%20Varligi%20C4%B1%C4%9F%C4%B1.pdf]. (Erişim Tarihi: 10 Kasım 2023)
- [2] Ünal S. "Orman yangınlarından doğan zarar düzeyinin hesaplanması üzerine araştırmalar". Journal of the Faculty of Forestry Istanbul University, 41(1), 2014.
- [3] Dayananda PWA. "Stochastic models forest fires". Ecological Modeling, 3, 309-313, 1977.
- [4] Küçük Ö, Sağlam B. "Orman yangınları ve hava halleri". Kastamonu Üniversitesi Orman Fakültesi Dergisi, 4(2), 220-231, 2004.
- [5] Altan G, Türkeş M, Tatlı H. "Çanakkale ve Muğla 2009 yılı orman yangınlarının Keetch-Byram Kuraklık İndisi ile klimatolojik ve meteorolojik analizi". In: 5th Atmospheric Science Symposium Proceedings Book: 263-274, 2011.
- [6] Cortez P, Morais ADJR. "A data mining approach to predict forest fires using meteorological data". 13th EPIA Portuguese Conference on Artificial Intelligence, 2007.
- [7] Bayat G, Yıldız K. "Comparison of the machine learning methods to predict wildfire areas". Turkish Journal of Science and Technology, 17(2), 241-250, 2022.
- [8] Chen Y, Zhang Y, Xin J, Wang G, Mu L, Yi Y, Liu D. "UAV image-based forest fire detection approach using convolutional neural network". IEEE conference on industrial electronics and applications (ICIEA), 2019.
- [9] Beşli N, Tenekeci E. "Uydu verilerinden karar ağaçları kullanarak orman yangını tahmini". Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, 11(3), 899-906, 2020.
- [10] Sayad YO, Mousannif H, Moatassime HA. "Predictive modeling of wildfires: A new dataset and machine learning approach". Fire Safety Journal, 104, 130-146, 2019.
- [11] Mohajane M, Costache R, Karimi F, Pham QB, Essahlaoui A, Nguyen H, Laneve G, Oudija F. "Application of remote sensing and machine learning algorithms for forest fire mapping in a Mediterranean area". Ecological Indicators 129:107869, 2021.
- [12] Kukuk SB, Kilimci ZH. "Comprehensive analysis of forest fire detection using deep learning models and conventional machine learning algorithms". International Journal of Computational and Experimental Science and Engineering, 7(2), 84-84, 2021.
- [13] Yıldırım O, Günay FB, Yağanoğlu M. "Makine öğrenmesi yöntemleriyle orman yangını tahmini". Journal of the Institute of Science and Technology, 13(3), 1468-1481, 2023.
- [14] Akıncı HA, Akıncı H. "Machine learning based forest

- fire susceptibility assessment of Manavgat district (Antalya), *Turkey*". *Earth Science Informatics*, 16(1), 397-414, 2023.
- [15] Iban MC, Sekertekin A. "Machine learning based wildfire susceptibility mapping using remotely sensed fire data and GIS: A case study of Adana and Mersin provinces, Turkey". *Ecological Informatics*, 69, 101647, 2022.
- [16] Kantarcioglu O, Kocaman S, Schindler K. "Artificial neural networks for assessing forest fire susceptibility in Türkiye". *Ecological Informatics*, 75, 102034, 2023a.
- [17] Kantarcioglu O, Schindler K, Kocaman S. "Forest Fire Susceptibility Assessment with Machine Learning Methods in North-East Turkey". *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48, 161-167, 2023b.
- [18] Ayanoğlu S, Dölarıslan M, Gül E. "Sadece bir yangın mı? Ekolojik ve sosyo-ekonomik açıdan orman yangınları". *Türk Bilimsel Derlemeler Dergisi*, 10(2), 32-35, 2017.
- [19] Oğuzlar A. "Veri ön işleme". *Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 21, 2003.
- [20] Keleş MB, Keleş A, Keleş A. "Makine öğrenmesi yöntemleri ile uçuş fiyatlarının tahmini". *Euroasia Journal of Mathematics, Engineering, Natural & Medical Sciences*, 7(11), 72-78, 2020.
- [21] Zilyas D, Yılmaz A. "Makine öğrenmesi yöntemleri ile eğitim başarısının tahmini modeli". *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 14(3), 437-447, 2023.
- [22] Kumaş E. "Türkçe twitter verilerinden duygu analizi yapılırken sınıflandırıcıların karşılaştırılması". *Eskişehir Türk Dünyası Uygulama ve Araştırma Merkezi Bilişim Dergisi*, 2(2), 1-5, 2021.
- [23] Stojanova D, Panov P, Kobler A, Džeroski S, Taškova K. "Learning to predict forest fires with different data mining techniques". In *Conference on data mining and data warehouses*, Ljubljana, Slovenia, 2006.
- [24] Albayrak YS, Yılmaz ÖK. "Veri madenciliği: Karar ağacı algoritmaları ve İMKB verileri üzerine bir uygulama". *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 14(1), 31-52, 2009.
- [25] Harrington P. *Machine Learning In Action. Shelter Island: Manning Publications*, 2012.
- [26] Aydın SG, Aydoğdu G. "Makine öğrenmesi algoritmaları kullanılarak Türkiye ve AB ülkelerinin CO2 emisyonlarının tahmini". *Avrupa Bilim ve Teknoloji Dergisi*, 37, 42-46, 2022.
- [27] Parlak MS, Kayri M. "Öğretmenlerin e-öğrenme hazır bulunuşluk düzeylerini etkileyen faktörlerin rastgele orman algoritması yöntemi ile incelenmesi". *Van Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 19(3), 670-696, 2022.
- [28] Harman G. "Destek vektör makineleri ve naive bayes sınıflandırma algoritmalarını kullanarak diabetes mellitus tahmini". *Avrupa Bilim ve Teknoloji Dergisi*, (32), 7-13, 2021.
- [29] Abdessemed F, Bouam S, Arar C. "Forest Fire prediction using Machine Learning Methods", *First National Conference in Computer Science Research and its Applications*, 2023.