

TOPLULUK-TABANLI MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE ORMAN YANGINI DUYARLILIK HARİTALARININ ÜRETİMİ: MANAVGAT (ANTALYA) ÖRNEĞİ

Halime Dilara Oğuz¹, İrem Üçler², Taşkın Kavzoğlu³, Alihan Teke⁴

¹Lisans Öğrencisi, Gebze Teknik Üniversitesi, Harita Mühendisliği Bölümü, 41400, Gebze, Kocaeli, h.oguz2020@gtu.edu.tr

²Lisans Öğrencisi, Gebze Teknik Üniversitesi, Harita Mühendisliği Bölümü, 41400, Gebze, Kocaeli, i.uculer2020@gtu.edu.tr

³Prof. Dr., Gebze Teknik Üniversitesi, Harita Mühendisliği Bölümü, 41400, Gebze, Kocaeli, kavzoglu@gtu.edu.tr

⁴Arş. Gör., Gebze Teknik Üniversitesi, Harita Mühendisliği Bölümü, 41400, Gebze, Kocaeli, a.teke2020@gtu.edu.tr

ÖZET

Orman yangınları, doğal veya insan kaynaklı nedenlerle ormanlık alanlarda çıkan ve geniş bölgelere yayılarak insan yaşamına, bitki örtüsüne ve ekosistemlere zarar veren afetlerdir. Son zamanlarda dünya çapında özellikle iklim değişikliği ve küresel ısınma gibi faktörlerin etkisi ile birlikte orman yangınlarının şiddeti ve sıklığında da artış görülmektedir. Orman yangını duyarlılık haritaları afet planlarının hazırlanması ve afet sonrası acil durum yönetimi için literatürde sıklıkla tercih edilen araçlardandır. Bu çalışmada, Antalya ili Manavgat ilçesi (Türkiye) için NGBoost, XGBoost ve Rastgele Orman olmak üzere üç makine öğrenmesi algoritması ile orman yangını duyarlılık haritaları üretilmiştir. Orman yangını envanter haritası, Orman Genel Müdürlüğü'nden temin edilmiş ve çalışma alanının ana karakteristiğini yansıtan toplam 11 parametre bağımsız değişken olarak kullanılmıştır. Doğruluk değerlendirme sonuçları, NGBoost'un %81.42 genel doğruluk ve % 86.23 eğri altında kalan alan (AUC) skoru ile XGBoost ve Rastgele Orman algoritmasına kıyasla daha yüksek tahminleme performansına sahip algoritma olduğunu göstermiştir. Üretilen duyarlılık haritalarının doğrulukları arasındaki farkın istatistiksel olarak anlamlı olup olmadıklarını belirlemek için McNemar testi uygulanmıştır. İstatistiksel analiz sonuçları NGBoost ve RF algoritmalarının performanslarının birbirinden istatistiksel olarak anlamlı olduğunu ancak diğer algoritmaların performansları arasındaki farkın istatistiksel olarak anlamlı olmadığını ortaya koymuştur. Orman yangını duyarlılık haritaları incelendiğinde, orman yangını duyarlılığının genellikle çalışma alanının orta ve güney bölgelerinde yüksek olduğu, buna karşın çalışma alanının kuzey ve kuzeybatı kısımlarında orman yangını duyarlılığının düşük ve çok düşük sınıflara karşılık geldiği gözlemlenmiştir.

Anahtar Sözcükler: duyarlılık haritalama, makine öğrenmesi, NGBoost, orman yangını, XGBoost

ABSTRACT PRODUCTION OF FOREST FIRE SUSCEPTIBILITY MAPS USING ENSEMBLE MACHINE LEARNING ALGORITHMS: CASE STUDY OF MANAVGAT, ANTALYA

Forest fires are disasters that typically occur in woodlands due to natural or human-induced reasons, spreading over large areas and causing damage to human life, vegetation, and ecosystems. Recently, there has been an increase in the intensity and frequency of forest fires across the world, particularly due to the climate change and global warming. Forest fire susceptibility maps are frequently used in the literature for preparing disaster plans and managing emergencies post-disaster when addressing forest fire management. In this study, forest fire susceptibility maps for the Manavgat district of Antalya province (Türkiye) were produced using three machine learning algorithms: NGBoost, XGBoost, and Random Forest. The forest fire inventory map was obtained from the General Directorate of Forestry, and a total of 11 parameters reflecting the main characteristics of the study area were used as independent variables. The accuracy assessment results revealed that the NGBoost, with an overall accuracy of 81.42% and an area under the curve (AUC) score of 86.23%, had higher predictive performance compared to the XGBoost and Random Forest algorithms. McNemar's test was applied to determine whether the differences in the accuracies of the produced susceptibility maps were statistically significant. The statistical analysis results showed that the performance difference between the NGBoost and RF algorithms was statistically significant, whereas the performance differences between the other algorithms were statistically insignificant. The thematic assessment of the forest fire susceptibility maps revealed that the forest fire susceptibility was generally higher in the central and southern regions of the study area, while it corresponded to low and very low classes in the northern and northwestern parts of the study area.

Keywords: forest fire, machine learning, NGBoost, susceptibility mapping, XGBoost.

1. GİRİŞ

Ormanlar yaşamın her alanında canlılar için en kritik ekolojik değerlerden biridir (Yıldırım vd., 2023). Dünyanın en büyük karasal ekosistemi olan ormanlar hava ve gürültü kirliliğinin azaltılmasından atmosferik döngünün sağlanmasına kadar doğal yaşamın sürdürülebilirliği için gerekli en önemli ekosistemlerden birisidir. Özellikle son zamanlardaki iklim değişikliği ve küresel ısınma gibi faktörlerin etkisi ile orman yangınları sıklığını, şiddetini ve etkisini en fazla gösteren doğal afetlerden birisi olarak karşımıza çıkmaktadır. Orman yangınları, hem doğal ekosistemler hem de insan toplulukları üzerinde yıkıcı etkilere sahip olan, hızla yayılan ve çoğu zaman kontrol altına

alınması zor doğal afetlerden birisidir. Orman yangınları biyolojik çeşitliliğin azalmasından toprağın verimliliğinin kaybolmasına, ekonomik kayıplardan sosyolojik sorunlara kadar geniş bir yelpazede ciddi zararlara sebebiyet verebilmektedir. Bunlara ek olarak, artan sıcaklıklar ve azalan yağışlar nedeniyle daha sık ve şiddetli hale gelen orman yangınları, sürdürülebilir kalkınma ve ekolojik denge açısından da önemli riskler taşımaktadır. Özellikle Akdeniz Bölgesi'nde küresel ısınma ve iklim değişikliği gibi faktörlerin de etkisi ile birlikte şiddetini ve sıklığını artıran orman yangınları, insan yaşamını ve ekosistemleri tehdit eden ciddi bir sosyal ve çevresel sorun haline gelmiştir (Kavzoğlu vd., 2021). Bu nedenle, orman yangınlarının neden olduğu tahribatı anlamak ve bu tür felaketlere karşı etkili önlemler geliştirmek, hem mevcut doğal kaynakların korunması hem de geleceğe yönelik sürdürülebilir bir yaşam için kritik öneme sahiptir.

Orman yangınlarından kaynaklanan yukarıda bahsedilen olumsuz etkiler araştırmacıları politika geliştirme (Xanthopoulos 2007) ve erken uyarı sistemi planlaması (Divya vd., 2019; Sefercik ve Kavzoglu, 2021) gibi bir dizi önlemler almaya yöneltmiştir. Bu öneriler arasında, duyarlılık haritalarının üretimi yukarıda belirtilen olumsuz etkileri rasyonel bir şekilde önlemenin yanında afet planlarının hazırlanması ve afet sonrası acil durum yönetimi için de önemli bir altlık olarak kullanılmaktadır. Bunlara ek olarak, duyarlılık haritalarının üretimi işleminin maliyeti çoğu zaman alternatiflerine göre çok daha düşüktür. Bu nedenlerden dolayı, yerel arazi koşullarına bağlı olarak belirli bir coğrafi konumda orman yangını meydana gelme olasılığını mekânsal olarak gösteren orman yangını duyarlılık haritalarının üretimi son zamanlarda alternatiflerine kıyasla literatürde en çok kullanılan yöntemlerden birisi haline gelmiştir (Tonbul vd., 2016; Iban ve Sekertekin, 2022; Bilucan vd., 2024).

Orman yangını aktivitelerine sebep olan faktörlerin birbirleriyle ilişkili karmaşık doğaları ve doğrusal olmayan davranışları göz önünde bulundurulduğunda, üretilen duyarlılık haritalarının etkinliği ve güvenilirliği esas olarak yangınların muhtemel konumsal dağılımlarını yansıtabilecek uygun ve etkin tekniklerin seçimine bağlıdır. Özellikle son yıllarda uzaktan algılama, yapay zeka ve coğrafi bilgi sistemlerindeki gelişmeler, orman yangını duyarlılık haritalama çalışmalarına olan ilgiyi de artırmıştır. Bu gelişmelere paralel olarak, araştırmacılar birçok farklı teknik kullanarak orman yangını duyarlılık haritaları üretimi için yoğun çaba sarf etmişlerdir. Bu teknikler istatistiksel, deterministik ve sezgisel olmak üzere üç kategoride özetlenebilir. Bunlara ek olarak, veri madenciliği ve makine öğrenmesi algoritmaları ise son zamanlarda orman yangını duyarlılık çalışmalarının odak noktası haline gelmiştir. Günümüze değin, karar ağacı (Beşli ve Tenekeci, 2020; Rihan vd., 2023), lojistik regresyon (Atmaca vd., 2022), destek vektör makineleri (Sayad vd., 2019; Xie vd., 2022), Naive Bayes (Ge vd., 2022; Piao vd., 2022) ve yapay sinir ağları (Trucchia vd., 2022; Akıncı ve Akıncı, 2023) olmak üzere birçok farklı makine öğrenmesi tekniği duyarlılık haritalarının üretiminde yaygın olarak kullanılmıştır. Geleneksel makine öğrenmesi algoritmaları bazı durumlarda belirli bir dereceye kadar başarı elde etmiş olsalar da, tatmin edici veya üstün başarıya ulaşma konusunda birtakım dezavantajlara sahiptirler (Dong vd., 2020). Özellikle tekil algoritmalar sınırlı veri kümesi boyutları ve özelliklerinin karmaşıklığı nedeniyle çoğu zaman genelleme yetenekleri sınırlıdır. Ayrıca, bu algoritmaların çoğu, yüksek boyutlu veri setlerinde hesaplama maliyetlerini ve zaman gereksinimlerini artırabilir. Veri üzerinde yeterince kapsamlı analiz yapamadıkları için, modelin doğruluğunu ve güvenilirliğini etkileyebilecek aşırı uyum (overfitting) veya yetersiz uyum (underfitting) sorunlarıyla karşılaşabilirler.

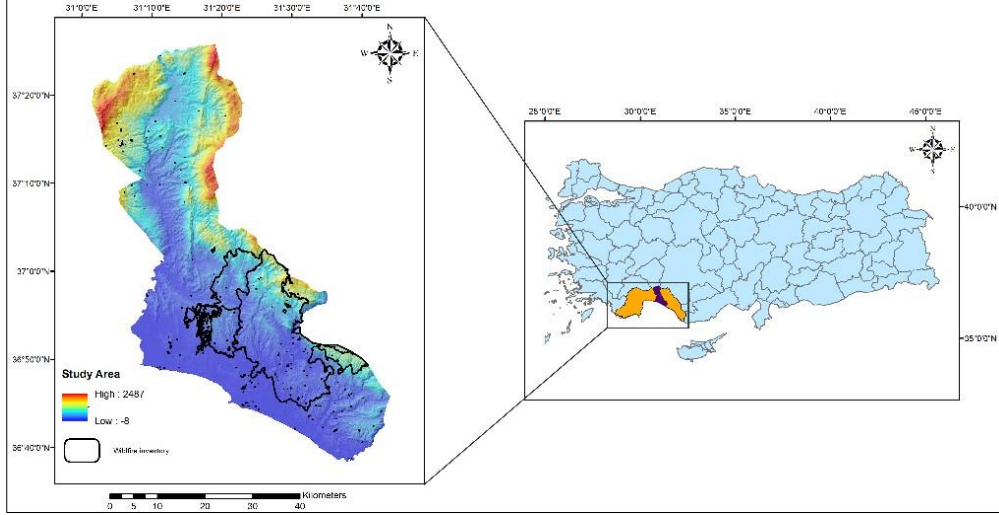
Birçok disiplinde uzun süredir var olan ancak son zamanlarda popüler bir kavram haline gelen topluluk öğrenimi, orman yangını duyarlılığı çalışmalarında önemli bir ivme kazanmıştır. Topluluk öğrenme tekniklerinin altında yatan temel ilke, daha iyi tahmine dayalı performanslar elde etmek için farklı oylama sistemleri kullanarak birden fazla makine öğrenmesi tekniğinin tahminlerini birleştirmektir. Böylece, herhangi bir geleneksel makine öğrenmesi tekniğinin hataları veya önyargıları muhtemelen diğer tekniklerle telafi edilecektir ve topluluk öğrenimi, tek makine öğrenmesi algoritmasından daha kararlı modeller elde edebilir. Bu çalışma kapsamında, 28 Temmuz 2021'de başlayıp yaklaşık iki hafta devam eden orman yangınlarıyla 47.824 ha ormanlık alanın yandığı Antalya'nın Manavgat ilçesi çalışma bölgesi olarak seçilmiştir. Bölgeye ait yangın duyarlılık haritası oluşturmak amacıyla NGBoost, XGBoost ve Rastgele Orman olmak üzere üç makine öğrenmesi algoritması ve orman yangınlarının oluşumunda önemli rol oynayan 11 faktör kullanılmıştır. İlk olarak, faktörler arasında potansiyel bir korelasyon olup olmadığı çoklu bağlantı testi ve korelasyon analizi uygulanarak araştırılmıştır. Daha sonra, üç farklı topluluk tabanlı makine öğrenmesi algoritmaları kullanılarak üretilen duyarlılık haritalarının doğrulukları arasındaki farkın istatistiksel olarak anlamlı olup olmadıklarını belirlemek için McNemar testi uygulanmıştır.

2. ÇALIŞMA ALANI VE VERİSETİ

Çalışma alanı olarak seçilen Manavgat ilçesi Türkiye'nin Batı Akdeniz bölgesinde bulunan Antalya ilinde yer almaktadır (Şekil 1). Yaz aylarında ortalama sıcaklık 28°C'ye kadar yükselirken, kış aylarında ise ortalama sıcaklık 8°C'ye kadar düşmektedir. Yıllık yağış miktarı ortalama 700 mm olan Akdeniz Bölgesi'nin yıllık ortalama sıcaklığı ise 18.5 °C'dir. Yazları sıcak ve kuru olan bölgede tipik olarak kış aylarında daha fazla yağış alır. Manavgat, Toros Dağlarının'nın eteklerine kadar uzanan bir ovanın üzerinde bulunmaktadır. Ormanlar, vadiler ve akarsuların olduğu bölgede kızılçam, fıstıkçamı ve diğer iğne yapraklı ağaç türleri bir araya gelerek zengin bir meşcere oluşturur. Çalışma alanı olan Manavgat ilçesinde yazların sıcak ve kuru geçmesi bölgeyi orman yangınları açısından daha riskli bir hale getirmektedir. 2019-2021 yılları arasında görülen orman yangınlarında yaklaşık 6.171 ha alan orman yanmıştır. Bu yangınların nedenleri nedeni bilinmeyen yangınlar, kasıtlı çıkarılmış yangınlar, ihmal ve dikkatsizlikten kaynaklanan yangınlar, kaza ile çıkarılan yangınlar ve doğal nedenlerden kaynaklanan yangınlardır. Bunların arasından en sık

görülmüş olanı ise nedeni bilinmeyen yangınlardır. İstatistiksel verilere göre çıkmış olan yangınların yaklaşık %50'si tehlikesiz yangın olarak bilinmektedir.

2019-2021 yıllarında gerçekleşen orman yangını alanlarını içeren envanter haritası, Orman Genel Müdürlüğü'nden temin edilmiştir. Bu veriye göre, çalışma alanı sınırları içerisinde 197 sayıda orman yangını aktivitesi gözlenmiş ve bu yangınlar toplamda yaklaşık olarak 6.171 hektar ormana zarar verdiği görülmüştür. Alansal olarak en büyük aktivite 42.725 hektarlık bir alana etki etmiş ve çalışma alanının güney kesiminde oluşmuştur. 427.25 km² ile en az alana etki eden aktivite ise çalışma alanının kuzey kesiminde gözlemlenmiştir.



Şekil 1. Çalışma alanının coğrafi konumu.

3.YÖNTEM

Literatürde duyarlılık haritaları üretiminde çok sayıda çeşitli yaklaşım bulunmaktadır. Bu çalışma kapsamında orman yangını duyarlılık haritası üretim sürecinde izlenen iş akışı toplamda beş temel adımdan oluşmaktadır. İlk aşama orman yangınına sebep olan ana faktörlerin ve parametrelerin test sahasının karakteristiğine uygun bir şekilde belirlenmesini ve envanter verisini oluşturmak için orman yangını olan ve olmayan bölgelerin hassas bir şekilde tanımlanarak ilgili örneklemelerin toplanmasını içermektedir. Bir sonraki adımda makine öğrenmesi modellerinin bir tahminleme mekanizması geliştirebilmesi ve performanslarının test edilebilmesi için veri seti eğitim (%70) ve test (%30) olmak üzere iki kısma ayrılmıştır. Üçüncü kısımda ise bağımsız değişkenler arasında makine öğrenmesi algoritmalarının performanslarını negatif olarak etkileyebilecek potansiyel bir korelasyon olup olmadığı çoklu bağlantı testi ve korelasyon analizi ile araştırılmıştır. Daha sonra, çalışma alanına ait orman yangını duyarlılık haritalarını üretebilmek için üç topluluk tabanlı makine öğrenmesi algoritması (Rastgele Orman, XGBoost ve NGBBoost) kullanılmıştır. Son olarak, üretilen duyarlılık haritalarının kaliteleri performans değerlendirme metrikleri ve istatistiksel anlamlılık testi ile analiz edilmiştir.

3.1 Rastgele Orman

Topluluk tabanlı sınıflandırma yöntemleri, bir sınıflandırıcı yerine birden çok sınıflandırıcı içeren ve sonrasında onların tahminlerinden alınan kararlar ile yeni veriyi sınıflandıran makinesi öğrenme algoritmalarıdır. Rastgele orman, en genel tabiriyle ağaç tipi sınıflandırıcılar topluluğu olarak tanımlanabilir. Rastgelelik özelliği eklenerek torbalama yönteminin geliştirilmiş bir versiyonudur (Breiman 2001). Bu rastgelelik, hem veri örnekleme sırasında hem de özelliklerin seçiminde kullanılır, bu da modelin genelleme yeteneğini artırır ve aşırı öğrenme riskini azaltır. Rastgele orman, tüm değişkenler arasından en iyi dalı kullanarak her bir düğümü dallara ayırmak yerine, her bir düğümde rastgele olarak seçilen değişkenler arasından en iyisini kullanarak her bir düğümü dallara ayırır. Her bir veri seti orijinal veri setinden yer değiştirmeli (bootstrap) olarak üretilir, yani bir veri örnekleme seçildikten sonra tekrar seçilebilmesi için orijinal veri setine tekrar dahil edilir. Daha sonra, rastgele özellik seçimi kullanılarak ağaçlar oluşturulur (Archer ve Kimes, 2008). Bu rastgele özellik seçimi, ağaçlar arasındaki korelasyonu azaltmayı ve böylece modelin içerisindeki çeşitliliği artırmayı amaçlar. Rastgele orman genellikle birçok sınıflandırma ve regresyon probleminde yüksek doğruluk ve hız sunar. Aynı zamanda, aşırı uyuma karşı dayanıklıdır ve modelin performansı kullanıcılar tarafından belirlenen sayıda bireysel ağaç kullanılarak optimize edilebilir.

3.2 Extreme Gradient Boosting (XGBoost)

XGBoost algoritması, gradyanı arttırılmış karar ağacı temeline dayanan topluluk tabanlı bir makine öğrenmesi algoritmasıdır. XGBoost, regresyon ve sınıflandırma problemlerinde yapılandırılmış (structured) veya tablo (tabular) halindeki veri setlerinde kullanılabilir. XGBoost, sistem optimizasyonu ve algoritmik geliştirmeler yoluyla temel gradyan artırma makineleri tekniğinin geliştirilmiş bir halidir. Algoritma, model içinde bulunan her bir yeni ağacı önceki ağaçların hatalarını düzeltmeye yönelik olarak eğitir. Süreç boyunca, her yeni ağaç, önceki ağacın tahmin hatalarına odaklanarak geliştirilir. Bu işlem, belirli bir hata eşiği veya maksimum ağaç sayısına ulaşılan kadar devam eder. Nihai model, tüm ağaçların ağırlıklı ortalamasını alarak tahmin yapar. Paralel hesaplama yapabilme yeteneği, eğitim süresini önemli ölçüde kısaltır ve bellek kullanımını etkin bir şekilde yönetir. XGBoost, diğer gradyan artırma algoritmalarına göre daha hızlı ve daha doğru performans gösterir (Chen ve Guestrin, 2016). Aynı zamanda, veri kümesini her biri belirli sayıda örnek içeren küçük parçalara böldükten sonra her bir parçadan karar ağacı oluşturur. Oluşturmuş olduğu karar ağaçlarını bir gradyan artırma algoritması kullanarak birleştirir.

3.3 Natural Gradient Boosting (NGBoost)

Natural Gradient Boosting (NGBoost), olasılıksal bir tahmin yapısı sağlamak amacıyla geliştirilen, yeni ve sağlam bir toplu makine öğrenimi algoritmasıdır (Kavzoglu ve Teke, 2022a). NGBoost, klasik gradyan artırma algoritmalarından farklı olarak, yalnızca belirli bir tahmin yapmak yerine, tahmin edilen değerlerin yanı sıra bu tahminin olasılık dağılımını da öğrenir. Yani, her bir iterasyonda, model hem tahmin edilen değeri hem de bu tahminin olasılık dağılımını optimize etmeye çalışır. NGBoost, hedef değişkenin olasılık dağılımını modellemek için geniş bir dağılım ailesini (örneğin, normal dağılım, log-normal dağılım, vb.) destekler. Her iterasyonda, model parametreleri doğal gradyan kullanılarak güncellenir. Her adımda, mevcut modelin hataları üzerinden yeni bir model oluşturulur ve bu yeni model mevcut modele eklenir. NGBoost, bir sonucu tahmin etmek için bir model eğitmez; bunun yerine tam bir olasılık dağılımı tahmin eder (Iban ve Bilgiliolu, 2023).

3.4 Hiperparametre Optimizasyonu

Makine öğrenmesi modelleri tasarlanırken kullanılan bu değişkenler veriden doğrudan elde edilen model parametreleri ve tasarımcı tarafından önceden belirlenen hiperparametreler olmak üzere iki gruba ayrılır. Model parametreleri genellikle veriden öğrenilir ve kullanıcının bunları ayarlaması gerekmez; bu parametreler, modelin bir parçası olarak kaydedilir. Hiperparametreler ise model parametrelerinden farklı olarak, veriden öğrenilmeyen ve tasarımcı tarafından ayarlanması gereken özelliklerdir (Tanyıldızı ve Demirtaş, 2019). Hiperparametreler, bir modelin başarılı bir şekilde performans göstermesi için optimize edilmesi gereken özelliklerdir. Bir başka ifade ile, makine öğrenmesi algoritmalarında algoritmanın performansını ve başarısını etkileyen, modelin kendisi tarafından öğrenilmeyen kullanıcı tarafından belirlenen modele ait özel parametrelerdir (Kavzoglu ve Teke, 2022b). Bunlar arasında modelin öğrenme oranı, yineleme sayısı, aktivasyon fonksiyonları, ağırlıklar, regülasyon parametreleri ve çıkış fonksiyonu gibi unsurlar bulunur. Hiperparametrelerin yanlış seçilmesi, modelin düşük performans göstermesine veya aşırı uyum sorununa yol açabilir.

Literatürde ızgara arama, rastgele arama ve Bayesian tabanlı algoritmalar gibi hiperparametre optimizasyon teknikleri bulunmaktadır. ızgara arama tekniği her ne kadar hesaplama karmaşıklığı açısından diğerlerine nazaran daha fazla olsa da optimum doğruluğu garanti ettiği için bu çalışmada kullanılmıştır. ızgara arama yönteminde, belirli bir hiperparametre uzayında belirli aralıklarla veya belirli değerlerle bir ızgara oluşturularak her bir kombinasyon denenir ve belirli bir performans değerlendirme metriği kullanılarak en iyi hiperparametre belirlenir. ızgara arama ile hiperparametre seçim işleminde; belirlenen aralıkta bulunan tüm değerlerin kombinasyonları için eğitim yapılır, sonuçlar gözlemlenerek elde edilen en iyi kombinasyon hiperparametre grubu olarak seçilir. Uygulama açısından kolay bir yöntem olup düşük boyutlu alanlarda iyi sonuçlar vermektedir. ızgara arama yönteminin dezavantajı ise çalışma süresinin uzun sürmesidir (Tanyıldızı ve Demirtaş, 2019).

3.5 İstatistiksel Anlamlılık Testi

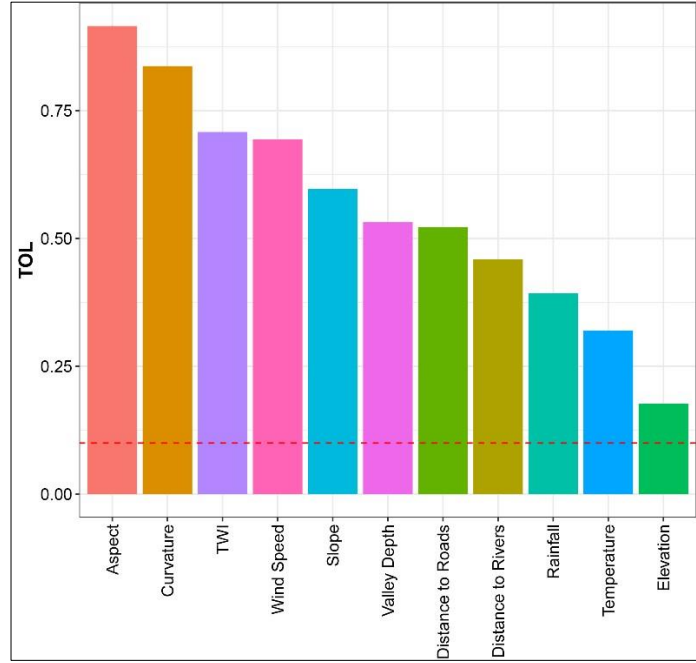
İstatistiksel anlamlılık, bir analizin sonuçlarının rastgele olup olmadığını değerlendirmek için kullanılır. Sonucun istatistiksel olarak anlamlı olması gözlemlenen farkın veya ilişkinin, rastgele kaynaklanmadığını, gerçek bir ilişkinin var olduğunu gösterir. Algoritmaların performansları arasındaki farklılıkların istatistiksel açıdan anlamlı olup olmadığını belirlemek için çeşitli istatistiksel anlamlılık testleri kullanılabilir. Karşılaştırmalı çalışmalarda sadece sayısal gözlem yerine model performansları arasındaki farkın istatistiksel olarak doğrulanması sonuçlardan daha sağlıklı çıkarımlar yapılabilmesi açısından büyük önem taşımaktadır (Kavzoglu ve Teke, 2022a). İstatistiksel anlamlılığı değerlendirmek için kullanılan testler arasında McNemar testi, Wilcoxon Signed-Rank testi ve Ki-kare testi gibi yöntemler bulunur.

Çalışma kapsamında, topluluk tabanlı makine öğrenmesi algoritmalarının performansları arasındaki istatistiksel olarak anlamlı bir fark olup olmadığı McNemar testi ile analiz edilmiştir. McNemar testi, hata matrislerinden hesaplanan standartlaştırılmış normal test istatistiğine dayanan parametrik olmayan bir testtir. McNemar testi ki-kare dağılımına

bağlıdır ve hesaplamalarda 2×2 boyutlu bir hata matrisi kullanır (Foody 2004). Doğruluk değerlendirmesi yapılırken doğruluk hesaplamaları arasındaki farkın istatistiksel önemi, kapa katsayısı ve McNemar testi kullanılarak belirlenebilir (Foody 2004).

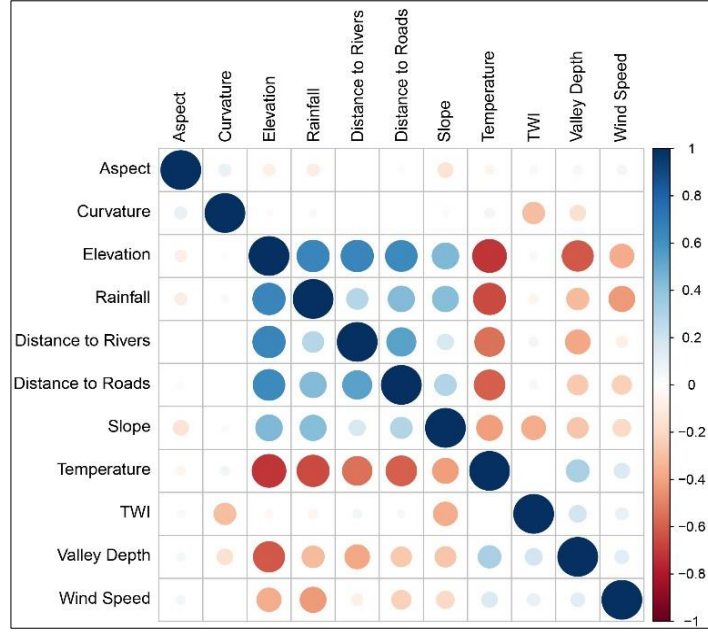
4. SONUÇLAR VE TARTIŞMA

Çalışma kapsamında ilk olarak faktörler arasında potansiyel istatistiksel olarak bir korelasyon olup olmadığını incelemek, eğer varsa makine öğrenmesi modellerinin performansını olumsuz yönde etkilemesini önlemek için çoklu doğrusallık testi ve korelasyon analizi olmak üzere iki farklı yaklaşım uygulanmıştır. Çoklu doğrusallık testinde VIF ve Tolerans olmak üzere iki farklı gösterge hesaplanmaktadır. Eğer hesaplanan VIF değeri 10'dan büyük ise veya Tolerans değeri 0.1'den küçük ise ilgili bağımsız değişkenler arasında yüksek düzeyde çoklu doğrusallık olduğunu gösterir. Elde edilen sonuçlar, maksimum Tolerans değerinin 0.915 ve minimum VIF değerinin ise 1.093 ile bakı parametresine ait olduğunu göstermiş; yani faktörler arasında istatistiksel çoklu doğrusallık sorununa rastlanmamıştır (Şekil 2).



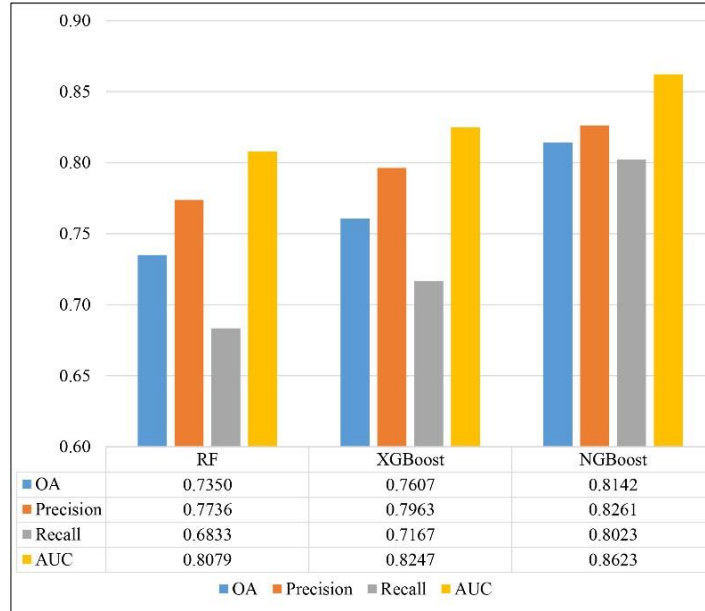
Şekil 2. Bağımsız değişkenler için tolerans değerleri.

Çoklu doğrusallık testine ek olarak, orman yangını duyarlılığında etkili 11 faktör arasındaki ilişkinin yönünü ve şiddetini tespit etmek için Pearson korelasyon analizi uygulanmıştır (Şekil 3). Korelasyon analizi maksimum korelasyon katsayısı değerinin 0.66 değeri ile nehirlere olan uzaklık ve yükseklik parametreleri arasında olduğunu, minimum korelasyon katsayısının ise -0.68 ile yükseklik ve sıcaklık faktörleri arasında olduğunu göstermiştir. Bulunan değerler 0.7 eşik değerinin altında olduğundan dolayı çalışmada belirlenmiş olan orman yangını duyarlılığını etkileyen 11 faktör tahminleme aşamasında bağımsız değişkenler olarak kullanılmıştır.



Şekil 3. Orman yangınına sebep olan parametreler arasındaki korelasyon matrisi.

Çalışma alanına ait orman yangını duyarlılık haritaları NGBoost, XGBoost ve Rastgele Orman olmak üzere üç farklı topluluk tabanlı makine öğrenmesi algoritması kullanarak üretilmiştir. Optimum tahminleme performansı elde edebilmek için her bir algoritmanın hiperparametre değerleri ızgara arama (grid search) yöntemi ile optimize edilmiştir. Optimizasyon sürecinde genel doğruluk fitness fonksiyonu olarak seçilmiş ve üç katlı çapraz doğrulama uygulanmıştır. Üretilen haritaların güvenilirlikleri ise genel doğruluk, eğri altında kalan alan, kesinlik ve duyarlılık olmak üzere dört farklı doğruluk değerlendirme metriği kullanılarak ölçülmüştür (Şekil 4). Sonuçlara göre, NGBoost %81.42 genel doğruluk skoru ile en yüksek orman yangını tahminleme performansına sahipken Rastgele Orman algoritması diğerlerine göre daha düşük bir genel doğruluk skoru (%73.50) üretmiştir.

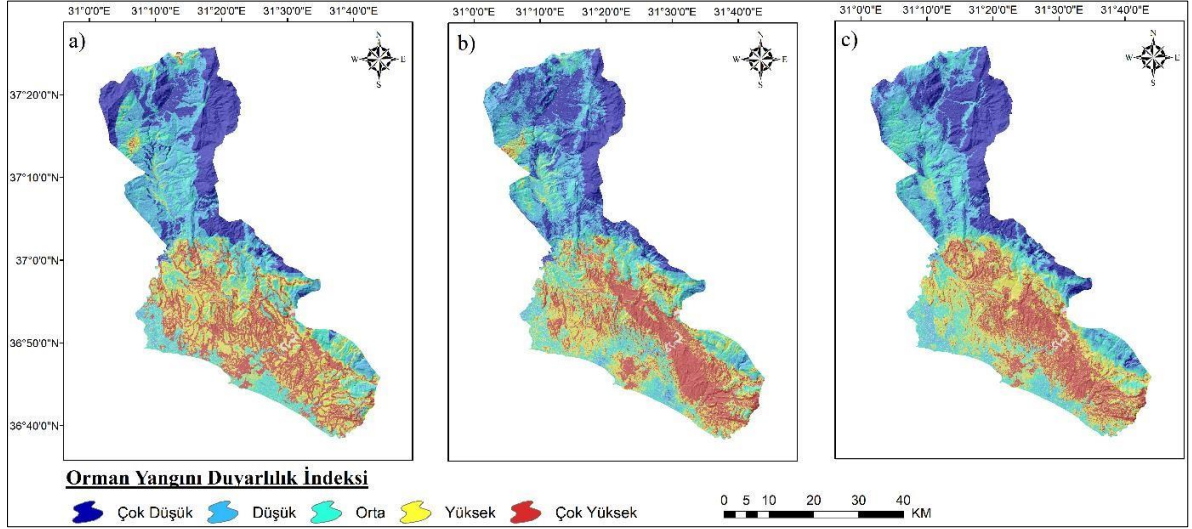


Şekil 4. Rastgele Orman, XGBoost ve NGBoost algoritmalarının doğruluk değerlendirme sonuçları.

Algoritmaların performansları arasındaki farkın istatistiksel olarak anlamlı bir değer olup olmadığı McNemar testi ile analiz edilmiştir (Çizelge 1). Bu farkın da istatistiksel olarak anlamlı olduğunu söylememiz için değer 3.84'den büyük olması gerekir. İstatistiksel analiz sonuçları NGBoost ve RF algoritmalarının performanslarının birbirinden istatistiksel olarak anlamlı olduğunu ancak diğer algoritmaların performansları arasındaki farkın istatistiksel olarak anlamsız olduğunu ortaya çıkarmıştır.

Çizelge 1. Çoklu doğrusallık testi analizi.

| | RF | XGBoost | NGBoost |
|---------|----|---------|---------|
| RF | - | 2.722 | 5.042 |
| XGBoost | - | - | 3.704 |
| NGBoost | - | - | - |



Şekil 5. Orman yangını duyarlılık haritaları: (a) NGBoost, (b) XGBoost, (c) Random Forest.

Son olarak üretilen orman yangını duyarlılık haritaları görsel olarak analiz edilmiştir. Orman yangını duyarlılık haritaları, quantile tabanlı yeniden sınıflandırma yaklaşımı yoluyla çok yüksek, yüksek, orta, düşük, çok düşük olmak üzere 5 aralığa bölünmüştür (Şekil 5). Orman yangını duyarlılık haritaları tematik olarak incelendiğinde, orman yangını hassasiyetinin genellikle çalışma alanının orta ve güney bölgelerinde yükseldiği, ancak çalışma alanının kuzey ve kuzeybatı kısımlarında orman yangını hassasiyetinin düşük ve çok düşük sınıflara karşılık geldiği gözlemlenmiştir. Çalışma alanının orta ve güney bölgesinde, rüzgâr hızı ve sıcaklığın yüksek olup ve bu alanın yağışı ilçenin diğer bölgelerine göre daha az olması orman yangını hassasiyetinin artmasına yol açan faktörler olarak değerlendirilmiştir. Elde edilen bulgular doğrultusunda, bu çalışmanın ana amacına uygun olarak en önemli sonuçlar aşağıdaki şekilde sıralanabilir.

- Çoklu doğrusallık testi kapsamında yapılan korelasyon analizi ve VIF/Tolerans değerleri, orman yangınına neden olan faktörler arasında potansiyel bir istatistiksel doğrusallık sorunu olmadığını ortaya koymuştur.
- Altı farklı doğruluk değerlendirme metriği göz önüne alındığında, NGBoost algoritması, orman yangını duyarlılığını tahmin etmede XGBoost ve Rastgele Orman algoritmalarına kıyasla genel doğrulukta %5 ile %8 oranında daha başarılı sonuçlar elde edilmiştir.
- İstatistiksel test sonuçları, NGBoost ile diğer makine öğrenme algoritmaları arasındaki performans farklarının %95 güven aralığında istatistiksel olarak anlamlı olduğunu göstermiştir. Bu da NGBoost algoritmasının üstünlüğünü bir kez daha ortaya koymuştur. Algoritmaların performansları arasında yaklaşık %8'e kadar performans farkları olduğu saptandı ve bu farkların istatistiksel olarak anlamlı olup olmadığı McNemar testi ile analiz edildi.

KAYNAKLAR

Akıncı, H. A. ve Akıncı, H., 2023. Machine Learning Based Forest Fire Susceptibility Assessment of Manavgat District (Antalya), Turkey. *Earth Science Informatics*, 16(1): 397-414.

Archer, K. J. ve Kimes, R. V., 2008. Empirical Characterization of Random Forest Variable Importance Measures. *Computational Statistics & Data Analysis*, 52(4): 2249-2260.

Atmaca, İ., Derakhshandeh, M., Işık Pekkan, Ö., Özenen-Kavlak, M., Tunca, Y. S. ve Çabuk, S. N., 2022. Lojistik Regresyon ve Coğrafi Bilgi Sistemleri Kullanılarak Orman Yangını Risk Modellemesi: Muğla-Milas Örneği. *Doğal Afetler ve Çevre Dergisi*, 8(1): 66-75.

Beşli, N. ve Tenekeci, E., 2020. Uydu Verilerinden Karar Ağaçları Kullanarak Orman Yangını Tahmini. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi*, 11(3): 899-906.

Bilucan, F., Teke, A. ve Kavzoglu, T., 2024. Susceptibility Mapping of Wildfires Using XGBoost, Random Forest and AdaBoost: A Case Study of Mediterranean Ecosystem. Bezzeghoud, M., vd. (Ed.), *Recent Research on Geotechnical Engineering, Remote Sensing, Geophysics and Earthquake Seismology*, Advances in Science, Technology & Innovation. Springer, Cham. syf: 99-101.

Breiman, L., 2001. Random forests. *Machine Learning*, 45: 5-32.

Chen, T. ve Guestrin, C., 2016. Xgboost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM: SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, syf: 785-794.

Divya, A., Kavithanjali, T. ve Dharshini, P., 2019. Iot Enabled Forest Fire Detection and Early Warning System. *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, Pondicherry, India, syf: 1-5.

Dong, X., Yu, Z., Cao, W., Shi, Y. ve Ma, Q., 2020. A Survey on Ensemble Learning. *Frontiers of Computer Science*, 14: 241-258.

Footy, G. M., 2004. Thematic Map Comparison. *Photogrammetric Engineering & Remote Sensing*, 70(5): 627-633.

Ge, X., Yang, Y., Peng, L., Chen, L., Li, W., Zhang, W. ve Chen, J., 2022. Spatio-Temporal Knowledge Graph Based Forest Fire Prediction with Multi Source Heterogeneous Data. *Remote Sensing*, 14(14): 3496.

Iban, M. C. ve Sekertekin, A., 2022. 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.

Iban, M. C. ve Bilgilioglu, S. S., 2023. Snow Avalanche Susceptibility Mapping Using Novel Tree-Based Machine Learning Algorithms (Xgboost, Ngboost, And Lightgbm) with Explainable Artificial Intelligence (XAI) Approach. *Stochastic Environmental Research and Risk Assessment*, 37(6): 2243-2270.

Kavzoğlu, T., Çölkesen, İ., Tonbul, H. ve Öztürk, M. Y., 2021. Uzaktan Algılama Teknolojileri ile Orman Yangınlarının Zamansal Analizi: 2021 Yılı Akdeniz ve Ege Yangınları. Kavzoğlu, T. (Ed.), *Orman Yangınları Sebepleri, Etkileri, İzlenmesi, Alınması Gereken Önlemler ve Rehabilitasyon Faaliyetleri*, Türkiye Bilimler Akademisi. syf: 219-251.

Kavzoglu, T. ve Teke, A., 2022a. Predictive Performances of Ensemble Machine Learning Algorithms in Landslide Susceptibility Mapping Using Random Forest, Extreme Gradient Boosting (XGBoost) and Natural Gradient Boosting (NGBoost). *Arabian Journal for Science and Engineering*, 47(6): 7367-7385.

Kavzoglu, T. ve Teke, A., 2022b. Advanced Hyperparameter Optimization for Improved Spatial Prediction of Shallow Landslides Using Extreme Gradient Boosting (XGBoost). *Bulletin of Engineering Geology and the Environment*, 81: 201.

Piao, Y., Lee, D., Park, S., Kim, H. G. ve Jin, Y., 2022. Multi-Hazard Mapping of Droughts and Forest Fires Using A Multi-Layer Hazards Approach with Machine Learning Algorithms. *Geomatics, Natural Hazards and Risk*, 13(1): 2649-2673.

Rihan, M., Bindajam, A. A., Talukdar, S., Naikoo, M. W., Mallick, J. ve Rahman, A., 2023. Forest Fire Susceptibility Mapping with Sensitivity and Uncertainty Analysis Using Machine Learning and Deep Learning Algorithms. *Advances in Space Research*, 72(2): 426-443.

Sayad, Y. O., Mousannif, H. ve Al Moatassime, H., 2019. Predictive Modeling of Wildfires: A New Dataset and Machine Learning Approach. *Fire Safety Journal*, 104: 130-146.

Sefercik, U. G. ve Kavzoglu, T., 2021. Orman Yangınlarında Uzaktan Algılama Teknolojileri ile Erken Uyarı, Tespit, İzleme ve Müdahale Stratejileri. Kavzoğlu, T. (Ed.), *Orman Yangınları Sebepleri, Etkileri, İzlenmesi, Alınması Gereken Önlemler ve Rehabilitasyon Faaliyetleri*, Türkiye Bilimler Akademisi. syf: 111-135.

Tanyıldızı, E. ve Demirtaş, F., 2019. Hiper Parametre Optimizasyonu Hyper Parameter Optimization. *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, Ankara, Türkiye, 2019, syf: 1-5.

Tonbul, H., Kavzoglu, T. ve S. Kaya., 2016. Assessment of Fire Severity and Post-Fire Regeneration Based on Topographical Features Using Multitemporal Landsat Imagery: A Case Study in Mersin, Turkey. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41: 763-769.

Trucchia, A., Izadgoshasb, H., Isnardi, S., Fiorucci, P. ve Tonini, M., 2022. Machine-Learning Applications in Geosciences: Comparison of Different Algorithms and Vegetation Classes' Importance Ranking in Wildfire Susceptibility. *Geosciences*, 12(11): 424.

Xanthopoulos, G., 2007. Forest fire policy scenarios as a key element affecting the occurrence and characteristics of fire disasters. *Proceedings of the 4th International Wildland Fire Conference*, Sevilla, Spain.

Xie, L., Zhang, R., Zhan, J., Li, S., Shama, A., Zhan, R., Wang, T., Lv, J., Bao, X. ve Wu, R., 2022. Wildfire Risk Assessment in Liangshan Prefecture, China Based on An Integration Machine Learning Algorithm. *Remote Sensing*, 14(18): 4592.

Yıldırım, O., Gunay, F. B. ve Yağanoğlu, M., 2023. Makine Öğrenmesi Yöntemleriyle Orman Yangını Tahmini. *Journal of the Institute of Science and Technology*, 13(3): 1468-1481.