

## Makine Öğrenmesi Algoritmalarının Modelleme ve Tahmin Performanslarının Karşılaştırılması: Türkiye'deki Orman Yangın Sayısı Örneği

### Comparison of Modeling and Prediction Performance of Machine Learning Algorithms: The Case of Forest Fire Count in Türkiye

İbrahim Edibali ATALAY<sup>1\*</sup> , Abdullah ALTAY<sup>2</sup> , Serdar NESLİHANOĞLU<sup>2</sup> 

<sup>1</sup> Artvin Çoruh Üniversitesi, Lisansüstü Eğitim Enstitüsü, Artvin

<sup>2</sup> Eskişehir Osmangazi Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Eskişehir

#### Eser bilgisi/Article info

Araştırma makalesi

DOI: 10.17474/artvinofd.1867212

\*Sorumlu yazar/Corresponding author

İbrahim Edibali ATALAY

e-mail: ibrahimatalay3@gmail.com

Gönderilme tarihi/Submission date

19.01.2026

Kabul tarihi/Acceptance date

07.04.2026

Yayınlanma tarihi/Publication date

01.05.2026

#### Anahtar kelimeler:

Makine öğrenmesi

Modelleme

Orman yangın sayısı

Türkiye

#### Keywords:

Machine learning

Modelling

Forest fire count

Türkiye

#### Öz

Dünya'nın astronomik hareketleri ve jeolojik aktivitelerin sonucunda oluşan küresel ısınma ve iklim değişikliği, günümüzde orman yangınları başta olmak üzere çeşitli doğal afetlerin artmasına neden olmaktadır. Türkiye'deki gelecek nesilleri orman yangınları konusunda bilinçlendirmek amacıyla orman yangınlarının geçmiş veriler ışığında analiz edilmesi ve tahminlerde bulunularak tedbirlerin alınmasının sağlanması bu araştırmanın temel odak noktası olmuştur. Bu amaç doğrultusunda günümüzdeki popülerliği gittikçe artan makine öğrenmesi yöntemlerinden CatBoost, LightGBM, LSTM, Random Forest ve XGBoost algoritmaları, orman yangın sayısını modellemek ve tahminini gerçekleştirmek için tercih edilmiştir. Bu çalışmada, ilk defa olarak Türkiye'deki günlük Kasım 2000-Ocak 2026 tarihleri arasındaki orman yangın sayısı verileri kullanılmıştır. Sonuç olarak, LightGBM modeli orman yangını sayısını modelleme daha etkin, Random Forest modeli ise gelecek tahmini konusunda daha etkin performansa sahiptir.

#### Abstract

Global warming and climate change, driven primarily by anthropogenic greenhouse gas emissions and various other factors, are today causing an increase in natural disasters, particularly forest fires. The primary focus of this research is to raise awareness among future generations in Türkiye about forest fires by analyzing them in light of historical data, making predictions, and ensuring that necessary precautions are taken. In this context, CatBoost, LightGBM, LSTM, Random Forest, and XGBoost algorithms, among the machine learning methods that have found widespread application today, were preferred to model and predict the count of forest fires. In this research, daily forest fire count data from Türkiye covering the period November 2000 – January 2026 were used for the first time. As a result, the LightGBM model demonstrated greater effectiveness in modeling the count of forest fires, while the Random Forest model exhibited higher performance in future prediction.

## GİRİŞ

Gezegemizin 4.5 milyar yıllık ömründe kıtaların hareketleri, volkanik patlamalar, küresel ısınma ve iklim değişikliği sonucunda orman yangınları, sel, kuraklık gibi doğal afetler meydana gelmiştir. Küresel ısınma ve iklim değişikliğinin nedenleri arasında son yıllarda doğal etkilerin yanında insani etkiler de görülmektedir. Tomanbay (2008) çalışmasında, 19. yüzyılın ortalarındaki sanayi devriminde, kömürün kullanılmaya başlanması ve geniş çaplı ormansızlaşmanın ortaya çıkması sonucu küresel ısınma gözlenmiştir. Sıcaklık artışının kaynağının orman yangınları ve tarlalardan çıkan metandan olduğu belirtilirken ormansızlaşma ise tomrukların ayrışması olarak belirtilmiş bununla beraber yaklaşık 8 yıl önce de aynı durumlarla karşılaşıldığı araştırmalar sonrasında görülmektedir.

Dünya üzerindeki ekonomik değer taşıyan en önemli doğal kaynaklarından birisi orman ekosistemidir. Ormansızlaşma sonucu ortaya çıkan çölleşme hem ormanları hem de su kaynaklarını tehdit eden diğer etmenlerdir. Ormansızlaşmayı engellemek adına alınacak önlemlerle beraber orman ekosisteminin sürdürülebilirliğini sağlanabilmesi için iklim değişikliği tehdidini ortadan kaldırmak gerektiği, araştırmacıların ortak görüşleri olarak belirtilebilir. Türkiye gibi orman varlığı bakımından zengin ve farklı iklim tiplerine ev sahipliği yapan ülkelerde, iklim değişikliğinin ormancılık faaliyetlerine

olan etkisi çok daha belirgindir. Orman ekosistemlerinin korunması, karbon yutak alanlarının artırılması ve iklim dirençli ormancılık politikalarının uygulanması, bu sürecin başarısı açısından kritik öneme sahiptir. Bu amaç doğrultusunda, Türkiye’de orman yangınları konusunda araştırmalar yapılmıştır.

Calda vd. (2020), artan kurak hava koşulları sebebiyle doğal yollardan, ihmal ve kasıt ve iklim elemanlarının etkisiyle büyüklükleri değişen orman yangınlarının meydana geldiğini belirtmişlerdir. Çalışmada, Antalya, Çanakkale ve Muğla Orman Bölge Müdürlükleri orman arazileri 2008 ile 2009 yıllarında çıkan Kanada Orman Yangın Hava İndisi (FWI) kullanılarak incelemişlerdir. Sonuç olarak, FWI değerleri bu yıllar için yangın riskini öngörmede başarılı olduğunu tespit etmişlerdir.

Özenen Kavlak vd., (2020), doğal veya insan faktörleri sonucu orman yangınlarının meydana geldiği belirtmişlerdir. Kütahya Orman İşletme Müdürlüğüne bağlı Ören Orman İşletmesini incelemişlerdir. Coğrafi Bilgi Sistemleri (CBS) yöntemleri kullanılan çalışmada orman yangını riskinin asgari düzeye indirilmesi, yangınlara karşı hazırlıklı olunması ve maddi ve fiziki olarak ortaya çıkacak zararların önlenmesini hedeflemişlerdir. Bu bağlamda Ağırlıklı Çakıştırma yöntemini ile oluşturulan risk haritaları alanlarına göre %36 çok riskli, %60 orta riskli ve %2 az riskli alanlar olarak tespit edilerek CBS’nin orman yangını planlamalarında etkin bir karar destek aracı olduğunu vurgulamışlardır.

Yılmaz vd. (2021), Türkiye’deki orman yangınlarının sebeplerini; kundaklama, dikkatsizlik ve ihmal, doğal ve faili meçhul olarak dört ana grupta toplandığını belirtmişlerdir. Çalışmalarında, Google Earth Engine (GEE) ve Coğrafi Bilgi Sistemleri (CBS) yardımıyla Uzaktan Algılama (UA) tekniğini kullanarak 9 Ekim 2020 tarihinde Hatay’ın Belen ilçesinde meydana gelen orman yangınının etkilerini incelemişlerdir. Sonuç olarak, GEE platformu ile yangın sonrası verilerine hızlı bir şekilde ulaşılabileceğini ve bu yöntem sayesinde orman yangınlarıyla mücadelede karar vericiler için kritik rehber olduğunu vurgulamışlardır.

Şahan ve Kaya (2022), Türkiye ve Avrupa’daki orman yangınları hakkında durum tespiti yaparak karşılaştırmalar yapmışlardır. Avrupa Orman Yangınları Bilgi Sistemi (EFFIS) tarafından yayımlanan yıllık veriler analiz edilmiştir. Bu karşılaştırmalar için 2010 ile 2019 dönemindeki yangın sayısı ve yanan alan miktarı, ormanlık alan miktarı ve yüzölçümüne oranla orman varlığı gibi değişkenler üzerinde analizler uygulanmıştır. Türkiye orman yangınları müdahalesinde birçok Avrupa ülkesine göre başarılı olduğu sonucu elde edilmiş olup, orman varlığı olarak yetersiz ve yüzölçüm bakımından orman alanının yetersiz olduğu sonucuna ulaşmışlardır.

Usta (2023), orman yangınlarını afet sınıflaması içinde yer aldığını, çıkış sebebini ihmal ve kaza olduğunu belirterek Türkiye’de 2005 ile 2020 yılları arasında çıkan orman yangınlarını incelemiştir. Çalışma sonucunda, orman yangını riskinin yüksek olduğu bölgede yaşayan insanların bilinçlendirilmesi, riskli bölgenin teknolojik araç ve gereçler yardımıyla denetlenmesi ve söndürme ekipmanları açısından donatılmasını bu kapsamda bilimsel veriler ışığında duyarlılık haritalarının geliştirilerek uygulanabilir modellerin geliştirilmesi önerilerinde bulunmuştur.

Yıldırım vd. (2023), makine öğrenmesi tabanlı ikili sınıflandırma modelleri kullanarak “yangın var, yangın yok” durumunun yüksek doğrulukla tahmin etmeyi hedeflemişlerdir. Çalışma verileri NASA’nın Oak Ridge Ulusal Laboratuvarı (ORNL) Dağıtılmış Aktif Arşiv Merkezi (DAAC) tarafından sağlanan 1983 ile 2016 yılları arasındaki Alaska ve Kanada’daki arazi örneklerinden oluşturulmuştur. Çalışmada nem, sıcaklık, rüzgâr hızı vb. 49 özniteliklere yer verilmiştir. Tahmin modellerinin oluşturulması 6 farklı denetimli makine öğrenmesi algoritmasıyla Destek Vektör Makineleri (DVM), Karar Ağaçları (KA), Rastgele Orman (RO), Gradyan Artırma (GA), K-En Yakın Komşu (KNN) ve Naive Bayes (NB) yapılmıştır. Sonuç olarak, makine öğrenmesi yöntemlerinin orman yangın tahmininde son derece etkili olduğunu tespit etmişlerdir.

Demirel ve Türk (2023), son yıllarda sıklık ve büyüklük açısından orman yangınlarında artış gözlemlendiğini belirterek orman yangınlarının analiz edilmesi bu durum doğrultusunda tedbirlerin alınmasının önemli olduğunu vurgulamışlardır. 2015-2022 yılları arasında Türkiye’de ortaya çıkan orman yangınlarının il sınırları seviyesine göre mekânsal ve zamansal dağılımı, yangının coğrafi konum ve çıkış nedeni arasındaki ilişkiyi Coğrafi Bilgi Sistemi ortamında analiz etmişlerdir. Sonuç olarak, kümelenmeler incelendiğinde Akdeniz ve Ege Bölgeleri etrafında yüksek konumsal kümelenmenin yüksek bir güven aralığı ile tekrar ettiğini tespit etmişlerdir.

Güngöroğlu vd. (2024), ülkemizdeki orman yangınlarının son on yılda hem sayısal hem de alansal olarak artış gösterdiğini kayıtlara göre ise ortalama yangın sayısının 2569, son 87 yıllık ortalamanın ise 1409 olduğunu belirtmişlerdir. Orman yangınlarının %90’ının insan kaynaklı olduğunu ve literatür çalışmalarına göre ise en büyük sorunlardan birisi olan “sebebi bilinmeyen” yangınların ise son on yılda %47.4 oranında arttığını belirtmişlerdir. Çalışma sonucunda yangın öncesine dair, yangın esnasına dair ve yangın sonrasında dair öneriler ve çözümler sunmuşlardır. Ateşoğlu (2016), Ormanları çeşitli tehlikelere karşı korumak, gerekli önlemleri zamanında almak ve bu önlemlere karşın meydana gelecek tehlikelerle savaşa hazır bulunmak gerektiğini belirtmiştir.

Türkiye’deki orman yangını konusunda literatürdeki araştırmalar incelendiğinde çalışmaların temelinde orman yangını sebebi ve yangın sonrasında dair öneriler ve çözümler sunulmasına odaklanıldığı görülmektedir. Fakat Türkiye’deki yangın sayısının modellenmesi ve gelecek tahmini konusundaki literatürün yetersizliği açık olarak görülmektedir. Bu amaç doğrultusunda, günümüzdeki popülerliği gittikçe artan makine öğrenmesi yöntemlerinden CatBoost, Random Forest, LightGBM, XGBoost ve LSTM’in 2000-2026 yılları arasındaki Türkiye’deki günlük orman yangın sayısının modellenmesine ve gelecek tahmini konusuna odaklanılacaktır.

Bu araştırmanın geri kalan kısımları: Bölüm 2’de tercih edilen modellerin metodolojisi ve Bölüm 3’te araştırma verisi istatistiklerine yer verilmiştir. Bölüm 4’te araştırma bulgularına ve Bölüm 5’te ise tartışma ve sonuçlara yer verilmiştir.

## MATERYAL VE YÖNTEM

Bu araştırmada kullanılan makine öğrenmesi modelleri bu bölümle tanıtılmıştır.

### CatBoost (Categorical Boosting)

CatBoost modeli hataları en aza indirmek ve tahminleri iyileştirmek için kategorik özelliklerle çalışmaları desteklerken bilgi kaybını en aza indirmek için tasarlanmıştır. CatBoost diğer modellere göre, daha az ön işleme ihtiyaç duymakta ve doğrudan veriyi kullanmaktadır. Dorogush vd. (2018) Ayrıca, modelin aşırı öğrenmesini engellemek için veri setinin rastgele permütasyonlarını kullanarak, o andaki veriden önceki verilerin istatistikleri hesaplanır. Buna ek olarak, CatBoost simetrik ağaç yapısını kullanır. Bu yapı, modelin tahmin hızını artırır ve daha kararlı sonuçlar üretmesine olanak tanır. CatBoost modeli aşağıdaki gibi formüle edilmiştir.

$$h^t = \arg L(F^{t-1} + h), h \in H \quad (1)$$

$$= \arg EL(y, F^{t-1}(x) + h(x)), h \in H$$

$$h(x) = \sum_j b_j I\{x \in R_j\} \quad (2)$$

Burada,  $y$  çıktığı temsil ederken,  $h$  ise  $H$  fonksiyon ailesinden seçilen bir gradyan adım fonksiyonudur.  $R_j$  ise karar ağacının yapraklarına karşılık gelen bölgeleri,  $b_j$  ise karşılık gelen bölgelerin tahmin değerlerini ifade ederken,  $I$  ise bir gösterge fonksiyonu olarak tanımlanabilir (Tunç ve Atasever, 2025).

### LightGBM (Light Gradient Boosting Machines)

LightGBM modeli, Microsoft tarafından büyük veri ve çok boyutlu özellikler üzerinde tahmin yapılabilmesi için tasarlanmıştır. Ayrıca, LightGBM modeli büyük verileri hızlı bir şekilde işleyip tahmin sunmaktadır. LightGBM, karar ağaçlarında öğreniminde seviye odaklı (level-wise or depth-wise) veya yaprak odaklı (leaf-wise) olarak iki strateji kullanılabilir. Seviye odaklı stratejide ağaç büyürken ağacın dengesi korunur, iken yaprak odaklı stratejide ise kaybı azaltan yapraklardan bölünme işlemi devam eder. LightGBM bu özelliğiyle diğer boosting algoritmalarından ayrılmaktadır. LightGBM modeli, daha hızlı öğrenme ve daha az hata ile tahmin için, her seferinde daha az yaprak ile daha iyi bir performans göstermiştir (Eşidir, 2025). LightGBM modeli aşağıdaki gibi formülize edilmiştir.

$$L^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \quad (3)$$

Burada,  $L^{(t)}$   $t$ . adımdaki toplam kaybı,  $l$  kayıp fonksiyonunu,  $y_i$   $i$ . gerçek değeri ve  $\hat{y}_i^{(t-1)}$ ,  $t - 1$  adımıdaki tahmini,  $f_t(x_i)$  ise  $t$  adımdaki model tarafından eklenen yeni fonksiyonu ve  $\Omega(f_t)$  yeni fonksiyonun cezasını temsil etmektedir.

### LSTM (Long Short Term Memory)

LSTM, öz yinelenmeli bir sinir ağı olarak zaman serisi verilerini işlemek için tasarlanmıştır (Zhang vd., 2019). Birçok farklı problemlerde yaygın olarak kullanılan ve uzun vadeli bağımlılık problemlerden kaçınmak için tasarlanan LSTM muadillerinin aksine zaman serisi içerisindeki gizli durumları yakalar, öğrenir ve önceki zaman verileriyle bağlantı kurabilmektedir. Zincir benzeri bir yapıya sahip olan LSTM'in en önemli noktası hücre durumu olmasıdır. Bu durum bilginin hiç değiştirilmeden zincir boyunca akması yani verinin adeta bir taşıma bandı olması gibi düşünülebilir. Uzun veriler için tercih edilen LSTM unutma kapısı, giriş kapısı, hücre durumu, çıkış kapısı olmak üzere 4 ana bileşenden oluşmaktadır (Çökekoğlu, 2023). LSTM ağının çalışma süreci aşağıdaki gibi formüle edilmiştir.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (4)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (5)$$

$$C_t = \tanh(W_C \cdot [h_{t-1}, x_t]) + b_C \quad (6)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot C_t \quad (7)$$

$$O_t = \sigma(W_O \cdot [h_{t-1}, x_t] + b_O) \quad (8)$$

$$h_t = o_t \cdot \tanh \tanh(c_t) \quad (9)$$

Burada,  $x_t$  girdiyi temsil ederken,  $h_{t-1}$  o andaki biriminin çıktısı,  $C_t$   $t$  anındaki yeni bilgiyi,  $\tanh$  aksiyon fonksiyonu ile  $C_t$  hücrenin güncellenmiş durumudur.  $\sigma$  verileri 0-1 aralığında aktivasyon işlemi için kullanmakta olup bu durum sigmoid aksiyonu olarak tanımlanmaktadır.  $O_t$  çıkış kapısının hesaplanmasını ve gizli durumdan ne kadarının çıkışa aktarılacağını belirlediği aşamadır.

## Random Forest

Random Forest, tekil karar ağaçlarının yapısal zaafalarını gidermek amacıyla geliştirilmiş topluluk öğrenme algoritmasıdır. Modelin temelinde orijinal veri setinden iadeli örnekleme yoluyla çok sayıda alt küme oluşmasına olanak sağlayan Bagging tekniği kullanılmaktadır. Ayrıca, tek bir karar ağacı modeline karşın; Random Forest bu ağaçlardan oluşan bir orman kurarak hatanın minimize edilmesine olanak sağlamaktadır. Buna ek olarak, her düğümde en iyi ayrımı bulmak için tüm özelliklere bakmak yerine, rastgele seçilen bir alt kümeye bakarak ağaçlar arasındaki korelasyonu düşürüp modelin genelleme kapasitesini artırır. Random forest modeli aşağıda formülize edilmiştir.

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B f_b(x) \quad (10)$$

## XGBoost (Extreme Gradient Boosting)

XGBoost, Gradient Boosting Machine (GBM) modelinin optimize edilmiş (Yeşilyurt ve Dalkılıç, 2021) ve yüksek performanslı bir ağaç tabanlı gradyan artırma kütüphanesidir (Chen ve Guestrin, 2016). XGBoost, sık karar ağaçlarını sıralı bir şekilde eğiterek güçlü bir tahminleyici oluşturma prensibine dayanır. Burada, her yeni ağaç, bir önceki ağacın hatalarını telafi edecek şekilde eğitilir. XGBoost yüksek tahmin gücü, azaltılmış karmaşıklık, eksik verilerin yönetilmesi, aşırı öğrenimin önlenmesi ve çalışma hızı gibi önemli avantajları sağlamaktadır. XGBoost aşağıdaki gibi formülize edilmiştir.

$$\hat{y}_i = \sum_{k=1}^k f_k(x_i) \quad (11)$$

$$L(\theta) = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{k=1}^k \Omega(f_k) \quad (12)$$

$$\Omega(f) = yT + \frac{1}{2} \lambda \|w\|^2 \quad (13)$$

Burada,  $\hat{y}_i$   $i$ . dönemdeki tahmin edilen değer,  $k$  toplamdaki veri sayısı,  $f_k(x_i)$   $x_i$  için  $k$ . ağaç tahminini temsil etmektedir.  $L$  modelin hedef değişkeni ne kadar iyi tahmin ettiğini ölçen kayıp fonksiyonu,  $\Omega(f_k)$  modelin karmaşıklığını cezalandıran düzeltme terimidir. Ayrıca,  $T$  ağaçtaki yaprak sayısını,  $y$  ve  $\Omega(f)$  modelin karmaşıklığını kontrol eden düzenleme parametresidir.

## Model Kıyaslama Kriterleri

Modelleme ve gelecek tahmini performanslarını karşılaştırmak için kullanılan Mean Absolute Error (MAE) ve Root Mean Square Error (RMSE) kriterleri aşağıdaki şekilde tanımlanmaktadır.

$$MAE_i = \sum_{t=1}^N \frac{|\hat{y}_{it} - y_{it}|}{N} \quad (14)$$

$$RMSE_i = \sqrt{\sum_{t=1}^N \frac{(\hat{y}_{it} - y_{it})^2}{N}} \quad (15)$$

Burada,  $y_{it}$  ve  $\hat{y}_{it}$  sırasıyla  $i$ . varlığın  $t$ . zamandaki gözlem değerini ve modelleme (gelecek tahmin) tahmini değerini göstermektedir. Bu araştırmada Python programlama dili kullanılmıştır.

### Veri

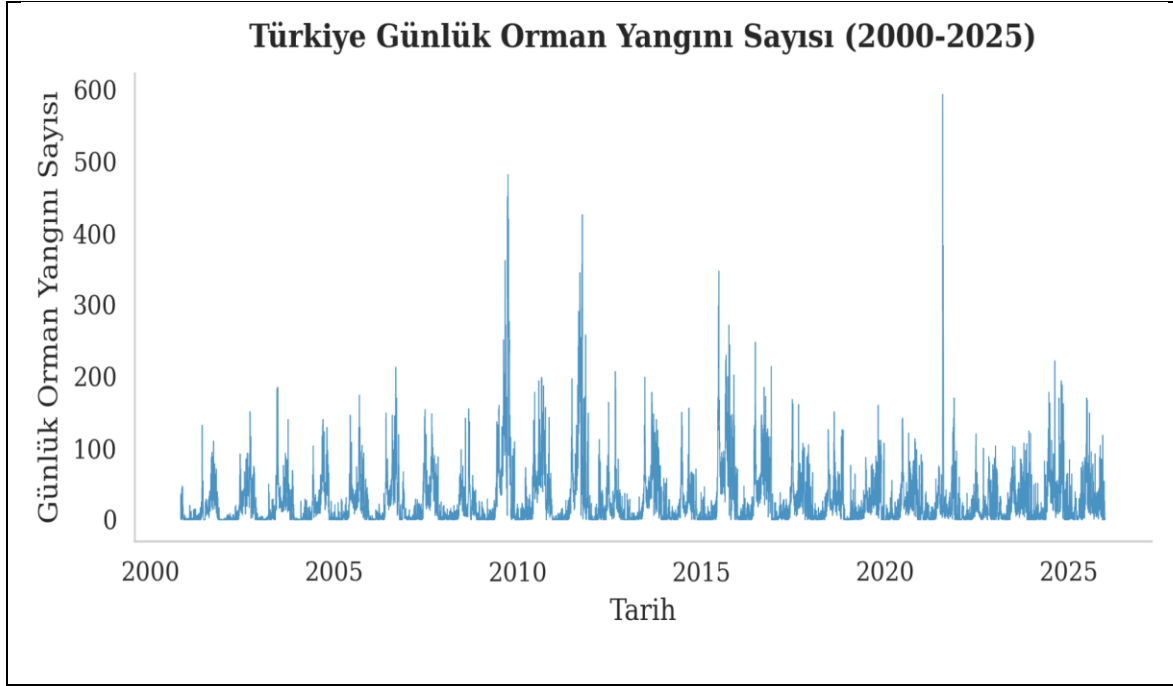
Bu çalışmada, Kasım 2000 – Ocak 2026 dönemi Türkiye’deki günlük orman yangını sayıları kullanılmıştır. Araştırma veriler, NASA tarafından yayınlanan FIRMS (NASA, 2026) (Fire Information for Resource Management System) veri bankasından alınmıştır. Türkiye’deki orman sayısı yangın sayısı istatistikleri Çizelge 1’de gösterilmiştir.

**Çizelge 1.** Orman yangını sayısı betimleyici istatistikleri

| İstatistik     | Değer  |
|----------------|--------|
| Gözlem Sayısı  | 9190   |
| Ortalama       | 27.37  |
| Standart Sapma | 40.44  |
| Minimum        | 0      |
| 1. Çeyrek (Q1) | 2      |
| Medyan (Q2)    | 12     |
| 3. Çeyrek (Q3) | 38     |
| Maksimum       | 594    |
| Çarpıklık      | 3.437  |
| Basıklık       | 20.488 |

Çizelge 1’de, günlük ortalama yangın sayısı 27.37 ve standart sapması 40.44 olarak gözlemlenmiştir. Standart sapmanın ortalamadan büyük olması bazı günlerde gözlenen orman yangını sayısının beklenenin çok üzerinde olduğunu göstermektedir. Bazı günlerde yangın gözlenmez iken en yüksek adetli orman yangın sayısının 594 olduğu görülmektedir. Çarpıklık katsayısının 3’ün üzerinde oluşunun verinin sağa çarpıklığa sahip olması anlamına gelmektedir.

Bu durum, günlük orman yangın sayısının beklenenden daha nadir sıklıkta yüksek sayılarda orman yangını meydana geldiğini göstermektedir.



**Şekil 1.** Günlük orman yangını sayısı zaman serisi grafiği

Türkiye’deki günlük orman yangını sayısı zaman serisi grafiği Şekil 1’de gösterilmiştir. Orman yangını sayısı en fazla 594 adet ile 2011 yılının 7. ayında gözlemlenmiştir. Buna ek olarak, yaz mevsimlerinde orman yangını sayılarında artış olduğu gözlenmektedir.

## BULGULAR

Bu bölümde, CatBoost, LightGBM, LSTM, Random Forest ve XGBoost modellerinin modelleme ve gelecek 2 yıllık tahmin performans karşılaştırmaları MAE ve RMSE kriterleri Çizelge 2’de sunulmuştur.

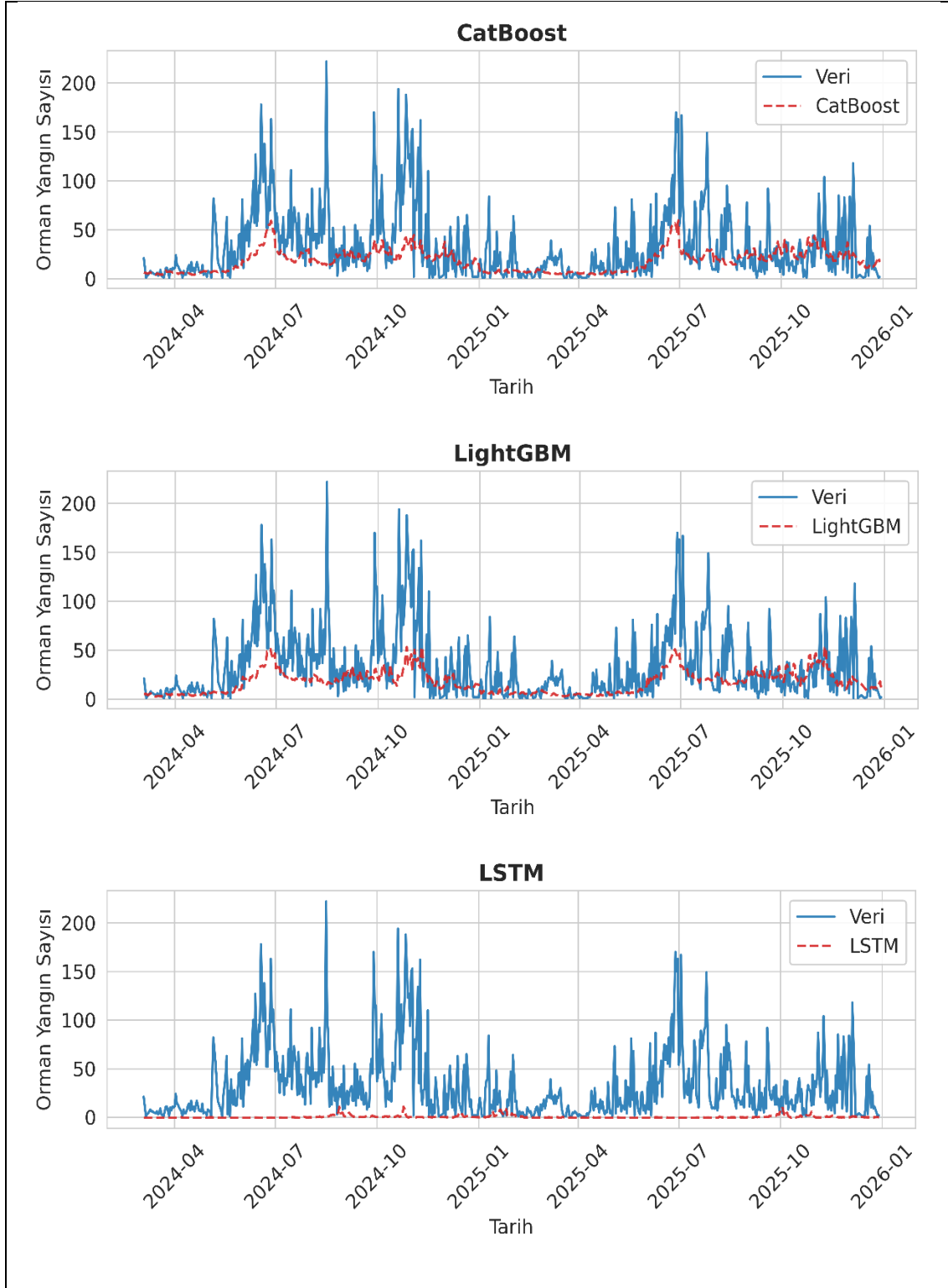
Modelleme aşamasında veriler MinMaxScaler ile [0,1] aralığına normalize edilmiştir. Aykırı değerler, üç katlı çeyrekler arası aralık (3xIQR) yöntemiyle temizlenmiş, oluşan eksik değerler ise zamansal sürekliliği korumak amacıyla ardışık doldurma (bfill+ffill) teknikleriyle tamamlanmıştır. Modeller, Nisan 2024 - Ocak2026 dönemini kapsayan eğitim setinde zaman serisine duyarlı çapraz doğrulama (TimeSeriesSplit, k=5) ve hiperparametre optimizasyonu (GridSearchCV ve Optuna) ile optimize edilmiştir. Gelecek tahmini periyodunda ise Nisan 2024 - Ocak 2026 dönemi için sabit pencere kayan pencere (rolling window walk-forward validation) yaklaşımıyla her gün için tek adım ileriye (one-step-ahead) tahmin yapılmış ve modeller yeniden eğitilerek test edilmiştir. Tahmin edilen ve gerçek orman yangını sayısı arasındaki MAE, RMSE ve  $R^2$  değerleri, bu süre zarfında hesaplanmış ve Çizelge 2’de verilmiştir.

**Çizelge 2.** Orman yangını sayısının modellenmesi ve tahmini için model karşılaştırma sonuçları

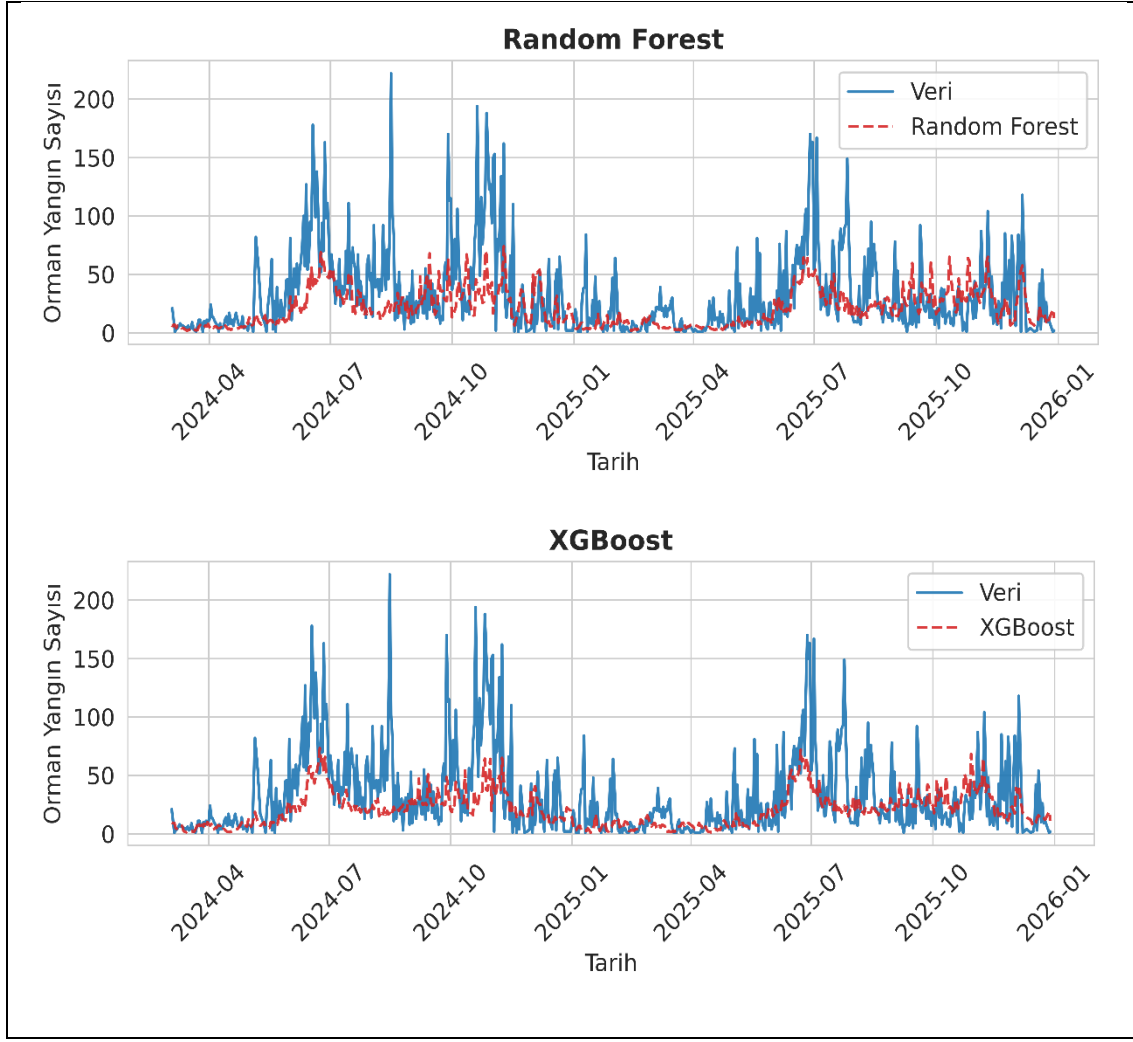
| Model         | Modelleme |        |       | Tahmin |        |       |
|---------------|-----------|--------|-------|--------|--------|-------|
|               | MAE       | RMSE   | $R^2$ | MAE    | RMSE   | $R^2$ |
| CatBoost      | 13.085    | 22.357 | 0.720 | 17.227 | 26.519 | 0.424 |
| LightGBM      | 10.892    | 18.775 | 0.802 | 17.076 | 26.380 | 0.430 |
| LSTM          | 15.141    | 28.301 | 0.551 | 18.716 | 28.922 | 0.318 |
| Random Forest | 12.883    | 22.733 | 0.710 | 16.775 | 26.166 | 0.440 |
| XGBoost       | 13.186    | 23.629 | 0.687 | 16.898 | 26.266 | 0.435 |

Çizelge 2’de görüldüğü gibi, LightGBM’nin orman yangını sayısını modelleme aşamasında en düşük MAE ve RMSE değerlerine sahip olmasından dolayı diğer modellerden daha iyi modelleme performansı gösterdiği söylenebilir. RMSE

(MAE) açısından, LightGBM'nin diğer modeller arasında en düşük performansa sahip LSTM'den %50.73 (%39.01) daha iyi olduğu görülmektedir. LightGBM'nin mevcut veriyi açıklama oranı %80.2 gibi yüksek bir yüzdelik orana sahiptir. Özetle, Türkiye'deki orman yangını sayısı modellenmesi için en iyi model LightGBM'dir.



Şekil 2. Modellerin orman yangın sayısı tahmin grafikleri



**Şekil 2.** Modellerin orman yangını sayısı tahmin grafikleri (devamı)

Random Forest'in günlük orman yangını sayısı tahmin performansının diğer modellerden daha iyi olduğu görülmüştür. Ayrıca, RMSE (MAE) açısından, tahmin prosedüründe 2 yıllık günlük orman yangını sayısı için Random Forest %10.53 (%11.57) oranında LSTM modelini geliştirmektedir. Random Forest'nin mevcut tahmin verisini açıklama oranı %44 ile yeterli olduğu söylenebilir. Sonuç olarak, Türkiye'deki orman yangını sayısı tahmini için en iyi model Random Forest'tir.

Çizelge 2'de  $R^2$  değerlerine göre genel olarak orman yangını sayısı verilerinin araştırmada kullanılan makine öğrenmesi modelleri ile modelleme tahmininin gelecek tahmini konusundaki performans sonuçlarına bakıldığında söz konusu makine öğrenmesi modellerinin gelecek orman yangını sayısı tahmini konusunda ani değişimler ve uç değerlerin tahmin edilmesinde yetersiz kaldığı görülmüştür. Bu durum yangın sayılarının sadece geçmiş verilere dayalı otokorelasyonla açıklanamayacak kadar stokastik bir yapıda olmasından dolayı makine öğrenmesi modellerinin yetersiz kalmasından daha çok, yangın sayısı konusunda veri setindeki kısıtların ve doğa olaylarının bir sonucu olarak değerlendirilmesi gerekmektedir.

Modellerin karşılaştırmalı orman yangını sayısı tahmin grafiği Şekil 2'de verilmiştir. Random Forest tahminleri diğer modellerin tahminlerine kıyasla daha etkin bir performans gösterdiği görülmüştür. Fakat, LSTM modelinin tahmin performansı konusunda yetersiz olduğu görülmüştür. Şekil 2'den de görüldüğü gibi genel olarak diğer modellerin performanslarının birbirine benzer olduğu söylenebilir.

## TARTIŞMA VE SONUÇ

Bu araştırmada, ülkemizdeki gelecek nesilleri orman yangınları konusunda bilinçlendirmek amacıyla ilk defa olarak Türkiye'deki Kasım 2000 ile Ocak 2026 tarihleri arasındaki orman yangını sayısının modellenmesi ve tahmini makine öğrenmesi yöntemlerinden CatBoost, LightGBM, LSTM, Random Forest ve XGBoost modelleri ile gerçekleştirilmiştir. Genel olarak, makine öğrenmesi yöntemlerinin orman yangını sayısı modellemesi ve gelecek tahmini konusunda etkin bir performansa sahip olduğu görülmüştür. Buna ek olarak, LightGBM modeli orman yangını sayısını modellemede ve Random Forest modeli ise gelecek tahmini konusunda daha etkin performansa sahip olduğu görülmüştür.

Bu araştırmanın bilimsel sonuçları ışığında alınacak bütüncül önlemler, hem mevcut ekosistem hizmetlerinin korunmasını sağlayacak hem de gelecek nesiller için yaşanabilir bir çevrenin teminatı olacaktır. Bu araştırmadaki gibi bilimsel anlamda makine öğrenmesi modellerinin kullanılması, erken uyarı sistemleri ile ormanları çeşitli tehlikelere karşı korumak, gerekli önlemleri zamanında almak ve meydana gelecek tehlikelere karşı orman yangını ile mücadelede hazır bulunmak konusunda araştırmacılara yön gösterici niteliktedir.

## KAYNAKLAR

- Ateşoğlu, A., Melemez, K., & Uğur, B. (2016). Orman yangınına hassas bölgelerde arasöz ile müdahale oranının belirlenmesi (Bartın Orman İşletmesi örneği). *Artvin Çoruh Üniversitesi Orman Fakültesi Dergisi*, 16(2), 132-143. <https://doi.org/10.17474/acuofd.24014>
- Calda, B., An, N., Turp, M.T., & Kurnaz, L. (2020). İklim değişikliğinin Akdeniz havzasındaki orman yangınlarına etkisi. *International Journal of Advances in Engineering and Pure Sciences*, 32(1), 15-32. <https://doi.org/10.7240/ieps.571001>
- Chen, T., & Guestrin, C. (2016). XGBoost: a scalable tree boosting system. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Vol:13-17, August-2016. Association for Computing Machinery, 8:785-794.
- Çökekoğlu, S. (2023). *Uzaktan algılama sistemleri ile yangına hassas bölgelerin tahmin edilmesi* (Yüksek Lisans Tezi). Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü, Isparta.
- Demirel, Y., & Türk, T. (2023). Türkiye'de 2015 ile 2022 yılları arasında meydana gelen orman yangınlarının coğrafi bilgi sistemleri ile zamansal ve mekânsal analizi. *Jeodezi ve Jeoinformasyon Dergisi*, 10(2), 136-150. <https://doi.org/10.9733/JGG.2023R0010.T>
- Dorogush, A.V., Ershov, V., & Gulin, A. (2018). CatBoost: gradient boosting with categorical features support. *Arxiv*, 1-7. <http://arxiv.org/abs/1810.11363>
- Eşidir, K.A. (2025). Makine öğrenimi modelleri ile yetişkin eğitimi analizi: modellerin karşılaştırmalı performansı. *Elektronik Sosyal Bilimler Dergisi*, 24(2), 946-964. <https://doi.org/10.17755/esosder.1589887>
- Güngöroğlu, C., Özkara, Z.U., & Tutmaz, V. (2024). Türkiye'de orman yangın yönetimi: sorunlar ve çözüm önerileri. *Memleket Siyaset Yönetim*, 19(43), 517-570.
- NASA. (2026). <https://firms.modaps.eosdis.nasa.gov/> Erişim Tarihi: 10.01.2026.
- Özenen Kavlak, M., Kurtipek, A., & Çabuk, S.N. (2020). Coğrafi bilgi sistemleri ile orman yangını risk haritası oluşturulması: Ören Örneği. *Resilience*, 4(1), 33-54. <https://doi.org/10.32569/resilience.597887>
- Şahan, C., & Kaya, İ. (2022). Türkiye ve Avrupa ülkeleri orman yangınlarının bazı değişkenler açısından değerlendirilmesi. *ODÜSOBİAD*, 12(1), 1-14. <https://doi.org/10.48146/odusbadiad.1013462>
- Tomanbay, M. (2008). *Dünyada su ve küresel ısınma sorunu*. Ankara: Phoenix Yayınevi.
- Tunç, B.N., & Atasever, Ü.H. (2025). CatBoost algoritmasının taşınmaz değerlemede kullanımı: Bayesian hiperparametre optimizasyonu ile karşılaştırmalı analiz. *Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi*, 14(2), 668-679. <https://doi.org/10.28948/ngumuh.1597288>
- Usta, G. (2023). Afet yönetimi odağında orman yangınlarının değerlendirilmesi. *Afet ve Risk Dergisi*, 6(1), 18-35.
- Yeşilyurt, S.E., & Dalkılıç, H.Y. (2021). XGBoost ve Gradient Boost Machine ile günlük nehir akımı tahmini. In *3rd International Symposium of III Engineering Applications on Civil Engineering and Earth Sciences*, 22-24 Eylül, Bartın.
- Yıldırım, O., Gunay, F.B., & Yağanoğlu, M. (2023). Makine öğrenmesi yöntemleriyle orman yangını tahmini. *Journal of the Institute of Science and Technology*, 13(3), 1468-1481. <https://doi.org/10.21597/jist.1249908>

- Yılmaz, O.S., Oruç, M.S., Ateş, A.M., & Gülgen, F. (2021). Orman yangın şiddetinin Google Earth Engine ve Coğrafi bilgi sistemleri kullanarak analizi: Hatay-Belen örneği. *Journal of the Institute of Science and Technology*, 11(2), 1519-1532. <https://doi.org/10.21597/jist.817900>
- Zhang, T., Song, S., Li, S., Ma, L., Pan, S., & Han, L. (2019). Research on gas concentration prediction models based on LSTM multidimensional time series. *Energies*, 12(1), 161. <https://doi.org/10.3390/en12010161>