Passive Defence
Vol. 17, No.1, Spring 2026, Serial No. 65
ISSN: 2980-8030 , E-ISSN: 2008-6849

Estimate the Burned Area Caused by Forest Fire with an Approach Based on Machine Learning

Ali Zhaleh Karimi[1], Mohammad Ali Javadzade[2*]

[1]Master's Student in artificial intelligence and robotics, Faculty of artificial intelligence and cognitive sciences, Imam Hossein Comprehensive University, Tehran, Iran. Email Address: azhkarimi@ihu.ac.ir

[2]Correspondence: Assistant Professor, Faculty of artificial intelligence and cognitive sciences, Imam Hossein Comprehensive University, Tehran, Iran Email Address: javadzade@ihu.ac.ir

ARTICLE INFO

Article history:

Article Type: Research paper

Received: 13 March 2025

Received in revised form: 2 June 2025

Accepted: 20 September 2025

Available online: 20 May 2026

Keywords:

Forest Fire

Crisis Management

Fire Management

Forest Meteorology

FWI

Regression

XGBoost

ABSTRACT

Forest fires are a serious threat to natural resources and critical infrastructure, causing extensive economic, environmental, and security damages annually. This research aims to strengthen passive defense and crisis management systems by presenting an innovative artificial intelligence-based approach for predicting burned area extent when facing forest fires. In this research, the XGBRegressor algorithm within machine learning framework was used to accurately predict the area damaged by fire using meteorological, geographical, and temporal data. The model achieves RMSE and MAE evaluation metrics of 61.602 and 12.273 respectively, enabling resource and equipment planning for fire suppression. The research findings demonstrate that through spatial-temporal fire analysis, critical points and high-risk periods can be identified. Additionally, the potential fire area can be estimated with minimal error, allowing targeted allocation of defensive resources. This approach, in line with intelligent territorial planning and proactive crisis management, enhances ability to protect critical infrastructures adjacent to forest areas. Moreover, this model can serve as a complement to an early warning system within the framework of passive defense and environmental resilience, playing a crucial role in reducing damages and preserving national resources. Although this study focuses on data from outside Iran, the presented research has the potential for localization to Iranian forests and can be used as a strategic tool in the country's passive defense.

Cite this article: A. Zhaleh Karimi and M. A. Javadzade, "Estimate the Burned Area Caused by Forest Fire with an Approach Based on Machine Learning," Journal of Passive Defence, vol. 17, no. 1, pp. 19-35, 2026.
DOI: https://doi.org/10.47176/pd.2026.1526

Publisher: Imam Hossein University.

Introduction

Forest fires represent one of the most serious environmental threats globally, causing extensive economic losses, environmental degradation, and security challenges annually. From the perspective of passive defense and crisis management, forest fires can have devastating impacts on critical infrastructure, including energy transmission lines, communication networks, and human settlements adjacent to forested areas. Effective crisis management requires intelligent decision support systems capable of estimating fire extent and optimizing resource allocation.

The relationship between meteorological conditions (temperature, wind, relative humidity, rainfall) and fire behavior has been well-established, as weather factors significantly influence both the ignition and spread of forest fires. The Canadian Fire Weather Index (FWI) system, which comprises six components (FFMC, DMC, DC, ISI, BUI, FWI), provides a standardized framework for rating fire danger based on fuel moisture codes and spread indices.

Previous research has explored various data mining and machine learning approaches for predicting burned areas. Cortez and Morais (2007) investigated five different data mining techniques including Support Vector Machines (SVM) and Random Forests on real-world data from Portugal's Montesinho Natural Park, finding that SVM with four meteorological features performed best for predicting small fires. Niranjan et al. (2019) proposed multiple machine learning approaches including Naïve Bayes, Decision Trees, SVR, Random Forests, Stochastic Gradient Descent, and Bagging for predicting burned areas using the same dataset.

This research aims to develop a novel artificial intelligence-based approach for predicting the burned area extent of forest fires using meteorological, geographical, and temporal data. The primary objectives are: (1) to perform comprehensive data analysis and preprocessing to handle non-numeric features and skewness issues; (2) to evaluate over 20 regression algorithms including linear models, tree-based methods, and ensemble techniques; (3) to optimize the XGBoost algorithm through hyperparameter tuning; and (4) to assess model performance using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) metrics. The ultimate goal is to provide an effective tool for passive defense systems to enable targeted resource allocation and proactive crisis management.

Results and Discussion

The study utilizes the Forest Fires dataset from the UCI Machine Learning Repository, containing data from Montesinho Natural Park in the Trás-os-Montes region of northeastern Portugal. This park features a supra-Mediterranean climate with average annual temperatures ranging from 8-12°C, high biodiversity, and data collected from January 2000 to December 2003, comprising 517 instances.

Statistical analysis revealed that the dataset contains no missing values (517 instances for all features). However, the target variable (Area) exhibits significant positive skewness (12.8469) and high kurtosis (194.1407), indicating that most fires are small (247 samples with zero area, meaning less than 100 m² burned), with few large fires. Similar skewness was observed in FFMC (-6.5756), ISI (2.5363), and Rain (19.8163) features.

Table 1: Dataset Features and Ranges

Feature	Description	Range
X	X-axis spatial coordinate	1 to 9
Y	Y-axis spatial coordinate	1 to 9
Month	Month of year	January to December
Day	Day of week	Monday to Sunday
FFMC	Fine Fuel Moisture Code	18.7 to 96.20
DMC	Duff Moisture Code	1.1 to 291.3
DC	Drought Code	7.9 to 860.6
ISI	Initial Spread Index	0 to 56.10
Temperature	Outside temperature (°C)	2.2 to 33.30
Relative Humidity	Outside relative humidity (%)	15 to 100
Wind	Wind speed (km/h)	0.40 to 9.40
Rain	Rainfall (mm/m²)	0 to 6.4
Area	Total burned area (hectares)	0 to 1090.84

To prepare the data for machine learning algorithms, several preprocessing steps were implemented:

1. Encoding categorical variables: Months and days were converted to numerical values using ordinal encoding
2. Skewness correction: Log transformation ($y = \ln(x+1)$) was applied to the Area feature and other skewed features to improve symmetry and normalize distributions
3. Feature selection: The dataset was divided into six subsets based on feature categories:
    - LD (Location-Date): X, Y, Month, Day
    - F (FWI components): FFMC, DMC, DC, ISI
    - M (Meteorological): Temperature, Relative Humidity, Wind, Rain

- LDF (Location-Date-FWI): Combined LD + F
- LDM (Location-Date-Meteorological): Combined LD + M
- LDFM (Full dataset): All features

The meteorological subset (M) was identified as most practical for real-world applications since these features can be obtained directly from weather sensors without accumulated calculations.

Over 20 regression algorithms were evaluated using 67% training and 33% testing split, with performance measured by MAE and RMSE (lower values indicate better performance). Table 5 in the original article presents comprehensive results across all subsets.

Table 2: Key findings from algorithm comparison:

Algorithm	Best MAE (RMSE)	Subset
XGBoost (Proposed)	12.2571 (61.6350)	LD
XGBoost (Full dataset)	12.2731 (61.6028)	LDFM
Support Vector Machine	12.1901 (61.7391)	M
Random Forest	12.9408 (60.5915)	M
Bagging	12.6547 (60.5791)	F
Gaussian Regression	12.2247 (61.8716)	LDFM

The XGBoost algorithm consistently outperformed all other algorithms across all six subsets, demonstrating superior predictive capability. The best performance was achieved on the LD subset (spatial-temporal features) with MAE = 12.2571 and RMSE = 61.6350. However, for practical deployment, the M subset (meteorological features only) is preferred due to real-time availability.

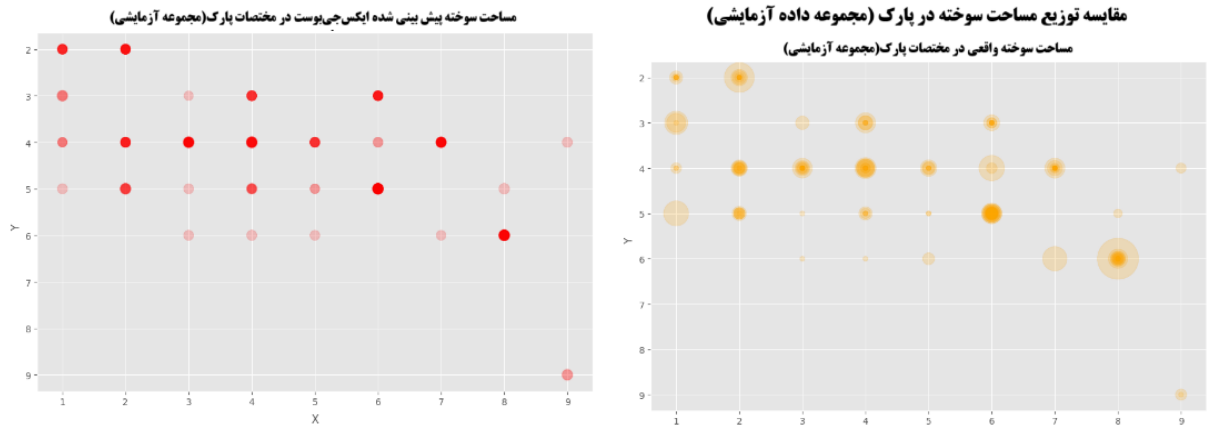


Figure 1 (adapted from original Figure 10) shows the XGBoost model's prediction performance compared to actual burned areas across park coordinates:

- Strengths: Good overall identification of fire-prone areas; acceptable estimation of fire intensity
- Weakness: Tendency for false predictions (predicting fire where actual fire is small or non-existent)

Conclusion

This research successfully demonstrates that machine learning, particularly the XGBoost algorithm, can effectively predict forest fire burned areas using meteorological, spatial, and temporal data. The key findings and contributions are:

1. Optimal Algorithm: XGBoost significantly outperformed over 20 alternative regression algorithms including linear models, SVM, Random Forest, and other ensemble methods across all feature subsets.
2. Best Performance Metrics: The optimized XGBoost model achieved RMSE = 61.602 and MAE = 12.273 on the full dataset, with even better performance (MAE = 12.257) on spatial-temporal features alone.
3. Practical Applicability: The meteorological-only subset (M) achieved strong performance (RMSE = 61.599, MAE = 12.276), enabling real-time predictions using only readily available weather sensor data without accumulated calculations.
4. Critical Features: Temperature, wind speed, relative humidity, and rainfall were identified as the most influential factors, in descending order of importance.
5. Spatial-Temporal Patterns: Fire incidence peaks in August-September and on Sundays, with specific high-risk grid zones identified within the park.

This research represents a significant step toward leveraging artificial intelligence for passive defense applications in environmental crisis management. The proposed XGBoost-based approach achieves high accuracy in predicting forest fire burned areas using readily available meteorological data. By localizing this technology for Iran's diverse forest ecosystems, it can substantially enhance the resilience of critical infrastructure against natural threats such as forest fires, contributing to national security and environmental protection. The model's ability to provide rapid, data-driven burned area estimates enables decision-makers to allocate limited firefighting resources more effectively, potentially reducing economic losses, environmental damage, and risks to human life.

پدافند غیرعامل

سال هفتم، شماره ۱، بهار ۱۴۰۵، (پیاپی ۶۵): صص ۳۵-۱۹

شاپای چاپی: ۶۹۴۹-۲۰۰۸ | **شاپای الکترونیکی:** ۸۰۳۰-۲۹۸۰

پیش‌بینی مساحت ناحیه سوخته ناشی از آتش‌سوزی در جنگل با رویکردی مبتنی بر یادگیری ماشین

علی ژاله کریمی[۱]، محمدعلی جوادزاده[۲*]

[۱] دانشجوی کارشناسی ارشد هوش مصنوعی و رباتیکز، دانشکده هوش مصنوعی و علوم شناختی، دانشگاه جامع امام حسین(ع)، تهران، ایران. رایانامه: azhkarimi@ihu.ac.ir

[۲] استادیار، دانشکده هوش مصنوعی و علوم شناختی، دانشگاه جامع امام حسین(ع)، تهران، ایران. (نویسنده مسئول). رایانامه: javadzade@ihu.ac.ir

چکیده

آتش‌سوزی جنگل یکی از تهدیدات جدی برای منابع طبیعی و زیرساخت‌های حیاتی کشور است که سالانه خسارات گسترده اقتصادی، زیست‌محیطی و امنیتی به همراه دارد. این پژوهش باهدف تقویت سامانه‌های پدافند غیرعامل و مدیریت بحران در مواجهه با آتش‌سوزی‌های جنگلی، رویکردی نوین مبتنی بر هوش مصنوعی برای پیش‌بینی مساحت ناحیه سوخته ارائه می‌دهد. در این پژوهش، از الگوریتم ایکس‌جی‌بوست (روش پیشنهادی) در چارچوب یادگیری ماشین برای پیش‌بینی دقیق مساحت مناطق آسیب‌دیده از آتش‌سوزی با استفاده از داده‌های هواشناسی، جغرافیایی و زمانی استفاده شده است. این مدل در معیارهای ارزیابی جذر میانگین مربعات خطاها و میانگین خطاهای مطلق به ترتیب مقادیر ۶۱/۶۰۲ و ۱۲/۲۷۳ را به دست می‌آورد؛ که امکان برنامه‌ریزی منابع و تجهیزات موردنیاز برای مهار آتش را فراهم می‌کند. یافته‌های این پژوهش نشان می‌دهد که با تحلیل مکانی - زمانی آتش‌سوزی‌ها، می‌توان نقاط بحرانی و زمان‌های پرخطر را شناسایی کرد، همچنین مساحت آتش‌سوزی احتمالی را با خطای اندکی برآورد نمود و منابع پدافندی را به‌صورت هدفمند تخصیص داد. این رویکرد در راستای آمایش سرزمینی هوشمند و مدیریت بحران پیشگیرانه، امکان حفاظت از زیرساخت‌های حیاتی مجاور مناطق جنگلی را تقویت می‌کند. همچنین، این مدل می‌تواند به‌عنوان مکمل برای یک سامانه هشدار سریع در چارچوب پدافند غیرعامل و تاب‌آوری زیست‌محیطی، نقش مهمی در کاهش خسارات و حفظ منابع ملی ایفا کند. گرچه این مطالعه بر داده‌های خارج از ایران متمرکز است، اما پژوهش ارائه شده قابلیت بومی‌سازی برای جنگل‌های ایران را دارد و می‌تواند به‌عنوان ابزاری راهبردی در پدافند غیرعامل کشور مورداستفاده قرار گیرد.

مشخصات مقاله

تاریخچه مقاله:

نوع مقاله: علمی پژوهشی

دریافت: ۱۴۰۳/۱۲/۲۳

بازنگری: ۱۴۰۴/۰۳/۱۲

پذیرش: ۱۴۰۴/۰۶/۲۹

ارائه آنلاین: ۱۴۰۵/۰۲/۳۰

کلیدواژه‌ها:

آتش‌سوزی جنگل

مدیریت بحران

مدیریت آتش‌سوزی

هواشناسی جنگل

FWI

رگرسیون

XGBoost

استناد: ژاله کریمی، علی، جوادزاده، محمدعلی، " پیش‌بینی مساحت ناحیه سوخته ناشی از آتش‌سوزی در جنگل با رویکردی مبتنی بر یادگیری ماشین"، نشریه پدافند غیرعامل، دوره ۱۷، شماره۱، صفحات ۳۵-۱۹، ۱۴۰۵. DOI: https://doi.org/10.47176/pd.2026.1526

ناشر: دانشگاه جام امام حسین(ع).

۱- مقدمه

یکی از نگرانی‌های مهم زیست‌محیطی وقوع آتش‌سوزی‌های جنگلی است که در جهان بسیار فراگیر است و باعث آسیب‌های اقتصادی و زیست‌محیطی می‌شود [۱]. چنین پدیده ای به دلایل متعدد طبیعی مانند رعد و برق و در برخی موارد عوامل انسانی ایجاد می‌شود و با وجود افزایش هزینه‌های دولت‌ها برای کنترل این فاجعه، سالانه میلیون‌ها هکتار جنگل در سراسر جهان از بین می‌رود [۲].

از منظر پدافند غیرعامل، آتش‌سوزی‌های جنگلی می‌توانند تأثیرات مخربی بر زیرساخت‌های حیاتی، خطوط انتقال انرژی و سکونتگاه‌های انسانی مجاور مناطق جنگلی داشته باشند [۳]. مدیریت بحران ناشی از این پدیده نیازمند سامانه‌های پشتیبان تصمیم هوشمند است که امکان برآورد گستره آتش و تخصیص بهینه منابع را فراهم سازد.

یکی از راه‌حل‌های کنترل پدیده آتش‌سوزی، استفاده از ابزارهای خودکار مبتنی بر حسگرهای محلی، مانند داده‌های ارائه شده توسط ایستگاه‌های هواشناسی است [۴]. در واقع، شرایط هواشناسی (مثلاً دما، باد و...) بر گستردگی و سرعت گسترش آتش‌سوزی‌های جنگل‌ها تأثیر می‌گذارند[۵].

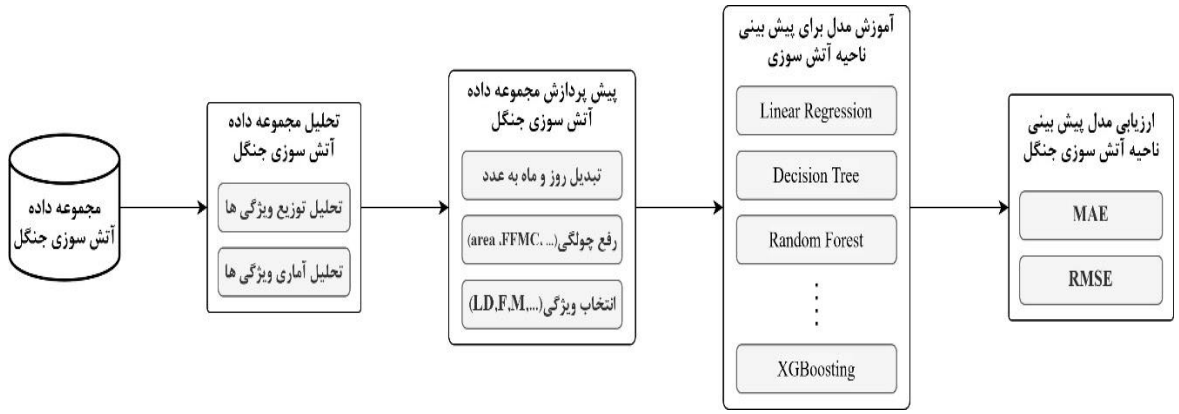
در مقاله کورتز و مورایس [۵] یک رویکرد داده‌کاوی برای پیش‌بینی منطقه سوخته آتش‌سوزی‌های جنگلی بررسی شده است. پنج تکنیک مختلف داده‌کاوی، به‌عنوان‌مثال ماشین‌های بردار پشتیبانی[1] و جنگل‌های تصادفی، و چهار تنظیم انتخاب ویژگی متمایز (با استفاده از اجزای مکانی، زمانی، شاخص آب‌وهوای آتش‌سوزی و ویژگی‌های آب‌وهوا)، بر روی داده‌های دنیای واقعی جمع‌آوری‌شده از پارک طبیعی مونتسینهو کشور پرتغال استفاده شده است. بهترین الگوریتم در این مقاله از یک ماشین بردار پشتیبان و چهار ویژگی هواشناسی (یعنی دما، رطوبت نسبی، باران و باد) استفاده می‌کند و می‌تواند منطقه سوخته آتش‌سوزی‌های کوچک را که فراوان‌تر هستند، پیش‌بینی کند.

نیرانجان و همکاران [۶] رویکردهای یادگیری ماشینی مختلفی مانند بیز ساده، درختان تصمیم، ماشین بردار پشتیبان رگرسیونی[2]، جنگل تصادفی، نزول گرادیان تصادفی و بگینگ را برای پیش‌بینی میزان زمین سوخته در جنگل پیشنهاد می‌کنند. در این مطالعه مدل پیش‌بینی با استفاده از داده‌های آتش‌سوزی ایجاد شده در پارک طبیعی مونتسینهو کشور پرتغال ساخته شده است.

ما در این مقاله، از روش‌های یادگیری ماشین در هوش مصنوعی برای برآورد مساحت منطقه سوخته آتش‌سوزی جنگل‌ها به کمک شرایط آب‌وهوایی استفاده شده است. برای رسیدن به برآورد مطلوب در یادگیری ماشین ابتدا باید به تحلیل مجموعه‌داده پرداخت و باتوجه‌به ویژگی‌های مجموعه‌داده پیش‌پردازش‌های لازم برای اعمال بر روی آن را انتخاب نمود. ممکن است برخی از مجموعه‌ها دارای داده‌های پرت و نامناسب برای الگوریتم باشند (مانند ویژگی‌های غیرعددی) در این صورت سعی می‌شود تاحدامکان این ویژگی‌ها به حالت‌های مطلوب برای الگوریتم تبدیل شود یا از مجموعه‌داده حذف شود. پس از پیش‌پردازش و آماده‌سازی مجموعه‌داده نوبت به آموزش مدل به کمک الگوریتم و ارزیابی آن با استفاده از معیارهای ارزیابی می‌رسد.

در این مسئله به دلیل ماهیت پیوسته وسعت منطقه سوخته، استفاده از راه‌حل‌های رگرسیونی برای برآورد و تخمین مساحت سوختگی مناسب است. برای رسیدن به راه‌حل مناسب این مسئله، انتخاب ویژگی بر روی مجموعه‌داده انجام شده است و عملکرد بیش از ۲۰ الگوریتم رگرسیون عبارت اند از رگرسیون خطی، ماشین بردار پشتیبان رگرسیونی ، درخت تصمیم، جنگل تصادفی، ایکس‌جی‌بوست[3] و... در این مقاله ارزیابی شده است. روش ایکس‌جی‌بوست یکی از راه‌حل‌های رگرسیونی است که معمولا نتایج بهتری ارائه می‌دهد[۷]. در این مقاله نیز مدل ایکس‌جی‌بوست با هایپرپارامترهای تنظیم شده، عملکرد خوبی نسبت به دیگر الگوریتم‌ها از خود نشان داده است. میزان خطای برآورد از ارزیابی این روش بر روی مجموعه‌داده بر اساس معیارهای جذر میانگین مربعات خطاها[4] و میانگین خطاهای مطلق[5] به ترتیب مقادیر ۶۱/۶۰۲ و ۱۲/۲۷۳ را نشان می‌دهد. استفاده از این مدل‌ها برای برآورد وسعت آتش‌سوزی تا حدی برای بهبود مدیریت منابع آتش‌نشانی مفید است و امکان فراهم‌کردن منابع و ابزارهای لازم برای مهار آتش را افزایش

[1] Support Vector Machine
[2] Support Vector Regression
[3] XGBRegressor
[4] RMSE: Root Mean Squared Error
[5] MAE: Mean Absolute Error



شکل (۱): روش پیشنهادی

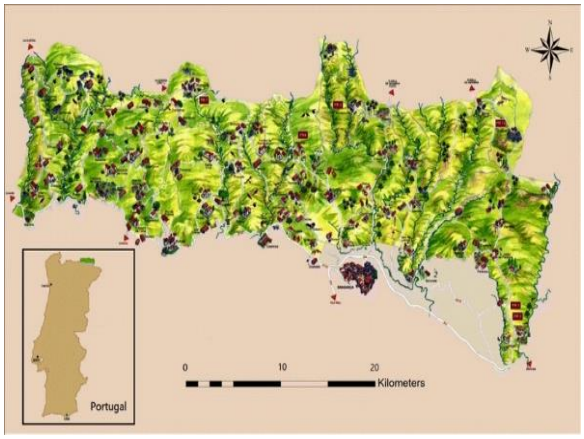
می‌دهد (به‌عنوان‌مثال اولویت بندی اهداف برای تانکرهای هوایی و تعداد نیروی زمینی) [۸].

۲- روش تحقیق

در این پژوهش فرایند حل مسئله به این صورت است که ابتدا بر روی مجموعه‌داده تحلیل آماری و تحلیل توزیع انجام می‌شود تا با استفاده از مشخصاتِ ویژگی‌ها، فضای کلی مجموعه‌داده به دست آید و داده‌های نامطلوب برای آموزش الگوریتم‌های یادگیری ماشین شناسایی شوند. در مرحله بعد پیش‌پردازش روی داده‌ها انجام می‌شود تا مقادیر غیرعددی مانند ماه و روز به مقادیر عددی تبدیل شود و چولگی[1] موجود در برخی ویژگی‌ها به کمک نرمال‌سازی رفع شود. سپس با استفاده از روش انتخاب ویژگی مجموعه‌داده به ۶ زیر مجموعه تقسیم می‌شود. در مرحله بعد الگوریتم‌های رگرسیون با پارامترهای پیش‌فرض از مجموعه‌داده‌ها برای یادگیری و برآورد وسعت منطقه سوخته استفاده می‌کنند. پس از ارزیابی همه الگوریتم‌ها با دو معیار میانگین خطای مطلق و جذر میانگین مربعات خطاها بهترین الگوریتم برای این مسئله انتخاب می‌شود (شکل ۱).

۲-۱- مجموعه‌داده

این مقاله داده‌های آتش‌سوزی جنگل[2] کتابخانه تنسور فلو از پارک طبیعی مونتسینهو[3]، از منطقه تراس او اس – مونتس[4] در شمال شرقی کشور پرتغال را موردبررسی قرار خواهد داد (شکل ۲). این پارک دارای تنوع گیاهی و جانوری بالایی است. در یک اقلیم فوق مدیترانه‌ای، میانگین دمای سالانه در محدوده ۸ تا ۱۲ درجه سانتی‌گراد است. داده‌های مورداستفاده در آزمایش‌ها از ژانویه ۲۰۰۰ تا دسامبر ۲۰۰۳ جمع‌آوری‌شده و با استفاده از دو منبع ساخته شده است [۵].



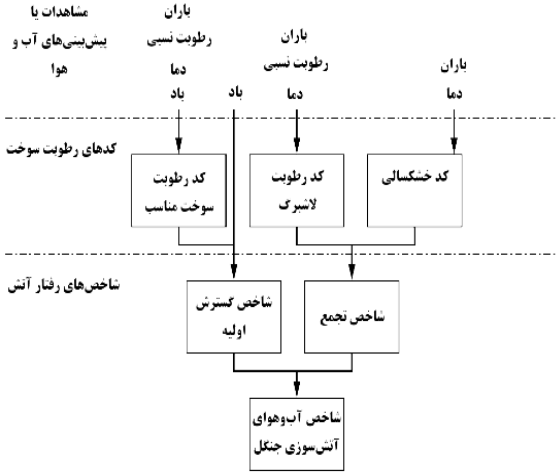
شکل (۲): نقشه پارک طبیعی مونتسینهو

در پایگاه‌داده اول به‌صورت روزانه، هر بار که آتش‌سوزی جنگل رخ می‌دهد، چندین ویژگی ثبت می‌شود، مانند زمان، تاریخ، موقعیت مکانی در یک شبکه ۹ * ۹ (محور x و y شکل ۲)، نوع پوشش گیاهی درگیر، شش مؤلفه سامانه شاخص آب‌وهوای آتش‌سوزی جنگل[5] و کل منطقه سوخته. پایگاه‌داده دوم شامل چندین مشاهده آب‌وهوایی (به‌عنوان‌مثال سرعت باد) بود که با یک دوره ۳۰ دقیقه ای توسط یک ایستگاه هواشناسی واقع در مرکز پارک مونتسینو ثبت شده. این دو پایگاه‌داده در ده‌ها صفحه گسترده مجزا، تحت شکل‌های مجزا ذخیره شدند، و یک تلاش دستی قابل توجه برای ادغام آن‌ها در یک مجموعه‌داده واحد با مجموع ۵۱۷ ورودی انجام شد [۵].

شاخص آب‌وهوای آتش‌سوزی جنگل یک سامانه کانادایی برای رتبه‌بندی خطر آتش‌سوزی است و شامل شش مؤلفه است

[1] Skewness
[2] https://www.tensorflow.org/datasets/catalog/forest_fires
[3] Montesinho natural park
[4] Tr´as-os-Montes
[5] Fire Weather Index: FWI

(شکل ۳): کد رطوبت سوخت مناسب[1] ، کد رطوبت لاشبرگ[2]، کد خشکسالی[3] ، شاخص گسترش اولیه[4]، شاخص تجمع[5] و شاخص آب‌وهوای آتش‌سوزی. سه مورد اول مربوط به کدهای سوخت هستند: کد رطوبت سوخت مناسب نشان دهنده میزان رطوبت سطح است و بر اشتعال و گسترش آتش تأثیر می‌گذارد، درحالی‌که کد رطوبت لاشبرگ و کد خشکسالی نشان دهنده میزان رطوبت لایه‌های آلی کم عمق و عمیق هستند که بر شدت آتش تأثیر می‌گذارند. شاخص گسترش اولیه امتیازی است که با گسترش سرعت آتش مرتبط است، درحالی‌که شاخص تجمع نشان دهنده مقدار سوخت در دسترس است. شاخص آب‌وهوای آتش‌سوزی شاخص شدت آتش‌سوزی است و دو مؤلفه قبلی را ترکیب می‌کند. اگرچه مقیاس‌های مختلفی برای هر یک از عناصر شاخص آب‌وهوای آتش‌سوزی استفاده می‌شود، مقادیر بالا نشان دهنده شرایط شدت سوختن است. همچنین، کدهای رطوبت سوخت به حافظه (تاخیر زمانی) شرایط آب‌وهوایی گذشته نیاز دارند: ۱۶ ساعت برای کد رطوبت سوخت مناسب، ۱۲ روز برای کد رطوبت لاشبرگ و ۵۲ روز برای کد رطوبت لاشبرگ [۹].



شکل (۳): ساختار شاخص آب‌وهوای آتش (FWI)[۹]

جدول (۱) توصیفی از ویژگی‌های داده انتخاب شده را نشان می‌دهد. چهار ردیف اول نشانگر ویژگی‌های مکانی و زمانی هستند. تنها دو ویژگی جغرافیایی در نظر گرفته شده، مقادیر محور X و Y که در آن آتش‌سوزی رخ داده است. دو ویژگی زمانی نیز متغیرهای زمانی ماه و روز هفته درنظر گرفته شده است. میانگین شرایط آب‌وهوایی ماهانه کاملاً متمایز است، درحالی‌که روز هفته نیز می‌تواند بر آتش‌سوزی‌های جنگلی تأثیر بگذارد (مثلاً روزهای کاری در مقابل آخر هفته) زیرا بیشتر آتش‌سوزی‌ها علت انسانی دارند. سپس چهار مؤلفه شاخص آب‌وهوای آتش‌سوزی می‌آیند که مستقیما تحت تأثیر شرایط آب‌وهوایی قرار می‌گیرند (شکل ۲، به‌صورت برجسته). در این مجموعه‌داده شاخص تجمع و شاخص آب‌وهوای آتش‌سوزی دور ریخته شده اند. از پایگاه‌داده ایستگاه هواشناسی، چهار ویژگی آب‌وهوایی دما، رطوبت نسبی، باد و باران انتخاب شده است. برخلاف تاخیرهای زمانی مورداستفاده توسط شاخص آب‌وهوای آتش‌سوزی، در این ویژگی‌ها مقادیر نشان دهنده رکوردهای لحظه ای هستند، به جز متغیر باران است که نشان دهنده بارش انباشته شده در ۳۰ دقیقه قبل است [۵].

جدول (۱): ویژگی‌های مجموعه‌داده

ویژگی	توصیف ویژگی	محدوده
X	مختصات محورX	۱ تا ۹
Y	مختصات محورY	۱ تا ۹
ماه	ماه سال	ژانویه تا دسامبر
روز	روز هفته	دوشنبه تا یکشنبه
FFMC	کد رطوبت سوخت مناسب	۱۸/۷ تا ۹۶/۲۰
DMC	کد رطوبت لاشبرگ	۱/۱ تا ۲۹۱/۳
DC	کد خشکسالی	۷/۹ تا ۸۶۰/۶
ISI	شاخص گسترش اولیه	۰ تا ۵۶/۱۰
دما	دمای بیرون (درجه سانتیگراد)	۲/۲ تا ۳۳/۳۰
رطوبت نسبی	رطوبت نسبی بیرون (%)	۱۵ تا ۱۰۰
باد	سرعت باد بیرون (کیلومتر بر ساعت)	۰/۴۰ تا ۹/۴۰
باران	باران بیرون (میلیمتر بر مترمربع)	۰ تا ۶/۴
مساحت	مساحت کل سوخته (به هکتار)	۰ تا ۱۰۹۰/۸۴

نمودار توزیع ناحیه سوخته در شکل (۷) نشان‌داده‌شده است که نشان دهنده یک چولگی مثبت است، در واقع اکثر آتش‌سوزی‌ها اندازه کوچکی دارند. در مورد مجموعه‌داده حاضر، ۲۴۷ نمونه با مقدار صفر وجود دارد. همه ورودی‌ها نشان دهنده وقوع آتش‌سوزی هستند و مقدار صفر به این معنی است که منطقه کمتر از ۱ هکتار / ۱۰۰ = ۱۰۰ مترمربع سوخته است [۵].

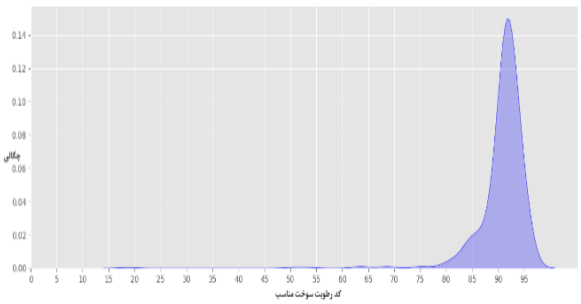
[1] Fine Fuel Moisture Code: FFMC
[2] Duff Moisture Code: DMC
[3] Drought Code: DC
[4] Initial Spread Index: ISI
[5] Buildup Index: BUI

۲-۲- تحلیل مجموعه‌داده

با بررسی تحلیل آماری (جدول ۲) و نمودار توزیع ویژگی‌ها (شکل ۴) می‌توان متوجه شد که امکان چولگی یا انحراف در برخی از ویژگی‌ها وجود دارد که در بخش پیش‌پردازش با نرمال‌سازی داده‌ها به حل این مشکل پرداخته می‌شود. همچنین باتوجه‌به تعداد مقادیر ۵۱۷ برای همه ویژگی‌ها، این مجموعه‌داده دارای مقادیر از دست رفته نمی‌باشد. برای تحلیل مجموعه‌داده ابتدا به تحلیل آماری آن پرداخته می‌شود. این تحلیل خلاصه‌ای از تمایل مرکزی، پراکندگی، و شکل توزیع یک مجموعه‌داده، به‌استثنای مقادیر غیرعددی را فراهم می‌شود. در خروجی این تحلیل میانگین، بیشینه، کمینه، انحراف معیار و میانه برای مقادیر همه ویژگی‌ها محاسبه می‌شود [۱۰](جدول۲).

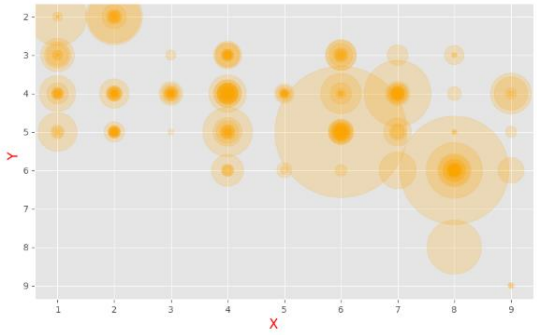
جدول (۲): اطلاعات آماری ویژگی‌های مجموعه‌داده

ویژگی	تعداد	میانگین	انحراف معیار	کمینه	میانه	بیشینه
X	۵۱۷	۴/۶۶۹۲	۲/۳۱۳۸	۱/۰	۴/۰	۹/۰
Y	۵۱۷	۴/۲۹۹۸	۱/۲۲۹۹	۲/۰	۴/۰	۹/۰
FFMC	۵۱۷	۹۰/۶۴۴۷	۵/۵۲۰۱	۱۸/۷	۹۱/۶	۹۶/۲
DMC	۵۱۷	۱۱۰/۸۷۲۳	۶۴/۰۴۶۵	۱/۱	۱۰۸/۳	۲۹۱/۳
DC	۵۱۷	۵۴۷/۹۴	۲۴۸/۰۶۶۲	۷/۹	۶۶۴/۲	۸۶۰/۶
ISI	۵۱۷	۹/۰۲۱۷	۴/۵۵۹۵	۰/۰	۸/۴	۵۶/۱
دما	۵۱۷	۱۸/۸۸۹۲	۵/۸۰۶۶	۲/۲	۱۹/۳	۳۳/۳
رطوبت نسبی	۵۱۷	۴۴/۲۸۸۲	۱۶/۳۱۷۵	۱۵/۰	۴۲/۰	۱۰۰/۰
باد	۵۱۷	۴/۰۱۷۶	۱/۷۹۱۷	۰/۴	۴/۰	۹/۴
باران	۵۱۷	۰/۰۲۱۷	۰/۲۹۶	۰/۰	۰/۰	۶/۴
مساحت	۵۱۷	۱۲/۸۴۷۳	۶۳/۶۵۵۸	۰/۰	۰/۵۲	۱۰۹۰/۸۴



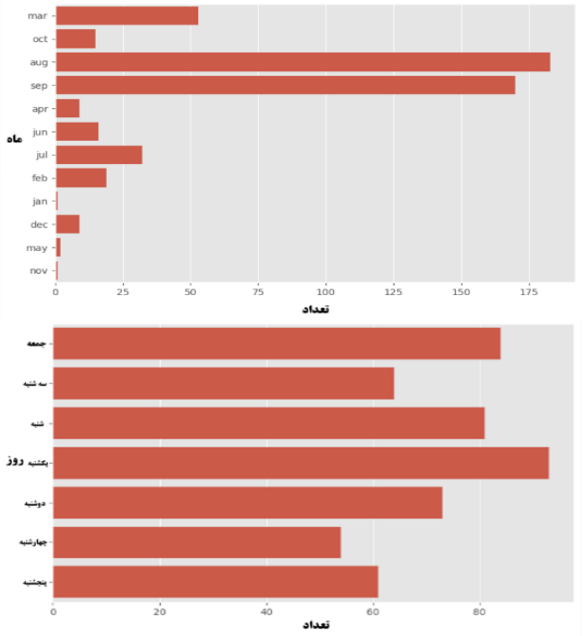
شکل (۴): نمودار توزیع ویژگی کد رطوبت سوخت مناسب

در تحلیل مختصات جغرافیایی (X ،Y) نیز بررسی مختصات آتش‌سوزی می‌تواند در شناسایی مناطق آتش‌خیز پارک کمک کند تا با مجهز کردن این مناطق از آتش‌سوزی بیشتر جلوگیری شود (شکل ۵).



شکل (۵): ناحیه‌های سوخته در مناطق مختلف پارک

برای تشخیص الگوی آتش‌سوزی در سال نیز باید به بررسی دو ویژگی زمانی ماه و روز بپردازیم. همان‌طور که در نمودار شکل (۶) مشاهده می‌شود آتش‌سوزی در ماه‌های سپتامبر و اگوست افزایش می‌یابد و روز یکشنبه به‌عنوان یک روز تعطیل در تقویم میلادی نسبت به روزهای دیگر هفته آتش‌سوزی بیشتری را تجربه کرده است.



شکل (۶): نمودار تعداد آتش‌سوزی در ماه (بالا) و روز (پایین)

۲-۳- پیش‌پردازش مجموعه‌داده

داده‌های غیرعددی مانند روزهای هفته و ماه‌های سال برای تأثیر بر روی الگوریتم نیاز به تبدیل‌شدن به عدد را دارند به همین منظور از تکنیک‌های مختلف برای تبدیل آن‌ها به اعداد استفاده می‌شود [۱۱]. در این مقاله از تکنیک شماره‌گذاری روزهای هفته و ماه‌های سال به ترتیب استفاده شده است.

برای افزایش کارایی الگوریتم ما نیاز به نرمال‌سازی (رفع چولگی‌های موجود) در مجموعه‌داده داریم. همان‌طور که در

بخش قبل بررسی شد در اینجا داده‌ها دارای چولگی راست یا چپ هستند. برای رفع چولگی عملیاتی مانند: ریشه مربع، ریشه مکعب، لگاریتم و غیره برای تبدیل داده‌ها انجام می‌شود. برای بررسی انحراف موجود در ویژگی‌ها از دو معیار چولگی و کشیدگی[1] استفاده شده است.

چولگی درجه اعوجاج از یک توزیع نرمال است. به این معنی که اقلیتی از ارزش‌های بسیار بزرگ وجود دارد. کشیدگی همه چیز در مورد دنباله‌های توزیع است - نه اوج یا تخت بودن. از آن برای توصیف مقادیر افراطی در یک دنباله در مقابل دنباله دیگر استفاده می‌شود. این در واقع اندازه‌گیری پرت‌های موجود در توزیع است. کشیدگی بالا در یک مجموعه‌داده شاخصی است که داده‌ها دارای دنباله‌های سنگین یا خارج از آن هستند [۱۲].

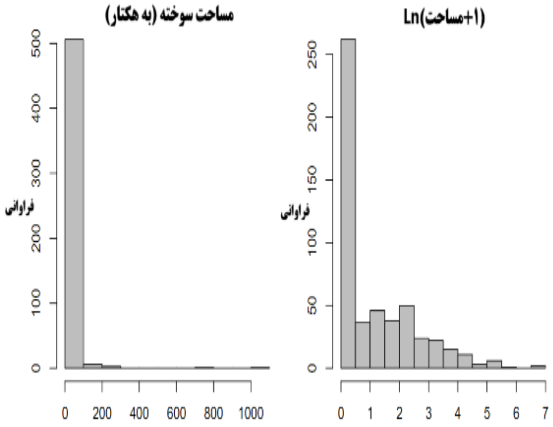
اگر چولگی مثبت باشد، داده‌ها به طور مثبت انحراف یا انحراف‌به‌راست دارند، به این معنی که دنباله راست توزیع طولانی‌تر از چپ است. اگر چولگی منفی باشد، داده‌ها به چپ منحرف یا انحراف‌به‌چپ دارند، به این معنی که دنباله چپ طولانی‌تر است [۱۳]. باتوجه‌به جدول (۳) موارد چولگی در ستون‌های ویژگی مساحت، کد رطوبت سوخت مناسب، شاخص گسترش اولیه و باران وجود دارد.

جدول (۳): مقدار چولگی و کشیدگی ویژگی‌های مجموعه‌داده

ویژگی	چولگی	کشیدگی
X	۰/۰۳۶۲	-۱/۱۷۲۳
Y	۰/۴۱۷۳	۱/۴۲۰۶
FFMC	-۶/۵۷۵۶	۶۷/۰۶۶۰
DMC	۰/۵۴۷۵	۰/۲۰۴۸
DC	-۱/۱۰۰۴	۰/۲۴۵۲
ISI	۲/۵۳۶۳	۲۱/۴۵۸۰
دما	-۰/۳۳۱۲	۰/۱۳۶۲
رطوبت نسبی	۰/۸۶۲۹	۰/۴۳۸۲
باد	۰/۵۷۱۰	۰/۰۵۴۳
باران	۱۹/۸۱۶۳	۴۲۱/۲۹۶۰
مساحت	۱۲/۸۴۶۹	۱۹۴/۱۴۰۷

برای کاهش چولگی و بهبود تقارن[2]، تابع لگاریتمی y=ln(x+1) که یک تبدیل رایج است و تمایل به بهبود نتایج رگرسیون برای اهداف راست چوله دارد، به ویژگی مساحت اعمال شده [۱۴] (شکل ۷). این تبدیل‌ها به ویژگی‌های دارای چولگی دیگر نیز اعمال می‌شوند تا چولگی و کشیدگی موجود در آن‌ها رفع شود.

[1] Kurtosis
[2] Symmetry



شکل (۷): نمودار هیستوگرام برای مساحت سوخته دارای چولگی (چپ)و تبدیل لگاریتم مربوط به آن (راست)[۵]

یکی از راه‌های بهبود عملکرد الگوریتم استفاده از انتخاب ویژگی (حذف متغیر) است. همچنین انتخاب ویژگی به درک داده‌ها و کاهش نیاز محاسباتی کمک می‌کند [۱۵]. مجموعه‌داده آتش‌سوزی جنگل دارای ۱۲ ویژگی مستقل است که به‌طورکلی به چهار گروه ویژگی‌های مکانی(X، Y)، ویژگی‌های زمانی(ماه، روز)، ویژگی‌های شاخص آب‌وهوای آتش‌سوزی (کد رطوبت سوخت مناسب، کد رطوبت لاشبرگ ، کد خشکسالی و شاخص گسترش اولیه) و ویژگی‌های هواشناسی(دما، رطوبت نسبی، باد، باران) تقسیم می‌شوند. ویژگی‌های شاخص آب‌وهوای آتش‌سوزی به دلیل این که از ویژگی‌های هواشناسی به وجود می‌آیند (شکل ۲) نسبت به آن‌ها دارای همبستگی می‌باشند.

جدول (۴): ویژگی‌های زیرمجموعه‌ها پس از انتخاب ویژگی

زیرمجموعه / ویژگی	LD[1]	F[2]	M[3]	LDF[4]	LDM[5]	LDFM[6]
X	✓	-	-	✓	✓	✓
Y	✓	-	-	✓	✓	✓
FFMC	-	✓	-	✓	-	✓
DMC	-	✓	-	✓	-	✓
DC	-	✓	-	✓	-	✓
ISI	-	✓	-	✓	-	✓
دما	-	-	✓	-	✓	✓
رطوبت نسبی	-	-	✓	-	✓	✓
باد	-	-	✓	-	✓	✓
باران	-	-	✓	-	✓	✓
مساحت	✓	✓	✓	✓	✓	✓

1: Location-Date 2:FWI 3:Meteorological 4:Location-Date-FWI 5:Location-Date- Meteorological 6:Full Dataset

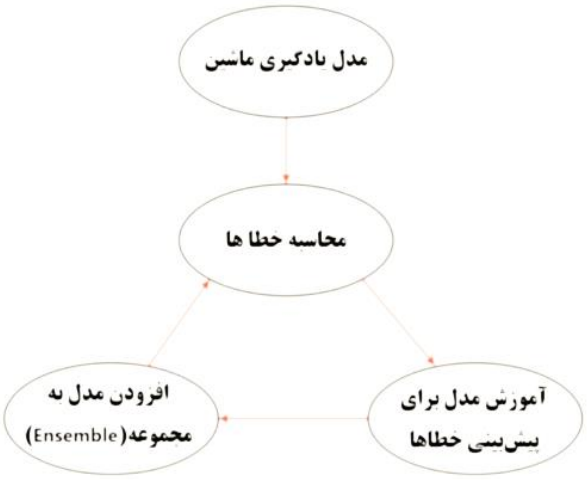
در جدول (۴) با انتخاب ویژگی، مجموعه‌داده به ۶ زیر مجموعه مجموعه مکانی- زمانی (LD)، مجموعه شاخص آب‌وهوای آتش‌سوزی (F) ، مجموعه هواشناسی (M)، مجموعه مکانی-زمانی-هواشناسی (LDM) ، مجموعه مکانی-زمانی- شاخص آب‌وهوای آتش‌سوزی (LDF) و خود مجموعه‌داده (LDFM) تقسیم شده است. شرایط مناسب برای برآورد مساحت حوادث آتش‌سوزی جنگل این است که فقط از مجموعه‌داده هواشناسی برای آموزش مدل استفاده شود چراکه ویژگی‌های هواشناسی را می‌توان بدون محاسبات انباشته[1] مستقیما از سنسورهای آب و هوا به دست آورد [۵].

۲-۴- آموزش مدل

پس از آماده‌سازی مجموعه‌داده نوبت به استفاده از الگوریتم‌ها برای یادگیری مدل می‌رسد. برای این کار ابتدا X(ویژگی‌ها) و y(برچسب) مشخص می‌شود. سپس مجموعه‌داده به دو زیر بخش آموزشی و آزمایشی تقسیم می‌شود به این صورت که ۶۷ درصد از داده‌ها به مجموعه آموزشی و ۳۳ درصد بقیه به داده‌های آزمایشی اختصاص می‌یابد. این تقسیم درصد تقسیم مجموعه‌های آموزشی و آزمایشی برای تمامی زیر مجموعه‌ها به همین صورت در نظرگرفته شده است.

باتوجه‌به این که ماهیت مساحت آتش‌سوزی جنگل به‌صورت پیوسته است؛ بنابراین برای حل این مسئله باید از روش‌های رگرسیون در یادگیری ماشین استفاده شود. در این مطالعه بیش از ۲۰ الگوریتم رگرسیون، از الگوریتم‌های پایه تا الگوریتم‌های ترکیبی با پارامترهای پیش‌فرض برای برآورد مساحت آتش‌سوزی بررسی می‌شود. در این مقاله بر روی الگوریتم‌های تقویت گرادیان[2] به خصوص ایکس‌جی‌بوست[3] تمرکز شده است و با انتخاب پارامترهای مناسب برای این الگوریتم سعی شده خطا در برآورد مساحت به کمترین مقدار خود برسد.

فرض کنید که مدل یادگیری ماشین ساده‌ای نظیر درخت تصمیم می‌تواند به آموزش مدل در مجموعه‌داده‌مان کمک کند. همان‌طور که گفته شد این مدل در امر پیش‌بینی استفاده می‌شود. برای بهبود نتایج ممکن است پارامترها را قدری تغییر دهیم یا امر داده‌افزایی را بررسی کنیم؛ اما در نهایت فقط از یک مدل استفاده می‌کنیم. حتی درصورتی‌که یک مجموعه از مدل‌ها بسازیم، همه مدل‌ها به‌صورت جداگانه آموزش‌داده‌شده و در داده‌هایمان به کار برده می‌شوند. از سوی دیگر کار که می‌توان برای بهبود نتایج انجام داد، به اجرا در آوردن عمل «تقویت[4]» در رویکردی تکراری است. مدل‌های بسیاری در این روش‌ها با یکدیگر ادغام می‌شوند، تا عملیات نهایی صورت گیرد، اما رویکرد هوشمندانه‌ای در پیش گرفته می‌شود. به‌جای آموزش کلیه مدل‌ها به‌صورت جدا از یکدیگر، عمل «تقویت» مدل‌ها را پشت‌سرهم آموزش می‌دهد. هر مدل جدید با این هدف آموزش داده می‌شود که خطاهای ناشی از مدل‌های پیشین تصحیح شوند. مدل‌ها تا زمانی به‌صورت متوالی افزوده می‌شوند که دیگر امکان پیشرفتِ بیشتر وجود نداشته باشد. مزیت این روش تکراری این است که مدل‌های اضافه‌شده در صدد تصحیح اشتباهاتی هستند که سایر مدل‌ها مرتکب شده‌اند. درصورتی‌که در روش کلاسه‌بندی جمعی استانداردی که مدل‌ها به‌صورت جداگانه آموزش داده می‌شوند، کلیه مدل‌ها ممکن است اشتباهات یکسانی را مرتکب شوند [۱۶].



شکل (۸): نحوه کار الگوریتم XGBoost[۱۷]

همان‌طور که گفته شد در این مقاله برای برآورد مساحت منطقه سوخته از روش تقویت گرادیان استفاده شده است. به‌طورکلی تقویت گرادیان به روشی اطلاق می‌شود که در آن، چند مدل یادگیری ماشین ساده مانند درخت تصمیم را پشت‌سرهم آموزش می‌دهد به‌طوری‌که مدل‌های جدید باهدف پیش‌بینی

[1] Accumulated calculations
[2] Gradient boosting
[3] XGBoost
[4] Boosting

باقی‌مانده‌های مدل‌های پیشین آموزش داده می‌شوند [۱۸]. جزئیات این روش در شکل (۸) نشان‌داده‌شده است. از چند الگوریتم روش تقویت گرادیان در این مطالعه استفاده شده است و بر روی الگوریتم رگرسیون ایکس‌جی‌بوست که از کتابخانه پایتونی ایکس‌جی‌بوست فراهم شده، تمرکز بیشتری انجام شده است.

۲-۵- ارزیابی مدل

برای ارزیابی مدل باید از معیارهای مناسب برای به‌دست‌آوردن دقت مدل‌های رگرسیونی استفاده شود. در این مقاله از دو معیار جذر میانگین مربعات خطاها و میانگین خطاهای مطلق برای ارزیابی دقت مدل استفاده شده است. در هر دو معیار، مقادیر کمتر منجر به مدل‌های پیش‌بینی بهتر می‌شود. بااین‌حال، جذر میانگین مربعات خطاها به خطاهای زیاد حساس‌تر است [۵]. فرمول ۱ و ۲ نحوه محاسبه دو معیار جذر میانگین مربعات خطاها و میانگین خطاهای مطلق را نشان می‌دهد:

$$MAD = \frac{1}{N}\sum_{i=1}^{N} y_i - \hat{y}_i \mid \quad (۱)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (۲)$$

نکته دیگر در ارزیابی مدل این است که مدل آموزش‌داده‌شده نباید بیش از حد به داده‌های آموزشی نگاشت شود. این کار باعث می‌شود عمل بیش برازش[1] رخ دهد و توانایی مدل در پیش‌بینی داده‌های جدید (آزمایشی) کاهش یابد [۱۹].

۳- نتایج و بحث

در این مقاله برای تمامی مراحل از سرویس COLAB شرکت گوگل با سخت‌افزار T4 GPU و 12.7G RAM همراه با Python3 استفاده شده است. جدول (۵) نتایج ارزیابی الگوریتم‌ها را توسط دو معیار جذر میانگین مربعات خطاها و میانگین خطاهای مطلق به‌صورت MAE(RMSE) نشان می‌دهد. بهترین نتایج در این جدول به‌صورت پررنگ نمایش داده شده است. نتایج حاصل از ارزیابی الگوریتم‌ها نشان می‌دهد که الگوریتم ایکس‌جی‌بوست باتوجه‌به معیارهای ارزیابی نسبت به الگوریتم‌های رگرسیونی دیگر عملکرد بهتری دارد. این الگوریتم برای تمامی زیر مجموعه‌ها، جذر میانگین مربعات خطاها و میانگین خطاهای مطلق کمتری را نتیجه می‌دهد. بهترین عملکردی که این الگوریتم ارائه می‌دهد برای زیر مجموعه مکانی-زمانی (LD) با میانگین خطاهای مطلق برابر با ۱۲/۲۵۷۱ و جذر میانگین مربعات خطاها برابر با ۶۱/۶۳۵۰ می‌باشد. الگوریتم‌های رگرسیون مقاوم[2]، ماشین بردار پشتیبان و جنگل تصادفی نیز در برخی زیر مجموعه‌ها عملکرد خوبی از خود نشان داده اند. به‌طورکلی کمترین مقدار برای معیار میانگین خطاهای مطلق توسط الگوریتم ماشین بردار پشتیبان بر روی مجموعه‌داده هواشناسی (M) به دست آمده که برابر با ۱۲/۱۹۰۱می‌باشد.همچنین کمترین مقدار برای معیار جذر میانگین مربعات خطاها توسط الگوریتم بگینگ[3] بر روی مجموعه‌داده شاخص آب‌وهوای آتش‌سوزی(F) به دست آمده که برابر با ۶۰/۵۷۹۱ می‌باشد.

همان‌طور که در بخش انتخاب ویژگی اشاره شد زیر مجموعه هواشناسی (M) به دلیل قابلیت تعمیم، نداشتن محاسبات انباشته و فراهم‌شدن مستقیم از سنسورها برای آموزش مدل مناسب‌تر است؛ بنابراین سعی شده با پیکربندی و تنظیم پارامترهای بهترین الگوریتم که همان الگوریتم ایکس‌جی‌بوست است، خطای برآورد کاهش یابد. به همین منظور شش ارزیابی برای انتخاب بهترین پارامترهای آموزش مدل با مقادیر مختلف پارامترها انجام شده است. تا بهترین نتیجه به‌عنوان پیکربندی الگوریتم برای یادگیری از مجموعه‌داده هواشناسی (M) انتخاب شود.

جدول (۶) نتایج ارزیابی پارامترها برای الگوریتم ایکس‌جی‌بوست را نشان می‌دهد. در این جدول از میان ارزیابی‌های صورت گرفته، ارزیابی ۴ جذر میانگین مربعات خطاها برابر با۶۱/۵۹۹۰ و میانگین خطاهای مطلق برابر با ۱۲/۲۷۵۸ به دست آورده است. بیشتر ارزیابی‌ها نیز جذر میانگین مربعات خطاها کمتر از ارزیابی ۴ به دست آورده اند اما در میانگین خطاهای مطلق نسبت به ارزیابی ۴ ضعیف‌تر عمل کرده‌اند. باتوجه‌به این که هرچه مقدار معیارهای جذر میانگین مربعات خطاها و میانگین خطاهای مطلق کمتر باشد مدل ایده آل تری ساخته شده است، پارامترهای ارزیابی ۴ به‌عنوان پارامترهای

[1] Overfitting

[2] Robust Regression
[3] Bagging Regression

آموزش مدل انتخاب شده است.

جدول (۵): نتایج حاصل از ارزیابی و مقایسه الگوریتم پیشنهادی(ایکس‌جی‌بوست) الگوریتم‌های رگرسیونی بر روی زیرمجموعه‌ها بر اساس معیارهای MAE (RMSE)

زیرمجموعه / الگوریتم	کل مجموعه‌داده (LDFM)	مجموعه مکانی - زمانی (LD)	مجموعه شاخص آب‌وهوای آتش‌سوزی (F)	مجموعه هواشناسی (M)	مجموعه مکانی - زمانی - شاخص هوای آتش‌سوزی (LDF)	مجموعه مکانی - زمانی - هواشناسی (LDM)
رگرسیون خطی	۱۲/۴۸۱۱(۶۱/۳۰۰۸)	۱۲/۴۷۸۵(۶۱/۴۷۰۴)	۱۲/۴۳۸۸(۶۱/۴۴۶۴)	۱۲/۴۶۳۵(۶۱/۴۳۲۴)	۱۲/۴۸۹۱(۶۱/۴۰۲۲)	۱۲/۴۶۶۵(۶۱/۳۹۰۶)
رگرسیون ریج[1]	۱۲/۴۸۶۵(۶۱/۳۰۳۹)	۱۲/۴۷۸۴(۶۱/۴۷۰۴)	۱۲/۴۴۹۲(۶۱/۴۴۴۴)	۱۲/۴۶۷۷(۶۱/۴۳۳۵)	۱۲/۴۹۴۴(۶۱/۴۰۰۸)	۱۲/۴۷۲۰(۶۱/۳۹۳۰)
رگرسیون لاسو[2]	۱۲/۴۵۴۵(۶۱/۴۵۳۲)	۱۲/۴۴۰۸(۶۱/۴۹۴۲)	۱۲/۴۵۴۵(۶۱/۴۵۳۲)	۱۲/۴۴۰۸(۶۱/۴۹۴۲)	۱۲/۴۵۴۵(۶۱/۴۵۳۲)	۱۲/۴۴۰۸(۶۱/۴۹۴۲)
شبکه الاستیک[3]	۱۲/۴۵۷۲(۶۱/۴۳۹۱)	۱۲/۴۴۰۸(۶۱/۴۹۴۲)	۱۲/۴۵۵۵(۶۱/۴۴۸۲)	۱۲/۴۴۱۴(۶۱/۴۹۰۳)	۱۲/۴۵۵۵(۶۱/۴۴۸۲)	۱۲/۴۴۱۴(۶۱/۴۹۰۳)
ماشین بردار پشتیبان	۱۲/۲۷۴۷(۶۱/۶۷۱۷)	۱۲/۲۹۷۳(۶۱/۷۵۱۱)	۱۲/۲۷۰۶(۶۱/۶۷۸۱)	**۱۲/۱۹۰۱**(۶۱/۷۳۹۱)	**۱۲/۲۷۳۸**(۶۱/۶۷۵۰)	**۱۲/۲۱۰۶**(۶۱/۷۲۳۷)
درخت تصمیم	۲۲/۶۵۱۰(۷۰/۴۶۱۸)	۲۰/۴۷۹۶(۶۸/۳۳۵۳)	۱۴/۵۱۹۱(۶۱/۳۰۰۳)	۲۲/۳۹۱۴(۷۱/۵۹۳۴)	۱۸/۰۴۱۸(۶۶/۷۵۸۹)	۲۲/۳۵۸۲(۷۲/۵۴۲۸)
جنگل تصادفی	۱۲/۶۸۷۴(۶۱/۳۷۵۵)	۱۴/۸۰۵۸(۶۲/۲۰۵۹)	۱۲/۷۹۵۷(۶۰/۷۷۸۵)	۱۲/۹۴۰۸(**۶۰/۵۹۱۵**)	۱۲/۸۹۵۴(۶۱/۲۷۰۰)	۱۲/۵۹۸۱(**۶۱/۱۱۳۹**)
درختان مازاد[4]	۱۲/۸۴۲۲(۶۱/۴۷۶۱)	۱۷/۶۰۴۱(۶۵/۳۵۴۱)	۱۳/۳۲۴۵(۶۰/۹۸۱۰)	۱۴/۳۶۸۲(۶۱/۹۶۳۴)	۱۲/۹۴۰۳(۶۱/۶۶۰۴)	۱۲/۸۰۷۴(۶۱/۵۱۵۳)
تقویت گرادیان	۱۲/۷۸۸۳(۶۱/۴۵۷۳)	۱۲/۸۶۸۳(۶۱/۵۰۸۹)	۱۲/۸۳۵۵(۶۱/۱۶۶۳)	۱۲/۹۱۲۳(۶۱/۱۹۹۸)	۱۲/۷۸۸۲(۶۱/۲۱۲۹)	۱۲/۴۴۲۲(۶۱/۲۵۷۸)
ایکس‌جی‌بوست (الگوریتم پیشنهادی)	**۱۲/۲۷۳۱(۶۱/۶۰۲۸)**	**۱۲/۲۵۷۱(۶۱/۶۳۵۰)**	**۱۲/۳۱۸۹(۶۱/۵۹۱۰)**	**۱۲/۲۸۰۸(۶۱/۶۲۳۲)**	**۱۲/۲۹۴۱(۶۱/۶۱۶۵)**	**۱۲/۲۷۴۵(۶۱/۶۳۵۲)**
آدابوست	۱۳/۴۴۸۷(۶۱/۳۵۶۰)	۱۳/۱۳۹۱(۶۱/۴۳۸۲)	۱۲/۸۳۵۵(۶۰/۵۱۰۶)	۱۳/۱۱۸۹(۶۱/۱۵۰۷)	۱۲/۹۹۱۶(**۶۱/۰۲۳۳**)	۱۳/۱۴۰۹(۶۱/۲۳۲۴)
k نزدیک‌ترین همسایه	۱۲/۷۳۹۰(۶۱/۴۸۹۶)	۱۳/۱۲۵۳(۶۱/۶۵۳۱)	۱۲/۵۷۸۶(۶۱/۱۷۰۱)	۱۲/۳۲۲۹(۵۶/۹۶۱۴)	۱۲/۶۷۴۷(۶۱/۲۱۴۵)	۱۲/۶۸۱۴(۶۱/۵۹۱۰)
گاوسی	**۱۲/۲۲۴۷**(۶۱/۸۷۱۶)	۱۸/۳۴۰۹(۶۵/۵۰۴۳)	۱۳/۱۸۸۱(۶۰/۹۸۲۶)	۳۵۸/۵۲۲۱(۴۴۶۵/۰۹۰۷)	۱۲/۴۴۰۴(۶۱/۸۲۱۴)	۱۲/۲۱۱۹(۶۱/۸۶۲۱)
بگینگ	۱۳/۳۲۸۳(**۶۱/۲۶۵۸**)	۱۵/۴۲۸۶(۶۲/۴۵۴۰)	۱۲/۶۵۴۷(**۶۰/۵۷۹۱**)	۱۳/۰۵۰۲(۶۱/۱۲۶۹)	۱۲/۷۱۶۷(۶۱/۴۴۸۹)	۱۲/۷۹۳۸(۶۱/۳۷۴۱)
رگرسیون‌مولفه‌های اصلی	۱۲/۵۰۷۹(۶۱/۳۱۴۷)	۱۲/۴۳۹۴(۶۱/۴۶۹۷)	۱۲/۴۵۰۰(۶۱/۴۴۴۲)	۱۲/۴۶۳۵(۶۱/۴۳۲۴)	۱۲/۵۰۰۴(۶۱/۳۹۸۱)	۱۲/۴۹۰۷(۶۱/۳۹۶۳)
بیزین	۱۲/۴۵۴۰(۶۱/۴۶۲۵)	۱۲/۴۴۰۵(۶۱/۴۹۳۶)	۱۲/۴۵۳۹(۶۱/۴۶۳۴)	۱۲/۴۴۱۳(۶۱/۴۸۹۶)	۱۲/۴۵۳۸(۶۱/۴۶۳۴)	۱۲/۴۴۱۲(۶۱/۴۸۹۶)
حداقل مربعات جزئی[5]	۱۲/۴۸۰۲(۶۱/۳۰۲۰)	۱۲/۴۷۸۲(۶۱/۴۷۰۰)	۱۲/۴۴۱۴(۶۱/۴۴۴۹)	۱۲/۴۶۳۵(۶۱/۴۳۲۴)	۱۲/۴۹۲۴(۶۱/۴۰۴۳)	۱۲/۴۶۷۳(۶۱/۳۹۰۸)
تایل - سین[6]	۱۲/۴۰۲۵(۶۱/۴۲۷۸)	۱۲/۲۷۷۹(۶۱/۶۲۶۳)	۱۲/۲۱۸۵(۶۱/۸۸۳۱)	۱۲/۲۳۲۵(۶۱/۷۱۶۳)	۱۲/۲۹۵۹(۶۱/۵۴۳۴)	۱۲/۳۵۹۱(۶۱/۵۲۷۵)
رگرسیون هوبر[7]	۱۲/۳۰۷۹(۶۱/۵۱۹۶)	۱۲/۳۱۱۸(۶۱/۶۰۵۷)	۱۲/۳۱۱۳(۶۱/۵۸۳۴)	۱۲/۲۸۲۶(۶۱/۵۹۰۱)	۱۲/۳۲۷۹(۶۱/۵۵۴۳)	۱۲/۲۸۷۵(۶۱/۵۷۰۳)
رگرسیون مقاوم	۱۲/۵۲۸۵(۶۱/۷۳۵۹)	**۱۲/۲۱۹۷**(۶۱/۸۸۱۰)	**۱۲/۲۱۷۳**(۶۱/۸۸۰۶)	۱۲/۲۳۴۶(۶۱/۸۷۰۸)	۲۱/۸۸۱۳(۶۸/۱۱۷۰)	۱۲/۳۰۸۰(۶۱/۸۵۵۷)
مدل‌جمعی‌تعمیم‌یافته[8]	۱۳/۳۲۶۸(۶۱/۵۳۶۵)	۱۲/۴۶۳۸(۶۱/۴۶۸۶)	۱۲/۸۳۲۴(۶۱/۴۷۷۰)	۱۲/۵۳۵۶(۶۱/۳۳۲۰)	۱۲/۹۶۶۹(۶۱/۴۳۴۰)	۱۲/۶۳۷۵(۶۱/۴۵۴۲)
رگرسیون پواسون[9]	۱۲/۴۴۰۸(۶۱/۴۹۴۲)	۱۲/۴۵۸۰(۶۱/۴۷۲۶)	۱۲/۴۴۰۸(۶۱/۴۹۴۲)	۱۲/۴۷۲۸(۶۱/۴۴۳۷)	۱۲/۴۴۰۸(۶۱/۴۹۴۲)	۱۲/۴۸۱۹(۶۱/۴۰۶۸)
مدل آمیخته گاوسی[10]	۱۲/۲۷۵۴(۶۱/۸۴۵۸)	۱۲/۸۶۲۲(**۶۱/۳۵۵۴**)	۱۲/۲۶۶۵(۶۱/۷۹۲۸)	۱۳/۴۸۹۹(۶۱/۵۵۴۳)	۱۲/۸۸۱۷(۶۱/۴۸۶۰)	۱۳/۴۸۹۸(۶۱/۴۷۹۷)

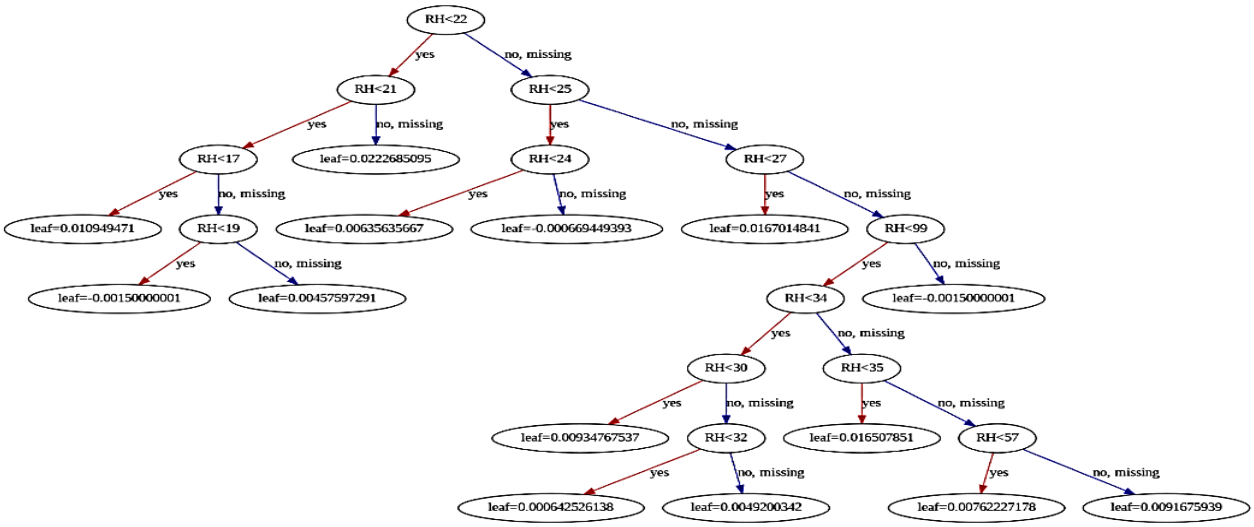
1: Ridge , 2: Lasso, 3: Elastic Net , 4: Extra Trees , 5: Partial Least Squares Regression, 6: Theil-Sen, 7: Huber, 8: Generalized additive model, 9: Poisson, 10: Gaussian Mixture Models

جدول (۶): نتایج حاصل از ارزیابی مدل منتخب ایکس‌جی‌بوست با زیر مجموعه هواشناسی(M)

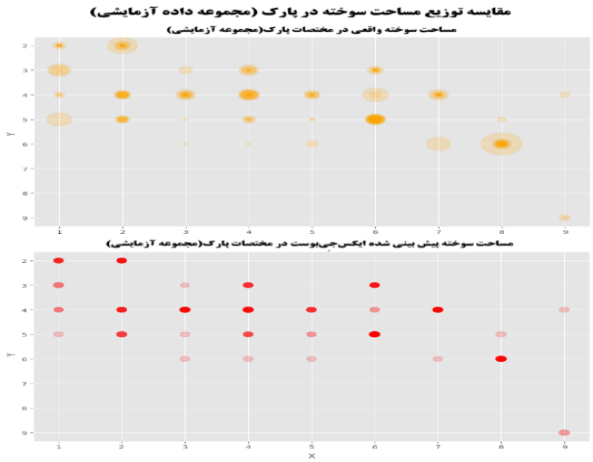
پارامتر	ارزیابی‌ها					
	۱	۲	۳	۴	۵	۶
نرخ یادگیری [۱]	۰/۲	۰/۱	۰/۰۱	**۰/۰۱**	۰/۰۱	۰/۰۱
عمق بیشینه [۲]	۱۰	۵	۳	**۷**	۳	۳
تعداد تخمینگر [۳]	۱۲	۵۲	۱۰۲	**۱۱۲**	۱۵۲	۲۰۲
MAE	۱۲/ ۴۴۹۶	۱۲/ ۸۰۴۵	۱۲/ ۲۸۰۸	**۱۲/ ۲۷۵۸**	۱۲/ ۳۴۴۰	۱۲/ ۴۱۴۶
RMSE	۶۱/ ۴۳۱۳	۶۱/ ۳۸۸۷	۶۱/ ۶۲۳۲	**۶۱/۵۹۹۰**	۶۱/ ۵۵۸۴	۶۱/ ۵۱۲۸

1: learning_rate, 2: max_depth, 3: n_estimators

در ارزیابی ۴ از داده‌های آموزشی و آزمایشی زیرمجموعه هواشناسی (M) و مدل مبتنی بر درخت تصمیم دودویی با نرخ یادگیری ۰,۰۱ ، حداکثر عمق ۷ و تعداد تکرار (مدل درخت تصمیم) ۱۱۲ همراه با معیار ارزیابی مکعب خطا استفاده شده است. بعد از تنظیم پارامترها آموزش مدل با استفاده از مجموعه‌داده‌های آموزشی انجام شده است و مدل با استفاده از داده‌های آموزشی درخت تصمیم مناسب با داده‌ها را ایجاد کرده است تا هنگام ارزیابی توسط داده‌های آزمایشی از آن برای برآورد ناحیه سوخته استفاده کند (شکل ۹).



شکل (۹): درخت تصمیم مدل آموزش دیده برای برآورد ناحیه آتش‌سوزی

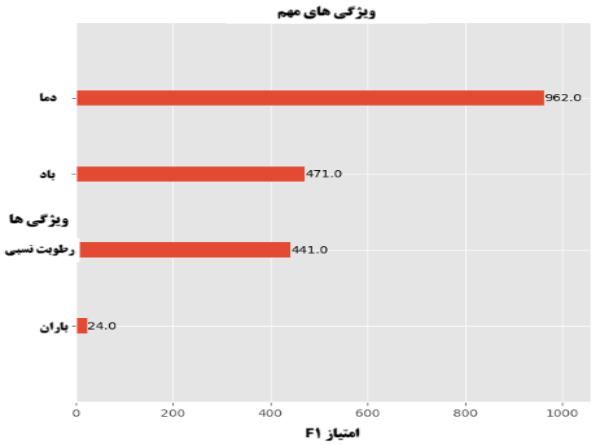


شکل (۱۰): نمودار مقایسه عملکرد ایکس‌جی‌بوست یعنی روش پیشنهادی در این مقاله(پایین) با واقعیت(بالا) در پیش‌بینی مساحت سوخته در مختصات پارک

عملکرد مدل ایکس‌جی‌بوست پیش‌بینی منطقه آتش‌سوزی و مساحت آن در مقایسه با داده‌های واقعی بر روی داده‌های آزمایشی در شکل (۱۰) نشان‌داده‌شده است. این مدل در شناسایی کلی مناطق مستعد آتش‌سوزی عملکرد خوبی دارد. همچنین، عملکرد قابل‌قبولی در تخمین شدت آتش‌سوزی‌ها ارائه می‌دهد. بااین‌حال، در جلوگیری از پیش‌بینی‌های کاذب (پیش‌بینی آتش در جایی که آتش واقعی کوچک است یا وجود ندارد) ضعف نشان می‌دهد.

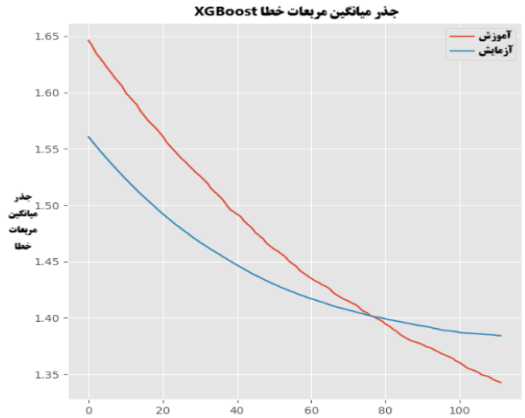
اگر با رویکرد یادگیری ماشین به اهمیت هر یک از ویژگی‌ها برای مدل ایکس‌جی‌بوست با پارامترهای ارزیابی ۴ توجه کنیم (شکل ۱۱). ویژگی‌های دما، باد، رطوبت نسبی و باران به ترتیب

بیشترین اهمیت را برای مدل بر اساس معیار امتیاز F1 کسب کرده‌اند. این ترتیب به این معنی است که مدل دما را تأثیرگذارترین عامل در برآورد مساحت حوادث آتش‌سوزی جنگل به‌دست آورده است.



شکل (۱۱): نمودار اهمیت ویژگی‌ها برای مدل

برای بررسی بیش برازش باید مدل از مجموعه‌داده‌های آموزشی نیز برای ارزیابی مدل استفاده شود. با این کار می‌توان تفاوت پیش‌بینی برای داده‌های آموزشی و آزمایشی را به دست آورد. نتایج مقایسه پیش‌بینی هر دو مجموعه‌داده با معیار جذر میانگین مربعات خطاها لگاریتمی نشان می‌دهد که این مدل عملکرد خوبی دارد چرا که اختلاف این پیش‌بینی‌ها اندک است [۱۹] (شکل ۱۲).



شکل (۱۲): نمودار مقایسه مقدار جذر میانگین مربعات خطاهای لگاریتمی برای داده‌های آموزشی (پررنگ) و آزمایشی (کم‌رنگ)

۴- نتیجه گیری

آتش‌سوزی جنگل‌ها به‌عنوان یک تهدید جدی برای امنیت ملی، نه‌تنها جان انسان‌ها را به خطر می‌اندازد، بلکه زیرساخت‌های حیاتی، خطوط انتقال انرژی و منابع طبیعی راهبردی کشور را با آسیب‌های جدی مواجه می‌سازد. این پژوهش مدلی هوشمند برای برآورد مساحت نواحی سوخته ارائه داده است که می‌تواند نقش مهمی در افزایش تاب‌آوری زیست‌محیطی و حفاظت از دارایی‌های ملی کشورها ایفا کند.

نتایج این پژوهش نشان می‌دهد که الگوریتم ایکس‌جی‌بوست بادقت قابل‌توجهی (جذر میانگین مربعات خطاها: ۶۱/۶۰۲، میانگین خطاهای مطلق: ۱۲/۲۷۳) قادر به برآورد مساحت حوادث آتش‌سوزی جنگل با استفاده از داده‌های هواشناسی است. این عملکرد نسبت به سایر پژوهش‌های انجام شده بر روی همین مجموعه‌داده، بهبود چشمگیری نشان می‌دهد. مزیت راهبردی چنین مدل‌هایی برای سامانه پدافند غیرعامل کشور در این است که به‌محض وقوع آتش‌سوزی، باتوجه‌به شرایط آب‌وهوایی می‌توان گستره احتمالی آتش را پیش‌بینی کرد و تصمیمات مناسبی برای تخصیص منابع و نیروهای عملیاتی اتخاذ نمود.

اگرچه این مطالعه بر اساس داده‌های پارک مونتسینهو در کشور پرتغال انجام شده است، مدل ارائه شده قابلیت بومی‌سازی و انطباق با شرایط اقلیمی و جغرافیایی مناطق مختلف ایران را دارد. برای ارتقای این سامانه در راستای پدافند غیرعامل کشور، پیشنهاد می‌شود:

- ایجاد پایگاه‌داده جامع از آتش‌سوزی‌های جنگلی ایران با ثبت دقیق پارامترهای هواشناسی و مشخصات آتش‌سوزی
- توسعه مدل‌های بومی با استفاده از داده‌های محلی و شرایط اقلیمی - جغرافیایی مناطق مختلف کشور

این پژوهش گامی در راستای بهره‌گیری از فناوری‌های نوین هوش مصنوعی در سامانه پدافند غیرعامل کشور محسوب می‌شود و زمینه را برای توسعه ابزارهای پیشرفته در مدیریت بحران فراهم می‌آورد. با بومی‌سازی این فناوری می‌توان تاب‌آوری زیرساخت‌های حیاتی کشور را در برابر تهدیدات طبیعی همچون آتش‌سوزی جنگل‌ها به میزان قابل‌توجهی افزایش داد.

۵- مراجع

[1]. P. Progias and G. C. Sirakoulis, "An FPGA processor for modelling wildfire spreading," Mathematical and Computer Modelling, vol. 57, no. 5, pp. 1436-1452, 2013/03/01/ 2013, doi: https://doi.org/10.1016/j.mcm.2012.12.005.

[2]. M. Ozbayoglu and R. Bozer, "Estimation of the Burned Area in Forest Fires Using Computational Intelligence Techniques," Procedia Computer Science, vol. 12, pp. 282–287, 12/31 2012, doi: 10.1016/j.procs.2012.09.070.

[3]. A. Azimpour, "Passive defense with a fire safety approach in the environment," presented at the 6rd International Conference

new ideas in Agriculture, Environment and Tourism, 2020. [Online]. Available: https://civilica.com/doc/1133040. (In Persian)

[4]. B. C. Arrue, A. Ollero, and J. R. M. d. Dios, "An intelligent system for false alarm reduction in infrared forest-fire detection," IEEE Intelligent Systems and their Applications, vol. 15, no. 3, pp. 64-73, 2000, doi: 10.1109/5254.846287.

[5]. P. Cortez and A. de J. R. Morais, “A data mining approach to predict forest fires using meteorological data,” Dec. 2007, Available: http://www3.dsi.uminho.pt/pcortez/fires.pdf

[6].T. Niranjan, D. Swetha, V. Charitha, and A. Stephen, “PREDICTING BURNED AREA OF FOREST FIRES,” IRJCS: International Research Journal of Computer Science, vol. 6, pp. 132-136, 2019. doi: 10.26562/IRJCS.2019.APCS10089.

[7]. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA, 2016. [Online].Available: https://doi.org/10.1145/2939672.2939785.

[8]. A. Alonso-Betanzos et al., “An intelligent system for forest fire risk prediction and fire fighting management in Galicia,” Expert Systems with Applications, vol. 25, no. 4, pp. 545–554, Nov. 2003, doi: 10.1016/s0957-4174(03)00095-2.

[9]. S. W. Taylor and M. Alexander, "Science, technology, and human factors in fire danger rating: the Canadian experience," International Journal of Wildland Fire, vol. 15, 03/28 2006, doi: 10.1071/WF05021.

[10]. A. Geron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media, Inc., 2019.

[11]. G. Montavon, W. Samek, and K.-R. Müller, "Methods for interpreting and understanding deep neural networks," Digital Signal Processing, vol. 73, pp. 1-15, 2018/02/01/ 2018, doi: https://doi.org/10.1016/j.dsp.2017.10.011.

[12]. D. N. Joanes and C. A. Gill, "Comparing Measures of Sample Skewness and Kurtosis," Journal of the Royal Statistical Society. Series D (The Statistician), vol. 47, no. 1, pp. 183-189, 1998. [Online]. Available: http://www.jstor.org/stable/2988433.

[13]. G. Hatem, J. Zeidan, M. Goossens, and C. Moreira, "Normality testing methods and the importance of skewness and kurtosis in statistical analysis," BAU Journal-Science and Technology, vol. 3, no. 2, p. 7, 2022.

[14]. S. Menard, Applied Logistic Regression Analysis, Thousand Oaks, California, 2002. [Online]. Available: https://methods.sagepub.com/book/applied-logistic-regression-analysis. Accessed on: 2024/01/30.

[15]. G. Chandrashekar and F. Sahin, "A survey on feature selection methods," Computers & Electrical Engineering, vol. 40, no. 1, pp. 16-28, 2014/01/01/ 2014, doi: https://doi.org/10.1016/j.compeleceng.2013.11.024.

[16]. J. H. Friedman, "Greedy function approximation: a gradient boosting machine," Annals of statistics, pp. 1189-1232, 2001.

[17]. D. Singh, A. H. Khan, and S. Meena, "Fake News Detection Using Ensemble Learning Models," in Proceedings of Data Analytics and Management, Singapore, A. Swaroop, Z. Polkowski, S. D. Correia, and B. Virdee, Eds., 2023// 2023: Springer Nature Singapore, pp. 53-66 .

[18]. I. D. Mienye and Y. Sun, "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects," IEEE Access, vol. 10, pp. 99129-99149, 2022, doi: 10.1109/ACCESS.2022.3207287.

[19]. B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. d. Freitas, "Taking the Human Out of the Loop: A Review of Bayesian Optimization," Proceedings of the IEEE, vol. 104, no. 1, pp. 148-175, 2016, doi: 10.1109/JPROC.2015.2494218.