Forest Fire Prediction Using Machine Learning

S Amrish[a] , V Sarankumar[b] , S Sri Kumaran[c] , C Gokila[d]

[a,b,c] *Student, Department of Electronics and Communication Engineering, Dr. Mahalingam College of Engineering and Technology, Pollachi-642003*

[d] *Assistant Professor, Department of Electronics and Communication Engineering, Dr. Mahalingam College of Engineering and Technology, Pollachi-642003*

Abstract: *A crucial element in fighting forest fires is forest fire prediction. This is a serious environmental issue that degrades the environment by endangering the natural resource landscape, upsetting ecosystem stability, raising the possibility of more natural disasters, and reducing the amount of resources like water that contribute to global warming and water pollution. An essential component of managing these kinds of situations is fire detection. It is anticipated that forest fire prediction would lessen the effects of future forest fires. There are numerous fire detection algorithms available, each taking a different approach to finding fires. Based on satellite photos, the fire-affected area is estimated in the current work procedures. The suggested approach uses meteorological characteristics including temperature, humidity, wind, and rain to predict the likelihood of a forest fire. We employed the RandomizedSearchCV algorithm to do hyperparameter tweaking and random forest regression. A number of decision trees are fitted to different subsamples of the dataset, and over-fitting is managed and prediction accuracy is increased by averaging. The investigation of the models that can simulate forest fire incidents using all of the chosen meteorological factors.This study compares several forest fire prediction models, including Decision Tree, Random Forest, Support Vector Machine.*

Keywords: **Forest fire, support vector machibe, classification, data processing, feature extracion.**

1. INTRODUCTION

Forest fires are a source for concern because they can seriously harm the surrounding area, personal belongings, and human life. Therefore, it's imperative that the fire at Associate in Nursing be discovered early. One of the main causes of the frequency of forest fires is global warming, which raises the average global temperature. On the other hand, human carelessness and lightning during thunderstorms are to blame. Every year, wildfires in the United States damage an average of 1.2 million acres of forest. The number of forest fires in the Asian country increased by a hundred and twenty-five percent between 2016 and 2018.These days, a variety of methods, including mathematical and physical models, are available for fireplace models that forecast the course of five flames. These models use a variety of facts from scientific lab experiments and forest fire simulations to describe and forecast the growth of fireplaces in various locations. Forest fires have been predicted by simulation tools more often recently; yet, these methods have encountered problems, such as computer file accuracy and tool execution time.

One possible subfield of artificial intelligence (AI) that deals with computers is machine learning.Two classes of machine learning exist: reinforcement learning and supervised, unsupervised learning. Regression, Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Decision Trees are examples of supervised machine learning methods. The information properties in the unattended learning don't appear to be tagged. This suggests that the labels should be outlined in the formula. The formula will discover the relationships between the options and the structure of the information set. • The primary goal of forest fire prediction is to support the Fire Management team's firefighters and allocate resources appropriately.

Weather conditions are one of the primary causes of fire. The nearest meteorological stations combine data from nearby sensors to obtain the climate. Land that may be at high danger of fire contains a number of markers that, when carefully examined, can be used to gauge the forecast. Each year, millions of hectares of land are destroyed by fire. Large tracts of land have been destroyed by these fires, and they produce more carbon monoxide than all traffic combined. Early fire detection and monitoring can significantly cut down on response times, possible damage, and firefighting expenses.

2. LITERATURE SURVEY

A. Forest Fire Prediction using Linear Regression

Mukhammad Wildan Alauddin et al. (2018) For forest fire prediction, multiple linear regression has been proposed. Temperature, humidity, wind, and rain are among the factors involved. Different techniques such as gauss-jordan, gaussseidel, and least-squares are used to calculate various linear regression coefficients. Comparative analysis of the methods is done and the results are discussed[2]. This Section discusses about the Literature survey

B. Forest Fire Prediction using Artificial Neural Network

Nizar HAMADEH LARIS EA et al. (2015), in this paper authors have considered an area called Lebanon to predict the occurrence of forest fire. Temperature, relative humidity, and wind speed are among the parameters. These parameters force Artificial Neural Networks to evolve in order to anticipate forest fires.

C. Forest Fire Prediction using Image Mining Technique

Divya T L et al. (2015) in their paper have presented the by analysing a series of pixel values, an image mining technique. can be used to predict the spread of a forest fire. The proposed model uses the satellite images for forest fire prediction.

D. Forest Fire Prediction using Artificial Intelligence

George E. Sakr et al. (2010), An approach to the study of forest fire prediction methods based on = intelligence has been suggested. Forest fire risk forecast algorithm is built on help vector machines. Lebanon data were used for the application of the algorithm and has proven the ability to correctly estimate the risk of fire.

E. Prediction of Forest Fire using Neural Network based on Extreme Learning Machines (ELM)

Mochammad Anshori(2019)This study focuses on preventing forest fires by predicting potential fire-prone areas based on meteorological conditions gathered from sensors. Key factors considered include temperature, wind speed, humidity, and rainfall. The method employed is a neural network trained with Extreme Learning Machines (ELM). Various tests are conducted to enhance the ELM method's performance, aiming for accurate predictions that can effectively mitigate fire spread before it escalates.

F. Mapping the risk of forest fires in Peru's Amazon and Andean Forest regions using the AdaBoost algorithm and Geographic Information Systems

Vincent Bax(2018) This study delves into the crucial aspect of forest fires as a significant driver of land-use change in tropical forest ecosystems. Focused on the Peruvian Amazon and tropical Andes regions, the research employs Kernel Density Analysis to pinpoint prevalent forest fire areas. Furthermore, it utilizes the AdaBoost algorithm to assess fire risk within two hotspot locations. The findings highlight the extensive nature of forest fires in Ucayali/Huánuco and San Martín departments, with distinct spatial patterns of susceptibility observed. The generated fire risk maps offer valuable insights for forest fire management strategies and research endeavors in these critical region's.

G. Extreme Learning Machine Approach for Prediction of Forest Fires using Topographical and Metrological Data of Vietnam

B.K. Singh(2019) Forest fires pose a significant threat to biodiversity, wildlife, and economic stability, making their prediction and mitigation crucial. This research employs Extreme Learning Machines (ELM) to predict forest fire occurrence in Vietnam using topographical and meteorological data. Factors such as slope, aspect, elevation, NDVI, and human proximity are considered. Historical data from 540 locations are utilized to establish relationships between fire-causing factors and occurrences. The study recommends the sigmoid activation function for accurate predictions, aiming to improve forest fire management strategies and minimize the impact of this calamity.

3. SOFTWARE DESCRIPTION

With pre-installed libraries like TensorFlow and PyTorch, Google Colab is a cloud-based platform for machine learning and data analysis that offers a collaborative environment. It provides smooth Google Drive integration, free access to GPU and TPU resources, and interactive tools for code execution, documentation, and visualization.

4. XG Boost Algorithm

Extreme Gradient Boosting Classifier, or XGBoost, is a popular and potent machine learning technique that excels in classification problems. It is built on the gradient boosting framework and falls under the group of ensemble learning. The XGBoost Classifier builds powerful predictive models by fusing together several weak learners, usually decision trees. In training, iteratively adding new trees that fix the mistakes of the old ones optimises an objective function. By using an iterative process, XGBoost can learn intricate patterns and predict class labels with a high degree of accuracy. The efficiency with which XGBoost handles big datasets is one of its main advantages. It is appropriate for big data workloads since it optimises computation speed and memory utilisation through the use of tree pruning and parallel processing techniques. Additionally, XGBoost provides a number of hyperparameters, including the learning rate, tree depth, and regularisation parameters, that can be adjusted to optimise the model's. XGBoost is a flexible option for a range of machine learning applications since it can handle regression, ranking, and recommendation tasks in addition to classification tasks. XGBoost Classifier is a well-liked option for data scientists and machine learning practitioners in several areas due to its resilience, scalability, and capacity to produce high-quality predictions. For a number of important reasons, the XGBoost (Extreme Gradient Boosting) classifier is considered by many to be among the top machine learning prediction algorithms. First off, XGBoost is an ensemble learning technique that builds a powerful and reliable predictive model by aggregating the predictions of several weak learners, usually decision trees. With the help of an ensemble method, XGBoost is able to identify intricate patterns and relationships in the data, which improves prediction task accuracy. Second, XGBoost can handle big datasets with hundreds or even millions of instances and features because of its excellent scalability and efficiency. Compared to conventional gradient boosting techniques, it uses less memory and has shorter training times thanks to the use of parallel processing and optimised tree construction algorithms. XGBoost also has the benefit of being able to efficiently tackle jobs involving both regression and classification. The model's predictive performance on unobserved data is improved by using regularisation techniques like L1 and L2 regularisation, which help to increase generalisation and prevent overfitting. Additionally, XGBoost provides a large variety of hyperparameters, such as regularisation parameters, tree depth, and learning rate, that may be adjusted to maximise model performance. Because of this flexibility, data scientists can experiment and tailor the algorithm to the unique properties of particular datasets in order to maximise prediction accuracy.

5. Data Set Information

The collection includes 244 examples from the northeastern district of Bejaia and the northwest region of Sidi Bel-abbes in Algeria. There are 122 incidents from each region, spanning the months of June through September of 2012. Eleven attributes and one output attribute (class) are present in this dataset. Two classes of cases are identified: fire (138 instances) and not fire (106 instances). Meteorological parameters including temperature, wind speed, humidity, and rainfall, as well as topographical characteristics like aspect, slope, and elevation, are important characteristics. These characteristics are essential for forecasting the likelihood of forest fires, which helps with the creation of efficient management plans and mitigation techniques for the areas being researched.

5.1 Attribute Information

The months of June through September of 2012 were the main emphasis of the weather data observations, which were made on a particular date and entered into the format DD/MM/YYYY. The temperature (Temp), which ranges from 22 to 42 degrees Celsius at noon, the relative humidity (RH), which is expressed in percentages from 21 to 90%, the wind speed (Ws), which is measured in kilometers per hour and ranges from 6 to 29 km/h, and the rainfall (Rain), which is measured in millimeters and has values between 0 and 16.8 mm, are among the attributes that are analyzed. These characteristics, which highlight fluctuations in temperature, humidity, wind speed, and precipitation levels, offer important insights into the meteorological conditions over the designated period. Such data is significant for meteorological studies, agricultural planning, and climate research since it is essential for comprehending weather patterns, anticipating trends, and evaluating environmental repercussions. A number of essential components make up the Fire Weather Index (FWI) system, which is used to evaluate possible wildfire behavior and fire danger. The Initial Spread Index (ISI), Buildup Index (BUI), Drought Code (DC), Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), and the overall Fire Weather Index are some of these components. The FFMC provides a range of readings from 28.6 to 92.5 that indicate the moisture content of fine fuels like grass and leaves. The DMC, which ranges from 1.1 to 65.9, evaluates moisture content in decomposed organic material. Deep-layer soil moisture conditions, which vary from 7 to 220.4, are assessed by the DC. The rate of fire propagation is predicted by the ISI, which ranges from 0 to 18.5. The BUI, which ranges from 1.1 to 68, shows the total amount of fuel that is accessible for combustion. Finally, these elements are combined by the FWI Index to produce an overall fire risk estimate that ranges from 0 to 31.1. These indices support attempts to manage and prevent wildfires by helping to categorise fire danger into two classes: fire and non fire. The terms "fire danger assessment," "wildfire behaviour prediction," "moisture levels," "fuel availability," and "fire risk classification" are synonymous with these FWI components.

5.2 Data collection and understanding

The UCI repository is the source of the Algerian forest fire dataset used in this project. It specifically collects data relating to forest fires and observations from the Bejaia and Sidi Bel-Abbes districts of Algeria. The dataset records important meteorological phenomena between June 2012 and September 2012. The main goal of the study is to use different classification algorithms to find important weather parameters that can accurately predict forest fires in these areas. The project's keywords are Bejaia region, Sidi Bel-Abbes region, classification algorithms, weather features, Algerian forest fires, and predictive modelling.

5.3 Data Preprocessing

In this stage, we will use Pandas for data analysis and Matplotlib and Seaborn for data visualization to undertake Exploratory Data Analysis (EDA) to extract insights from the dataset and determine which features have contributed more to the prediction of forest fires. Understanding the data first and attempting to extract as many insights as possible from it is always a smart idea. **Importing filter warnings to ignore warning messages** and **Import Required Library.**

5.4 Exploratory Data Analysis

An essential first stage in data analysis is exploratory data analysis (EDA), which entails looking at and visualising data to comprehend its properties, trends, and connections. In the dataset, EDA seeks to reveal patterns or trends, find abnormalities, and unearth insights. To get a thorough grasp of the structure and underlying patterns of the data, it usually includes correlation analysis, summary statistics, and data visualisation tools like histograms, scatter plots, and box plots.

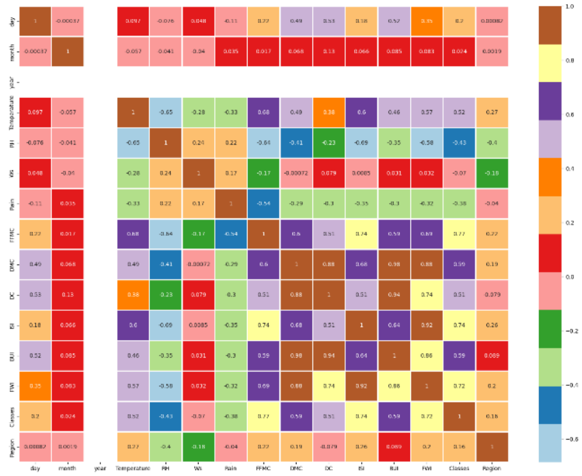
EDA is essential for providing guidance for later data modelling and decision-making procedures, which in turn helps to guarantee the precision and dependability of analytical outcomes.

5.5 Creating a copy of original dataset

For a number of reasons, making a duplicate of the original dataset is crucial to machine learning. First of all, it ensures that any alterations or preprocessing operations don't impact the original dataset, protecting the data's originality and integrity. Because researchers and practitioners may always refer back to the undisturbed dataset if needed, this is essential for reproducibility and traceability.

5.6 Graph and correlation

To run regression analysis, I must convert all of the features from object datatype to integer**.** The dataset has no duplicate data, and since there are no longer any missing data, we may go on to the analysis stage. A correlation matrix was shown, allowing us to assess the dependence between two variables and determine how they move together. It also displays columns with more than one individual value and checking counts to add classes column in matrix.



We can read the data now, and we can plot the data to see the data visually.

• A distribution graph (bar graph or histogram) was created for the column data that showed the distinct values

.

5.7 Selection of Algorithm

For machine learning applications, XGBoost is the best option for predicting forest fires because of a number of important criteria. First of all, because of its capacity to manage deep, nonlinear relationships within data, it is an excellent tool for capturing the complex interactions between different climatic parameters like temperature, humidity, wind speed, and vegetation density that all play a role in forest fires. Furthermore, XGBoost's ensemble learning method improves its predictive accuracy and robustness by combining the predictions of several weak learners—in this case, decision trees. This is especially important for predicting forest fires because many factors and their interactions might affect the results. Fast training and prediction are also made possible by XGBoost's effective implementation, which is crucial for real-time or near-real-time applications where prompt identification and containment of possible fire outbreaks are critical. Because of its adaptability to handle regression and classification problems, it can be used to support proactive fire management techniques by predicting not just the occurrence of fires but also their severity or size. Overall, XGBoost is a great option for forest fire prediction in machine learning projects due to its combination of flexibility, accuracy, speed, and scalability.

5.8 Classifier Model

The Classification algorithm, a Supervised Learning technique, is used to classify new observations based on the training data. In classification, software learns how to classify new observations into different classes or groups by utilizing the dataset or provided observations. Labeled training data is provided to the classifier, and the classifier is used to train the model until the model is able to classify the data. The classifier may accept unlabeled datasets and begin classifying output once it has received sufficient training. Classifier algorithms utilize sophisticated mathematical and statistical techniques to estimate the likelihood that a given data input will be classed in a specific way.

6. Results and Discussions

In this machine learning project that makes use of the XGBoost method, analyzing the results thoroughly is essential to comprehending the model's functionality and deriving significant insights. Start by assessing the accuracy metrics, which offer information about the model's capacity to accurately categorize fire occurrence events. These measures include precision, recall, F1-score, and accuracy itself. Furthermore, think about using assessment methods such as confusion matrices to see how well the model performs in various classes and see any possible biases or areas for development. Additionally, evaluate the feature importance scores produced by XGBoost to determine which environmental factors are most important in predicting forest fires. By educating stakeholders about the main causes of fires, this analysis can help direct future resource allocation and mitigation initiatives. Additionally, think about performing sensitivity assessments to evaluate how resilient the model is to changes in input data or parameter settings, guaranteeing its dependability under a range of environmental circumstances. Finally, to add context and demonstrate the superiority of your XGBoost model, compare its performance with other machine learning algorithms or conventional statistical techniques. You can further scientific knowledge and useful management practices in areas vulnerable to wildfires by carrying out a thorough result analysis and provide insightful feedback on the effectiveness and applicability of the XGBoost algorithm for forest fire prediction.

7. Conclusion

In order to have a variable number of training instances set and evaluation instances set for forest fire prediction, experiments are concluded. In this project, the factors contributing to the frequency of fires are examined. Temperature, relative humidity, and wind speed are considered meteorological factors. High wind speeds, moderate humidity, and extreme temperatures all greatly increase the risk of burning. Additionally, it is discovered that there are more fires in forests than in other surface regions. Data mining methods need to be employed for fire prediction since the risk of forest fires in the forest rises dramatically. This project can be developed further to provide better results, including greater effects and better-equipped models. We might also have a user interface designed to offer some real-time functionality for the application.
The user may input their zip code and local address in the UI model's process. We will use the zip code to obtain latitude and longitude using any API, ingest the coordinates as arguments, and obtain the weather information for that day, including the highest and lowest temperatures, humidity levels, wind speeds, and other factors.

8. Reference

1. *A. Alonso-Betanzos, O. Fontenla-Romero, B. Guijarro-Berdi nas, E. Hern´andez-Pereira, M. Inmaculada Paz Andrade, E. Jim´enez, J. LuisLegido Soto, and T. Carballas, "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, 2003.*

2. *N. Aronszajn, Introduction to the theory of Hilbert spaces. Stillwater, Oklahoma:Reasearch [sic] Foundation, 1950.*

3*.T. Cheng and J. Wang, "Applications of spatio-temporal data mining and knowledge for forest fire," in ISPRS Technical Commission VII Mid Term Symposium, Enschede, 2006, pp. 148–153.*

4. *"Integrated Spatio-temporal Data Mining for Forest Fire Prediction," Transactions in GIS, vol. 12, no. 5, pp. 591–611, 2008*

5. *K. Clarke, J. Brass, and P. Riggan, "A cellular automaton model of wildfire propagation and extinction," Photogrammetric Engineering and Remote Sensing, vol. 60, no. 11, pp. 1355–1367, 1994.*

6. *Z. Li, Y. Kaufman, C. Ithoku, R. Fraser, A. Trishchenko, L. Giglio, J. Jin, and X. Yu, "A review of AVHRR-based active fire detection algorithms: Principles, limitations, and recommendations," Global and Regional Vegetation Fire Monitoring from Space.*

7. *J. Han, K. Ryu, K. Chi, and Y. Yeon, "Statistics Based Predictive Geo-spatial Data Mining: Forest Fire Hazardous Area Mapping Application," Lecture notes in computer science, pp. 370–381, 2003.*

8. *R. Jaiswal, M. Saumitra, D. Kumaran, and S. Rajesh, "Forest fire risk zone mapping from satellite imagery and gis," International Journal of Applied Earth Observation and Geo-information, vol. 4, pp. 1–10,2002.*

9. *G. Mitri and I. Gitas, "A semi-automated object-oriented model for burned area mapping in the Mediterranean region using Landsat-TM imagery," International Journal of Wildland Fire, vol. 13, no. 3, pp. 367–376, 2004.*

10. *A. Muzy, T. Marcelli, A. Aiello, P. Santoni, J. Santucci, and J. Balbi, "An object oriented environment applied to a semi-physical model of fire spread across a fuel bed," in Actesde la conf´erence ESS 2001 conference, 2001, pp. 641–643.*