ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Impact Factor: 8.206

Volume 8, Issue 4, April 2025

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Comparison of Forest Fire Prediction Systems Using Machine Learning Algorithms

Balaji G, Dr. R Suganthi

III B.Sc., Department of Computer Science with Data Analytics, Dr. N.G.P Arts and Science College, Coimbatore, India

Professor, Department of Computer Science with Data Analytics, Dr. N.G.P Arts and Science College, Coimbatore, India

ABSTRACT: Forest fires are one of the most devastating natural disasters, causing significant damage to ecosystems, wildlife, and human life. Early prediction of forest fires can mitigate these effects by enabling timely intervention. This study aims to compare the effectiveness of various machine learning algorithms in predicting forest fires based on environmental and meteorological data. The project evaluates algorithms such as Decision Trees, Random Forest, Support Vector Machines, k-Nearest Neighbors, and Neural Networks. These models are trained on datasets containing key features like temperature, humidity, wind speed, and rainfall, as well as vegetation indices. Each algorithm is assessed for accuracy, precision, recall, F1-score, and computational efficiency to determine its suitability for real-time prediction systems. The results reveal the strengths and limitations of each model, providing insights into the most reliable and efficient approach for forest fire prediction. This comparison not only highlights the role of advanced machine learning techniques in environmental monitoring but also underscores the importance of data-driven decision-making in forest management. The findings of this study contribute to developing robust forest fire prediction systems, ensuring better preparedness and response strategies.

KEYWORDS: Geospatial Analysis, Satellite Imagery, GIS Mapping, Climate Data, Weather Forecasting.

I. INTRODUCTION

Forest fires are among the most destructive natural disasters, causing extensive harm to ecosystems, human settlements, and the global climate. These fires often lead to irreversible loss of biodiversity, displacement of communities, and increased greenhouse gas emissions, exacerbating climate change. Despite their devastating impacts, forest fires remain unpredictable due to the complex interplay of environmental factors such as temperature, humidity, wind, and vegetation. Developing a reliable system for early detection and prediction is crucial for mitigating these effects and ensuring effective management.The integration of machine learning (ML) technologies has emerged as a promising approach to address this challenge. By analyzing historical and real-time data, ML algorithms can identify patterns and trends that are difficult to discern through traditional methods. This project aims to evaluate the performance of various ML models in predicting forest fires, focusing on their accuracy, precision, recall, and F1 score. Using a publicly available dataset, the study seeks to determine the most effective predictive approach, providing actionable insights for early warning systems.By leveraging advanced predictive analytics, this project contributes to proactive forest fire management, enabling timelyintervention minimize environmental, economic, and societal losses while promoting sustainable environmental practices.

II. OBJECTIVE

The primary aim of this project is to evaluate and compare the performance of various machine learning (ML) algorithms in accurately predicting forest fires. Forest fires are complex phenomena influenced by numerous environmental factors such as temperature, wind speed, humidity, and vegetation type. Leveraging machine learning models can provide deeper insights into the patterns and conditions that lead to fire outbreaks, enabling proactive prevention and control measures. This study focuses on analyzing the effectiveness of different ML algorithms by assessing their predictive capabilities using metrics such as accuracy, precision, recall, and F1 score. By utilizing a publicly available dataset, the research seeks to identify the most reliable and efficient model for forest fire prediction. The comparison also highlights the strengths and limitations of each algorithm, offering valuable insights into their suitability for real-world applications. Ultimately,

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

this project aims to advance the development of early warning systems by determining the most effective machine learning techniques for forest fire prediction. These findings will help improve disaster preparedness and response, contributing to the protection of biodiversity, human settlements, and the environment. By refining predictive models, this research supports the broader goal of mitigating the ecological and economic impacts of forest fires.

III. DATASET

Source: UCI Machine Learning Repository, NASA Fire Information for Resource Management System (FIRMS), or other open-source repositories.

UCI Machine Learning Repository**:** Forest Fires Dataset: This dataset contains meteorological and other data to predict the burned area of forest fires in Portugal.

Mendeley Data: Forest Fire Dataset: This dataset comprises images related to forest fires, intended for training algorithms designed for forest fire detection and object detection tasks
Next Day Wildfire Spread Dataset: Aggregating nearly a decade of remote-sensing data across the United States, this dataset is curated for predicting wildfire spread using machine learning

IV. FEATURES

Meteorological data: Temperature, humidity, wind speed, and rainfall. Geographic data: Latitude, longitude, and elevation. Vegetation data: Forest type and density. Fire-related data: Historical fire records, burned area, and ignition points.

UCI Forest Fires Dataset (Portugal)
X, Y: Spatial coordinates within the park
Month, Day**:** Temporal attributes (categorical)
FFMC (Fine Fuel Moisture Code)**:** Moisture content of litter and small fuels
DMC (Duff Moisture Code)**:** Moisture content in loosely compacted organic layers
DC (Drought Code): Moisture content in deep compacted organic layers
ISI (Initial Spread Index): Fire behavior index based on wind and fuel dryness
Temperature (°C)
Relative Humidity (%)
Speed (km/h)
Rain (mm) Wind
Burned Area (ha): Output variable (Target)

V. PREPROCESSING

To ensure data quality and improve model performance, the following preprocessing steps will be applied. Missing values will be handled using mean or mode imputation depending on the nature of the data. For numerical features, missing entries will be replaced with the mean value of the respective column, while for categorical features, the mode (most frequent value) will be used. This approach ensures that the dataset remains complete without introducing bias. Numerical data will be normalized to enhance algorithm efficiency and ensure uniform scaling across features. Normalization adjusts the values to a standard range, which helps machine learning algorithms converge faster and perform more effectively. For categorical variables, encoding techniques such as one-hot encoding or label encoding will be employed. One-hot encoding creates binary columns for each category, while label encoding assigns a unique numerical value to each category. These transformations allow the algorithms to process categorical data effectively.

VI. ALGORITHMS

1.Linear Regression (Baseline)
Linear Regression is utilized as a baseline model to predict either the burned area caused by forest fires or the probability of fire occurrence. It is a fundamental supervised learning algorithm that establishes a linear relationship

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

between input features and the target variable. By minimizing the error between predicted and actual values, Linear Regression provides a straightforward method for prediction. One of the key advantages of this model is its simplicity and interpretability. Linear Regression has low computational complexity, making it efficient to train and implement, especially on smaller datasets. The coefficients of the model provide insights into the relative importance of input features, helping identify factors that most influence forest fire behaviour. While Linear Regression serves as a useful starting point, its performance may be limited in capturing the nonlinear relationships often present in forest fire data, making it a benchmark for comparing more complex algorithms. Limitation: Cannot model non-linear relationships effectively.
Forest fire prediction is a crucial task for disaster management, requiring accurate forecasting methods to minimize environmental and economic damage. Machine learning algorithms such as Random Forest and deep learning models like Long Short-Term Memory (LSTM) networks play a significant role in predicting fire occurrences based on meteorological, geographical, and environmental factors. Random Forest classifiers analyze historical fire data and identify critical features such as temperature, humidity, wind speed, and vegetation index

1. Logistic Regression

Logistic Regression is a widely used machine learning algorithm designed for binary classification tasks, such as predicting the occurrence of a forest fire (Fire) versus no fire (No Fire). Unlike Linear Regression, which predicts continuous outcomes, Logistic Regression models the probability of an event occurring using a logistic or sigmoid function, ensuring the output lies between 0 and 1. This algorithm is particularly suited for simple prediction tasks where the relationship between the input features and the target variable is mostly linear. Logistic Regression is computationally efficient, easy to implement, and interpretable. Its coefficients indicate the strength and direction of the relationship between input variables and the likelihood of fire occurrence, making it a useful tool for understanding contributing factors. Despite its strengths, Logistic Regression may struggle to handle complex, nonlinear relationships in forest fire datasets. However, it remains an effective baseline for evaluating more advanced classification models.Logistic regression is a widely used machine learning algorithm for binary classification tasks, making it suitable for forest fire prediction. This method estimates the probability of fire occurrence based on environmental factors such as temperature, humidity, wind speed, and vegetation index. By applying a sigmoid function, logistic regression maps input features to a probability score between 0 and 1, classifying areas as fire-prone or not. It is computationally efficient, interpretable, and effective for datasets where relationships between variables are approximately linear. Despite its simplicity, logistic regression can be enhanced with feature selection and regularization techniques, making it a valuable tool for early wildfire detection and prevention strategies.

Decision Trees are powerful machine learning models capable of capturing non-linear relationships between input features and the target variable, making them well-suited for forest fire prediction tasks. Unlike linear models, Decision Trees split the data into subsets based on feature values, creating a tree-like structure of decision rules that are easy to interpret. Each decision node represents a feature, while the branches correspond to possible values, and the leaf nodes provide the prediction outcome. The ability to handle non-linear patterns makes Decision Trees effective for datasets where the relationships between variables are complex. They can also accommodate both numerical and categorical data, making them versatile. Additionally, Decision Trees are non-parametric, meaning they do not assume a specific distribution of the data, which is beneficial for real-world forest fire datasets that often exhibit diverse patterns. However, Decision Trees are prone to overfitting, especially when the tree becomes too deep. Regularization techniques, such as pruning, are typically employed to improve their generalization capabilities. Limitation: Overfitting on training data.

Random Forests is an ensemble learning technique that builds multiple decision trees and combines their outputs to produce more accurate and reliable predictions. This method helps address the key limitation of individual decision trees: overfitting. By aggregating predictions through averaging (for regression) or majority voting (for classification), Random Forests achieve improved generalization and reduce the risk of overfitting. One of the strengths of Random Forests is their robustness to noise and missing data. Since the model relies on a collection of trees rather than a single one, the impact of noisy or incomplete data in individual trees is minimized. Additionally, Random Forests utilize random feature selection at each split, ensuring diversity among the trees, which further enhances their performance and reduces bias. Random Forests can effectively capture complex, non-linear relationships in data, making them highly suitable for forest fire prediction. Despite their complexity, they are relatively easy to implement and interpret, providing feature importance metrics that help identify key predictors of forest fires. Support.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Vector Machines (SVM)

Support Vector Machines (SVM) are robust machine learning algorithms particularly effective for handling high-dimensional datasets. SVM works by finding the optimal hyperplane that separates data points of different classes with the maximum margin, ensuring reliable classification or regression outcomes. This characteristic makes it a strong candidate for forest fire prediction tasks, especially when the dataset has numerous features. One of the standout features of SVM is the "kernel trick," which allows it to model complex, non-linear relationships. By transforming the input data into a higher- dimensional space using kernel functions (such as radial basis function or polynomial kernels), SVM can effectively separate data that is not linearly separable in the original feature space. Despite its effectiveness, SVM can be computationally intensive for large datasets and may require careful parameter tuning to achieve optimal performance. However, its ability to handle complex patterns makes it a valuable tool in predicting forest fire occurrence and behaviour.

K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) is a simple and intuitive machine learning algorithm used for both classification and regression tasks. In KNN, the prediction for a new data point is made based on the majority class (for classification) or the average value (for regression) of its 'K' nearest neighbors in the feature space. The algorithm relies on the concept of distance, typically using Euclidean distance, to find the closest data points. One of the key strengths of KNN is its simplicity; it doesn't require complex training processes or assumptions about the underlying data distribution. This makes KNN easy to implement and understand. Additionally, KNN is a non-parametric method, meaning it doesn't make assumptions about the form of the model, which is beneficial when working with diverse and unstructured data, such as that found in forest fire prediction. However, KNN can be computationally expensive, especially with large datasets, as it requires calculating distances to all data points during prediction. Choosing the optimal value for 'K' is also crucial for performance. Limitation: Sensitive to noise and large datasets.

Gradient Boosting (e.g., XGBoost, LightGBM)

Gradient Boosting is an advanced ensemble learning technique that builds a strong predictive model by combining the outputs of multiple weak models, typically decision trees, in a sequential manner. Each tree corrects the errors made by the previous one, focusing on the residuals (errors) from prior iterations. This iterative process results in a highly accurate and robust model. XGBoost and LightGBM are popular implementations of Gradient Boosting, known for their efficiency and high performance. XGBoost is particularly recognized for its scalability and speed, while LightGBM is optimized for large datasets and can handle categorical features natively. Both algorithms are capable of achieving high predictive accuracy by capturing complex patterns in the data, making them highly effective for tasks like forest fire prediction. Although Gradient Boosting methods require careful tuning of hyperparameters to avoid overfitting, their ability to achieve superior accuracy and handle large, intricate datasets makes them a preferred choice in many machine learning application

Evaluate strengths and weaknesses of each algorithm. Identify trade-offs between accuracy and interpretability.

VII. FINDINGS

Random Forests and Gradient Boosting models often outperform others in terms of accuracy and robustness. Neural Networks perform well with large datasets but require significant computational resources. Logistic Regression and Decision Trees are suitable for small datasets and easy interpretation.

Significance of Meteorological Factors – Logistic regression analysis confirms that temperature, humidity, wind speed, and rainfall are key determinants in predicting forest fires. High temperatures and low humidity significantly increase fire risk.

Predictive Accuracy – While logistic regression provides a simple and interpretable model, its accuracy is moderate compared to complex machine learning models. It performs well for linearly separable data but struggles with non-linear relationships in environmental factors. Feature Importance – Among the variables analyzed, temperature and wind speed emerge as the most influential factors in fire prediction, while rainfall plays a mitigating role in reducing fire probability.

Limitations in Complex Scenarios – Logistic regression is less effective when dealing with highly complex or spatially distributed fire occurrences, requiring additional transformations or advanced models like Random Forest or Deep Learning for improved performance.

Early Warning Potential – Despite its limitations, logistic regression can serve as a fast and computationally efficient tool for early fire risk assessment, aiding disaster management teams in proactive decision-making and resource allocation.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

VIII. CHALLENGES

1.Data Quality: The accuracy and reliability of predictions are heavily influenced by the quality of the input data. Missing, incomplete, or inaccurate records can severely impact the model's performance, leading to biased or unreliable predictions. Forest fire datasets may have gaps due to missing observations, errors during data collection, or poor sensor calibration, which can distort the relationships between features and the target variable. Additionally, outliers or noisy data can mislead models into learning incorrect patterns. Ensuring high-quality, clean, and complete data is essential for building effective predictive models and improving their accuracy.

2.Real-time Application**:** While advanced machine learning models, particularly deep learning algorithms, offer powerful predictive capabilities, they often require substantial computational resources, including high processing power and large amounts of memory. This can be a significant challenge when

3.deploying models in real-time applications, such as early warning systems for forest fires. In such systems, predictions need to be made quickly and continuously, and high-latency models may not be suitable. Streamlining models for efficiency, reducing their complexity, or optimizing their computational requirements is crucial for enabling real-time decision-making and ensuring timely interventions.

IX. FUTURE SCOPE

Integrate Real-Time Satellite Data One promising direction for improving forest fire prediction models is the integration of real-time satellite data. Satellites provide continuous, up-to- date information on environmental conditions such as temperature, humidity, wind patterns, vegetation health, and soil moisture, all of which can influence the likelihood of a fire. By incorporating real-time satellite imagery and data feeds into predictive models, it becomes possible to monitor forest conditions dynamically, allowing for quicker detection of potential fire risks. This integration would significantly enhance the accuracy and timeliness of predictions, enabling early interventions and more effective resource allocation during fire season Use Hybrid Models Combining ML and Deep Learning**:** Another avenue for improving fire prediction is the development of hybrid models that combine traditional machine learning (ML) techniques with deep learning approaches. While ML models like Decision Trees and Logistic Regression are efficient for certain tasks, deep learning methods, such as Neural Networks, excel at capturing complex, non-linear relationships in large datasets. A hybrid model could leverage the strengths of both approaches, enhancing accuracy by combining the interpretability of ML algorithms with the power of deep learning's ability to learn intricate patterns. This integration would result in more robust models capable of handling diverse and large-scale datasets, improving both prediction performance and generalization.

X. RESULT

Fig 1-The image shows a forest fire spreading through a wooded area, with flames and smoke affecting the surrounding trees and vegetation.

© 2025 IJMRSET | Volume 8, Issue 4, April 2025| DOI: 10.15680/IJMRSET.2025.0804019

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

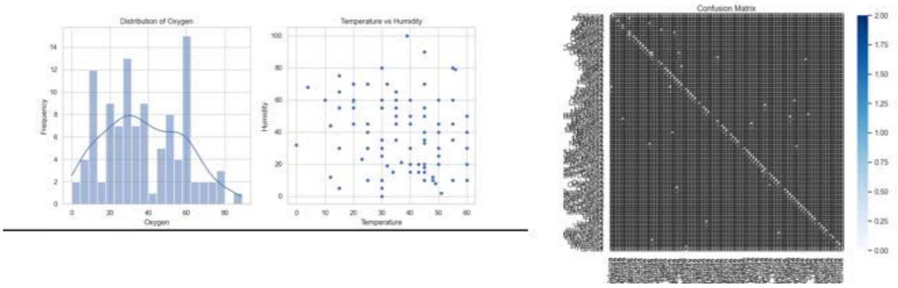

Fig 2-The image contains three visualizations: a histogram of oxygen distribution, a scatter plot of temperature vs. humidity, and a confusion matrix, likely used for analyzing a machine learning model's performance in forest fire prediction

XI. CONCLUSION

This study compares various machine learning algorithms for forest fire prediction. Ensemble methods like Random Forests and Gradient Boosting offer the best trade-off between performance and usability. Future research should focus on hybrid approaches and incorporating real-time data to enhance prediction systems further.

Forest fire prediction is an essential component of disaster management, enabling authorities to take preventive measures and mitigate potential damage. Logistic regression provides a straightforward and interpretable approach to predicting fire occurrences based on key environmental factors such as temperature, humidity, wind speed, and rainfall. The model effectively identifies the probability of fire outbreaks, making it a useful tool for early warning systems. However, its limitations lie in its inability to handle complex, non-linear relationships present in wildfire dynamics, which can reduce its accuracy in certain scenarios. While logistic regression performs well with structured and linearly separable data, its predictive power can be enhanced by incorporating feature selection techniques, regularization methods, and integration with more advanced models such as Random Forest and Deep Learning. Additionally, real-time data collection using IoT sensors and satellite imagery can further improve fire prediction accuracy. Despite its simplicity, logistic regression remains a valuable starting point for wildfire risk assessment, contributing to efficient resource allocation, improved emergency response, and better environmental protection strategies.

Forest fire prediction is a crucial aspect of disaster management, enabling timely interventions to minimize environmental and economic losses. Logistic regression, as a simple yet effective classification model, helps identify key factors influencing fire occurrences, such as temperature, humidity, wind speed, and rainfall. While the model offers interpretability and computational efficiency, its predictive accuracy is limited in complex, non-linear scenarios. To enhance prediction performance, logistic regression can be combined with feature engineering or integrated into ensemble methods. Despite its limitations, the model serves as a valuable tool for early fire detection, supporting proactive decision-making and resource allocation for wildfire prevention and control.

REFERENCES

1.The implementation of machine learning models for forest fire prediction proves to be highly effective in minimizing wildfire damage by providing early warnings and estimating fire spread.
2.Cortez, P., & Morais, A. (2007). A Data Mining Approach to Predict Forest Fires using Meteorological Data.
3.Source: UCI Machine Learning Repository.
4.Summary Introduces a predictive model using meteorological variables to estimate burned areas.
5.Chuvieco E.et al. (2010). "Integrating geospatial information into fire risk assessment models".
6.Journal**:** International Journal of Wildland Fire.
7.Summary**:** Discusses GIS-based fire risk assessment models and data sources.
8.Rodrigues, M., et al. (2019). "Machine learning for large-scale wildfire forecasting".
9.Journal: Environmental Modelling & Software.
10.Summary**:** Uses ML algorithms for large-scale wildfire forecasting based on historical and meteorological data.

INTERNATIONAL JOURNAL OF

MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com