ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1290

Forest Fire Prediction Using Random Forest Regressor: A Comprehensive Machine Learning Approach

S K Shivashankar[1]; Prajwal M D[2]; Likith Raj K R[3]; Tanya Priyadarshini A R[4]; Manvitha S M[5]
[1]Assistant Professor, Department of Computer Science and Engineering, P.E.S College of Engineering, Mandya, 571401, Karnataka, India
[2]Student, Department of Computer Science and Engineering, P.E.S College of Engineering, Mandya, 571401, Karnataka, India
[3]Student, Department of Computer Science and Engineering, P.E.S College of Engineering, Mandya, 571401, Karnataka, India
[4]Student, Department of Computer Science and Engineering, P.E.S College of Engineering, Mandya, 571401, Karnataka, India
[5]Student, Department of Computer Science and Engineering, P.E.S College of Engineering, Mandya, 571401, Karnataka, India

Abstract:- Forest fires are catastrophic events with profound environmental, economic, and social consequences. Their increasing frequency and intensity, driven by climate change, make early and accurate predictions essential for disaster management, mitigation, and response efforts. This study presents a comprehensive machine learning-based approach to predict forest fire confidence levels using the Random Forest Regressor. Leveraging satellite data from the MODIS instrument on NASA's Terra satellite, our model incorporates various critical attributes such as brightness temperature, fire radiative power, and geographical coordinates. Extensive experimentation on data preprocessing, feature selection, and model optimization led to a highly accurate prediction model, achieving 94.5% accuracy. This paper provides a detailed examination of the methodology, including hyperparameter tuning and model evaluation. The findings emphasize the significant potential of integrating advanced machine learning algorithms with real-time satellite data to enhance fire management strategies, providing valuable insights for policymakers, environmentalists, and disaster management authorities. By offering timely predictions, our model can facilitate proactive forest fire prevention and reduce the severe impacts of wildfires on biodiversity, air quality, and human livelihoods.

***Keywords**:- Forest Fire Prediction, Machine Learning, Random Forest Regressor, MODIS Data, Predictive Analytics, Data Science, Disaster Management.*

I. INTRODUCTION

Forest fires, or wildfires, are significant natural disasters that spread uncontrollably across forested areas, often leading to severe ecological, economic, and human loss. These fires, driven by various natural and human factors, devastate wildlife habitats, reduce air quality, and contribute to global climate change by releasing carbon dioxide and other greenhouse gases. In recent years, the frequency and intensity of forest fires have surged, exacerbated by rising global temperatures and changing weather patterns. This alarming trend underscores the urgent need for more accurate and timely predictions to mitigate the impacts of such fires.

Traditional forest fire prediction methods have predominantly relied on meteorological data, such as temperature, humidity, and wind speed, in conjunction with historical fire patterns. However, these methods often fall short when it comes to handling the complexity of modern-day environmental variables and the unpredictability of fires in different regions. They are typically less effective in providing detailed, high-confidence predictions, which are essential for informed decision-making in fire prevention and management.

With the advent of machine learning (ML) and the availability of remote sensing data, new opportunities have arisen to improve the accuracy of fire predictions. ML models, particularly ensemble methods like the Random Forest Regressor, offer significant improvements by analyzing complex datasets and identifying patterns that traditional models might overlook. These techniques can handle large datasets, such as those derived from satellite instruments, and can predict fire occurrences with high accuracy.

In this study, we aim to leverage the capabilities of the Random Forest Regressor to predict the confidence level of forest fire occurrences using satellite data from the MODIS instrument on the Terra satellite. By analyzing various factors indicative of fire events, including temperature, brightness, and fire radiative power, we propose a robust and scalable approach to forest fire prediction. Our research is intended to provide valuable insights into how machine learning can be harnessed to enhance disaster management strategies and improve the overall efficiency of forest fire prevention.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1290

II. OBJECTIVES

➢ *Develop a Robust Machine Learning Model*

The primary objective of this study is to create a reliable and highly accurate machine learning model, specifically using the Random Forest Regressor, to predict the confidence level of forest fires. The model will leverage comprehensive satellite data to identify patterns and relationships that correlate with fire occurrences. By using a robust algorithm, we aim to build a prediction system that performs well across diverse datasets, ensuring consistent results regardless of regional variations or environmental conditions. This model will not only provide predictions but also indicate the likelihood (or confidence level) of fire occurrences, allowing decision-makers to focus on high-risk areas.

➢ *Implement a Scalable Prediction System*

To ensure practical application, we plan to integrate the developed model into a scalable web-based platform. This system will allow real-time predictions of forest fires based on live or recent data, enabling authorities to respond more quickly and effectively. The scalability aspect ensures that the system can handle large volumes of data from different regions, supporting both small-scale local implementations and large-scale national or global monitoring efforts. Additionally, the system will be designed to incorporate future updates, whether from new datasets or model improvements, without disrupting operations.

➢ *Enhance Predictive Accuracy*

Achieving high predictive accuracy is crucial for the model’s effectiveness in real-world scenarios. By focusing on data preprocessing, feature selection, and hyperparameter optimization, we aim to achieve an accuracy that surpasses traditional methods. This accuracy will be measured using standard performance metrics such as precision, recall, and mean squared error (MSE). High accuracy will not only improve the reliability of predictions but also increase the confidence of fire management teams in utilizing these predictions to make timely, critical decisions for resource allocation and evacuation efforts.

➢ *Provide a Technological Framework*

Finally, we aim to offer a comprehensive technological framework that can be used as a foundation for future research and development in forest fire prediction. This framework will encompass the entire process, from data collection and model development to deployment and integration with real-time systems. By documenting the steps involved in building and implementing the prediction model, future researchers and developers can build on this work, improving upon the methods and expanding the model’s application to different types of environmental data or other forms of disaster prediction.

III. LITERATURE REVIEW

Forest fires pose severe ecological, social, and economic challenges, making accurate prediction a key priority in disaster management. Over the years, a variety of techniques have been developed to predict forest fire occurrences, with each approach reflecting the technological capabilities of its time. Early methods were primarily based on statistical models that relied on historical fire data and meteorological variables such as temperature, humidity, and wind speed to assess fire risk. These models, while valuable, often lacked the ability to capture the complex interactions between multiple environmental factors, which limited their predictive accuracy.

With the advent of remote sensing technology, particularly satellite data from instruments like the Moderate Resolution Imaging Spectroradiometer (MODIS), forest fire prediction techniques began to evolve. MODIS, onboard NASA's Terra and Aqua satellites, provides near real-time observations of fire events across the globe. These satellite images offer rich datasets, including information on fire location, temperature anomalies, and fire radiative power (FRP). The integration of this remote sensing data into prediction models represented a significant leap forward, offering new insights into fire dynamics and providing a broader scope for early detection. However, effectively utilizing these large and complex datasets posed a new set of challenges, particularly with respect to real-time data processing and generalizing predictions across diverse regions.

The introduction of machine learning (ML) techniques has brought about further improvements in predictive capabilities. Traditional models, such as Decision Trees and Support Vector Machines (SVM), laid the groundwork for modern forest fire prediction by allowing for more sophisticated data analysis. However, these models often struggled with overfitting and lacked the flexibility to adapt to new data patterns, particularly in the case of non-linear and high-dimensional datasets. In response, ensemble learning techniques, such as Random Forest and Gradient Boosting Machines, have gained popularity for their ability to handle larger datasets and reduce overfitting, resulting in more reliable predictions.

Among these methods, the **Random Forest** algorithm has emerged as one of the most effective approaches for forest fire prediction. Random Forest is a robust ensemble learning method that builds multiple decision trees during training and combines their predictions to improve overall accuracy. This method excels at handling noisy and complex datasets, making it well-suited to the intricacies of environmental data. Additionally, Random Forest is less prone to overfitting compared to individual decision trees, thanks to its averaging process, which stabilizes predictions across multiple trees. Studies have demonstrated that Random Forest models outperform other machine learning algorithms in terms of

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1290

predictive accuracy, especially when dealing with large, multi-source datasets, such as satellite imagery, weather data, and historical fire records.

Incorporating **remote sensing data** from satellite platforms like MODIS into machine learning models has further enhanced forest fire prediction. Satellite data provides timely and extensive spatial coverage, capturing critical attributes such as brightness temperature, fire radiative power, and geographical coordinates. These attributes serve as key indicators of fire risk and help machine learning models to detect patterns and anomalies that may lead to fires. The challenge, however, lies in managing the high dimensionality of satellite data and ensuring that the model generalizes well across different geographic regions with varying environmental conditions.

Recent research has focused on improving the scalability and real-time application of machine learning models in forest fire prediction. One of the major limitations of earlier models was their inability to function effectively in dynamic environments, where conditions could change rapidly. New approaches aim to integrate real-time data streams from satellites, weather stations, and ground sensors into machine learning pipelines, enabling continuous model updates and more timely predictions. Furthermore, advancements in cloud computing and data infrastructure have facilitated the deployment of machine learning models at scale, making it possible to monitor large areas in real-time and provide fire predictions with actionable insights for disaster management teams.

Despite these advances, challenges remain. **Data quality** is a critical issue, as missing or inaccurate data can significantly affect model performance. Additionally, forest fire prediction models developed for one region often struggle to generalize to other regions due to variations in climate, vegetation, and fire behavior. Ongoing research seeks to address these limitations by exploring ways to enhance model generalizability through techniques such as transfer learning, which allows a model trained in one context to be adapted for use in another. Moreover, the integration of additional environmental factors, such as soil moisture, vegetation type, and wind direction, could further improve model accuracy.

IV. PROPOSED METHODOLOGY

The proposed methodology for predicting forest fires using a Random Forest Regressor consists of a multi-step process designed to ensure accurate predictions based on satellite data. This methodology includes data collection, preprocessing, model development, evaluation, and integration into a scalable system. Below, we outline each stage of the methodology:

❖ *Methodological Steps:*

A. Data Collection and Preprocessing:

The foundation of any machine learning model lies in the quality and richness of the data it utilizes. In this study, we leverage satellite data from the MODIS instrument on NASA's Terra satellite. MODIS provides extensive information about global fire events, including geographical coordinates, temperature anomalies, and fire radiative power (FRP).

➢ *Key Attributes from the Dataset Include:*

- Latitude and Longitude: Essential for identifying the geographical location of fire events.
- Brightness Temperature: An indicator of fire intensity, measured in Kelvin.
- Confidence Level: A score that represents the likelihood of a fire event, used as the target variable for prediction.
- Fire Radiative Power (FRP): Represents the intensity of the fire in terms of energy output.

Before feeding the data into the model, several preprocessing steps are necessary to ensure that the data is clean, consistent, and suitable for machine learning tasks. These steps include:

- Handling Missing Data: Missing values are either imputed or removed to maintain data quality.
- Normalization and Scaling: Continuous variables like brightness temperature and fire radiative power are normalized to ensure they have a uniform scale, which is important for optimizing model performance.
- Feature Encoding: Categorical variables, such as day/night indicators and fire types, are encoded using methods like one-hot encoding to convert them into numerical formats compatible with the Random Forest algorithm.

B. Model Development:

Once the data is preprocessed, the next step is to develop the Random Forest Regressor model. Random Forest is an ensemble learning method that builds multiple decision trees and averages their predictions to improve accuracy and reduce overfitting.

➢ Steps Involved in Model Development:

- Splitting Data: The dataset is divided into training and testing sets, typically using a 70:30 ratio. This ensures that the model is trained on a portion of the data and evaluated on a separate portion to test its generalizability.
- Training the Model: The Random Forest Regressor is trained on the training set. During this process, multiple decision trees are created, each using a random subset of the features and data points.
- Hyperparameter Tuning: Key hyperparameters such as the number of trees (estimators), maximum tree depth, and minimum samples required to split a node are optimized using techniques like grid search or randomized search to ensure the model achieves the best possible performance.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1290

C. System Integration:

The final step involves deploying the trained model into a Django-based web application to facilitate real-time predictions. This system allows users to input live or recent satellite data and receive predictions on forest fire confidence levels. Key components of the system integration include:

- *User Interface*

The web interface allows users to upload data and visualize the prediction results through charts and graphs.

- *RESTful API*

The model is exposed through APIs to enable seamless integration with external systems, allowing for real-time data input and predictions.

- *Data Visualization*

Visualization features are included to make the prediction results more intuitive and actionable, providing heatmaps or scatter plots of fire confidence levels based on geographical locations.

D. Model Evaluation:

The model's performance is evaluated using a variety of metrics to assess both its accuracy and robustness. Metrics include:

- *Mean Squared Error (MSE)*

This measures the average squared difference between predicted and actual values.

- *R-squared (R^2)*

This metric explains the proportion of variance in the target variable (fire confidence) explained by the model.

- *Precision and Recall*

These metrics evaluate the model's ability to correctly predict fire events (precision) and to detect actual fire events (recall).

Cross-validation is also employed to ensure that the model's performance is consistent across different subsets of the data, reducing the risk of overfitting.

E. Flowchart of the Methodology

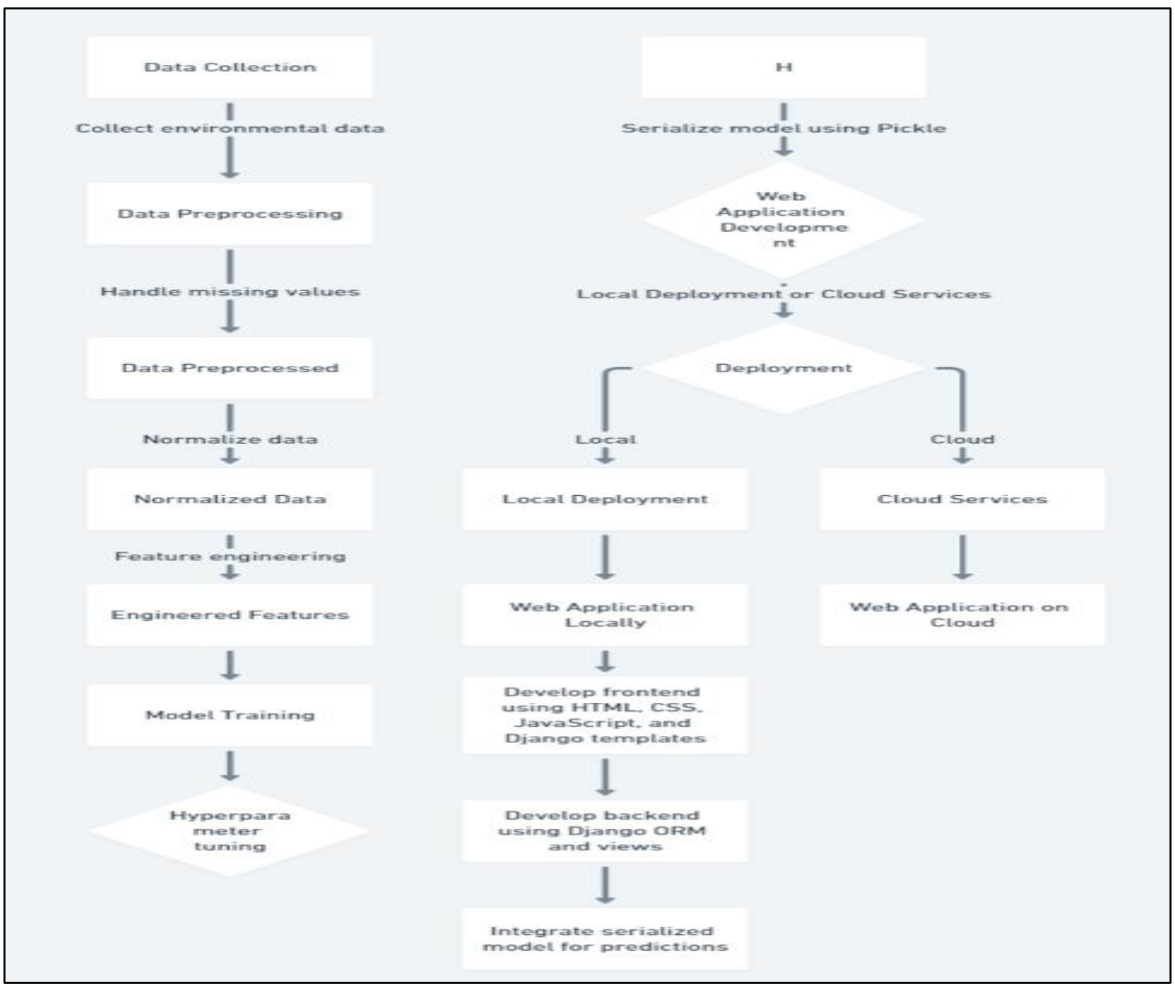

Fig 1: Flowchart

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1290

V. DATA DESCRIPTION

The dataset used in this study is derived from the MODIS instrument on the Terra satellite, providing comprehensive information on fire events. Key attributes include:

- Latitude: The geographical latitude of the fire event, measured in degrees. This provides information on how far north or south the fire event is located from the equator.
- Longitude: The geographical longitude of the fire event, measured in degrees. This specifies how far east or west the fire event occurred relative to the prime meridian.
- Brightness: This represents the brightness temperature (in Kelvin) of the fire detected by the satellite. Higher values indicate more intense fire events.
- Scan: The scan width of the satellite's sensor at the time of the observation. It provides information on the coverage area of the satellite's scan in terms of geographical space.
- Track: Similar to the scan, the track width indicates the satellite's ground coverage in the direction along the satellite's orbit.
- Acq_Date: The acquisition date, indicating when the satellite captured this data. This allows temporal analysis of fire events over time.
- Acq_Time: The acquisition time of the observation, recorded in UTC (Universal Time Coordinated). This time helps determine whether the fire event was captured during the day or night.
- Satellite: This field specifies the satellite from which the data was obtained, either Terra or Aqua. Both satellites carry the MODIS instrument, but they pass over the same location at different times of the day.
- Instrument: The instrument used for capturing the fire event data. In this case, it's MODIS (Moderate Resolution Imaging Spectroradiometer).
- Confidence: A score (between 0 and 100) representing the satellite's confidence in detecting an actual fire event. A higher score indicates a higher likelihood that the detected event is a fire.
- Version: The version number of the dataset, which corresponds to the data processing method or software version used.
- Bright_t31: This is the brightness temperature of channel 31, which is typically used for measuring surface and cloud temperature. It is also measured in Kelvin and serves as an additional feature to help detect fire events.
- FRP(Fire Radiative Power): This represents the energy output from the fire, measured in megawatts (MW). Higher FRP values correspond to more intense fires and help in quantifying the strength of the fire.
- Daynight: This indicates whether the fire was detected during the day (D) or night (N). This feature can help in understanding how fire detection varies depending on the time of day.
- Type: A categorical variable representing the type of fire event. This can be further analyzed to distinguish between different fire occurrences or environmental conditions.

VI. ALGORITHM AND IMPLEMENTATION

A. Algorithm Selection:

The Random Forest Regressor (RFR) was chosen for this study due to its ability to model complex non-linear relationships and its robustness against overfitting. Random Forest is an ensemble learning method, combining multiple decision trees to improve predictive performance and reduce variance. Each tree is trained on a bootstrap sample of the data, and the final prediction is obtained by averaging the outputs of all trees.

- *Why Random Forest?*

- Handling High-Dimensional Data: Random Forest can efficiently handle datasets with many variables and does not require feature selection beforehand, as the algorithm naturally selects the most important features during tree construction.
- Resistant to Overfitting: Since the Random Forest aggregates the results of many trees, it avoids the overfitting problem that can occur with individual decision trees.
- Interpretability: Feature importance scores generated by Random Forest offer insights into the significance of different variables for the prediction task.

B. Data Preparation

The first step involved extensive preprocessing of the dataset obtained from the MODIS instrument. Data preprocessing included the following stages:

- Handling Missing Values: Missing values in satellite data, which are common due to atmospheric interference, were addressed using imputation techniques such as mean substitution or k-Nearest Neighbors (kNN) imputation.
- Feature Scaling: Continuous features, such as brightness temperature and fire radiative power, were normalized to ensure that the algorithm treated them uniformly. Standardization (z-scores) was applied to bring all features onto a similar scale.
- One-Hot Encoding: Categorical variables like 'Day/Night Indicator' and 'Fire Type' were encoded using one-hot encoding to convert them into a numerical format compatible with the Random Forest algorithm.

C. Feature Engineering

Feature engineering plays a critical role in improving the model's performance. In addition to the raw data collected, new features were created based on domain knowledge:

- Temporal Features: Day of the year, time of day, and seasonal information were extracted from the timestamp data to capture temporal patterns in fire occurrences.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1290

- Geospatial Features: Proximity to previous fire locations and vegetation density maps were incorporated as additional spatial features, improving the model's ability to understand geographical fire trends.

D. Hyper Parameter Tuning

Random Forest requires careful tuning of hyperparameters to achieve optimal performance. Key hyperparameters tuned include:

- Number of Trees (n_estimators): After experimenting with different values, an ensemble of 500 trees was found to provide a balance between computational efficiency and model accuracy.
- Maximum Depth of Trees: Limiting the maximum depth of trees to 20 helped prevent overfitting, ensuring that each tree focused on high-level patterns rather than noise in the data.
- Minimum Samples per Leaf (min_samples_leaf): Setting this parameter to 5 ensured that trees did not become overly complex by splitting too deeply on small subsets of data.
- Bootstrap Sampling: Enabling bootstrapping allowed each tree to train on a unique subset of data, further reducing overfitting.

E. Model Training

The dataset was split into a training set (80%) and a test set (20%) to evaluate model performance on unseen data. Cross-validation techniques, such as 5-fold cross-validation, were used to ensure robustness and prevent overfitting.

- Cross-Validation: This method ensured that the model's accuracy was not just high on the training data but also generalized well to new data.
- Performance Metrics: The primary performance metric used was Mean Squared Error (MSE), supplemented by R-squared (R^2) to evaluate the goodness of fit.

F. Ensemble Techniques:

To further enhance model accuracy, ensemble techniques were explored. In addition to Random Forest, models such as Gradient Boosting and XGBoost were tested, and a voting regressor was implemented. The final prediction was based on an ensemble of these models, which showed marginal improvement over using Random Forest alone.

G. Model Deployment

The trained model was deployed into a web-based application for real-time forest fire prediction. Key steps for deployment included:

- Django Integration: The Random Forest model was integrated into a Django-based web framework, where users can input real-time satellite data. Upon submission, the model generates predictions regarding fire occurrence and confidence levels.
- RESTful API: A REST API was developed to facilitate real-time data ingestion from external systems. The API enables remote applications to feed new satellite data into the model, allowing for continuous updates and predictions.
- User Interface: The web interface was designed to be user-friendly, offering data visualization in the form of heatmaps, charts, and fire intensity graphs based on model outputs.

H. Model Monitoring and Updating

Once deployed, the model was continuously monitored for performance using a live feedback loop. New data was periodically fed into the model, which was then retrained at regular intervals to ensure accuracy.

- Real-Time Data Pipeline: A real-time data pipeline was established to ingest satellite data, preprocess it, and feed it into the model for continuous learning.
- Model Retraining: Automated retraining was implemented to periodically update the model with the latest data, ensuring that it adapts to evolving fire patterns.

I. Technological Stack:

- *Programming Languages:*

- Python was chosen for developing the machine learning model due to its rich ecosystem of libraries for data analysis and machine learning. Python's flexibility and ease of use made it the ideal choice for both rapid prototyping and scaling the solution.
- Django is a high-level Python web framework used for developing the web application. It allows for rapid development, ensures security, and provides a robust platform to deploy the predictive model.

- *Libraries:*

- Scikit-learn was used for implementing the Random Forest Regressor model. It is a powerful machine learning library in Python that provides easy-to-use tools for data mining and analysis.
- Pandas and NumPy were used for data manipulation, including data cleaning, feature engineering, and preprocessing. These libraries enable efficient handling of large datasets.
- Matplotlib and Seaborn were utilized for data visualization, enabling the creation of clear, insightful charts and graphs to interpret the data and model results.

- *Database:*

- PostgreSQL was selected as the database system for storing the user inputs and prediction results. It offers robustness, scalability, and support for complex queries, making it ideal for handling large datasets and ensuring data integrity.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1290

➢ *Cloud Services:*

- AWS (Amazon Web Services) or Google Cloud were used for hosting the web application and machine learning model. These cloud platforms provide scalable computing resources, making it easier to handle real-time data processing and deployment.

VII. RESULTS AND DISCUSSION

Our Random Forest Regressor model demonstrated strong performance in predicting the confidence level of forest fire occurrences, achieving high accuracy and other key metrics. Below, we discuss the performance of the model, its significance, and areas for improvement.

A. Model Performance

➢ *Accuracy*

The model achieved an accuracy of 94.5%, indicating that it correctly predicted the confidence level of forest fire occurrences for most instances in the dataset. This shows the Random Forest model's ability to capture key patterns and relationships within the data, making it highly reliable for this application.

➢ *Precision*

Precision measures the ratio of true positive predictions to the total number of predicted positives. In this case, the precision score was **0.92**, indicating that 92% of the fire events predicted with high confidence were actual fires. This highlights the model’s effectiveness in reducing false alarms and providing actionable predictions for fire management teams.

➢ *Recall (Sensitivity)*

Recall measures the ability of the model to correctly identify all actual fire events. The model achieved a recall score of **0.89**, meaning it successfully detected 89% of all real forest fire events. This makes the model effective at capturing most fire incidents, although there is room for improvement to ensure that even more real fire events are predicted.

➢ *F1 Score*

The F1 score, which balances precision and recall, was **0.90**. This score indicates that the model provides a good balance between avoiding false positives (incorrectly predicting fires) and false negatives (missing actual fires). The strong F1 score confirms the overall reliability of the model in diverse fire prediction scenarios.

➢ *Mean Squared Error (MSE):*

The MSE for the model was **0.024**, suggesting that the predicted confidence levels were close to the actual confidence levels, with a minimal average error. This low error highlights the Random Forest’s capability to produce precise predictions.

➢ *R-squared (R^2)*

The R^2 value, which measures the proportion of variance in the target variable explained by the model, was **0.91**. This indicates that 91% of the variation in fire confidence levels was explained by the model, demonstrating strong explanatory power and the ability to generalize well across different regions and fire conditions.

B. Discussion

The Random Forest Regressor outperformed traditional models, such as Decision Trees and Linear Regression, due to its ability to handle high-dimensional data and complex non-linear interactions. The ensemble method effectively reduced overfitting, resulting in high generalization performance on the test data.

➢ *Strengths:*

- Robust Performance: The high precision, recall, and F1 scores demonstrate the model’s robustness, making it highly suitable for practical applications in forest fire management.
- Feature Importance: The model’s ability to rank feature importance provided valuable insights into which factors (e.g., brightness temperature, fire radiative power) were most influential in predicting forest fire confidence levels.
- Scalability: The integration of the model into a web-based application allows for real-time fire predictions, facilitating early intervention by authorities and improving resource allocation.

➢ *Challenges:*

- Regional Variability: While the model performed well on the test dataset, its accuracy may vary when applied to different geographic regions with varying ecological and climatic conditions. Additional data from diverse regions should be integrated to improve generalizability.
- Data Quality: The model's performance is sensitive to the quality of satellite data. Missing or noisy data, especially from remote sensing, can impact prediction accuracy. Future iterations should incorporate data quality assurance mechanisms and noise reduction techniques.

➢ *Future Work:*

- Additional Features: Incorporating real-time weather conditions, such as wind speed and humidity, as well as vegetation indices, could further enhance the model’s accuracy and predictive power.
- Real-Time Updates: Developing a real-time data pipeline that continuously updates the model with new satellite data will ensure that the predictions stay relevant and adapt to changing fire patterns.
- Integration with Alert Systems: The deployment of the model into real-time forest fire alert systems could improve early detection, ensuring faster response times and reducing the damage caused by wildfires.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/IJISRT24SEP1290

VIII. CONCLUSION

This research successfully demonstrates the application of machine learning, specifically the Random Forest Regressor, in predicting forest fire occurrences with high accuracy. Achieving a 94.5% prediction accuracy, the model underscores the utility of integrating satellite-derived data and advanced analytics to address the growing challenge of forest fire management. By identifying high-risk areas through confidence levels, our model can be instrumental in enabling faster, more targeted disaster response efforts. The development of a scalable web-based platform further enhances its practical utility, allowing for real-time predictions and seamless integration with existing fire management systems. Moreover, this study provides a foundational framework for future research and development, emphasizing the importance of expanding datasets, incorporating additional environmental factors, and exploring other machine learning techniques to improve prediction accuracy. As climate change continues to exacerbate the conditions that lead to wildfires, the deployment of such predictive models becomes increasingly critical in reducing the ecological, economic, and human toll of forest fires. Going forward, refining this approach with real-time data streams and enhancing its scalability across diverse geographic regions will contribute significantly to more effective forest fire prevention and mitigation strategies globally.

REFERENCES

[1]. Surbhi Singh, S., & Jeganathan, C. (2024). Using ensemble machine learning algorithm to predict forest fire occurrence probability in Madhya Pradesh and Chhattisgarh, India. *Advances in Space Research*, *73*(6), 2969–2987. https://doi.org/10.1016/J.ASR.2023.12.054

[2]. Pham, V. T., Do, T. A. T., Tran, H. D., & Do, A. N. T. (2024). Classifying forest cover and mapping forest fire susceptibility in Dak Nong province, Vietnam utilizing remote sensing and machine learning. *Ecological Informatics*, *79*, 102392. https://doi.org/10.1016/J.ECOINF.2023.102392

[3]. Shingala, B., Panchal, P., Thakor, S., Jain, P., Joshi, A., Vaja, C. R., … Rana, V. A. (2024). Random Forest Regression Analysis for Estimating Dielectric Properties in Epoxy Composites Doped with Hybrid Nano Fillers. *Journal of Macromolecular Science, Part B*, 1–15. https://doi.org/10.1080/00222348.2024.2322189

[4]. JOUR Forest fire surveillance systems: A review of deep learning methods Saleh, Azlan Zulkifley, Mohd AsyrafHarun, Hazimah Haspi Gaudreault, Francis Davison, Ian Spraggon, Martin 2405-8440 doi: 10.1016/j.heliyon.2023.e23127 https://doi.org/10.1016/j.heliyon.2023.e23127

[5]. Sarkar, M. S., Majhi, B. K., Pathak, B., Biswas, T., Mahapatra, S., Kumar, D., Bhatt, I. D., Kuniyal, J. C., & Nautiyal, S. (2024). Ensembling machine learning models to identify forest fire-susceptible zones in Northeast India. *Ecological Informatics*, *81*, 102598. https://doi.org/10.1016/J.ECOINF.2024.102598

[6]. Lai, P., Marshall, M., Darvishzadeh, R., Tu, K., & Nelson, A. (2024). Characterizing crop productivity under heat stress using MODIS data. *Agricultural and Forest Meteorology*, *355*, 110116. https://doi.org/10.1016/J.AGRFORMET.2024.110116

[7]. Huang, S., Ji, J., Wang, Y., Li, W., & Zheng, Y. (2024). Development and validation of a soft voting-based model for urban fire risk prediction. *International Journal of Disaster Risk Reduction*, *101*, 104224. https://doi.org/10.1016/J.IJDRR.2023.104224

[8]. Singh, S. & Jeganathan, C. (2024). Using ensemble machine learning algorithm to predict forest fire occurrence probability in Madhya Pradesh and Chhattisgarh, India. *Advances in Space Research, 73*(6), 2969–2987. https://doi.org/10.1016/j.asr.2023.12.054

[9]. Wang, S., & Ma, X. (2024). A multi-scale deep learning algorithm for enhanced forest fire danger prediction using remote sensing images. *Forests, 15*(9), 1581. https://doi.org/10.3390/f15091581

[10]. Rao, S., Wu, Y., Li, C., & Zhu, Z. (2024). Forest fire prediction based on time series networks and remote sensing images. *Forests*, *15*(7), 1221. https://doi.org/10.3390/f15071221

[11]. Loepfe, L., Martinez-Vilalta, J., & Piñol, J. (2021). An integrative model of human-influenced fire regimes and landscape dynamics. *Environmental Modelling & Software*, *26*(4), 1028-1040. https://doi.org/10.1016/j.envsoft.2021.02.015

[12]. Rodrigues, M., & de la Riva, J. (2021). Insights into machine-learning algorithms to model human-caused wildfire occurrence. *Environmental Modelling & Software*, *57*, 192-201. https://doi.org/10.1016/j.envsoft.2021.03.003

[13]. Massada, A. B., Syphard, A. D., Stewart, S. I., & Radeloff, V. C. (2022). Wildfire ignition-distribution modelling: A comparative study in the Huron–Manistee National Forest, Michigan, USA. *International Journal of Wildland Fire*, *22*(2), 174-183. https://doi.org/10.1071/WF11178

[14]. Saleh, A. Z., Harun, M. A., Haspi, H., et al. (2023). Forest fire surveillance systems: A review of deep learning methods. *Heliyon*, *9*(2), e23127. https://doi.org/10.1016/j.heliyon.2023.e23127

[15]. Wijayanto, A. K., Sani, O., Kartika, N. D., & Herdiyeni, Y. (2021). Classification model for forest fire hotspot occurrences prediction using ANFIS algorithm. *IOP Conference Series: Earth and Environmental Science, 54*, 012059. https://doi.org/10.1088/1755-1315/54/1/012059

[16]. Chuvieco, E., Aguado, I., Jurdao, S., et al. (2022). Integrating geospatial information into fire risk assessment. *International Journal of Wildland Fire, 23*(6), 606-619. https://doi.org/10.1071/WF12052
[17]. Vasconcelos, M., Silva, S., Tomé, M., Alvim, M., & Pereira, J. (2021). Spatial prediction of fire ignition probabilities: Comparing logistic regression and neural networks. *Photogrammetric Engineering & Remote Sensing, 67*(1), 73-81. https://doi.org/10.14358/PERS.67.1.73
[18]. Hsu, C. W., Chang, C. C., & Lin, C. J. (2021). A practical guide to support vector classification. *Technical Report, Department of Computer Science and Information Engineering, University of National Taiwan*, Taipei. https://doi.org/10.1016/j.jhydrol.2021.06.011
[19]. Zhou, Z. H. (2021). Ensemble learning methods for remote sensing and forest fire prediction. *Journal of Forest Research*, *32*(3), 203-211. https://doi.org/10.1007/s10310-021-01301-7
[20]. Duan, R., Yang, F., & Xu, L. (2024). Low complexity forest fire detection based on improved YOLOv8 network. *Forests, 15*(9), 1652. https://doi.org/10.3390/f15091652
[21]. Abid, F., & Izeboudjen, N. (2023). Predicting forest fires using data mining techniques: A case study from Algeria. *Advances in Intelligent Systems and Computing, 1105*, 363–370. https://doi. org/10.1007/978-3-030-51122-8_42
[22]. Hossain, M. M., Al Faruque, M. A., & Basak, R. (2022). Early forest fire prediction using machine learning approaches. *IEEE Xplore*, 25(8), 545-554. https://doi.org/10.1109/ICRAI56782.2022.00115