# Forest Fire Prediction Using Supervised Machine Learning Models

Jakkamsetti Bharath
Master Of Computer Applications
Dept of IT & CA
Andhra University college of Engineering
AU, Visakhapatnam, India
bharathjs44@gmail.com

Dr. K.Swapna
Assistant Professor (Ad-hoc)
Dept of IT & CA
Andhra University college of Engineering
AU, Visakhapatnam, India
swapnakethireddy24@gmail.com

*Abstract* -- **Forest fires are a critical environmental concern, with devastating effects on ecosystems, economies, and human life. Accurate prediction of forest fires is essential for early warning systems and effective resource management. This project explores the use of supervised machine learning techniques, including Random Forest Classifier, Decision Tree Classifier, and Logistic Regression, to predict the occurrence of forest fires based on Fire Weather Index (FWI) components and meteorological parameters. The FWI, a widely used indicator in fire danger systems, includes sub-indices such the Initial Spread Index (ISI), Duff Moisture Code (DMC), and Fine Fuel Moisture Code (FFMC). These components, along with additional meteorological parameters like temperature (in Celsius), relative humidity, and speed of the wind serve as input features for the predictive models. A machine learning model is trained using this data, specifically Random Forest Classifier, to identify patterns and relationships between environmental factors and fire occurrences. The likelihood of a fire can then be predicted by the trained model, based on real-time environmental data. This information can be used by forest authorities to prioritize areas with a high probability of fire and allocate resources accordingly. To facilitate user interaction and accessibility, a concept web application is developed. This project draws attention to the potential of integrating the machine learning with environmental indices to enhance predictive capabilities, providing timely and accurate forecasts that can significantly aid in mitigating the impact of forest fires.**

***Keywords*** **– Wildfire Prediction, Random Forest Classifier, Decision tree, Machine learning, Supervised learning.**

## I. INTRODUCTION

Wild fires are a global concern that is becoming worse, because they destroy ecosystems, wildlife habitats, and human settlements. The increasing frequency and severity of forest fires have significant environmental, economic, and social impacts. Predicting forest fires is crucial for taking preventive measures and minimizing their impact. 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. The Forest Survey of India (FSI) claims that, an estimated 3.7 million hectares of forest are lost every year due to forest fires. This translates to:

- 2012: 4.3 million hectares (approx. 10.6 million acres) of forest lost
- 2017: 4.8 million hectares (approx. 11.9 million acres) of forest lost
- 2020: 5.1 million hectares (approx. 12.6 million acres) of forest lost.

In recent years, various technologies have emerged for fire place models to evaluate the spread of forest fires. For the purpose of defining and forecasting fire development in various places, these models depend on data gathered from laboratory experiments and forest fire models. Physical models, such as Rothermel's Fire Spread Model and the FARSITE model, are based on the physical principles of heat transfer and combustion, and take into account factors such as fuel type, moisture content, wind speed, and topography. Mathematical models, including logistic regression and Markov chain models, are also used to predict forest fires, leveraging statistical and mathematical principles to analyse large datasets and identify patterns and relationships between variables.

The use of simulation tools in addition to these conventional methods has increased in the prediction of forest fires. Simulation tools have, nevertheless, encountered several difficulties, such as input data accuracy and tool execution time. The accuracy of the input data is critical, as any errors or inaccuracies can significantly impact the reliability of the simulation results. Furthermore, the execution time of the simulation tool can be a limiting factor, particularly when dealing with large and complex datasets. To address these challenges, researchers have been exploring new approaches, such as artificial intelligence and by using machine learning.

Artificial intelligence (AI) has a subset called machine learning that allows systems to learn from data without explicit programming. Three general categories may be used to describe machine learning: reinforcement learning, unsupervised learning, and supervised learning. Supervised machine learning algorithms are trained on labelled data, where the target output is already known. Examples of supervised machine learning algorithms include regression, Random Forest, Artificial Neural Networks(ANN), and Decision Tree Classifier. In contrast, unsupervised machine learning algorithms are trained on unlabelled data, where the goal is to discover patterns, relationships, or structure in the data. Since the data attributes are not tagged, the algorithm must identify the underlying labels or categories itself

## II. LITERATURE SURVEY

### *A. Forecasting Forest Fires through Wireless Sensor Network*

M. Hefeeda et al [1] In this project, they showcases a wireless sensor network architecture that utilizes the widely recognized Fire Weather Index (FWI) System to identify forest fires early on. Modeling the detection problem as a k-coverage problem, the research analyses the FWI components to optimize sensor network architecture. An approach to data aggregation that uses the FWI system is proposed to prolong network lifetime by transmitting only relevant data. Simulation results validate the effectiveness of the proposed design in terms of coverage, data aggregation efficiency, and overall fire detection performance.

### *B . Forest Fire Prediction Using Fuzzy AHP-GIS*

S. Eskandari et al. [2] This study models fire risk in Iran's Hyrcanian forests using a fuzzy AHP-GIS approach. It takes into account 17 sub-criteria in addition to the four main criteria (topography, biology, climate, and human variables), allocating weights in accordance with professional judgment. Spatial data are converted to fuzzy maps and overlaid to create a fire risk map. Validation against actual fire data shows the model accurately predicts 80% of fire occurrences. This fire risk map serves as a decision support tool for predicting future fires.

### *C. Forest Fire Prediction Using Image Mining Techniques*

In their project, Divya T L et al. (2015) demonstrated how an image mining technique may be used to anticipate the development of a forest fire by analysing a series of pixel values. Using satellite imagery, the proposed method makes predictions about forest fires.

### *D. An IOT-Powered Smart System to Regulate CO2 Emissions*

Raj Kumar D.M.N. at el [6] This project's primary goal is to use the Raspberry Pi, a CO2 sensitive device, to detect CO2 emissions from industry, public transportation, and forest fires. The city's most polluting location is determined by tracking the amount of carbon dioxide released on a regular basis. Incorporate an intelligent technology to promptly identify wildfires or forest fires. The Internet of Things (IOT), which is more secure and has a wealth of services available, is then interconnected with these. This would allow a Simple Notification Service (SNS) to be sent to the phone in the event that the specific location is contributing to elevated CO2 levels.

### *E. Logistic Regression Based Forest Fire Prediction*

Mukhammad Wildan Alauddin et al. (2018) Multiple linear regression has been proposed for the purpose of forest fire prediction. Among the variables are wind, rain, humidity, and temperature. To compute different linear regression coefficients, several methods including gauss-Jordan, gauss-seidel, and least-squares are utilized. The techniques are compared, and the outcomes are talked about.

## III. PROPOSED SYSTEM

The suggested system seeks to utilize machine learning techniques such as logistic regression, decision tree and random forest classifier to predict the likelihood of forest fire occurrences based on meteorological parameters and FWI components.

By utilizing historical data and real-time meteorological information, the system seeks to provide timely alerts and insights to fire management teams, enabling them to take proactive measures in fire prevention and control.

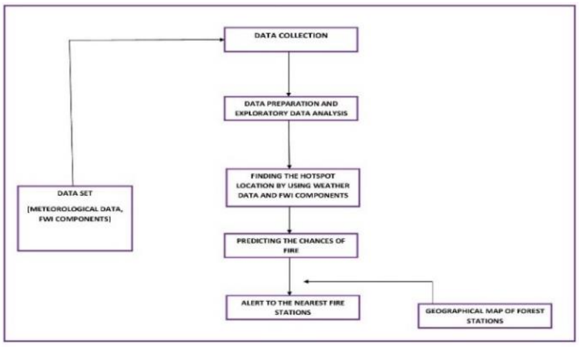

DATA COLLECTION
DATA PREPARATION AND EXPLORATORY DATA ANALYSIS
FINDING THE HOTSPOT LOCATION BY USING WEATHER DATA AND FWI COMPONENTS
DATA SET [METEOROLOGICAL DATA, FWI COMPONENTS]
PREDICTING THE CHANCES OF FIRE
ALERT TO THE NEAREST FIRE STATIONS
GEOGRAPHICAL MAP OF FOREST STATIONS

*Figure 1. Flow Diagram Of Proposed System*

The proposed system block diagram describes how we collected the meteorological and FWI components data set from UCI. Next, we conducted an exploratory analysis which involved pre-processing to try to eliminate noisy data and convert the categorical data to numerical data, making the dataset easier to understand. Following the preprocessing method, the position of the hotspot is determined using the weather data included in the data set. Models are then used to estimate the likelihood of a fire occurring and notify the closest fire station.

## IV. METHODOLOGY

### *A. SYSTEM ANALYSIS AND DESIGN*

A task in software engineering called requirement analysis fills the gap between software design and software allocation at the system level. It gives the system engineer the ability to define the function and performance of the program, show how the software interacts with other system components, and set requirements that the software must adhere to.

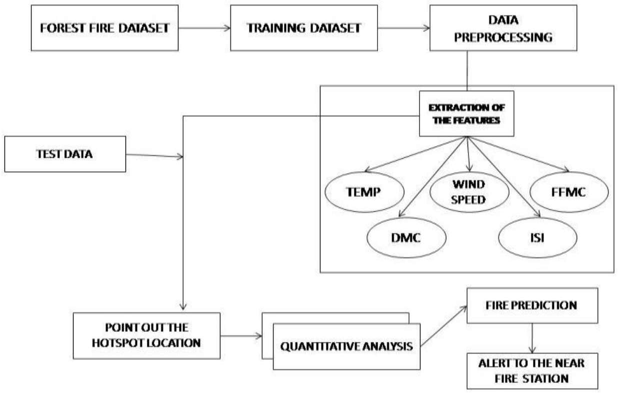


*Figure 2. High Level Diagram For Predicting Forest Fire*

Here are some key points regarding the detailed design of my project:

I utilized a dataset on Algerian Forest Fires sourced from UCI, it contains information and observations from the Algerian districts of Sidi Bel-Abbes and Bejaia. The time frame covered by this dataset is June 2012–September 2012.

The focus of this project was to determine whether specific weather features could be used to predict forest fires in these regions, employing various classification algorithms.

As part of the task, I choose to approach the problem as a classification challenge to predict forest fires.

- The first step involved data collection, followed by data preprocessing to ensure the dataset was formatted consistently.
- After preparing the data, I selected an appropriate model based on the characteristics of the dataset.
- We are using a variety of classification approaches, including Random Forest (RF), Decision Tree (DT), and Logistic Regression, for prediction in this research.
- After implementing these models, we will proceed to evaluate their performance.
- This involves making predictions with each model and assessing their accuracy to determine which model performs best in predicting forest fires.

### *B. SYSTEM ARCHITECTURE*

The system architecture for the proposed forest fire prediction system is designed to facilitate data collection, processing, model training, prediction, and user interaction. It integrates various components to create a cohesive framework that enables accurate forecasting of forest fire occurrences based on meteorological data.

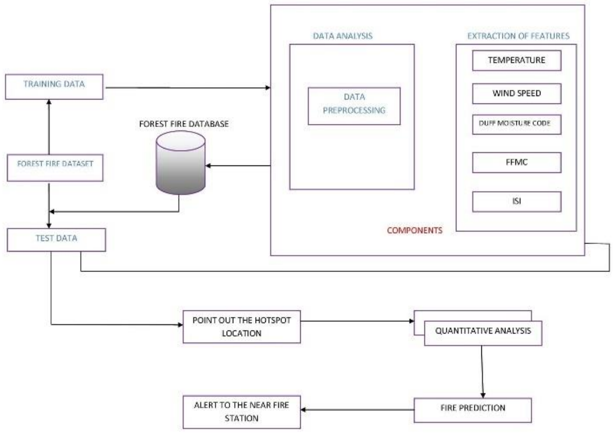


*Figure 3. System Architecture Level Diagram Of Forest Fire Prediction*

A visual depiction of the system is given by an architectural diagram, which shows the relationships, constraints, and divisions among the software's many components as well as the general organization of the program.

### *C. EXPLORATORY DATA ANALYSIS*

In this phase, we will conduct Exploratory Data Analysis (EDA) to derive insights from the dataset and identify which features most significantly contribute to predicting forest fires. This involves data analysis using Pandas and data visualization with Matplotlib and Seaborn. Understanding the data thoroughly and extracting as many insights as possible is considered best practice. The following tasks will be performed during the EDA process:

- Importing Libraries: Load the necessary libraries for data manipulation and visualization.
- Data Cleaning for EDA Report: Perform cleaning and preprocessing of the data to ensure its readiness for analysis.
- Conducting Exploratory Data Analysis (EDA): Analyse all features in the dataset to uncover patterns and insights relevant to forest fire predictions.
- Displaying the Dataset Head: The first five records are shown using the head() function., giving a glimpse of the data structure.
- Inspecting the Columns: The columns attribute allows us to view the dataset's structure, showing that it has 518 observations and 13 variables.
- Checking for Null Values: To ensure data integrity, we use isnull().sum() to display null values in the dataset.
- Dataset Info: The info() function provides comprehensive details about the dataset's dimensions and data types.
- Visualization: We plot distribution graphs (e.g., histograms or bar graphs) to visualize the frequency of unique values across various columns.
- Correlation Graph : Heatmaps are used to plot a correlation matrix, which shows the correlations between the dataset's

various properties, showing how variables depend on one another.

### *D. PREPROCESSING ANALYSIS*

One of the most important steps in the data analysis pipeline is data preprocessing, which creates clean datasets from raw data, enabling efficient analysis. The preprocessing stage encompasses various operations:

**Handling Missing Values**: Initial identification of missing values is followed by the removal of rows containing nulls with dropna().

**Correlation Analysis**: After preprocessing, the correlation matrix is plotted again to analyse the relationships between meteorological features such as humidity, speed of wind, and temperature, informing us about their dependencies.

**Understanding Correlations**: By examining correlation coefficients, we can infer relationships. For instance, a positive sum indicates a positive correlation, while a negative value suggests a negative correlation.

**Data Splitting**: Finally, A 70:30 ratio is used to divide the dataset into training and testing sets ,allowing for model training and evaluation.

We then applied the Random Forest classifier to the dataset to make predictions.

### *E. MACHINE LEARNING TECHNIQUES*

The machine learning algorithms employed in this project are as follows:

1. *RANDOM FOREST CLASSIFIER:* In machine learning, a reliable tree-based learning technique is the Random Forest algorithm. During the training phase, it generates many Decision Trees in order to function.

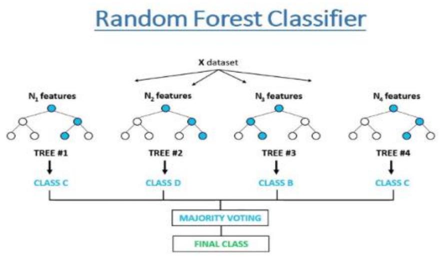


*Figure 4. Structure Of Random Forest*

Every tree is constructed with a randomly chosen subset of the dataset, and at each split, its value is determined by a randomly chosen feature set. The diversity among the trees produced by their intrinsic randomness reduces overfitting and improves overall prediction accuracy.

In order to provide predictions, the algorithm averages (for regression problems) or votes (for classification problems) over all of the tree outputs. By utilizing the insights from several trees, this collaborative decision-making method produces reliable and accurate results. Random forests are widely used in both classification and regression applications. They are well-known for their capacity to handle intricate datasets, reduce overfitting, and provide accurate predictions in a variety of scenarios.

2. *DECISION TREE CLASSIFIER :* A decision tree is a diagrammatic representation resembling a flowchart, used for making decisions or predictions. It comprises nodes that signify decisions or evaluations of specific attributes, branches that illustrate the outcomes of these evaluations, and leaf nodes that denote final results or predictions. Every internal node symbolizes an attribute test, every branch denotes the test's result, and every leaf node correlates to a continuous value or a class label.

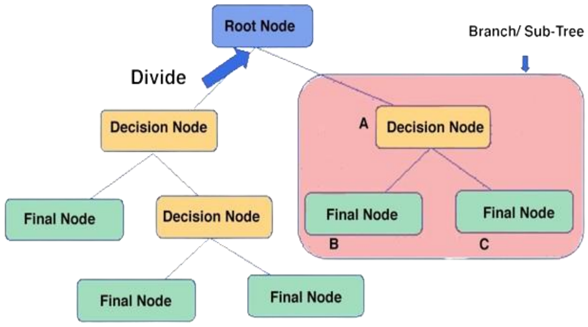


*Figure 5. Structure Of Decision Tree Classifier*

3. *LOGISTIC REGRESSION:* A supervised machine learning approach used for classification problems is called logistic regression. Its goal is to determine the likelihood that a given example falls into a certain class. The link between two variables is investigated using this statistical technique. The sigmoid function, which accepts independent variables as input and produces a probability value ranging from 0 to 1, is mostly used for binary classification.

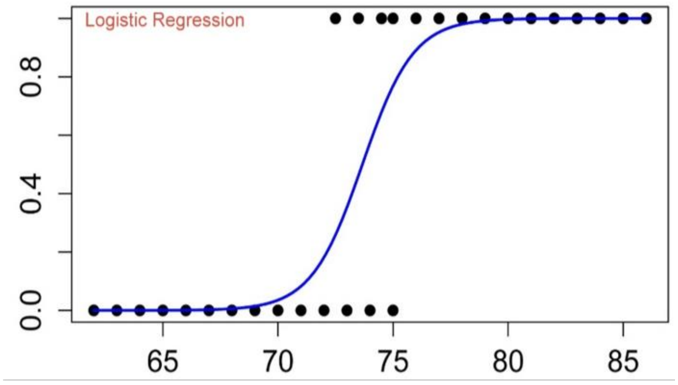


*Figure 6. Curve Of Logistic Regression*

## V. RESULTS ANALYSIS AND OUTPUTS

Exploratory Data Analysis & Results:

1.Plotting the density plot for the all the available features that are present in the dataset.

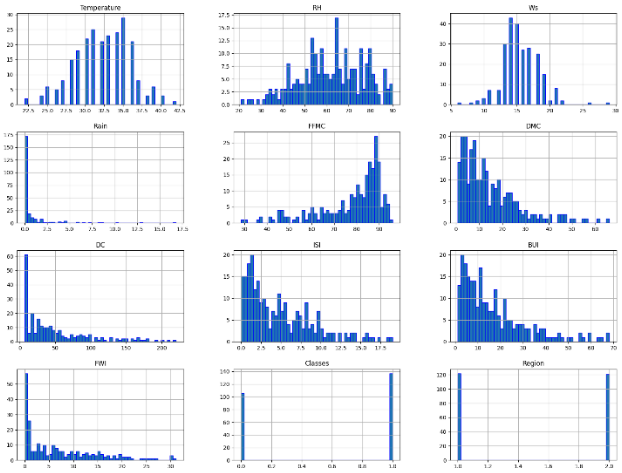


*Figure 7. Density Plot*

2.Plotting the pie chart for the fire and non-fire regions

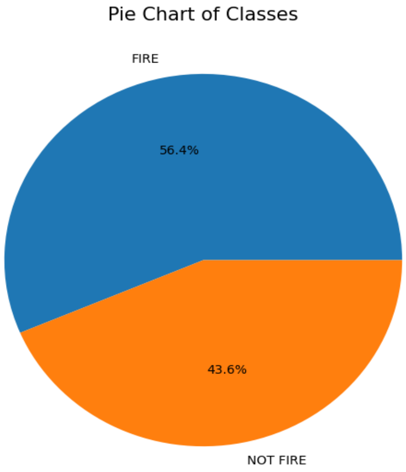


*Figure 8. Pie Chart*

3. Checking multi collinearlity by using pearson correlation and remove highly correlated features which has correlation more than 0.75

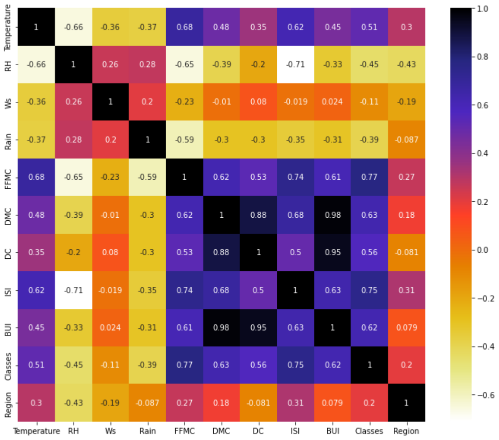


*Figure 9. Correlation Graph*

Data Preprocessing:

4. Standardizing the independent features in the data within a predetermined range by using feature scaling and draw a box plot to the data set to understand the effect of standard scaler.

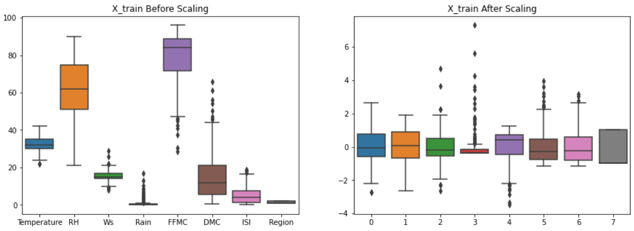


*Figure 10. Box Plot*

5.Confusion Matrix: Based on the model's predictions, the number of accurate and inaccurate instances is displayed using a confusion matrix representation.

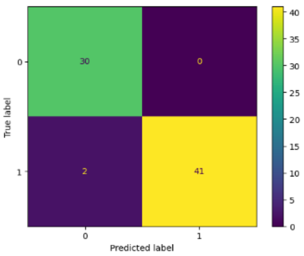


*Figure 11.1 Random Forest Classifier*

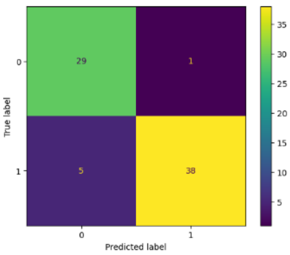
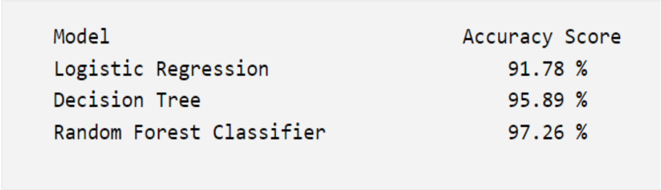


*Figure11.2 Decision Tree*

6.Table of comparision of different models with their accuracy scores



```
Model                         Accuracy Score
Logistic Regression               91.78 %
Decision Tree                     95.89 %
Random Forest Classifier          97.26 %
```

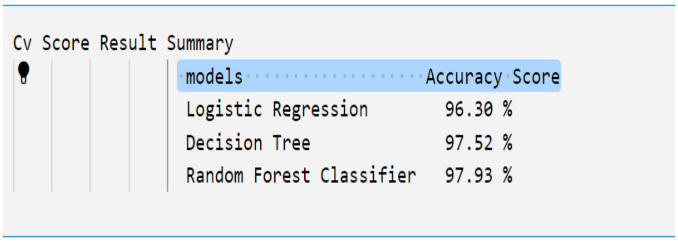
7. Hyper Parameter Tuning with RandomizedSearchCv:

Finding the ideal values for the hyperparameters of a machine learning model is known as hyperparameter tuning. Unlike parameters of model, which are learned during the training process, hyperparameters are typically set before training begins and cannot be directly learned. These hyperparameters play a crucial role in defining key characteristics of the model, such as its complexity and the rate at which it learns.

The RandomizedSearchCV class offers methods like "fit" and "score," along with additional capabilities such as "score_samples", "predict", "predict_proba", "decision_function," "transform," and "inverse_transform," provided they are supported by the designated estimator. This approach allows for the optimization of the estimator's parameters by performing a cross-validated search over a range of parameter configurations.

Model Selection:

Applying Stratified Kfold Cross-Validation to know the exact Mean CV Accuracy Score for all models. In this instance, the training and test sets are guaranteed to contain the same percentage of the feature of interest as the original dataset by applying the stratified K-fold sampling principle in cross-validation. This ensures error-free, high precision when done with the target variable.

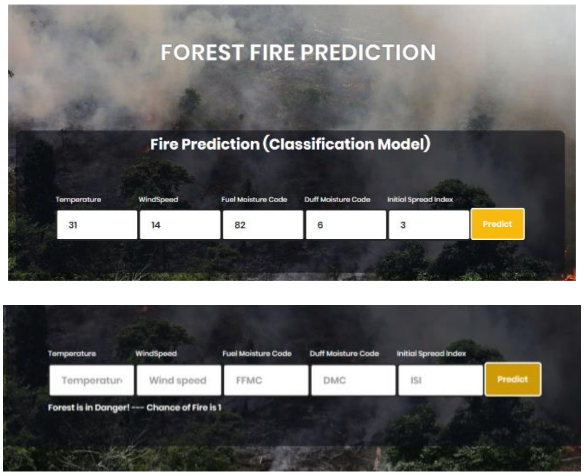


```
Cv Score Result Summary
models                  Accuracy Score
Logistic Regression       96.30 %
Decision Tree             97.52 %
Random Forest Classifier  97.93 %
```

Hence from the above score results, the Random Forest Classifier has given better result. So, we proceed further with that model for model deployment.

WEB APPLICATION:

Finally the above Random Forest Classifier model is applied and integrated with the web development, we can create a web application by using flask, that simply takes 5 inputs from the user to get the forest fire probability.



FOREST FIRE PREDICTION

Fire Prediction (Classification Model)

Temperature: 31
WindSpeed: 14
Fuel Moisture Code: 82
Duff Moisture Code: 6
Initial Spread Index: 3

Temperature: Temperatur
WindSpeed: Wind speed
Fuel Moisture Code: FFMC
Duff Moisture Code: DMC
Initial Spread Index: ISI

Forest is in Danger! --- Chance of Fire is 1

*Figure 12.Web application*

The above image shows the occurrence of fire and how our model predicting the forest fire by using the FWI components and meteorological parameters.

FINAL REPORT:

**Temperature:** Highest Fire counts happened between **31-38degree Celsius**

**Wind Speed** :Highest Fire count happened when the wind speed were between **12 to 20Km/hr.**

**Fine Fuel Moisture Code(FFMC):**The value of this index varies from 28.6 to 92.5, here **above 74** has higher chance of forest fires.

**Duff Moisture Code (DMC): index** which ranges between 1.1 to 65.9, here 1.1-10 has lower chance of Forest fires whereas above **10-30 DMC** has very high evidence of Forest fires in past.

**Initial Spread Index (ISI) index** which ranges between 0 to 18, here 0-3 has lower Forest fires and **above 3 ISI** has higher chance of Forest fires.

## VI. CONCLUSION AND FUTURE SCOPE

Hence, machine learning presents a promising solution for predicting the likelihood of forest fires depending on important meteorological parameters like temperature and wind speed and

FWI components. By analysing historical data, we can train models that accurately assess fire risks, enabling proactive measures to be taken. This prediction can be used for calculating if the fire is possible at the location based on inputs.The integration of these predictive models with web or app development can facilitate real-time monitoring and alerts for forest authorities and citizens to better manage fire risks and protect valuable ecosystems.

This project may be developed further to provide better results, including greater effects and better-equipped models. We may also have a user interface designed to offer some real-time functionality for the program.The user may input their zip code and local address in the UI model's process. We will utilize the coordinates as inputs, use the zip code to get latitude and longitude using any accessible APIs, and get the weather at that moment.

## REFERENCES

[1] M. Hefeeda and M. Bagheri, "Wireless Sensor Networks for Early Detection of Forest Fires," *2007 IEEE International Conference on Mobile Adhoc and Sensor Systems*, Pisa, Italy, 2007, pp. 1-6

[2] S. Eskandari, “A new approach for forest fire risk modeling using fuzzy AHP and GIS in Hyrcanian forests of Iran,” *Journal of Geosciences*, vol. 10, no. 8, p. 190, 2017.

[3] G. Demin, L. Haifeng, J. Anna and W. Guoxin, "A forest fire prediction system based on rechargeable wireless sensor networks," *2014 4th IEEE International Conference on Network Infrastructure and Digital Content*, Beijing, China, 2014, pp. 405-408.

[4] V. Suresh Babu, V. S. K. Vanama, A. Roy and P. R. Prasad, "Assessment of forest fire danger using automatic weather stations and MODIS TERRA satellite datasets for the state Madhya Pradesh, India," *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Udupi, India, 2017.

[5] V. Kumar, A. Jain, and P. Barwal, “Wireless sensor networks: security issues, challenges and solutions,” *International Journal of Information and Computation Technology (IJICT)*, vol. 4, no. 8.

[6] D. M. N. Rajkumar, M. Sruthi, and D. V. V. Kumar, “Iot based smart system for controlling co2 emission,” *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 2, no. 2, p. 284, 2017.

[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] Divya T.L., Manjuprasad B., Vijayalakshmi M.N. and A. Dharani, "An efficient and optimal clustering algorithm for real-time forest fire prediction with," *2014 International Conference on Communication and Signal Processing*, Melmaruvathur, India, 2014, pp. 312-316.

[9] K. V. Murali Mohan, A. R. Satish, K. Mallikharjuna Rao, R. K. Yarava and G. C. Babu, "Leveraging Machine Learning to Predict Wild Fires," *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*, Trichy, India, 2021, pp. 1393-1400.

[10] S. Pawar, K. Pandit, R. Prabhu, R. Samaga and Geethalaxmi, "A Machine Learning Approach to Forest Fire Prediction Through Environment Parameters," *2022 International Conference on Artificial Intelligence and Data Engineering (AIDE)*, Karkala, India, 2022, pp. 1-7.