TOPIC-LEVERAGING TO PREDICT FOREST FIRES USING MACHINE LEARNING

Major project report submitted
in partial fulfillment of the requirement for award of the degree of

Bachelor of Technology
in
Computer Science & Engineering

By

GOLI PAVAN	(19UECS0313)	**(VTU 14400)**
V C JAGAPATHI BABU	(19UECS1107)	**(VTU 13952)**
G KRISHNA CHAITANYA	(19UECS0281)	**(VTU 12658)**

Under the guidance of
Ms R.VAISHANAVI,M.E.,
ASSISTANT PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
SCHOOL OF COMPUTING

VEL TECH RANGARAJAN DR. SAGUNTHALA R&D INSTITUTE OF SCIENCE & TECHNOLOGY

(Deemed to be University Estd u/s 3 of UGC Act, 1956)
Accredited by NAAC with A++ Grade
CHENNAI 600 062, TAMILNADU, INDIA

April, 2023

TOPIC-LEVERAGING TO PREDICT FOREST FIRES USING MACHINE LEARNING

Major project report submitted
in partial fulfillment of the requirement for award of the degree of

Bachelor of Technology
in
Computer Science & Engineering

By

GOLI PAVAN	(19UECS0313)	**(VTU 14400)**
V C JAGAPATHI BABU	(19UECS1107)	**(VTU 13952)**
G KRISHNA CHAITANYA	(19UECS0281)	**(VTU 12658)**

Under the guidance of
Ms R.VAISHANAVI,M.E.,
ASSISTANT PROFESSOR

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
SCHOOL OF COMPUTING

VEL TECH RANGARAJAN DR. SAGUNTHALA R&D INSTITUTE OF SCIENCE & TECHNOLOGY

(Deemed to be University Estd u/s 3 of UGC Act, 1956)
Accredited by NAAC with A++ Grade
CHENNAI 600 062, TAMILNADU, INDIA

April, 2023

CERTIFICATE

It is certified that the work contained in the project report titled ” TOPIC-LEVERAGING TO PREDICT FOREST FIRES USING MACHINE LEARNING ” by ”GOLI PAVAN (19UECS0313), V C JAGAPATHI BABU (19UECS1107), G KRISHNA CHAITANYA (19UECS0281)” has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

Signature of Supervisor
Ms.R.Vaishanavi
Assistant Professor
Computer Science & Engineering
School of Computing
Vel Tech Rangarajan Dr.Sagunthala R&D
Institute of Science & Technology
April, 2023

Signature of Head of the Department
Computer Science & Engineering
School of Computing
Vel Tech Rangarajan Dr. Sagunthala R&D
Institute of Science & Technology
April, 2023

Signature of the Dean
Dr. V. Srinivasa Rao
Professor & Dean
Computer Science & Engineering
School of Computing
Vel Tech Rangarajan Dr. Sagunthala R&D
Institute of Science & Technology
April, 2023

DECLARATION

We declare that this written submission represents my ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

(GOLI PAVAN)

Date: / /

(V C JAGAPATHI BABU)

Date: / /

(G KRISHNA CHAITANYA)

Date: / /

APPROVAL SHEET

This project report entitled " TOPIC-LEVERAGING TO PREDICT FOREST FIRES USING MACHINE LEARNING" by (GOLI PAVAN (19UECS0313),(V C JAGAPATHI BABU(19UECS1107), (G KRISHNA CHAITANYA (19UECS0281) is approved for the degree of B.Tech in Computer Science & Engineering

Examiners **Supervisor**

Ms R.Vaishanavi,M.E.,

Date: / /

Place:

ACKNOWLEDGEMENT

We express our deepest gratitude to our respected **Founder Chancellor and President Col. Prof. Dr. R. RANGARAJAN B.E. (EEE), B.E. (MECH), M.S (AUTO),D.Sc., Foundress President Dr. R. SAGUNTHALA RANGARAJAN M.B.B.S.** Chairperson Managing Trustee and Vice President.

We are very much grateful to our beloved **Vice Chancellor Prof. S. SALIVAHANAN,** for providing us with an environment to complete our project successfully.

We record indebtedness to our **Professor & Dean, Department of Computer Science & Engineering, School of Computing, Dr. V. SRINIVASA RAO, M.Tech., Ph.D.,** for immense care and encouragement towards us throughout the course of this project.

We are thankful to our **Head, Department of Computer Science & Engineering, Dr. M. S. MURALI DHAR, M.E., Ph.D.,** for providing immense support in all our endeavors.

We also take this opportunity to express a deep sense of gratitude to our Internal Supervisor **Ms VAISHANAVI.R,M.E.,** for his cordial support, valuable information and guidance, he/she helped us in completing this project through various stages.

A special thanks to our **Project Coordinators Mr. V. ASHOK KUMAR, M.Tech., Ms. C. SHYAMALA KUMARI, M.E.,** for their valuable guidance and support throughout the course of the project.

We thank our department faculty, supporting staff and friends for their help and guidance to complete this project.

GOLI PAVAN	**(19UECS0313)**
V C JAGAPATHI BABU	**(19UECS1107)**
G KRISHNA CHAITANYA	**(19UECS0281)**

ABSTRACT

Forest Fire Prediction is a key component of forest fire control.The simulation tools have been used to forecast forest fires, but they have encountered issues with data quality and tool execution time. Machine learning, a form of artificial intelligence, has been used to teach computers to predict forest fires. There are two categories of machine learning: supervised and unsupervised, and reinforcement learning. In supervised learning, the entire collection of data is fully labeled, while in unsupervised learning, the algorithm defines the labels.using the machine learning algorithms and meteorological variables, such as temperature, rain, wind, and humidity, forest fires can be predicted, and their future effects can be lessened.Fire Detection is a key element for controlling such incidents. Prediction of forest fire id expected to reduce the impact of forest fire in the future.Many fire detection algorithms are available with different approach towards the detection of fire.In the existing work processes the fire affected region is predicted based on the satellite images.To predict the occurrences of a forest fire the proposed system processes using the meteorological parameters such as temperature, rain, wind and humidity were used. Random forest regression and Hyperparameter tuning using Random Forest algorithm that combines multiple decision trees to make a more accurate prediction. In the case of forest fire prediction, the algorithm can take in various environmental and meteorological data such as temperature, humidity, wind speed, precipitation, and other variables that may affect forest fire risk.

Keywords: Vector Machines, Global Warming,Meteorological Parameters

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATIONS

AI	Artificial Intelligence
ML	Machine Learning
SVM	Support Vector Machine
RFG	Random Forest Algorithm
RFR	Random Forest Regression

TABLE OF CONTENTS

Chapter 1

INTRODUCTION

1.1 Introduction

Forest fires (also known as wildfires) are one of the recent calamities that occur most often. Several acres of forest are being burned as a result of these flames. The primary causes of forest fires are global warming because of the a rise in the earth's average temperature and carelessness on the part of people. Forest fires are brought on by human activities like animal husbandry and agriculture in Africa, South America, Southeast Asia, and New Zealand. The spread of flames may now be predicted using a variety of technologies, including mathematical and physical models. In order to specify and forecast fire development in various places, these models rely on data collecting during forest fires, modelling, and lab experimentation.

The spread of flames may now be predicted using a variety of technologies, including mathematical and physical models. For the purpose of describing and forecasting fire growth in several places, these models rely on data gathered during forest fires, modelling, and lab trials. In recent years, simulation tools have been employed to forecast forest fires, however these systems have encountered issues with data quality and tool execution time. A form of artificial intelligence called machine learning is used to teach computers. Two categories of machine learning exist: supervised, unsupervised, and reinforcement learning.Algorithms for unsupervised machine learning include self-organizing maps and k-means clustering. Reinforcement learning punishes the learning system for incorrect actions and rewards it for the right ones. By using machine learning algorithms and meteorological variables, such as temperature, rain, wind, and humidity, forest fires can be predicted, and their future effects can be lessened.

1.2 Aim of the project

The main goal is to use machine learning techniques to anticipate forest fires based on a few parameters.

1.3 Project Domain

Machine learning is the topic of the project.Artificial intelligence known as machine learning enables software programmes to improve their predictive abilities without having been expressly trained to do so. In order to forecast new output values, machine learning algorithms use past data as input. Internet search engines, spam-filtering email software, websites that offer personalised recommendations, banking software that spots suspicious transactions, and many phone apps like speech recognition all employ machine learning.

1.4 Scope of the Project

This study uses machine learning techniques to forecast forest fires. For more accuracy, this project employs a random forest regressor. Training and test data will be separated from the forest fire exploration dataset. The model for supervised learning receives input from the training data.

Chapter 2

LITERATURE REVIEW

Dieu Tien Buia et al.,[1]patial pattern analysis and prediction of forest fire using new machine learning approach of Multivariate Adaptive Regression .The authors collected data on various environmental and topographical factors and used them to train their MARS-DFP model. The model was able to accurately predict forest fires in the study area, with an accuracy of 90.25 % . The authors demonstrate the effectiveness of their approach and highlight the importance of considering spatial patterns in predicting forest fires.

Gao X et al.,[2] Forest fire detection based on deep learning and multiple features. In this paper, the authors propose a forest fire detection system based on deep learning and multiple features. The system uses convolutional neural networks (CNN) to extract features from images and then employs support vector machines (SVM) to classify the images. The results show that the proposed system outperforms other methods in terms of accuracy and efficiency.

Javed S et al.,[3] Forest Fire Prediction Using Machine Learning and Internet of Things. This paper proposes a forest fire prediction system using machine learning and internet of things (IoT) technologies. The system collects real-time environmental data using IoT sensors and feeds it into machine learning algorithms for prediction. The authors demonstrate the effectiveness of their approach through experimental results.

Karimi B et al., [4] Prediction of forest fires using machine learning algorithms". This paper presents a study on the prediction of forest fires using machine learning algorithms. The authors use meteorological variables, such as temperature, humidity, wind speed, and rainfall, to train different machine learning models, including decision tree, random forest, and support vector machines. The results indicate that the random forest model outperforms other models in terms of prediction accuracy.

Lim C et al.,[5] Predicting forest fire occurrences using data mining techniques: a case study in South Korea," applied random forest and other machine learning techniques to predict forest fires in South Korea. The authors found that random forest produced the most accurate predictions, and identified the most important predictor variables.

Mohamed F et al.,[6]Detection of Hydrogen Leakage Using Different Machine Learning Techniques: pure hydrogen, its leakage poses a serious safety risk since it can cause fire or explode if it comes into contact with the air. In this study, hydrogen leakage in a form of a buoyant jet is investigated using Random Forest Approach.

Neri A et al., [7] Fire risk forecasting using machine learning models based on meteorological data In this paper, the authors propose a fire risk forecasting model based on machine learning algorithms and meteorological data. They apply their model to different forest areas in Italy and find it to be effective in predicting the risk of forest fires.

Singh A et al., [8] explored the use of random forest to predict forest fires in India. The authors compared the performance of random forest with other machine learning algorithms and found that random forest produced the highest accuracy. They also identified the most important predictor variables for forest fires.

Salis M et al., [9] proposed a model for predicting the spread and behavior of wildfires in Mediterranean landscapes. Their study focused on developing a system that could predict the likelihood of wildfire ignition and spread in Mediterranean areas using a combination of topographical and meteorological data. The authors used a machine learning-based approach to develop their model and tested it in different Mediterranean areas.

Tymstra D et al., [10] used random forest to predict the number of forest fires in Ontario, Canada. The authors found that random forest outperformed other statistical methods, and identified the most important predictors for forest fires, including weather variables and human activity.

Chapter 3

PROJECT DESCRIPTION

3.1 Existing System

Smoke alarms and heat alarms are being used to detect fires. One module is not enough to monitor all of the potential hot spots for fires, which is the fundamental drawback of smoke sensor alarms and heat sensor alarms. The only way to avoid a fire is to exercise constant caution. Even if they are deployed in every nook and cranny, it still won't be enough to constantly produce an efficient output. The price will rise by a multiple as the number of smoke sensors required rises. Within seconds of an accident or fire, the suggested method can provide reliable and extremely accurate alarms. One piece of software powers the whole monitoring network, which lowers costs. Data scientists and machine learning experts are actively conducting research in this area. The key problem is reducing inaccuracy in fire detection and issuing notifications at the appropriate time.

3.2 Proposed System

We attempted to anticipate the burnt area in the Forest Area in this study. This investigation made use of the Forest Fires Data Collection. The information was clustered. Methods of stepwise regression were used to choose the top predictor. It will be interesting to observe which one affects the burnt area in each cluster the most. Starting with the first recorded fire perimeter and creating predictions for all gathered perimeters for the next day, Fire Cast gets the identical input variables as the training fires for the testing fire. Fire Cast selects a sample of POIs at random from the area surrounding the active fire and assigns a prediction value p, such that p [0, 1]. Each pixel shows the possibility that, given the input factors, the pixel will be confined inside the ensuing fire perimeter. Higher numbers indicate a higher expected probability of burn. the model's training data set. Moreover, it makes it simple to integrate Deep Learning models because of its capacity to produce step-by-step photos of a fire's course. A straightforward model called the Fire Assessment Learner assumes that all nearby trees have an equal chance of catching fire.

3.3 Feasibility Study

An analysis that considers all of the important aspects of a project is known as a feasibility study. The goal of the feasibility study is to logically and objectively identify the advantages and disadvantages of the current project, the opportunities and threats presented by the earth, the resources needed to get there, and finally the chances for advancement. The following three factors are crucial to the feasibility analysis:

1. Economic Feasibility
2. Technical Feasibility
3. Social Feasibility .

3.3.1 Economic Feasibility

Economic feasibility is a type of cost-benefit analysis that determines if the project under consideration can be implemented. By objectively and logically evaluating a project's ability to help the decision-making process, its strengths, weaknesses, opportunities, and related risks are assessed and analysed. The economic viability of forest fire detection is cost effective.

3.3.2 Technical Feasibility

The technological resources at hand are the focus of technical feasibility. This project utilises Anaconda navigator to run and identify the fire and requires numerous datasets that demonstrate the fire detection by taking into account various variables.

3.3.3 Social Feasibility

A thorough examination of interpersonal interactions inside a system or organisation is known as social feasibility. To comprehend the scope and influence of the project, social impact analysis is a process that identifies and examines such affects. Forest In order to prevent forest fires in the future, fire detection is utilised.

The act is legitimate. Yet, everyone should adhere to socially acceptable data collection practises as a responsible person.

3.4 System Specification

3.4.1 Hardware Specification

- Processor: Intel i3.
- Disk space: 160 GB.
- Ram: 4 GB.
- Internet Connection.

3.4.2 Software Specification

- Windows 7 or Windows or higher versions of OS.
- Any Latest Web Browser. (Preferably, Mozilla Firefox, Google chrome).
- Python for User Interface.
- web based google colab.
- Anaconda Navigator.

3.4.3 Standards and Policies

- IEEE 829 - Software Test Documentation.
- IEEE 830 - Software Requirements Specifications.
- IEEE 1012 - Software verification and validation.
- IEEE 1016 - Software design description.

Chapter 4

METHODOLOGY

4.1 General Architecture

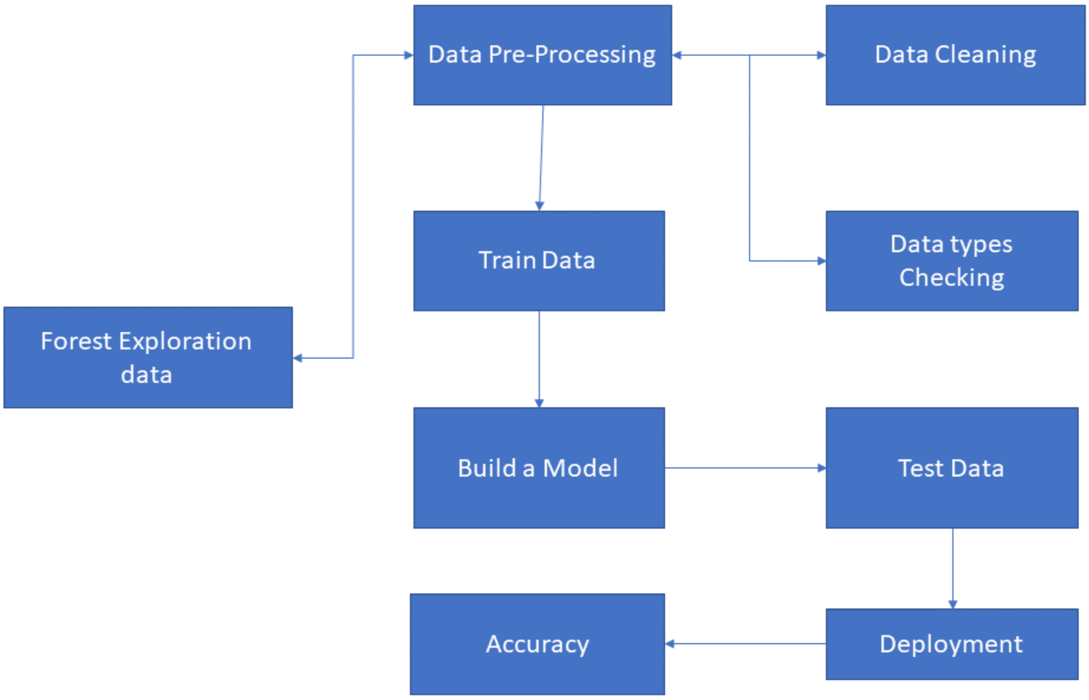
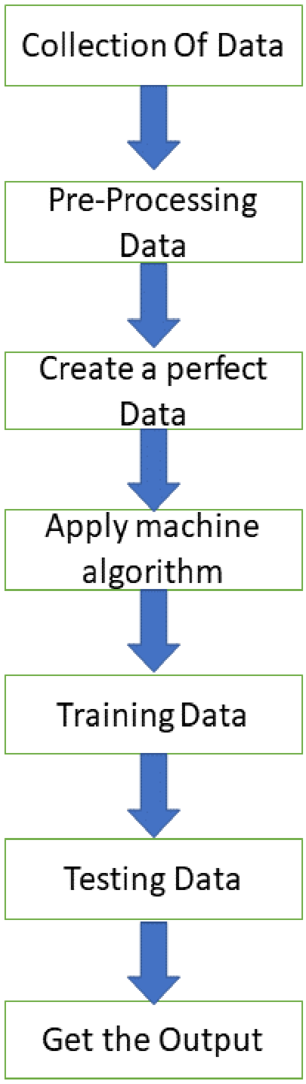


Figure 4.1: **Architecture Diagram for Prediction of Forest fire**

In Figure 4.1 The information gathered during forest exploration is pre-processed for improved analysis during the prediction of forest fire. Data cleansing and data type validation are part of this pre-processing. The data will then be trained, and the model will be created using the gathered data, following the Pre-Processing. The data will then be evaluated before being deployed. The data will be sent during the deployment phase to a location where an action may be carried out on it. Then accuracy is determined, which is the fundamental idea of prediction.

4.2 Design Phase

4.2.1 Data Flow Diagram



.

Figure 4.2: **Data Flow Diagram for Prediction of Forest fire**

In Figure 4.2 shows It states that there are 7 phases involved in the prediction of forest fires. Data will first be gathered, and then it will go through pre-processing. In order to produce flawless data, all the acquired information will be equally segregated in this stage. The data will then be subjected to the Machine Learning Algorithm after that. The data will then be trained, and when the trained data has been tested, output will be produced.

4.2.2 Use Case Diagram

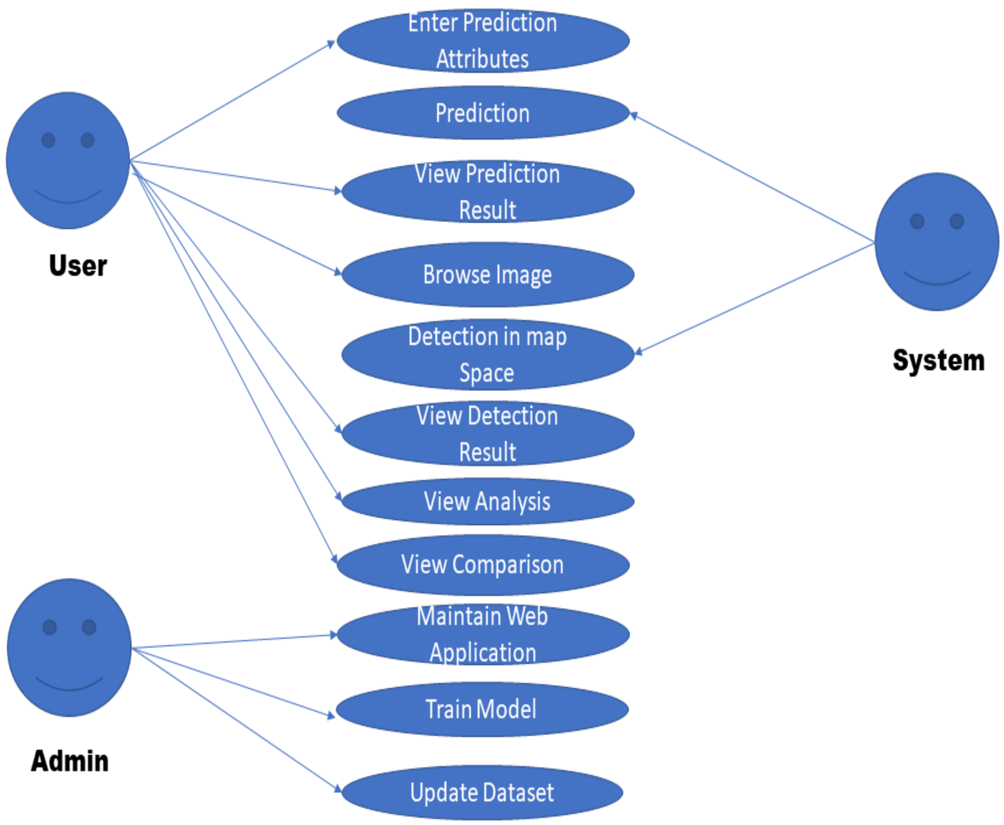


Figure 4.3: **Use Case Diagram for Prediction of Forest fire**

In Figure 4.3 shows how user inputs and outputs interact with one another. the approaches and modules involved in achieving better results and making judgements. Inevitably, the internal phases in the process that take place between input and output are displayed. The input data goes through pre-processing and processing processes that are necessary for filtering and cropping for a later analysis that aids in identifying normal and abnormal features.

4.2.3 Class Diagram

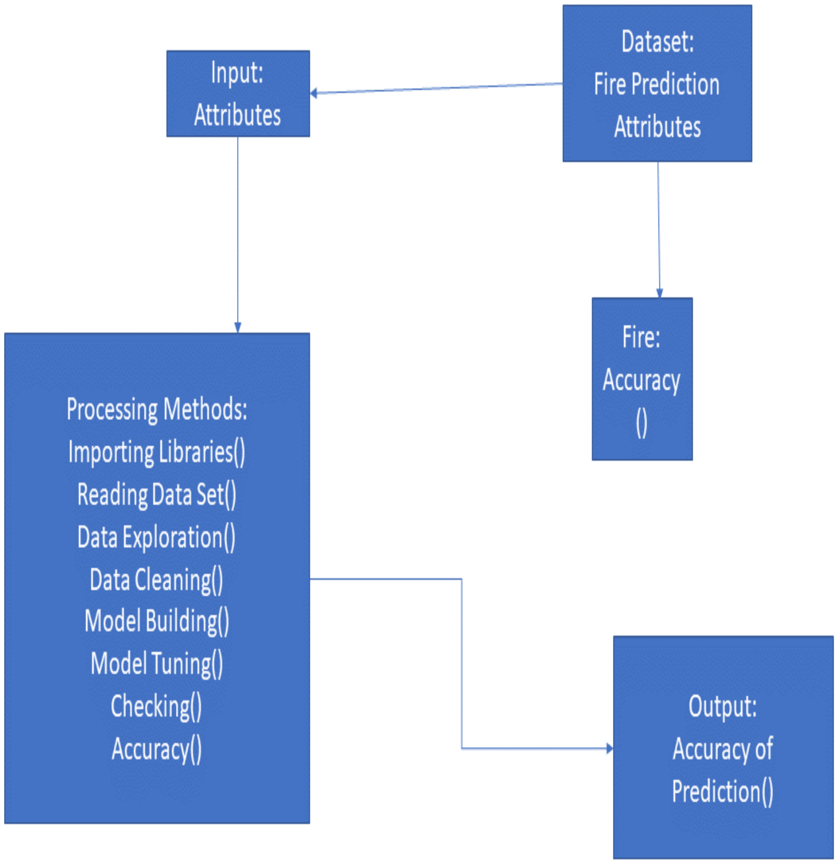


Figure 4.4: **Class Diagram for Prediction of Forest fire**

In Figure 4.4 The static structural representation of the project or technique of problem-solving that conveys the classes, sub-classes with suitable characteristics, and methods or functions assuming the relationship among the objects mentioned. It defines the architecture by doing so. The segmentation techniques for the given blocks relate to them. The class diagram, which serves as the conceptual model for the project, is helpful in organising the application or project in detail and in turning the class model into programmable code. The classes listed show the fundamental elements and connections between them.

4.2.4 Sequence Diagram

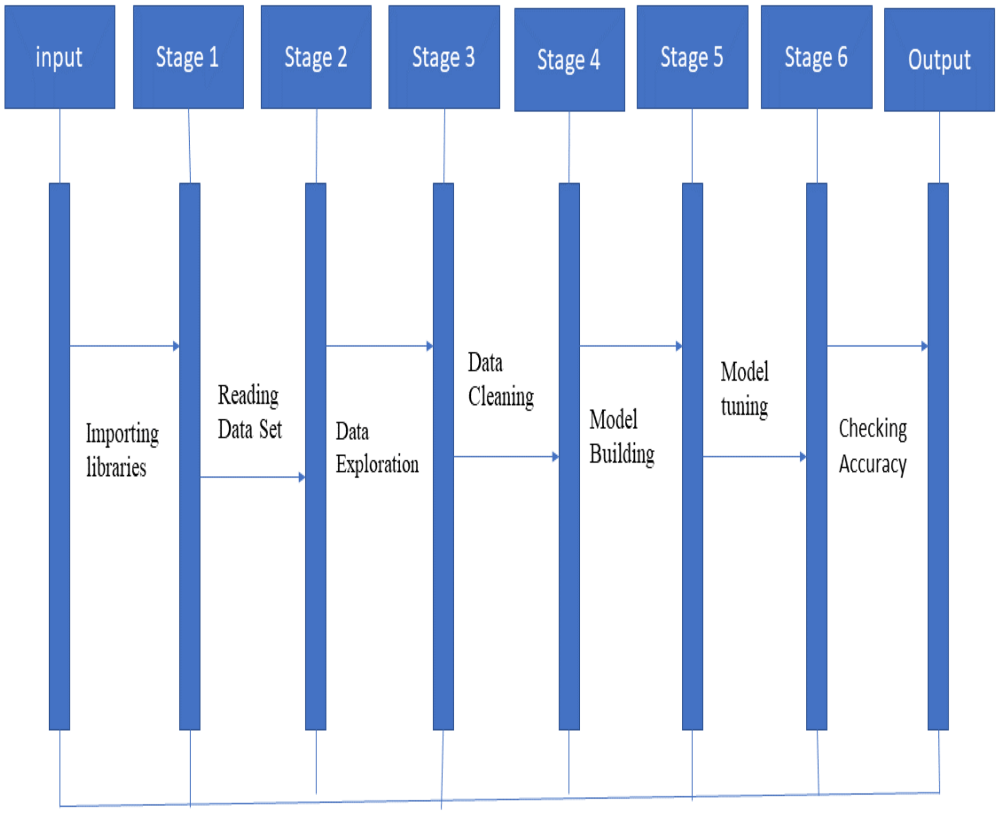


Figure 4.5: **Sequence Diagram for Prediction of Forest fire**

In Figure 4.5 The modules based on the time sequence is depicted in the project's sequence diagram. The data translation between each job or sequence in this situation is spread in real time. The tasks that must be accomplished in each module are laid down in a time order and are then traded with time management. On the basis of the identification of each situation, a logical viewpoint is developed. The well-organized horizontal and vertical flow of the figure demonstrates the project's visible segmentation process.

4.2.5 Collaboration diagram

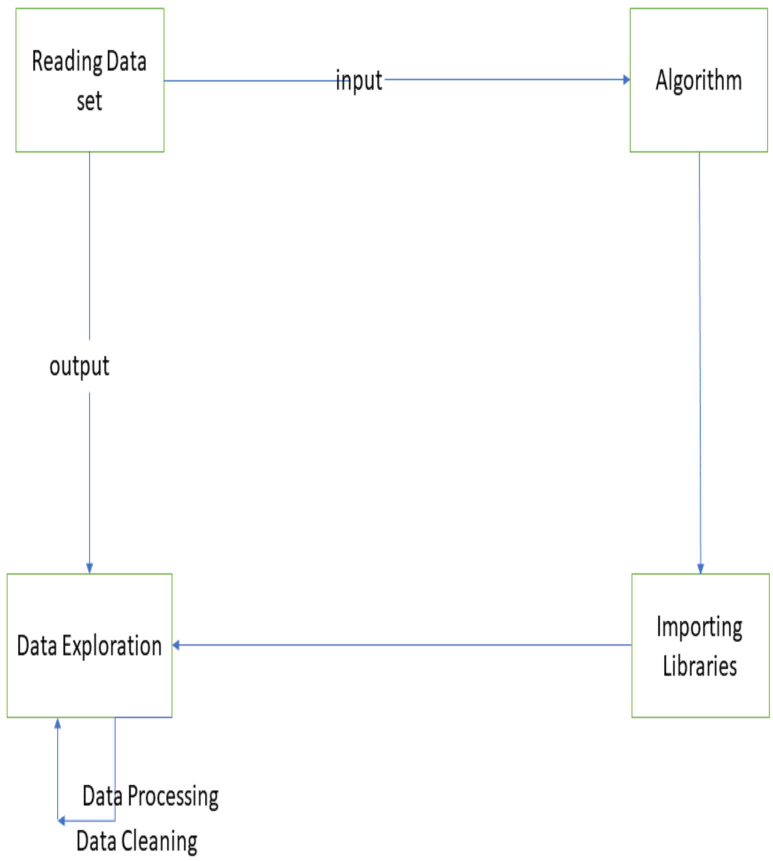


Figure 4.6: **Collaboration diagram for Prediction of Forest fire**

In Figure 4.6 shows how the classes are related. Collaboration illustrates the logic flow for easier comprehension as opposed to the architectural structure, just as the methodology's architecture does. The thing has a lot of features. It is easier to convey the suggested notion when these many elements are connected to the appropriate characteristic level of defining the system. Diagram of communication is another name for this partnership. This aids in illustrating the forest fire prediction system and machine learning techniques.

4.2.6 Activity Diagram

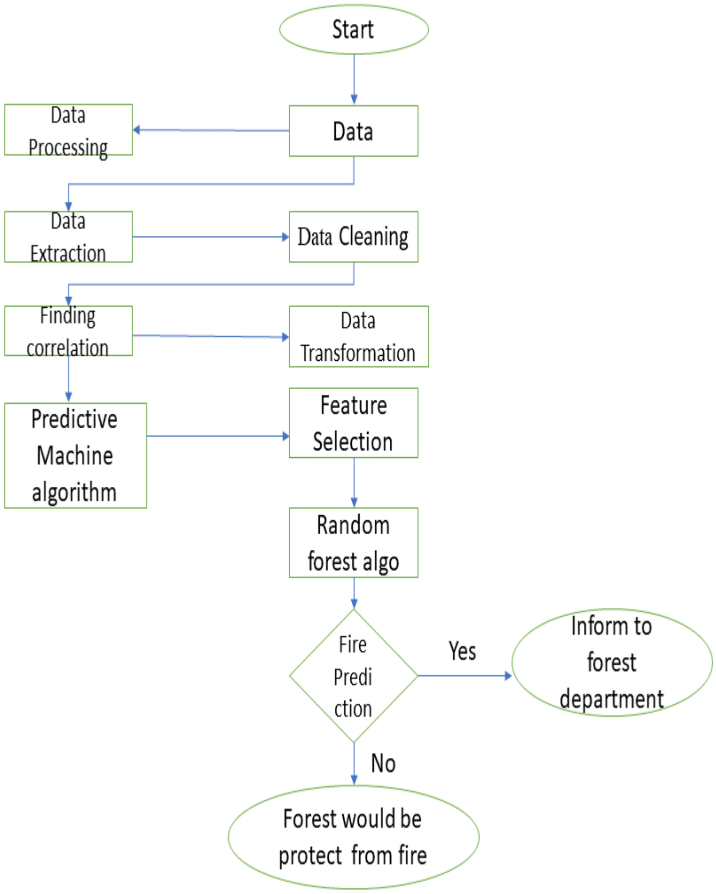


Figure 4.7: **Activity Diagram for Prediction of Forest fire**

In Figure 4.7 flowchart of the activities and represents the workflow among the multiple activities. It shows the actual process and flow of the different activities and their mechanism in the system. The execution flow from the start of the system to the end of the system is clearly depicted. The activity nodes helps in differentiating each activity from each other.

4.3 Algorithm & Pseudo Code

4.3.1 Random Forest Algorithm

Random Forest is a well-known machine learning algorithm from the supervised learning technique. It may be applied to both classification and regression issues in machine learning. It is built on the notion of ensemble learning, which is a method that involves integrating several classifiers to solve a complicated issue and enhance the model's performance. ”Random Forest is a game,” as the name implies.A classifier that uses a number of decision trees on different subsets of a given dataset and averages their results to enhance the predicted accuracy of that dataset. Rather of depending on a single decision tree,the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

4.4 Module Description

4.4.1 Data Collection and Pre processing

Gather information on historical forest fires, including their locations and any variables that may have influenced their occurrence (such as weather, human activity) Cleanse the data, examine it for outliers or missing numbers, and then convert it into a format that machine learning algorithms can utilise.

4.4.2 Feature Selection

Do a data analysis to pinpoint the key elements that are most likely to have an impact on the occurrence of forest fires. Choose a portion of these features to include in the model, being careful to avoid having them substantially associated with one another.

4.4.3 Model Development

Separate the data into training and testing sets and Use the training data to train the random forest algorithm. to Assess the model's performance on the testing set and adjust the algorithm's parameters as necessary.

4.4.4 Model Deployment

To create predictions on new data, use the trained random forest algorithm. and Install the model in a user-friendly interface (e.g., a web app) so that users may enter information about their location and other relevant factors and obtain a prediction about their probability of being affected by a forest fire.

4.4.5 Model Monitoring and Maintenance

Monitor the model's performance over time to ensure that it continues to make correct predictions. Retrain the model on new data on a regular basis and tweak the feature selection and model parameters as needed to increase accuracy.

4.5 Steps to execute/run/implement the project

4.5.1 Importing Libraries

In this step we import/install some required libraries.

4.5.2 Data Exploration

In this data exploration, we go through the dataset to find specific patterns which helps to understand insights and implement new policies. It checks for the null values in the forest fire prediction dataset.

4.5.3 Data Cleaning

In this process, we prepare the data for analysis by removing or modifying the data that is incomplete, incorrect, irrelevant, duplicated, or improperly formatted.

4.5.4 Model Modelling

In Machine Learning, the model is built by learning and generalizing from the data that is trained, and then applying the acquired knowledge to new data it has never seen before to make predictions and fulfill its purpose. Lack of data will prevent us from building the model, and access to data is not enough.

4.5.5 Model Tuning

Model Tuning is also known as hyperparameter optimization.The hyperparameters are variables that control the training process. These are configuration variables that do not change during a Model training job. Model tuning provides optimized values for hyperparameters, which maximize the model's predictive accuracy.

4.5.6 Checking Accuracy

Machine Learning model accuracy is the measurement used to determine which model is best at identifying relationships and patterns between variables in a dataset based on the input, or training of the data. Here the accuracy of the Trained data is 95.55 percentage and for the Tested data, the accuracy is 68.02 percentage.

Chapter 5

IMPLEMENTATION AND TESTING

5.1 Input and Output

5.1.1 Input Design

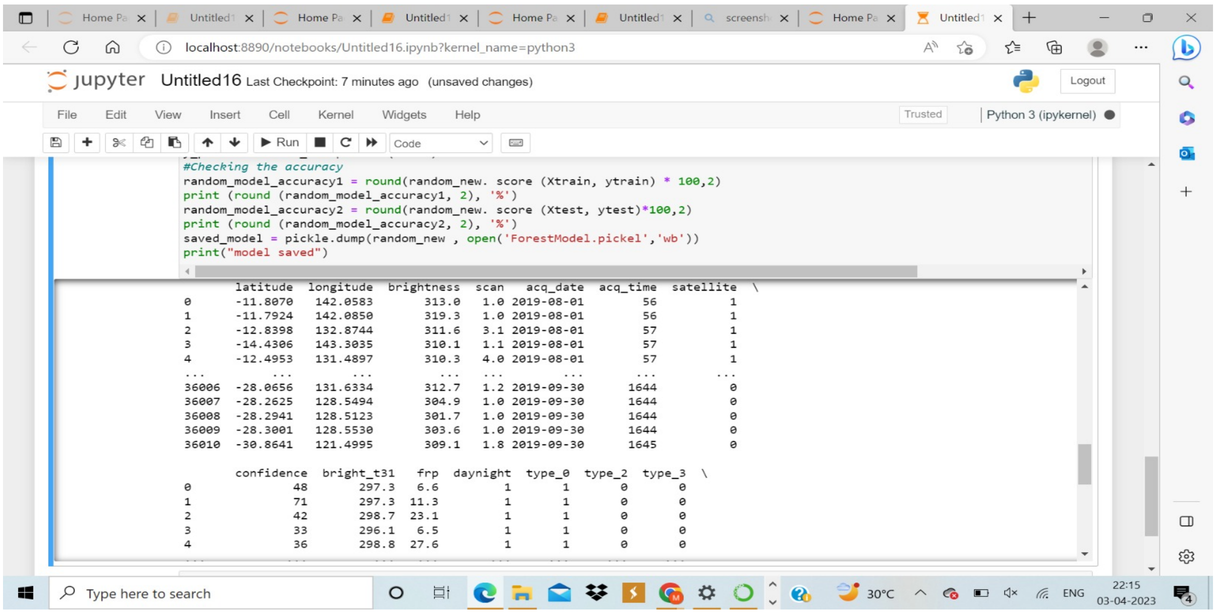


Figure 5.1: **Trained Data of Forest Samples**

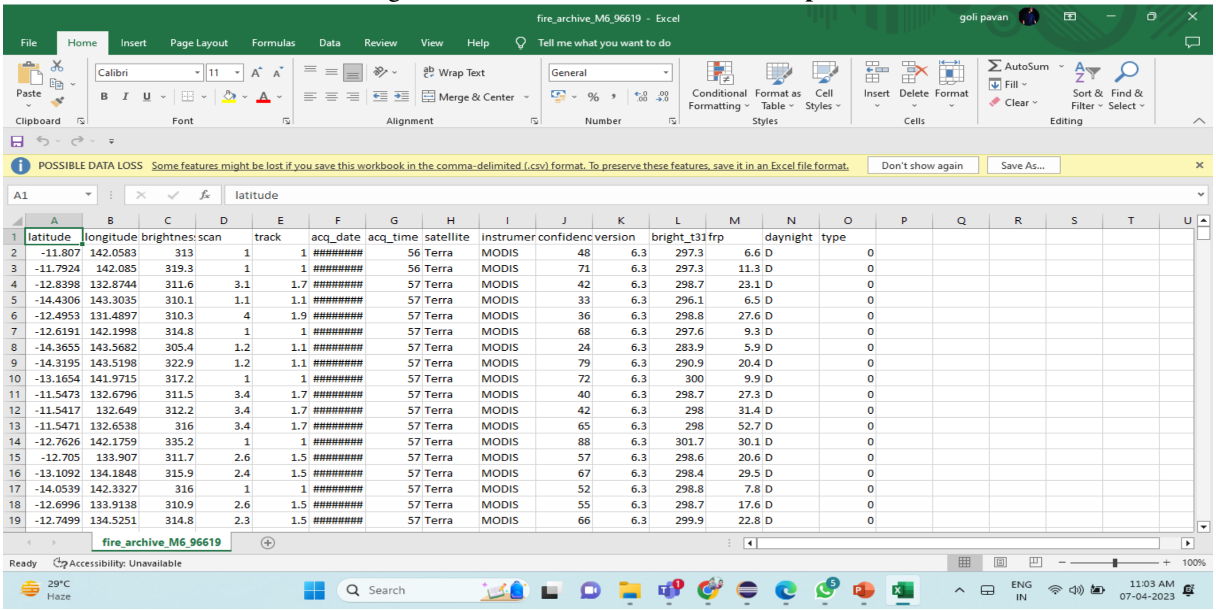


Figure 5.2: **Raw Data of forest Samples**

5.1.2 Output Design

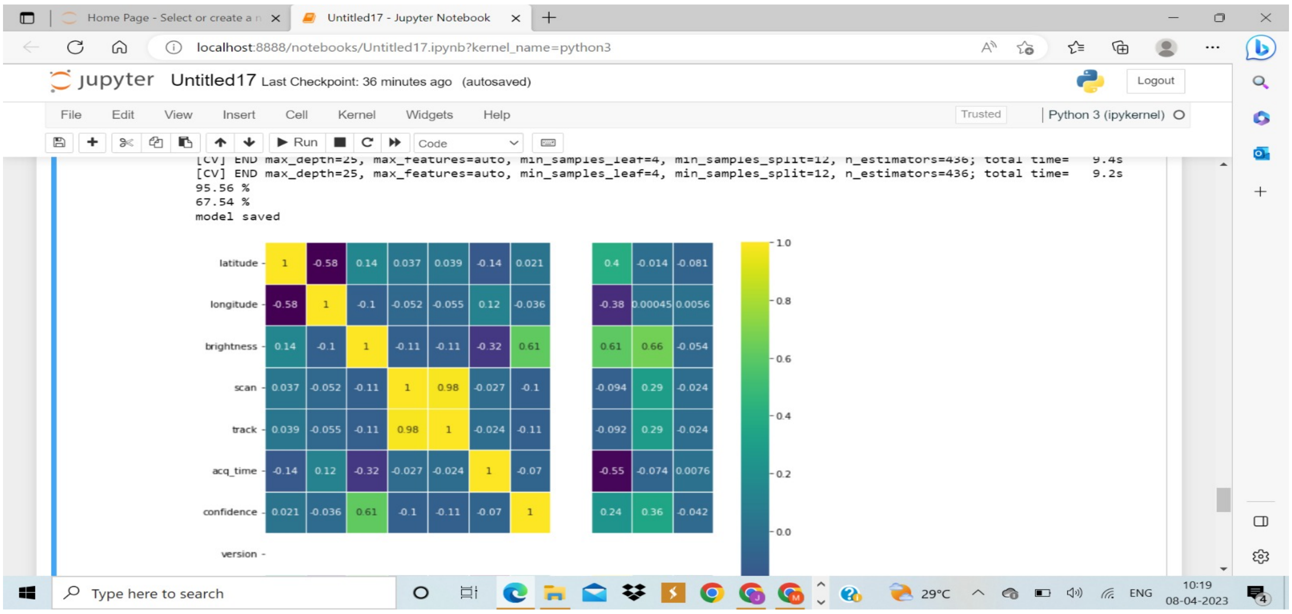


Figure 5.3: **Output of prediction and accuracy**

5.2 Testing

Data generated or chosen to fulfil the execution preconditions and inputs required to run one or more test cases.Security testing, performance testing, and regression testing are all receiving a lot of attention.

5.3 Types of Testing

5.3.1 Unit Testing

Unit testing is a type of software testing in which individual programme components or units are tested. The primary goal of unit testing is to examine the performance of each individual unit in software. The data components such as latitudes, longitudes, climate, temperature, light, and segmentation algorithms are evaluated and completed in this project strategy to ensure that no such random unsuitable values are given.

5.3.2 Integration Testing

White box testing is a software test approach in which the internal schema of the item is known to the tester.The tester pick out the input attributes that brought up more in case of predicting the data that is incomplete, incorrect, irrelevant, duplicated, or improperly formatted which creates abnormalities to practicing paths through the code and concludes the predicted rate accurately.

5.3.3 System Testing

The logical and syntax errors have been pointed out by program testing. A syntax error is an error in a program statement that in violates one or more rules of the language in which it is written. An improperly defined field dimension or omitted keywords are common syntax error. These errors are shown through error messages generated by the computer. A logic error on the other hand deals with the incorrect data fields, out-off-range items and invalid combinations. Since the compiler will not deduct logical error, the programmer must examine the output. Condition testing exercises the logical conditions contained in a module. The possible types of elements in a condition include a Boolean operator, Boolean variable, a pair of Boolean parentheses A relational operator or on arithmetic expression. Condition testing method focuses on testing each condition in the program the purpose of condition test is to deduct not only errors in the condition of a program but also other a errors in the program.

5.3.4 Test Result

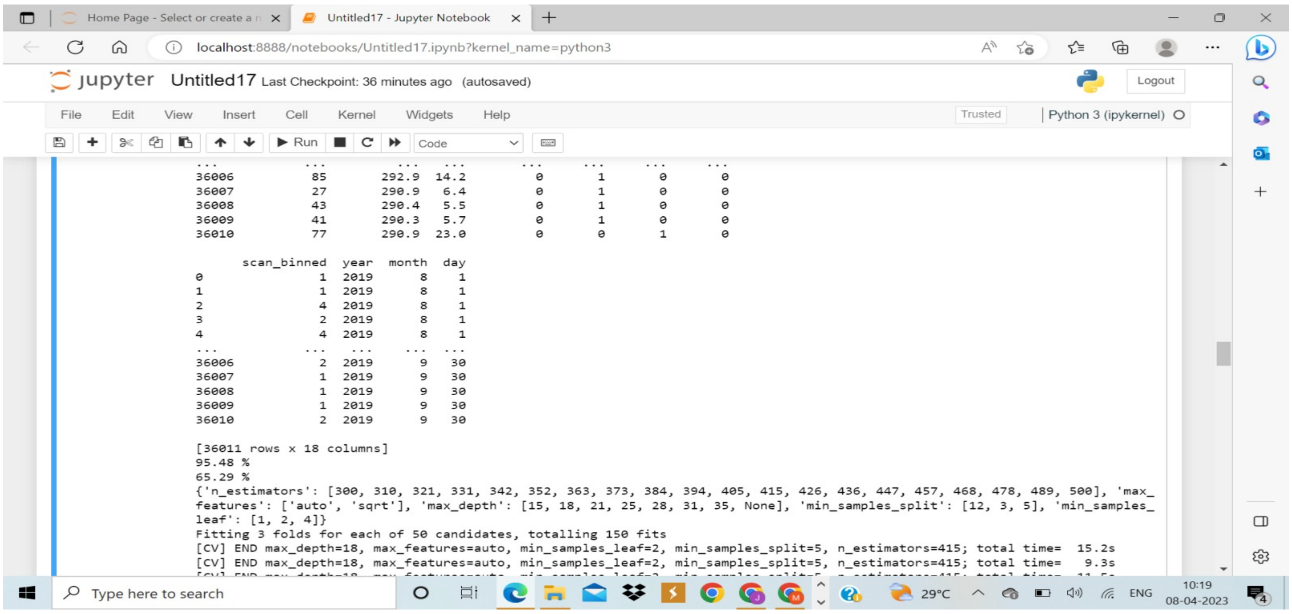


Figure 5.4: **Data processing samples**

Chapter 6

RESULTS AND DISCUSSIONS

6.1 Efficiency of the Proposed System

The proposed system is based on the Random forest Algorithm that creates many decision trees. Accuracy of proposed system is done by using random forest gives the ouput approximately 76 to 78 percent. Random forest implements many decision trees and also gives the most accurate output when compared to the decision tree. Random Forest algorithm is used in the two phases. Firstly, the RF algorithm extracts subsamples from the original samples by using the bootstrap resampling method and creates the decision trees for each testing sample and then the algorithm classifies the decision trees and implements a vote with the help of the largest vote of the classification as a final result of the classification. The random Forest algorithm always includes some of the steps as follows: Selecting the training dataset:Using the bootstrap random sampling method we can derive the K training sets from the original dataset properties using the size of all training set the same as that of original training dataset. Building the random forest algorithm: Creating a classification regression tree each of the bootstrap training set will generate the K decision trees to form a random forest model, uses the trees that are not pruned. Looking at the growth of the tree, this approach is not chosen the best feature as the internal nodes for the branches but rather the branching process is a random selection of all the trees gives the best features.

6.2 Comparison of Existing and Proposed System

Existing system:

In the Existing system, we implemented a decision tree algorithm that predicts whether to grant the loan or not. When using a decision tree model, it gives the training dataset the accuracy keeps improving with splits. We can easily overfit the dataset and doesn't know when it crossed the line unless we are using the cross validation. The advantages of the decision tree are model is very easy to interpret we can know that the variables and the value of the variable is used to split the data. But the accuracy of decision tree in existing system gives less accurate output that is less when compared to proposed system.

Proposed system:

Random forest algorithm generates more trees when compared to the decision tree and other algorithms. We can specify the number of trees we want in the forest and also we also can specify maximum of features to be used in the each of the tree. But, we cannot control the randomness of the forest in which the feature is a part of the algorithm. Accuracy keeps increasing as we increase the number of trees but it becomes static at one certain point. Unlike the decision tree it won't create more biased and decreases variance. Proposed system is implemented using the Random forest algorithm so that the accuracy is more when compared to the existing system.

6.3 Sample Code

```
import datetime as dt
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, classification_report
from sklearn.ensemble import RandomForestRegressor
forest = pd.read_csv('forestfires.csv')
forest.head ()
forest.shape
forest.columns
forest.isnull().sum ()
forest.describe ()
plt.figure(figsize=(10, 10))
sns.heatmap(forest.corr(),annot=True, cmap='viridis', linewidths =.5)
forest=forest.drop(['track'], axis = 1)
print("The scan column")
print (forest ['scan']. value_counts ())
print ()
print("The aqc time column")
print (forest ['acq_time']. value_counts())
print ()
print("The satellite column")
print (forest['satellite']. value_counts ())
print ()
print ("The instrument column")
```

```
print (forest['instrument']. value_counts ())
print ()
print ("The version column")
print (forest ['version']. value_counts())
print ()
print("The daynight column")
print (forest['daynight']. value_counts ())
print ()
forest = forest.drop(['instrument', 'version'], axis = 1)
forest.head ()
daynight_map = {"D": 1, "N": 0}
satellite_map = {"Terra": 1, "Aqua": 0}
forest['daynight'] = forest ['daynight']. map (daynight_map)
forest['satellite'] = forest ['satellite']. map(satellite_map)
forest. head ()
forest['type']. value_counts ()
types = pd.get_dummies (forest['type'])
forest = pd.concat([ forest, types], axis=1)
forest = forest.drop(['type'], axis = 1)
forest.head ()
forest = forest.rename(columns = {0: 'type_0', 2: 'type_2', 3: 'type_3'})
bins = [0, 1, 2, 3, 4, 5]
labels = [1,2,3,4,5]
forest ['scan_binned'] = pd. cut (forest['scan'], bins = bins, labels = labels)
forest.head ()
# Converting the datatype to date type from string or numpy.
forest['acq_date'] = pd.to_datetime (forest ['acq_date'])
forest['year'] = forest ['acq_date']. dt. year
forest.head()
forest['month'] = forest ['acq_date'].dt. month
forest ['day'] = forest['acq_date'].dt. day
forest.shape
y = forest['confidence']
print(forest)
fin = forest.drop(['confidence', 'acq_date', 'acq_time', 'bright_t31', 'type-0'], axis =1)
plt.figure(figsize=(10, 10))
sns.heatmap(fin.corr(), annot=True, cmap='viridis', linewidths =.5)
fin. head ()
Xtrain, Xtest, ytrain, ytest = train_test_split (fin.iloc[:, :500], y, test_size =0.2)
random_model = RandomForestRegressor (n_estimators =300, random_state= 42, n_jobs = -1)
random_model. fit(Xtrain, ytrain)
y_pred = random_model.predict(Xtest)
#Checking the accuracy
random_model_accuracy = round(random_model.score(Xtrain, ytrain)*100,2)
print (round (random_model_accuracy, 2), '%')
random_model_accuracy1= round(random_model.score(Xtest, ytest)*100,2)
print (round (random_model_accuracy1, 2), '%')
import pickle
saved_model = pickle.dump(random_model, open('ForestModelOld. pickle', 'wb'))
random_model.get_params()
```

```
from sklearn. model_selection import RandomizedSearchCV
n_estimators= [int(x) for x in np. linspace (start = 300, stop= 500, num = 20)]
# Number of features to consider at every split
max_features = ['auto' , 'sqrt']
# Maximum number of levels in tree
max_depth= [int(x) for x in np. linspace (15, 35, num = 7)]
max_depth.append(None)
# Minimum number of samples required to split a node
min_samples_split = [12, 3, 5]
# Minimum number of samples required at each leaf node
min_samples_leaf = [1, 2, 4]
# Create the random grid
random_grid = {'n_estimators' : n_estimators ,
                'max_features' : max_features,
                 'max_depth'    :max_depth,
                 'min_samples_split' :min_samples_split,
                 'min_samples_leaf'  :min_samples_leaf,
                }
print(random_grid)
rf_random = RandomizedSearchCV(estimator = random_model ,param_distributions = random_grid,
n_iter =50,cv=3,verbose =2 ,random_state=42)
#Fit the random search model
rf_random.fit(Xtrain,ytrain)
rf_random.best_params_
random_new = RandomForestRegressor (n_estimators = 394, min_samples_split = 2, min_samples_leaf=1,
max_features = 'sqrt',max_depth = 25, bootstrap =True)
random_new.fit (Xtrain, ytrain)
y_pred1 = random_new.predict(Xtest)
#Checking the accuracy
random_model_accuracy1 = round(random_new. score (Xtrain, ytrain) * 100,2)
print (round (random_model_accuracy1, 2), '%')
random_model_accuracy2 = round(random_new. score (Xtest, ytest)*100,2)
print (round (random_model_accuracy2, 2), '%')
saved_model = pickel.dump(random_new , open('ForestModel.pickel','wb'))
print("model saved")
```

Output

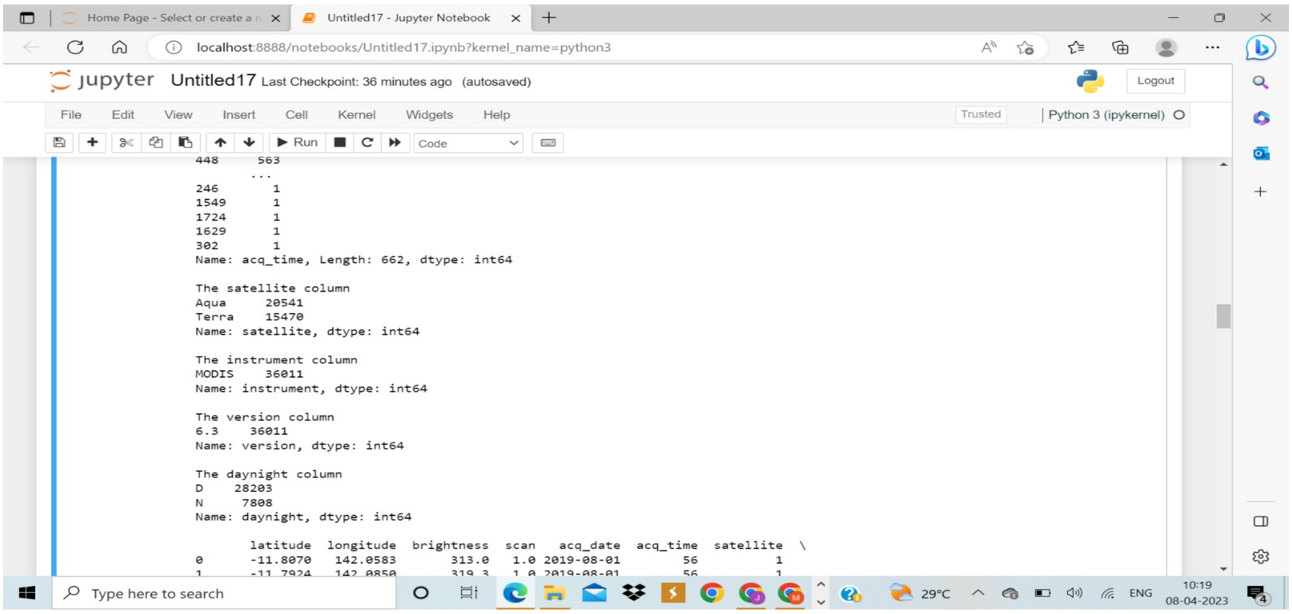


Figure 6.1: **Processing Data**

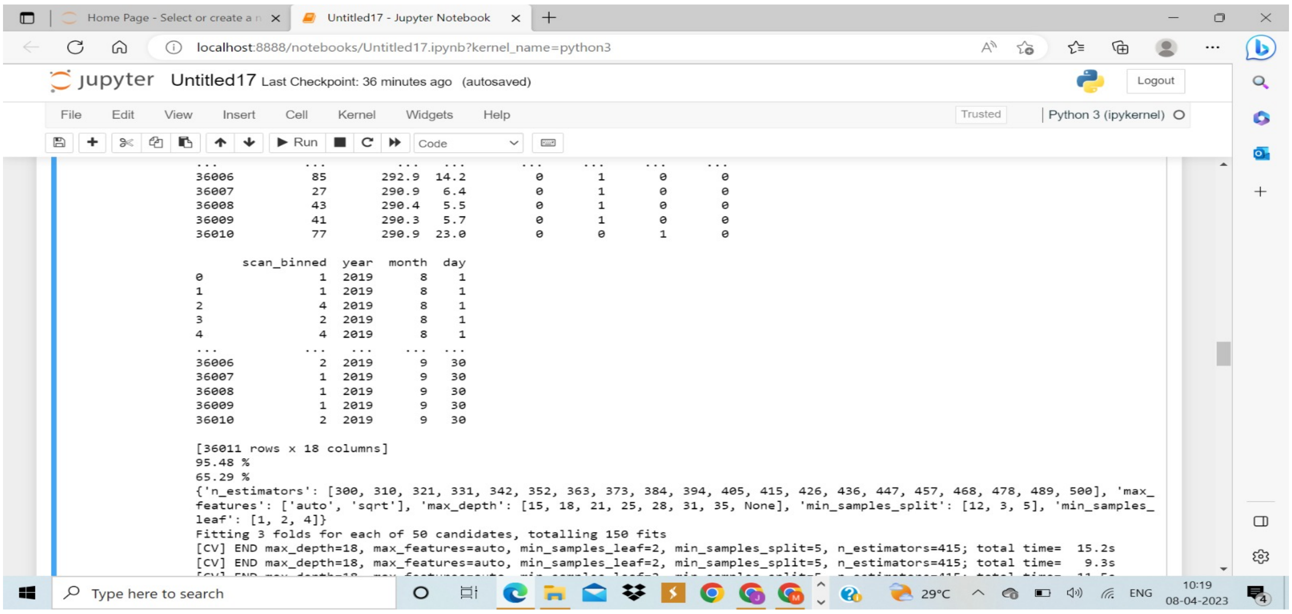


Figure 6.2: **Predicted output**

Chapter 7

CONCLUSION AND FUTURE ENHANCEMENTS

7.1 Conclusion

Machine learning's creative and difficult approach to fire prediction. It is conceivable to employ surveillance systems to stop loss and damage brought on by random fire accidents if this system with a low mistake rate can be applied on a broad scale, such as in major companies, homes, or woods. The algorithm shows considerable potential in terms of environment adaptation. The experimental findings demonstrate that, even with an unstable measure, the suggested technique still yields high and steady outcomes. Throughout the flora and wildlife, forest fires are a common occurrence. The globe loses millions of hectares of forest each year.Due to the priceless human lives that were lost, this also caused enormous environmental devastation. The threat of forest preservation is posed by forest fires, which also impair the economy, the ecology, and people's quality of life. Less fire damage can be achieved by quick fire detection and response. To enhance early fire prediction, several investigations are carried out. Our study was done, and we developed a system that uses user-provided meteorological information on temperature, oxygen content, and humidity to anticipate the proportions of fires that may occur. The amount of flames that might occur is properly predicted by the technique.

7.2 Future Enhancements

Forest fire detection is an ongoing project that may always be improved and upgraded by the addition of new features. Analyses of structure and function are giving new information on fire prediction. The new features that will be included in the future aid in not only fire prediction but also fire prevention. In the future, we may have an extra alerting mechanism that, by simply uploading a station's weather data, might be utilised to notify a certain department. By doing this, we can aid in mitigating forest fires before they completely damage the forest. Because forest fires can be easily predicted, prevention is much simpler. We intend to enhance the accuracy and speed of our model in the future. Moreover, we would like to provide forecasts that are real-time, local, and based on satellite imagery.

Chapter 8

INDUSTRY DETAILS

8.1 Industry Name

Boston IT solutions india pvt limited.

Duration of Internship : Feb 18 - May 20

Duration of Internship in months: 3 months

Industry Address: No. 64, Ground Floor, Railway Parallel Road, Kumarapark West, Bengaluru, 560020. India

8.2 Internship offer letter

BOSTON
TRAINING
ACADEMY

Date: 16/2/23

INTERNSHIP LETTER

Dear Student,

Congratulations! we are happy to inform you that you are selected for Internship. Please treat this email as our official acceptance for internship at Boston IT Solution PVT LTD with BTA team. Kindly reply on the same email threat for your acceptance.

Please follow below formalities :
Start date: 18/02/23
End date: 20/05/23

Formalities :

- Candidates needs to submit a bonified copy from the college where it states that the candidates is a students of the University
- Candidates are expected to wear professional attire whenever visiting the office.
- Candidates are expected to wear face mask in the office premises and follow covid guideline in the office.
- Candidates need to submit their latest resume and Aadhar card as address proof on the first day of internship
- The candidate needs to visit office and present their work to line manager or engineer team as per the schedule given by the BTA team.

Thanks
Laxmi Nageswari
Laxmi Nageswari
Global Head AI Education & Solutions
Boston IT Solutions India Pvt. Ltd.

ADDRESS
No 64, Ground Floor,Railway Parallel Road, Kumara Park West, Bengaluru, 560020, India

PHONE
+91 80 4308 4000

EMAIL
sales@bostonindia.in

WEB
www.bostonindia.in

8.3 Project Commencement Form

Vel Tech
Rangarajan Dr. Sagunthala
R&D Institute of Science and Technology
(Deemed to be University Estd. u/s 3 of UGC Act, 1956)

Project Commencement Form

Name of the Industry: Boston IT solutions india pvt limited

Address :Bangalore

Team Details:

S.No	ID No	Student Name	Degree & Branch
1.	14400	Goli Pavan	BTech &CSE
2.	12658	Gali Krishna Chaitanya	
3.	13952	V Jagapathi Babu	

Date of reporting for project work: Feb 18

Name of the Industry Supervisor : Rethishwar

Department : IT

Designation : Junior Data Analyst

Contact Number : 6369358668

Email ID : rethishvaar.selvakumar@bostonindia.in

Name of the Internal Supervisor : Ms.R Vaishnavi

Contact No. : 8148552249

Email ID : vaishnavir@veltech.edu.in

Tentative Project Title / Project domain: Topic-leveraging to predict forest fires using machine Learning

Brief Project/task description: Forest Fire Prediction is a key component of forest fire control. This is a major environmental problem that creates ecological destruction in the form of a threatened landscape of natural resources that disrupts the stability of the ecosystem, increases the risk for other natural hazards, and decreases resources such as water that causes global warming and water pollution. Fire Detection is a key element for controlling such incidents. Prediction of forest fire id expected to reduce the impact of forest fire in the future. Many fire detection algorithms are available with different approach towards the detection of fire. In the existing work processes the fire affected region is predicted based on the satellite images. To predict the occurrences of a forest fire the proposed system processes using the meteorological parameters such as temperature, rain, wind and humidity were used. Random forest regression and Hyperparameter tuning using Random Forest algorithm that combines multiple decision trees to make a more accurate prediction

8.4 Internship Completion certificate

Chapter 9

PLAGIARISM REPORT

Check Plagiarism

PLAGIARISM SCAN REPORT

Date	April 08, 2023
Exclude URL:	NO

Unique Content	100	Word Count	566
Plagiarized Content	0	Records Found	0

CONTENT CHECKED FOR PLAGIARISM:

A crucial part of controlling forest fires is forest fire prediction. This is a serious environmental issue that results in ecological degradation in the form of a vulnerable ecosystem, raises the risk of climate change, and lowers water resources, increases the danger for other natural calamities, and contributes to water pollution and global warming. A crucial component for managing such occurences is fire prediction. Future effects of forest fire should be lessened thanks to forest fire prediction. There are several fire detection algorithms available with various fire detection approaches. The region impacted by the fire is
forecasted using satellite imagery in the current work procedures.
The suggested system procedures employed meteorological
variables including temperature, rain, wind, and humidity to anticipate when a forest fire will occur.
The spread of flames may now be predicted using a variety of technologies, including mathematical and physical models. In order to specify and forecast fire development in various places, these models rely on data collecting during forest fires, modelling, and lab experimentation.
In recent years, simulation tools have been employed to forecast forest fires, however these systems have encountered issues with data quality and tool execution time. A form of artificial intelligence called machine learning is used to teach computers.
Two categories of machine learning exist: supervised, unsupervised, and reinforcement learning. In supervised learning, a supervisor is there to inform the algorithm as to whether a choice or action is wise or not. With supervised learning, the entire collection of data is fully labelled. Linear

Chapter 10

SOURCE CODE & POSTER PRESENTATION

10.1 Source Code

```
import datetime as dt
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, classification_report
from sklearn.ensemble import RandomForestRegressor
forest = pd.read_csv('forestfires.csv')
forest.head ()
forest.shape
forest.columns
forest.isnull().sum ()
forest.describe ()
plt.figure(figsize=(10, 10))
sns.heatmap(forest.corr(),annot=True, cmap='viridis', linewidths =.5)
forest=forest.drop(['track'], axis = 1)
print("The scan column")
print (forest ['scan']. value_counts ())
print ()
print("The aqc time column")
print (forest ['acq_time']. value_counts())
print ()
print("The satellite column")
print (forest['satellite']. value_counts ())
print ()
print ("The instrument column")
print (forest['instrument']. value_counts ())
print ()
print ("The version column")
print (forest ['version']. value_counts())
print ()
print("The daynight column")
print (forest['daynight']. value_counts ())
print ()
```

```
forest = forest.drop(['instrument', 'version'], axis = 1)
forest.head ()
daynight_map = {"D": 1, "N": 0}
satellite_map = {"Terra": 1, "Aqua": 0}
forest['daynight'] = forest ['daynight']. map (daynight_map)
forest['satellite'] = forest ['satellite']. map(satellite_map)
forest. head ()
forest['type']. value_counts ()
types = pd.get_dummies (forest['type'])
forest = pd.concat([ forest, types], axis=1)
forest = forest.drop(['type'], axis = 1)
forest.head ()
forest = forest.rename(columns = {0: 'type_0', 2: 'type_2', 3: 'type_3'})
bins = [0, 1, 2, 3, 4, 5]
labels = [1,2,3,4,5]
forest ['scan_binned'] = pd. cut (forest['scan'], bins = bins, labels = labels)
forest.head ()
# Converting the datatype to date type from string or numpy.
forest['acq_date'] = pd.to_datetime (forest ['acq_date'])
forest['year'] = forest ['acq_date']. dt. year
forest.head()
forest['month'] = forest ['acq_date'].dt. month
forest ['day'] = forest['acq_date'].dt. day
forest.shape
y = forest['confidence']
print(forest)
fin = forest.drop(['confidence', 'acq_date', 'acq_time', 'bright_t31', 'type-0'], axis =1)
plt.figure(figsize =(10, 10))
sns.heatmap(fin.corr(), annot=True, cmap='viridis', linewidths =.5)
fin. head ()
Xtrain, Xtest, ytrain, ytest = train_test_split (fin.iloc[:, :500], y, test_size =0.2)
random_model = RandomForestRegressor (n_estimators =300, random_state= 42, n_jobs = -1)
random_model. fit(Xtrain, ytrain)
y_pred = random_model.predict(Xtest)
#Checking the accuracy
random_model_accuracy = round(random_model.score(Xtrain, ytrain)*100,2)
print (round (random_model_accuracy, 2), '%')
random_model_accuracy1= round(random_model.score(Xtest, ytest)*100,2)
print (round (random_model_accuracy1, 2), '%')
import pickle
saved_model = pickle.dump(random_model, open('ForestModelOld. pickle', 'wb'))
random_model.get_params()
from sklearn. model_selection import RandomizedSearchCV
n_estimators= [int(x) for x in np. linspace (start = 300, stop= 500, num = 20)]
# Number of features to consider at every split
max_features = ['auto' , 'sqrt']
# Maximum number of levels in tree
max_depth= [int(x) for x in np. linspace (15, 35, num = 7)]
max_depth.append(None)
# Minimum number of samples required to split a node
```

```
min_samples_split = [12, 3, 5]
# Minimum number of samples required at each leaf node
min_samples_leaf = [1, 2, 4]
# Create the random grid
random_grid = {'n_estimators' : n_estimators ,
                'max_features' : max_features,
                 'max_depth'    :max_depth,
                 'min_samples_split' :min_samples_split,
                 'min_samples_leaf'  :min_samples_leaf,
                }
print(random_grid)
rf_random = RandomizedSearchCV(estimator = random_model ,param_distributions = random_grid,
n_iter =50,cv=3,verbose =2 ,random_state=42)
#Fit the random search model
rf_random.fit(Xtrain,ytrain)
rf_random.best_params_
random_new = RandomForestRegressor (n_estimators = 394, min_samples_split = 2,
min_samples_leaf=1,max_features = 'sqrt',max_depth = 25, bootstrap =True)
random_new.fit (Xtrain, ytrain)
y_pred1 = random_new.predict(Xtest)
#Checking the accuracy
random_model_accuracy1 = round(random_new. score (Xtrain, ytrain) * 100,2)
print (round (random_model_accuracy1, 2), '%')
random_model_accuracy2 = round(random_new. score (Xtest, ytest)*100,2)
print (round (random_model_accuracy2, 2), '%')
saved_model = pickel.dump(random_new , open('ForestModel.pickel','wb'))
print("model saved")
```

10.2 Poster Presentation

Topic-leveraging to predict forest fires using machine Learning

Department of Computer Science & Engineering
School of Computing
1156CS701 – MAJOR PROJECT
WINTER SEMESTER 2022-2023

ABSTRACT

Forest fires have devastating effects on the environment, leading to ecological degradation, climate change, water resource depletion, and an increase in natural calamities. Fire prediction is a critical component of managing these occurrences. In recent years, simulation tools have been used to forecast forest fires, but they have encountered issues with data quality and tool execution time. Machine learning, a form of artificial intelligence, has been used to teach computers to predict forest fires. There are two categories of machine learning: supervised and unsupervised, and reinforcement learning. By using machine learning algorithms and meteorological variables, such as temperature, rain, wind, and humidity, forest fires can be predicted, and their future effects can be lessened. Forest fires have devastating effects on the environment, leading to ecological degradation, climate change, water resource depletion, and an increase in natural calamities. Fire prediction is a critical component of managing these occurrences. In recent years, simulation tools have been used to forecast forest fires, but they have encountered issues with data quality and tool execution time. Machine learning, a form of artificial intelligence, has been used to teach computers to predict forest fires. By using machine learning algorithms and meteorological variables, such as temperature, rain, wind, and humidity, forest fires can be predicted, and their future effects can be lessened.

TEAM MEMBER DETAILS

<1. vtu14400/ Goli Pavan >
<2. vtu13952/ V.C Jagapathi Babu >
<3. vtu12658/G Krishna Chaitanya>
<9346014976>
<9347119524>
<9347068609>
<vtu14400@veltech.edu.in>
<vtu13952@veltech.edu.in >
<vtu12658@veltech.edu.in >

INTRODUCTION

Forest fires (also known as wildfires) are one of the recent calamities that occur most often. Several acres of forest are being burned as a result of these flames. The primary causes of forest fires are global warming because of the a rise in the earth's average temperature and carelessness on the part of people. Forest fires are brought on by human activities like animal husbandry and agriculture in Africa , india . The spread of flames may now be predicted using a variety of technologies, including mathematical and physical models. In order to specify and forecast fire development in various places, these models rely on data collecting during forest fires, modelling, and lab experimentation.
The spread of flames may now be predicted using a variety of technologies, including mathematical and physical models. For the purpose of describing and forecasting fire growth in several places, these models rely on data gathered during forest fires, modelling, and lab trials. In recent years, simulation tools have been employed to forecast forest fires, however these systems have encountered issues with data quality and tool execution time. A form of artificial intelligence called machine learning

METHODOLOGIES

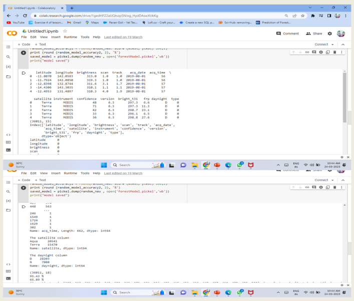
In this project we tried to make a prediction for the burned area in Forest Area. Forest Fires Data Set was used for this analysis. The data was clustered. Stepwise regression methods were applied to choose one best predictor. It is interesting to see, which one of them has the biggest impact on the burned area in each cluster.

Fire Cast receives the same input variables as the training fires for the testing fire, starting with the first recorded fire perimeter and making next day predictions for all collected perimeters.

RESULTS

The final results show that the proposed forest fire detection method detects fire-like items with a high detection rate (92%) and a low false-alarm rate (8%). Our model has a precision value of 0.84 and recall value of 0.97. These findings suggest that the proposed technologies accurate and suitable for application in automatic forest-fire detection systems. The method's accuracy could be enhanced in the future by extracting more fire features and expanding the training data set.

Input and Output.



STANDARDS AND POLICIES

IEEE 830-Software Requirements Specifications
IEEE 829-Software Test Documentation
IEEE 1012-Software verification and validation

Figure 1. Wild fire in forest

Figure 2. fire caused in forest

CONCLUSIONS

From this review we have done the detail execution of random forest machine learning model. The model was developed on training data and tested for new data. 95.55% of accuracy was obtained using this algorithm.

ACKNOWLEDGEMENT

1. Project Supervisor Name/Designation: Ms.R.Vaishanavi / Assistant Professor
2. Project supervisor Contact No: 8148552249
3. Project supervisor Mail ID: vaishnavir@veltech.edu.in

References

[1] Dieu Tien Buia, Nhat-Duc Hoangc, Pijush Samuid(2019). “Spatial pattern analysis and prediction of forest fire using new machine learning approach of Multivariate Adaptive Regression Splines and Differential Flower Pollination optimization ”: A case study at Lao Cai province (VietNam)”,Journal of Environmental Management, Volume 237.

[2] Gao, X., Wei, Y., Zhao, Y. (2020). ”Forest fire detection based on deep learning and multiple features”. Applied Sciences, 10(15), 5172.

[3] Javed, S., Hussain, M., Riaz, S., Abbas, Q. (2020). ”Forest Fire Prediction Using Machine Learning and Internet of Things. Journal of Sensors”, Volume 2020, Article ID 7415124.

[4] Karimi, B., Ghaemi, A. (2020). Prediction of forest fires using machine learning algorithms. Geomatics, Natural Hazards and Risk, 11(1), 1744-1759.

[5] Lim, C. H., Kim, M. J., Lee, W. K., Lee, D. H. (2018). ”Predicting forest fire occurrences using data mining techniques”: a case study in South Korea. Journal of Forestry Research, 29(6), 1649-1659.

[6] Mohamed F. El-Amin, ”Detection of Hydrogen Leakage Using Different Machine Learning Techniques”, 2023 20th Learning and Technology Conference (LT), pp.74-79, 2023.

[7] Neri, A., Angelini, M. G. (2019). ”Fire risk forecasting using machine learning models ” based on meteorological data. Environmental Monitoring and Assessment, 191(9), 581.

[8] Singh, A., Singh, G. (2020). ”Predicting Forest Fires Using Machine Learning”: A Comparative Study. IEEE Access, 8, 71662-71669.

[9] Salis M, Arca B, Alcasena F, Arianoutsou M, Bacciu V, Duce P, Duguay B, Koutsias N, Mallinis G, Mitsopoulos I, Moreno JM(2019). “Predicting wildfire spread and behaviour in Mediterranean landscapes”, International Journal of Wildland Fire, volume 25, pages 1015 -1032.

[10] Tymstra D, C., Woolford, D. G., Martell, D. L., Scott, J. (2015). ”Predicting forest fires in Ontario, Canada using random forests”. Forest Science, 61(5), 909-919.