International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

FOREST FIRE PREDICTION USING MACHINE LEARNING TECHNIQUES

Mr. ARUNKUMAR V[1] MSc Data Science, Ms. Kanimozhi.V[2] MTech,(Ph.D.)
[1]Student, Rathinam College of Arts and Science
[2]IT Assistant Professor Department of Computer Science,
Rathinam College of Arts and Science

***Abstract*—In this paper, we examine the use of numerous strategies for predicting the occurrence of forest fires across the country of India. Here, we'll look at a few classification techniques, including decision trees, logistic regression, and support vector machines (SVM). We employ the Random Forest with Bagging technique for purposes of comparison and if it gives good accuracy then it will take out to the deployment. By using their zip code, these algorithms will notify them about the forest fire alert and fire zone on the dedicated website. Python is utilized for all calculations and model construction.**

***Index Terms*—SVM, Bagging, Random Forest, AI.**

I. INTRODUCTION

Forest fire is an important environmental world phenomenon and affected the environment, infrastructures, and human life. Forest fires have recently become one of the most common disasters that have been recorded to cause the destruction of hectares of forests. They threaten the forest resources and the entire regime, the fauna, and plants, which seriously disrupt biodiversity, the ecosystems, and the environment.

In summer when there is no rain for months, it is packed with dry senescent leaven and low humidity on the surface it will give the fire way to spread easily on the forest. There are numerous technologies

for fireplace models to predict the unfolding of fires, like physical models and mathematical models. These models rely upon the knowledge assortment throughout forest fires simulations and science laboratory experiments to specify and predict fireplace growth in several areas. Recently simulation models wont predict forest fires it have some issues like the accuracy of the compute file and simulation tool the simulation tool take too much of time to predict the fire . Machine learning could be used for the effective prediction of fires Artificial intelligence's area of machine learning enables computer systems to automatically learn from experience and get better over time without explicit programming. Machine learning may be divided into 2 classes: supervised, unattended and reinforcement. Supervised machine learning algorithms are as regression, Support Vector Machine (SVM), Artificial Neural Networks (ANN) and Decision trees. In unattended learning, the information attributes don 't seems to be tagged. This leads that the formula should outline the labels.

- The Main aim of forest fire prediction is to provide proper information of fire before and after fire happened.
- Main Factors for fires are temperature, Humidity, Wind Speed.
- A land possible to fire may have indicators, a land will selected by zip code.
- Every year, fire destroys millions of hectares of land these fires erupting more carbon monoxide than vehicle because of green area of forest also affected by fire.

II. Literature survey

A. Forest Fire Prediction using Artificial Intelligence George E. Sakr and colleagues (2010) A methodology for studying artificial intelligence-based forest fire prediction approaches has been proposed. Help vector machines are the foundation of the forest fire risk forecast method. Lebanon data were utilised to apply the algorithm, and it has demonstrated its accuracy in estimating the risk of fire.

B. Prediction of Forest Fire Occurrence in peatlands using Machine Learning Approaches

The development of early warning systems for fire detection has been one of the major achievements in the fight against forest fires. This study demonstrates that the machine learning approach, whether the traditional approach or the more modern and advanced approach, may be employed for peatland fire occurrence detection.

C. Forest Fire Prediction using Linear Regression Mukhammad Wilden Alauddin et al. (2018)The use of multivariate linear regression has been suggested for predicting forest fires. A few of the variables are temperature, humidity, wind, and precipitation. Different linear regression coefficients are computed using different methods including gauss-Jordan, gauss-seidel, and least-squares. The findings of a comparative comparison of the methodologies are discussed.

D. *An Integrated Approach to the Regression Problem in Forest Fires Detection*

For measuring the burned area in relation to the challenge of detecting forest fires, a complicated methodology was created. Data from Montensinho Park was used to test the suggested methodology. This study compares the created approach to traditional regression algorithms and demonstrates its benefits. It is shown that the proposed solution has a noticeably higher quality

III.PROPOSED SYSTEM

This section discusses the working architectures of System

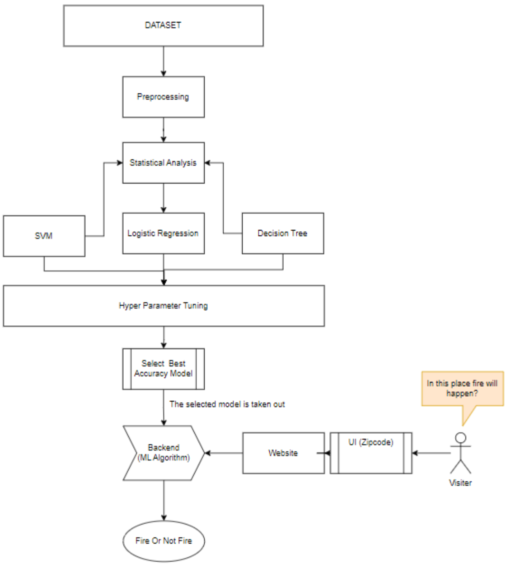


Fig. 1. Working Arch. Of System.

The proposed architecture explains how we obtained the Forest Fire data dataset from Kaggle and used exploratory data analysis methods along with feature engineering (pre-processing), where we tried to omit unnecessary data to make the dataset easier to understand. After applying the pre-processing technique, hotspot locations are determined using meteorological data and forest fire data, and then models are applied to that dataset to predict the occurrence of fire. Finally, using metrics, all models are compared, and the model that provides the prediction notifies the user and fire station about the Temperature and Fire.

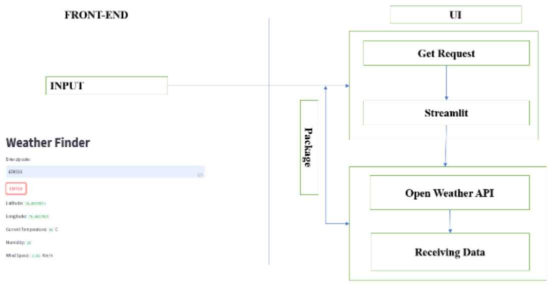


Fig.2. API connection

The Python method will request to open the weather API when a user inputs a zip code, and it will then receive information about the temperature, wind speed, longitude, and latitude.

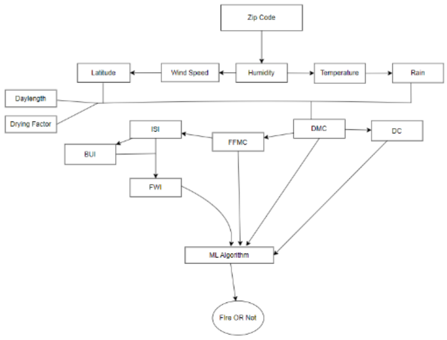


Fig.3. Workflow After API request

The process is based on the Canadian paradigm for preventing forest fires. Calculations are made to determine the FWI (Forest Weather Index) and its associated values. when the model has received the inputs.

- The information used in this paper was gathered from Kaggle. It includes The collection contains 244 occurrences that group data from two locations in Algeria, namely the Sidi Bel-abbes region in the northwest and the Bejaia region in the northeast..

- After data preparation, an appropriate model should be chosen based on accuracy.

IV.METHODOLOGY

1) *Exploratory Analysis:* Exploratory Data Analysis is currently one of the top techniques employed in data science. People often don't realise the distinction between exploratory data analysis and data analysis until later in their careers. Exploratory data analysis pays homage to inferential statistics, which employs arbitrary data to guard against formulae and rules that are generally somewhat strict. Exploratory data

analysis will be used to examine a data set. We load libraries to get this exploratory analysis started, and then we construct data plotting routines in Matplotlib.

2) *Pre-processing Analysis*: When we go through and talk about data, normally we think about any big databases with a massive number of rows and columns. Although this is likely to be the case, it is not necessarily the case that the data may be in too many different forms: Structured tables, images, audio files, videos, etc. Data: Data Pre-processing is the stage in which the data is converted, or encoded, to get it to such a state that now the computer can quickly parse it. Pre-processing is one of the features that offers a number of functions and transform classes to translate raw data vectors into representation and also to convert raw data into a clean data set i.e. (when data is gathered from different sources, in turn, it is collected from the raw format which is not feasible for the analysis).

- To show the association between the meteorological parameters such relative humidity, wind speed, temperature, and rain, a correlation matrix was shown for the pre-processed data. By doing so, we may gauge how two variables interact and how dependent they are on one another..
- This finding indicates that there is less of a correlation between relative humidity and temperature..
- There is a positive association if the total is bigger than 0. Since the correlation has a negative value, we can assume that there is less of a relationship between relative humidity and temperature. A negative correlation is one where the value is less than zero.

3) To ensure that the answer for "fire is 1" and not "fire is 0" was label encoding.

4) We split our dataset into training and testing datasets in an 80:20 ratio after encoding.Analogous Regression: The logistic function has an S-shaped curve with a low initial value, a sharp middle climb, and a levelling out as it gets closer to 1.0. The logistic function is expressed as follows: $p(y=1|x) = 1 / (1 + e^{(-z)})$

where $p(y=1|x)$ is the probability of the binary outcome (y=1) given a set of explanatory variables (x), and z is a linear combination of the explanatory variables and their associated weights:

$z = b0 + b1x1 + b2x2 + ... + bn*xn$.

where b0 is the intercept term and b1, b2, ..., bn are the weights associated with each of the explanatory variables (x1, x2, ..., xn).

The logistic function converts the values of z into probabilities by mapping them into the range [0, 1]. The chance of the binary outcome approaches 1.0 when z is large and positive, and it approaches 0.0 when z is large and negative. The moment of greatest uncertainty occurs when z is zero, which corresponds to the binary outcome's probability of 0.5.

4)Decision Tree: Within the category of tree structure, the decision tree builds a categorization model. The data set is divided into smaller paired degrees and smaller subsets. The knowledge is divided into subsets of instances with the same values (homogenous), and the decision tree is generated top-down from the root node.

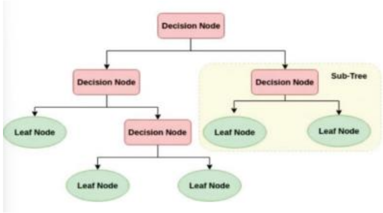


Fig.4. Decision Tree

4) *Support Vector Machine*: to locate the best border, or hyperplane, that can classify data points into several groups. In the case of binary classification, this hyperplane divides the dataset into two parts, one of which contains data points from one class and the other of which contains data points from the other class. SVMs search for the hyperplane that maximises the margin between the two classes. The margin is the distance between the closest data points in each class and the hyperplane. By keeping the hyperplane as far away from the data points as

possible, SVMs are able to perform more generally.

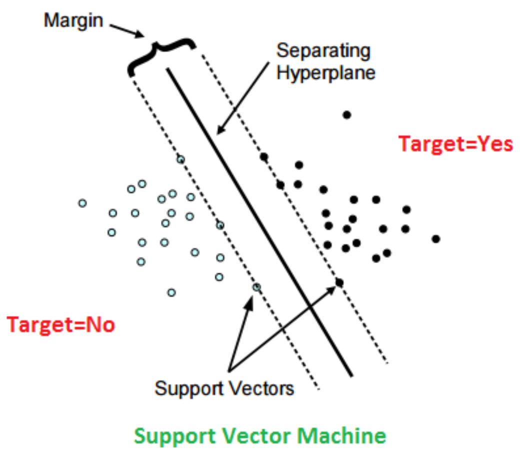


Fig.6. Support Vector Machine

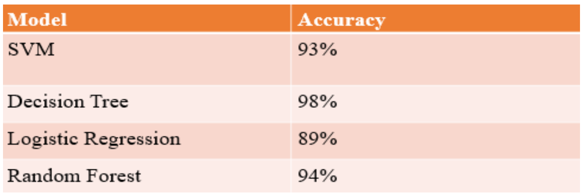
5) *Random Forest:* A supervised learning rule called random forest that uses ensemble learning for classification and regression may exist. The area unit and the trees in the random forest are parallel. With certain helpful adjustments, a random forest may be a meta-estimator (i.e., integrates the results of many forecasts) that combines various call trees: There is a maximum amount of options that can be shared across each node as a percentage of the total.

6) *Hyper Parameter Tuning:* The hyperparameters of a model can be changed using the machine learning technique known as grid search cross-validation (CV). Hyperparameters are parameters that are defined by the user before the model is trained rather than being inferred from the data. Examples include the regularisation parameter in logistic regression and the number of trees in a random forest.The core concept behind grid search CV is to train and test the model on many subsets of data to evaluate how well it performs for different combinations of hyperparameters. For this, the performance of the model is evaluated for each possible configuration of the hyperparameters using a grid of all possible hyperparameter values.

7) *Bagging:* Bagging, often referred to as Bootstrap Aggregating, is an ensemble method used in machine learning to improve the stability and accuracy of a prediction model. The underlying idea behind bagging is to train several instances of the same model using different subsets of the training data, then combine the results from all the models to get a final prediction. Several bootstrap samples are generated from the training data using the replacement approach in bagging, and these samples are then used to train various models. This sampling technique works to increase the overall accuracy of the model by ensuring that each model is trained on a slightly different subset of the training data.

V. RESULT ANALYSIS

The information gathered is used to train the system and generate predictions. By examining the temperature, humidity, rain, wind speed, and other relevant parameters, we can predict a forest fire. There are four types of regression techniques used for prediction: *Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and Decision Tree (DT*). The intended models were implemented using the Python platform. A decision tree, a Support Vector Classifier Logistic Regression model, and a random forest with bagging are used for implementation once the results from the model coaching and testing are compared.



Model	Accuracy
SVM	93%
Decision Tree	98%
Logistic Regression	89%
Random Forest	94%

Fig.7. Accuracy

Accuracy: The accuracy statistic determines the proportion of cases out of all instances that are correctly classified. The amount is determined by dividing the total cases by the number of occurrences that were successfully classified.

Precision : is a measurement of the proportion of accurate positive forecasts among all positive forecasts. By dividing the total number of true positives by the sum of all true positives and false positives, it is calculated.

Recall: The recall quantifies the proportion of correct positive predictions among all instances of positive outcomes. By dividing the total number of true positives by the sum of all true positives and false negatives, it is calculated.

A. Logistic Regression

- Accuracy:0.89
- AUC:0.9442

B. SVM

- Accuracy:0.93
- AUC:0.9223

C. DT

- Accuracy:0.98
- AUC:0.9324

D. Random Forest

- Accuracy:0.94
- AUC:0.9455

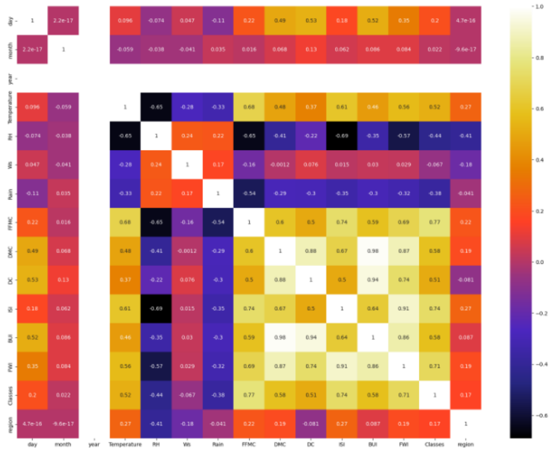


Fig.8. Correlation graph

This graph shows the dependency between two variables.

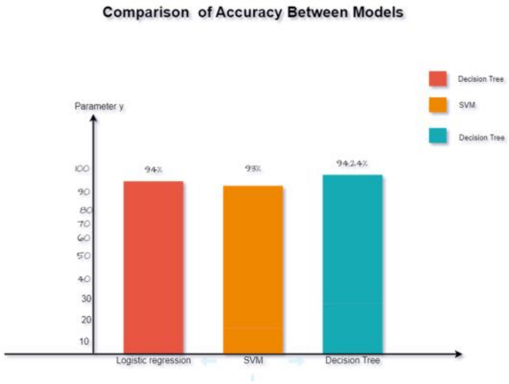


Fig.9. Distribution Graph

Accuracy of Models are compared selected to deployment.

VI.CONCLUSION AND FUTURE SCOPE

The factors causing fires are Temperature, Relative, Humidity, Windspeed, Latitude and Longitude. Experiments are completed with different *zip codes* in order to have good training, instances set, and evaluation instances set for forest fire prediction. *High temperature and low humidity are the reason for forest fires*. The website integrated with Random Forest is used as the number of fires and the death toll of residents and tourists rises daily. We may use Real time sensors in the deep forest and the government will be incorporated with tourists who enters in particular forest, If Government Takes this website, it will give the government ease way to track passengers that will helps in Resource Management and rescue Personals.

REFERENCES

[1] D. T. Buia, Q.-T. Buib, Q.-P. Nguyenc, B. Pradhand, H. Nampak and P. Trinh, "A hybrid artificial intelligence approach using GIS-based neural-fuzzy inference system and particle swarm optimization for forest fire susceptibility modeling at a tropical area", Agricultural and Forest Meteorology, vol. 233, pp. 32-44, February 2017

[2] P. Cortez and A. Morais, "A Data Mining Approach to Predict Forest Fires usingMeteorological Data", In: Neves J, Santos M F, Machado J (eds.) EPIA 2007, pp.512-523, 2007

[3] G. F. Shidik and K. Mustofa, "Predicting size of forest fire using hybrid model", Proceeding ICT EurAsia Conference 2014, pp 316- 327, 2014

[4] D. Rosadi and W. Andriyani, "Prediction of forest fire using ensemble method" Presented in ICMSE 2020, Semarang, Indonesia, October 6th 2020,

[5] J. Zhu, H. Zou, S. Rosset and T. Hastie, "Multi-class AdaBoost", Statistics and Its Interface, vol. 2, pp.349–360, 2009

[6]Dedi Rosadi,Widyastuti and Deasy Arisanty,'Prediction of Forest Fire Occurrence in Peatlands using Machine Learning Approaches',vol.3,ISRITI,2020

[7] Kajol R Singh, K.P. Neethu, K Madhurekaa, A Harita, Pushpa Mohan,' Parallel SVM model for forest fire prediction', ELSEVIER,Letr3,2020

[8] Detection of Forest Fires using Machine Learning Technique: A Perspective Aditi Kansal1, Yashwant Singh2, Nagesh Kumar3, Vandana Mohindru4 Department of Computer Science & Engineering Jaypee University of Infonnation Technology Waknaghat, Solan- 173234, (H.P), India 1 aditi.kansaI4@gmail.com, 2yashu _ want@yahoo.com