PREDICTION OF INFERNO USING ML TECHNIQUES

Ankitha N
Master of Computer Applications
M S Ramaiah Institute of Technology
Bengaluru, India
ankitharaj142@gmail.com

Varshitha K
Master of Computer Applications
M S Ramaiah Institute of Technology
Bengaluru, India
Varshithashetty03@gmail.com

Ms. Nithya B N

Master of Computer Applications
M S Ramaiah Institute of Technology
Bengaluru, India

Abstract— The key component of controlling fire occurrence in nature (Inferno) is main motto of this prediction mode, that has been presented in this paper. Everyone's lives and the environment are significantly impacted by fire occurrence in nature. It is not always possible to contain a fire and get to the wooded region in time. As a result, there is frequently a lot of destruction. Hence, it is crucial to be prepared for fires and act quickly when they occur. Millions of hectares of forests and animals are lost each worldwide due to fire occurrence in nature, not to say of human lives and cause significant economic and ecological harm. Many ecosystems, including grasslands and temperate forests, depend heavily on fire occurrence in nature. It is expected that forest fire prediction could decrease the effects of forest fire in the future. The region affected by the fire is forecasted using satellite imagery in the current work processes. The proposed system processes used meteorological variables like temperature, rain, wind, and humidity to predict when a forest fire will occur. Using meteorological data, this paper then predicts forest fire danger using a machine learning method. Based on the available literature and limitations, we can conclude that various studies have revealed the amount of burnt area caused by fire occurrence in nature, and many have presented various models to predict fire occurrence in nature. This work compares various models for predicting fire occurrence in nature, such as Decision Tree, Random Forest, Support Vector Machine, MLP Classifiers and Artificial Neural Network (ANN), XGBoost classifier algorithms. The primary goal of this work is to forecast the likelihood of a forest fire and its intensity in specified atmospheric circumstances in each region.

Keywords—Machine Learning, MLP Classifiers, prediction, XGBoost classifier, ANN.

I. INTRODUCTION

Forests play an essential role to the earth's ecological balance. Fire occurrence in nature is a source of concern because they cause extensive damage to the environment, property, and human lives. As a result, it is critical to detect the fire at an early stage. One of the most common causes of fire occurrence in nature is warming, which is caused by an increase in global average temperature. However, these natural riches are under threat from fires caused by both natural and human forces. Fire occurrence in nature, not only burn trees, but they also inflict significant harm to the forest ecosystem, habitat, and forest structure, as well as changes in the physical and chemical qualities of soil. When soil loses its ability to penetrate and retain water, the underground water level in the forest increases, under severe situations, soil pollution develops, followed by a decline in forest biomass and an imbalance in the ecosystem. Every year, around one-tenth of the world's forests have become barren because of these reasons. Total forests are lost by fire, with over 200,000 fire occurrences in nature occurring in yearly. Predicting forest fires is primarily done to enable the fire management team's firefighters. These predictions help the fire fighters to have a fair idea about where the fire is going to start so that very effective actions can be taken beforehand to avoid calamities.

Every year, Fire obliterates millions of hectares of land. These flames have consumed enormous swaths of land and produce more carbon monoxide than total transportation. Meteorological conditions are the primary causes of fire. Climate data is obtained from adjacent sensors, which are combined in the nearest meteorological stations. Monitoring potentially dangerous places and detecting fires quickly can cut down on response times, damage risk, and firefighting expenses. Machine learning techniques appear to be a viable alternative for streamlining the entire dataset. There are several machine learning techniques that effectively do the job for us. We have used many methods one of them is to collect all features from data set and then do training with the help of machine learning algorithms to classify the data into target variable and predictor variable. The behavior of these models has been encouraging, with excellent accuracy in predicting fire occurrence in nature.

In order to help a team managing a forest fire, know where the fire is most likely to be started, this work suggests a model for forest fire prediction in a specific area. The dataset used is the Prediction of fire occurrence is UCI dataset https://archive.ics.uci.edu/dataset/162/forest+fires. The databases offer details on the various factors that affect the prediction of forest fires. To find the optimum model for each dataset, machine learning techniques such decision trees, random forests, logistic regression, support vector classifiers, MLP classifiers, and XGBoost classifiers are utilized. In order to decide which machine learning algorithm is appropriate for the dataset under consideration, confusion measures are used to analyze and evaluate each model from every possible aspect. The goal is to increase the accuracy and speed of predicting the forest fire, ultimately leading to earlier intervention. So when we are able to detect the fire before it occurs, we would help the firefighters get it under control faster and more efficiently.

LITERATURE REVIEW

Nizar HAMADEH LARIS EA et al. (2015) in this paper the author has tried to predict the occurrence of fires in forests in Lebanon. Among the parameters are relative humidity, temperature, and wind speed. Artificial Neural Networks were used on these criteria in order predict fire occurrence in nature. The result is this paper is about 94% accuracy.[2]

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 utilized to apply the algorithm, and it has demonstrated its accuracy used for a two-class prediction of fire risk with a very high accuracy of up to 96% in estimating the risk of fire. [1]

Mukhammad Wildan Alauddin et al. (2018) The use of multiple 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 such gauss-jordan, gaussseidel, and least-squares. The findings of a comparative comparison of the methodologies are discussed. This section talks about the review of the literature. [3]

Preeti T et al. (2021) in this paper they predicted the forest fire occurrence using different algorithms like using regression techniques used for prediction are Random Forest (RF), Decision Tree (DT) and Support Vector Regression (SVR) and Naive Bayes and the data set they have used is from the natural park of Montesano in the European republic. [4]

II. METHODOLOGY

A. System Design

A computational paradigm known as system architecture or system design describes the structure, behavior, and perspectives of the system. System modules and subsystems that work together to operate the whole system can make up a system architecture.
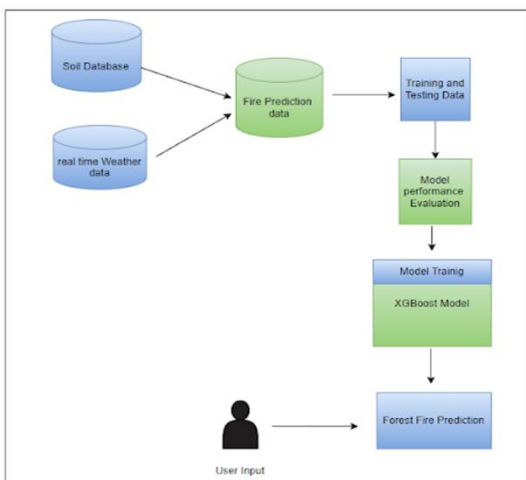

Figure 1. System architecture

B. The Dataset

The main dataset for the Machine learning has been taken from the UCI Machine Learning Repository
https://archive.ics.uci.edu/dataset/162/forest+fires

The forest fire dataset is a multivariate dataset, which is a term used to describe a data set that includes two or more than two variables. It has 13 attributes having 2500 instances.

The attributes are explained as follows:

1. X - x-axis the location of the fire on the Montesinho park map, in the range from 1 to 9.

2. Y - y-axis the location of the fire is shown by the y-axis geographical coordinates: 2 to 9 within the Montesinho park map.

3. month - The year's months are "January to December."

4. Day - day: 'Monday to Sunday' of the week

5. FFMC - 18.7 to 96.20 for the FFMC index from the FWI system. The Fine Fuel Moisture Code indicates the moisture level of surface litter, which affects fire spread and ignition.

6. DMC - The Duff Moisture Code, which indicates the moisture content of shallow and deep organic layers and influences fire intensity, ranges from 1.1 to 291.3 in the FWI system.

7. DC- The Drought Code, which measures the moisture content of superficial and deep organic layers, which determine fire intensity, ranges from 7.9 to 860.6 in the FWI system.

8. ISI - The FWI system's ISI index ranges from 0.0 to 56.10. It is determined by and correlated with fire velocity spread by the initial spread index.

9. Temp - Celsius range of temperatures for the area: 2.2 to 33.30.

10. Relative humidity (RH) in percent: in the air, 15.0 to 100.

11. Wind - during the time of the fire, wind speeds ranged from 0.40 to 9.40 km/h.

12. Rain - 0.0 to 6.4 mm/m2 of outdoor rainfall.

13. Area - The forest area (in ha): 0.00 to 1090.84 that burned during the forest fire.

C. Data Preprocessing

Data preprocessing is first and foremost stem in converting raw values into acceptable values for ml algorithm. There are a total of 2500 occurrences with 8 attributes in the Forest Fire dataset. Out of the 13 attributes 6 attributes are float data type, 2 are of string data type.
Missing Values: Identify and handle missing values using functions like isnull(), dropna(), and fillna() in Python libraries like pandas. Detect and remove duplicate rows using duplicated() and drop_duplicates() functions. Identify outliers using statistical methods or visualizations and handle them through replacement, winsorization, or removal

Convert data types, handle inconsistent values, and correct inaccurate data using appropriate transformations and functions.

Skewed Data: Address skewness through log transformation or Box-Cox transformation for normalization in Python using libraries like pandas and NumPy.

XGBOOST Classifier

Machine learning techniques like supervised learning can be used to solve classification and regression issues. Ensemble learning, where numerous classifiers are integrated to address complicated issues and increase model performance, is one internal well-liked strategy in supervised learning. This study focuses on using the XGBoost Classifier algorithm, an ensemble learning strategy that seeks to outperform other algorithms in terms of performance measures.

Algorithm

Algorithm XGBoost()

{

Input: Data set with size S which consists on [mxn] values

Output: Predicting output "Test set result"

　　{

1.　Data Pre-processing of Dataset D

　　// import the libraries
　　// import the dataset
　　　　　　D =Dataset
　　// extracting the independent and Dependent variable
　　　　　　X -> Independent variables
　　　　　　Y -> Dependent variable
// splitting data into training and testing data in 6:4 ratios
　　　　　　Such as
　　　　　　 X_train, x_test, y_train, y_test

2. Fitting the XGBOOST classifier algorithm to the training set

　　　　　　// with activation as relu and solver as adam classifier = XGBOOSTClassifier (activation='relu',
　　16,16),n_iter_no_change=100,solver='adam')
　// importing XGBOOST classifier function to classifier variable.
　　　　classifier.fit (x_tain, y_train)

3.Predicting the Test set result

y_pred -> classifier.predict (x_test)

//output

4. Creating Confusion Matrix for Accuracy
cm -> confusion_matrix (y_test, y_pred)

　　　　　　}

}

D. Data Analysis

In this section, an analysis was performed to inspect the association between each attribute and its influence on other attributes. The below pie charts shows that there is 53.2% fire occurred and 46.8% was safe in dataset.
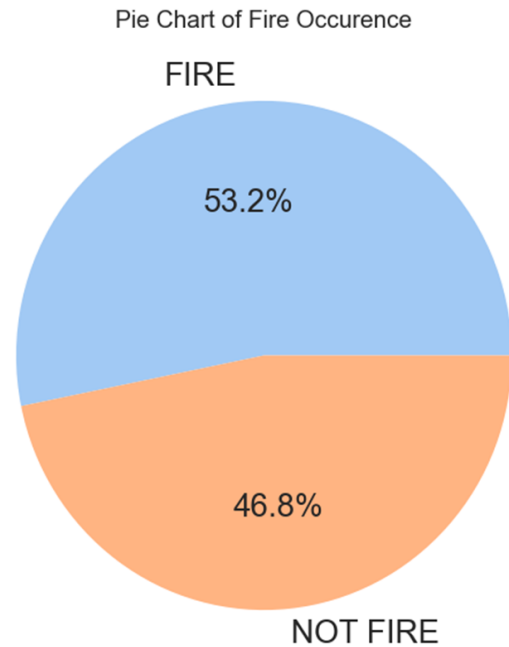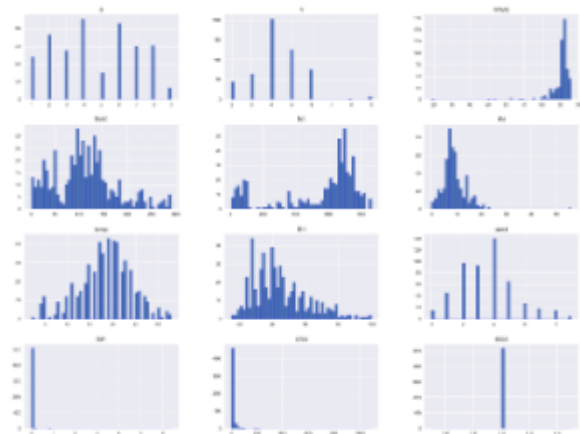

Figure 2. Occurrence of fire

Figure 3. The density of each attribute used is visualized in subplots

Figure 4. Fire analysis in month wise

The month wise analysis of fire occurrence shows in month of September the fire occurrence is more and as well as august it has both highest count in occurrence of fire and not occurrence of fire
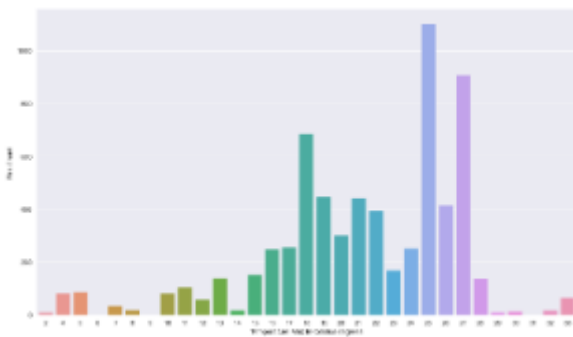

Figure 5. Temperature record

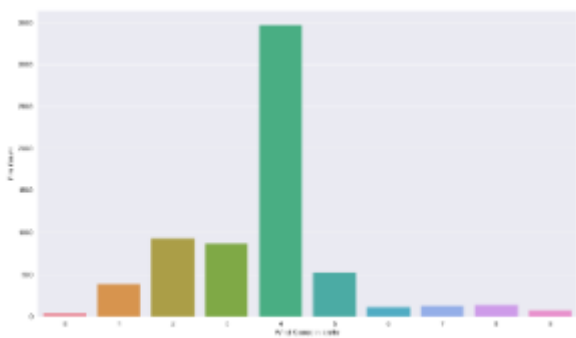The maximum temperature recorded was 25 degrees Celsius.

Figure 6. Wind speed bar chart

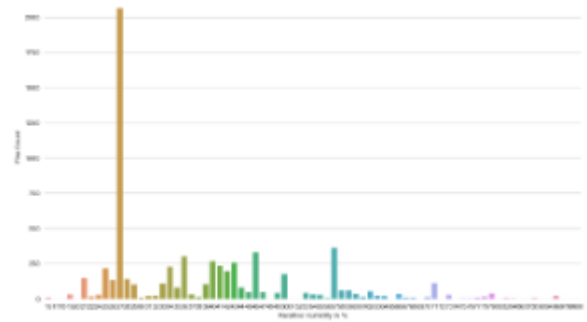The highest recording of wind speed per kilometer is 4 km/hr.

Figure7. Relative humidity values

The 27% is highest value of Relative humidity

E. *Data Prediction*

After pre-processing the data uses "sklearn" preprocessing techniques and replaces all NaN values with "0". The dataset is partition into two sets: "test data" and "training data" in a 2:8 ratio. The learning set of data is used to examine many machine learning algorithms to identify the prominent algorithm so it can foretell the fire occurrence in nature. The correlation relationship between each attribute is shown in the co-relation heatmap below, with light-toned color shades indicating higher correlations between attributes. This figure indicates that the dataset is not over fitted. It shows how each attribute is unique.
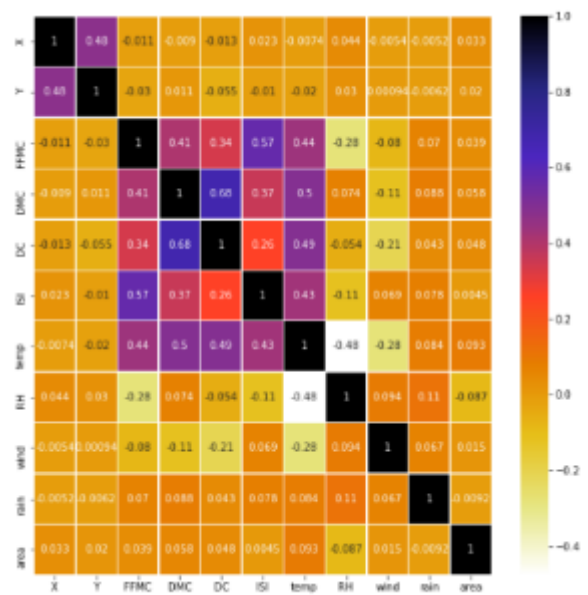

Figure 8. Correlation of Attributes.

III. PERFORMANCE METRICS

A. **True Positive (TP)**
   It is number of values are predicted true and it is true.
B. **False Negative (FN)**
   It is number of values are predicted false, but it is true.

C. **True Negative (TN)**
It is number of values are predicted false and it is false.

D. **False Positive (FP)**
It is number of values are predicted true, but it is false.

E. **Accuracy**
It is the ratio of number of correct predictions to the total number of predictions.

$$Accuracy = \frac{Number\,of\,Correct\,Predictions}{Total\,Number\,of\,Predictions}$$

F. *Sensitivity / Recall*
Sensitivity is the fraction of individuals who will be correctly predicted as positive class. High sensitivity means a greater number of individuals have been correctly predicted to have ASD.

$$Sensitivity = \frac{TP}{TP+FN}$$

G. *Specificity*
Sensitivity is the fraction of individuals who will be correctly predicted as negative class. High specificity means a greater number of individuals have been correctly predicted not to have ASD.

$$Specificity = \frac{TN}{TN+FP}$$

IV. RESULTS

Four discrete machine learning algorithms, including Support Vector Machine, Random Forest, Logistic Regression, and Naïve Bayes, have been utilized to analyze the preprocessed dataset. Among all the algorithms, XGBOOST classifier exhibited the highest accuracy, and specificity based on various performance metrics. Therefore, XGBOOST classifier has been hand-picked for further prediction of the fire occurrence in nature.

Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)
MLP Classifier	0.96	0.9512	0.910
Random Forest	0.95	0.9579	0.881
Logistic Regression	0.90	0.9589	0.861
XGBOOST classifier	0.97	0.9726	0.912

Table 1. Performance Metrics on different ML Algorithm

Here we try to implement flask framework for implementation of website locally.
Where they can give the data for the values and it will predict whether there is fire or not in that particular area. If it shows 0 then fire is not there, if it shows 1 then the fire as occurred.

Figure 9 Home page

Figure 10 Prediction page for fire

V. CONCLUSION

In this work, the implementation of the XGBoost algorithm for forest fire prediction has demonstrated impressive results with an accuracy of 97%. This high accuracy signifies the algorithm's effectiveness in accurately forecasting forest fires, enabling proactive measures to be taken for prevention and mitigation. The utilization of XGBoost's advanced ensemble learning techniques has significantly improved the prediction accuracy compared to traditional methods. This accomplishment contributes to enhancing forest fire management strategies, reducing potential risks, and protecting valuable ecosystems. The 97% accuracy rate provides reliable and actionable insights for forest fire management authorities, aiding in timely decision-making and resource allocation. Further research and development in this field could explore integrating real-time data sources and expanding the algorithm's capabilities to provide even more accurate predictions. Ultimately, the successful application of XGBoost in forest fire prediction highlights its potential as a valuable tool for safeguarding forests and preventing devastating wildfires.

VI. FUTURE ENHANCEMENT

As we have used 13 attributes in our proposed system, we can add more attributes by collecting from different forests and wildlife and understand those forest geographically climate collect the data and pre-process it to get accuracy. Next, we have XGBOOST classifier which we have used to build our ML model so with this XGBOOST we get about 97% of accuracy therefore we can explore on more classifiers or algorithms which can hence our model. Then we used python flask for building our web application on our model we can make it interactive using FastAPI or deploy it on Heroku where Heroku can be used for Platform as a Service offering that enables to carry out hassle-free application deployment, scaling, and management.

As for proposed project we can more add more components like forest and wildlife awareness to general public and make they understand the importance of preservation forest and its exotic flora and fauna. And we can add 24/7 AI chat-bot to interact with user and understand the flow usage of web application. And we can we have one more component to notify the government forest officer and fire extinguisher department so they can take instant measures the forest and wildlife.

Finally, we can make this ML project blend among another machine like IOT. In IOT we can add more sensors takes input from forest and give to our model for prediction. Or add surveillance camera to detect fire and help extinguish the forest fire and help preserve our precious environment.

VII. REFERENCES

[1] George E. Sakr, Imad H. Elhajj, George Mitri and Uchechukwu C. Wejinya. "Artificial Intelligence for Forest Fire Prediction" IEEE/ASME International Conference on Advanced Intelligent Mechatronics Montréal, Canada, July 6-9, 2010

[2] Nizar Hamadeh, Alaa Hilal, Bassam Daya, Pierre Chauvet "Studying the factor affecting the risk of forest fire occurrence and applying neural networks for prediction" SAI Intelligent system, 2015.

[3] A. Alonso-Betanzos, O. Fontenla-Romero, B. Guijarro-Berdi nas, E. Hern´andez-Pereira, M. Inmaculada Paz Andrade, E. Jim´enez, J. LuisLegido Soto, and T. Carballas, "An intelligent system for forest fire risk prediction and firefighting management in Galicia,"Expert systems with applications, vol. 25, no. 4, pp. 545–554, 2003.

[4] Preeti T, Dr.Suvarna Kanakaraddi, Aishwarya Beelagi, Sumalata Malagi, Aishwarya Sudi KLE Technological University " Forest fire prediction using machine learning techniques" International Conference on Intelligent Technologies, 2021.

[5] PranatiRakshit, Srestha Sarkar, Sambit Khan, Pritam Saha,Sonali Bhattacharyya,Nilarpan Dey " Prediction of forest fire using machine learning algorithm" International conference ,2021.

[6] R. Jaiswal, M. Saumitra, D. Kumaran, and S. Rajesh, "Forest fire risk zone mapping from satellite imagery and gis," International Journal of Applied Earth Observation and Geo-information, vol. 4, pp. 1–10,2002.

[7] Z. Li, Y. Kaufman, C. Ithoku, R. Fraser, A. Trishchenko, L. Giglio, J. Jin, and X. Yu, "A review of AVHRR-based active fire detection algorithms: Principles, limitations, and recommendations," Global and Regional Vegetation Fire Monitoring from Space.

[8] M. Wiering and M. Dorigo, "Learning to control forest fires," in Proceedings of the 12th international Symposium on Computer Science for Environmental Protection.

[9] T. Cheng and J. Wang, "Applications of spatio-temporal data mining and knowledge for forest fire," in ISPRS Technical Commission VII Mid Term Symposium, Enschede, 2006, pp. 148–153.