Forest Fire Prediction using Deep Learning

Harshal Pazare[1], Pratik Raut[2], Mihir Dhanorkar[3], Sameer Dhage[4]
Students, Department of Computer Science & Engineering[1,2,3,4]
G. H. Raisoni Institute of Engineering and Technology, Nagpur, Maharashtra, India
An Autonomous Institute Affiliated to Rashtrasant Tukdoji Maharaj Nagpur University, Nagpur, Maharashtra, India
Accredited by NAAC With A+ Grade

Abstract: *Wildfire prediction is an essential component of wildfire management. Accurate fire prediction is critical to mitigating its effects. It plays an important role in resource allocation, mitigation, and recovery. Wildfires rob the natural world of life and destroy the land they live on. Building forecast models will allow authorities to estimate the long-term effects of climate change on local forest distribution. Human motives include clearing and other agricultural activities, grassland care for livestock, extraction of non-timber forest products, industrial development, resettlement, hunting, neglect, and arson. Lightning strikes are the main cause of fires only in very remote areas of Canada and the Russian Federation. These wildfires are often man-made or caused by mother nature through varying weather conditions and wind. Wildfires economic and environmental impacts of wildfires are significant, and predictions can prevent long-term damage to sensitive forest ecosystems.*

Keywords: Graphical User Interface, MLP Classifier, Flask, Forest Fire Prediction

I. INTRODUCTION

Wildfires are one of the most serious disasters facing our society today, causing irreversible damage as well as having recently become one of the most common disasters destroying acres of forest. They seriously disturb not only forest resources, but also biodiversity, ecosystems, and local environments, posing threats to entire ecosystems, flora, and fauna. Wildfires have a mixed ecological and economic role, as some ecosystems rely on wildfires to maintain dynamism, biodiversity, and productivity. But wildfires burn down millions of acres of forest and other vegetation each year, claim the lives of many people and animals, and cause immeasurable economic damage, both in terms of resource destruction and containment cost. Human motives includes clearing and other agricultural activities, grassland care for livestock, extraction of non-timber forest products, industrial development, resettlement, hunting, neglect, and arson.

II. LITERATURE SURVEY

A wildfire hazard zone is a place where fires are likely to start and from which they can easily spread to other areas. Predicting the factors that influence fire initiation and understanding fire dynamic behavior are important aspects of fire management. Accurate assessment of forest fire problems and determination of solutions can only be adequately made if maps of fire risk zones are available. Forest fire endangered zones were demarcated by assigning subjective weights to all strata classes according to their susceptibility to fire or fire-inducing ability. Four wildfire risk categories were automatically derived, ranging from very high to low.

The developed GIS-based wildfire risk model for the study area was found to be in strong agreement with actual fire-affected locations. The methodology used in this research paper demonstrates that logistic regression modeling and geographic information system are suitable for determining the forest fire risk zone, and therefore management of it.

III. METHODOLOGY

3.1 Data Collection

The dataset used is a public dataset that is available for research purposes. This dataset was created by Paulo Cortez and Anibal Marais. The forest fire dataset is a multivariate dataset which is a dataset consisting of two or more two variables is referred to as a multivariate dataset. It has 10 attributes having 517 instances.

The total file size of the Kaggle Forest fire dataset is 19KB, which consists of various parameters responsible for the occurrence of a forest fire. Kaggle is a data science and machine learning community where students, professionals, researchers, and enthusiasts compete and share machine learning techniques as well as datasets. The dataset represented a collection of information from the forest fire department.

Oxygen	Temperature	Humidity	Fire Occurrence	FFMC	Wind	X	Y	DMC	Rain
40	45	20	1	91	6.7	7	5	26.2	-0.4
50	32	10	1	86.2	0.9	7	4	35.4	-0.6
44	20	12	1	90.6	1.3	7	4	43.7	0
60	45	70	1	90.6	4	8	6	33.3	-1.9
30	48	10	1	91.7	1.8	8	6	51.3	-1.1
50	15	30	0	89.3	5.4	8	6	85.3	5.9
5	35	35	0	92.3	3.1	8	6	88.9	4.2
5	20	70	0	92.3	2.2	8	6	145.4	5.6
60	32	19	1	91.5	5.4	8	6	129.5	-0.7
30	50	45	1	91	4	7	5	88	-1.4
40	55	80	1	92.5	7.2	7	5	88	0
30	39	100	0	92.5	4	7	5	73.2	4.5
10	12	12	0	92.8	6.7	6	5	70.8	5.2
20	23	23	0	63.5	2.2	6	5	126.5	3.9
60	34	15	1	90.9	4.5	6	5	133.3	0
80	4	68	0	92.9	5.4	6	5	141.2	0
25	56	79	1	93.3	5.4	5	5	35.8	-0.1
12	45	90	0	91.7	4.9	8	5	32.8	4.1

3.2 Data Preprocessing and Feature Selection

Data preprocessing is a process of preparing the raw data and making it suitable for a learning model. It is the first and crucial step while creating a supervised learning model. Data preprocessing is the concept of transforming raw data into a clean dataset.

Datasets are preprocessed to check for missing values, noisy data, and other inconsistencies before running the algorithm.

There were zero null values in the dataset, we then checked for skewed data, there was some skew in the data, which was been removed by interquartile ranging the data. As well as an outlier of the data has been checked while preprocessing the data. Bar plots and pie plots were used to see the correlation between two attributes of the dataset, from which we came across a detailed understanding of the dataset. To understand the detailed correlation of the attributes of the dataset, we applied correlation coding on the whole dataset, from which we were able to understand that oxygen, humidity, temperature, wind, and drought moisture code (DMC) have a better correlation between each other than other parameters of the dataset, so for further training of our prediction model, so for further training of our prediction model, these five parameters are been basically used as they have better correlation between each other and made a new CSV file and stored our selected parameters in this file for further process of our prediction model.
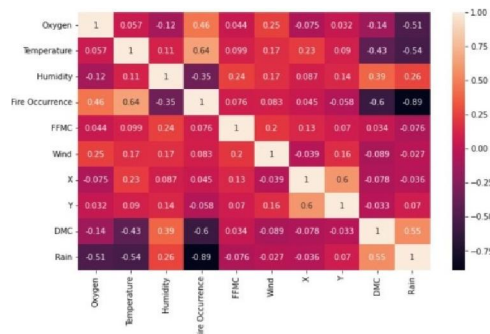

Fig. Correlation of Data

466

3.3 Training and Testing Algorithms

After performing data collection and data preprocessing for our model, we then further tried to use mainly two types of predictive algorithm regression and classification. Regression is a powerful tool for statistical inference and has also been used to try to predict future outcomes based on past observations were as classification algorithm categorizing something into a certain group or system based on certain characteristics. A program has been executed consisting of regression and classification algorithm simultaneously, which was deployed in such a way that without performing regression or classification algorithm separately, we can just select regression or classification algorithm in one code itself, which will reduce size of programming as well as save minute amount of time while programming, we used such coding in our programming because as a development procedure our aim is also to reduce time and make code small as well as efficient. Algorithm which was applied on our prediction model are Linear Regression, Random Forest, Logistic Regression and MLP classifier.

1. **Linear Regression** : Linear regression analysis is basically used to predict the value of one variable based on the value of another variable. The variable we want to predict is also called the dependent variable. A variable that is used to predict the value of another variable is called an independent variable.

2. **Random Forest (RF)** : Random Forest is a commonly used machine learning algorithm trademarked by Leo Breiman and Adele Cutler that combines the outputs of multiple decision trees to produce a single result. Its ease of use and flexibility are driving its adoption as it can handle both classification and regression problems.

3. **Logistic Regression** : Logistic regression is used for estimates the probability of an event occurring. For example, whether they are elected based on a particular set of independent variables. The dependent variable is constrained between 0 and 1 because the outcome is a probability.

4. **Multilayer Perceptron Classifier (MLP)**: MLP Classifier stands for Multi-Layer Perceptron Classifier, the name itself connects to a neural network. Unlike other classification algorithms such as Support Vector and Naive Bayes Classifier, MLP Classifier relies on an underlying neural network to perform the classification task.

In which almost every algorithm had some accuracy. Depending on the performance of the algorithm we have chosen the appropriate algorithm which in our case was the MLP classifier (Multi-layer Perceptron Classifier). We tried the 70:30 ratio in our model first to see the result of our model, but after researching about the train and split ratio required for the prediction model, we understand that the 80:20 ratio would be a perfect ratio for our training and splitting.
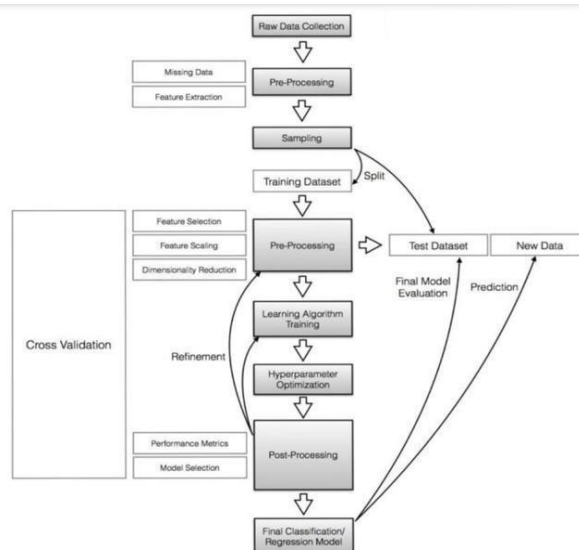

Fig. Block diagram of Architecture detail

Then we have applied MLP classifier algorithm on our split data, and made our model to have 16 hidden layers in it. We then applied the algorithm, r2 score has been checked of the algorithm as r2 score tells us how accurate our model works while processing our prediction task, the r2 score for the MLP classifier task was good, we then further move ahead to perform deployment of our model in to a web based graphical user interface.

3.4 Performance Metrics Details

Sr no.	Algorithm	R2 Score
1	Linear regression	0.43
2	Random Forest	0.62

Table 1: Performance of regression task

Sr no.	Algorithm	Accuracy
1	Logistic regression	86.21 %
2	MLP classifier	96.38 %

Table 2: Performance of Classification Task

As per performance observed for different algorithm MLP Classifier has better accuracy than other algorithm. Due to which we use MLP Classifier algorithm in our forest fire prediction model.

3.5 Pickle File

Pickle in Python is primarily used in serializing and deserializing a Python object structure. This is a process of converting a Python object into a byte stream to store in a file/database, maintain program state between sessions, transfer data over the network, etc. After testing and evaluating the MLP Classifier algorithm we store the trained algorithm in a pickle file. This pickle has been used to store trained algorithms in it, which is then been used in deploying of prediction model into a web GUI, by storing it in the backend. We use the pickle file because the pickle module keeps track of the objects it has already serialized, so that later references to the same object will not be serialized again, thus allowing for faster execution time. We prefer a pickle file over a CSV file for our prediction model because Pickle is about 11 times faster than a CSV file when uncompressed. Compression is a big issue when reading and saving files. Pickle file saves learned algorithm in it using the dump method. uncompressed. Compression is a big issue when reading and saving files. Pickle file saves learned algorithm in it using dump method.

3.6 Web Application Using Flask

Flask is a lightweight Python web framework that provides convenient tools and features for building web applications in the Python language. It is a framework that provides flexibility to developers and is accessible to new developers by allowing web applications to be quickly created with just a single Python file.

To perform proper interaction between user and prediction model decided to build a graphical based user interface (GUI) because GUIs are more intuitive to most of us than text-based interfaces, enough that even those with extremely limited knowledge of computers can use them without learning a coding language or computer command. To build User Interface (UI) we installed packages and extensions of HTML and CSS in our virtual environment. We build a proper interface for users using frontend programming and build a prediction button whose path was linked with our app.py file, so when user would enter values of parameters on screen and click on prediction button, our whole app.py and forest.py file connected to html file would run simultaneously and perform prediction. We have to click on terminal of IDE which and have to type python app.py, which runs whole programming and provide us with server link, on clicking on that server link we are being showed our user interface in which we give input values to our

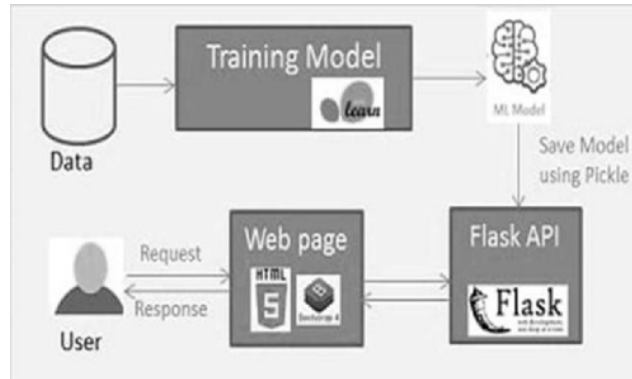model and in return our model gives predictive result.

Fig. External Working of Prediction Model

IV. RESULT AND DISCUSSION

When the preprocessing of the data has been completed, the data will be in operational form. And the 10 features which are obtained after preprocessing are taken for the remaining study. Among that, 80% of the data will be used for training and the remaining 20% will use for testing as random. The performance of classifiers varies when using different feature selection methods. The best features selected to for getting accuracies is through analytical forest.csv files are Oxygen, Temperature, Humidity, Drought Moisture Code (DMC), Wind, and Fire Occurrence. We have used various supervised learning model in our model like Linear Regression, Random Forest, Logistic Regression and MLP Classifier. We have used the python library Flask which is used for web applications. We are using Flask to create web page for our sample dataset to enter various features like Oxygen, Humidity, Temperature, Drought Moisture Code (DMC), and Wind. By entering various features in our web page we can identify that forest is in danger or not.
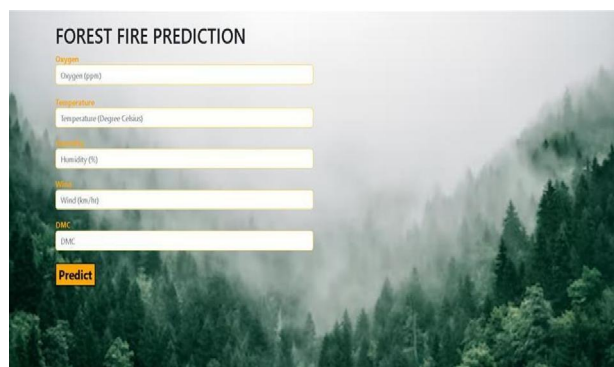

Fig. Generation of Server for User Interface

Fig. Graphical User Interface

V. CONCLUSION

Although the causes of wildfire outbreaks are not always predictable, wildfire risk is an understanding of how climate, geography, weather, and land cover affect fire behavior and likelihood of spreading. can be predicted to some extent.

Using deep learning and machine learning techniques, the monitor can detect natural disasters in real time. The world is moving toward automation, and in the age of big data, more solutions to complex problems need to be developed. Forest fires, while threatening human life, can cause significant environmental damage. A great deal of effort has been expended over the past two decades to develop automated detection tools that can support fire management systems (FFS).

It was decided to create a graphical-based user interface (GUI) for proper interaction between the user and the predictive model. This is because the GUI is more intuitive than the text-based interface. You don't need to learn programming languages or computer commands to use it. Installed HTML and CSS packages and extensions in the virtual environment for creating user interfaces (UI). Use front-end programming to create a user interface that suits your users, and create a prediction button whose path is linked to your app.py file. So when the user enters the values of the parameters on the screen and clicks the prediction button, the whole app .py and forest.py files associated with the html file are executed and the prediction is performed at the same time.

REFERENCES

[1]. Alonso-Benzos, O. Fontella-Romero, B. Guijarro-Berdi ˜nas ,E. Hernandez-Pereira, M. Inmaculada Paz Andrade, E. Jimenez, J. Luis Legion Soto, and T. Carballo's, "An intelligent system for forest fire risk prediction and fire fighting management in Galicia," Expert systems with applications, vol.25, no. 4, pp. 545–554, 2003.

[2]. V. Kecman, Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models. London: MIT Press, 2001.

[3]. D.R. Brillinger 、H.K. Preisler, J. W. Benoit, "Risk Assessment: The Wildfire Example," Statistics and Science: a Festschrift for Terry Speed, ser. IMS Lecture Notes Monograph Series, D. Goldstein and T. Geschwindigkeit, eds. Beachwood, Ohio: Institute of Mathematics and Statistics, 2003, vol. 40, pp. 177–196.

[4]. Aslam, Fankar & Mohammed, Hawa & Lokhande, Prashant. (2019). Efficient Way

[5]. Of Web Development Using Python And Flask. International Journal of Advanced Research in Computer Science.

[6]. T. Cheng and J.Wang, "Applications of Spatiotemporal Data Mining and Wildfire Knowledge", ISPRS Technical Commission VIIMMid Term Symposium, Enschede, 2006, pp. 148-153.

[7]. K Clark, J. Blass, and P. Riggan, "Cellular Automata Models for Wildfire Propagation and extinguishing," Photogrammetric Engineering and Remote Sensing, vol. 60, no. 11, p. 1355–1367, 1994.

[8]. G. Mitri and I. Gitas, "Semi- Automated Object-Oriented Model for Mapping Burned Areas in theMediterranean Using Landsat TM Imagery," International Journal of Wildland Fire, Vol. 13, No. 3, pp. 367-376, 2004.

[9]. J. Han, K. Ryu, K. Chi およびY. Yeon,Statistics Based Predictive Geo- spatial Data Mining: Forest Fire Hazardous Area Mapping Application, Research Workshop in Computer Science, 370-381, 2003.

[10]. R. E. Schapire. The Boosting Approach to Machine Learning – An Overview. MSRI Workshop on Nonlinear Estimation and Classification, 2002. Y. Freund and R. E. Schapire.