International Journal of Advance Research in Science and Engineering
Volume No. 12, Issue No. 04, April 2023
www.ijarse.com
IJARSE
ISSN 2319 - 8354

Prediction of Forest Fires Using Machine Learning

M.Harika , K.Naveena, K.Vinod Kumar , K.Om Keerthi

Student

Mr.S.Anil Kumar

Associate Professor, Project Coordinator

Department of Computer Science, Tirumala Engineering College, Narasaraopet, Andhra Pradesh

ABSTRACT

One of the most occurring disasters in recent times is forest fires(wildfires). Due to these wildfires, a lot of acres of forest area are getting destroyed. The significant reasons that lead to the occurrence of forest fires is warming due to the increase in the average temperature of the earth and human negligence. The forest fire plays an important role in everyone's lives and also in our environment. Forest fire is an integral part of many ecosystems such as grassland, temperate forest etc. The ability to predict the area where the forest fire may occur will help in optimizing the situation. The paper presented the prediction of forest fire risk with the help of a machine learning algorithm by using meteorological data. The main objective of this paper is to predict the possibility of forest fire with its intensity in specific atmospheric conditions in a given location. We have made comparison of the performance analysis of the different machine learning classifiers. At the end of the abstract, we got the highest AUC value of 0.99 and classification accuracy of 0.98 using Decision Tree to predict the same.

Keywords— Forest fire, Prediction, Machine Learning, meterological data, Decision Tree

I. INTRODUCTION

Forest fires are a pre-eminent intrinsic phenomenon, especially occurring during summer [1][2]. Forest fires, frequently effective in regions with hot, sun-baked or Mediterranean climates, present a threat to life and enduring infrastructure. In addition, the after-effects of forest fires can have other wide-ranging repercussions. For instance, forest fires have a widespread effect on many chemical, mineralogical, and biological soil possessions[3][4]. Fatalistic effects caused by high levels of burn extremity include noteworthy removal of biological matters, worsening of both soil structure and permeability, notable loss of nutrients through evaporation, lixiviation, and erosion. Also, the release of precarious chemicals remarkably impacts human health and escalate the risk of future diseases. Therefore, the capability to prognosticate fire progression and area burnt is critical to all forest conflagrations' immediate and comprehensive consequences. Prompt diagnosis is the main element for a fortunate firefighting[5]. After all, conventional human monitoring is costly and is hugely affected by impressionistic factors, there always has been a significance to come out with automatic solutions[6][7]. These can be classified into three major groups: satellite-based, infrared/smoke scanners and local sensors (e.g. meteorological).

Satellites have accession costs, and the resolutions of the satellite photographs are not sufficient for all cases. In addition, scanners are also having high equipment and perpetuation costs. Arguably temperature and air humidity are the two most important factors which influences fire contingency[8][7]. Now that meteorological stations are available, data can be accumulated in real-time and are very cost effective[9][10]. Earlier, all these meteorological data used to be integrated into numerical indices, which are used for interception of fire danger and are used for interception of fire danger and support fire management resolutions such as prioritizing the necessary guidelines for safe and secure firefighting. In earlier days, different weather parameters were mostly collected manually in weather stations[9][11]. These indexes have been highly used worldwide in recent times, even if the climate differs from that of Canada. The FWI system corresponds with the fire activities in most of the European countries[12][13]. The dataset we have used is the meteorological data collected from Montesinho Natural Park, Portugal.

Objective: As there is no such literature that predict the depth of risk for this forest fire specifically the objective of this work is to predict the risk of forest fire by identifying the particular area as highly prone, moderately prone, low prone and no fire prone area.

Present Research: In this paper, we have used different Machine learning classifiers to predict the high risk zone as well as moderate, low and no risk zone also. Here a comparative study also has been done to understand which algorithm works best to predict forest fire with greater accuracy.

Contribution: This work can detect highly prone, moderately prone, low prone and no fire prone area. And it can predict for the same if different attribute are input to it.

Structure of the Paper:

Section 2 describes the background with a literature survey, AI models, and statistical parameters in this paper. section 3 describes the methodology, section 4 describes results with discussion, and section 5 concludes the present work.

II. BACKGROUND

2.1. Literature Review

In this section, we reviewed some literature and then we explained some AI models, specifically Machine learning models for classification. In the last part of this section, some statistical parameters are specified, which will be used for evaluation. Bui et al. [14] proposed a new ML methodology that is used to analyze and predict Forest fire spatial patterns with the case study of Vietnam. They have used the accuracy, AUC and some other statistical parameters to evaluate their model. They got an accuracy of 86.57% and AUC as 0.91. Tehranyet al. [15] designed a susceptibility map of a forest fire with some comparative case studies of some area of Vietnam. They used LEDT (LogitBoost ensemble-based decision tree) and got a good result. Phamet al. [16] evaluated different classifiers ‘ability to predict forest fire susceptibility. They got the AUC value of 0.96 using Bayes network and 0.94 using the Decision Tree model. These two are the first and second highest values of their evaluation parameter. They claimed that their model is very robust. There is no such literature which predicts the depth of risk for this forest fire specifically. For that reason, the objective of this present work is to predict the risk of forest fire by identifying the particular area as like highly prone, moderately prone, low prone and no fire prone area.

2.2. AI Models

Decision Trees

The tree has two entities, namely nodes and leaves. The leaves are the final decisions or the last results. In the places where the data is divided, the nodes are present.

KNN

K-Nearest Neighbor is a very basic Machine Learning algorithm that is used for classification. It works in the supervised learning domain.

SVM

SVM is used to train the model with training data by supervised learning, and the algorithm gives an optimal hyper-plane as output which gives good results for new test examples.

Naïve Bayes

Naive Bayes is a very powerful algorithm for all types of the classification task. Even if we have been working with millions of data with very few attributes, it is told to apply Naive Bayes. Naive Bayes gives very good results when it is used for textual data analysis, such as NLP.

2.3 Statistical Parameters

Following are the few statistical parameters that will be used to evaluate the classification result of this present work.

AUC: Area under curve.

CA: Classification accuracy.

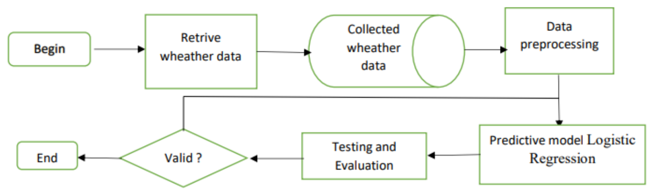
F-1: Harmonic mean of precision and recall, which is weighted.

Precision: The ratio of true positives among positive instances.

Recall: The ratio of true positives among all positive instances in the data.

III. METHODOLOGY

The methodology of machine learning technique



IJARSE
ISSN 2319 - 8354

3.1 Dataset description

For the current Research work We are building our own dataset. Factors in the dataset include oxygen, temperature, humidity, and fire occurrence. We summarized the attributes and corresponding values.

Applying machine learning techniques

Once the data is ready for modelling, we apply the most popular prediction In logistic regression, we fit a "S" shaped logistic function, which predicts two maximum values, rather than a regression line (0 or 1). algorithm Logistic Regression to predict the forest fire.

Logistic Regression

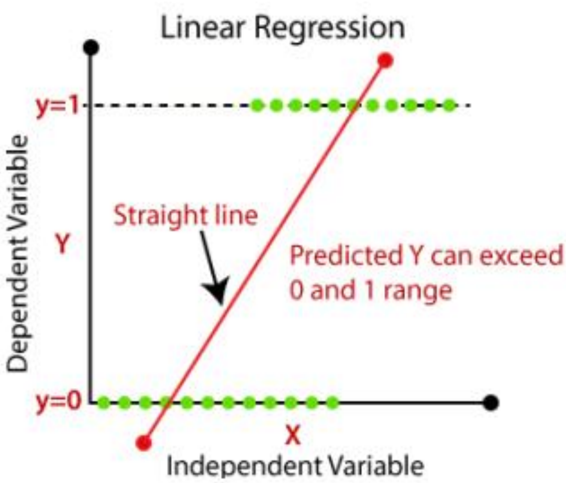
One of the most often used Machine Learning algorithms, within the category of Supervised Learning, is logistic regression. With a predetermined set of independent factors, it is used to predict the categorical dependent variable.

The output of a categorical dependent variable is predicted via logistic regression. The result must thus be a discrete or categorical value. It can be either True or False, Yes or No, 0 or 1, etc., but rather than providing the precise values of 0 and 1, it provides the probability values that fall between 0 and 1.

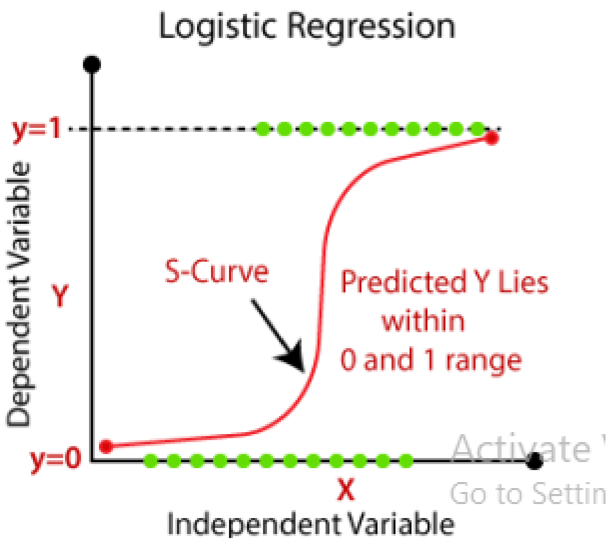
The main difference between linear regression and logistic regression is how they are used. In order to solve regression issues, one uses linear regression, whereas classification difficulties are solved using

Logisticregression. In logistic regression, we fit a "S" shaped logistic function, which predicts two maximum values, rather than a regression line (0 or 1).

Although it goes by the name of logistic regression and is used to classify samples, classification algorithms are what it comes under since it employs the notion of predictive modelling as regression.



IJARSE
ISSN 2319 - 8354



Types Of Logistic Regression

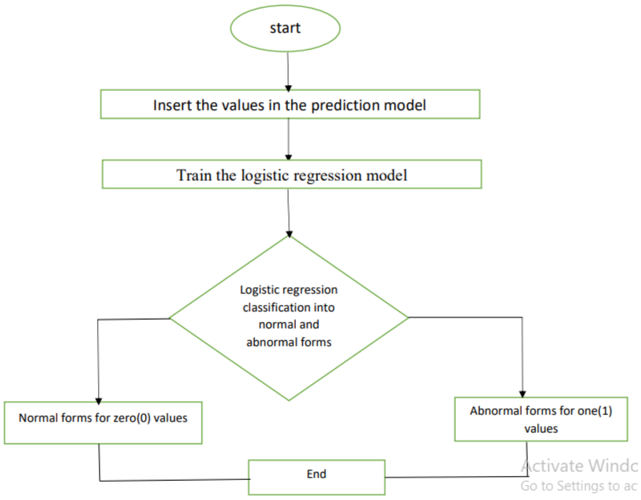
On the basis of the categories, Logistic Regression can be classified into three types:

Binomial: In a binomial logistic regression, the dependent variables can only be one of two potential kinds, such as 0 or 1, Pass or Fail, etc.

Multinomial: In multinomial logistic regression, the dependent variable may be one of three or more potential unordered kinds.

Ordinal:In ordinal Logistic regression, there can be 3 or more possible ordered types of dependent variables, such as "low", "Medium", or "High".

Algorithm of Logistic Regression



International Journal of Advance Research in Science and Engineering
Volume No. 12, Issue No. 04, April 2023
www.ijarse.com
IJARSE
ISSN 2319 - 8354

Dataset preprocessing

A. Identify the Headings

Data pre-processing is a data analysis technique used to convert the raw, manageable data into a useful and organized format.

1. Data Cleaning:

The data generally has many irrelevant and useless parts in it, and sometimes data can also be empty. To handle this, data cleaning is applied. It takes missing data and noisy data etc.

2. Data Transformation:

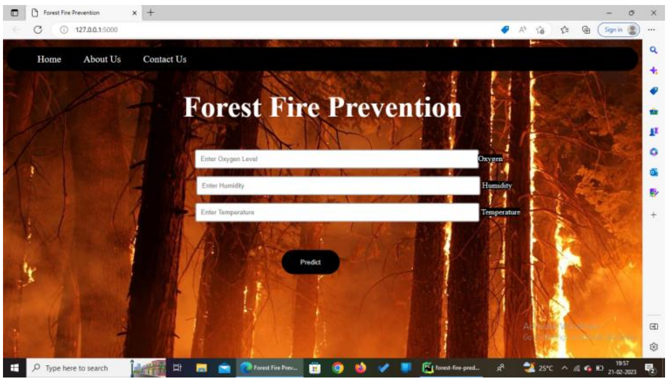
Data transformation transforms the data in the correct format suitable for the data mining process.

3. Data Reduction:

Data mining, a technique, handles a huge amount of data which is sometimes very hard to work with in some cases. To get rid of these unnecessary data, we use data reduction. It increases the storage efficiency and reduces analysis costs.

IV. RESULT & DISCUSSION

The dataset contains factors like oxygen, temperature, and humidity, which are the main factors that contribute to forest fires. By providing these values to the webpage (or) algorithm, the machine predicts the forest fire and outputs values between 0 and 1. This raises the question: Why does the machine predict values between 0 and 1 when the majority of probability is between 0 and 1? Because the algorithm we used is logistic regression, which predicts values between 0 and 1, the machine will respond The user must input the values into the model in order for it to predict the values and inform you whether the climate is nice or not. If the prediction is below 0.5, it is low; if it is above 0.7, there is a greater likelihood of a wildfire; and between 0.5 and 0.7, it is moderate. In this research, the machine only simply predicts the values; it does not intimate you whenever you want. A wildfire is more likely to occur in the summer and less likely to occur in the winter.



V. CONCLUSION

In this research project, we utilised logistic regression to forecast forest fires. Since logistic regression primarily relies on probability, the numbers that should be entered into the webpage we constructed in order for the algorithm to function properly are temperature, humidity, and oxygen.Because of temporal variations, these values do not remain constant. Because of the dynamic environment, unpredictability, and change of factors like humidity and temperature, predicting forest fires is a difficult endeavour. Accurate fire probability forecasting helps forest managers create more effective firefighting plans and restructure existing regulations for the sustainable management of forest resources.

REFERENCES

[1] V. Attri, R. Dhiman, and S. Sarvade, "A review on status, implications and recent trends of forest fireanagement," *Arch. Agric. Environ. Sci.*, vol. 5, no. 4, pp. 592–602, 2020.

[2] W. W. F. Deutschland, "Impacts of rewetting and reforestation on greenhouse gas emissions and removals in Sebangau National Park, Central Kalimantan, Indonesia, 2015-2018.," 2019.

[3] D. Pant, S. Verma, and P. Dhuliya, "A study on disaster detection and management using WSN in Himalayan region of Uttarakhand," in *2017 3rd International conference on advances in computing, communication & automation (ICACCA)(Fall)*, 2017, pp. 1–6.

[4] A. K. Gupta and S. S. Nair, "Environmental Knowledge for Disaster Risk Management, ekDRMSecretariat (GIZ- NIDM)." P, 2011.

[5] S. Parkash and S. S. Nair, "National Disaster Statistical System–An Initiative for Disaster Information Management with particular reference to Landslides," in *Proceedings of Central Building Research Institute (CBRI), Diamond Jubilee Conference on Landslide Present Scenario & Future Direction*, 2008, pp. 55–64.

[6] [6]A. K. Sharma, V. Joshi, and K. Kumar, "Use of remote sensing and GIS in disaster management in Gangtok area, Sikkim," *GIS Dev.*, vol. 6, no. 16, 2010.

[7] K. Konca-Kędzierska and K. Pianko-Kluczyńska, "The influence of relative humidity on fires in forests of Central Poland," *For. Res. Pap.*, vol. 79, no. 3, pp. 269–279, 2018.

[8] T. Cannon, J. Twigg, and J. Rowell, "Social vulnerability, sustainable livelihoods and disasters," *Rep. to DFID Confl. Humanit. Assist. Dep. Sustain. livelihoods Support Off.*, p. 93, 2003.

[9] E. Irzmańska, "The microclimate in protective fire fighter footwear: foot temperature and air emperature and relative humidity," *Autex Res. J.*, vol. 16, no. 2, pp. 75–79, 2016.

[10] K. V. S. Babu, "Developing Forest Fire Danger index using geo spatial techniques," 2019.

[11] D. De Rigo, G. Libertà, T. H. Durrant, T. A. Vivancos, and J. San- Miguel-Ayanz, "Forest firedanger extremes in Europe under climate change: variability and uncertainty." Publications Officeof the

European Union, 2017.

[12] M. Milenković, A. A. Yamashkin, V. Ducić, V. Babić, and Z. Govedar, "Forest fires in Portugal-the connection with the atlantic multidecadal oscillation (AMO)," *J. Geogr. Institute" Jovan Cvijic", SASA*, vol. 67, no. 1, pp. 27–35, 2017.

[13] L. Gigović, G. Jakovljević, D. Sekulović, and M. Regodić, "GIS multi-criteria analysis for identifying and mapping forest fire hazard: Nevesinje, Bosnia and Herzegovina," *Teh. Vjesn.*, vol. 25, no. 3, pp. 891–897, 2018.

[14] Bui, Dieu Tien, Nhat-Duc Hoang, and Pijush Samui. "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 (Viet Nam)." *Journal of environmental management* 237 (2019): 476-487.

[15] Tehrany, Mahyat Shafapour, et al. "A novel ensemble modeling approach for the spatial prediction of tropical forest fire susceptibility using logitboost machine learning classifier and multi-source geospatial data." *Theoretical and Applied Climatology* 137.1 (2019): 637-653.

[16] Pham, Binh Thai, et al. "Performance evaluation of machine learning methods for forest fire modeling and prediction." *Symmetry* 12.6 (2020): 1022.