IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

IJARSCT
Impact Factor: 6.252

Volume 2, Issue 5, May 2022

Performance Evaluation of Machine Learning for Forest Fire Modeling and Prediction

Vaidya Kalyani[1], Suryawanshi Shruti[1], Mulay Shrddha[1], Dr. N. P. Kulkarni[2]

Students, Department of Information Technology[1]

Guide, Department of Information Technology[2]

Smt. Kashibai Navale College of Engineering, Pune, Maharashtra, India

Abstract: *Nowadays, forest fires became one of the foremost important problems that cause damage to several areas around the world. The paper displays machine learning regression techniques for predicting forest fire-prone areas. The data set used in this paper is presented within the UCI machine learning repository that consists of climate and physical factors of the Montesano's park in Portugal. This research proposes machine learning approaches linear regression, Decision Tree, Neural Network and Random Forest algorithm with data set size 517 entries and 13 features for each row. This paper uses two versions, all features are included in the first, and 70% of the features were included in the second. The paper uses a training set which is 70% of the data set, and the test set is 30% of the data set. The accuracy of the linear regression algorithm gives more accuracy than other algorithms. So we proposed a system with the help of machine learning techniques and algorithms like Linear Regression, Neural Network Logistic Regression, Decision Tree and Random Forest to predict percentage of fire occurrence based on different parameters like temperature, wind, rain and oxygen data entered by the user in the front end..*

Keywords: Machine Learning, Linear Regression, Neural Network, Decision Tree, Random Forest, Forest Fire**.**

I. INTRODUCTION

One of the most extremely 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 are warming due to the increase in the average temperature of the earth [1], and human negligence. Dynamic Integrated Model of Climate and the Economy (DICE) indicates that the economy will lose about $23 trillion in the next 80 years due to the change in climate [6]. In Africa, South America, Southeast Asia, and New Zealand, forest fires occur due to human factors like husbandry of animals and agriculture [7]. Nowadays, there are various technologies for fire modelling to predict the spread of fires, such as Physical models and mathematical models [8]. These models depend on data collection during forest fires, simulation, and lab experiments to specify and predict fire growth in many regions. Recently, simulation tools have been used to predict forest fires, but simulation tools faced some problems such as the accuracy of input data and simulation tool execution time [7]. Machine learning is a sub-branch of Artificial Intelligence (AI) to learn computers aspect. Machine learning can be divided into two classes: supervised, unsupervised and reinforcement. In supervised learning, a supervisor is existed to give insights to the learning algorithm on how a decision or an action is bad or good. In supervised learning, the whole the data set is labelled completely.

Supervised machine learning algorithms are as linear regression, Random Forest, Neural Networks (NN) and Decision Trees. In unsupervised learning, the data set is not labelled. This leads that the algorithm must define the labels. The structure of the data set and the relationship between the features will be learned by the algorithm. Unsupervised machine learning algorithms are as k-means clustering and Self-Organizing Map (SOM). In reinforcement learning, the learning algorithm gets punished in case of a wrong action and gets rewarded in case of correct action.

In forest fire prediction, however, it is difficult to compile sufficient amounts of spatially explicit geo-environmental data, particularly over large-scale forests, due to field survey difficulties and budgetary constraints. Over the past decade, machine learning methods have successfully reached the primacy as a replacement to the traditional field-survey methods for the prediction of forest fire by elucidating the relationship between historic fire events and different explanatory variables in order to predict future fires [9]. Examples for machine learning methods suggested and used for forest fire prediction include decision tree based classifiers, artificial neural network (ANN).

II. RELATED WORKS

Dacre HF, Crawford BR, Charlton-Perez AJ, Lopez-Saldana G, Griffiths GH, Veloso JV, "Chilean wildfires [2] state that the forecast is a system that combines Artificial Intelligence (AI) and Geographic Information Systems (GIS). The forecast obtained more accurate results when compared to other random prediction models. A machine-learning algorithm based on Wireless Sensor Networks (WSN) to predict forest fires[2].

Al_Janabi S, Al_Shourbaji I, Salman MA [3] explains that There were three steps to design the probabilistic model. In step 1, the probabilistic model of forest fires was built from data of weather forecast and historical satellite. In step 2, the prediction of forest fires was produced using the data of the weather forecast as an input in the model of forest fires. In step 3, the warnings of forest fires were transported on different levels based on the need of the user. Machine learning models to predict the size of forest fires at the time of their inflammation [3]

Pérez-Sánchez J, Jimeno-Sáez P, Senent-Aparicio J, Díaz-Palmero JM, Cabezas-Cerezo JD [4] narrates that the results obtained using that intelligent system were better than using standard genetic programming. A novel system called forecast to predict the spread of forest fires in the future[4]

Jean-Luc Kouassi 1, Narcisse Wandan and Cheikh Mbow [5] elaborates that the modeling and the forecasting of the number of wildfires and burned areas at the watershed and ecoregion scale using SARIMA models indicate that the proposed models provide satisfactory results. These models present a good description and acceptable prediction performance of the pyro logical variables. Observed and fitted values of the numbers of wildfires and burned areas were very close, and the forecast errors highlight the accuracy of the models regarding their low RMSE. Forecasts for the numbers of wildfires and burnt areas over four years (48 months) seemed to be increasing slightly. These relevant and reliable forecasts should be used as a basis for sensitizing local populations and developing a decision-support tool for the management of the studied ecosystem and wildfire prevention in this ecosystem. Wildfire managers should share these forecasts with the local population and sensitize them to avoid activities that could ignite and spread wildfires[5]

Binh Thai Pham , Abolfazl Jaafari , Mohammadtaghi Avand , Nadhir Al-Ansari , Tran Dinh Du 5 Hoang Phan Hai Yen , Tran Van Phong [6] states that the accurate prediction of fire probability aids forest managers in drafting more efficient fire-fighting strategies and also helps to reorganize policies for sustainable management of forest resources. To achieve these, we evaluated and compared four fire predictive models derived from the BN, NB, DT, and MLR machinlearning methods for predicting and mapping fire susceptibility in the Pu Mat National Park, Vietnam. We formulated our modeling methodology based on processing the information from the historical fires and a set of spatially explicit explanatory variables. The outcome of the ROC-AUC method and several other performance metrics revealed that all four models developed in this study had high accuracy in predicting future fire susceptibilities (AUC > 0.90) in the Pu Mat National Park, although the BN model performed slightly better than the others[6]

George E. Sakr, Imad H. Elhajj, George Mitri and Uchechukwu C. Wejinya :The paper presented a forest fire risk prediction mechanism, based only on meteorological data and independent of any weather prediction mechanism. The results demonstrate the ability to predict forest fire risk with a limited amount of data and has shown that support vector machines can be used for a two-class prediction of fire risk with a very high accuracy of up to 96% for August as well as four classes prediction with a low error on the number of firesas well as on the predicted scale.[15]

A K Wijayanto, O Sani, N D Kartika, Y Herdiyeni**:** Here they proposed ANFIS algorithm can be used to generate expert system for predicting hotspot occurrences. The proposed method provided low error for training result (error = 0.0093676) and also low error testing result (error = 0.0093676). Attribute NEAR ROAD is the most determine factor that influences the probability of true and false alarm where the level of human activities in this attribute is high. The higher level of human activities, the higher probability of hotspot to be forest fire. This classification model can be used to build early warning system for forest fire.[16]

Mauro Castelli, Leonardo Vanneschi , and Ales Popovic**:** Here they showed a new intelligent GP-based system that makes use of these operators to examine burned area. The main objective was the development of a system for predicting the amount of area that will be burned during a forest fire, based on explicit relationships between meteorological data, forest-related data, and the amount of burned area. The comparatively small MAE obtained from experimental results showed that geometric semantic genetic programmingoutperforms.[17]

A.Kansal, Y. Singh, N. Kumar and V. Mohindru**:** In this paper, an algorithm for detection of fire has been proposed by using regression and dividing the datasets according to months. The algorithm achieves low root mean square error and

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 6.252

Volume 2, Issue 5, May 2022

high R-squared. The beauty of the algorithm lies in the way that it can give the result without doing the computation on whole dataset. In future, this approach can be extended by for other disasters as well. Application of certain transformation might also improve the model efficiency.[18]

L. Yu, N. Wang, and X. Meng: Ensemble learning is implemented at all the cluster heads which are used to collect and store the information that they acquire from the respective cluster. SVM which is a supervised machine learning technique is applied at the base station with polynomial kernel function. The sensors which are deployed can sense carbon dioxide, temperature, humidity and carbon monoxide. The data in tabular form or clustered data is generated by cluster. The SVM is applied after this to detect fire.[19]

Guruh Fajar Shidik and Khabib Mustofa**:** This research has proposed an alternative hybrid model capable of predicting the size of forest fire by combining Fuzzy C-Means and Back-Propagation Neural Network method. The model which incorporates meteorological and forest weather index variables (FFMC, DMC, DC, ISI, temperature, RH, wind and rain) has been shown to be successfully classify the level of burning into three categories: No Burn Area, Light Burn and Heavy Burn. The evaluation of the proposed model has showed promising results with accuracy of confusion matrix around 97.50% and Kappa 0.961.[20]

III. PROPOSED METHODOLOGY

The sequential steps to achieve the presented approach have been elaborated in the section below.

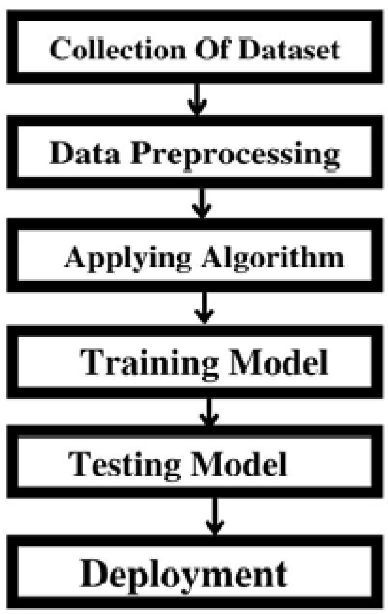

Fig.1 Flow of forest fire prediction Implementation

2.1 Collection of Dataset

The proposed technique requires the input of the user as well as the Forest Fire dataset for forest fire prediction. The data set is presented in the UCI machine learning repository. The size of the data set is 517 instances, and some attributes are 13. For this purpose a dataset containing reviews has been extracted from the URL

http://archive.ics.uci.edu/ml/machine-learning-databases/forest-fires/

The reviews in the dataset need to be preprocessed before providing it to the system to reduce the incidences of any error or redundancy that may impact the performance negatively. The preprocessing approach has been elaborated in the next step of the approach.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
IJARSCT
Impact Factor: 6.252
Volume 2, Issue 5, May 2022

2.2 Data Preprocessing

The preprocessing approach is the initial logical step of the approach that is designed to facilitate the processing of the reviews before it can be provided to the system. The extracted dataset achieved in the previous step is provided as an input in this step of the procedure. The presence of any unnecessary data can be detrimental to the system as it could result in an error reducing the accuracy.

Meteorological factors include the factors that are concerned with the processes and phenomenon of the atmosphere like relative humidity, rain, temperature. Topographical factors include the factors related to the arrangement or accurate representation of physical distribution of features of an area like land surface temperature (LST), burnt area. Vegetation factors like Normalized Difference Vegetation Index (NDVI) that can be used to analyse remote sensing measurements that can be used for assessing whether the target being observed contains live green vegetation. The components of Fire Weather Index (FWI) are meteorologically based components used worldwide to estimate fire danger. It consists of different components that account for the effects of fuel moisture and wind on fire behavior and spread. Calculation of these components is based on consecutive daily observations of temperature, relative humidity, wind speed, and 24-hour precipitation. They include Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), Drought Code (DC), and Initial Spread Index (ISI).

III. APPLYING ALGORITHM

Machine learning algorithms are programs that can learn from data and improve from experience, without human intervention. Learning tasks may include learning the function that maps the input to the output or learning the hidden structure in unlabeled data. Supervised learning uses labelled training data to learn the mapping function that turns input variables (X) into the output variable (Y). There are different machine learning techniques like Linear Regression, Decision Tree, Random Forest ,Neural Network. Here, the accuracy of these algorithms is measured based on the data using ten-fold method, as a result of which Linear Regression gives the best performance with highest accuracy range. Therefore, Linear Regression algorithm is used to build the proposed solution.

3.1Random Forest Algorithm

Random Forest algorithm is used to train and classify the prediction model. Random forests are an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and output the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

Algorithm 1 : Random Forest Algorithm

Input: Forest size (Number of decision trees)
Output: Predicted class (1 for fire predicted and 0 for no fire predicted)
Step 1 − Selection of random sample from a given dataset (Bootstrapping) of a given size.
Step 2 − The algorithm constructs a set decision trees (random forest) for bootstrap dataset based on the size of forest. Here the model is trained.
Step 3 − For every test instance, voting will be performed on each of the decision tree to generate the prediction result.
Step 4 − At last, select the most voted prediction result as the final prediction result.

3.2 Decision Tree

Decision Tree is a supervised learning technique that can be used for both classification and regression problems, but mostly it is preferred for solving classification problems. It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 5, May 2022
Impact Factor: 6.252

3.3 Neural Network

Neural networks, also known as artificial neural networks (ANNs) or simulated neural networks (SNNs), are a subset of machine learning and are at the heart of deep learning algorithms. Their name and structure are inspired by the human brain, mimicking the way that biological neurons signal to one another.

The equation for the neural network is a linear combination of the independent variables and their respective weights and bias (or the intercept) term for each neuron. The neural network equation looks like this:

$$\mathbf{Z = Bias + W_1X_1 + W_2X_2 + \ldots + W_nX_n}$$

where, Z is the symbol for denotation of the above graphical representation of ANN.
W is, are the weights or the beta coefficients
X is, are the independent variables or the inputs, and
Bias or intercept = W_0

There are three steps to perform in any neural network:

1.We take the input variables and the above linear combination equation of $Z = W_0 + W_1X_1 + W_2X_2 + \ldots + W_nX_n$ to compute the output or the predicted Y values, called the Y_{pred}.

2.Calculate the loss or the error term. The error term is the deviation of the actual values from the predicted values.

3.Minimize the loss function or the error term.

3.4 Traning Model

A training model is **a dataset that is used to train an ML algorithm**. It consists of the sample output data and the corresponding sets of input data that have an influence on the output. The training model is used to run the input data through the algorithm to correlate the processed output against the sample output. Here 70% dataset is used to train the dataset.

3.5 Testing Model

Once your machine learning model is built (with your training data), you need unseen data to test your model. This data is called testing data, and you can use it to evaluate the performance and progress of your algorithms' training and adjust or optimize it for improved results. Here 30% dataset is used to test the dataset.

VI .RESULT

The required packages of Python programming language has been installed and the code was implemented using jupyter notebook and the python code we have developed runs in Python 3.7, and below are the outputs of our forest fire prediction system model.

6.1 Forest is in Danger

Here user give the input values to FFMC=18.7, DMC=1.1, DC=7.9 , ISI =56.10, Temp=45 ,RH= 15.0, Wind= 0.50, Rain =0.0 parameters it shows the output Forest is in Danger that is fire is occuring .

Figure 2: Output Screen showing forest is in Danger

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Volume 2, Issue 5, May 2022

Impact Factor: 6.252

6.2 Forest is Safe

Here user give the input values to FFMC=50,DMC=200, DC=500, ISI =30,Temp=35, RH=50, Wind=8, Rain=2 parameters it shows the output Forest is safe that is fire is not occurring.

Figure 2: Output Screen showing forest is Safe.

V. CONCLUSION

In this research , the main idea is to perform machine learning algorithms to predict forest fires. The data set is presented in the UCI machine learning repository. The size of the data set is 517 instances, and some attributes are 13. Linear regression, Neural Network, Decision Tree and Random Forest machine learning algorithm are implemented to perform the prediction process. These algorithms are applied using two scenarios. In the first scenario, all attributes of the data set were included and in the second scenario, 70% of the attributes were included. The training set is 70% of the data set and the set and the test set is 30% of the data set it was 1,0.93,0.67 and 0.87 on linear regression , Neural Network, Decision Tree and Random Forest respectively , in testing data set it was 1, 0.90, 0.64 and 0.80 on linear regression , Neural Network, Decision Tree and Random Forest respectively . The experimental results demonstrated that the linear regression algorithm presented the best result.

REFERANCES

[1]. László F, Rajmund K, "Characteristics of forest fires and their impact on the environment", Academic and Applied Research in Military and Public Management Science, vol.15, pp.5-17, 2016.

[2]. Dacre HF, Crawford BR, Charlton-Perez AJ, Lopez-Saldana G, Griffiths GH, Veloso JV, "Chilean wildfires: probabilistic prediction, emergency response, and public communication", Bulletin of the American Meteorological Society, vol. 99, pp. 2259-2274, 2018

[3]. Al_Janabi S, Al_Shourbaji I, Salman MA, "Assessing the suitability of soft computing approaches for forest fires prediction", Applied Computing and Informatics, vol. 14, pp. 214-224, 2018

[4]. Pérez-Sánchez J, Jimeno-Sáez P, Senent-Aparicio J, Díaz

[5]. Palmero JM, Cabezas-Cerezo JD, "Evolution of burned area in forest fires under climate change conditions in southern Spain using ANN", Applied Sciences, vol. 9, pp. 4155, 2019.

[6]. Jean-Luc Kouassi 1 , Narcisse Wandan and Cheikh Mbow" Predictive Modeling of Wildfire Occurrence and Damage in a Tropical Savanna Ecosystem of West Africa" ,2020.

[7]. Binh Thai Pham , Abolfazl Jaafari , Mohammadtaghi Avand , Nadhir Al-Ansari , Tran Dinh Du 5 Hoang Phan Hai Yen , Tran Van Phong Performance Evaluation of Machine Learning Methods for Forest Fire Modeling and Prediction Duy Huu Nguyen , Hiep Van Le , Davood Mafi- Gholami , Indra Prakash , Hoang Thi Thuy and Tran Thi Tuyen , "Performance Evaluation of Machine Learning Methods for Forest Fire Modeling and Prediction" ,2020.

[8]. Ahmed M. Elshewey, Amira. A. Elsonbaty Forest Fires Detection Using Machine Learning Techniques ISSN No : 1006-7930 Volume XII, Issue IX, 2020

[9]. Ananya Shahdeo1, Aditi Shahdeo1, Prakruthi S Reddy1, Chaitra K2 Wildfire Prediction and Detection using Random Forest and Different Color Models -ISSN: 2395-0072 Volume: 07 Issue: 06 | June 2020

[10]. Jaafari, A.; Gholami, D.M.; Zenner, E.K. A Bayesian modeling of wildfire probability in the Zagros Mountains, Iran. Ecol. Inform. 2017, 39, 32–44.

[11]. Rodrigues, M.; de la Riva, J. An insight into machine-learning algorithms to model human-caused wildfire occurrence. Environ. Model. Softw. 2014, 57, 192–201
[12]. Zhang, G.; Wang, M.; Liu, K. Forest Fire Susceptibility Modeling Using a Convolutional Neural Network for Yunnan Province of China. Int. J. Disaster Risk Sci. 2019, 10, 386–403
[13]. Cortez P, Morais AD, "A data mining approach to predict forest fires using meteorological data, Environmental Science, 2007.
[14]. Özbayoğlu AM, Bozer R, "Estimation of the burned area in forest fires using computational intelligence techniques", Procedia Computer Science, vol.12, pp. 282-287, 2012.
[15]. Salis M, Arca B, Alcasena F, Arianoutsou M, Bacciu V, Duce P, Duguay B, Koutsias N, Mallinis G, Mitsopoulos I, Moreno JM, "Predicting wildfire spread and behaviour in Mediterranean landscapes", International Journal of Wildland Fire, vol. 25, pp. 1015-1032, 2016.
[16]. George E. Sakr, Imad H. Elhajj, George Mitri and Uchechukwu C. Wejinya "Artificial Intelligence for Forest Fire Prediction" 2010 IEEE/ASME International Conference on Advanced Intelligent Mechatronics Montréal, Canada, July 6-9, 2010.
[17]. A K Wijayanto , O Sani , N D Kartika , Y Herdiyeni "Classification Model for Forest Fire Hotspot Occurrences Prediction Using ANFIS Algorithm" 2017 IOP Conf. Ser.: Earth Environ. Sci. 54 012059.
[18]. Mauro Castelli, Leonardo Vanneschi , and Ales Popovic "PREDICTING BURNED AREAS OF FOREST FIRES: AN ARTIFICIAL INTELLIGENCE APPROACH" Fire Ecology Volume 11, Issue 1, 2015 .
[19]. A. Kansal, Y. Singh, N. Kumar and V. Mohindru, "Detection of forest fires using machine learning technique: A perspective," *2015 Third International Conference on Image Information Processing (ICIIP)*, Waknaghat, 2015, pp. 241-245, doi: 10.1109/ICIIP.2015.7414773.
[20]. L. Yu, N. Wang, and X. Meng "Real-time forest fire detection with Wireless Sensor Networkss," in Wireless Communications,Networking and Mobile Computing, 2005.
[21]. Proceedings. 2005 International Conference on, vol. 2. IEEE, 2005, pp. 1214--1 217.
[22]. Guruh Fajar Shidik and Khabib Mustofa "Predicting Size of Forest Fire Using Hybrid Model "ICT-EurAsia 2014, LNCS 8407, pp. 316– 327, 2014..
[23]. Salis M, Arca B, Alcasena F, Arianoutsou M, Bacciu V, Duce P, Duguay B, Koutsias N, Mallinis G, Mitsopoulos I, Moreno JM, "Predicting wildfire spread and behaviour in Mediterranean landscapes", International Journal of Wildland Fire, vol. 25, pp. 1015-1032, 2016