AN XGBOOST-BASED REGRESSION MODEL FOR WILDFIRE IMPACT PREDICTION

Mrs. Priyadharshini M[1], Chrisolus Timonsingh J[2], Insuvai V[3], Jenvin Shirly R[4]

[1] Assistant Professor, Department of Computer Science and Engineering
[2] UG Scholar, Department of Computer Science and Engineering
[3] UG Scholar, Department of Computer Science and Engineering
[4] UG Scholar, Department of Computer Science and Engineering
SRM Valliammai Engineering College, Chengalpattu, Tamil Nadu, India

---***---

Abstract - *AI is a powerful decision-making tool that makes use of decision makers to do intensive prediction and association tasks. The forest fire predictor plays a vital role in forest fire management. Timely prediction reduces the number of areas affected by this fire, lowering the cost of fire extinguishment and forest damage. This project presents a forest fire prediction mechanism based on artificial intelligence. Prediction can be done with the help of a supervised learning algorithm that entails a set of inputs with an expected output result being fed into the model, enabling it to be trained to identify trends and patterns. The machine learning algorithm works based on previous weather conditions in order to predict the fire hazard level for the day. Given that forest fires are rare, there exists only a few dataset instances, prompting us to devise a method for producing a reasonable prediction using a small and frequently skewed dataset. The easily measurable features are chosen in order to make the prediction, thus effectively reducing the cost of the system. In the past, meteorological data has been incorporated into numerical indices, which can be used for prevention and fire management The Canadian Forest Fire Weather Index (FWI) system, in particular, was designed in the 1970s when computers were scarce, necessitating only simple calculations using look-up tables with readings from four meteorological observations (temperature, relative humidity, rain, and wind) that could be manually collected in weather stations.*

Key Words: Impact Prediction, Risk Prediction, Logistic Regression, XGBoost, Wild Fire, Forest Fire, Initial Spread Index, Build Up Index

1.INTRODUCTION

[11] Prediction of events has always been a challenging task especially when it comes to natural events. Nature has always been tough to predict, which kicked off the curiosity to explore the predictability of wild-fires.[6] Wildfires are not common events, but unfortunately lead to costly damages and death when they occur. Meteorological data and national fire records show that the prime factor for wildfires is climate driven. [12] Prediction of occurrence of wild fire is proved feasible using the meteorological factors like temperature, humidity, wind speed and rain. Occurrence of wild-fires in Indian forests is not uncommon. [13] Most of the fires is of lesser intensity and generally put out by rain. But when there is less or no rain due to dry weather (summer), the fire spreads with ease. Most notable wild-fire events in India takes place in Uttarakhand, Karnataka and Odissa during the months of January to May.

The simplest practice is to keep track on the meteorological factors that heavily affects the ignition of fire. These factors include temperature, humidity, rain and wind speed. With the help of these factors, it is proposed that the prediction of wildfire is feasible. [14] Wildfire could be caused due to various factors. Most of the fire is caused by humans because of their carelessness. But during dry seasons, the fire is initiated by nature and when occurred it could cause serious damage. In India, forest fire frequently takes place but are mostly considered harmless. Most significant wildfire takes place in the place of Uttarakhand, Karnataka and Odissa which affected more than 10,000 acres of forest cover and caused damage to wildlife and vegetation in India. These fires are suspected to be caused due to dry weather. These fires could have been prevented or the damage caused could have been subsided if the occurrence of the fire and its impact was known before. From the research papers, we concluded that prediction of forest fire and its scale is possible using the meteorological factors. We consider 3 moisture codes as FFMC, DMC and DC which can be used to predict the occurrence of fire. In addition to this, there exist 2 more indices namely Initial Spread Index (ISI) and Build-Up index (BUI) that could influence the spread of fire. With the help of these 2 indices, it is possible to predict the area that could be affected by the fire even before it occurs.

1.1 System Overview

Wild-fire risk and impact prediction system consist of 4 components: Forest Department Database, Indices Calculator, Risk Predictor and Impact Predictor. The Forest Department Database acts as the repository for meteorological data of the forest cover. It should contain temperature, humidity, wind speed and rain in various part of the forest. It also contains the previous day's moisture codes such as Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC) and Drought Code

(DC), which plays a major role in calculating the current moisture codes which is carried out by the second component: Indices calculator. It not only contains the logic for calculation of moisture codes, but also can calculate the indices like Initial Spread Index (ISI) and Build-Up Index (BUI) which plays a major role in prediction of the impact of forest fire. The Risk Predictor and Impact Predictor are two main components of the system that works based on Machine Learning Models. The Risk Predictor is a simple classification model called as logistic regression model that predicts the likelihood of the ignition of fire using the moisture codes in a particular area. The output of this model is a probability of the occurrence of the fire. If the probability value is greater than 70-80 %, an alert is sent to the department through mail. This mail carries the information about the probability and the area that could be affected by the fire if it ever occurs. The Impact Predictor is also a machine learning model based on eXtreme Gradient Boosting (XGB) algorithm that follows ensemble technique of sequential learning which fits distributed gradient boosted decision trees. This model is responsible for the prediction of the area covered by the fire. This is possible by considering the ISI and BUI which are calculated from the moisture codes and meteorological dataFor demonstrating the working of the prediction model, an user interface is designed that could accept the weather data as user inputs as we currently doesn't have the actual database of the forest department. The figure – 1 represents the architecture of the proposed system.

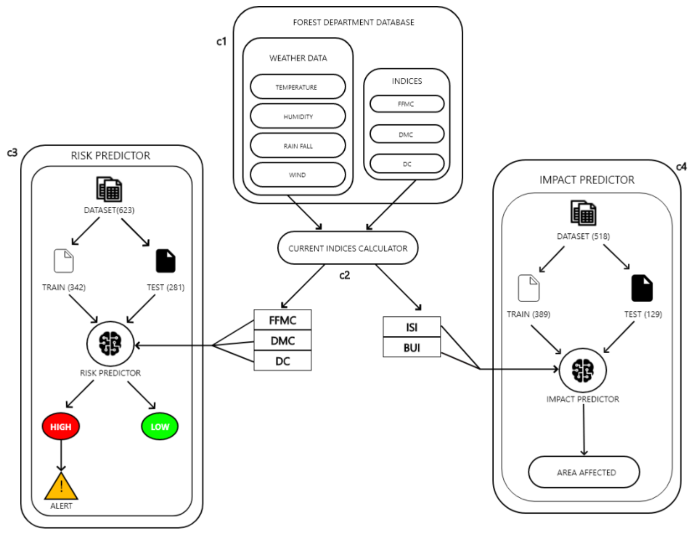

Fig – 1: System Architecture

1.2 Related Work

[1] Research on Event Detection for Forest Fire uses a Distributed Fixed Partitioning SVM at base stations to detect the events. The method presented above uses the compact representation model for categorization, which leads to energy efficiency in a distributed environment. The WSN constructed in combination of CluStream and SVM is efficient in performance and time with respect to other related works.

[2] As the result of a research conducted in Iberian Peninsula during the period of 2010-2014, proposed that a raise in temperature and increase in the duration and intensity of drought leads to larger wildfires, which threatened the environment as well as human lives. Upon further analysing, they managed to develop a model that performs prediction based on the surface temperature and soil moisture yielding accuracy of 83.3%.

[3] Fire detection systems have been promoted immensely in the past few years and have helped in the safety of people and property against fire hazards. The detection of fire hazards on the other hand can lead to unnecessary false alarms that can be very expensive if the occurrence happens in a commercial building. As well, false fire alarms have been a nuisance to the fire department and cause tie ups in resources and needless commotion that leads to panic. The problem that was addressed by this work was to detect fires and reduce the occurrence of false positives in a kitchen environment.

[4] Experiments are concluded for three partitions having a different number of training set instances and testing set instances for forest fire Prediction. Results that best accuracy achieved from model trained with sigmoid activation function is for dataset having5 40 training instances and 540 testing instances, with SinC function it is 85.42%, with radial basis function it is 84.95.

[5] Prediction of wild-fire using the meteorological factors has been proved feasible through a study that revealed the relationship between them. It will be helpful for forest department to prevent and rescue wildlife and resources by taking effective and appropriate measures in accordance to the scale of fire as predicted at its initial stage.

[6] Part-1 of a two-part paper is intended to review and categorise research in different fields of science and industrial projects that attempt to address wildfire issues. The topics include prediction and prevention means, detection methods, monitoring and surveillance techniques, suppression methods, allocation and mapping algorithms.

[7] In a decision tree-based system, they exposed an HLS based hardware implementation of a decision tree classifier as an IP core for forest fire prediction purposes. The designed DT_IP includes an AXI interface that allows its integration into several architectures based on various processors e.g., the ARM and Micro Blaze processors and also Open Sources processors, such as the Open RISC and NEO (by means of the wishbone to AXI4 protocol). The developed DT_IP has been integrated into the MicroBlaze based FFP SoC. The hardware implementation results show that the decision tree classifier is suitable for our purpose, since the

hardware implementation of the DT classifier requires few resources. In the other hand it gives significant performance.

[8] Proposal of effective detection of forest fire using Neural Network, analysed three algorithms U-Net, U2-Net and EfficientSeg. Models were trained using data augmentation techniques and two loss functions. Using Corsican Fire Dataset, EfficientSeg, U-Net and U2-Net showed F1 score of 0.95, 0.94 and 0.92 respectively. Upon analysing the result, the research concluded with proposing EfficientSeg as the best performing model for forest fire detection.

[9] AI-based 6 layered deep architecture model was evaluated by the RMSE score. The experimental works showed that LSTM based deep learning approach has potential in the use of the prediction of forest fires. Besides, the experiments showed that the proposed approach outperformed the other machine learning predictors.

[10] Either the classical or more recent machine learning approaches could be used for fire occurrence detection of peatlands. These approaches in general can be used in various types of fires, including bush fires, forest fires or peatland fires. When the data is unbalanced between classes, the accuracy of the prediction can be improved by pre-processing the data using SMOTE approach, to obtain a balanced sample, together with the application of the ensemble classification approach.

1.3 Existing System

Early research papers analysed the relation between meteorological factors like temperature, relative humidity, wind speed and rain and the occurrence, growth and spread of wild-fire. Upon analysing the event, it is found that the above-mentioned factors highly influence the fire. They also proved that the prediction of wild fire using such factors is feasible. Most of the existing systems were focused on detecting the fire after it is ignited and already started destroying the resources. It is done by collecting the forest fire images and applying various machine learning algorithms like Neural Network, Decision Trees etc. Some researches were focused on creating a curated, large-scale dataset using historical wild-fire aggregating nearly a decade of remote sensing data using which they analysed the performance of various machine learning models.

1.4 Proposed System

We believe that detection and prediction of the impact of the fire after its occurrence is not of great use, as most of the resources might already be damaged. So, we propose a prediction system based on machine learning technique to predict the risk of wild-fire even before the occurrence and computes the probability of fire using meteorological data. This helps the forest department to take necessary actions to control the fire if it ever occurs. In addition to this, we also propose an impact prediction system that could estimate the area that could be affected by the fire if the probability is higher. This describes the impact of the wild-fire even before it occurs. It is observed from the historical data, that not all wild-fire causes serious damage, as most of them covers lesser area. By predicting the impact, we can estimate the seriousness of the fire.

2. Implementation

The user is provided with a portal to predict the wildfire in an area. It can be done by inputting the meteorological factors like temperature, humidity, rain and wind speed. These meteorological factors can be used to calculate the 3 moisture codes (FFMC - Fine Fuel Moisture Code, DMC – Duff Moisture Code and DC – Drought Code) and 2 indices (ISI – Initial Spread Index and BUI – Build Up Index).

Table – 1: Indices and it's uses

Indicators	Full Form	What it Represents	Use	Range	Factors used for calculation
FFMC	Fine Fuel Moisture Code	Moisture content of the surface litter and other fine fuel	Assess potential Risk of Forest Fire	0 – 101	Temperature Humidity Rain Wind
DMC	Duff Moisture Code	Moisture content of organic soil layer or duff layer	Assess potential Risk of Forest Fire	0 – 100	Temperature Humidity Rain
DC	Drought Code	Drought Condition of forest cover	Assess potential Risk of Forest Fire	0 – 1000	Temperature Rain
ISI	Initial Spread Index	Rate of fire spread	Assess the fire spread	0 – 100	FFMC Wind
BUI	Build-Up Index	Amount of fuel available for spread	Assess the fire spread	0 – 300	DMC DC

The moisture codes affect the ignition in the forest covers and the indices help predicting the spread area of the fire. Prediction can be done using many Machine Learning Models. But before that we must process the data that we obtained from the internet. Using the processed dataset, we must train the Machine Learning model. We could use logistic regression and linear regression model for simple classification and regression respectively. After prediction is done, the result is displayed to the users on the same portal. To achieve this, we consider 6 modules: Data Collection and pre-processing, Indices Calculation, Building Risk Prediction Model, Building Impact Prediction Model, Building User Interface and Email Alert.

2.1 Data Collection and Pre-processing

Since our prediction system consists of two machine learning models, we have to collect suitable datasets for both of them, in-order to train them. The risk prediction model must be a binary classifier which means the result of the prediction should be 0 or 1; 0 representing no and 1 representing yes. The impact prediction model must be a regressor that predicts the area that could be affected by the fire. Since these two models highly differ from each other, the datasets that are used to train them has to be processed and transformed to make them suitable for each model. For training the classifier, we use a combination of two datasets that were found on Kaggle and UCI Machine learning repository. A suitable dataset should contain the following features. FFMC, DMC, DC and class.

First dataset which is obtained from Kaggle is Montesinho Natural Park Dataset that contains 337 records of historical wild-fire happened in 2007 and 13 features. However, the dataset cannot be directly used to train the classifier as the target feature is not of binary. Instead, it represents the area covered by the fire. We can conclude that, when the value is 0, there is no fire occurred. When the area > 0, we can infer that the fire occurred. So, during pre-processing, we convert the values in the feature from numerical to binary. Also, we remove the irrelevant and unwanted features like x-coordinate, y-coordinate, day, month, temperature, humidity, rain, wind speed and ISI. Now we are left with 4 features: FFMC, DMC, DC and class.

Second dataset which is obtained from UCI machine learning repository is The Algerian Forest Fire Dataset that contains 246 records. This dataset is obtained from two regions: Bejaia and Sidi-Bel Abbes during 2012. It contains 14 features. After removing the irrelevant features, we would be left with 4 features: FFMC, DMC, DC and class. Now we can combine these two datasets to create the suitable dataset which contains 623 record that can be used to train the classifier. For training the regressor, we make use of previously obtained dataset from Kaggle. Since the dataset is heavily skewed, the regression becomes difficult. Also, the dataset lacks an important feature, BUI, that highly affects the spread of wild-fire. It has to be included along the other features. This can be done by calculating BUI for every record in the dataset and appending the value to a new column named as BUI. Now after processing the dataset, we must remove the irrelevant features which results in the new dataset with 6 features: FFMC, DMC, DC, ISI, BUI and area.

2.2 Indices Calculation

The indices under consideration are based on Canadian forest fire weather index which include the 3 moisture codes and 2 indices. The calculation of these indices is provided by Canadian Forestry Services in 1984. The formulated equations make use of the meteorological data to compute the indices.

2.3 Building Risk Prediction Model

Since risk prediction is a simple classification problem, logistic regression model is used. The deciding factor of this classifier works based on the probability of the event. When the probability crosses the threshold, the result would be 1. If the probability stays behind the threshold, the result would be 0. Generally, for logistic regression model the threshold value is 0.5.

The dataset is divided into training and testing sets, containing 342 and 281 records respectively. After the model is trained, it is fed with the testing set and the prediction is compared with the actual result.

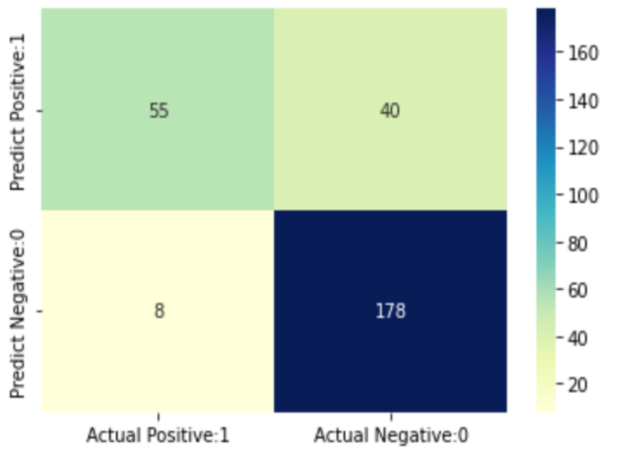

Fig-2: Confusion Matrix of Prediction Model

The above figure shows the confusion matrix of the model which can be used to deduce the performance of the trained model. Using this confusion matrix, the accuracy of the model can be calculated as Accuracy = (TP+TN)/(TP+TN+FP+FN). The accuracy of the risk prediction model is found to be 82.92 %

2.4 Building Impact Prediction Model

During risk prediction, the model produces a result that contains the probability of occurrence of the fire. If the probability is higher than 85%, then the fire is most likely to take place. So, we design an impact prediction model that could estimate the area that could be affected by the fire, if it ever occurs.

As mentioned before, the dataset obtained is heavily skewed; Out of 6 features, 4 were skewed. To reduce the skewness, we applied various transformation algorithms and observed the result. Boxcox transformation is found to be an effective transformation algorithm when I come to normalizing the left skewed data: FFMC and DC. ISI is slightly skewed to the right which can be handled by square root transformation. Area, a feature of chief importance which is considered as the target feature for the regressor, is heavily skewed to right. It is handled by logarithmic transformation.

Impact prediction is considered to be a simple regression problem, so we use linear regression with XGB framework, as XGB is proved to out-perform various other algorithms. After testing the model, it is found that the RMSE = 0.321307

2.5 Building User Interface

User has to be provided with an UI with which they could interact and perform prediction. To build an interface, we use web development tools such as HTML, CSS and JavaScript. It consists of 7 input fields: Temperature, Relative humidity, Wind Speed, Rain, FFMC, DMC and DC. The request and response to the server is handled by the python Flask library which is a web framework used for backend of the application. The values from the input are processed and fed to the model to make predictions and the result is reflected on the UI for the user to observe.

2.6 Email Alert

After prediction, if the probability of fire is found to be high, an alert is sent to the department through email containing the information about the impact and probability of the fire. This is achieved by integrating the Mail-Gun API in the Flask application. With a free subscription of $0 per month, mail gun allows the user to send 5000 mails/ month with a fixed domain name that cannot be changed unless we upgrade the plan. For a demo, there is no need for a perfect domain name. So, we continued with the free plan.

3. CONCLUSION

The proposed system uses logistic regression model to predict the risk of fire using the meteorological factors like temperature, humidity, rainfall and wind speed which are gathered from a particular area. Using the four collected meteorological factors, Fire weather Index (FFMC, DMC and DC) is calculated which represents the moisture content of various layers of the forest cover. Using this data, the model could predict the probability of ignition with an accuracy of 82.9%. If the risk is high an alert is sent to the provided email. It is observed that the probability of ignition is directly proportional to the temperature and wind speed and inversely proportional to the humidity and rainfall.

When the obtained probability from the risk prediction model is greater than 85%, the fire is most likely to occur and thus the impact prediction model that is based on linear regression, trained using XG-Boost algorithm, is used to predict the scale of the fire if it ever occurs. The RMSE value of the trained regression model is found to be 0.382.

The result of the proposed system is helpful in forest fire prevention and rescue. Fire Fighters will be able to take effective and appropriate measures if the risk of wildfire in an area is predicted. Thus, the proposed system makes use of the available meteorological factors collected from two different parts of the world to predict the impact of the wild-fire.

3.1 Future Scope

In addition to predicting the impact of the forest fire, we can also detect if there is any wildlife present in the possible spread area of the wildfire. This can be done by using remote sensing images in the predicted area along with the Neural Network to detect any wildlife in that area. If present, they can be rescued or drove to safer place. This helps reducing loss of life even before the fire occurs.

REFERENCES

[1] Yashwant Singh & et.al.,(2013), Distributed Event Detection in Wireless Sensor Networks for Forest Fires

[2] David Chaparro & et.al.,(2016).Predicting the Extent of Wildfires Using Remotely Sensed Soil Moisture and Temperature Trends

[3] K B Deve & et.al.,(2016), Design of a smart fire detection system

[4] B. K. Singh & et.al.,(2019).Extreme Learning Machine Approach for Prediction of Forest Fires using Topographical and Meteorological Data of Vietnam.

[5] Hao Liang & et.al., (2019), A Neural Network Model for Wildfire Scale Prediction Using Meteorological Factors

[6] Saeed Jazebi & et.al.,(2020).Review of Wildfire Management Techniques—Part I: Causes, Prevention, Detection, Suppression, and Data Analytics

[7] Faroudja ABID & et.al.,(2020), Decision Tree based System on Chip for Forest Fires Prediction.

[8] Rafik Ghali1 & et.al.,(2021).Forest Fires Segmentation using Deep Convolutional Neural Networks.

[9] Naaman Omar & et.al.,(2021), Deep Learning Approach to Predict Forest Fires Using Meteorological Measurement

[10] Dedi Rosadi & et.al.,(2021).Improving Machine Learning Prediction of Peatlands Fire Occurrence for Unbalanced Data Using SMOTE Approach

[11] Mauro Castelli & et.al., (2015) Predicting Burned Areas of Forest Fires: an Artificial Intelligence Approach

[12] Mutasim Mahmoud Al-Kahlout(2020) Neural Network Approach to Predict Forest Fires using Meteorological Data

[13] Krishna Chandra, Atul Kumar Bhardwaj (2015) Incidence of Forest Fire in India and Its Effect on Terrestrial Ecosystem Dynamics, Nutrient and Microbial Status of Soil

[14] Josep Piñol & et.al., (1998) Climate Warming, Wildfire Hazard, and Wildfire Occurrence in Coastal Eastern Spain