International Research Journal of Computer Science (IRJCS) ISSN: 2393-9842
Issue 04, Volume 6 (April 2019) www.irjcs.com

PREDICTING BURNED AREA OF FOREST FIRES

T.Niranjan Babu
Computer Science & Engineering,
JNTUA College of Engineering, Pulivendula, INDIA
tniranjanbabu@gmail.com

D.Swetha
Computer Science & Engineering,
JNTUA College of Engineering, Pulivendula, INDIA
swethadandu01@gmail.com

V.Charitha
Computer Science & Engineering,
JNTUA College of Engineering, Pulivendula, INDIA
vacharitha@gmail.com

A.J.Stephen
Computer Science & Engineering,
JNTUA College of Engineering, Pulivendula, INDIA
alphonsojostephen0406@gmail.com

Manuscript History
Number: **IRJCS/RS/Vol.06/Issue04/APCS10089**
Received: 04, April 2019
Final Correction: 10, April 2019
Final Accepted: 18, April 2019
Published: April 2019
Citation: T.Niranjan, D.Swetha, V.Charitha & A.J.Stephen (2019). PREDICTING BURNED AREA OF FOREST FIRES. IRJCS:: International Research Journal of Computer Science, Volume VI, 132-136.
doi://10.26562/IRJCS.2019.APCS10089
Editor: Dr.A.Arul L.S, Chief Editor, IRJCS, AM Publications, India
Copyright: **©2019 This is an open access article distributed under the terms of the Creative Commons Attribution License, Which Permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited**

Abstract: Forest fire causes serious damage to the Flora and fauna of a country. This is one of major environmental concern. Early prediction of fires saves large number of Flora and fauna and prevents the ecosystem. By predicting the area burnt we can also classify whether the fire into small or big. The key motivation for this prediction is to help fire management team in proper resource allocation and to help the fire-fighters in a best possible way. The meteorological conditions of the forest are the key factors of the forest fire. These climatic data can be obtained using the local sensors which are incorporated in the nearest meteorological stations. This research proposes various Machine learning approaches such as Naïve Bias, Decision Trees, SVR, Random forest, Stochastic Gradient Descent and Bagging for predicting the amount of land burnt in the forest. Here the predictive model is build using the outbreaks of fire caused in the northeast region of Portugal.

Keywords: Forest Fire, Weather, Artificial Neural Network, Climatic data, Portugal, Semantics

I. INTRODUCTION

The wildfire is one of the serious events that needs to be controlled which can cause enormous damage. During summer forest fire is known to be a prominent event. Around 4 million hectares of land were known to be burned every year and in the Mediterranean belt an average of 550 ha of land were burned. Smoke that emitted from the burned forest can serious health issues such as nausea, mental illness, nausea, heart attack and even death. The fire fighters are the people who are exposed to these gases easily so they should be provided with all the necessary safety measures. They are the one who fights to save the people life and ecosystem of the forest. From 1990 to 2006 21.9 % of fire fighters death occur red due to heart attack .The main idea of this paper is to help the (FMS)Fire Management System and the earlier prediction of fire in the forest help the forest managers in proper resource allocation such as providing enough air tankers and ground crews. There are many factors that can influence forest fire such as human intervention, Climatic changes, Forestry operations, Power line failure Improper forest policies , Lack of knowledge among the people to prevent the forest. This approach uses the FWI system in Canada, it requires only small calculations and vlook up tables from the variables such as temperature, Wind speed, Relative humidity and rainfall. These variables details can be collected easily from nearby weather stations. The application of this data mining predictive model helps in analyzing the logical statements if the fire occurred in region Y then its most likely to spread towards regionX. In this approach the data is modeled as a regression task where various algorithms are used and its efficiency is calculated by splitting the data into train and test.

© 2014-19, IRJCS- All Rights Reserved

II. LITERATURE SURVEY

Many Data mining and Machine learning techniques has been developed for the early detection of forest fire and estimation of the burned area in the forest. Forest fire is one major environmental issue that can cause serious damage to the ecosystem in various countries. The CFFDRS is one of the fire danger rating system and it has been under development since 1968.The first introduced subsystem of CFFDRS is (FWI) which provides the numerical readings purely based on the weather. In 1990 the second major subsystem was evolved by CFFDRS called Fire behavior system (FBP). The (FOP) The fire occurrence prediction is the recently introduced sub-system by CFFDRS. These are the systems used in the area of forest fire. Machine learning is a discipline that deals with programming the systems so as to make them automatically learn and improve with experience. Here, learning implies recognizing and understanding the input data and taking informed decisions based on the supplied data. It is very difficult to consider all the decisions based on all possible inputs. To solve this problem, algorithms are developed that build knowledge from a specific data and past experience by applying the principles of statistical science, probability, logic, mathematical optimization, reinforcement learning, and control theory. Machine Learning (ML) is an automated learning with little or no human intervention. It involves programming computers so that they learn from the available inputs. The main purpose of machine learning is to explore and construct algorithms that can learn from the previous data and make predictions on new input data.

III. ARCHITECTURE

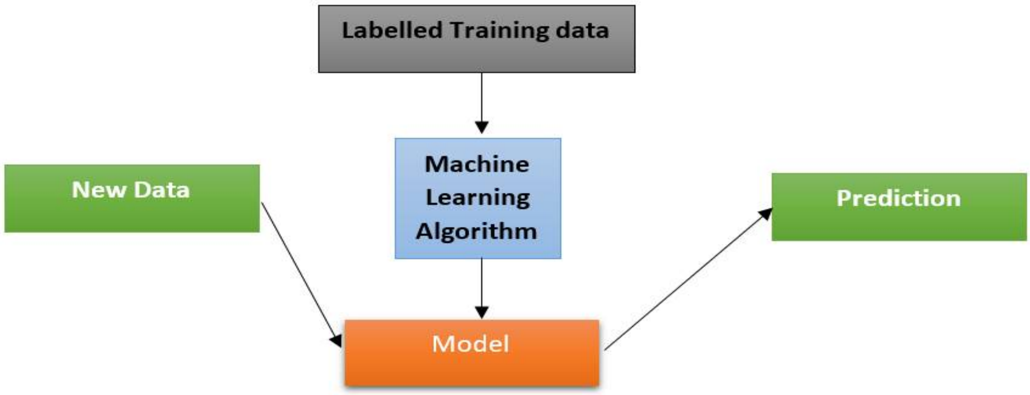

Fig A:Architecture

Forest Fire Data The forest Fire Weather Index (FWI) is the Canadian system for rating fire danger and it includes four components: Fine Fuel Moisture Code (FFMC),Duff Moisture Code (DMC), Drought Code (DC), Initial Spread Index (ISI). The first three are related to fuel codes: the FFMC denotes the moisture content surface litter and influences ignition and fire spread, while the DMC and DC represent the moisture content of shallow and deep organic layers, which affect fire intensity. The ISI is a score that correlates with fire velocity spread.. Although different scales are used for each of the FWI elements, high values suggest more severe burning conditions. Also, the fuel moisture codes require a memory (time lag) of past weather conditions: 16hours for FFMC, 12 days for DMC and 52 days for DC. This study will consider forest fire data from the Montesinho natural park, from theTr´as-os-Montes northeast region of Portugal. This park contains a high flora and fauna diversity. Inserted within a supra-Mediterranean climate, the average annual temperature is within the range 8 to 12◦C. The data used in the experiments was collected from January 2000 to December 2003 and it was built using two sources. The first database was collected by the inspector that was responsible for the Montesinho fire occurrences. At a daily basis, every time a forest fire occurred, several features were registered, such as the time, date, spatial location within a 9×9 grid (x and y axis), the type of vegetation involved, the six components of the FWI system and the total burned area. The second database was collected by the Braganc¸ a Polytechnic Institute, containing several weather observations (e.g. wind speed) that were recorded with a 30 minute period by a meteorological station located in the center of the Montesinho park. The two databases were stored in tens of individual spreadsheets, under distinct formats, and a substantial manual effort was performed to integrate them into a single dataset with a total of 517 entries. Number citations consecutively in square brackets [1].

IV. DATA ANALYSIS

The objective of data analysis step is to increase the understanding of the problem from the data. There are two approaches to describe a given dataset. Summarizing and Visualizing data. This dataset is public available for research. The details are described in [Cortez and Morais, 2007].The data can be used to test regression (difficult task), feature selection or outlier detection methods. There are 517 instances and 13 attributes in my dataset.

© 2014-19, IRJCS- All Rights Reserved

Dataset information

The attributes in the dataset include:
X - x-axis spatial coordinate within the Montesinho park map: 1 to 9
Y - y-axis spatial coordinate within the Montesinho park map: 2 to 9
Month - month of the year: "jan" to "dec"
Day - day of the week: "mon" to "sun"
FFMC - FFMC index from the FWI system: 18.7 to 96.20
DMC - DMC index from the FWI system: 1.1 to 291.3
DC - DC index from the FWI system: 7.9 to 860.6
ISI - ISI index from the FWI system: 0.0 to 56.10
Temp - temperature in Celsius degrees: 2.2 to 33.30
RH - relative humidity in %: 15.0 to 100
Wind - wind speed in km/h: 0.40 to 9.40
Rain - outside rain in mm/m2 : 0.0 to 6.4
Area - the burned area of the forest (in ha): 0.00 to 1090.84

Algorithms and Techniques

Regression: In regression tasks, the machine learning program must estimate – and understand – the relationships among variables. Regression analysis focuses on one dependent variable and a series of other changing variables – making it particularly useful for prediction and forecasting. The algorithms which are used in this regression problem are:

1. Linear Regression (Supervised Learning/Regression):

Linear regression is the most basic type of regression. Simple linear regression allows us to understand the relationships between two continuous variables. Real time example: Linear Regression can be used to predict the sale of products in the future based on past buying behaviour. Strength and Weakness: The main advantage is, the best fit line is the line with minimum error from all the points ,it has high efficiency but sometimes this high efficiency created disadvantage which is prone to over fitting of the data (i.e some noisy data also considered as useful data), and also it can't be used when the relation between dependent and independent variable is not linear.

2. Decision Trees (Supervised Learning – Classification/Regression)

A decision tree is a flow-chart-like tree structure that uses a branching method to illustrate every possible outcome of a decision. Each node within the tree represents a test on a specific variable – and each branch is the outcome of that test.

Real time example: Direct Marketing, Fraud Detection

Strengths: It is very easy to understand and interpret. The data for decision trees require minimal preparation.

Weaknesses: Sometimes decision tree may become complex. The outcomes of decisions can be based mainly on your expectations. So this can lead to unrealistic decision trees. Since a decision tree can handle both numerical and categorical data, it’s a good choice of algorithm. The goal is to create a model that predicts the value of target variable by learning simple decision rules.

3. Random Forests (Supervised Learning – Classification/Regression)

Random forests or ‘random decision forests’ is an ensemble learning method, combining multiple algorithms to generate better results for classification, regression and other tasks. Each individual classifier is weak, but when combined with others, can produce excellent results. The algorithm starts with a ‘decision tree’ (a tree-like graph or model of decisions) and an input is entered at the top. It then travels down the tree, with data being segmented into smaller and smaller sets, based on specific variables. Real Time Example: Random forest model can be applied in medical domain to identify a disease based on symptoms. Example: detection of Alzheimer's disease. Strengths and weaknesses: Random forest runtimes are quite fast, and they are able to deal with unbalanced and missing data. Random Forest weaknesses are that when used for regression they cannot predict beyond the range in the training data, and that they may over fit data sets that are particularly noisy.

4.Support Vector Machine Algorithm (Supervised Learning -Classification)

Support Vector Machine algorithms are supervised learning models that analyse data used for classification and regression analysis. They essentially filter data into categories, which is achieved by providing a set of training examples, each set marked as belonging to one or the other of the two categories. The algorithm then works to build a model that assigns new values to one category or the other. Strengths and Weakness of the model: It has a regularisation parameter, which makes the user think about avoiding over-fitting. It uses the kernel trick, so you can build in expert knowledge about the problem via engineering the kernel. It is defined by a convex optimisation problems (no local minima) for which there are efficient methods. The parameters for a given value of the regularisation and kernel parameters and choice of kernel, kernel models can be quite sensitive to over-fitting the model selection criterion.

© 2014-19, IRJCS- All Rights Reserved

5. Lasso Model: LASSO (Least Absolute Shrinkage Selector Operator): It uses L1 regularization technique. It is generally used when we have more number of features, because it automatically does feature selection. The black point denotes that the least square error is minimized at that point and as we can see that it increases quadratically as we move from it and the regularization term is minimized at the origin where all the parameters are zero. Strength and Weakness: The biggest issue of L1 penalization is that, as any dimensionality reduction algorithm, it might lose some relevant independent variables along the way. This mainly depends on how much penalized the system multi collinearity. L1 penalization tends to select one in a group of highly correlated variables. For many problems highly correlated variables should be selected or discarded as a group.

V. METRICS

Here we use mean squared error, variance and accuracy score as evaluation metric for predicting the best algorithm for this dataset.

Mean Squared Error: Mean Squared Error (MSE) it takes the average of the square of the difference between the original values and the predicted values. The advantage of MSE being that it is easier to compute the gradient. As we take square of the error, the effect of larger errors become more pronounced then smaller error; hence the model can now focus more on the larger errors.

Variance:

Variance is the amount that the estimate of the target function will change if different training data was used. The target function is estimated from the training data by a machine learning algorithm, so we should expect the algorithm to have some variance. Ideally, it should not change too much from one training dataset to the next, meaning that the algorithm is good at picking out the hidden underlying mapping between the inputs and the output variables. Machine learning algorithms that have a high variance are strongly influenced by the specifics of the training data.

Low Variance: Suggests small changes to the estimate of the target function with changes to the training dataset.

High Variance: Suggests large changes to the estimate of the target function with changes to the training dataset.

These metrics are helpful for this problem because of the following reasons:

i. It is a Regression based problem.

ii. RMSE will give an idea about how accurate the predictions are to actual values.

VI .PREDICTION

The implementation process can be split into two main stages. On data before processing and on data after processing & Normalization. Later after normalization of data we reached finally a best suited algorithm i.e., Linear Regression algorithm where we achieved least Mean Squared Error and high variance. Later we have converted variables in to binary classes so that we can find accuracy of different models like Decision tree, Naviebayes, SVM, SGD, Random forest and we find highest accuracy indecision tree with 62.820.

Model	Accuracy
SGD	51.28205128205128
Decision Tree	62.82051282051282
Navie Bayes	55.12820512820513.
SVM	58.97435897435898.
Random Forest	58.97435897435898.

Fig B: Accuracy result

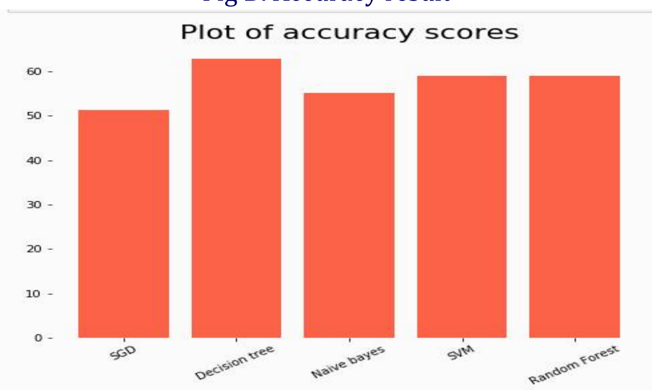

Fig C: Graph analysis

© 2014-19, IRJCS- All Rights Reserved

VII. CONCLUSION

After finding the best suitable regression then converting the target values to binary classes and finding accuracy among them, we conclude that Decision tree is best suited algorithm to predict the burn area of forest fire. The future work in this project can be done by creating probabilistic models that can identify the origin of fire by using some conditions. Those probabilistic models should be integrated with the model provided in this study to handle more risky conditions in the case of large or big fires. The use of GIS (Geographical Information System) data and satellite view can also be included with this model which provides better accuracy.

REFERENCES

1. L. Breiman. Random Forests. Machine Learning, 45(1):5–32, 2001.
2. L. Breiman, J. Friedman, R. Ohlsen, and C. Stone. Classification and Regression Trees. Wadsworth, Monterey, CA, 1984.
3. European-Commission. Forest Fires in Europe. Technical report, Report N-4/6, 2003/2005.
4. S. Taylor and M. Alexander. Science, technology, and human factors in fire danger rating: the Canadian experience. International Journal of Wildland Fire, 15:121–135, 2006.
5. D. Viegas, G. Biovio, A. Ferreira, A. Nosenzo, and B. Sol. Comparative Study of various methods of fire danger evalutation in southern Europe. International Journal of WildlandFire, 9:235–246, 1999.

© 2014-19, IRJCS- All Rights Reserved