

Forest Fire Prediction using Machine Learning Models based on DC, Wind and RH

Tanishka Jain, Nayan Sharma

Abstract: The paper points out forest fire prediction using machine learning models on the basis of viz. DC, Wind, RH out of the several machine learning classifier algorithms, It is relevant that random forest algorithm generates optimum accuracy(99.61%).

Keywords : forest fire, machine learning, DC, Wind, RH.

I. INTRODUCTION

A forest fire is an uncontrolled fire in an area of combustible vegetation occurring in forest areas[1]. The forest fire weather index(FWI) is an estimation of the risk of wildfire. The essential information needed to calculate this index is – Humidity, Temperature, Precipitation and max speed of avg. wind [2]. The factors includeFFMC, DMC, ISI, DC, Temp, RH, Wind. The FWI varies according to the region for eg., In France it ranges between 0-20 and in Canada it ranges from 0-30[3].

In the last 10 years there have been 67,000 wildfires annually and 7 million acres burned annually[4].

The description of the classifiers that we have used are:-

- **Random Forest**-It works on the principal of grouping individual decision trees and operating on it. Each tree outputs a class prediction and the class with highest number of votes becomes the model's prediction.
- **Random Tree** - A Random Tree builds on a random subset of columns .
- **HoeffdingTree** - It is an incremental decision tree algorithm that is used in learning from huge amounts of data streams.
- **K-Star**-It is an algorithm used for classification. We have to provide some labeled data at the start on which the model tries to understand and make sense.
- **LWL** - Locally Weighted Learning is a function approximation method. The prediction occurs using almost same local model around ongoing point of activity.
- **ADA BOOST** - It was basically called Ada Boost M1 by its authors and it is used to advance the performance of decision trees on binary classification questions.
- **Bagging** -It is a technique that groups meta-estimator that fit base classifiers on random subsets of fore runner dataset.

- **Decision Table**-They are algorithms whose result is a set of behavior. The data communicated in these tables could be shown as a decision tree.
- **PART** - It is a limited decision tree algorithm.
- **Logistics** - In this type of algorithm the output variable can hold discrete values for a set of inputs.

II. METHODOLOGY

The dataset for the project is taken from the UCI Machine Learning Repository.

<http://archive.ics.uci.edu/ml/datasets/Forest+Fires>. The dataset comprises of 517 instances and 12 input attributes plus 1 class attribute. The output is based on results on whether an area would catch fire or not.

For our study we have used Weka 3.8.4 tool to classify data for forest fire. On the basis of area we have mapped the values for DC, Wind and RH. Also Microsoft Excel has been used for graphical representation of the same.

III. EXPERIMENTAL RESULT AND ANALYSIS

To classify forest fire dataset various data processing algorithms have been used in this study. The dataset comprises of 517 instances of different factors involved in forest fires. The following table was generated to determine the accuracy of the dataset.

Table 1: Classification results using Weka 3.8.4

Classifier	CC	MAE	RMSE	RAE	RRSE	KS	CCI
trees.RandomForest	0.9961	0.0145	0.0724	0.2349	0.4178	0.9314	515
trees.RandomTree	0.9903	0.0104	0.0988	0.1677	0.5706	0.8435	512
trees.REPTree	0.9825	0.0225	0.1197	0.3633	0.6909	0.7183	508
trees.HoeffdingTree	0.9903	0.015	0.0969	0.2419	0.5593	0.8435	512
lazy.IBK	0.9883	0.0132	0.1076	0.2129	0.621	0.8176	511
lazy.Kstar	0.9941	0.0059	0.0762	0.0962	0.4398	0.9002	514
lazy.LWL	0.9845	0.0251	0.1156	0.4057	0.6676	0.742	509
meta.AdaBoostM1	0.9941	0.0085	0.0717	0.1379	0.414	0.9002	514
meta.AttributeSelectedClassifier	0.9922	0.0144	0.0877	0.2332	0.5064	0.8533	513
meta.Bagging	0.9845	0.0267	0.1072	0.4314	0.6187	0.742	509
rules.DecisionTable	0.9941	0.038	0.959	0.6155	0.5537	0.9002	514
rules.PART	0.9903	0.0132	0.097	0.2141	0.5601	0.8337	512
functions.Logistic	0.9883	0.0119	0.1078	0.1918	0.6223	0.8274	511

where,

CC(Correlation Coefficient)-

It is a measure that considers the power of relationship among the movements of a couple of variables.

MAE(Mean Absolute Error)-

It is measure of change among a couple of variables. It is also an average of absolute errors.

RMSE(Root Mean Squared Error)-

It is a constantly used measure of change among values anticipated by a model and the values detected.

Revised Manuscript Received on March 5, 2020.

Tanishka Jain*, Techno India NJR Institute of Technology Udaipur Rajasthan 313003 jaintanishka022@gmail.com

Nayan Sharma, Techno India NJR Institute of Technology Udaipur Rajasthan 31300

RAE(Root Absolute Error)-

It measures the moderate degree of errors in a set of predictions without taking into account their order.

RRSE(Root Relative Square Error)-

It is relative to what it supposed to be if a predictor had been used.

KS(Kappa Statistics)-

It is a measure that takes a roundabout figure into account by subtracting it from the predictor’s achievements.

CCI(Correctly Classified Instances)

The best classification result is observed from Random Forest classification algorithm with CCI equal to 515, also the value of Kappa Statistics > 0.8(i.e. 0.9314) and the Mean Absolute Error is 0.0145 which is less than that of REPTree(0.0225).

IV. GRAPHICAL REPRESENTATION OF ANALYSIS

a. DC

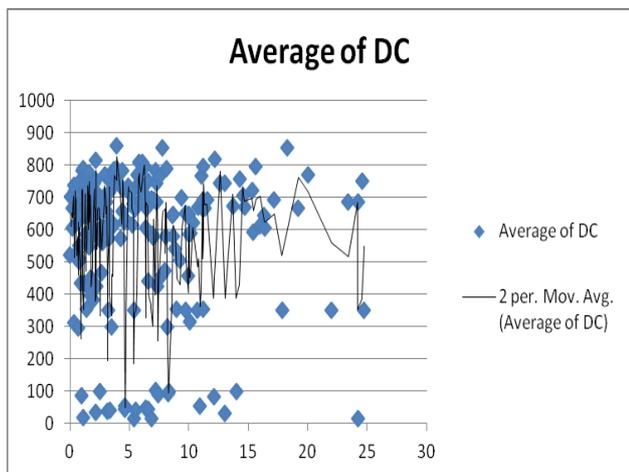


Figure 1: Average of DC w.r.t. Area.

From the above graph it is observed that when DC ranges from 400-800 the chances of a larger forest fire are prevalent whereas a DC ranges from 0-100 indicates that chances of a smaller forest fire.

b. WIND

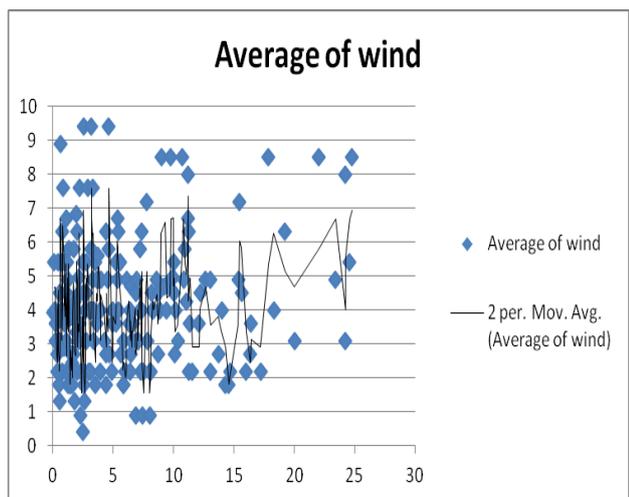


Figure 2: Average of Temperature w.r.t. Area

From the above graph it is observed that when the average wind speed is above 4.5 the chances of forest fire are prevalent whereas when average of wind speed is below 4.5 the forest fires tend not to grow in strength.

c. RH

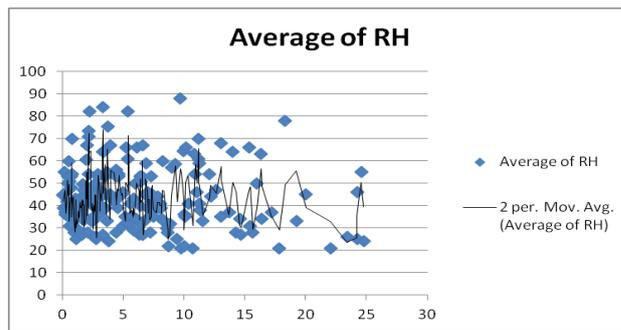


Figure 3: Average of RH w.r.t. Area

From the above graph it is observed that when RH increases the chances of forest fire are prevalent whereas when RH decreases the forest fires tend not to grow in strength.

V. ABBREVIATIONS

Abbreviation used in this paper are as follows:

- FFMC- Fine Fuel Moisture Code
- DMC- Duff Moisture Code
- ISI- Initial Speed Index
- DC- Drought Code
- Temp- Temperature
- RH- Relative Humidity
- Wind- Wind
- Area- Area

VI. CONCLUSION

The conclusion suggests that Random Forest algorithm gives the best result (99.61%) for the study, Also it portrays that DC, Wind and RH are the major factors for the study of Forest Fires.

REFERENCES

1. Cambridge Advanced Learner's Dictionary (Third ed.). Cambridge University Press. 2008. ISBN 978-0-521-85804-5. Archived from the original on 13 August 2009.
2. J.A. Turner, and B.D. Lawson (1978). Weather in the Canadian Forest Fire Danger Rating System. A user guide to national standards and practices. Victoria, British Columbia: Environment Canada, Centre de recherches forestières du Pacifique.
3. "Canadian Forest Fire Weather Index (FWI) System". Background Information. Natural Resources Canada. 2009. Archived from the original on 2011-07-06. Retrieved 2009-09-13
4. fas.org/sgp/crs/misc/IF1-244.pdf

AUTHORS PROFILE



Tanishka Jain, Department of Computer Science & Engineering, Techno India NJR Institute of Technology, Udaipur 313003



Nayan Sharma, Department of Computer Science & Engineering, Techno India NJR Institute of Technology, Udaipur 313003

