SUMMER INTERNSHIP PROJECT REPORT

FOREST FIRES

Submitted by(GRYFFINDOR)

Dinesh

Sowjanya

Archana

Jagadish

Aseeth Vardan

CERTIFICATE

This is to certify that , team GRYFFINDOR enrolled in the SIP programme offered by SMARTBRIDGE. We have successfully completed a summer project entitled ‘Forest Fires’ during the time period from 6th May, 2019 to 27th May, 2019 under the guidance of Dr.Mohan in the field Machine learning with python held at JNUTUH.

TABLE OF CONTENTS

1. Abstract
2. Introduction
3. Exploratory data analysis
4. Results and Discussion
5. Conclusion
6. Future Perspective
7. References

ABSTRACT

Forest fires importantly influence our environment and lives. The ability of accurately predicting the area that may be involved in a forest fire event may help in optimizing fire management efforts. Given the complexity of the task, powerful computational tools are needed for predicting the amount of area that will be burned during a forest fire. The purpose of this study was to develop an intelligent system based on genetic programming for the prediction of burned areas, using only data related to the forest under analysis and meteorological data. We used geometric semantic genetic programming based on recently defined geometric semantic genetic operators for genetic programming. Experimental results, achieved using a database of 517 forest fire events between 2000 and 2003, showed the appropriateness of the proposed system for the prediction of the burned areas. In particular, results obtained with geometric semantic genetic programming were significantly better than those produced by standard genetic programming and other state of the art machine learning methods on both training and out-of-sample data.

INTRODUCTION

A forest fire is a natural disaster consisting of a fire which destroys a forested area, and can be a great danger to people who live in forests as well as wildlife. Forest fires are generally started by lightning, but also by human negligence or arson, and can burn thousands of square kilometers.

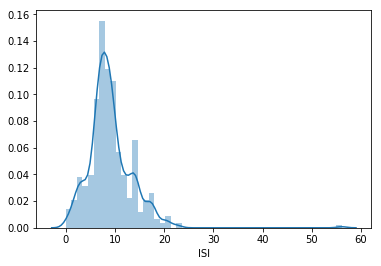
Forest fires, also known as wildfires, vegetation fire, grass fire, brush fire or bush fire, is common in vegetated areas of Australia, South Africa, United States and Canada, where climates are sufficiently moist to allow the growth of trees, but feature extended hot and dry periods.

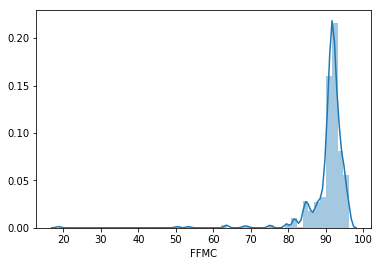
Forest fires are caused by the drying out of branches, leaves and therefore becomes highly flammable.

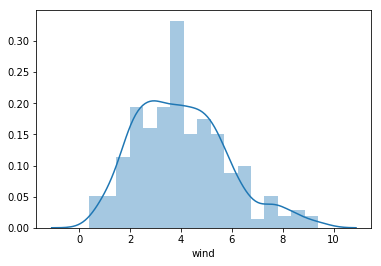


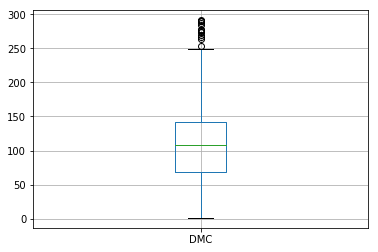
Exploratory data analysis :

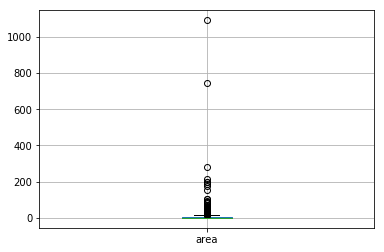
1.Distribution plots

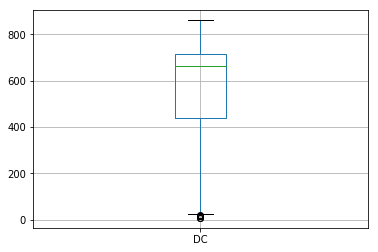




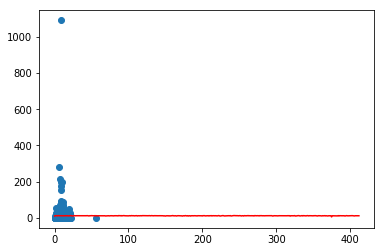
Boxplots:

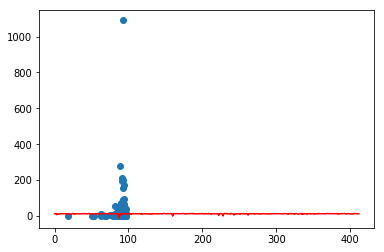


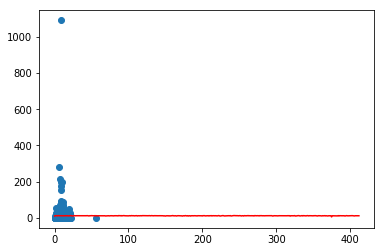
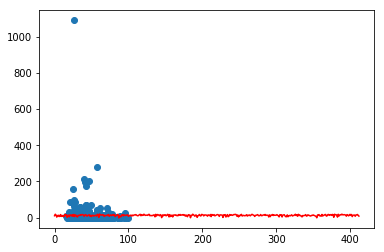




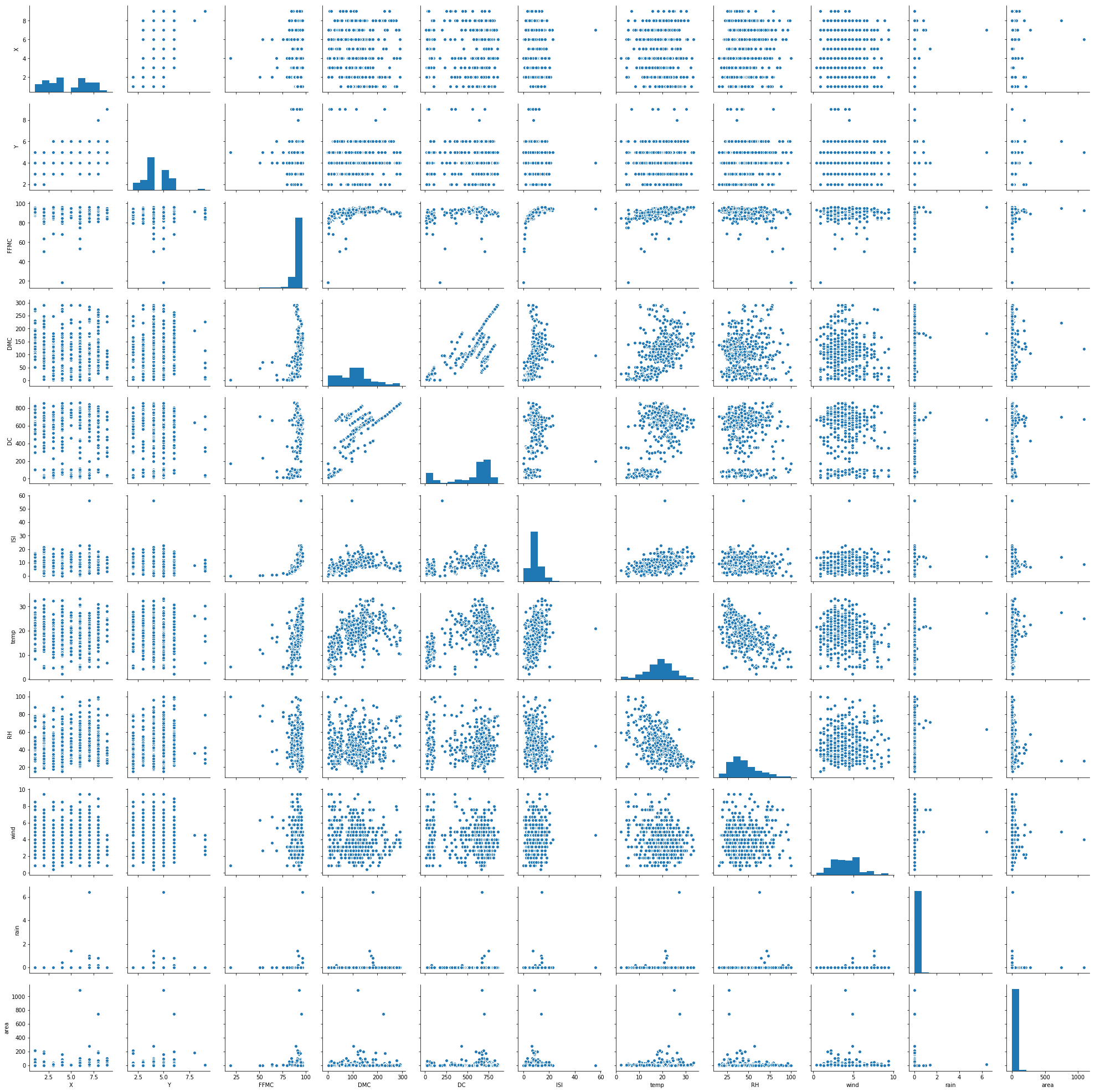
Scatter plots:





Pairplots:



Result Analysis:

Linear Regression:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| s.no | Features | Accuracy | RMSE | R2score |
| 1 | FFMC | 0.24 | 62.86 | 0.24 |
| 2 | DMC | 1.05 | 62.61 | 1.05 |
| 3 | DC | 0.25 | 62.86 | 0.25 |
| 4 | ISI | -0.11 | 62.98 | -0.11 |
| 5 | temp | 1.66 | 62.41 | 1.66 |
| 6 | RH | 0.85 | 62.67 | 0.85 |
| 7 | wind | 0.01 | 62.94 | 0.01 |
| 8 | rain | -1.7 | 63.5 | -1.7 |

Multi Linear Model:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| s.no | Features | Accuracy in % | RMSE in % | R2score in % |
| 1 | DC,temp | 1.22 | 75.87 | 1.22 |
| 2 | DMC,RH | 1.52 | 75.75 | 1.5 |
| 3 | X,Y,DMC | 1.23 | 75.86 | 1.23 |
| 4 | FFMC,DMC,RH | 2.98 | 61.86 | 2.98 |
| 5 | FFMC,RH,temp,wind | 2.26 | 62.86 | 2.2 |
| 6 | FFMC,RH,temp,wind,dmc | 3.26 | 61.99 | 3.26 |
| 7 | FFMC,RH,temp,wind,  DMC,ISI,rain,DC | 3.69 | 61.7 | 3.61 |

Considering the regression models we cannot get best accuracy because of unnormalized data. So the output variable area is taken into two classes. The area which is equal to zero in one class and greater than zero in another class and thus it forms a dichotomous data. Therefore classification technique can be applied.

The results using classification techniques are:

|  |  |
| --- | --- |
| Algorithm | Accuracy |
| Logistic Regression | 54.92 |
| SVM | 40.38 |
| Random Forest | 55.92 |

Conclusion:

Finally the best model is retrieved in Linear Regression after normalizing the data i.e. removing all the outliers.

The normalization is done using random forest classifier.

Entire data is considered in this scenario. because of splitting there is a lot of deviation in accuracy.

The accuracy after applying RFC is 88.7%

RMSE error is 34.41

Future perspective

Till now, the study was mainly focused on identifying the burnt area in the forest and in knowing how far it affected the forest and results in deforestation and forest degradation as well. The features considered are FFMC,DC,DMC,ISI, temperature, wind, rain and humidity.

In future, more analysis needs to be done on the prevention of this forest fires. Afforestation techniques are to be implemented strictly.

References :

1. [www.kaggle.com](http://www.kaggle.com)

2.<http://www3.dsi.uminho.pt/pcortez/forestfires/forestfires-names.txt>

3. <http://cwfis.cfs.nrcan.gc.ca/background/summary/fwi>

4. <https://towardsdatascience.com/>