Prediction on Amazon Wildfire

Ahmad Zulfahmi bin Harum, Hifdzul hadi bin Daud, Izzat Asyraf bin Haniz,

Muhammad Amirul Aizat bin Zaidi

Department of Computer Science, Kulliyyah of Information and Communication technology,
International Islamic University Malaysia, Kuala Lumpur, Malaysia.

Abstract— For over a decade, the year 2019 saw the worst fires to hit the Amazon Basin. The Amazon wildfire have become a hot topic that has gained much attention on the international stage because today the environmental issues are becoming increasingly political and is concerned by many. Why did this fire happened? How does a forest fire even starts in a rainforest? Are there ways to prevent the fire from happening before it even starts? The aim of this project is to make prediction regarding the occurrence of forest fires in the Amazon forest so that early prevention actions can be taken in order to ensure the unwanted losses are not happening. To make this aim a reality, data analytics approach will be implemented in this project. The algorithm that will be utilized is linear regression which will carry out the predictive analysis on the dataset that had been obtained from the website that offers various public datasets to be used for analysis purpose.

Keywords— Prediction on Amazon Wildfire, Wildfire, Linear Regression, Prediction.

I. INTRODUCTION

The Amazon Rainforest that covers mostly of Amazon basin gives a big impact on the ecosystem of the earth. The moist broadleaf tropical rainforest is part nine big countries; Peru, Colombia, Venezuela, Ecuador, Bolivia, Guyana, Suriname, French Guiana and Brazil[1]. It is located in a big part of the Brazil and cover small parts of Venezuela, Ecuador, Bolivia, Guyana, Suriname and French Guiana. The forest is widely known for its 5,500,00km squared and the biodiversity tract. It is a home to 390 billion individual trees from 16,000 species[2]. Hence, it is the biggest natural air purifier that produces oxygen regularly for mortal to breath in.

This Eocene era formed rainforest is also well known for wet tropical type of forest, which contributed to its characteristic of most biodiverse rainforest. It homes one tenth species of animals that makes it the place for the largest collection of living plants and animal species in the world[3]. Moreover, it is important to preserve our nature for the earth and living things.

In 2019, the rainforest has faced a big catastrophe of wildfires that damaged 906 hectares of its area [4]. Due to the catastrophe, lots of plants has been vanished from the forest and the animals that occupied it has lost their natural home. The causes of the wildfires was said to be climate change and the deforestation that turned the forest to be in a very hot temperature and has sparked fires in the wet tropical rainforest. Humans are also affected by this nature apocalypse that has caused reduction in oxygen production from the rainforest.

This report is to study the wildfires data that have been collected and produces predictions on future wildfires occurence that will cost a lot of damage to nature and the community using the data analytics methodology. By producing a prediction on the future wildfires, related authorities can make use of the data to take prevention and preservation steps to stop wildfire from occurring in the Amazon rainforest from this study. Furthermore, the authorities can also be prepared if the predicted wildfire occurrence happens and prevent it from producing a massive damage to the forest.

The algorithm that will be used for this research is the linear regression that will do the predictive analysis on the data acquired by the team through online dataset engine. A linear regression model that is built for this study assumes the relationship between the dependent variable y and the p-vector of regressors x is linear. It is formularized for prediction purposes and it suits the objective of this project to predict the occurrences of the wildfire in Brazil Amazon forest.

II. RELATED WORKS

Year	Author/ Title	Main Techniques	Results	Future works
2018	Sun, Ping, and Yunlin Zhang A Probabilistic Method Predicting Forest Fire Occurrence Combining Firebrands and the Weather-Fuel Complex in the Northern Part of the Daxinganling Region, China	Probabilistic Method Predicting Fire Occurrence Probability	daily fire probability ranges from 3.76 to 5.5%.	Use of probabilistic method to predict the fire rate occurrences in based on a specific area
2011	Youssef Safi and Abdelaziz Bouroumi A Neural Network	Artificial Neural Network	25% error rate, 72% accuracy	Use genetic algorithms in order to

	Approach for Predicting Forest Fires			optimize the architectural parameters of the network
2018	Hevelyne Henn da Gama Viganó , Celso Correia de Souza , José Francisco Reis Neto , Marcia Ferreira Cristaldo, Leandro de Jesus Prediction and Modeling of Forest Fires in the Pantanal	Multivariate analysis	41% of variance of the number of foci	non-efficient technique for prediction, Avoid multivariate analysis for forest fire prediction
2017	A K WijayantoSani, N D Kartika , Y Herdiani Classification Model for Forest Fire Hotspot Occurrences Prediction Using ANFIS Algorithm	Classification	low error for training result (error = 0.0093676) and also low error testing result (error = 0.0093676)	Use of the attributes that might determine probability of forest fire occurrences like land use type, land cover type, and soil type

III. DATA DESCRIPTION

For this project, we have obtained our dataset from https://www.kaggle.com/ , this website offers thousands of public datasets that can be shared. We choose dataset named "Forest Fires in Brazil" by Luis Gustavo Modelli which basically recorded about the number of forest fires reported in Brazil according to each state. The number of fire reports in this dataset are collected from year 1998 to 2017.

The dataset only have five features with 6454 rows. The features are Year, State in Brazil, Month, Number of forest fire reported, and Date reported.

IV. EXPERIMENTAL SETUP

A. Algorithm – Linear Regression

Linear regression attempts to model the relationship between two variables by adjusting the linear equation to the observed data. One variable is considered to be an explanatory variable, while the other is considered to be a dependent variable. For example, a modeler might want to relate the weights of individuals to their heights using a linear regression model.

The representation is a linear equation that combines a specific set of input values (x) the solution to which is the predicted output for that set of input values (y). As such, both the input values (x) and the output values are numeric.

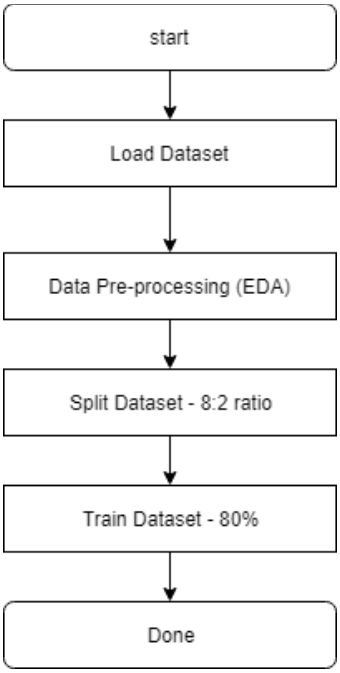
The linear equation assigns one scale factor to each input value or column, called a coefficient and represented by the capital Greek letter Beta (B). One additional coefficient is also added, giving the line an additional degree of freedom (e.g. moving up and down on a two-dimensional plot) and is often called the intercept or the bias coefficient.

B. Data Cleaning

The datasets to be used is first examined, and it is found that the data is already clean because there were no missing values in the datasets. Thus the data does not need to be clean.

V. MODELLING

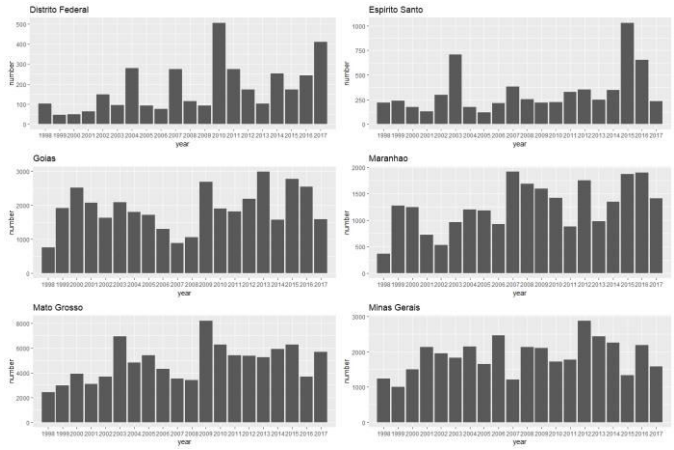
After cleaning the datasets, and choosing the sample to be used, we will proceed to the modelling part. The algorithm used for modelling the dataset is linear regression. Several steps are taken for the modelling process:



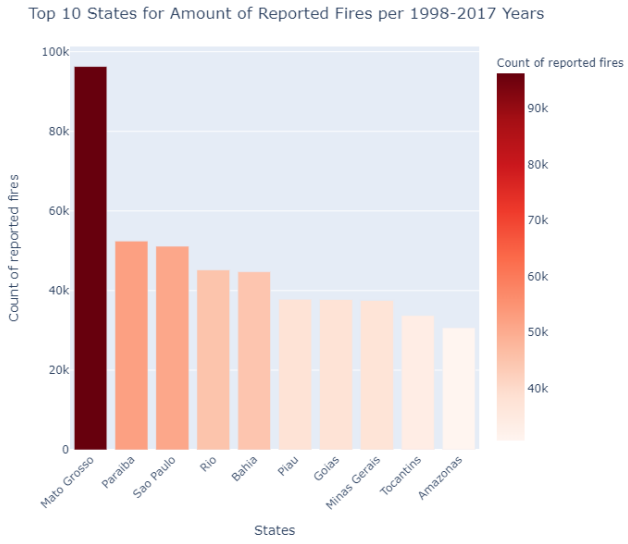
Based on the above flowchart, the first step is to load the dataset. Before modelling starts, data pre-processing needs to be done to know more about the dataset. Exploratory Data Analysis is done first to identify outliers, and the distribution of the data. After finishing with EDA, the dataset is split randomly into training set and test set. The training set is to be 80% of the whole dataset, whereas the test set is 20%.

A. Exploratory Data Analysis (EDA)

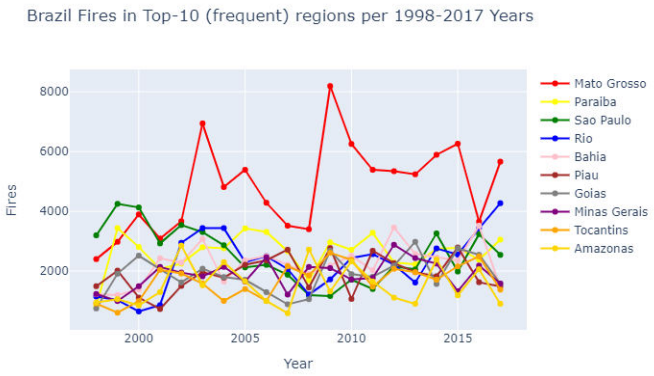
Exploratory Data Analysis is done to understand the dataset, to learn what we can gain from the dataset. The data is classified into each state and the number of occurence in recorded based on each year from 1998 to 2017.



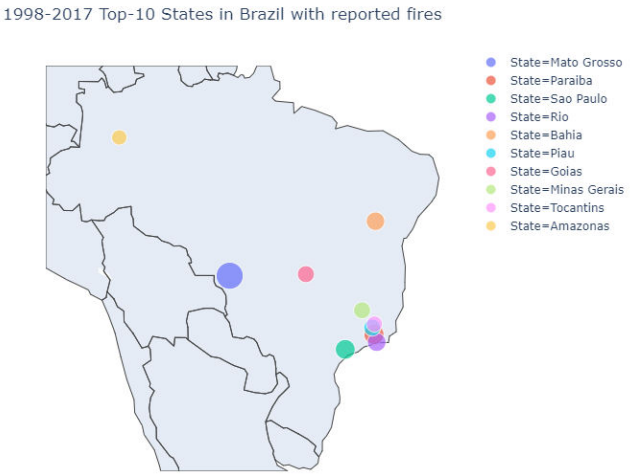
Depiction of fire occurrences in Brazil from year 1998 to 2017. We can see an increment in number of Forest Fire since 2007. It has increased Steadily. We can also see year 2003 as an exception or outliers. But since 2007 Forest Fire is surging.



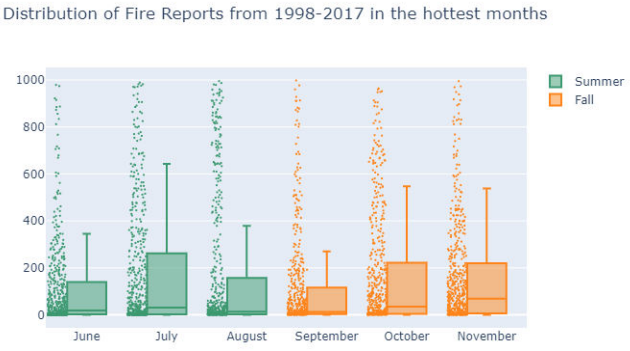
Fire occurrences reported in every state in Brazil



Frequencies of fire occurrences in states of Brazil according to years



Fire occurrences depiction in Brazil through geographical map.

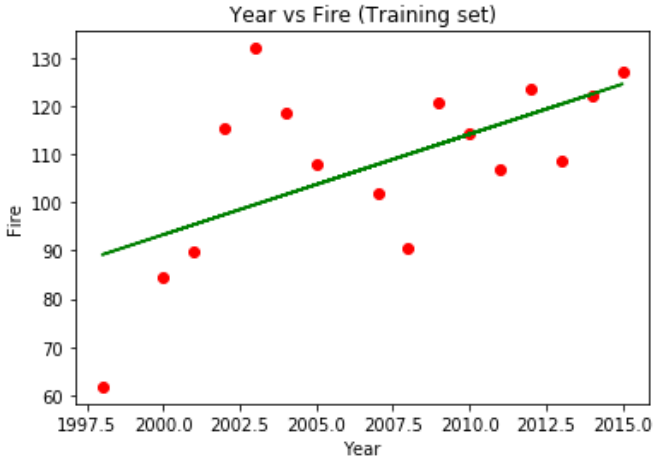


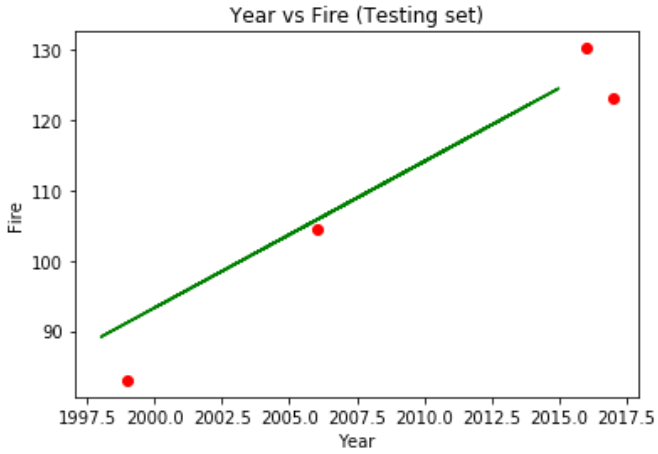
Fire report distribution according to months in years recorded.

B. Train Dataset

After analysing the dataset with EDA, the dataset is then split into two sets; one set for the training phase and another one set for the testing phase. The data is split into 8:2 ratio, where 80% of the data is to be the training set, and the rest as test set. Doing it this way, we will have the model predicted values for the 20% data (test) as well as the actuals (from the original dataset). By calculating Mean Absolute Error, Mean Squared Error, Root Mean Squared Error, we can find out the accuracy for the linear regression model.

```
Mean Absolute Error: 4.72595692217002
Mean Squared Error: 28.72340171616578
Root Mean Squared Error: 5.359421770691852
```





VI. Analysis of Results

A. Hypothesis Testing

The hypothesis that we came up with is the occurence of forest fire will increase for the next few years because of the massive fire that happened in 2019. Based on our prediction using training and testing data from our linear regression model. Our model predicted that there will be an increase in the number of fire from 2018 to 2022. The predicted result is shown in table below.

Year	Predicted number of fire
2018	131
2019	132
2020	135
2021	137
2022	139

VII. Prescriptive Analytics

Prescriptive analytics enable us to transform descriptive data into action that can be taken in the future. Based on our project, we can identify which locations that have the highest risk of forest fire using predictive analytics, then estimate the impact of forest fire in certain location in Brazil. This data can be used to prioritize forest fire management efforts in areas that have the most impact and also identify which area is most suitable for development so that future forest fire can be prevented and will reduce the impact of the economy in Brazil.

VIII. Conclusion

Study of the fire forest occurence in Brazil has given a better insight on the matter of fire forest. The result of the study found out that there will be an increase in the future year of until 2022. The analysis was made on the data that are acquired through a trusted and updated data. However, the value of the data does not really match the criterias to be taken as a characteristic. The only useful attribute that is available for the study is the number of fire occurrences in every state of Brazil. With this data, a linear regression model are used to be predicting the future fire occurrences.

The prediction that is the outcome of this project might not be accurate and precise as there are very few data are available. However, this study can be a warning or precautions study for the involved parties that have the fire occurrences problem in their hands. In the future, academic like students and data scientists can use this study as a reference to a fire occurrences problem.

ACKNOWLEDGMENT

By the end of this data analytics project, we would like to express our gratitude and appreciation to all those who contribute and help out in completing this project. We would especially thank Dr. Raini Binti Hassan as our Big Data Analytics (CSC 3303) lecturer for kindly assist and guide us, in addition our completion of this project could not have been accomplished without the support from Dr. Raini. At last but not least, our parents are also an essential inspiration as they give us full support to complete this analytics project.

REFERENCES

[1] "WNF: Places: Amazon". Retrieved June 4, 2016 from http://www.worldwildlife.org/places/amazon

[2] "Field Museum scientists estimate 16,000 tree species in the Amazon". *Field Museum*. October 17, 2013. Retrieved October 18, 2013 from http://www.eurekalert.org/pub_releases/2013-10/fm-fms101413.php

[3] "Photos / Pictures of the Amazon Rainforest". Travel.mongabay.com. Archived from the original on December 17, 2008. Retrieved December 18, 2008 from http://travel.mongabay.com/brazil/brazil_amazon_index.htm .

[4] "Brazil's Bolsonaro says he will accept aid to fight Amazon fires". CBS News. August 27, 2019. Retrieved August 30,2019 from https://www.cbsnews.com/news/amazon-wildfires-brazil-spurns-20-million-aid-offer-from-g-7-nations-today-2019-08-27/ .

[5] Lombardo, C. (2015). Pros and Cons of Forest Fires. Retrieved October 10, 2019 from http://visionlaunch.com/pros-and-cons-of-forest-fires/

[6] The Predictive Analytics Process. (n.d.). Predictive Analytics, Data Mining and Big Data. doi: 10.1057/9781137379283.0010o

[7] Safi, Y., Bouroumi, A., & Bouroumi, A. (2011). A neural network approach for predicting forest fires. 2011 International Conference on Multimedia Computing and Systems. doi: 10.1109/icmcs.2011.5945716

[8] “Basics of linear regression”. Retrieved November 6, 2018 from https://medium.com/datadriveninvestor/basics-of-linear-regression-9b529aeaa0a5