**ASSIGNMENT FOR DATA ANALYTICS**

Summited by: Abusomwan Amadevborho

K00266018

This project was a big step to my learning in Data Analytics. The data used in this project was a public dataset from Kaggle without any constraint likewise the private dataset which has some restrictions. However, working on this project was a bit easier because we worked in pairs and hence, creating more ideals working as a team.

Furthermore, this project delves deep into the significant earthquakes, a time series data from 1965-2016 which consists of some interesting key attributes such as longitude, latitude, magnitude, depth etc. Nevertheless., this was a large dataset consisting of 23,412 rows, therefore making the task at hand a bit more challenging. Rapid Miner which is a popular data analytics software with a variety of multiple statistical operators was used for this task. Hence, with this we achieved a lot from doing less.

The first step was planning, myself and my teammate brainstormed on ideals which would be feasible of realising the goals in advance of the project. Fortunately, we came up with a good plan after looking up and referring to previous project that was published on the Kaggle website. Hence, referring and comparing the published work of other authors on the Kaggle website.

The second step was building a model for our project using a linear regression operator. Here, we were more interested in the depth while filtering the irrelevant attributes, thereby making the magnitude, longitude, and latitude the explanatory, predictors, covariates or regressor variable which was used in making predictions for the depth. However, we discovered that the location factor which is the longitude and latitude has an inverse or negative relationship with the depth while the magnitude has a positive relationship with the depth.

The third step, myself and my teammate thought it was better to do a classification of the data to get a valid accuracy of our model. A cross validation was done taking a subset of the data while we get the depth predictions in a class of two categories. So, after adjusting the parameters we realised a good accuracy score and a Roc curve spanning in the range of 0.75 which is a good fit.

In conclusion, I realised a lot of analysis can be done by using statistical operators, thereby making data analysis easy. This project helped in building my team spirit, achieving goals and making decision.