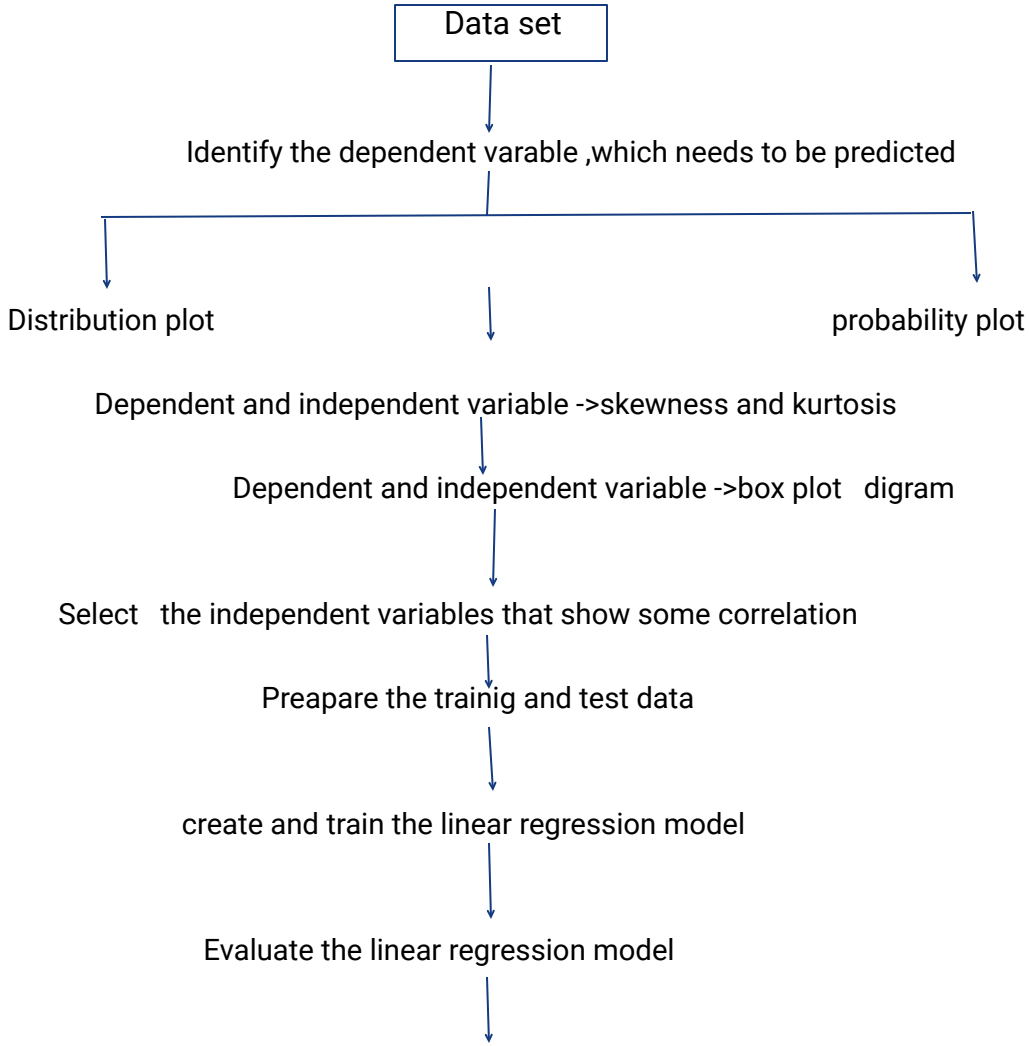
REPORT ON HEALTH INSURANCE PREDICTION

INTRODUCTION:

The idea is to present a machine learning model that predicts the premium cost with the best accuracy.By deriving the correlation from the dataset,it is obsrved that there is a strong relation between charges and the category whether the person is smoker or not.This helps a person in focusing more on the health aspect of an insurance rather than the futile part.

Health insurance is a necessity nowadays, and almost every individual is linked with a government or private health insurance company.So by using this alogrithm we can predict the amount of increase or decrease in their insurance amount.

THEORITICAL ANAYLSIS :



Use the model to predict insurance cost

To develop this model we require a PC and softwares like eclipise IDE and weka gui .By using these softwares we can develop a machine learning algorithm with linear regression .

EXPERIMENTAL INVESTIGATIONS :

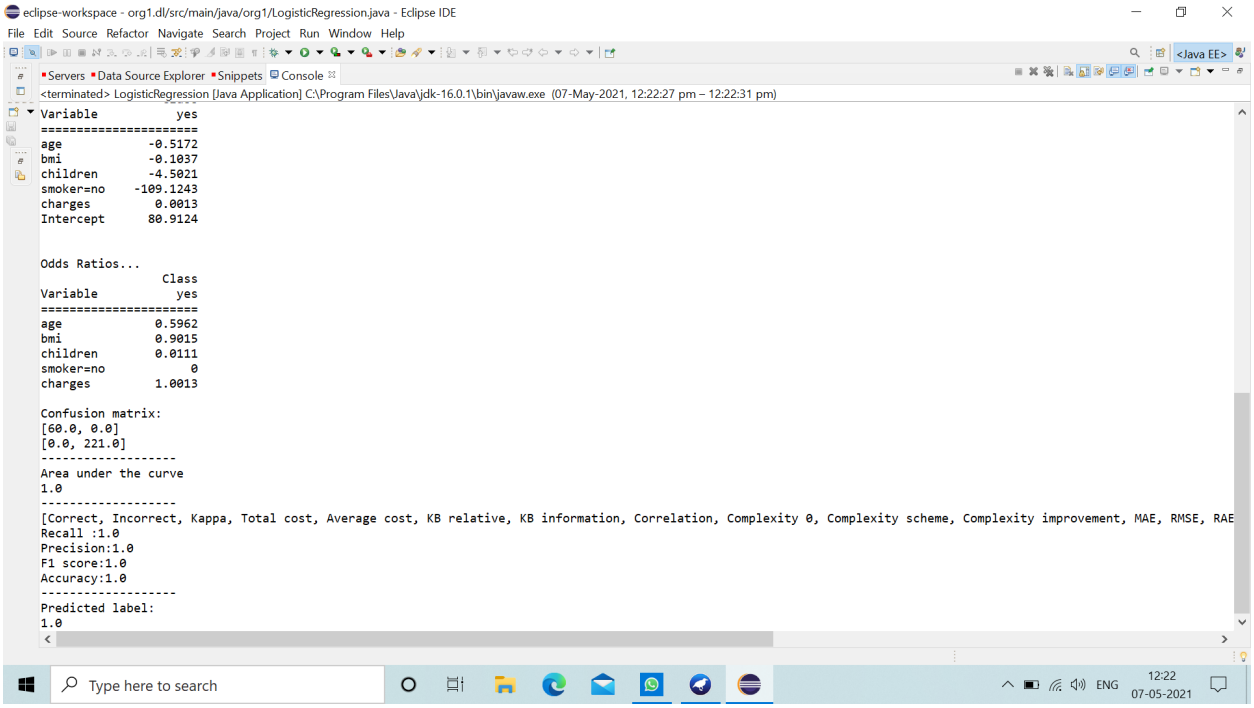
The first step in this is data preparation and cleaning,.First the data has been imported from kaggle website. The data included various attributes such as age, gender, body mass index, smoker and the charges attribute which will work as the label for the project.The data was in structured format and was stores in a csv file format.

The presence of missing, incomplete, or corrupted data leads to wrong results while performing any functions such as count, average, mean etc. These inconsistencies must be removed before doing any analysis on data. The data included some ambiguous values which were needed to be removed.

We are using weka gui software to process the data and visualising the data .weka software is the very helpful in data visualization and to build train and test data .
To build linear regression algorithm we use eclipse IDE ,and finally training and testing the model is done through the developed machine learning model.

RESULT :

we finally able predict the health isurance amount of a person accurately through the model build.



CONCLUSION :

Various factors were used and their effect on predicted amount was examined. It was observed that a persons age and smoking status affects the prediction in algorithm applied. Attributes which had no effect on the prediction were removed from the features.
The effect of various independent variables on the premium amount was also checked. The attributes also in combination were checked for better accuracy results.

FUTURE SCOPE :

Premium amount prediction focuses on persons own health rather than other companys insurance terms and conditions. The models can be applied to the data collected in coming years to predict the premium. This can help not only people but also insurance companies to work in tandem for better and more health centric insurance amount.

BIBILOGRAPHY :

1. https://en.wikipedia.org/wiki/Healthcare_in_India

2. https://www.kaggle.com/mirichoi0218/insurance

3 . https://www.zdnet.com/article/the-true-costs-and-roi-of-implementing-ai-in-the-enterprise