CPSC 483, Project 1

Team Name: ‘AKMMR’

Team Members and contribution percentage:

Asit Kumar Dash, 893423350

Kashyap Sodha, 893260984

Melika Amini, 893423350

Milan Ghori, 893585703

Rishabh Sharma, 893463695

b) The project will be done in python 3 using the sci-kit learn library. The use of library was done to reduce the number of lines of the code and quickly test different models. Sci-Kit Learn is a machine learning library that supports python. The library provides different kind of regression and classification model including linear regression, SVM (Support vector machines) and even perceptron-based regression techniques. The above features of the library prompted the authors to use this library.

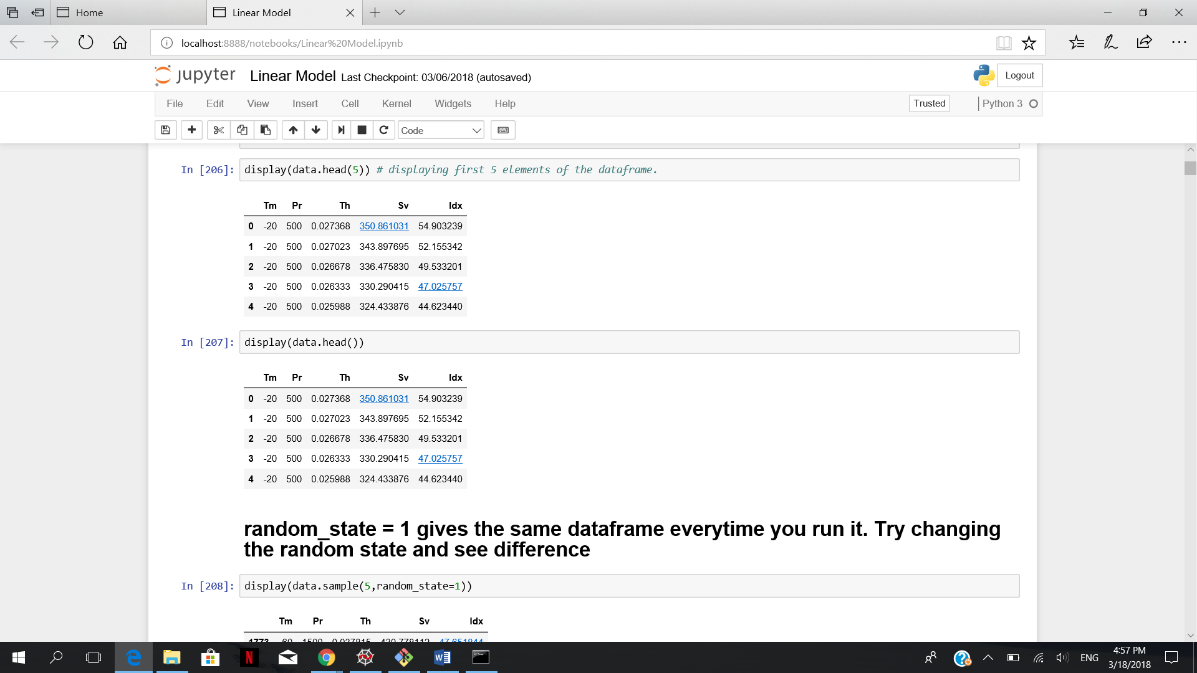
c) The authors used Linear Model and used mean squared error and mean absolute error. We observed that when we used a linear model Y = MX + C. The error in mean squared approach was 1.28 and the error in mean absolute error was also 1.28. For the Non-Linear model. The linear model was tested and build by Rishabh Sharma, Kashyap Sodha and Asit Kumar Dash and Milan Ghori. Asit Kumar Dash and Rishabh Sharma tested different regression models like Ridge regression, Lasso Regression, Bayesian Regression with the same loss functions. Milan Ghori tested the Elastic Regression Model which had a error of 2.30 and hubber regressor which had a loss of 6.00. Bayesian Regression was getting a complete error of 6.0. The best model that were obtained was Linear Regression.

Rishabh Sharma decided the loss function for all the models. The validation split in the Linear Model is 75% training and 25% testing.

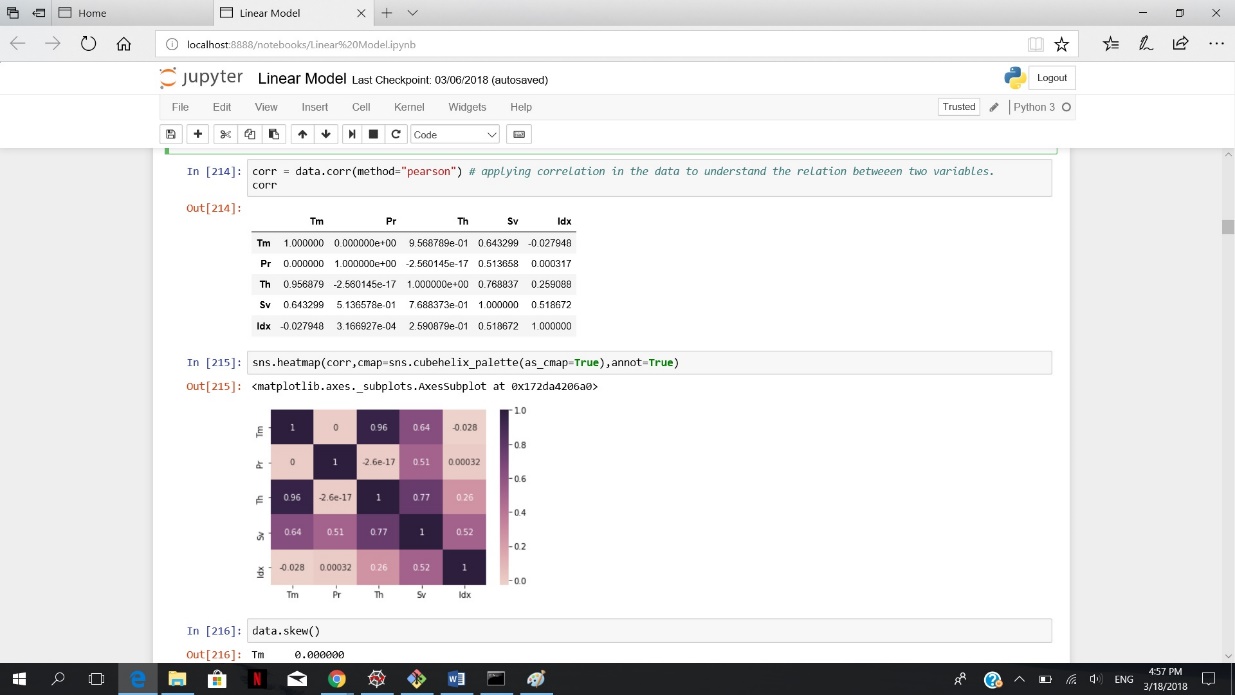
The validation split in the Non-Linear model is of 80% training and 20% testing.

For Non-Linear Models the equation used are Y = M1X^2 + M2X +C that is the with the expansion the degree of the model is 2 and there are total 15 weights along with an intercept, Squared error obtained for this model is 0.1024 and the variance score is 1. Another model with a degree of 4. The total coefficients are 69 and an intercept, the mean squared error for this model is 0.0708 and the Variance score is 1. After the experiments the team members decided to use the Non-Linear model with the degree 4 because the computation time is not very high and the accuracy is fairly high. The Non-Linear model was designed by Rishabh Sharma and the same model was taken by other team members and new loss functions were tested. The experiments and results showed that the best lost function was the mean squared error. Asit Kumar Dash was actively involved in the Non-Linear model. Melika Amini tested the Non-Linear model with the absolute error.

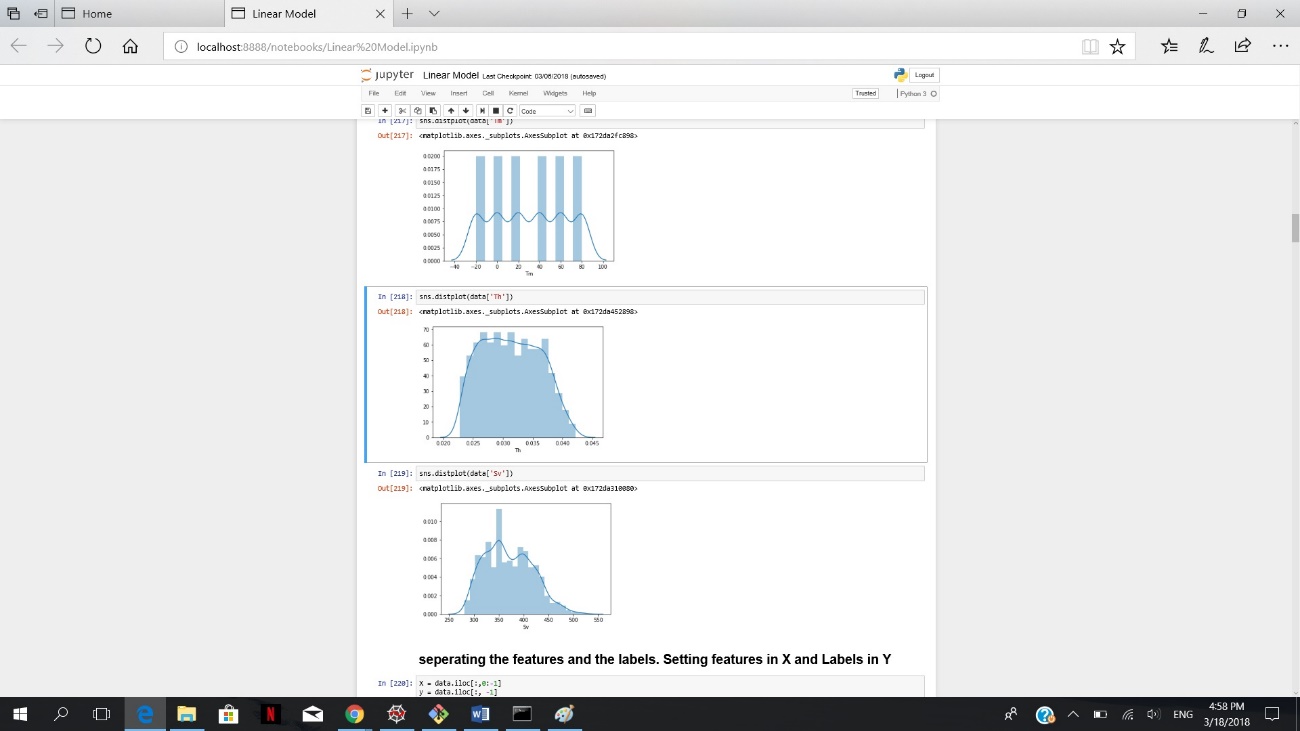
d)The models were tested, and the data was designed in the Jupyter notebook. Please find the attached screen shots of the jupyter notebook.



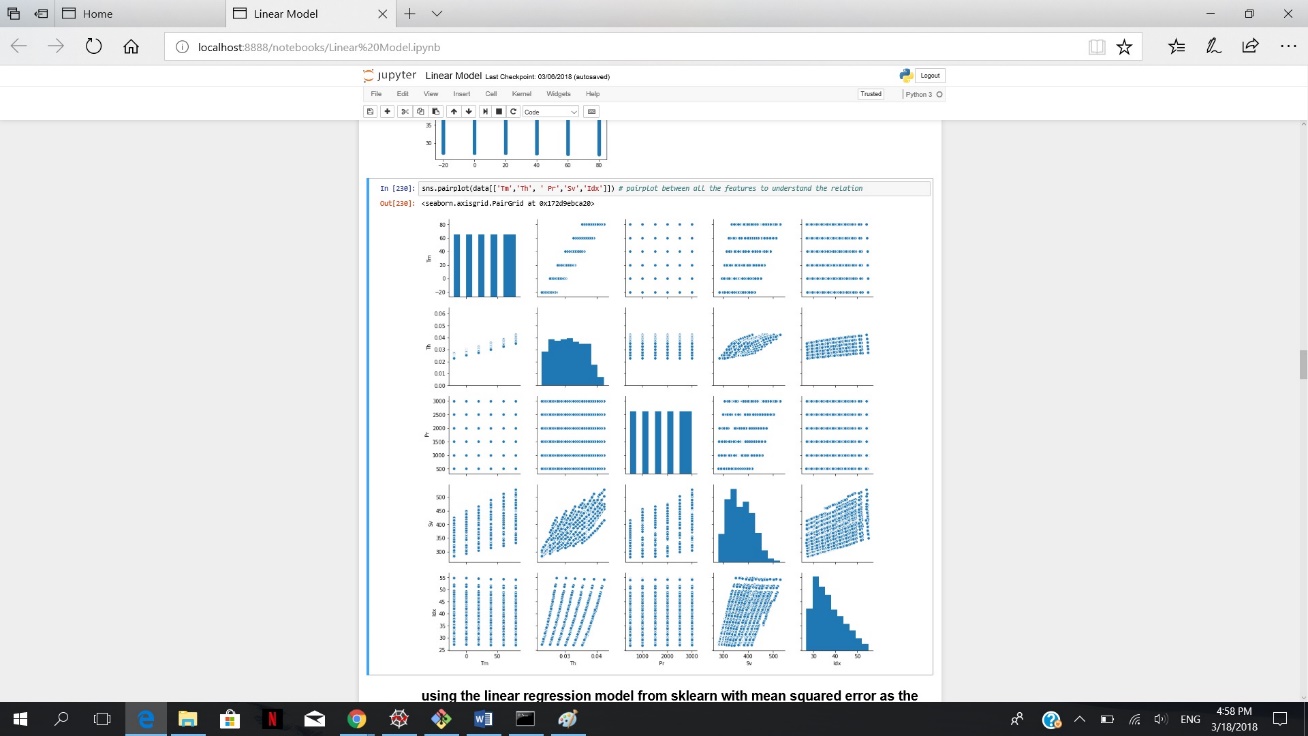
The data image



The Correlation Matrix



Data Distribution



Pair Distribution of the features.

To run the files and to see the results please install the required libraries, numpy, sci-kit learn.

To run the linear model use the command in terminal: python Linear\_Model.py.

To run the non-linear model in python use the command: python Non\_linear\_model.py