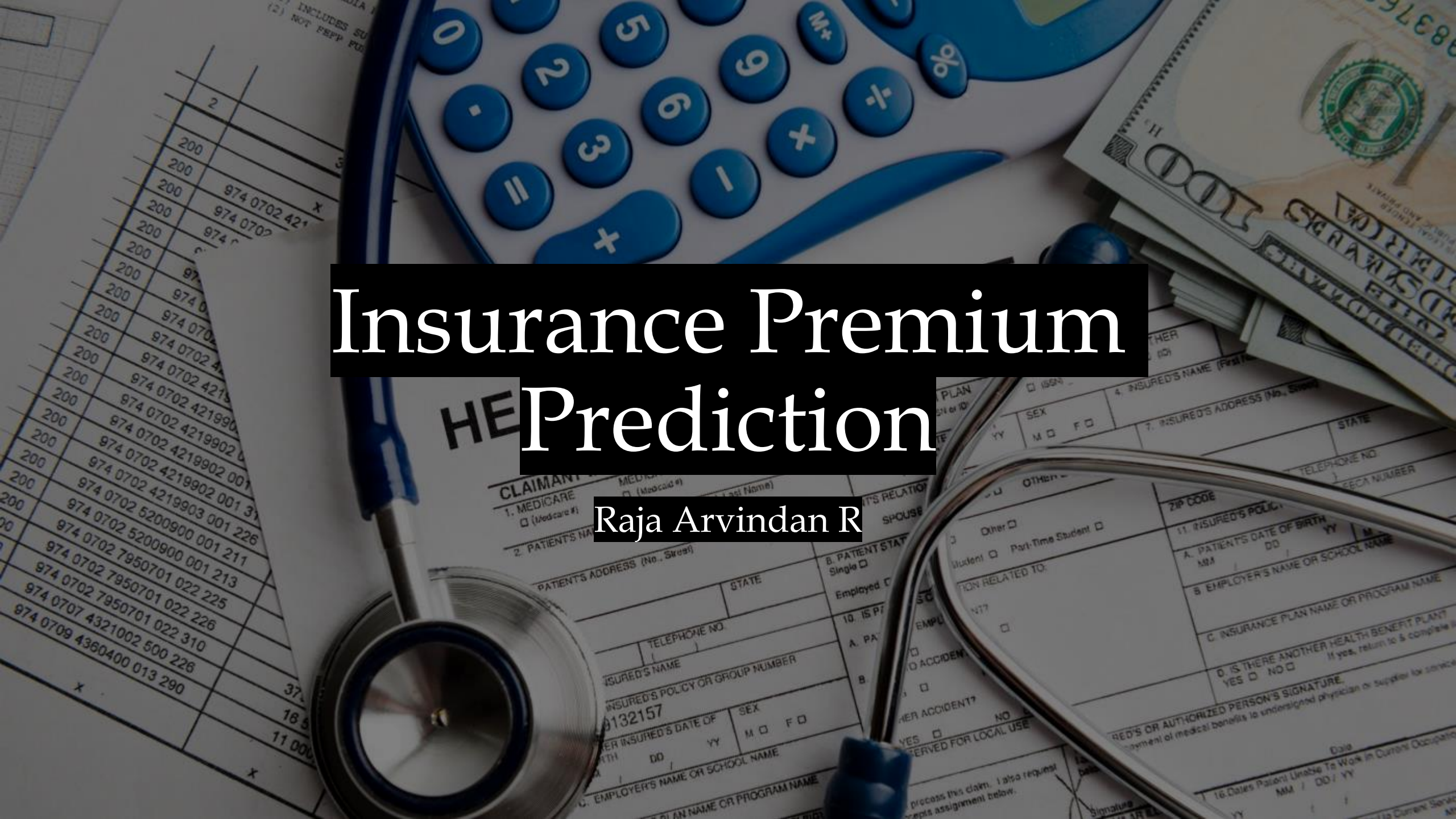


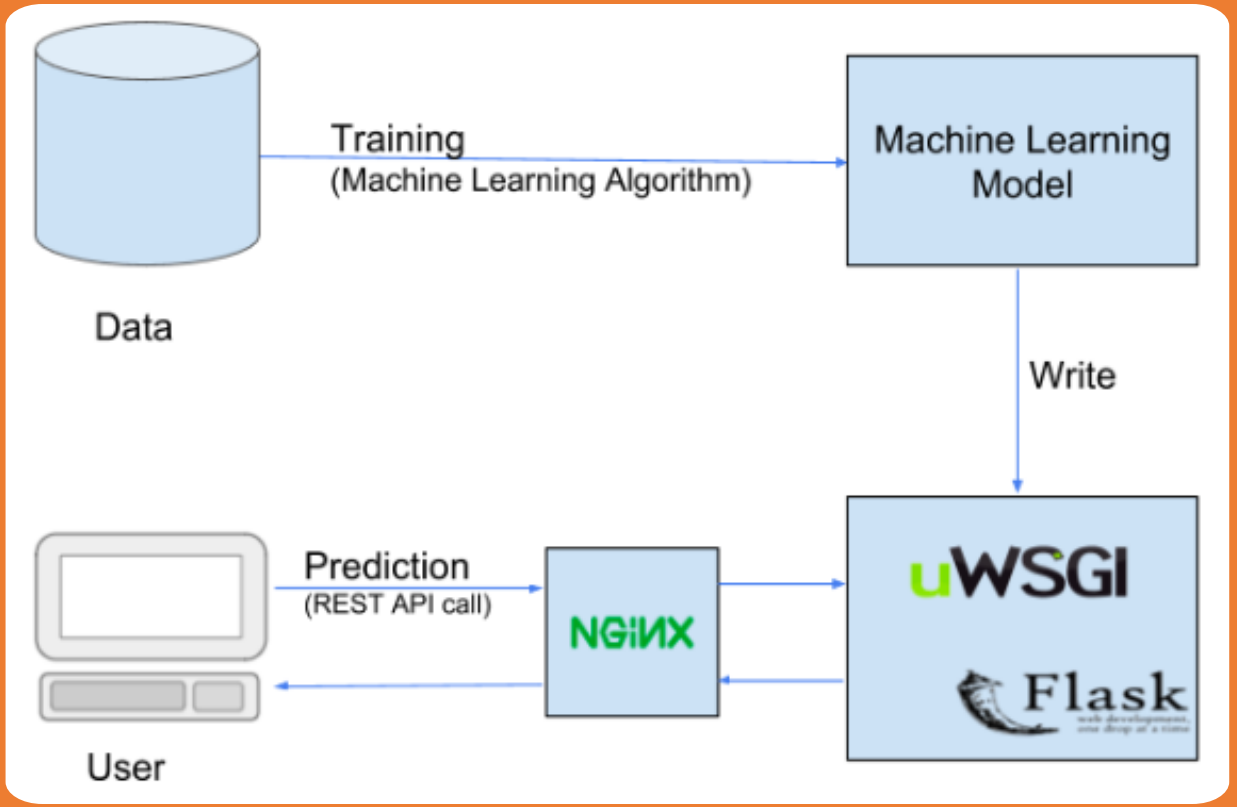
Insurance Premium Prediction

Raja Arvindan R

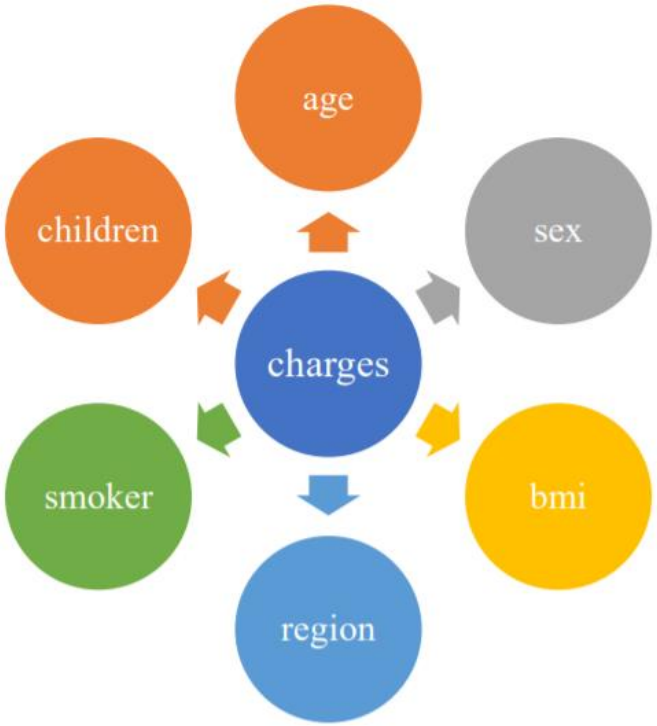
Objective:

- Judicious use of predictive analysis has empowered health insurers to improve their premium pricing accuracy, create customized health insurance plans and services, and build stronger customer relationships.
- The main goal of this project is to predict the insurance premiums based on the behavioral data collected from the individuals so that insurance companies can make useful and accurate predictions.

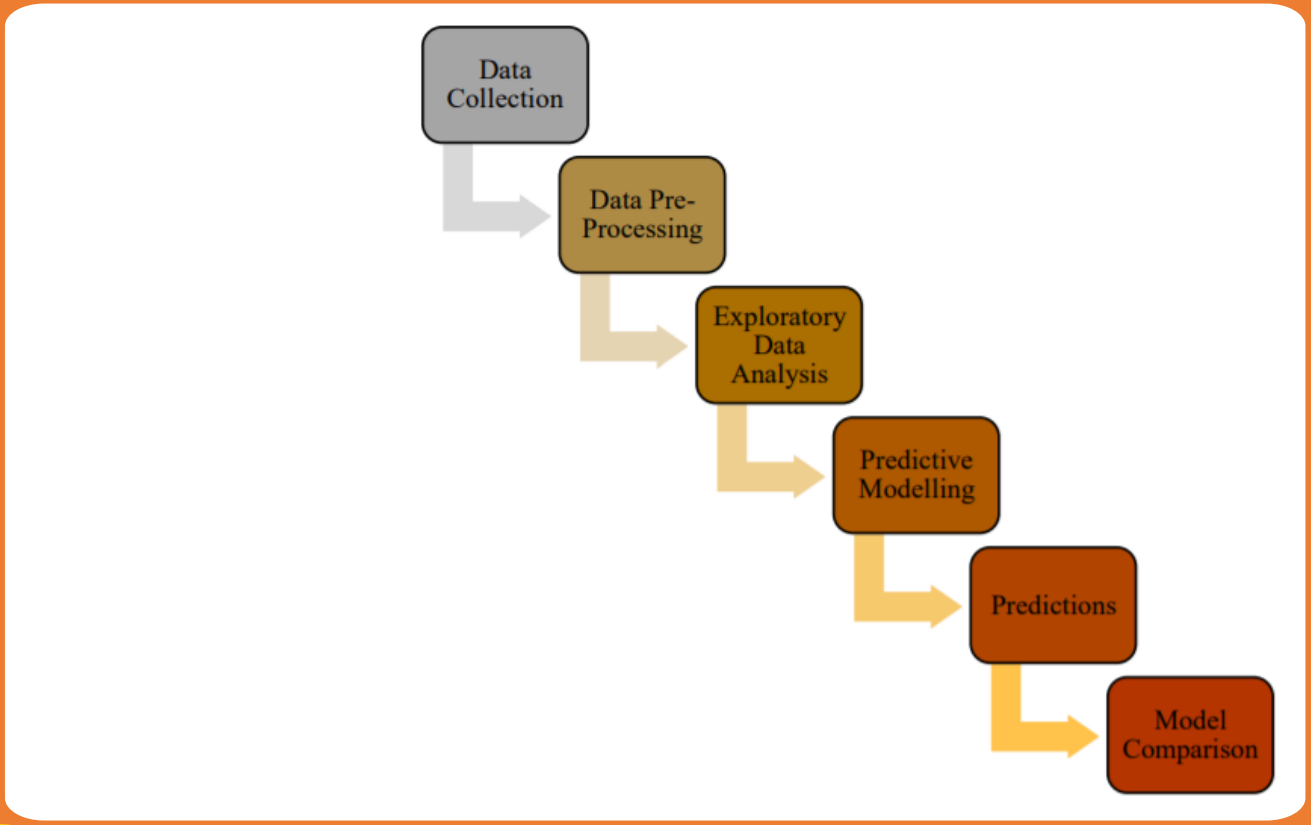
Architecture:



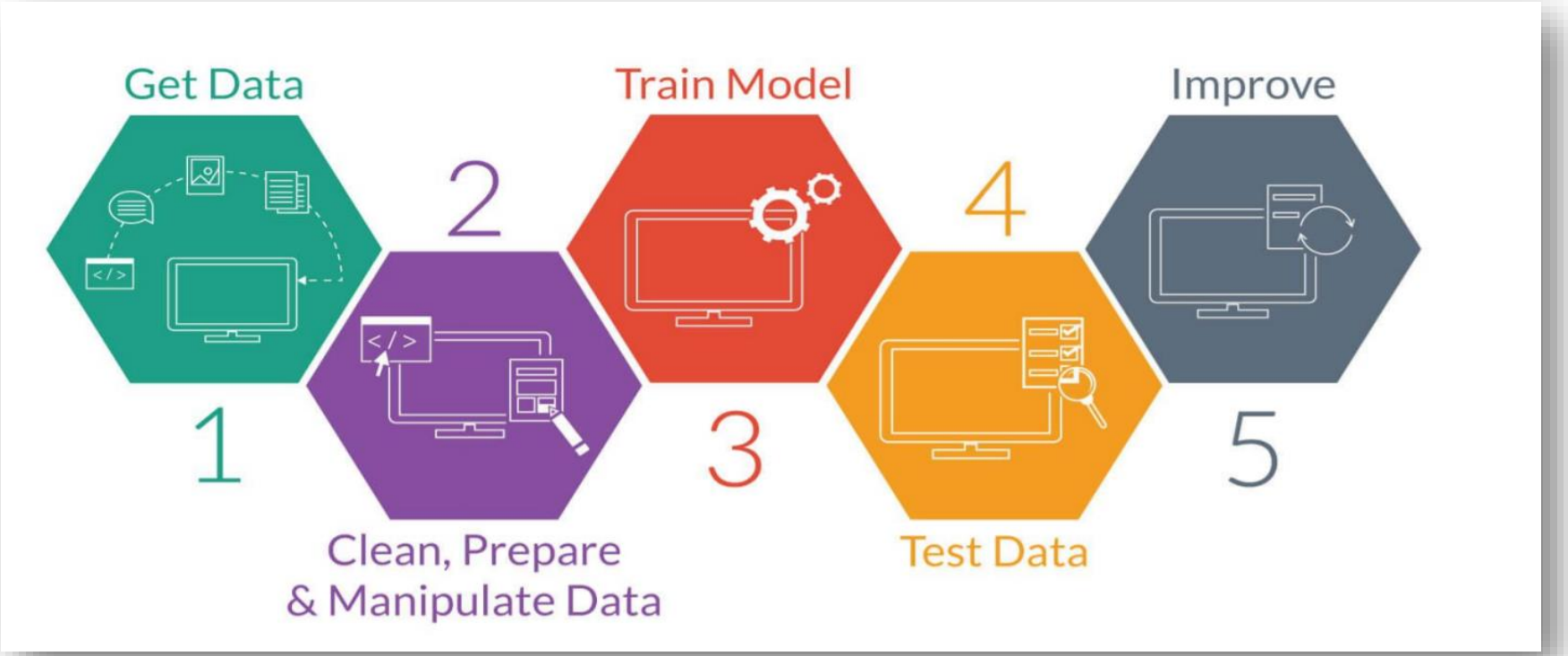
Dataset:



Data Analysis Steps:



Steps to Predictive Modelling:



Gradient Boosting Regression:

This algorithm for Boosting Trees came from the application of boosting methods to regression trees. The basic idea behind this is to compute a sequence of simple trees, where each successive tree is built for the prediction residuals of the preceding tree. For predictive models, gradient boosting is considered as one of the most powerful techniques. Gradient boosting involves three elements:

1. An optimized loss function.
2. An additive model to add weak learners to minimize the loss function.
3. A weak learner to make predictions

Training:

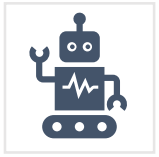
- Once training data is in a suitable form to feed to the model, the training and testing phase of the model can proceed. During the training phase, the primary concern is the model selection. This involves choosing the best modelling approach for the task, or the best parameter settings for a given model. In fact, the term model selection often refers to both of these processes, as, in many cases, various models were tried first and best performing model (with the best performing parameter settings for each model) was selected.

Predictions:

The model was used to predict the insurance amount which would be spent on their health.

The model used the relation between the features and the label to predict the amount.

Accuracy defines the degree of correctness of the predicted value of the insurance amount.

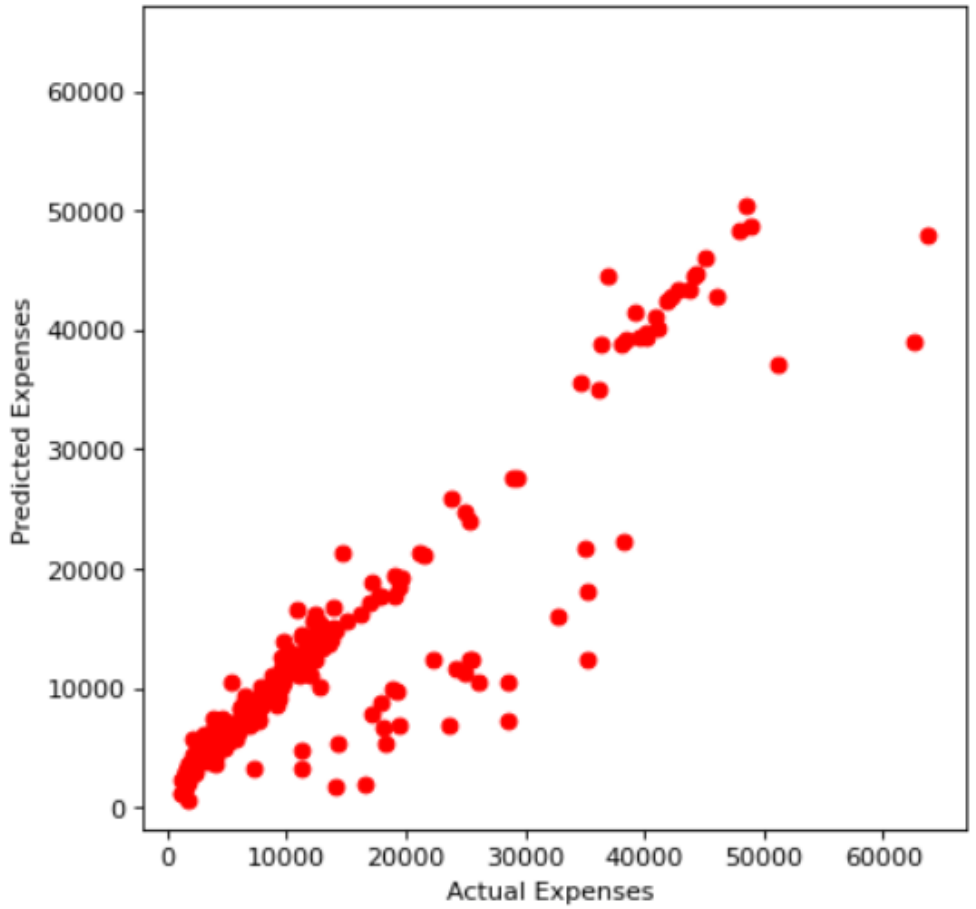


The model predicted the accuracy of model by using different algorithms, different features and different train test split size.

The size of the data used for training of data has a huge impact on the accuracy of data. The larger the train size, the better is the accuracy.

The model predicts the premium amount using multiple algorithms and shows the effect of each attribute on the predicted value.

Actual vs Predicted:



- Raja Arvindan R