Final Project: Group 3

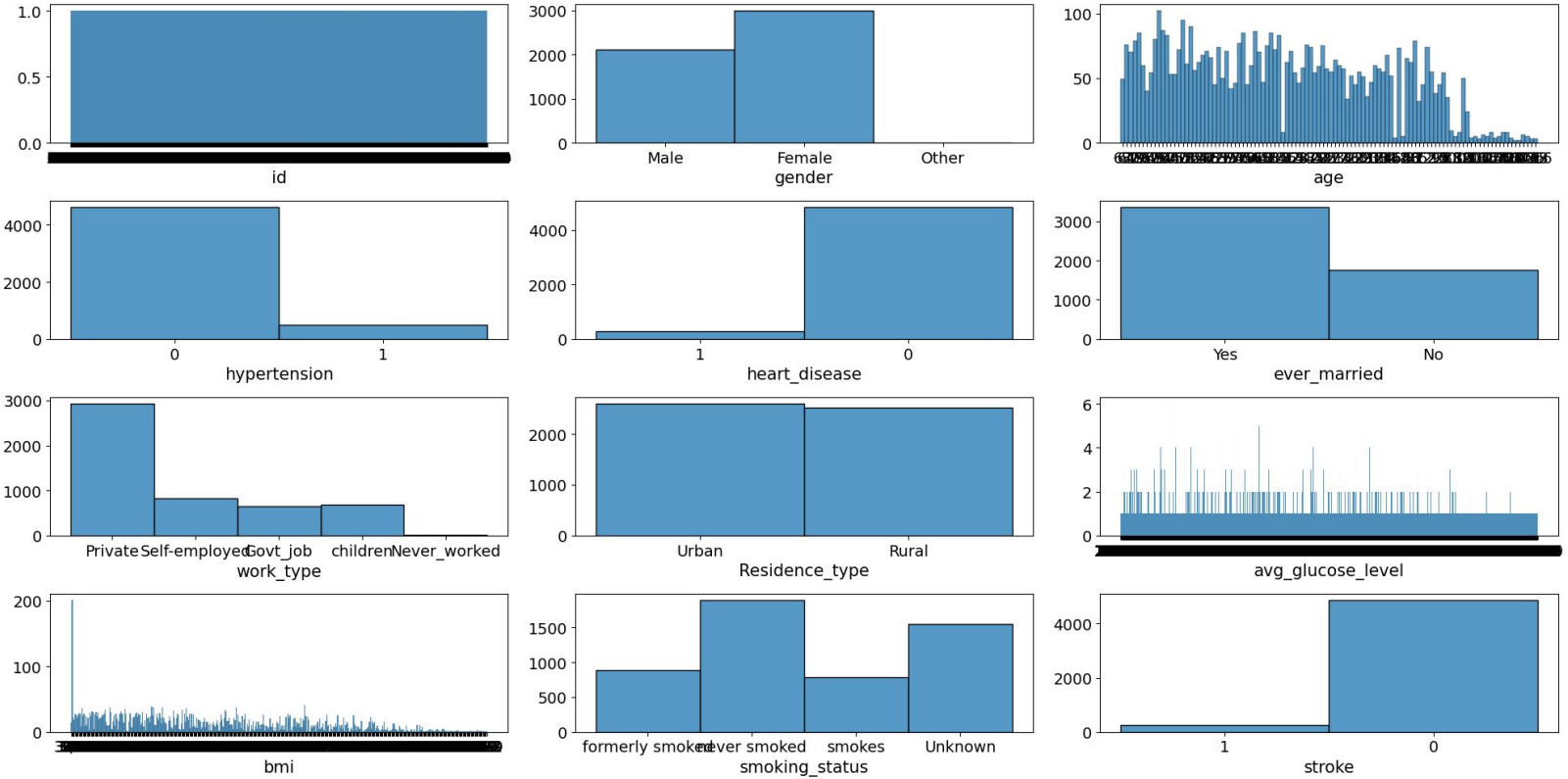
CS 301- Summer 2023

Prof. Islam

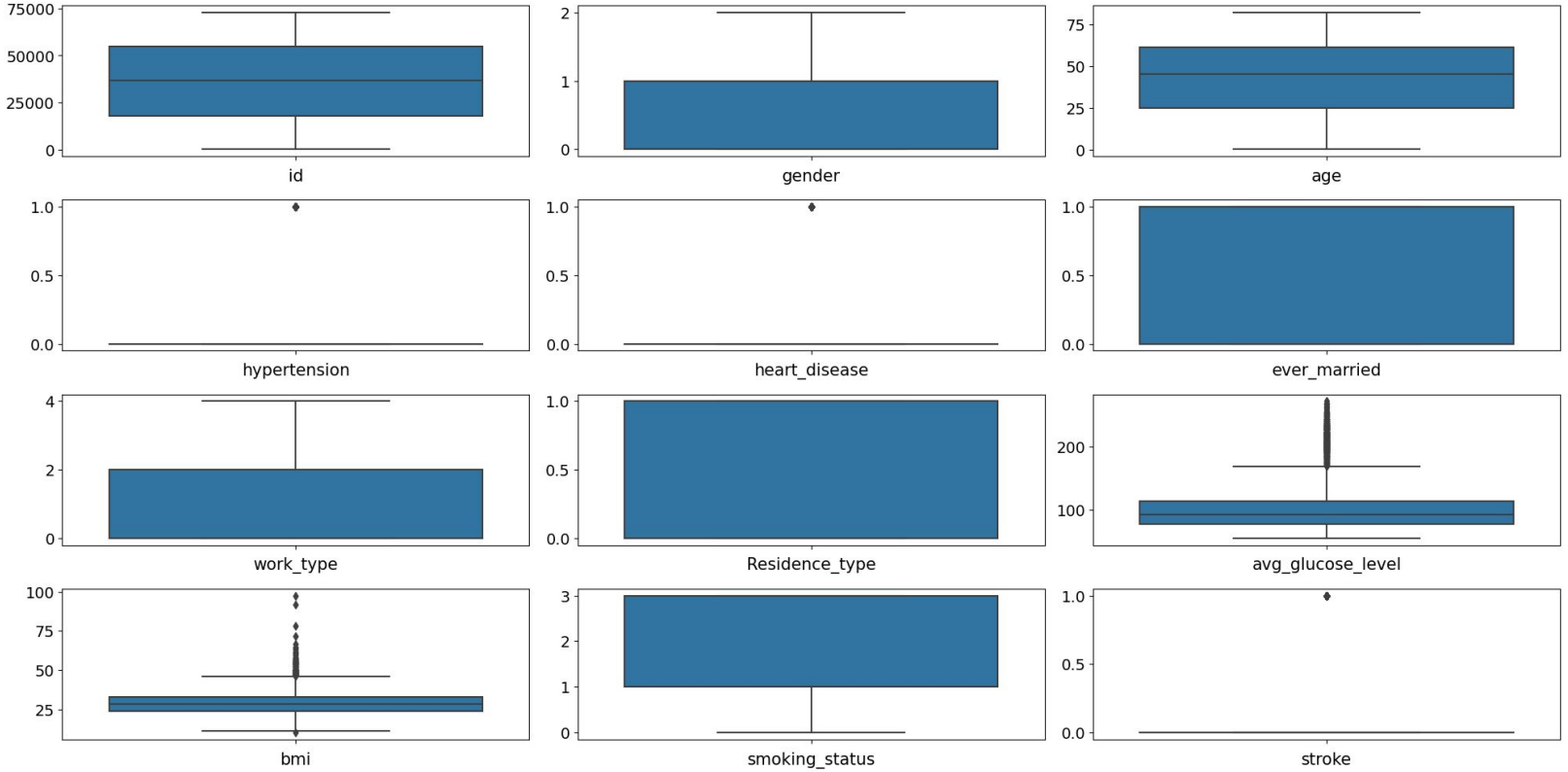
Exploratory Analysis
Healthcare Dataset

Sarthak Basu, Dan Kabore, Rohan Shanbhag, Vibhav Ohri

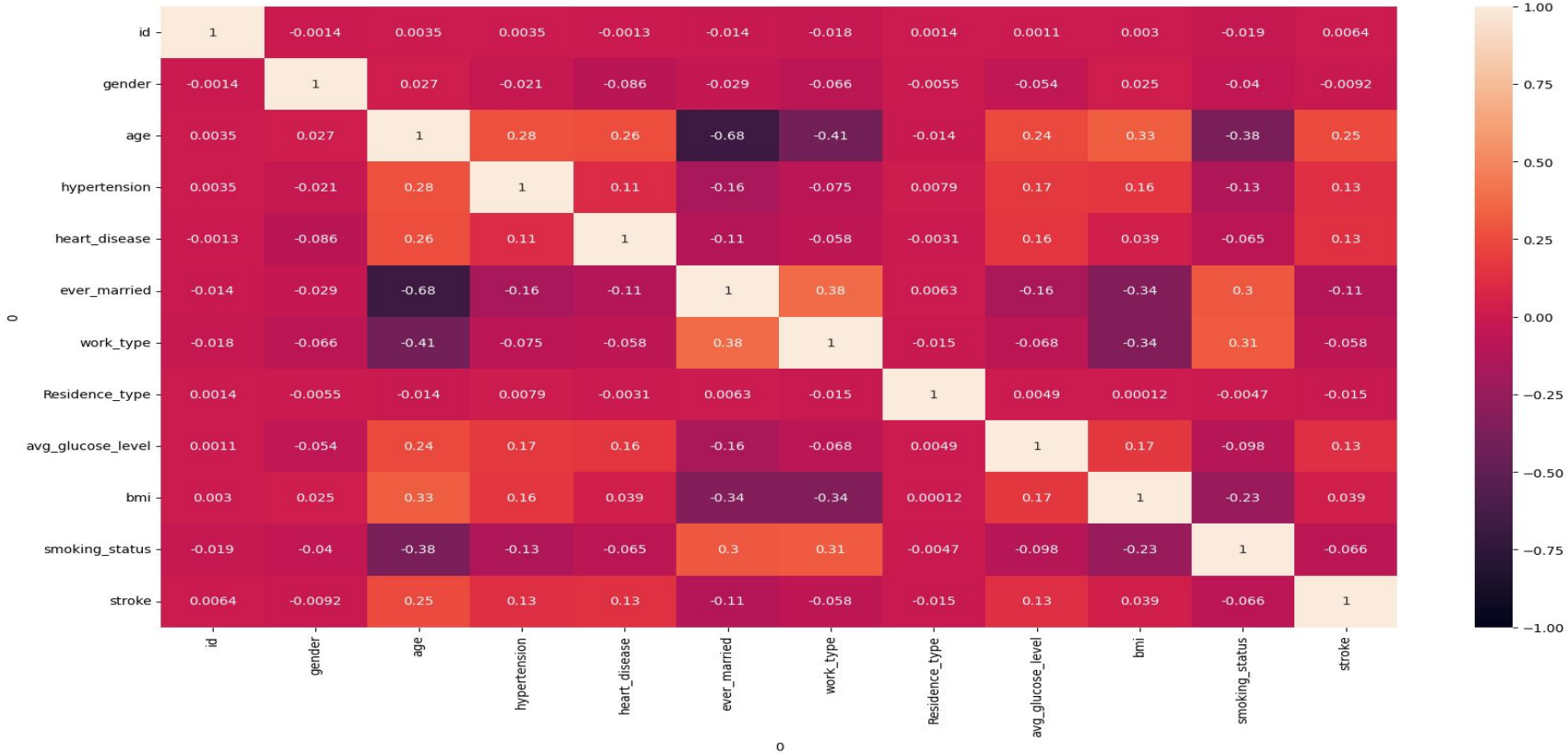
Histograms for Variables



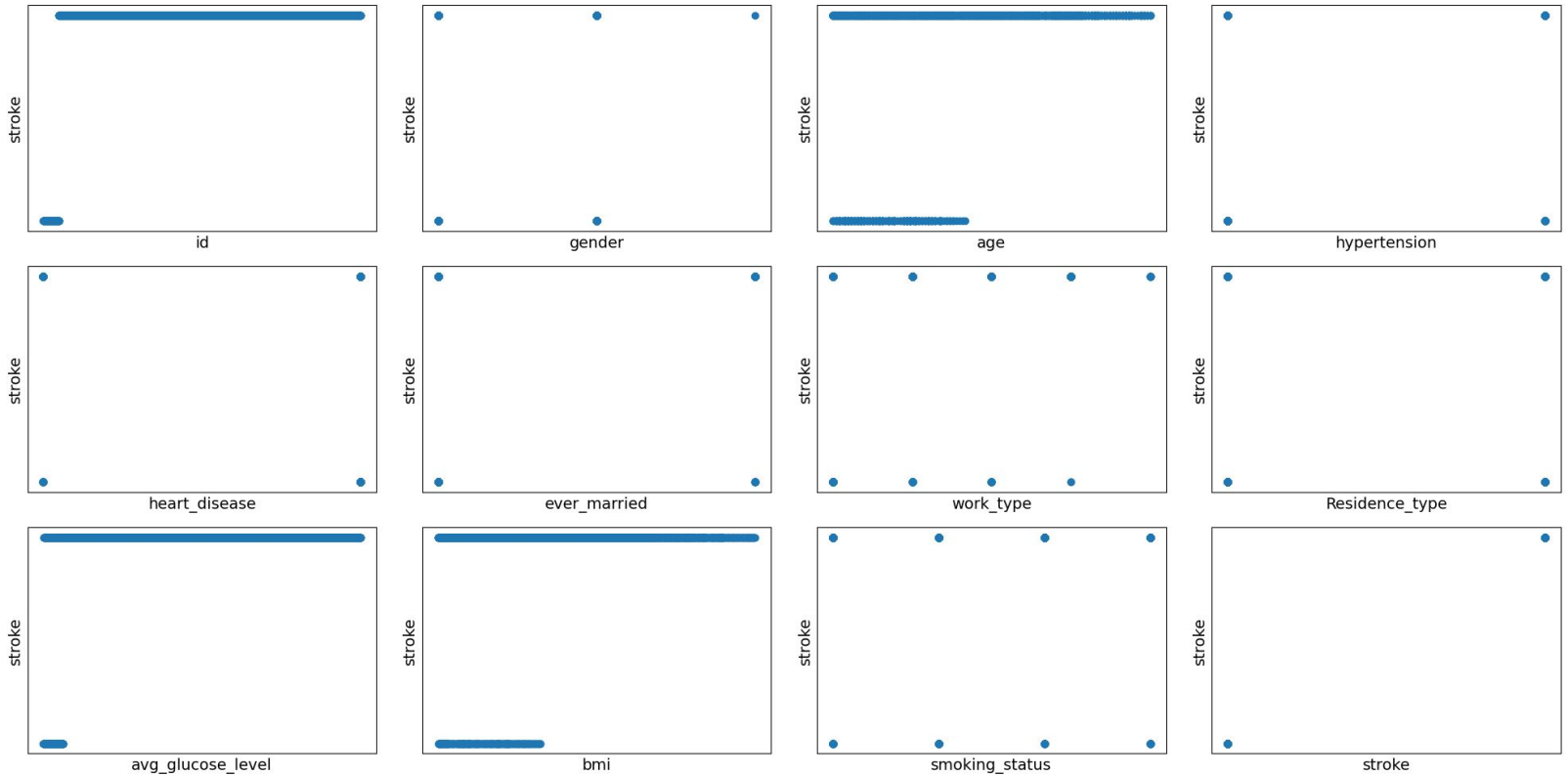
Boxplots for Variables



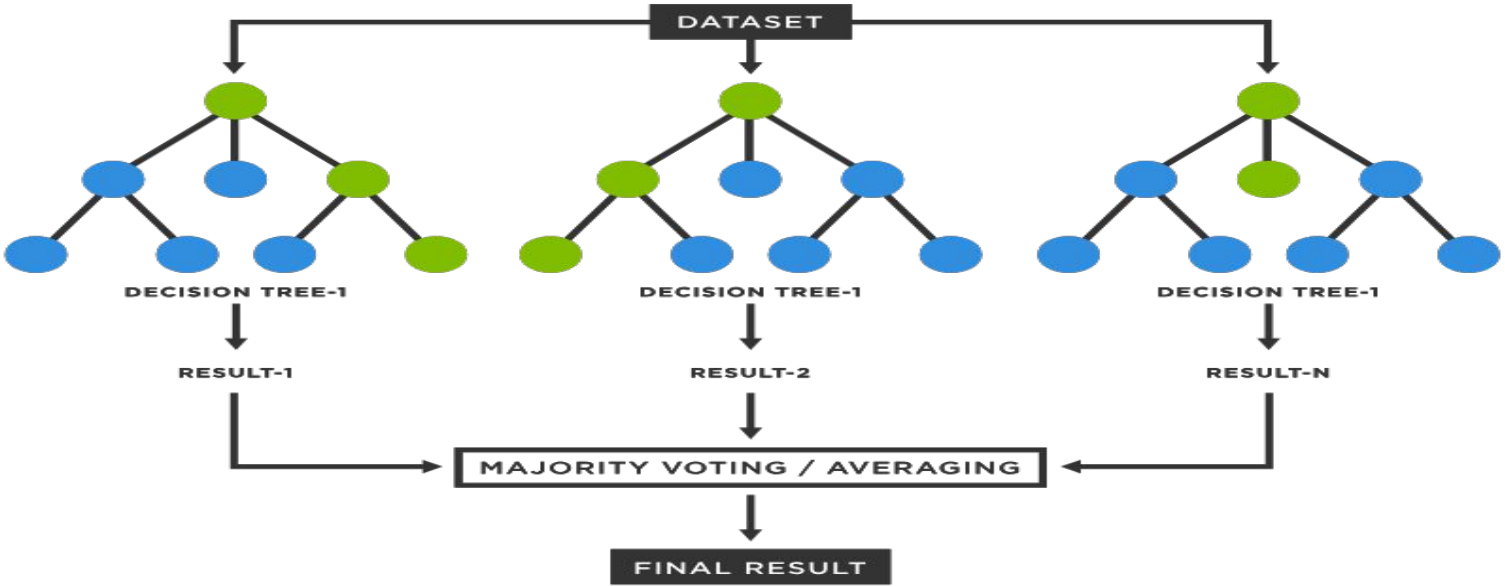
Correlation between Variables



Scatterplots of the Variables



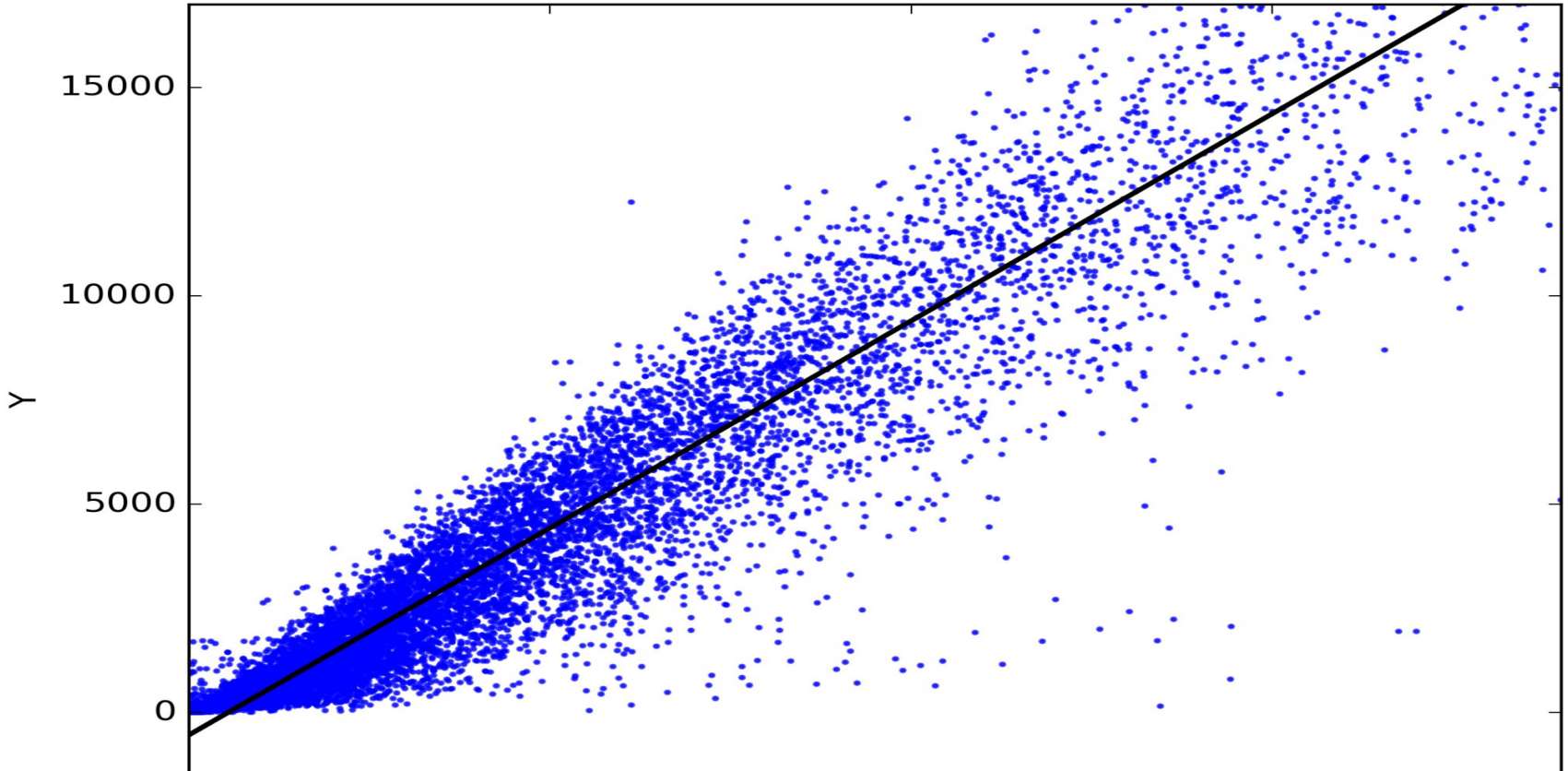
Random Forest Model



What is a Random Forest Model

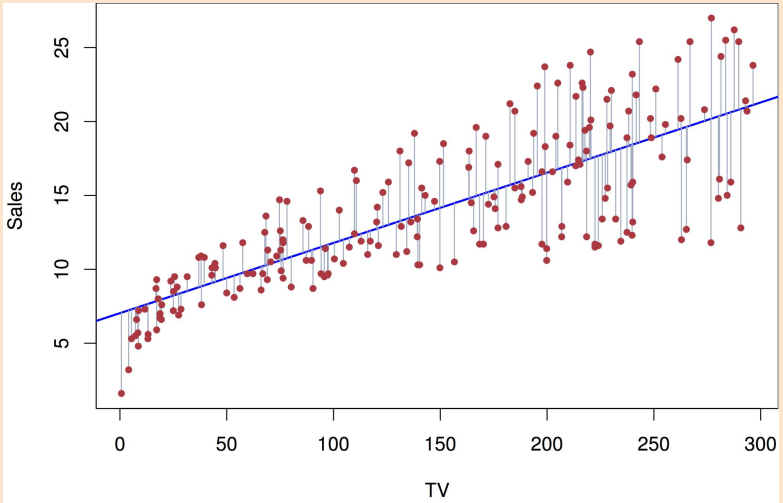
- A random forest model is an algorithm in machine learning which uses classification and regression together and combines the result of all the trained multiple models and decision trees to come up with with a final output result
-

Linear Regression



- What is linear regression

- Difference type of linear regression model



Optimisation Algorithm

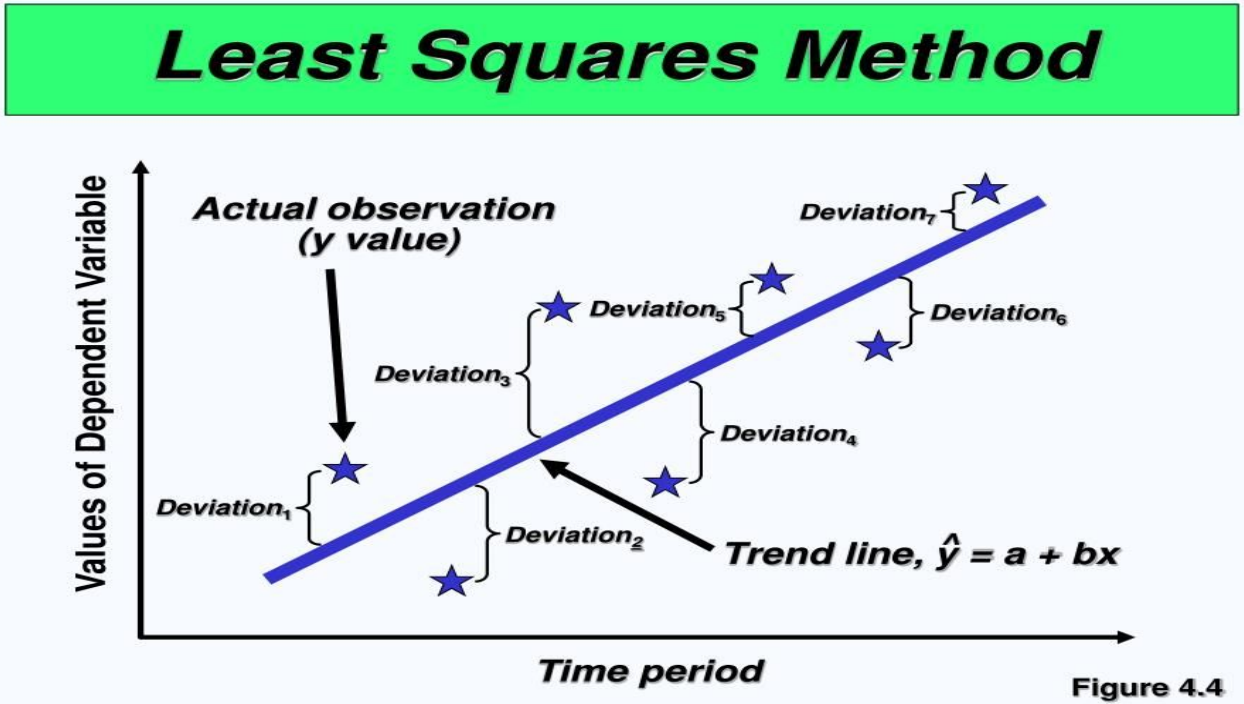
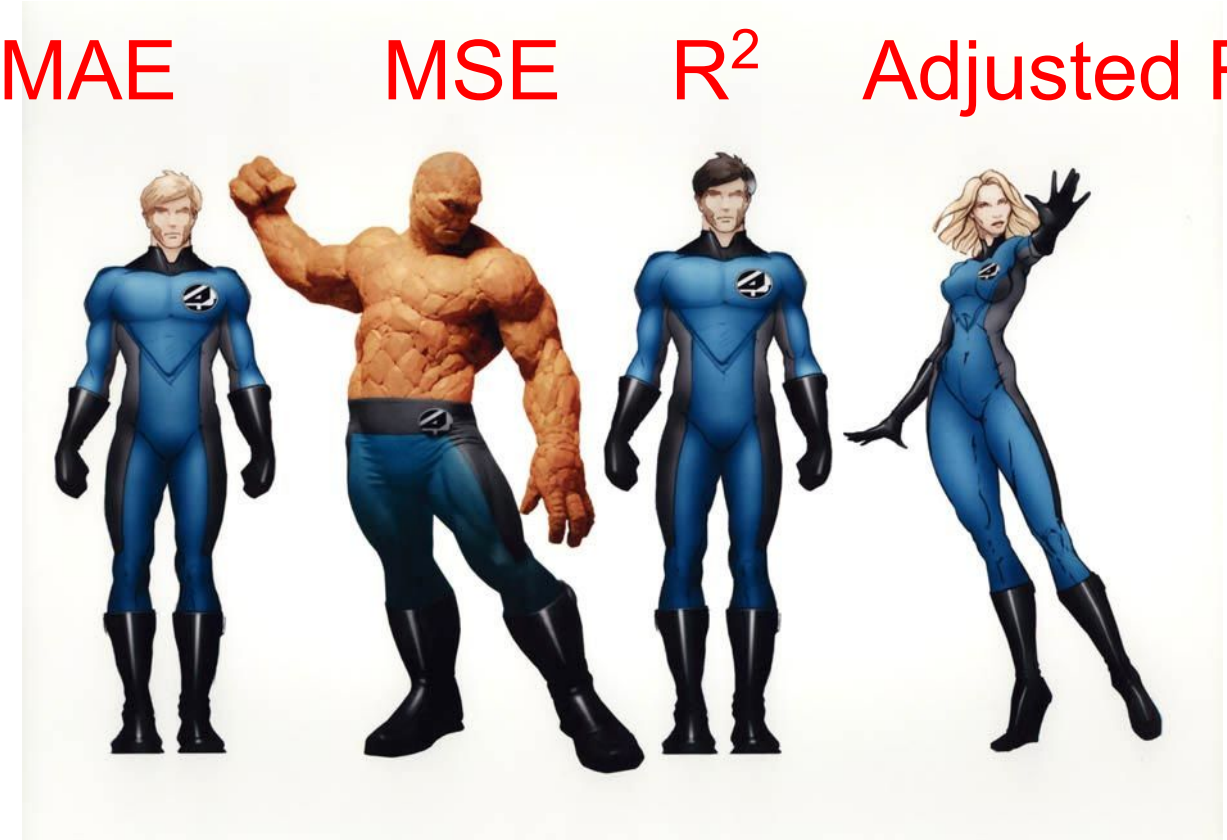


Figure 4.4

Part 3: Evaluation

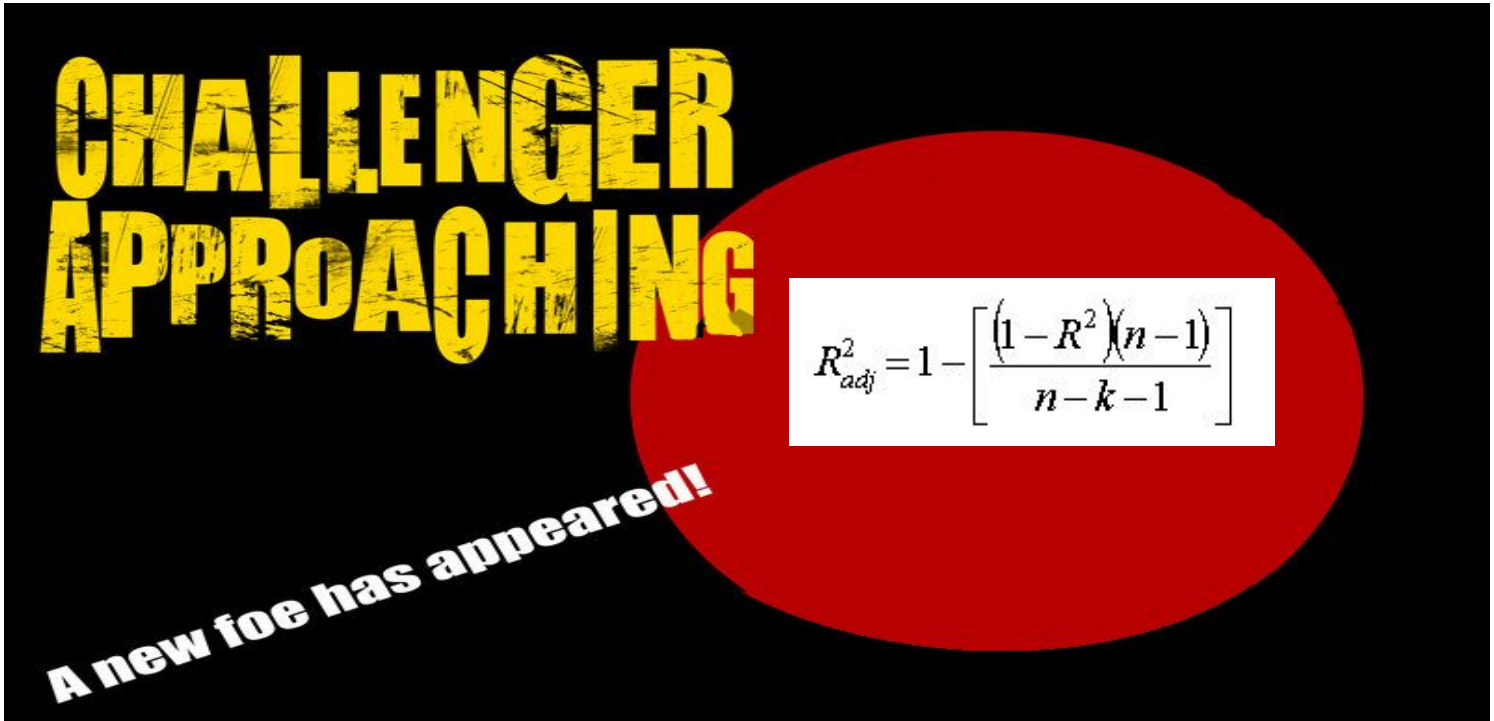
MAE MSE R^2 Adjusted R^2



MAE vs MSE

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - y_j|$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$


CHALLENGER APPROACHING
$R^2_{adj} = 1 - \left[\frac{(1-R^2)(n-1)}{n-k-1}\right]$
A new foe has appeared!

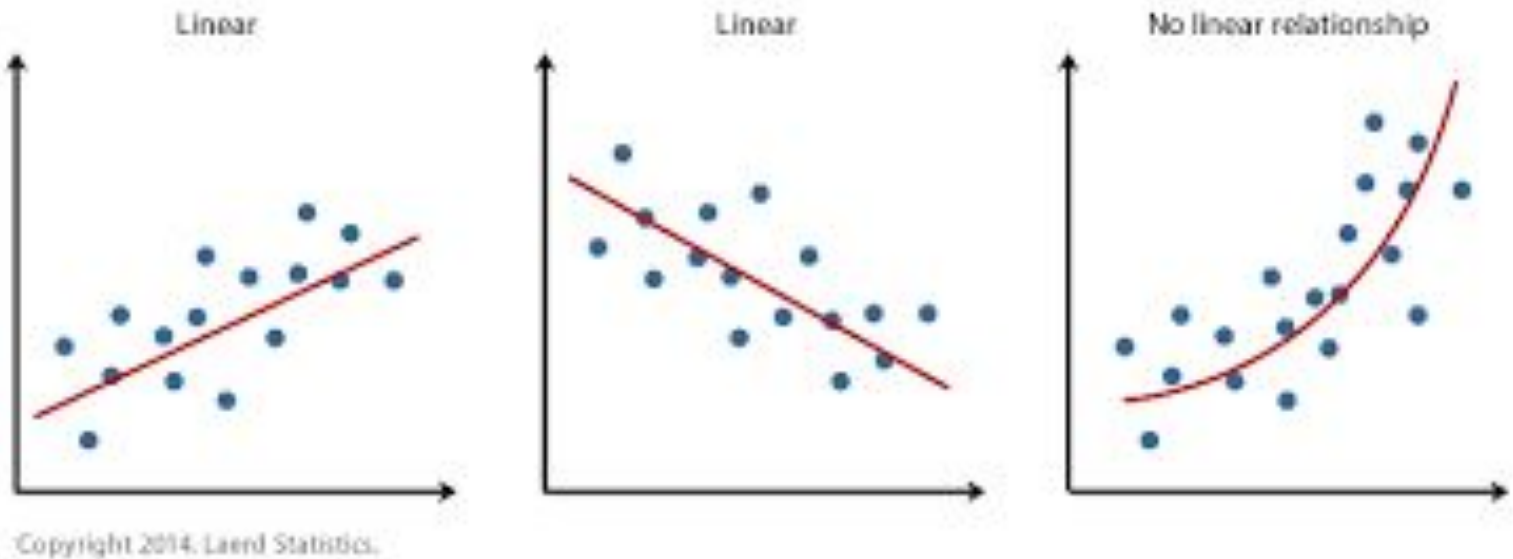
R^2 vs R^2_{adj}

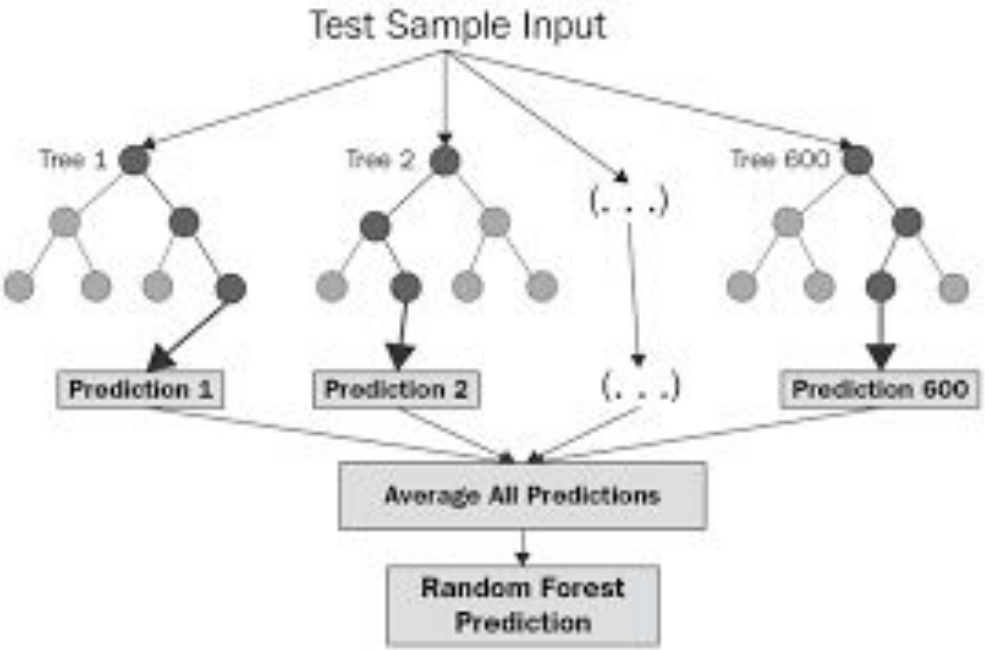
$$R^2 = \frac{SSR}{SST} = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2}$$

$$R^2_{adj} = 1 - \left[\frac{(1-R^2)(n-1)}{n-k-1}\right]$$

Evaluation Recap

Part 4: Rohan Shanbhag




Test Sample Input
Tree 1
Tree 2
(. . .)
Tree 600
Prediction 1
Prediction 2
(. . .)
Prediction 600
Average All Predictions
Random Forest Prediction

Compare

Mean Average Error:

Mean Squared Error:

R^2 value: