



## MCTA 4362: Machine Learning Random Forest Homework

Name: Muhamad Nurhakimie Thaqif Bin Abdullah Matric No: 2213217

Answer ALL Questions.

### Question 1 (20 Marks)

Please refer to the data set attached which shows the salary of an employee in a company with respect to their position in the company

A) Perform Random Forest Regression on this data and fill in the following table

Number of Trees	R Squared Value
10	0.9704434230386582
50	0.9564799890903117
100	0.9378393331453065
200	0.9528325486829341
300	0.952881406729914
400	0.9522705577500628
500	0.9479703752374872

B) Did increasing the number of trees result in a better regression model? Justify your answer (why regression accuracy did or did not increase).

Increasing the number of trees should improve it until a certain point, beyond that point the model would hit a diminishing return it can be overfitting. We can see it especially at 500 where the accuracy from 10 to 500 has significant gap. This tells us the patterns model already learns most of pattern needed for prediction. But it could be due to noise as the model seems a bit sensitive cause there is fluctuation around 100-200 trees