INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

## MCTA 4362:Machine Learning
## Assignment 1

Name:………………………………………………… Matric No:……………………

**Answer <u>ALL</u> Questions.**

Question 1 (20 Marks)

Please refer to the data set attached which was obtained from a census done throughout various locations in California. The dataset contains the latitude and longitude of a particular suburb/region/location, how old the houses in that area are (housing median age), total number of rooms in that location, total number of bedrooms, the population, number of families (households), the median income of the families (median income), the proximity to the ocean (Near Bay, 1 hour away,etc) and the median house value in that area. Your objective is to develop regression models to approximate the median house value in that area (dependant variables) based on the independent variables that are available

A) Based on this data set select which parameters (independent variables) will you feel will be suitable/useful to form your matrix of features. Justify your answer.

B) Perform the necessary data preparation for applying supervised machine learning algorithms, Describe what data preparation was necessary (approximate missing data, removing outliers, One Hot Encoding, scaling, etc)

C) Perform Multiple Linear Regression on this data. Evaluate and describe your model.

D) Perform Polynomial Linear Regression on this data. Evaluate and describe your model.

E) Perform Support Vector Regression on this data. Evaluate and describe your model.

F) Perform Decision Tree Regression on this data. Evaluate and describe your model.

G) Perform Random Forest Regression on this data. Evaluate and describe your model.

H) Based on your results, determine the best model for this data and justify your answer,