PROBLEM STATEMENT-2(SD03Q02) HARSITHA.N

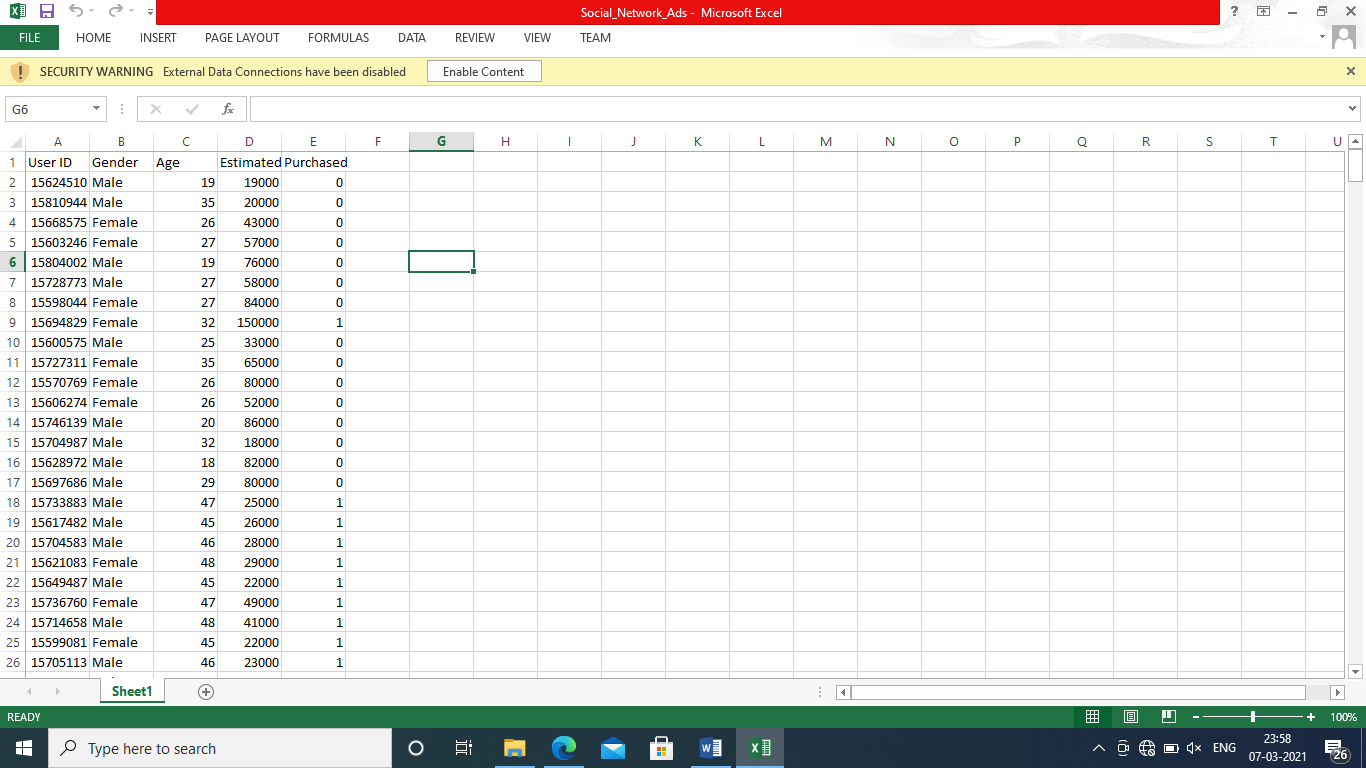
1832022

TASK-2

PROBLEM STATEMENT:

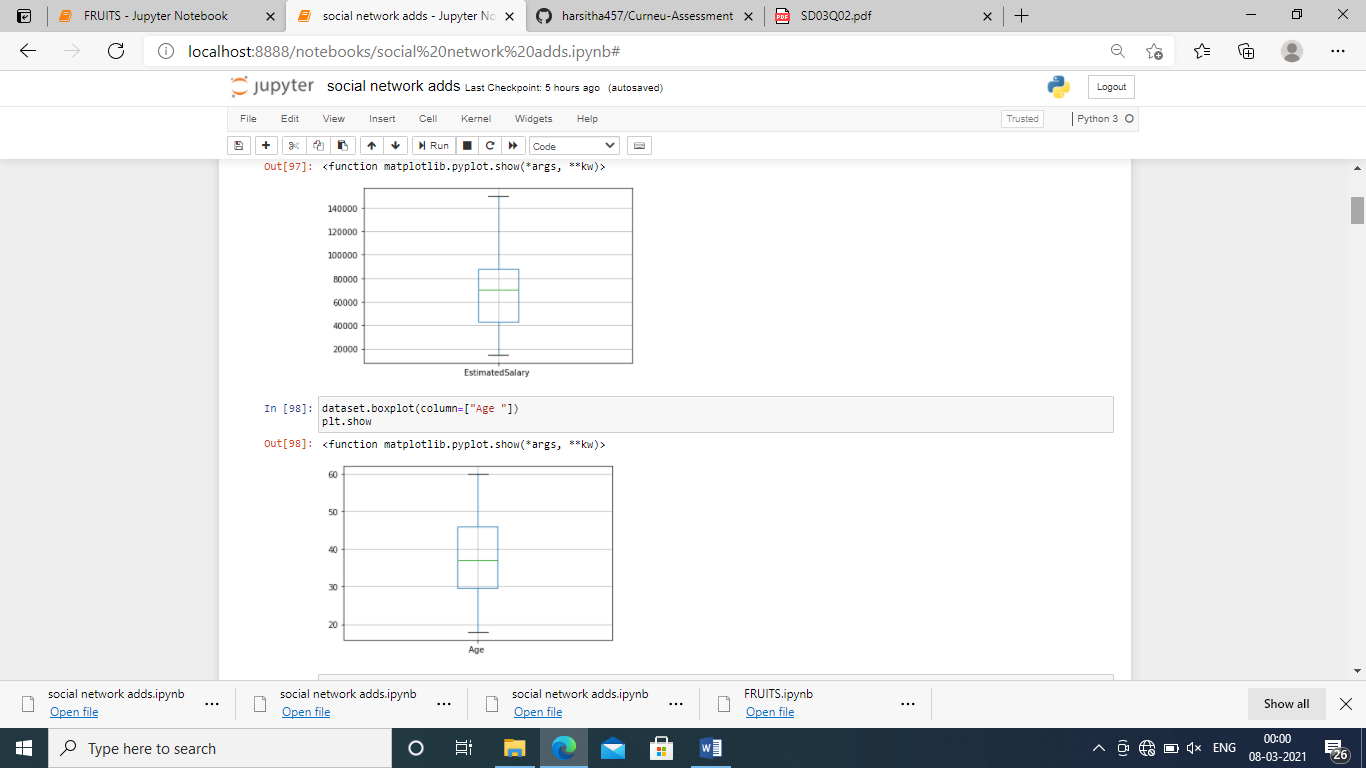
Try to understand the dataset of Social\_Network\_Ads.csv and try to find the best suitable ML algorithm and write the code in python for algorithm from scratch and try to achieve the below output plot.

DATASET SAMPLE:



DATA PREPROCESSING:

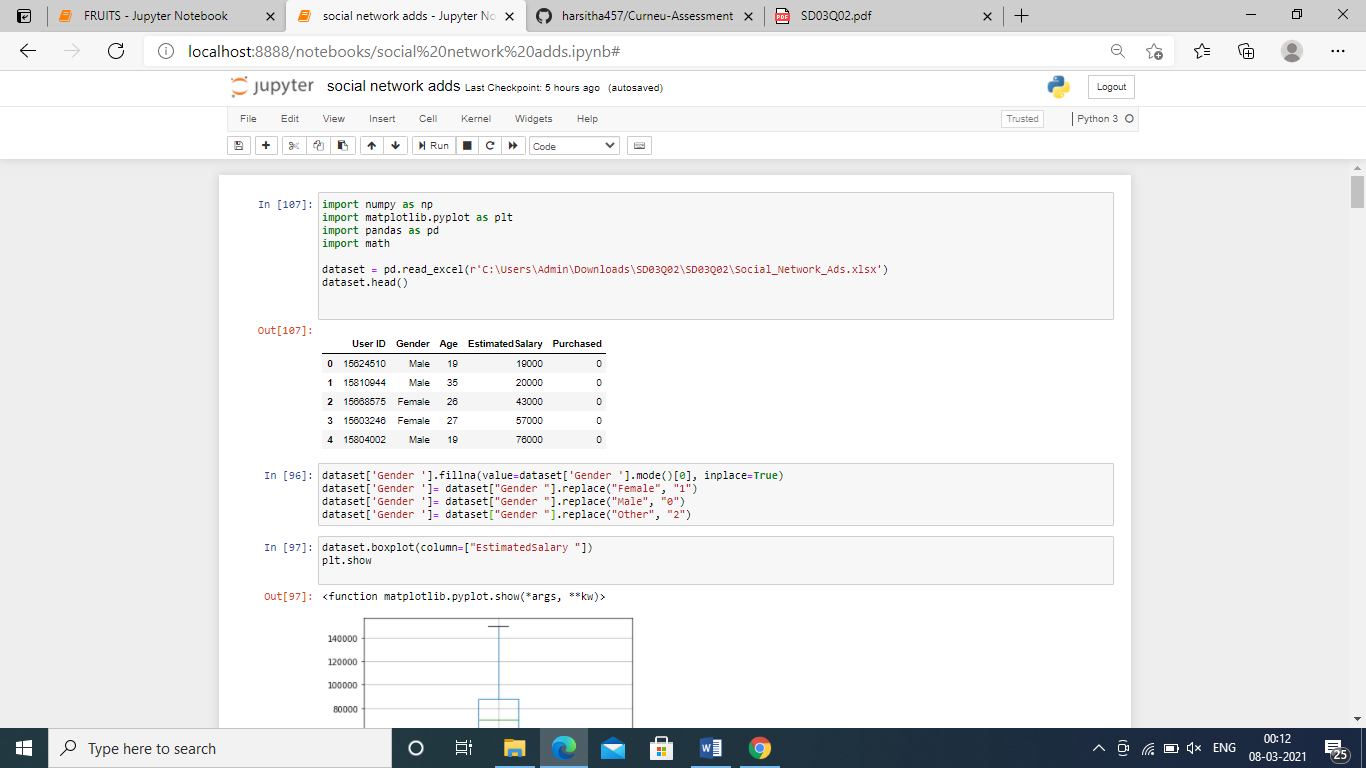
The box plot for each attribute is plotted ( age and estimated salary)



The above graph shows that there are no outliers present in the data.

ENCODING:

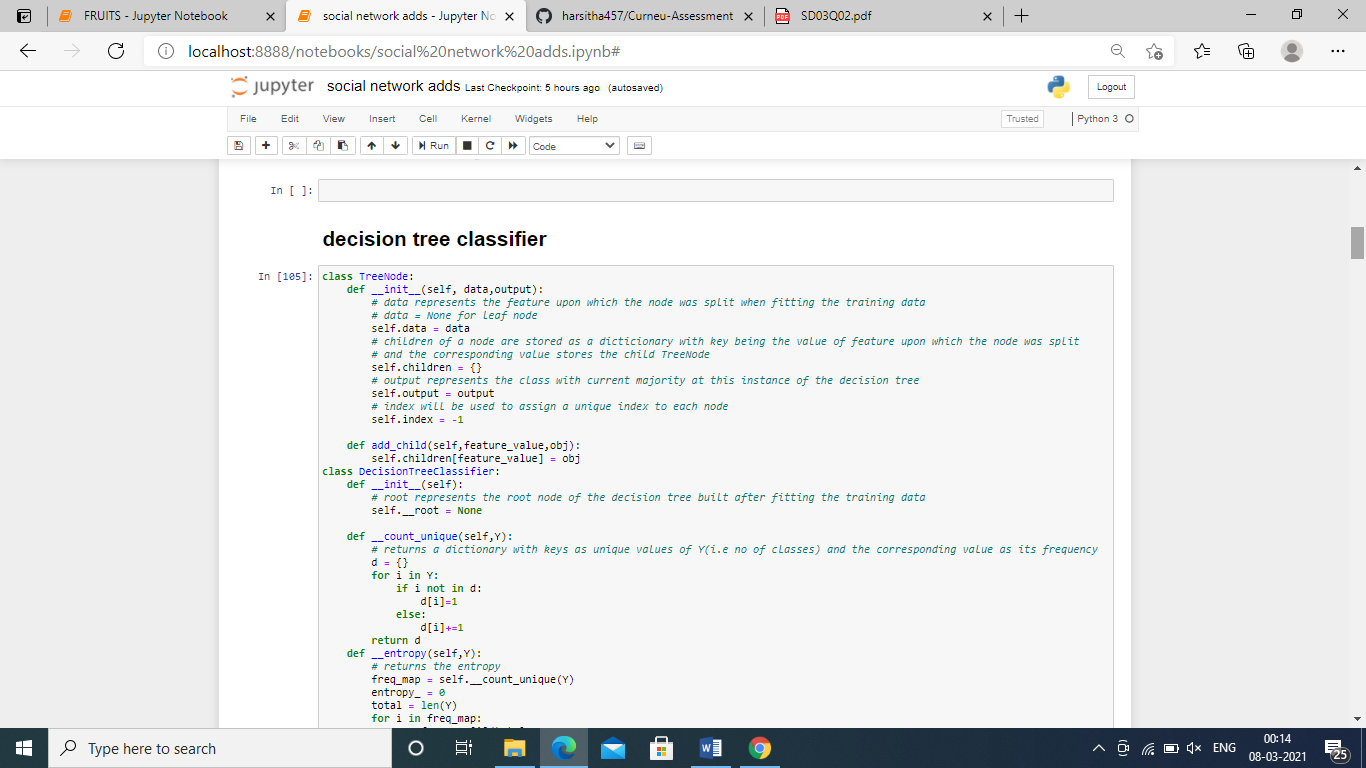
The encoding is done to convert categorical values to numerical values for gender



There is no null values present in the data

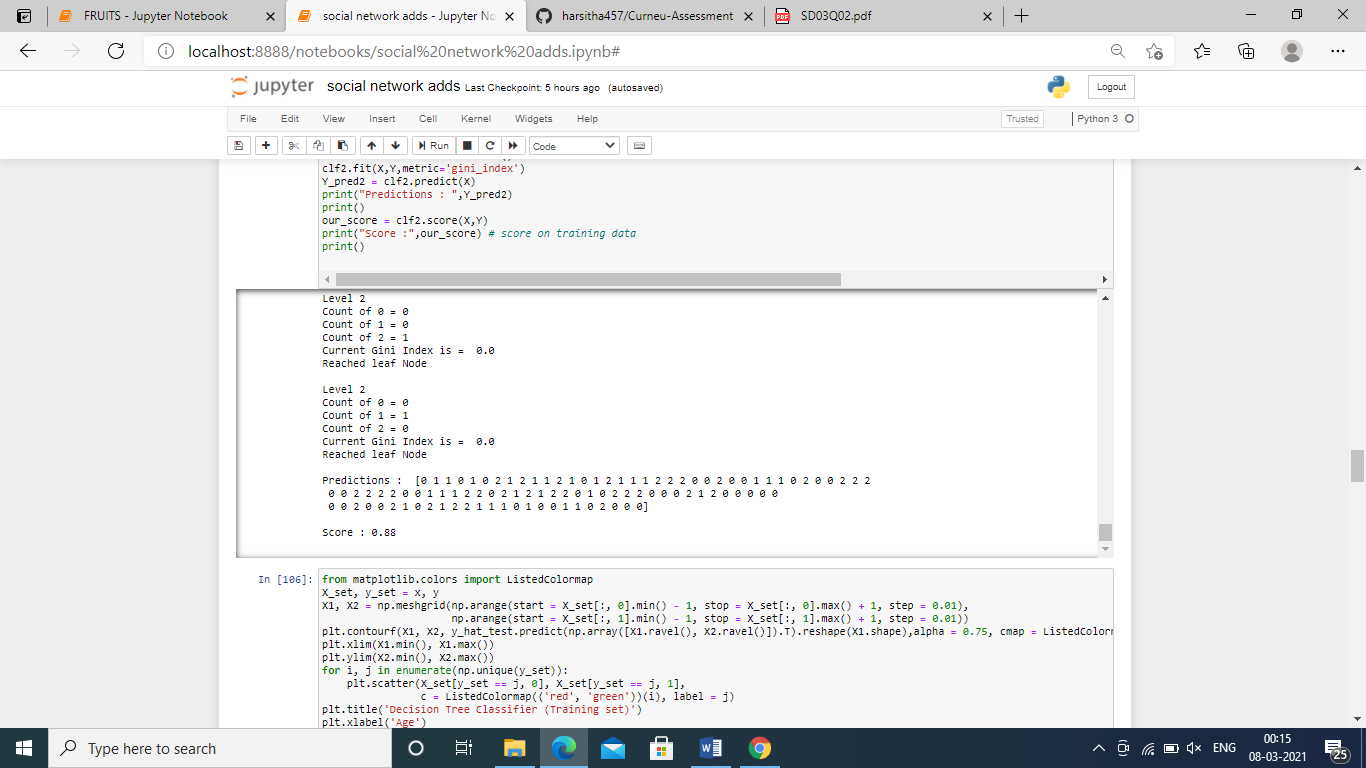
MODEL IMPLEMENTATION:

THE DECISION TREE CLASSIFIER :

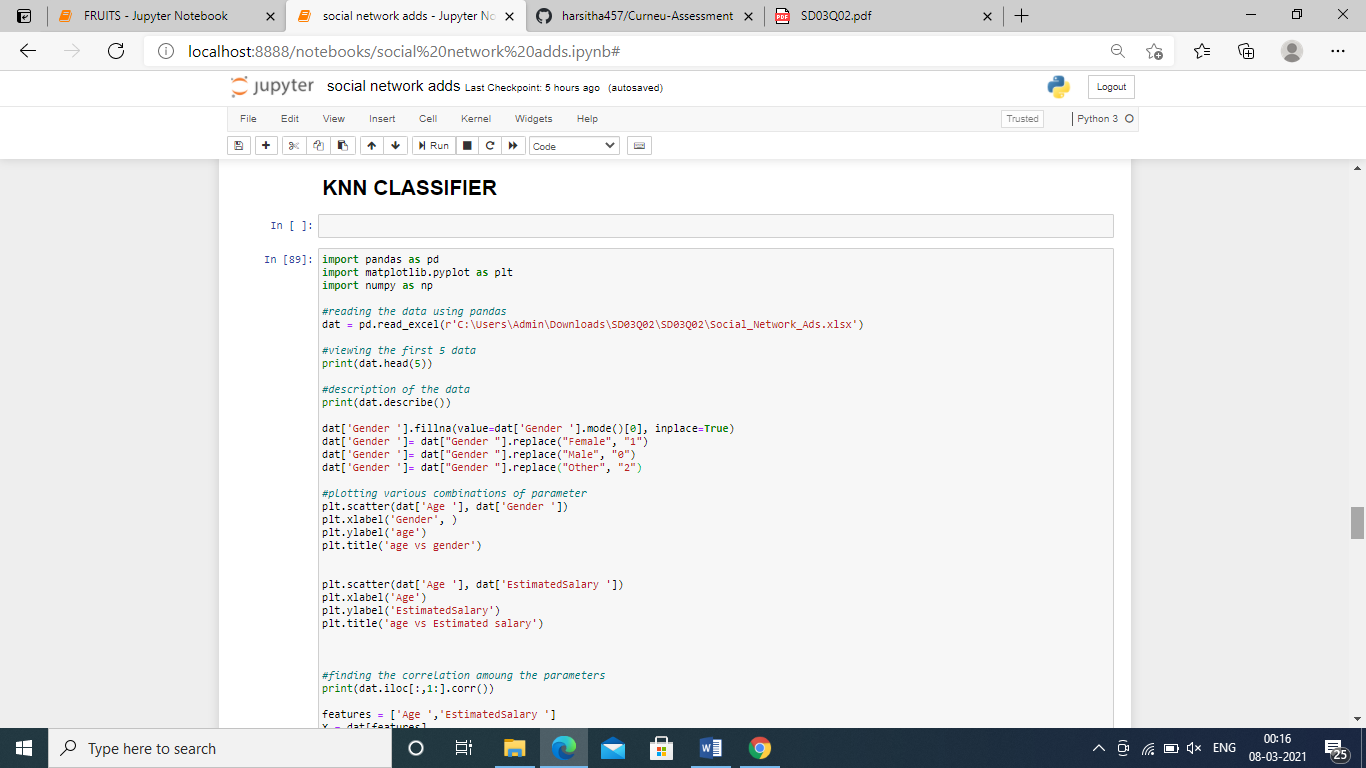


OUTPUT:

The accuracy score is 0.88 with the decision tree classifier

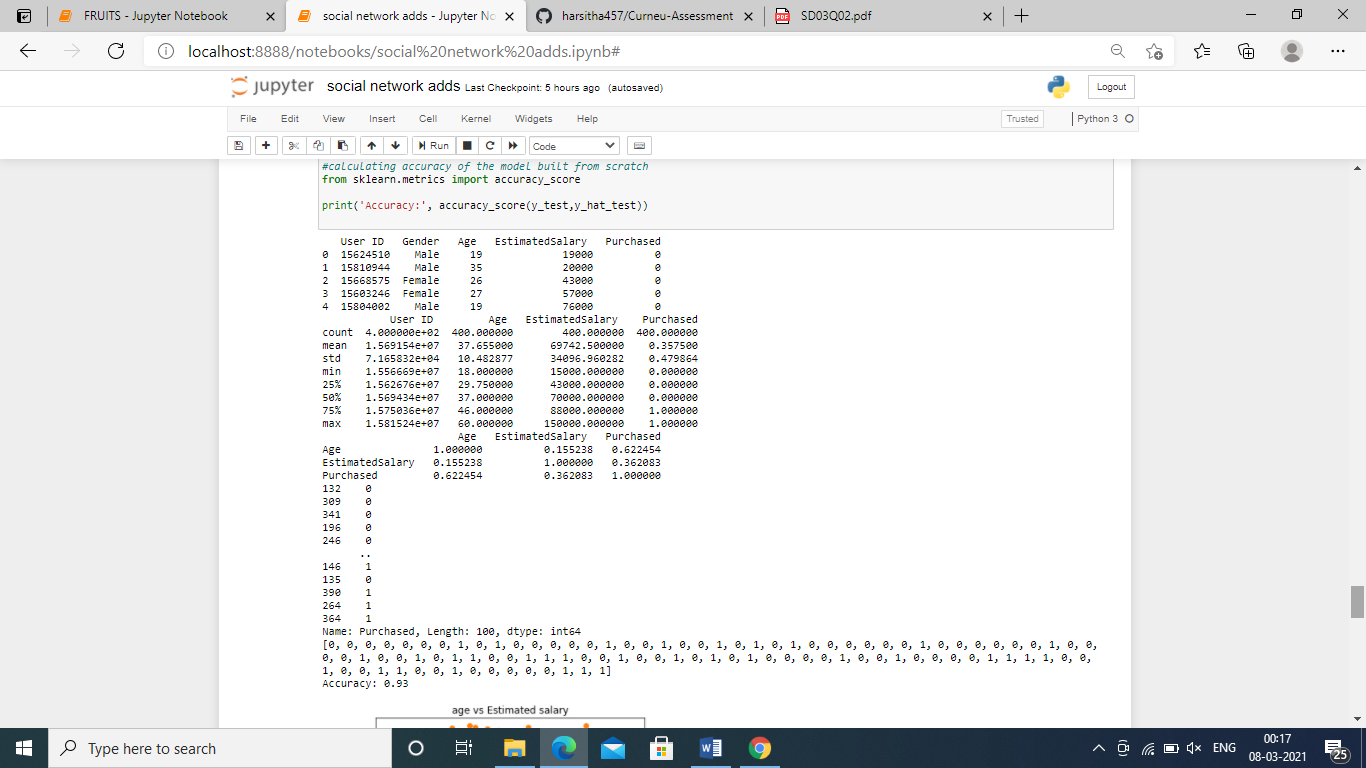


KNN CLASSIFIER:



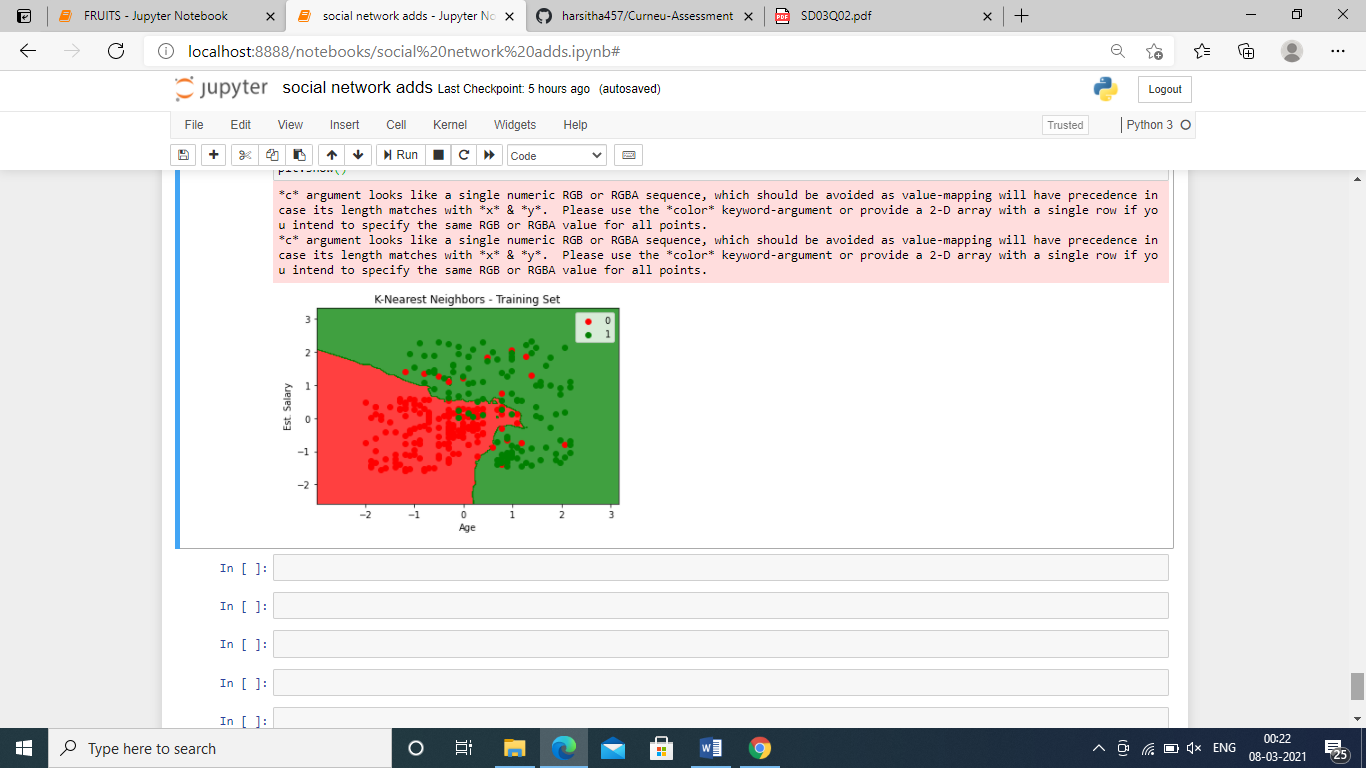
OUTPUT:

The accuracy score obtained by knn model is found to be 0.93



REQUIRED PLOT OBTAINED :

Through knn algorithm the required plot for the dataset is obtained for training data:



CONCLUSION:

The plot achieved through knn classifier as the knn model and it is considered as best model among decision tree and knn model. The knn model yeilds more accuracy than decision tree model hence the knn model classifies the data well.