Curneu MedTech Innovations Assessment -Task 2

SOCIAL NETWORK ADS ANALYSIS

- MANOJ KUMAR P (1832032)

AIM:

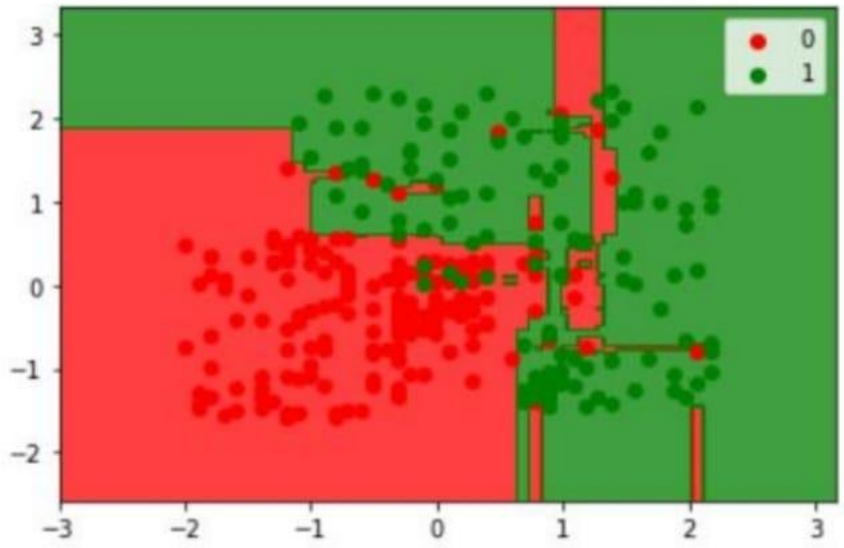
To predict whether the customers purchase the products listed on social network based on other components in the dataset.

DESCRIPTION:

The data consist of attributes such as User id, gender, age, estimated income, purchased. Purchased is the dependent variable. Here we need to find whether the user purchased the product based on the given information. In order to predict the output we need to find the key factors and perform data pre processing and we need to fit the best model for this data.

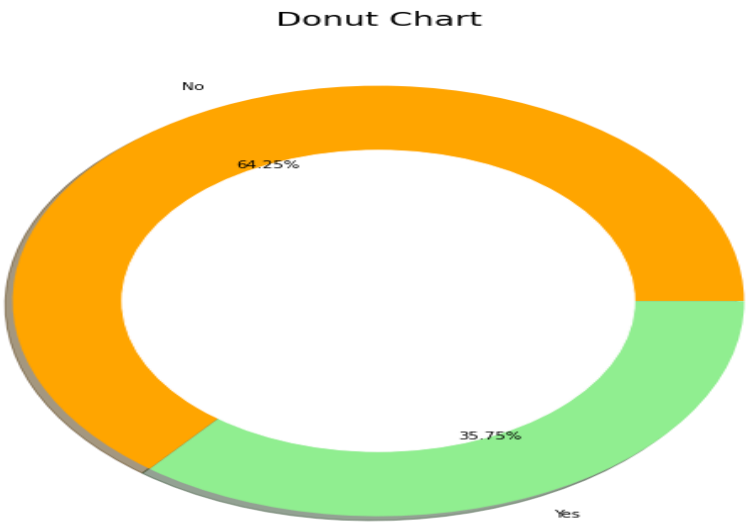
PROBLEM STATEMENT:

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.

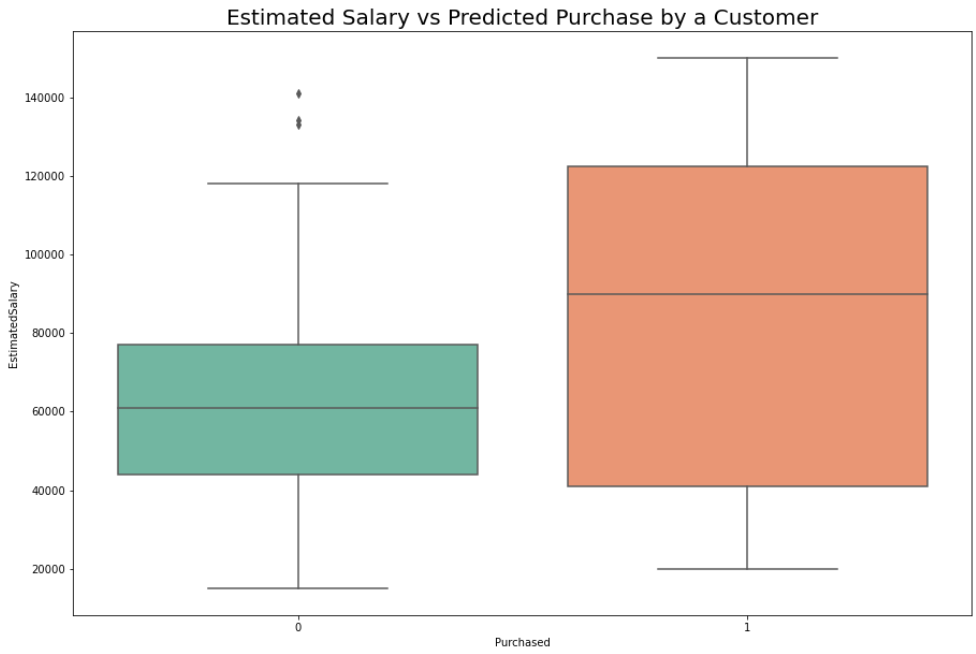


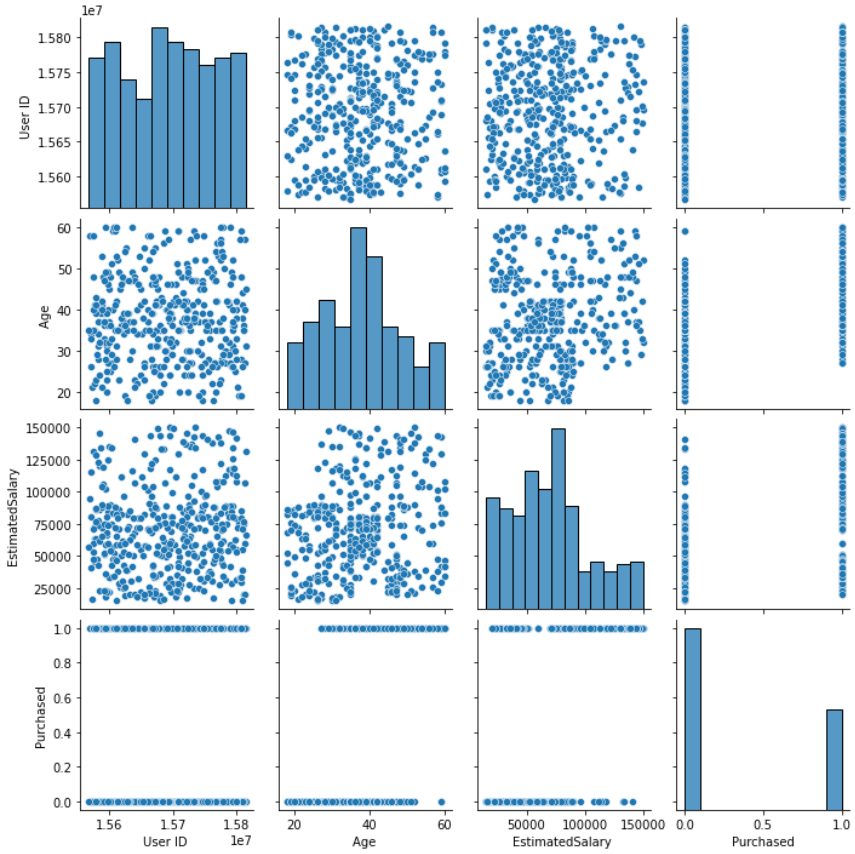
APPROACH TO THE PROBLEM:

First the given dataset (social_network_xlsx) is read using pandas package.The dataset has one dependent attribute (purchased) and remaining independent attributes.The donut chart is made to know the percentage of purchased and not purchased people.

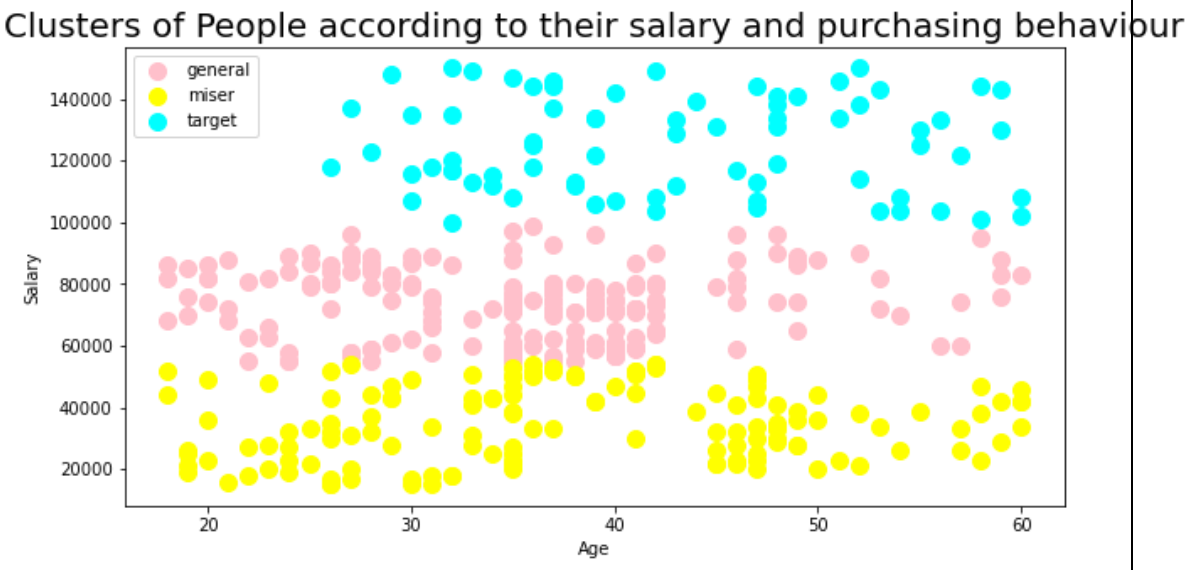


And then we used boxplot to now whether there are any outliers.



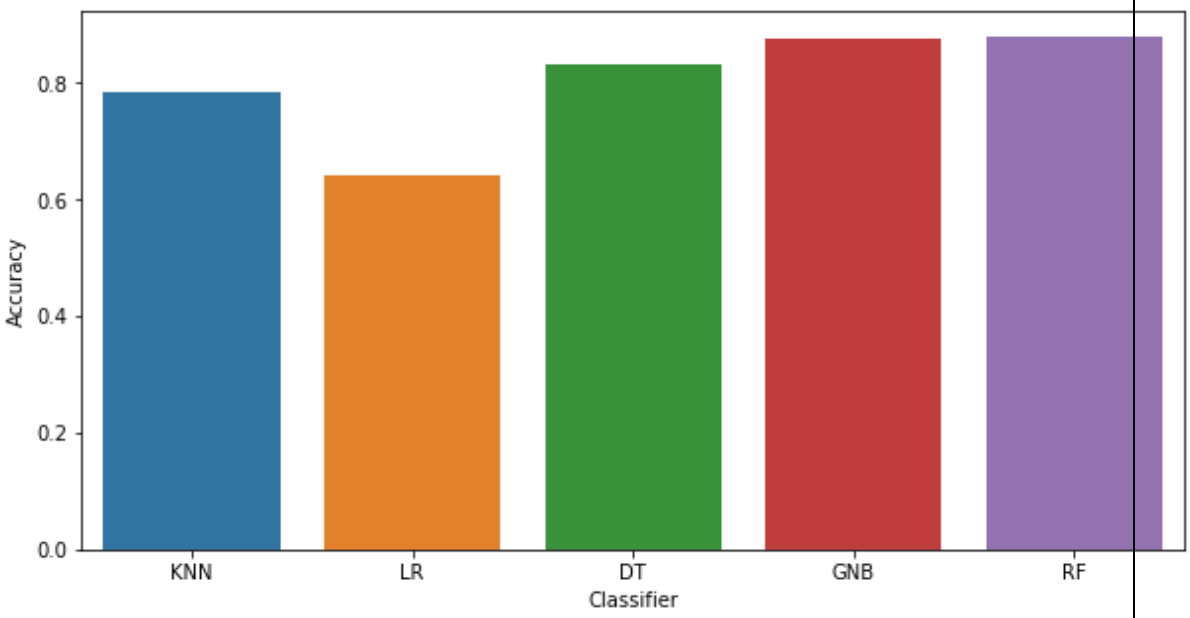


From the pairplot I looked whether homoscedascity exists



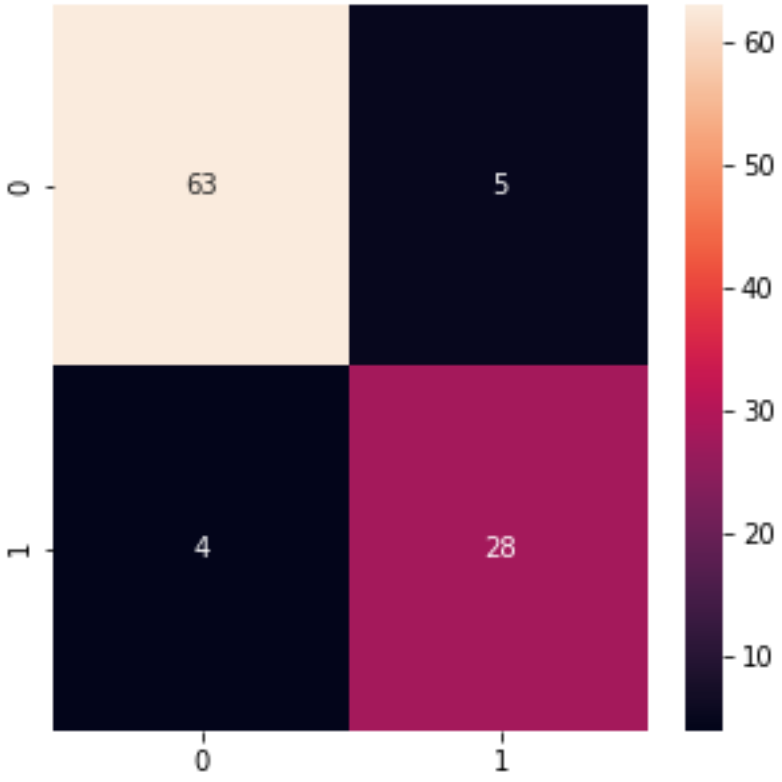
With this cluster graph,we can see how any types of people purchasinng behaviour.

And then to check which algorithm suits better we trie every algorithm and found ththe best fitting algorithm



From this graph we can understoodd that random forest has great accuracy over other algorithms such as knn,logistic regression,decision tree,gausian naïve bayes.

And after bulding the random forest from scratch we checked for accuracy



This is the confusion matrix obtained from the random forest

Then to understand it visually we plotted th graph of the prdicions of random forest

Hence , the given graph is obtained