**PROBLEM STATEMENT 2**

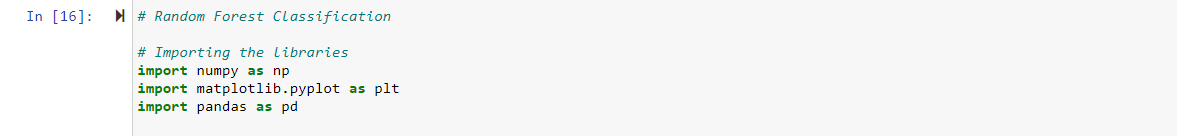
Try to understand the dataset of Social\_Network\_Ads.csv and try to find the best suitable ML algorithm and write the code in python for algorithm from scratch and try to achieve the below output plot.

**INTRODUCTION:**

Random forests or random decision forests are an [ensemble learning](https://en.wikipedia.org/wiki/Ensemble_learning) method for [classification](https://en.wikipedia.org/wiki/Statistical_classification), [regression](https://en.wikipedia.org/wiki/Regression_analysis) and other tasks that operate by constructing a multitude of [decision trees](https://en.wikipedia.org/wiki/Decision_tree_learning) at training time and outputting the class that is the [mode](https://en.wikipedia.org/wiki/Mode_(statistics)) of the classes (classification) or mean/average prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of [overfitting](https://en.wikipedia.org/wiki/Overfitting) to their [training set](https://en.wikipedia.org/wiki/Test_set). Random forests generally outperform [decision trees](https://en.wikipedia.org/wiki/Decision_tree_learning), but their accuracy is lower than gradient boosted trees. However, data characteristics can affect their performance.

**CODE:**

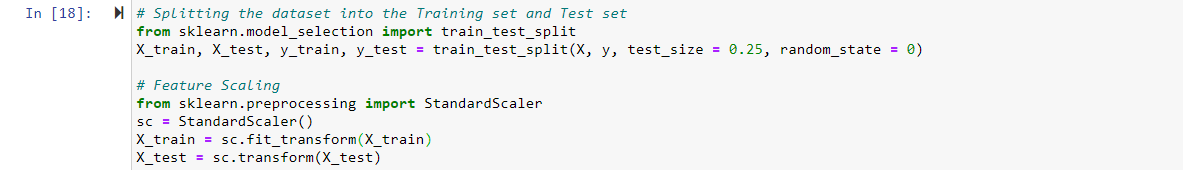
**IMPORTING THE LIBRARIES:**



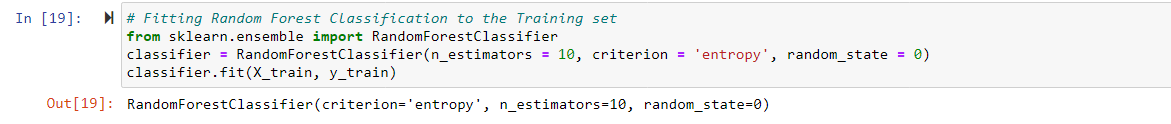
**IMPORTING THE DATASET:**



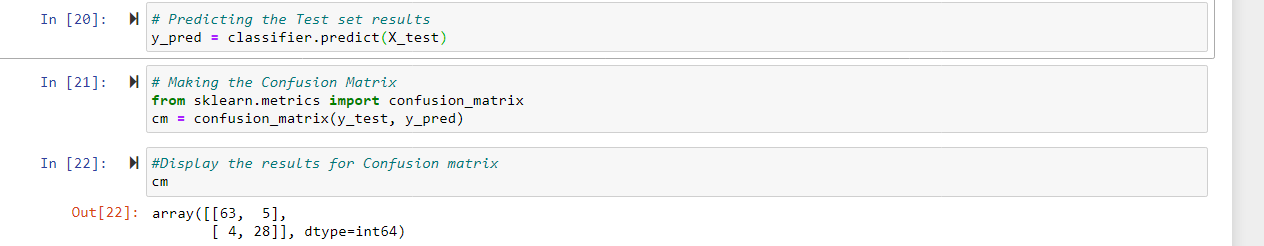
**SPILTTING THE DATASET INTO TEST AND TRAINING DATA**



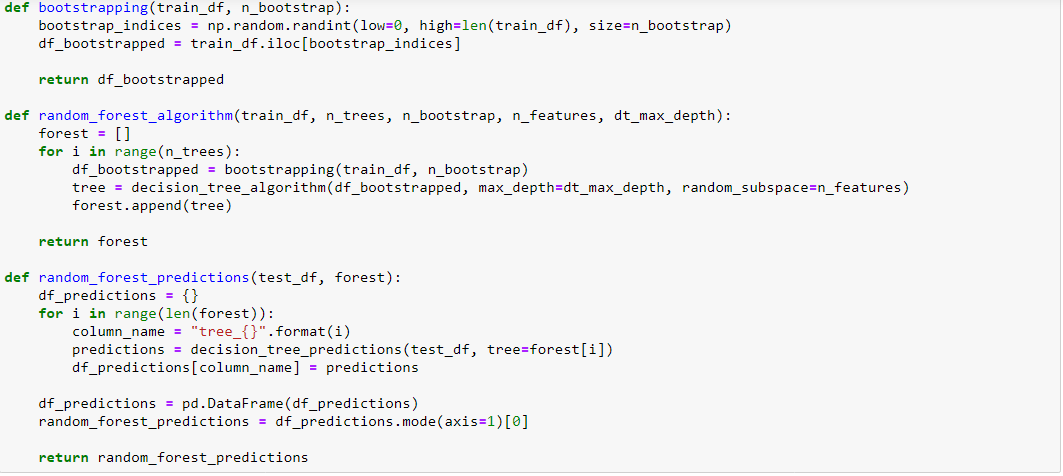
**FIITING THE RANDOM FOREST CLASSIFIER**



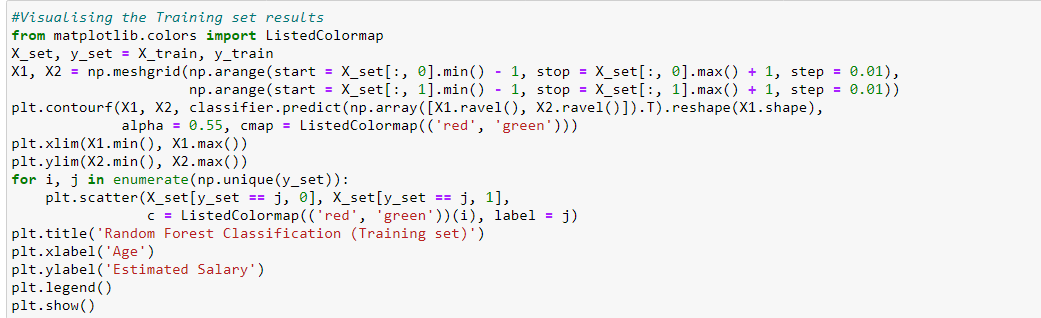
**PREDICTING THE RESULTS AND MAKING CONFUSION MATRIX**

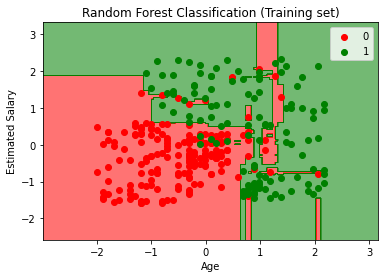


**RANDOM FOREST IMPLEMENTATION FROM SCRATCH**

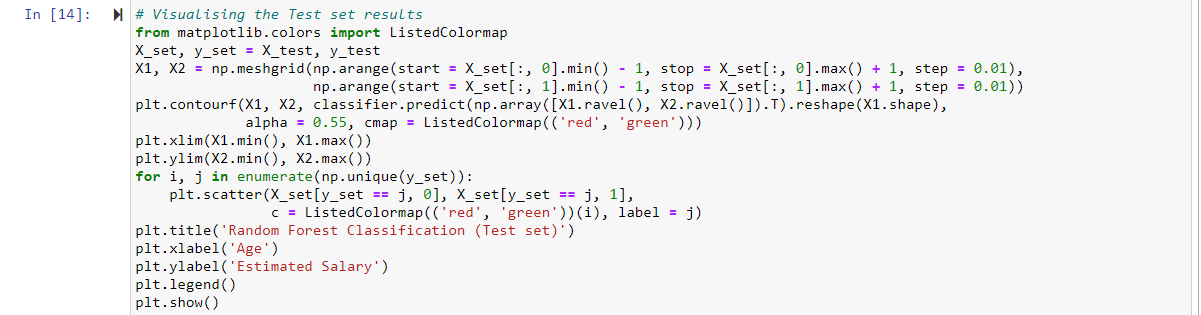


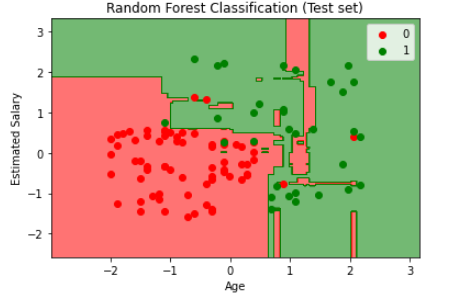
**VISUALISING THE TRAINING SET RESULTS**



****

**VISUALISING THE TEST SET RESULTS**





**CONCLUSIONS:**

The random forest counted the number of trees that counted Yes(user buys suv) and number of trees that counted NO(user doesn’t buy suv) and then takes the prediction that was voted the most times. In this test data set visualization , we can observe that the model captured irrelevant patterns which doesn’t have user’s corresponding to that particular region. This is called overfitting. Overall, the classifier is good for predictions.