**SOCIAL\_NETWORK\_ADS**

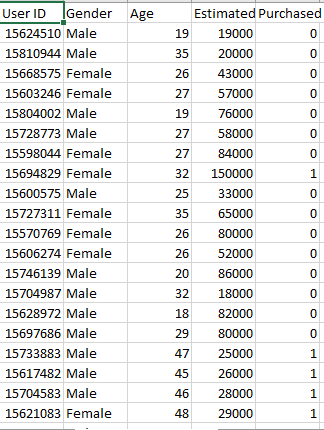
**AIM:**

I have used random forest classifier for the Social\_Network\_Ads problem.

**IMPLEMENTING THE RANDOM FOREST CLASSIFIER:**

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. Here we have done the same that is we constructed a decision tree first and then we call the function into the random forest.The greater number of trees in the forest leads to higher accuracy and prevents the problem of over fitting.

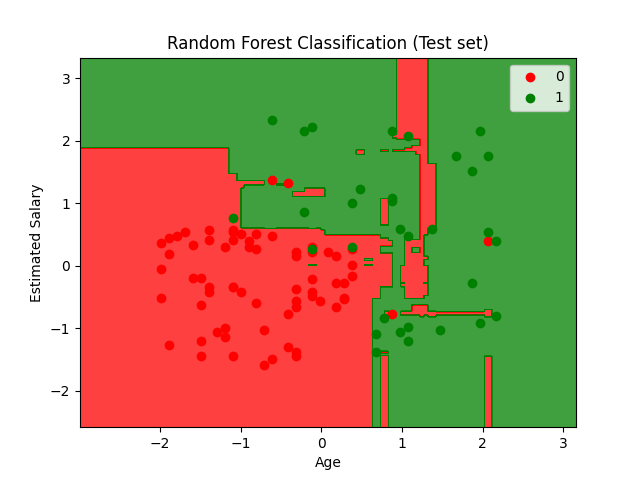
**SAMPLE DATASET:**

****

**DATASET DESCRIPTION:**

The dataset contains the real time data of the Social\_Network\_Ads with the corresponding factors that indicates that the patient has diabetes. This data set contains 5 columns “User ID ”, “Gender ”, “Age ”, “EstimatedSalary ”, “Purchased” Containing these features the “Purchased” feature is the target variable which tells us whether the user has purchased the product or not. “0” corresponds to the decision that the user not purchased and “1” represents that the user has purchased.

**OUTPUT:**

****

**INFERENCE:**

From the graph we can infer that most of the data points are predicted correctly and rarely a few data points are misinterpreted. But we can see that there are some 12 red spots in the green region which infers that the classifier needs more essential attributes to perform well and boost up the classifier.

**CONCLUSION:**

But while there is always room for improvement, we can be satisfied with this model as our final product. Our accuracy is high, but not so high that we need to be suspicious of any over fitting. We can safely say that an increase in both Age and Estimated Salary will lead to a higher probability of clicking the advertisement. As new users sign-up for the website, we can use this model to quickly determine whether or not to expose them to this particular ad or choose another that is more relevant to their profile