**Social Network Ads**

**Problem Statement:**

This project tells whether a user on Social Networking site after clicking the ad’s displayed on the website end’s up buying the product or not. This could be helpful for the company selling the product. This is done with the help random forest classifier algorithm.

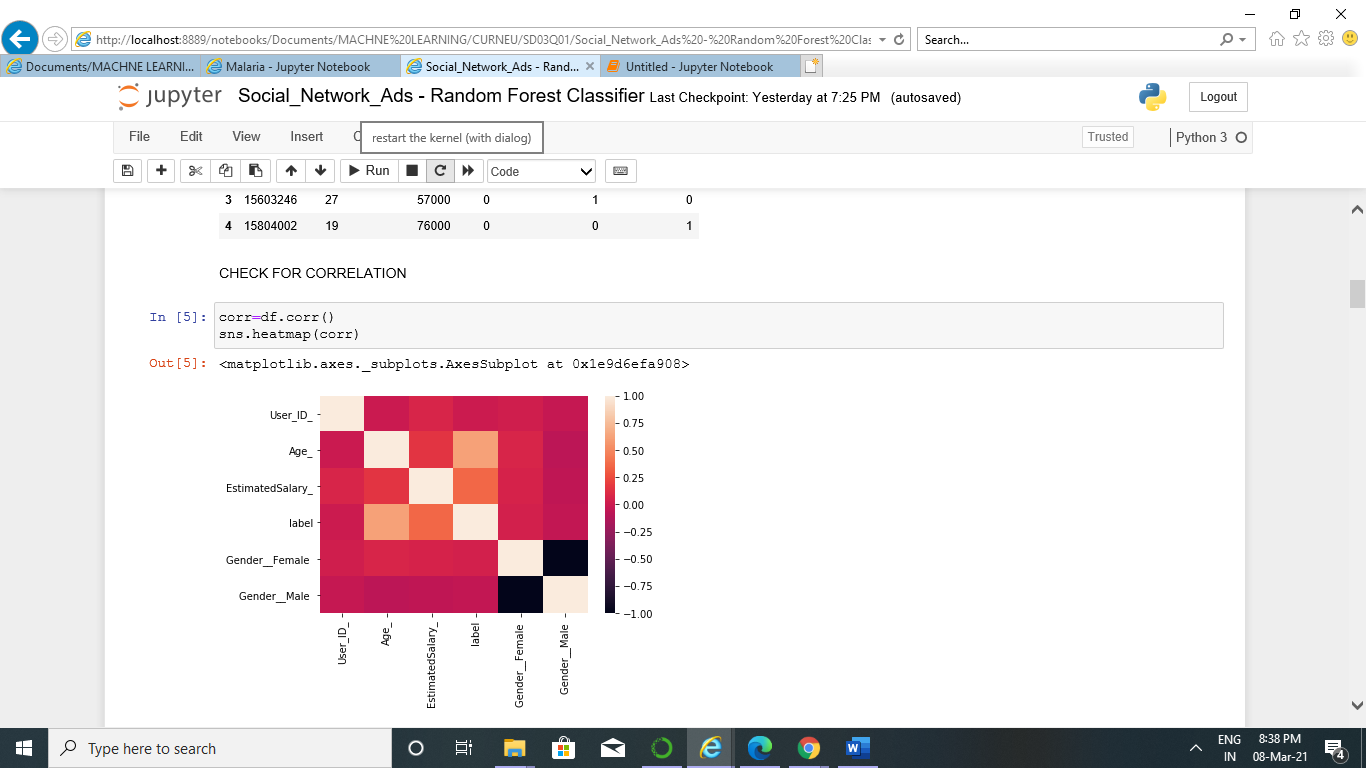
**Dataset:**

This dataset consists of five features

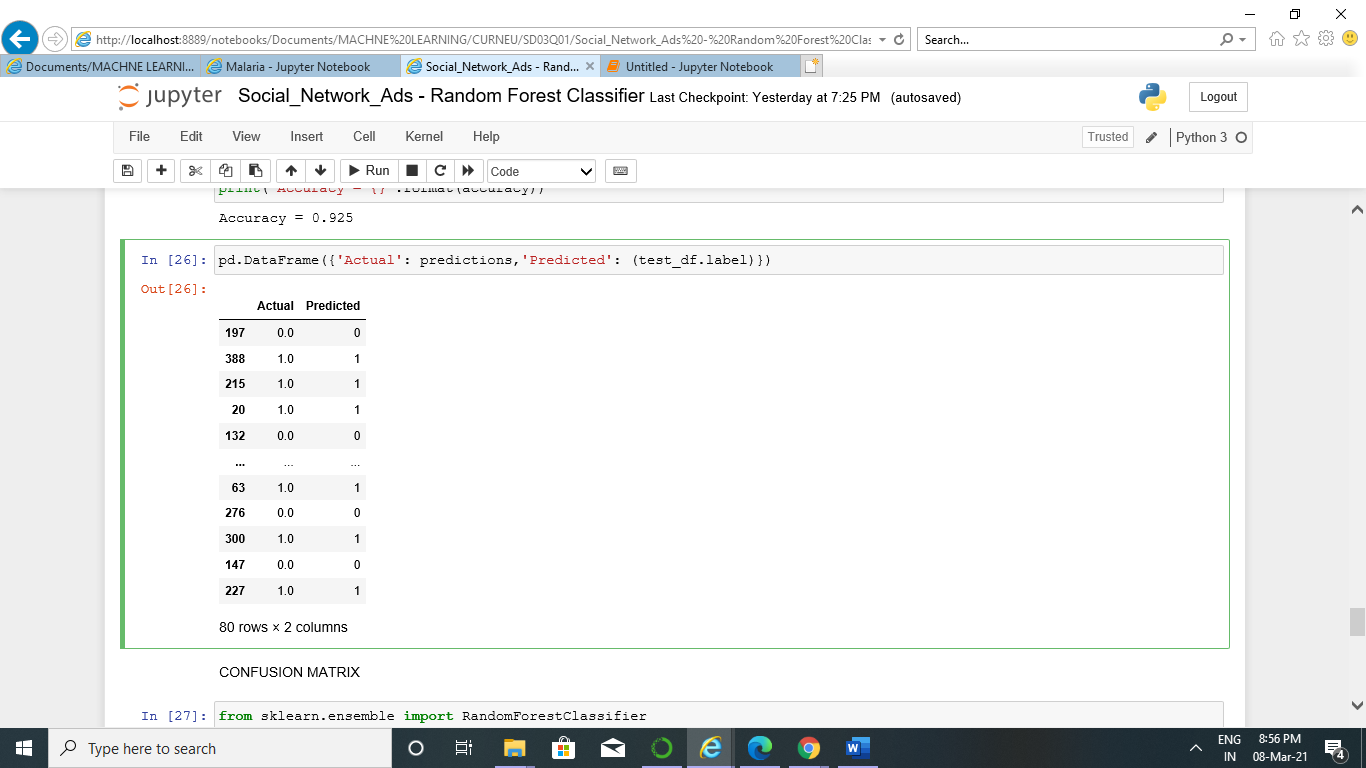
1. User ID
2. Gender
3. Age
4. Salary
5. Purchased

**Interpretation:**

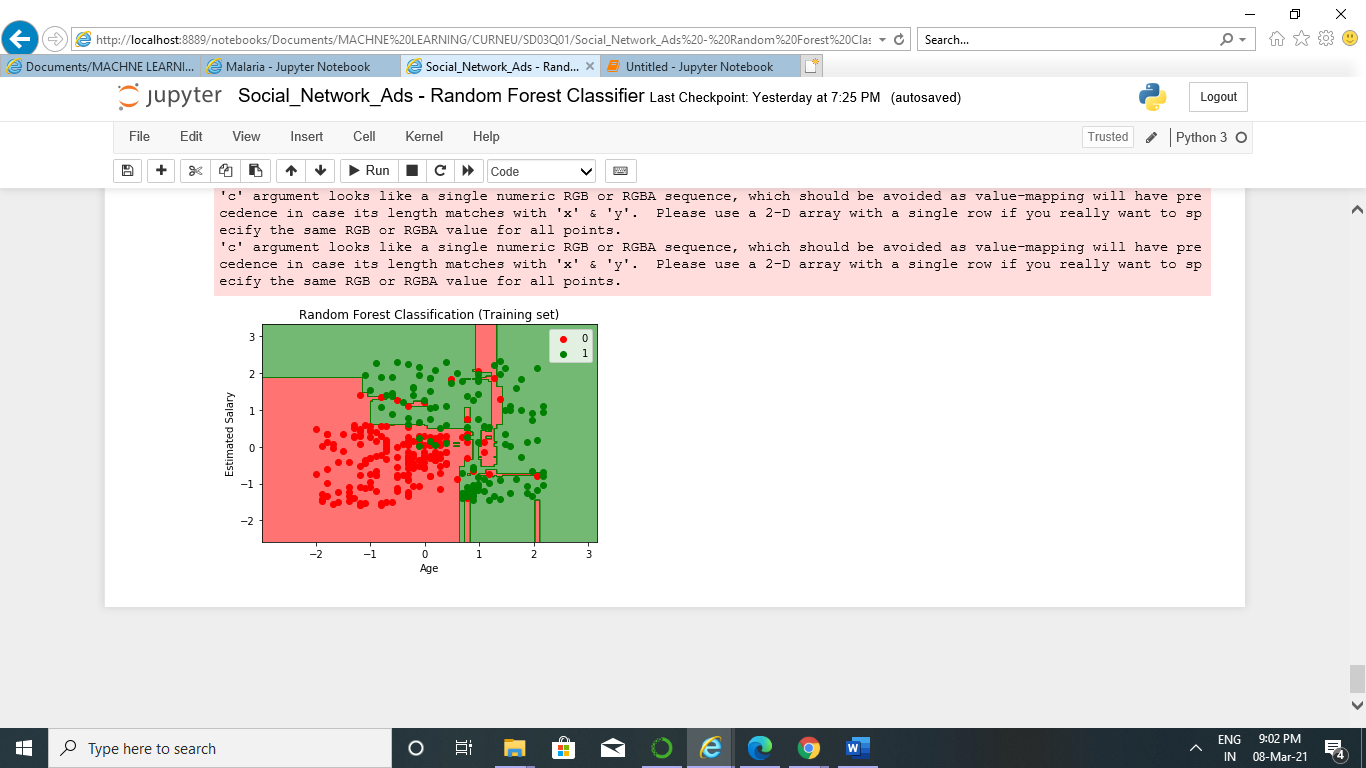
* Libraries are imported and the data is loaded using the pandas library.
* We check whether the dataset has null values. The dataset doesn’t have any null values hence we proceed with further operations.
* Looking into the dataset we infer that it has some categorical values hence it is converted to numeric values using get dummies function.
* On checking with the correlation of the features we see that the gender and user id are not correlated, and it seems like it doesn’t affect the data. And hence we drop both the gender and the user id column from the dataset.



* Some basic operations are to be included in the random forest algorithm. This is done by defining the operations in python definitions. This includes definition for checking the purity, accuracy, to classify data, getting potential splits, to calculate entropy, overall entropy, to determine best split.
* After this the decision tree algorithm is coded from scratch since this the base for random forest classifier. Then decision tree algorithm is also coded including the prediction definition.
* Then we predict the results from the random forest classifier model and we see that the accuracy of our model is 92.5%.



* Then the mesh grid plot is plotted using the random forest model.



* Since the accuracy is 92.5% we can infer that random forest classifier is the best fit for the given dataset social network ads. Hence using this algorithm we can predict like which customer would buy the product and who will not purchase the product.