PROJECT – iPHONE PURCHASE

CASE:

Suppose that a survey is conducted in a mobile store where we collect information about the customers who buy iPhone and who do not buy an iPhone. This information is our dataset. The ultimate intent of this project is to build classifier models which predict if a customer who walks into the store will buy an iPhone or not.

In this project, I have built 7 different types of classifier models namely **GaussianNB, Decision Tree, K- Nearest Neighbors, Logistic regression, Support Vector Machine (SVM), Random Forest** and **XGBoost**. To evaluate these models, I have used 2 different types of evaluation metrics namely, Accuracy Score and Confusion Matrix.

The following are the steps involved:

1. Importing the data
   This is a crucial step. After successfully importing the data, we can import the libraries whenever necessary
2. Convert Gender to number
   Since this project involves dealing with customers who are either male of female, therefore we have to convert the data. Gender is a categorical variable which we have to convert to numerical variable
3. Feature selection and data split
4. Building seven types of classification models
5. Evaluating the built classification models using the evaluation metrics

CONCLUSION:

After successfully building the seven classification models and passing each of them through two evaluation metrics, it is clear from the accuracy score that all are models predict with atleast 90% accuracy.

CLASSIFIER MODEL	ACCURACY SCORE
GaussianNB	0.91
Decision Tree	**0.93**
K- Nearest Neighbors	0.9
Logistic regression	0.9
Support Vector Machine (SVM)	**0.93**
Random Forest	**0.93**
XGBoost	0.92

Carefully observing the confusion matrix and comparing all the accuracy scores, it is clear that decision tree, SVM and random forest are the most accurate models whereas KNN and logistic regression predict with least accuracy.