Assignment Report.

- Surajkumar Yadav

*[please have a look on MAIN\_NB.ipynb file for indepth documentation]*

**● Step 1 : Understanding the business objective or business problem.**

* The main aim was to build a predictive model that can predict whether the customer will buy or not the Term Deposit using Past data. **The data is of a Portuguese bank's Marketing Camping where they comomunicated with their Customers on a phone call for selling a term deposit.**

**● Step 2 : Analyzing and performing EDA on the dataset.**

* The data had around 4521 datapoints about customers and their attributes. There were no null values and no duplicate values in the dataset. Plotted some graphs for the variables and learned about them.
* Also found that the dataset was imbalanced with more values for customers not subscribing to plan than customers subscribing to it.

**● Step 3 : Data Preprocessing.**

* First converted the categorical columns to numerical columns using replace method by pandas and LabelEncoder from sklearn.
* Scaled the values using Standardscaler.

**● Step 4 : Feature Selection.**

* Used Various techniques to find the dominant features that contribute to the target variable.
* Techniques such as correlation, ensemble methods [Extratree Classifier], Information gain based feature selection.
* Dropped some irrelevant features.

**● Step 5 : solving the imbalanced dataset problem.**

* Used SMOTE Method to generate new samples for the Yes Class.

**● Step 5 : Model Building.**

* Build various Classification models and calculated their scores using F1score metric and found that Gradient Boosting performed the best.
* Then found the best Hyperparameters for model using GRIDSEARCHCV .
* Build the final model which had 0.91 F1 score and predicted the result for the test data and saved the file in csv form(Submission.csv).