Bank Marketing Campaign

Your Deep Learning Partner

Taimoor Razi
Akshar Chaklashiya
Ogwu Augustine

Date: 15-09-2022

Outline

- Executive Summary
- Problem Statement
- Approach
- Model Preparation
- Model Building
- Model Results
- Recommendations

Executive Summary

ABC Bank wants to sell it's term deposit product to customers and before launching the product they want to develop a model which help them in understanding whether a particular customer will buy their product or not (based on customer's past interaction with bank or other Financial Institution).

Bank wants to use ML model to shortlist customer whose chances of buying the product is more so that their marketing channel (tele marketing, SMS/email marketing etc) can focus only to those customers whose chances of buying the product is more.

Problem Statement

Data Glacier Objective : Provide actionable insights to help XYZ firm in identifying the right customers for targeting the marketing campaign.

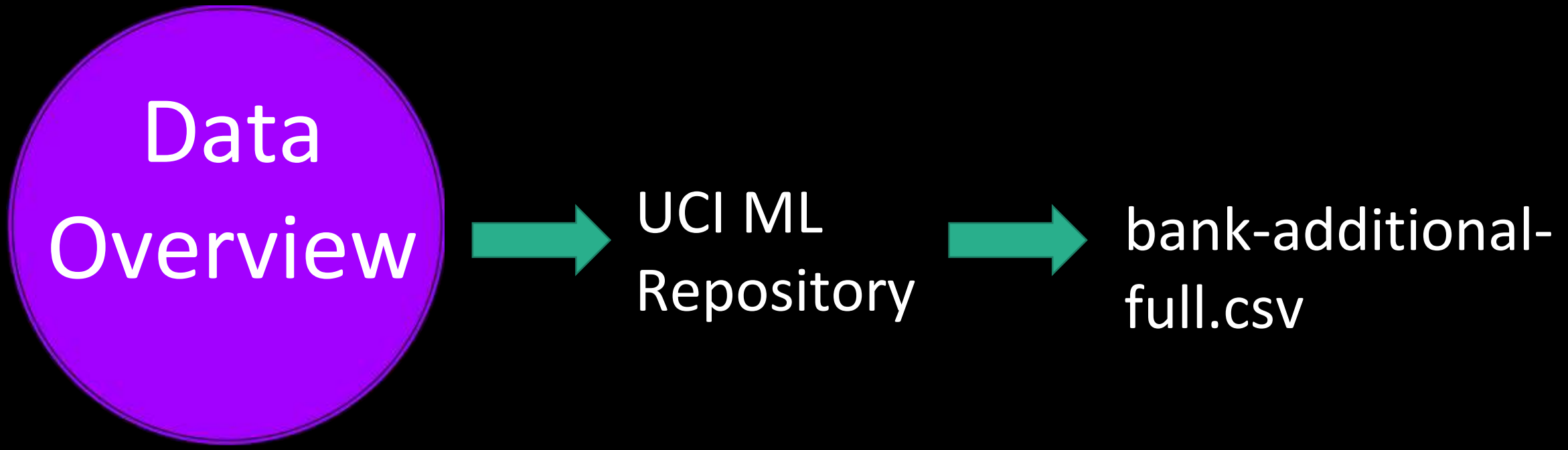
This will save resource and their time (which is directly involved in the cost (resource billing)).

Data Glacier did a 1 month pilot focusing on these tasks:

- Data Intake Report
- EDA Notebook
- ML Model Proposal
- Presentation to ABC’s Executive team (Today)

Approach

1. Data Understanding and Extraction
2. Data Cleaning/Modelling
3. **Model Preparation**
4. **Model Building**
5. **Model Results**


Data Overview
UCI ML Repository
bank-additional-full.csv

Data Overview

Bank Data

- 21 Features
- 41188 data points
- 5834924 bytes

Model Preparation

- Removed Default column
- Handle Outliers: box plot to detect outlier.
- 6 of the categorical variables have an "unknown" value.
- Replaced them with the most frequent category for the column

Model Preparation

- Logarithmic Transformation is applied to the "age" variable in order to get a more Gaussian-like distribution.

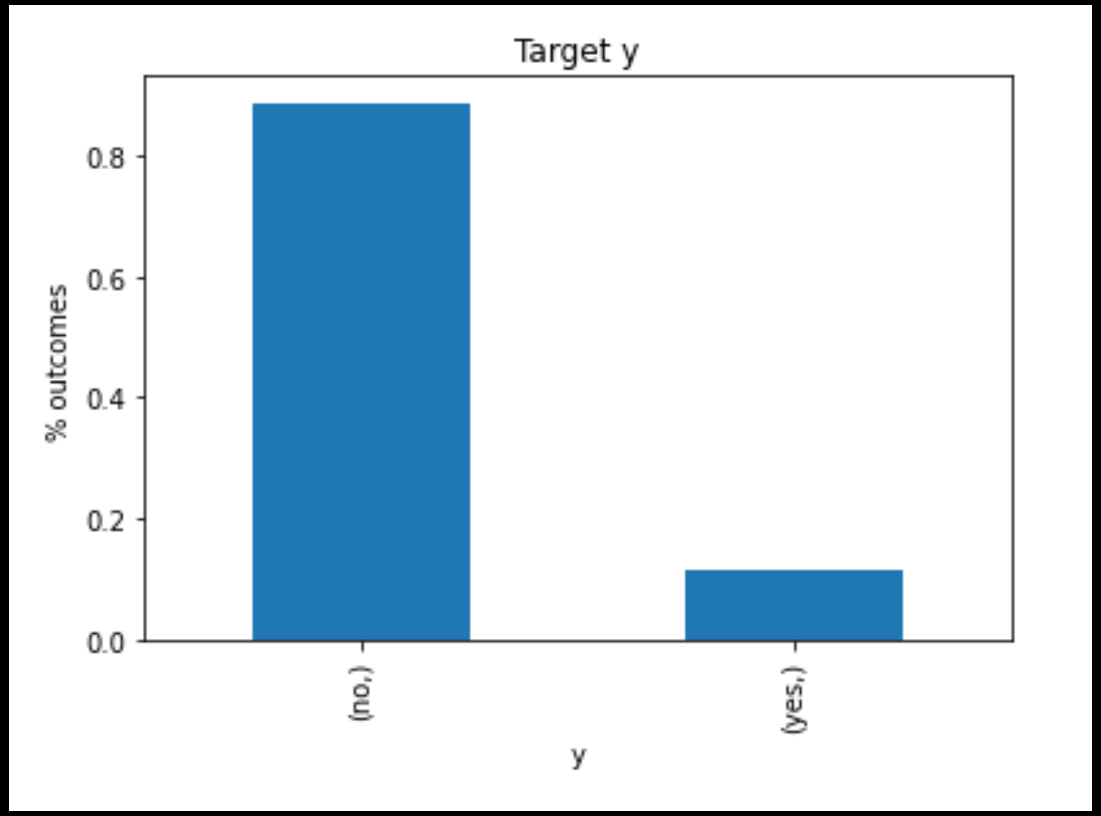
- Yeo-Johnson transformation is applied to duration variable to obtain gaussian like distribution.

- Binary features, house and loan, values of “yes” and “no” are replaced with “1” and “0”.

- Ordinal Encoding is done for education column

Model Preparation

- Imbalance Data: Over Sampling



Model Building

Random Forest

The process of this method is:

1. Take a sample of size n from the training dataset;
2. Randomly choose p variables from all the variables available;
3. Train a single big tree on the sample dataset and using p variables;
4. Repeat the step above B times;
5. Take a majority vote of the results for all of the B trees.

Model Building

Random Forest

- RandomForestClassifier class from sklearn is used to create the model.
- Hyperparameter tuning is done over n_estimators and max_depth of the ensemble through grid search with five-fold cross-validation.

Model Building

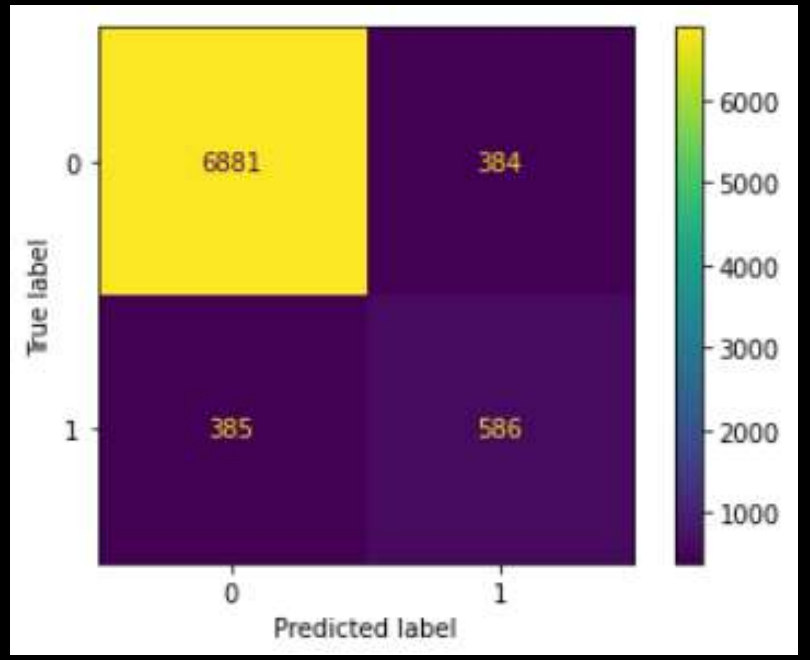
Random Forest

- RandomForestClassifier class from sklearn is used to create the model.
- Hyperparameter tuning is done over n_estimators and max_depth of the ensemble through grid search with five-fold cross-validation.

Model Results

Random Forest

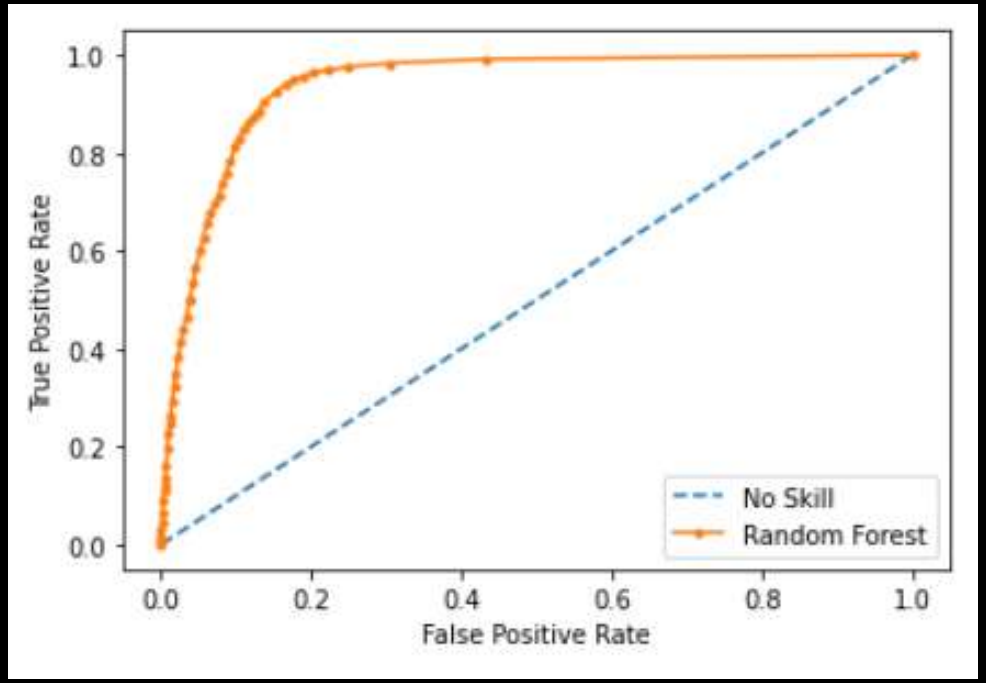
- The test accuracy of 90.66% was obtained with a ROC AUC score of 93.8%.
- The confusion matrix



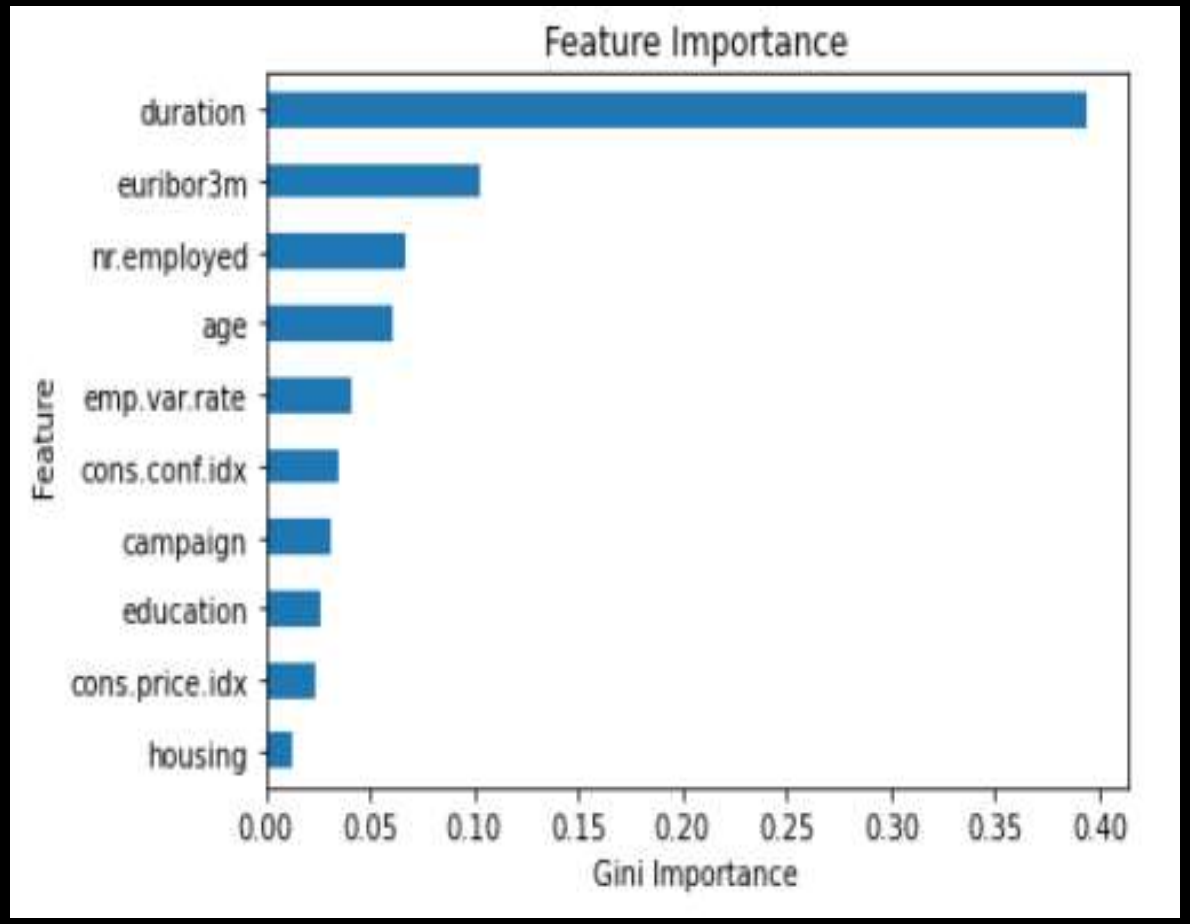
Model Results

Random Forest

- The test accuracy of 90.66% was obtained with a ROC AUC score of 93.8%.
- The ROC AUC Curve



Feature importance for Random Forest



Model Building

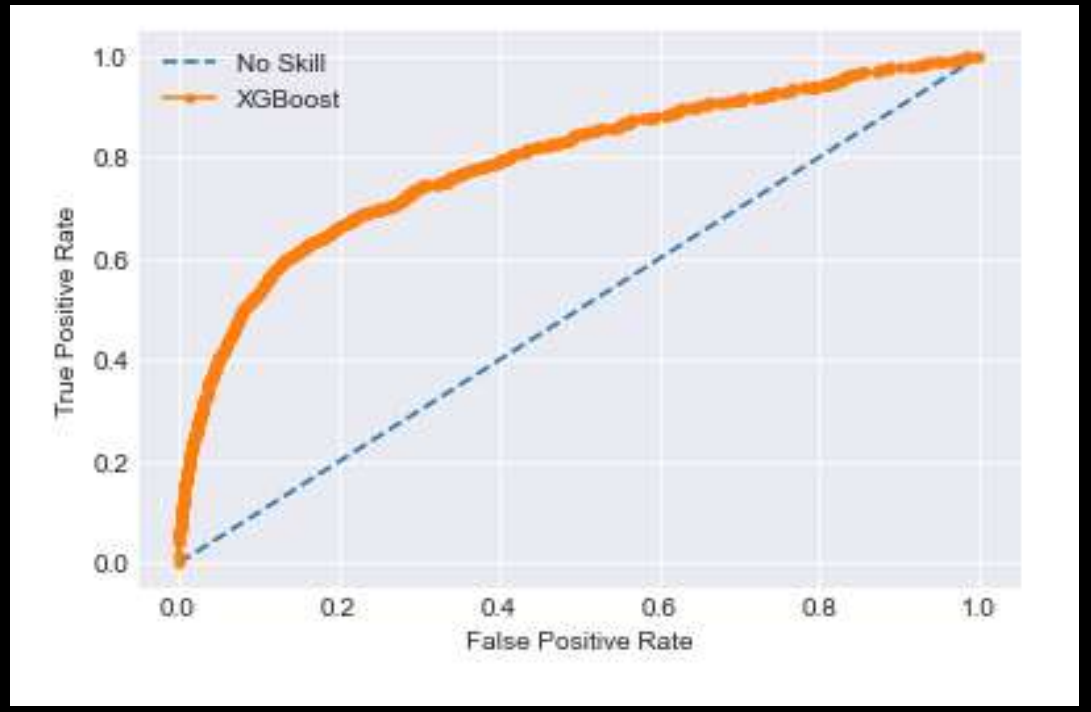
Extreme Gradient Boosting

- Popularly called XGBoost;
- model works to accurately predict a target variable by combining the estimates of a set of simpler, weaker models;
- Tree based ensemble machine learning algorithm- scalable machine learning system for tree boosting;
- Hyperparameter tuning - carried out over learning_rate, n_estimators, max_depth, colsample_bytree, and subsample through random search with a 10-fold cross validation;

Model Results

Extreme Gradient Boosting

- Test accuracy : 82.4% , Train Accuracy: 78.9%
- ROC AUC score: Without hyper-param: 62.9%
- ROC AUC score: with hyper-param: 60%



Model Building

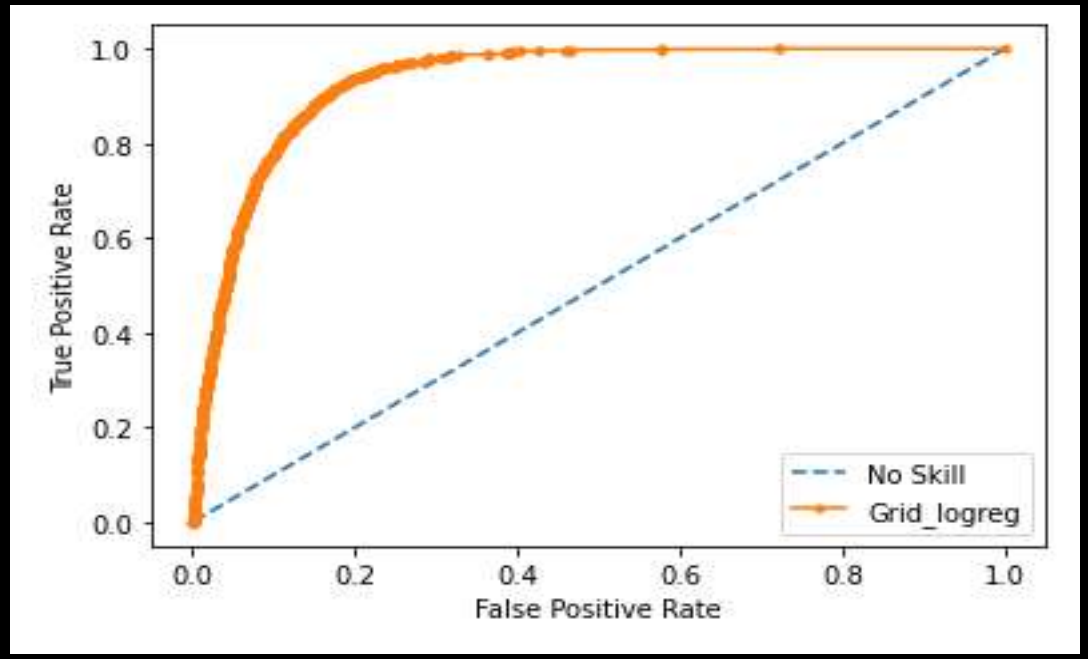
Logistic Regression

- one of the most popular Machine Learning algorithms;
- It is used for predicting the categorical dependent variable using a given set of independent variables
- It predicts the output of a categorical dependent variable;
- we have set an upper limit to train our model up to 3000 iteration, not to overfit our model;
- Then we used the GridSerchLogistic algorithm as our final training algorithm.

Model Results

Logistic Regression

- Test accuracy : 85.4%, Train Accuracy: 87.9%
- ROC AUC score: 86.4%



Model Summary - Insights

Random Forest

Best Test Accuracy

Random Forest

Best ROC AUC Score

euribor3m

Most influential variable after duration

Recommendations

ABC bank should:

- Hire more people to work for them
- Improve the quality of conversation on the phone
- Run their campaigns when interest rates are high and macroeconomic environment is stable.
- Target old age groups

Reference

1. How to deal with outliers in python -. ProjectPro. (n.d.). Retrieved September 30, 2022, from https://www.projectpro.io/recipes/deal-with-outliers-in-python

2. Detecting and treating outliers: How to handle outliers. Analytics Vidhya. (2022, July 21). Retrieved September 30, 2022, from https://www.analyticsvidhya.com/blog/2021/05/detecting-and-treating-outliers-treating-the-odd-one-out/%20

3. Handling outliers in python. Handling Outliers in Python. (n.d.). Retrieved September 30, 2022, from https://www.datasciencesmachinelearning.com/2018/11/handling-outliers-in-python.html

4. Aleksandradeis. (2019, February 8). Bank marketing analysis. Kaggle. Retrieved September 30, 2022, from https://www.kaggle.com/code/aleksandradeis/bank-marketing-analysis

5. Sklearn.linear_model.logisticregression. scikit. (n.d.). Retrieved September 30, 2022, from https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

Thank You!