

BANK MARKETING CLASSIFICATION

Presented by:

Nada & Rawan

INTRODUCTION

The data is related to direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same client was required, in order to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed.

PROJECT WORKFLOW

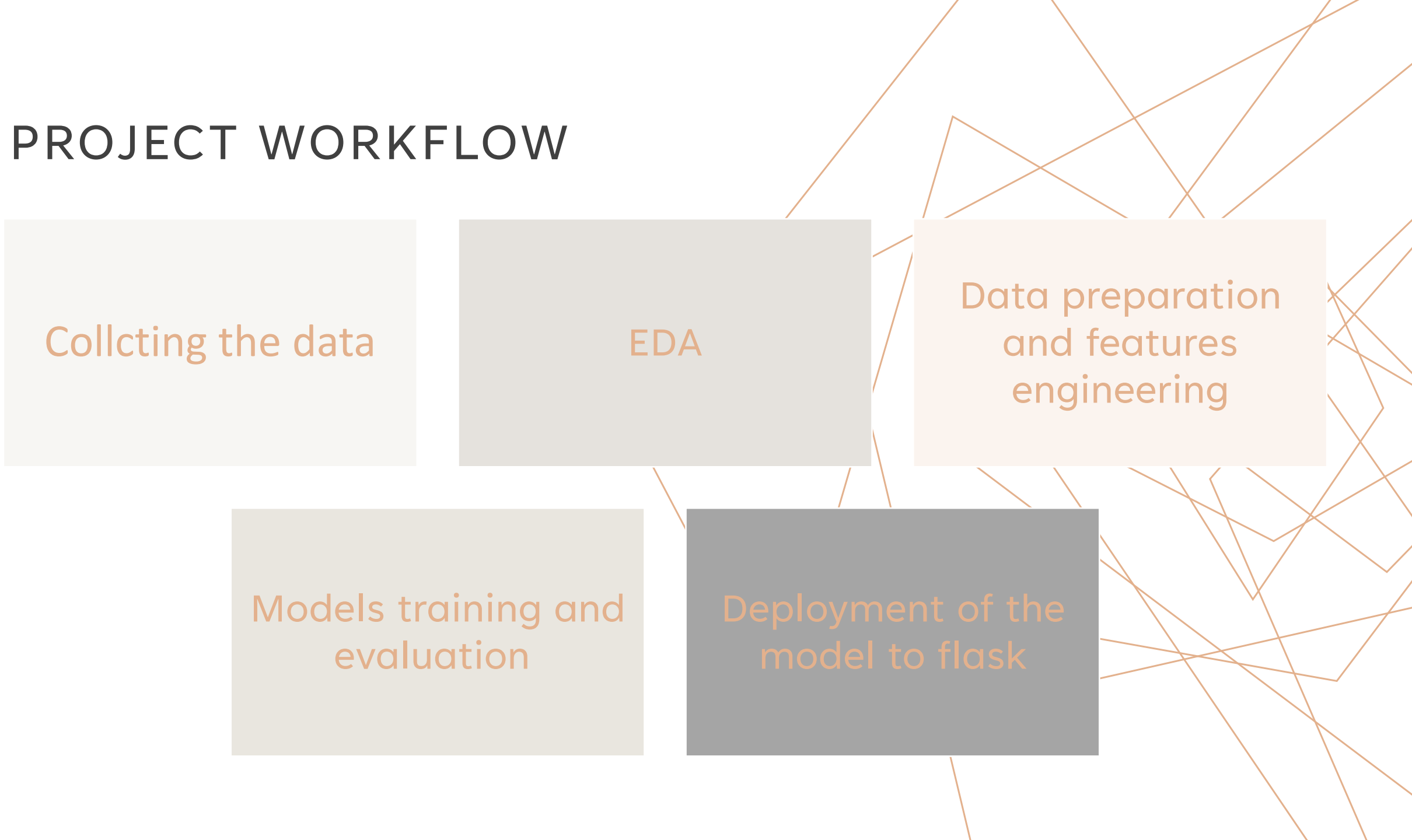
Collecting the data

EDA

Data preparation
and features
engineering

Models training and
evaluation

Deployment of the
model to flask



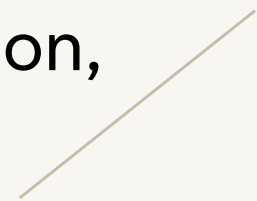


DATA DESCRIPTION

We collected the data from uci.edu website

We had 45212 rows and 17 columns in total

Age, job, marital, education, default, housing, loan, Contact, month, day_of_week, duration, campaign, pdays, previos, balance



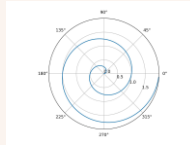
TOOLS:



Jupyter



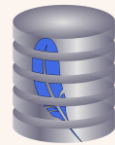
Git



Matplotlib



Seaborn



SQLite

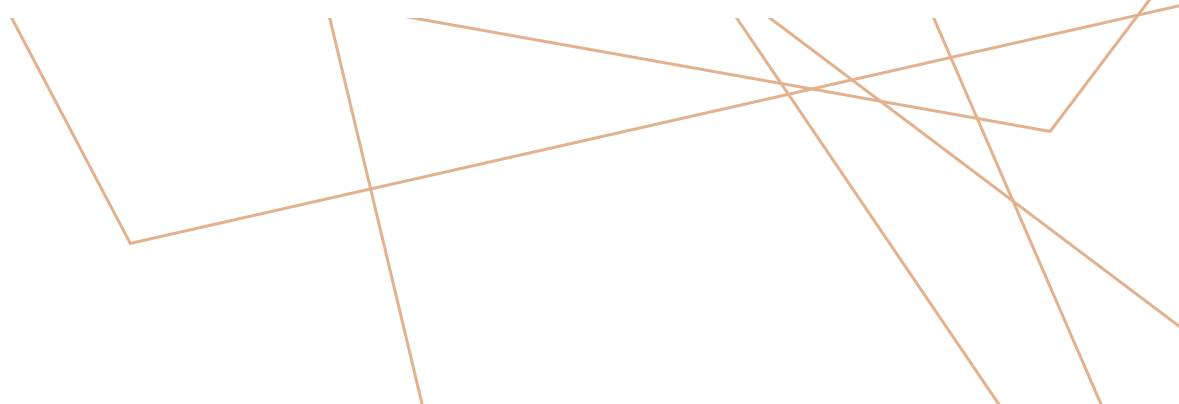


Flask

PICKLE

An abstract geometric pattern composed of several overlapping triangles. The triangles are rendered in two colors: a light, muted blue and a light, muted orange. They are set against a plain white background. The lines forming the triangles are thin and intersect to create a complex, layered composition. Some triangles are partially obscured by others, creating a sense of depth and movement. The overall effect is minimalist and modern.

Turn categorical features into dummies



DATA PREPARATION AND FEATURES ENGINEERING



We looked at the correlation values between columns to find features that we can engineer to raise the score of the models:

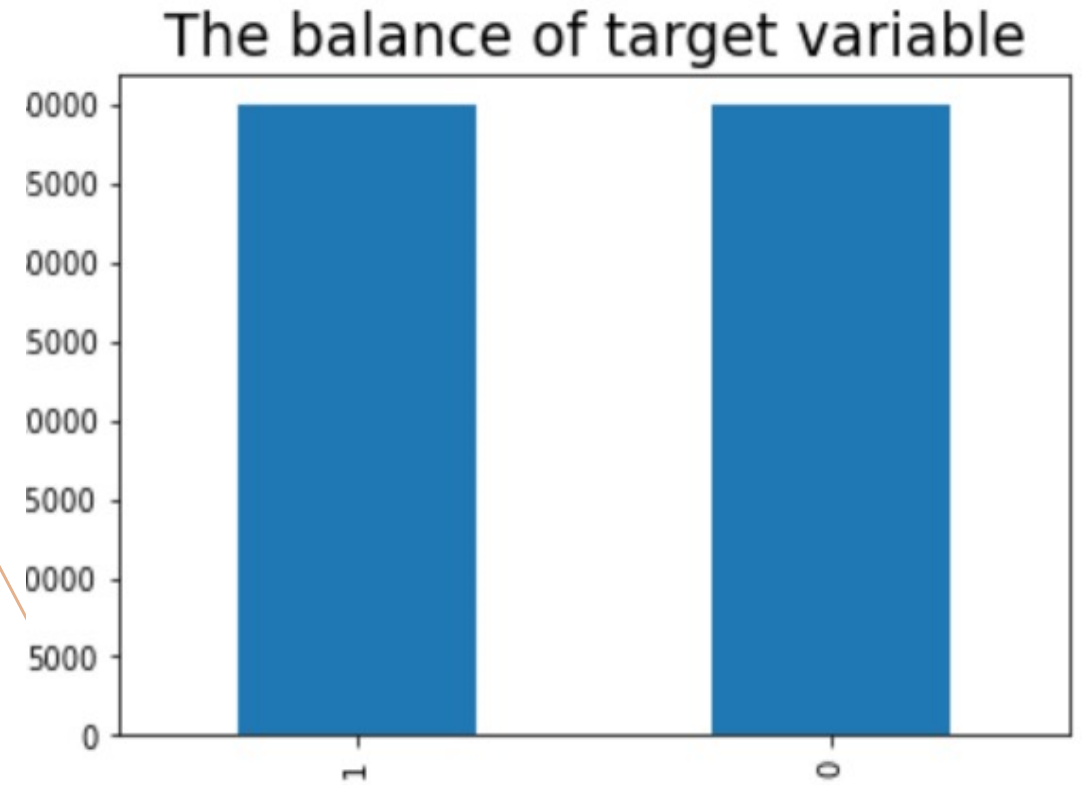
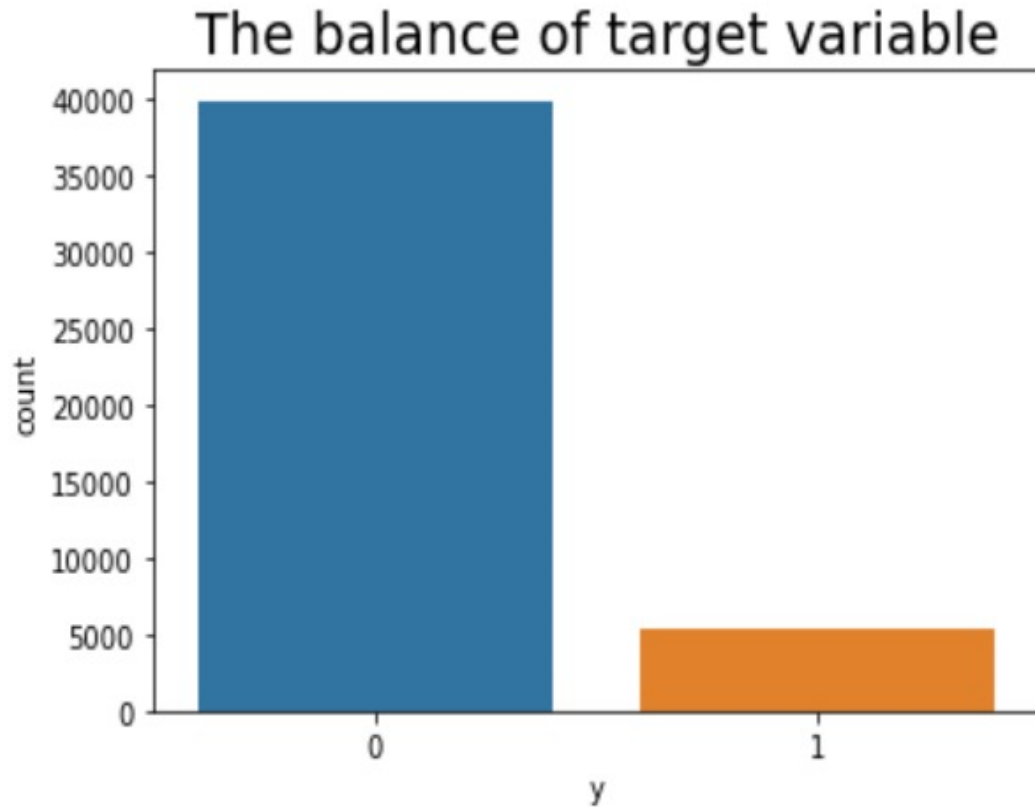


We added a new column to the data to improve the scores

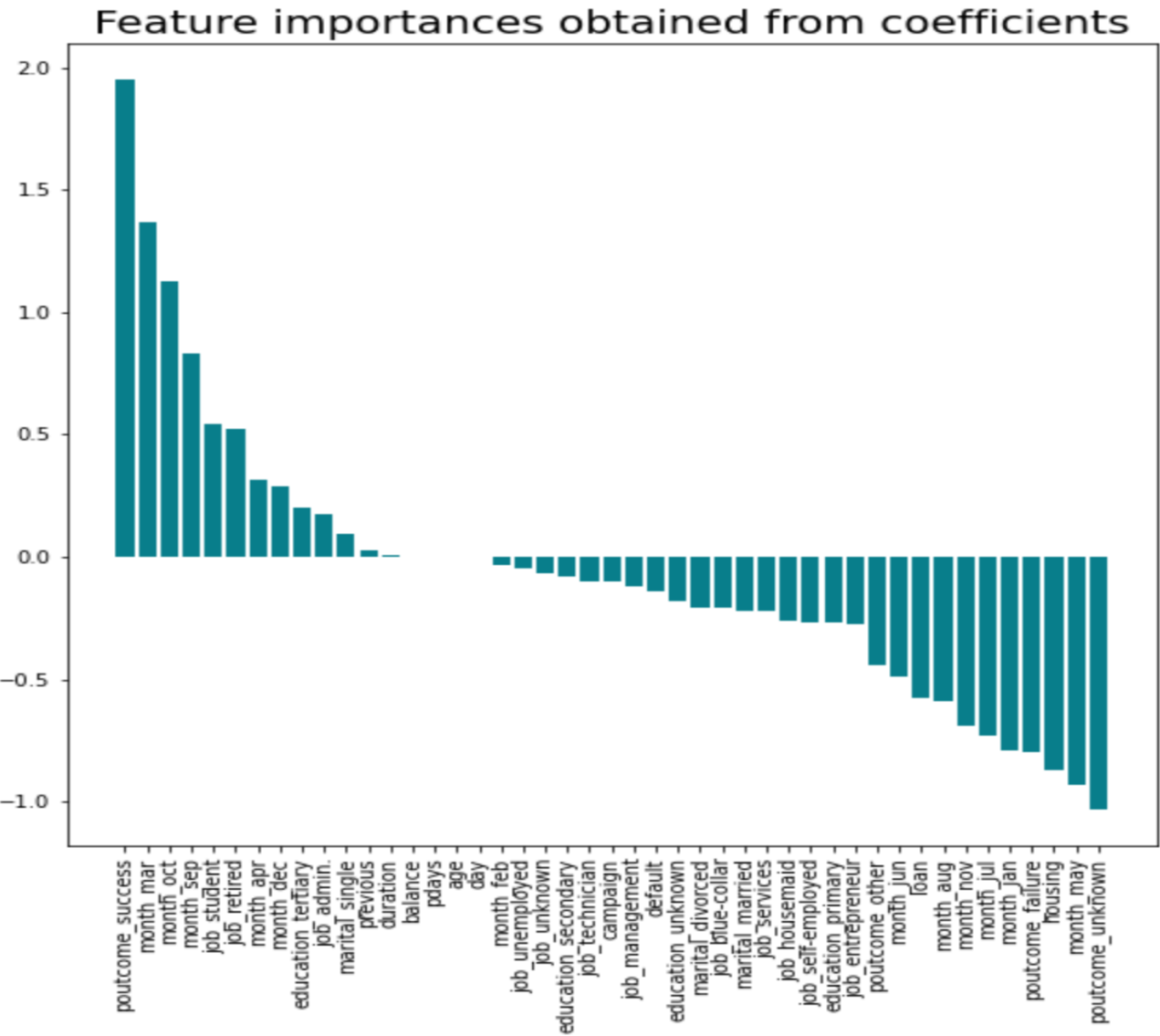


We did the feature importance for most of the models to improve the scores

SOLVING IMBALANCED DATA



FEATURES ENGINEERING



ACCURACY OF THE MODELS

model	training	Validation	test
LogisticRegression	0.825	0.835	0.865
KNeighborsClassifier	0.944	0.716	
DecisionTreeClassifier	1.0	0.766	
RandomForestClassifier	1.0	0.795	
XGBClassifier	0.983	0.829	
BaggingClassifier	0.999	0.801	
NaiveBayes	0.778	0.783	

MODELS TRAINING AND EVALUATION

Out[96]:

	Model	Accuracy	F1	Precision	Recall	ROC Score
0	LogisticRegression	0.8658	0.5247	0.4580	0.6141	0.7572
1	KNeighborsClassifier	0.7763	0.3531	0.2711	0.5060	0.6597
2	RandomForestClassifier	0.8983	0.5267	0.6002	0.4693	0.7132
3	DecisionTreeClassifier	0.8679	0.4392	0.4500	0.4290	0.6785
4	XGBClassifier	0.8853	0.5909	0.5187	0.6865	0.7996
5	BaggingClassifier	0.8944	0.5497	0.5660	0.5344	0.7391

LOGISTIC REGRESSION CLASSIFIER

Result:

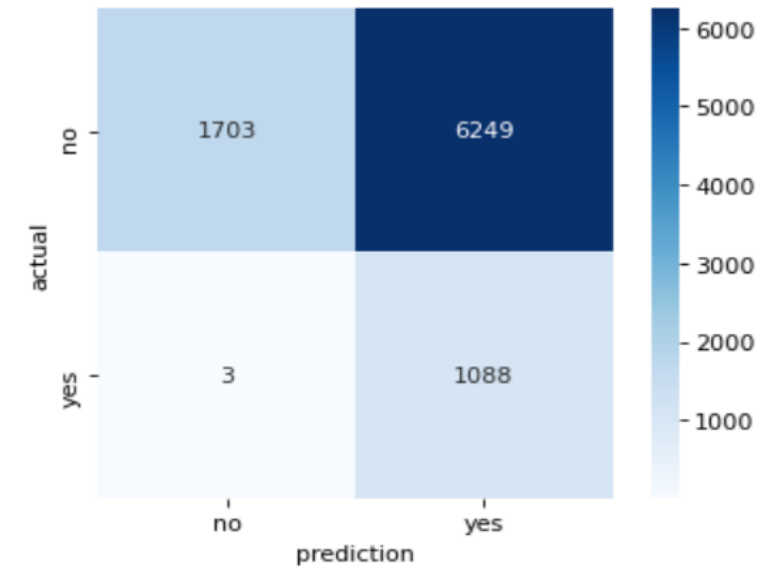
Training accuracy: 82%

Validation accuracy: 83%

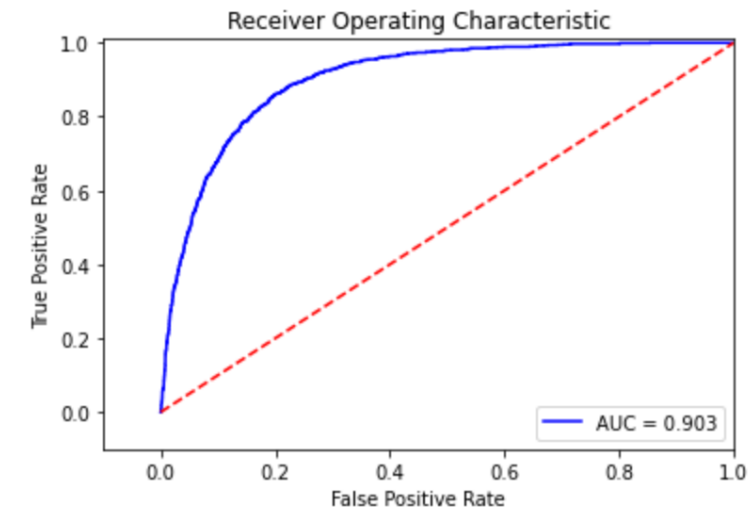
Testing accuracy: 86%

F1 Score: 71%

ROC-AUC score: 90%



ROC AUC score = 0.9030663126329008



BANK MARKETING

[HOME](#)

BANK MARKETING

START NOW

A series of thin, light brown lines on the left side of the slide, forming an abstract geometric pattern of overlapping polygons and intersecting lines.

THANKS FOR LISTENING