

Bank Marketing Project

Walid Mohamed Tawfik, Adham Moataz Mohamed, Malek fouda, Abdullah Eid

June 6, 2022

1 Abstract

The main target of this project is to improve the marketing campaign of a bank by implementing machine learning models that will be able to accurately and successfully predict whether the customers would deposit their money in the bank or not. We have successfully achieved 85% accuracy using the Support Vector Machine model.

2 Introduction

In this project we are targeting to improve the marketing campaign of a bank that find it hard to attract new customers to deposit their money in it. To achieve this we are going to implement different machine learning models which are (Logistic Regression, Neural Network, SVM, KNN, Decision Tree) and we are going to train these models on the dataset collected from the previous campaigns so that we could improve the next marketing campaign by successfully predicting whether or not a certain customer would deposit his money in the bank based on certain features.

3 Dataset

Our dataset consists of 11162 rows and 17 columns, 16 of them represents the features in which 7 of them are Integers (age, balance, day, duration, campaign, pdays, previous) and 6 of them are of type string (job, marital, education, contact, month, poutcome) and 3 of them are of type boolean (default, housing, loan) and the last column is the (deposit) column which is the label and it is of boolean type.

The following image shows an example of few records from the dataset:

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	outcome	deposit
0	59	admin.	married	secondary	no	2343	yes	no	unknown	5	may	1042	1	-1	0	unknown	yes
1	56	admin.	married	secondary	no	45	no	no	unknown	5	may	1467	1	-1	0	unknown	yes
2	41	technician	married	secondary	no	1270	yes	no	unknown	5	may	1389	1	-1	0	unknown	yes

Figure 1: dataset example.

There are various preprocessing steps that were needed which are:

1. Encoding The Categorical Data:

We have encoded the categorical columns of the features using One Hot Encoder and encoded the label column using Label Encoder.

2. Feature Scaling:

We have performed feature scaling on the numerical columns using the Standardisation method.

3. Splitting The Dataset Into Training Set and Testing Set:

We have split the dataset into 80% training set and 20% testing set.

4 Experimental Results

Table 1: Performance Results

Model	Accuracy	Precision	Recall	F1 Score
SVM	0.85	0.82	0.88	0.85
Artificial Neural Network	0.82	0.82	0.79	0.80
Logistic Regression	0.82	0.83	0.79	0.81
KNN	0.82	0.83	0.79	0.81
Decision Tree	0.82	0.79	0.84	0.81

As the above table1 shows, the Support Vector Machine classifier (SVM) has achieved the highest accuracy of 85% while the accuracy of all the other classifiers was around 82% and all the classifiers had near performance in terms of Precision, Recall and F1 Score.