Sri Lanka Institute of Information Technology

Bank Term Deposit Customer Prediction

Statement of Work Document

FDM Mini Project 2020

Submitted by:

1. IT18225570 – (A. M. Azhar)
2. IT18223422 – (M. G. S. M. Maweekumbura)
3. IT18210620 – (R. M. W. Gunathilake)
4. IT18197006 – (Maddumaarachchi M.A.T.B)
5. IT18215502 – (K. Priyanka)
6. IT18217896 – (M. Mayuriya)
7. IT18227550 – (R. A. K. Wandana)

24/09/2020

Table of Contents

Background

The mini project requires to find out a real-world problem and propose a solution for it using Data Mining and Machine Learning techniques. After discussing with the group and by searching on the web as well, the team came across a problem that may be an issue in many banks and financial institutions.

The real world scenario that we identified is that when a bank needs to promote its bank term deposit accounts (also referred as fixed deposit) with customers who have normal savings accounts, they do not know whom to approach exactly. They will waste a lot of time as well as resources by contacting all customers and that would be very futile.

By analysing historical data of customers, we can mine these data and identify patterns that will finally conclude if a customer will create a fixed deposit account or not.

This will help in finding out what type of customers a bank needs to approach, so that they can promote their Bank Term deposit accounts.

The dataset that is being used is the Bank Marketing dataset, found on the UCI Machine Learning Repository. The data is related with direct marketing campaigns of a Portuguese banking institution.

The dataset will be used to create a model that can predict if the customer will create a term deposit account or not, by splitting the dataset to training and testing datasets, and doing the necessary analysis.

Scope of work

In our dataset we have data of 45211 current customers in the bank. There are 17 features that have been collected as below,
1.   age (numeric)
2. Job (categorical:'admin.','bluecollar','entrepreneur','housemaid','management','retired','self-employed','services','student','technician','unemployed','unknown')
3.   marital (categorical: 'divorced','married','single','unknown')
4.   education(categorical: 'basic.4y','basic.6y','basic.9y','high.school','illiterate','professional.course','university.degree','unknown')
5.   default: has credit in default? (categorical: 'no','yes','unknown')
6.   balance (numerical)

7. housing: has housing loan? (categorical: 'no','yes','unknown')
8. loan: has personal loan? (categorical: 'no','yes','unknown')
9. contact: contact communication type (categorical: 'cellular','telephone')
10. day: (numerical)
11. month: last contact month of year (categorical: 'jan', 'feb', 'mar', ..., 'nov', 'dec')
12. duration: last contact duration, in seconds (numeric)
13. campaign: number of contacts performed during this campaign and for this client (numeric, includes last contact)
14. - pdays: number of days that passed by after the client was last contacted from a previous campaign (numeric)
15. previous: number of contacts performed before this campaign and for this client (numeric)
16. poutcome: outcome of the previous marketing campaign (categorical: 'failure','nonexistent','success')
Desired Target Feature
17. y - has the client subscribed a bank term deposit? (binary: 'yes','no')

We are going to analyze these data to predict if the customer will subscribe to a Bank Term deposit or not.
When roughly considering the above features, we hope to remove some columns like day, month, contact and campaign since they may not affect the prediction. This will be correctly concluded by doing the necessary data analysis by the Data Scientists of the project. We will focus on removing rows which contain null values in the data preprocessing level.
The dataset will be split into training dataset and testing dataset. Classification algorithms will be used to and train the dataset to build the model. The testing dataset will be used to validate the model and to check the accuracy of the predictions. Using the model, we will predict if a new customer will subscribe to a Bank Term deposit or not.

Activities

We started off by identifying a problem that is faced by many banking institutions and then downloaded a Bank Marketing Dataset in UC Irvin Machine Learning Repository.
Preparation of Statement of work.
Data preprocessing is performed to transform the raw data in a useful and efficient format. We will clean the dataset by removing some unnecessary columns that are not useful for us to come up with solutions (date, month, campaign and contact).
Defining the entities and using models such as the Classification model and Decision tree model to train the datasets.
Designing a user friendly and most effective front-end and back-end and connecting them so that by using this, any bank institution can determine if a customer will obtain a fixed deposit or not.
Interpretation and evaluation of the outputs will be displayed in a form that can be understood by anyone easily, by graphs and charts. Identification of hidden data patterns can be done here.

Approach

In this mini project we hope to use some machine learning techniques and algorithms for the analysis of bank marketing data to produce predictions of if a customer will create a Bank Term deposit account or not. Accordingly, we will make use of data mining methods, data preprocessing methods and specifically classification methods to get highly accurate predictions.

We desire to utilize Angular to develop the front-end of the project and Python for Data analysis and backend development.
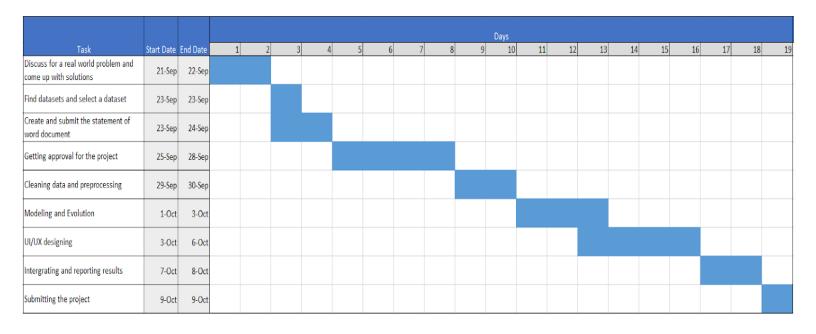
This prediction will help as a solution in finding out what type of customers a bank needs to approach, so that they can promote their Term Deposit accounts.

Deliverables

The outcome of this project will predict whether the client will subscribe to a bank term deposit or not.
This will help for the banks to promote their term deposits for the specifically identified customers who are more likely to open new term deposits.

Project Plan & Timeline

Task	Start Date	End Date	Days																		
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Discuss for a real world problem and come up with solutions	21-Sep	22-Sep	▓	▓																	
Find datasets and select a dataset	23-Sep	23-Sep			▓																
Create and submit the statement of word document	23-Sep	24-Sep			▓	▓															
Getting approval for the project	25-Sep	28-Sep					▓	▓	▓	▓											
Cleaning data and preprocessing	29-Sep	30-Sep									▓	▓									
Modeling and Evolution	1-Oct	3-Oct											▓	▓	▓						
UI/UX designing	3-Oct	6-Oct													▓	▓	▓				
Intergrating and reporting results	7-Oct	8-Oct																	▓	▓	
Submitting the project	9-Oct	9-Oct																			▓

Assumptions

We assume that all the data provided in the dataset are accurate.
We assume that the columns that were removed do not affect the prediction of the final output.
We assume that all the data are taken from current customers in the bank.

Project team, roles and responsibilities

Role	Team members	Responsibilities
Data Scientist	A.M. Azhar (IT18225570) Maweekumbura M.G.S.M. (IT18223422) R.A.K.Wandana (IT18227550)	Mine and analyze data from company databases. Develop custom data models and algorithms to apply to data sets. Improvement of product development, marketing techniques and business strategies. Use predictive modeling to increase and optimize customer experiences and other business outcomes. Coordinate with different functional teams to implement models and monitor outcomes.

Data Engineer	K.Priyanka (IT18215502)	Test the reliability and performance of each part of a system
	R. M. W. Gunathilake (IT18210620)	Designing the architecture of a data platform.
	M.Mayuriya (IT18217896)	Managing the data stored and structuring it properly via database management systems
		Manage meta-data, Set standards for data transformation/processing and Develop machine learning models.
UX/UI engineer	Maddumaarchchi M.A.T.B (IT18197006)	Gathering and evaluating user requirements in collaboration with Data engineer and Data Scientist.
	M.Mayuriya (IT18217896)	Designing graphic user interface elements, like menus, tabs and widgets.
	A.M. Azhar (IT18225570)	Design websites and applications with the users' trend
Analysing and Reporting	All members	Analyse the final results, create graphs to visualise the trends and identify patterns.
		Create final report on all trends identified and report on the accuracy of the model.