BANK CUSTOMER TERM DEPOSIT PREDICTION

Predicting Bank Customer Behavior When Marketed for Term Deposit Subscription by Bank Officers

KAIZEN TEAM

Team Members

Names
Stanley Zebulon
Sri Sai Venkata Kameshwar Rao Dhulipala

August 11, 2025

Data Mining Project Overview

The goal of the project is to identify and define a problem the team want to solve using the data mining techniques. For example, using the data related to understand the trends that which types of cars can be grouped together to categorize the car for customers. This will be helpful for the dealers to meet customer satisfaction. Also, the used cars data can be analyzed to identify the most significant factors in predicting the price (numerical target) of the car or predicting whether the car is expensive or inexpensive (categorical target).

For a project, teams will identify the problem, obtain the data, and define the scope of the problem. Team needs to identify a data set with at least 8000 observations or records. There should be at least two problem questions. One of the questions needs to be addressed using predictive analysis (using supervised techniques) and the other using exploratory analysis (unsupervised techniques). The team will use visualization for data exploration and preparation. Then, the team will implement several supervised and unsupervised data mining techniques to solve the identified problem. Appropriate performance evaluation metrics will be used to evaluate the performance of these supervised and unsupervised models to establish the significance of the developed models with highest accuracy. Finally, scoring the predictive models (supervised models) on new data will be performed. Scoring plays a crucial role in data mining as it allows in confirming the reliability of the developed model by deploying the model on new data.

To assist teams in solving a feasible and interesting problem using several data mining techniques, each team should obtain the instructor's approval regarding the scope and nature of the problem, the data set, and the nature of the analysis. *Teams will use Decision Trees, Regression, and Neural Network supervised techniques for comparing and identifying the best predictive model (can be adapted from predictive analysis assignment)* The *unsupervised techniques of Cluster Analysis will be used by team for exploratory analysis (can be adapted from exploratory analysis assignment)*. Overall, integrating the learning outcomes from individual course Assignments (*data exploration and visualization, exploratory analysis, and predictive analysis*) will be helpful to implement the project. **The team will use SAS Enterprise Miner to implement the project (exploratory and predictive analysis). Tableau or Excel can be used for data preparation and visualization.**

Project Proposal

1. **What is (are) the problem(s) for investigation? Briefly discuss the problem you want to investigate. Additionally, discuss why this problem is important for a data mining project.**

The purpose of our project is to utilize decision tree methodologies to analyze and predict bank customer behavior. Specifically, the project seeks to develop a classification model to predict whether customers will subscribe to a term deposit and a regression model to estimate the duration of time spent with the customer by the bank official. Using the SAS Miner and our multivariate dataset, we would derive insights to assist the bank officials make informed decisions regarding the customer to target and the amount of time to be spent with each customer.

2. **Develop tentative question(s) you want to address in your project** [*Be very specific and **write the questions**. There should be at least two questions, one for predictive analysis and the other for exploratory analysis. These questions will evolve and can change while working on the project.*]

For this purpose, we have the following research questions:

RQ1. What are the significant/important factors that can be used to determine if a customer will subscribe to the term deposit of the ban or not?

RQ2. What are the significant/important factors that can be used to predict the optimal time a bank official should spend with a customer to make them subscribe to the term deposit?

3. **Information about the dataset. [***Team can obtain the secondary data set of their choice from various web sources. Kindly refer to Appendix for some recommended sources. If team member has any data set, they want to use, please feel free to use the data set with appropriate approval and usage protocol.***]**

3.1 What is the source of the dataset? [*Be specific. Provide the URL if applicable.*]

Our dataset source is http://archive.ics.uci.edu/ml/datasets/Bank+Marketing

3.2 What is the dataset about? Explain briefly. Conduct some literature search to identify where and how the data set you have selected for your project has been used. This will help you to identify the gap and address the problem in a new and/or different way [*Not more than two paragraphs. Include references and in-text citations using APA. The data and variable description are usually available at the data source in the form of data dictionary and/or codebook.]*

This dataset is based on "Bank Marketing" UCI dataset. The data is related with direct marketing campaigns of a Portuguese banking institution. The marketing campaigns were based on phone calls. Often, more than one contact to the same customer was required, to access if the product (bank term deposit) would be ('yes') or not ('no') subscribed (Moro et al., 2012).

Reference

Moro,S., Rita,P., and Cortez,P.. (2012). Bank Marketing. UCI Machine Learning Repository. https://doi.org/10.24432/C5K306.

3.3 Which variable in your dataset will be your target variable for predictive analysis? Provide a brief description (at least one line) of your target variable. [*Be specific. You can only have ONE target variable. The target variable is what you are going to predict in your analyses. If you are going to do unsupervised analysis, which does not have a target variable, please say so*].

Target Variable Name	**Variable Data Type** *[e.g. Numeric, Nominal, Ordinal, Binary]*	**Description** *[can be obtained from data dictionary and/or codebook]*
Y (yes or no)	Binary	Output variable

3.4 Which variables in your dataset will be your predictor variables for predictive and exploratory analysis? Provide brief description (at least one line) of each predictor variable. [*Be specific. The predictor variables are the variables that will help you predict the target variable. These predictor variables might evolve over time but based on your preliminary analysis of data chose some predictor variables. The same or different predictors/input variables can be used to perform the exploratory analysis*].

Input/Predictor Variable(s) Name	**Used for Exploratory Analysis or Predictive Analysis or Both**	**Variable Data Type** *[e.g. Numeric, Nominal, Ordinal, Binary]*	**Description** *[can be obtained from data dictionary and/or codebook]*
Job	Both	Categorical	Type of Job
Education	Both	Categorical	"Basic.4y", "basic.6y", "basic.9y", "high. school", "illiterate", "professional. course", "university. degree", "unknown"
Duration	Both	Numeric	last contact duration, in seconds
Campaign	Both	Numeric	number of contacts performed during this campaign and for this client
Age	Both	Numeric	Customer Age
Marital status	Both	Categorical	"divorced","married","single","unknown"; note: "divorced" means divorced or widowed
Default	Both	Categorical	has credit in default? "no","yes","unknown"
Housing Loan	Both	Categorical	has housing loan? "no","yes","unknown"
Personal Loan	Both	Categorical	has personal loan? "no","yes","unknown"
Contact	Both	Categorical	contact communication type "cellular","telephone"

Month	Both	Categorical	last contact month of the year: "Jan", "Feb", "Mar", ..., "Nov", "Dec")
Day_of_week	Both	Categorical	last contact day of the week: "mon","tue","wed","thur","fri")
pdays	Both	Numeric	number of days that passed by after the client was last contacted from a previous campaign; 999 means client was not previously contacted
previous	Both	Numeric	number of contacts performed before this campaign and for this client
poutcome	Both	Categorical	outcome of the previous marketing campaign: "failure","nonexistent","success"
emp.var.rate	Both	Numeric	employment variation rate - quarterly indicator
cons.price.idx	Both	Numeric	consumer price index - monthly indicator
cons.conf.idx	Both	Numeric	consumer confidence index - monthly indicator
euribor3m	Both	Numeric	euribor 3 month rate - daily indicator
nr.employed	Both	Numeric	number of employees - quarterly indicator
Y	Outcome	Categorical	has the client subscribed a term deposit? (binary: "yes","no")

Note: Add more rows to the table as needed

4. Provide the timetable (Project Plan) for your project.

SCHEMATIC OF THE DATA MODELING PROCESS

Milestone	Activities	Timeline
Project Proposal	Definition of Purpose, determination of research questions, dataset definition, and scope of the project.	June 29 - July 9, 2024
Data Collection	Obtain relevant data sources.	June 29 - July 9, 2024
Data Cleaning and Preparation	Data exploration and transformation.	July 10 - July 13, 2024
Determination of Data Mining Task	Establish mining objectives and select relevant variables and techniques.	July 14 - July 17, 2024
Choosing & Implementing Data Mining Methods	Implement various data mining techniques (supervised and unsupervised).	July 14 - July 17, 2024
Model Selection	Compare the performance of different models and select the best-performing one.	July 14 - July 17, 2024
Performance Evaluation	Evaluate model performance using appropriate metrics.	July 14 - July 17, 2024
Model Deployment	Deploy with the test dataset	July 18 - July 20, 2024
Project Report	Finalize project report and submit.	21-Jul

References

Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., and Lichtendahl Jr, K. C. 2018. *Data Mining for Business Analytics: Concepts, Techniques, and Applications in R*. John Wiley & Sons.

Appendix: Recommended Data Sources

Potential data sources can include, but are not limited to the Internet (e.g., search for "datasets for data mining"), current or past employers (with permission), public databases, university datasets, datasets that you collected in the past. Some sources of publicly accessible datasets are as follows. You might need to create a profile on some of these websites to access the content. Using university email ids to create the profile might provide additional benefits.

- https://www.kaggle.com/datasets
- https://www.datasciencecentral.com/profiles/blogs/big-data-sets-available-for-free
- http://www.kdnuggets.com/datasets/index.html
- https://www.kdnuggets.com/datasets/government-local-public.html
- https://aws.amazon.com/datasets?_encoding=UTF8&jiveRedirect=1
- Google Dataset Search
- https://www.cdc.gov/data.html
- https://data.world/datasets/open-data
- https://www.nhtsa.gov/research-data/crash-injury-research