WEST TEXAS A&M UNIVERSITY

DATA MINING FINAL PROJECT:

PREDICTING CUSTOMER ACCEPTANCE OF NEW TERM DEPOSIT THROUGH TELE-MARKETING

Team 2 Data Miners:

Chidinma Aniekwena

Chukwuebuka Dibie

Fougnigue Sefon

Mariam Adegbindin

Tokoni Forun

I/ EXECUTIVE SUMMARY

II/ INTRODUCTION

From the early days of the banking sector, effective communication with potential customers has been crucial, with telemarketing emerging as a vital strategy for promoting banking products such as term deposits. Telemarketing, a form of direct marketing, involves contacting prospective clients via telephone to offer services or products, capitalizing on the personal touch of voice communication to enhance customer engagement and response rates. For banks, this technique has become increasingly important, serving not just to boost sales but also to establish and maintain customer relationships.

With telemarketing, banks and financial institutions are tasked with the dual challenge of effectively engaging potential customers while understanding and adapting to their preferences and behaviors. In this competitive environment, accurately predicting customer behaviors emerges as a crucial opportunity. To seize this, we plan to employ data mining techniques aimed at predicting customer responses to new term deposit offers via telemarketing. This approach is part of our broader strategy to leverage data mining to drive business intelligence. We aim to unveil the underlying patterns that influence acceptance rates. This project not only helps in refining marketing approaches but also aligns with the broader goal of Data Miners: to discover meaningful patterns and rules and empower organizations with the knowledge to make informed decisions.

Our dataset was retrieved from the UC Irvine Machine Learning repository (https://archive.ics.uci.edu/dataset/222/bank+marketing). It contains data on customer responses to new term deposits via phone calls, along with comprehensive demographic and behavioral attributes. It provides a rich foundation for analyzing customer behavior towards bank term deposits.

To achieve our objective, we will undertake a data analytics process, examining the characteristics of customers who have either accepted or declined the telemarketing offer. Our initial step involves acquiring the dataset from the UC Irvine Machine Learning Repository and preparing it to ensure cleanliness and usability. We will scrutinize the dataset for highly correlated attributes, eliminating those deemed redundant, and conduct a thorough review of summary statistics to verify the absence of missing data and address any outliers.

Our data will then be segmented using the holdout method, allocating 70% for training and the remainder for testing the models. We will employ various modeling techniques, including Decision Trees, Naïve Bayes, Logistic Regression, and Neural Networks, to construct our predictive framework. After developing these models, we will assess their performance using critical metrics such as accuracy, recall, and precision. This comprehensive evaluation will enable us to discern the strengths and weaknesses of each model in accurately forecasting customer responses to telemarketing campaigns.

III/ DATA DESCRIPTION

The Bank Marketing dataset, created by Paulo Cortez (University of Minho) and Sérgio Moro (ISCTE-IUL) in 2012, forms the foundation of our analysis. Comprising data from direct marketing campaigns via phone calls of a Portuguese banking institution from May 2008 to November 2010, the dataset's primary aim is to forecast whether clients will subscribe to a term deposit. Containing 45,211 records, the dataset encompasses 16 input attributes plus the target output attribute. These attributes include bank client data such as age, job type, marital status, education level, credit default, average yearly balance, housing loan, and personal loan status. Additionally, it details the last contact of the current campaign, including contact type, day and month of last contact, and duration in seconds. Other attributes cover the campaign's contact counts, days since the last contact from a previous campaign, previous contacts count, and the outcome of the previous marketing campaign. The output variable 'y' is our dependent variable and indicates whether the client subscribed to a term deposit.

Attributes	**Defintiion**	**Type**	**Values**
Age	Client's age	numeric	"18","19",...,"95"
Job	type of job	categorical	"admin.","unknown","unemployed","management"," housemaid", "entrepreneur","student","blue-collar","self-employed","retired", "technician","services"
Marital	marital status	categorical	"married","divorced","single"
Education	Level of education	categorical	"unknown","secondary","primary","tertiary"
Default	has credit in default?	binary	"yes","no"
Balance	average yearly balance	numeric	"-8019","-8018",..., "102127"

Housing	has housing loan?	binary	"yes","no"
Loans	has personal loan?	binary	"yes","no"
Contact	contact communication type	categorical	"telephone","cellular"
Day	last contact day of the month	numeric	"1","2",..., "31"
Month	last contact month of year	categorical	"jan", "feb", "mar", ..., "nov", "dec"
Duration	last contact duration	numeric	"0",...,"4918"
Campaign	number of contacts performed during this campaign and for this client	numeric	"1","2",...,"63"
pDays	number of days that passed by after the client was last contacted from a previous campaign	numeric	"-1","0",...,"871"
previous	number of contacts performed before this campaign and for this client	numeric	"0","1",...,"275"
pOutcome	outcome of the previous marketing campaign	categorical	"other","failure","success"

Y	has the client subscribed a term deposit?	binary	"yes","no"

IV/ DATA PREPARATION

V/ MODELING

VI/ EVALUATION

VII/ DISCUSSION

VIII/ REFERENCES

VIII/ APPENDIX