



Final Project: Dataset Proposal

ALY6040

Date: April 17, 2024

Group 5

Team Members: Nidhi Jeetesh Patel

Yukang Lin

Ramish Fatima

Introduction:

The purpose of this project is to leverage predictive analysis techniques to forecast customer subscriptions to term deposits based on historical data from direct marketing campaigns conducted by a Portuguese banking institution.

Business Rationale:

The chosen dataset focuses on direct marketing campaigns for bank term deposits conducted by a Portuguese banking institution. The business rationale for selecting this dataset lies in the importance of optimizing marketing campaigns to improve subscription rates and maximize the return on investment. By analyzing customer attributes and campaign outcomes, businesses can gain insights into the factors influencing subscription decisions and develop targeted strategies to enhance customer engagement and drive conversions. Moreover, with the increasing competition in the banking industry, understanding customer preferences and behavior is essential for maintaining competitiveness and achieving sustainable growth.

Business Quadrant:

In the APLC quadrants, this dataset primarily aligns with the "Product" quadrant, emphasizing the importance of understanding customer behavior and optimizing marketing campaigns to achieve business objectives. By leveraging machine learning techniques on this dataset, businesses can develop predictive models to identify potential subscribers more accurately, thereby improving campaign effectiveness and reducing marketing costs. Additionally, insights derived from the dataset can inform strategic decision-making processes, such as resource allocation and product development, leading to enhanced customer satisfaction and loyalty.

Data Engineering Quadrant:

From a data engineering perspective, this dataset presents opportunities for data collection, preprocessing, and model training to extract actionable insights. With features such as customer demographics, campaign attributes, and previous contact outcomes, effective data management

and feature engineering are crucial for building robust predictive models. Data engineers need to implement pipelines for data cleaning, transformation, and integration to ensure the quality and consistency of the dataset. Additionally, model training and evaluation require scalable infrastructure and algorithms to handle the large volume of data and optimize predictive performance. By addressing these data engineering challenges, organizations can unlock the full potential of the dataset and drive meaningful outcomes for marketing optimization and customer relationship management.

Description of Proposed Dataset:

- Number of Records: 45211 (Train dataset)
- Number of Variables: 11
- Variables of Interest:
 - Age
 - Job
 - Marital Status
 - Education
 - Default
 - Balance
 - Housing Loan
 - Contact Communication Type
 - Day
 - Duration
 - Campaign Contacts Count
 - pdays (number of days passed since previously contacted from the previous campaign)
 - poutcome (outcome from the previous marketing campaign)

Initial Interest and Relevance:

The initial interest in this dataset stems from the opportunity to leverage customer data and campaign outcomes to optimize marketing strategies and improve subscription rates for bank term deposits. By analyzing historical campaign data and customer attributes, businesses can identify patterns and trends that influence subscription decisions, allowing for targeted outreach and personalized messaging. Furthermore, the inclusion of test data enables the evaluation of predictive models, facilitating continuous improvement and refinement of marketing tactics. Overall, this dataset offers valuable insights into customer behavior and campaign

effectiveness, empowering businesses to make data-driven decisions and achieve better outcomes in their marketing initiatives.