Final Report: Predicting Term Deposit Subscriptions for a Portuguese Bank

Name: Data Science : Bank Marketing (Campaign)
Report date: Feb 28 2024
Internship Batch: LISUM28
Version:1.0
Data intake by:Minseok Kim
Data intake reviewer:Minseok Kim
Data storage location: https://github.com/N0VA-code/Data-Glacier-Data-Based-Consulting-Project.git

Executive Summary

A Portuguese banking institution is preparing to launch a term deposit product. The objective is to develop a predictive model to identify potential customers who are likely to purchase this product. By leveraging machine learning, the bank aims to optimize their marketing strategies, focusing resources on customers with a higher propensity to subscribe, thus saving on time and cost.

Problem Statement

The bank requires an understanding of the likelihood of a customer purchasing their term deposit product based on past interactions with the bank or other financial institutions. This predictive insight will enhance the efficiency of marketing channels like telemarketing and email campaigns.

Importance of ML Model

The utilization of a machine learning model will enable the bank to shortlist customers with higher probabilities of purchasing the product. This targeted approach is expected to conserve resources and reduce costs associated with marketing efforts.

Data Set Overview

The dataset contains information from direct marketing campaigns conducted through phone calls. Success of these campaigns was measured by whether the client subscribed to a term deposit, which is the target variable 'y' for prediction.

Business Understanding

The success of the term deposit marketing campaign directly impacts the bank's revenue and growth. Accurate prediction models will aid in enhancing the conversion rate, resulting in a more efficient allocation of the marketing budget.

Data Understanding and Preparation

The dataset underwent thorough cleaning, including the removal of outliers using z-score normalization. This approach was adopted to maintain data integrity and improve the model's predictive accuracy.

Data Intake Report

Tabular data details:

Total number of observations	42780
Total number of files	1
Total number of features	19
Base format of the file	.csv
Size of the data	4.6MB

Note: Replicate same table with file name if you have more than one file.

Exploratory Data Analysis

An in-depth analysis revealed key insights into customer behavior and significant predictors of term deposit subscriptions. This guided the subsequent feature selection and model-building phase.

Model Building and Selection

Various models, including Logistic Regression, Random Forest, and Gradient Boosting, were evaluated. Gradient Boosting emerged as the superior model due to its highest accuracy.

Handling Imbalanced Data

Techniques to address data imbalance were implemented, enhancing the model's performance by ensuring a more representative learning process. Methods that suggested discarding available data were avoided.

Performance Reporting

The Gradient Boosting model demonstrated the highest accuracy among the models tested. The decision to prioritize accuracy as the primary metric was driven by the business goal of maximizing correct predictions of subscription behavior.

Model Deployment

The selected Gradient Boosting model was deployed into a production environment, enabling real-time predictions to support the bank's marketing campaigns.

Business Metric Translation and Communication

The model's machine learning metrics were translated into business metrics, providing actionable insights. These results were communicated to stakeholders, focusing on how the model's predictions could enhance marketing efforts and return on investment.

Presentation for Non-Technical Audience

A comprehensive presentation was prepared, elucidating the model's findings in an accessible manner for non-technical stakeholders, ensuring a clear understanding of the model's business

value.

Conclusion

The development and deployment of the Gradient Boosting model mark a significant step forward for the bank's marketing strategy. The model's predictions will play a crucial role in refining customer targeting, ultimately saving resources and increasing the success rate of the term deposit product launch.