

Predictive Modeling for Bank Marketing Campaigns: A Comparative Analysis

Sangram Jagtap

September 30, 2023

Abstract

This paper presents a comparative analysis of multiple machine learning models applied to a bank marketing dataset. The objective is to identify predictive factors influencing customers' decisions to subscribe to a term deposit. The models' performance is evaluated using various metrics, and insights are derived to aid future marketing strategies.

1 Introduction

Bank marketing campaigns aim to promote banking products and services, with a focus on maximizing subscriptions to term deposits. Understanding customer behavior and identifying factors influencing their decisions are crucial for optimizing campaign effectiveness. This study employs machine learning models to analyze a bank marketing dataset and predict the likelihood of customers subscribing to a term deposit.

2 Methodology

2.1 Dataset

The dataset comprises features such as age, job type, marital status, education level, and previous campaign outcomes. The target variable is binary, indicating whether a customer subscribed to a term deposit.

2.2 SEMMA Methodology

The SEMMA methodology was adopted, encompassing Sample, Explore, Modify, Model, and Assess phases, to guide the data science process.

2.3 Models Used

- Logistic Regression
- Decision Tree Classifier

- Random Forest Classifier
- Support Vector Machine (SVM)
- Gradient Boosting Classifier

3 Results

3.1 Model Performance

The models were evaluated based on accuracy, precision, recall, and F1-score. The Random Forest model demonstrated superior performance among the models tested.

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	78.95%	-	-	-
Decision Tree	76.40%	-	-	-
Random Forest	83.52%	81%	86%	83%
SVM	81.15%	-	-	-
Gradient Boosting	82.18%	-	-	-

Table 1: Model Performance Comparison

3.2 Feature Importance

Feature importance analysis revealed significant predictors such as education level, job type, and previous campaign outcomes.

4 Conclusion

This study showcased the application of machine learning models to analyze bank marketing data and predict customer subscription to term deposits. The Random Forest model emerged as the most accurate, providing valuable insights into influential factors. The findings have implications for tailoring marketing strategies and targeting specific customer segments.

5 Future Work

Future studies could explore advanced models, address class imbalance, conduct hyperparameter tuning, and integrate real-time feedback for model updates. The exploration of additional features and external data sources could enhance model performance and insights.

6 References

1. Moro, S., Cortez, P., & Rita, P. (2014). A Data-Driven Approach to Predict the Success of Bank Telemarketing. *Decision Support Systems*, Elsevier, 62:22-31.

2. Hastie, T.; Tibshirani, R.; Friedman, J. (2009). The Elements of Statistical Learning. Springer.
3. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning. Springer.