Predictive Insights from Portuguese Bank Marketing Data

Gulnar Armour

Springboard

October 29, 2023

Introduction

- Modern banking
  - Bank customers are faced with numerous options
  - Banks to engage customers through data-driven marketing approaches

- Data analysis and Machine learning
  - Patterns in customer data
  - Algorithms to process and interpret data

Introduction

- Business Outcomes Insights:
  - Deposit Subscriptions
  - Loan approvals

- Likelihood of clients subscribing to term deposits
  - ‘Term deposits’: *fixed investment where funds are deposited for a specific period at a fixed interest rate*
  - Models
  - Thresholds
  - Profits

Dataset Overview

- Direct marketing phone campaigns for a Portuguese bank (2008 to 2010)
  - UC Irvine Machine Learning Repository dataset
- Data:
  - 45,000 observations
  - 16 input variables
- Data wrangling process
  - Variable identification
  - No missing values

Dataset Overview

- Input variables:

  1. Bank client data (*age, job, marital status, education, default, balance, housing loan, personal loan)*

  2. Last contact of the current campaign (*contact, day, month, duration)*

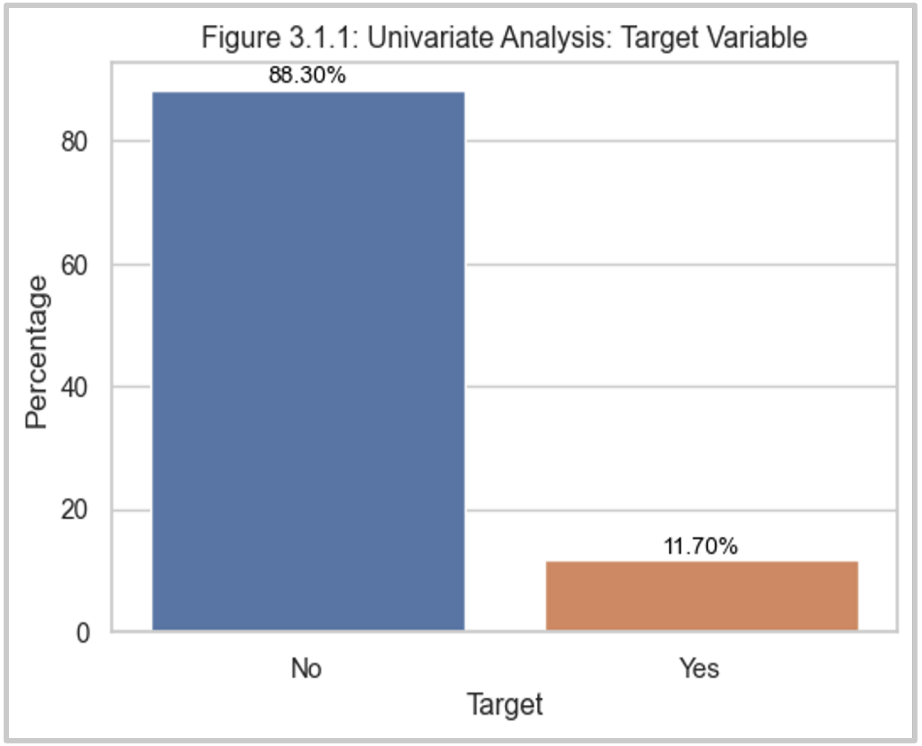
  3. Other attributes (*campaign, pdays, previous, poutcome*)

- Output variable:

  'target' is the client's subscription to a term deposit (binary: "yes", "no")

Univariate Analysis: Target Variable

- Minority class is "yes" at 11.7%
- Majority class is "no" at 88.3%
  - Class imbalance effect
  - Predictive model performance
  - Biased towards the majority class
  - Specialized modeling approach

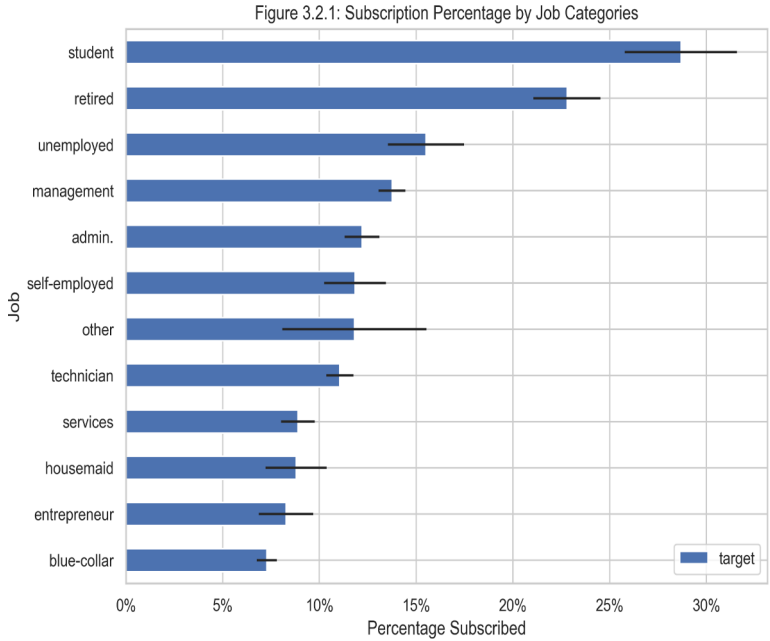


Bivariate Analysis: Job Categories

- Subscription across job categories:
  - students (28.68%)
  - retirees (22.79%)
  - management (13.76%)
  - entrepreneurs (8.27%)

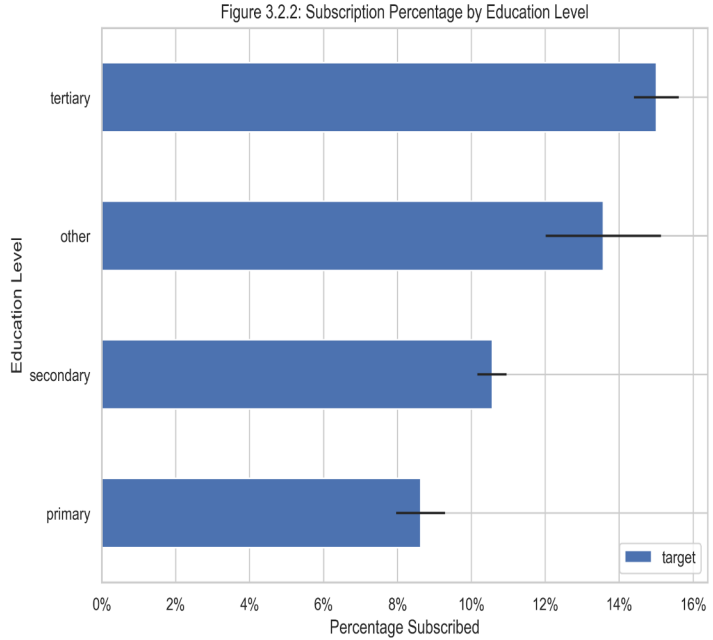
- Students and retirees are more inclined to subscribe
  - Financial independence

- Subscription rates vary by job
  - Tailored marketing approaches



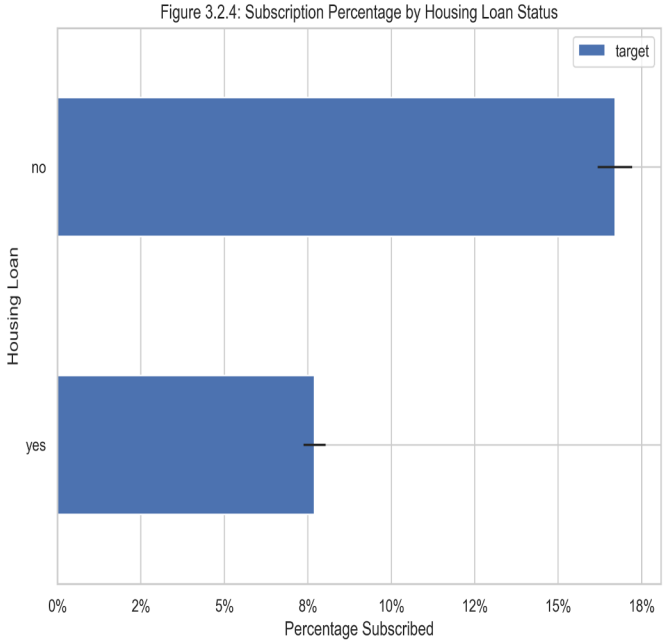
Bivariate Analysis: Education Level

- Tertiary/advanced degree at 15.01%
- Other education at 13.57%
- Secondary at 10.56%
- Primary at 8.6%
  - Higher education - greater likelihood to subscribe
  - Higher education - higher income or financial literacy



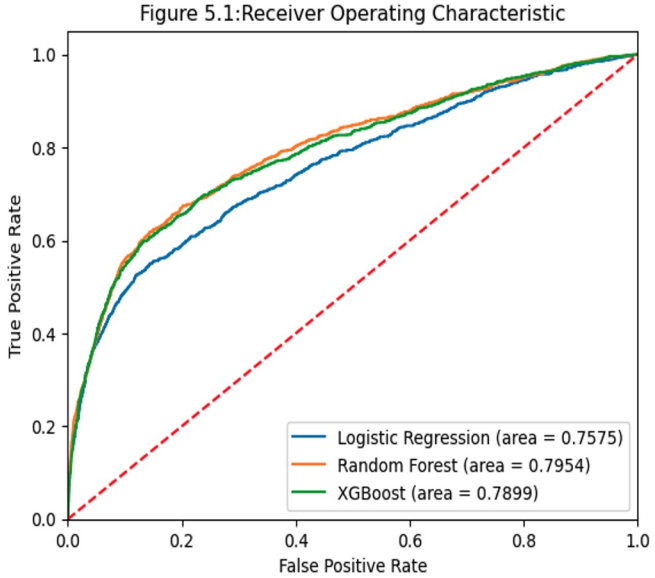
Bivariate Analysis: Housing Loan Status

- Customers with housing loan at 7.70%
- Customers without housing loan at 16.70%
- Housing loan significantly lowered subscription rates
- Customizing marketing strategies based on loan ownership status



Model Performance

- ROC AUC assesses model's ability to differentiate between positive and negative classes
- Based on predictive accuracy:
  - Random Forest (0.7941)
  - XGBoost (0.7899)
  - Logistic Regression (0.7575)
- Selected model for the analysis:
  - Random Forest



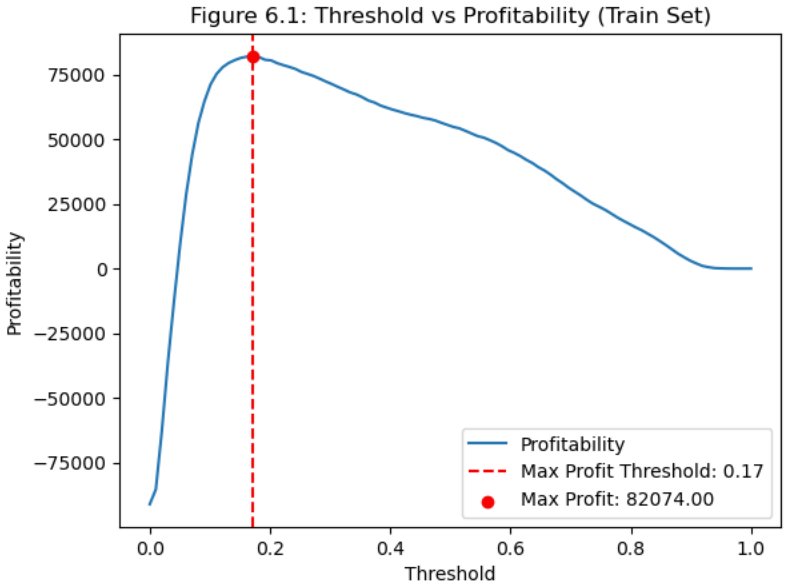
Feature Importance

- Model identified top 5 most important features for predicting subscriptions

Top 5 Features				
balance	*age*	*day*	*campaign*	*pdays*
↓ spending power	↓ financial preferences	↓ pay cycles or budget	↓ increase awareness	↓ contact recency

Threshold vs Profitability

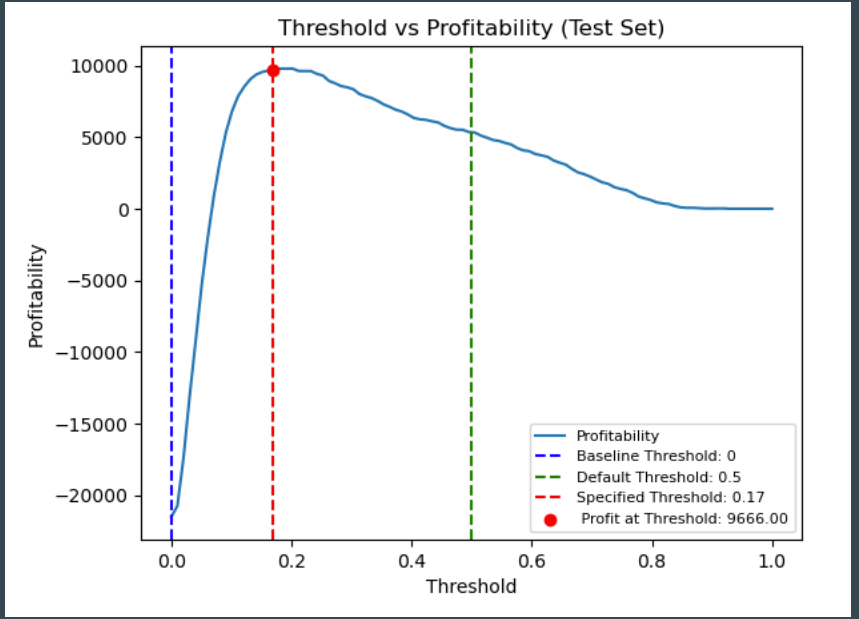
- Graphical representation of relationship between decision threshold and profit
- Banks weigh the number of sales calls against the likelihood of customer sign-ups
- Graph is based on values:
  - A deposit amount of €1000
  - A net investment margin of 3%[1]
  - Revenue per subscription: €30
  - Cost per call: €6[2]



1. Source: www.investopedia.com
2. Source: www.qualtrics.com

Model Evaluation on Test Set

- **Threshold vs Profitability:**
- Baseline threshold at 0 (i.e., calling everyone)
- Default threshold at 0.5
- Specified threshold at 0.17
- **Classification Report:**
- High precision and recall for “No” class
- Lower precision and recall in “Yes” category



Classification Report

Target:	Precision:	Recall:	F1-Score:	Support:
“No”	0.94	0.88	0.91	7952
“Yes”	0.40	0.58	0.47	1091

Conclusion

- Random Forest model achieved ROC AUC of 0.7941
- Key features like *balance, age, day* impact subscription likelihood
- Threshold-profitability analysis found optimal threshold of 0.17
- Addressing class imbalance through threshold optimizes profit
- Model provides predictive capability to optimize marketing
- Future work includes incorporating more features and techniques

Q&A

Thank you!