Experimentation & Model Training with Portuguese Bank Data

Naomi Buell[1] and Richie Rivera[1]

[1]CUNY School of Professional Studies

October 19, 2025

1 Introduction

We analyzed a bank marketing dataset that records the outcomes of a Portuguese bank's telemarketing campaign for term deposits. Our goal was to apply three tree-based machine learning techniques to test six experiments on this dataset, comparing algorithms to choose the optimal model.

2 Exploratory Data Analysis and Pre-processing

We used the `bank-additional-full.csv` data with all examples, ordered by date (from May 2008 to November 2010) (Moro, Rita, and Cortez 2014). The dataset contains 41,188 records and 21 attributes, including both client-level variables (e.g., age, job, marital status) and economic indicators (e.g., employment variation rate, consumer confidence). Initial correlation analysis showed certain macroeconomic features were highly collinear.

We addressed missingness by removing columns with high levels of missingness and weak predictive power (`default`, whether the client has credit in default, was 21% missing missing. `pdays`, the number of days that passed by after the client was last contacted from a previous campaign, was 96% missing). We also removed missing observations of features that were important and had with minor missingness (`education`, `housing`, `loan`, `job`, and `marital`); Since the missingness is low, we will not lose much information by dropping these rows.

We removed the `duration` column since this attribute highly affects the output target (e.g., if `duration` = 0 then `y` = `"no"`) and was not recommended for use with predictive models per the data notes.

We also performed feature engineering on to map the ordinal categorical variable `education` to numeric values corresponding to the number of years of schooling a client received, the `month` variable to a number from 1-12, and the `day_of_week` variable to a number from 1-7, starting on Sunday. We also transformed the rest of the nominal categorical variables using one-hot encoding.

There were 1,946 duplicated observations (5.1%), which we dropped from our dataset, assuming those were erroneous.

Experiments

For all experiments, we split the data into 30% testing/70% training sets. We performed two decision tree experiments: (1) a decision tree with a max depth of four and (2) a decision tree with a max depth of 2. Both models predicted the correct class of the test data with high accuracy (about 90% of test data were correctly classified), but the pruned decision tree with only 2 levels was slightly more accurate. Neither model showed signs of high variance or bias, as training accuracy was similar to test accuracy, and accuracy scores were generally high.

We also performed two random forest experiments: (1) a random forest with 25 trees and (2) a random forest optimized with a grid search to find the best combination of the number of trees and their max depth. The first random forest model appeared to overfit the data, as it was highly accurate in predicting training data (99.14%) but not the test data (much lower accuracy at 89.38%). The recall of the model is very poor, only correctly identifying 30.15% of true positives (subscribers). The second model using a grid search still did not address the overfitting problem (training accuracy = 98.79%, test accuracy = 89.02%) and recall performance was marginally better than the first, correctly identifying 30.46% of true subscribers. This showed that hyperparameter tuning alone was not enough to overcome the challenge of the significant class imbalance in the dataset.

Lastly, we performed two AdaBoost model experiments: (1) a baseline AdaBoost model with 50 estimators and a learning rate of 1.0, and (2) an optimized AdaBoost model where we first balanced the dataset by undersampling the majority class and then performed an extensive grid search to find the best hyperparameters. The initial baseline model performed very poorly, achieving a recall of only 17.38%, confirming that the class imbalance was the primary obstacle to building an effective model. The second experiment, however, was highly successful. By first addressing the data imbalance through undersampling, the model was able to properly learn the patterns of the minority class. The subsequent grid search identified the optimal hyperparameters, resulting in a final model that achieved a recall indicating we could correctly identify 70.24% of actual customers who would subscribe to a term deposit. This marked a significant breakthrough in the project, as we finally created a model that could effectively identify the majority of customers who would subscribe to a term deposit, successfully meeting our primary objective. This success came with a trade-off as precision fell to 0.25 (meaning that many of our predictions are false positives), but it accomplished the goal of maximizing our ability to find positive cases.

Across all modeling methods, the optimized AdaBoost model which used undersampling to balance the dataset performed the best. This approach was the only one to successfully overcome the significant class imbalance and meet our primary objective. After performing an extensive grid search on the balanced training data, the optimal hyperparameters were found to be a learning rate of 0.85 and 20 estimators.

Results

Table 1 summarizes results across all six experiments. In this section, we also present business insights gleaned from these results. Decision Tree 1 recommended focusing marketing efforts during times of lower employment to potentially increase subscription rates. Decision Tree 2's recommendation is the same as above, adding a focus on marketing efforts for clients who responded positively in prior campaigns. It also suggests avoiding heavy marketing during times of high consumer confidence and strong employment. Random Forest 1 indicated the bank should focus on priority age groups (e.g., 55 plus) and running campaigns during favorable economic periods in terms of the Euro Interbank Offered Rate (Euribor). This model is not recommended for business use. It fails to identify nearly 70% of potential subscribers (Recall = 30.46%), making it unreliable for targeted marketing campaigns. The first AdaBoost model (AdaBoost 1) recommends that the bank data focus on employment-related indicators `emp.var.rate` and `nr.employed`, suggesting that these economic indicators are key factors in customers' decisions to subscribe to term deposit. However, it is unsuitable for deployment, since its inability to find potential customers (Recall = 17.38%) means it offers no improvement to marketing ROI. The second AdaBoost model (AdaBoost 2) is the recommended model, with a recall of 70.24%.

Figure 1 displays a preview of the first three decision trees in the recommended model, AdaBoost 2, ensemble. AdaBoost 2's estimators split on a variety of features, both economic (`nr.employed`, `euribor3m`) and otherwise (`poutcome_success`, the outcome of the previous marketing campaign), suggesting that multiple factors can influence subscription likelihood. For maximum impact, marketing campaigns should prioritize clients with a successful outcome in a previous campaign and be timed during periods of lower national employment rates.

An analysis of this model's feature importances provides further strategic direction. **Figure 2** displays the most important features in AdaBoost 2. The most significant predictor by far is a successful outcome in a previous campaign (poutcome_success), indicating that retargeting past subscribers should be the highest priority. Additionally, the number of people employed (nr.employed) and the 3-month Euribor interest rate (euribor3m), are also highly influential. This suggests that marketing efforts will be most effective when timed with periods of economic uncertainty, as customers are more likely to seek safer investments.

Model	Variation	Eval. Metrics
Decision Tree 1	Max depth = 4	Training accuracy = 90.06% Test accuracy = 89.95%
Decision Tree 2	Max depth = 2	Training accuracy: 90.02% test accuracy = 89.96%
Random Forest 1	Number of trees = 25	Training accuracy = 99.14% test accuracy = 89.38% Recall = 30.15%
Random Forest 2	Grid search for optimal trees & depth	Training accuracy = 98.79% test Accuracy = 89.02% Recall = 30.46%
AdaBoost 1	N estimators=50, learning rate=1	Recall = 17.38% Precision = 0.70
AdaBoost 2	Undersampling + Grid search (n estimators = 20, learning rate = 0.85)	Recall = 70.24% Precision = 0.25

Table 1: Summary of Results by Experiment

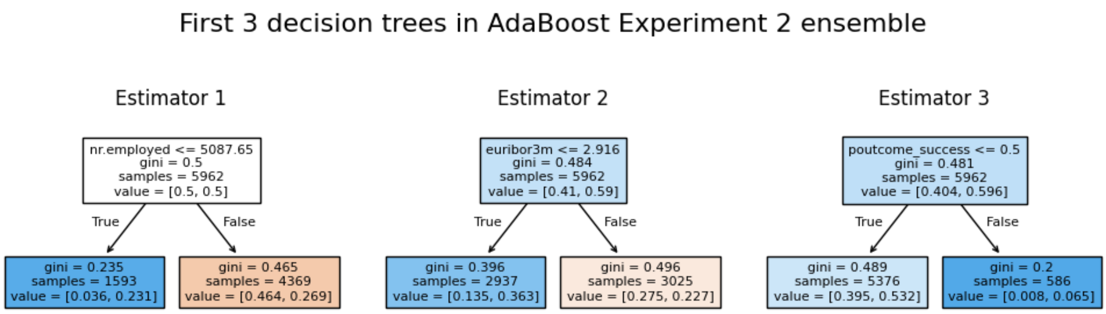

Figure 1: Preview of trees in AdaBoost 2 ensemble

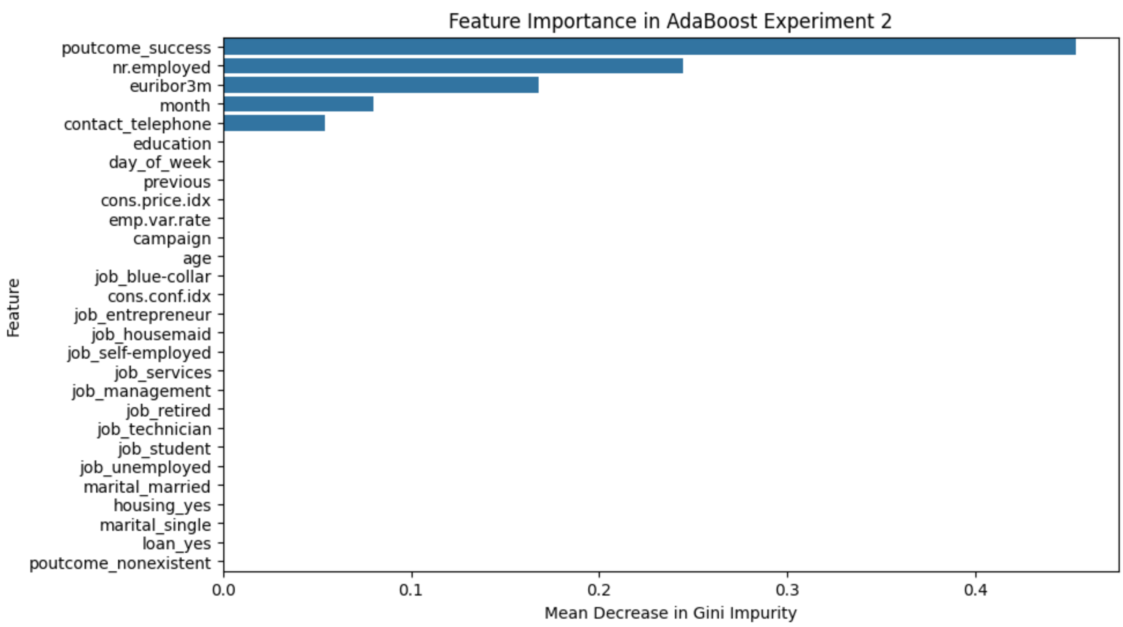

Figure 2: Feature importance in AdaBoost 2

3 Conclusion

In this analysis, the initial models, while demonstrating high accuracy (around 90%), were ultimately misleading due to the class imbalance in the dataset. The key challenge was not achieving high accuracy, but correctly identifying the small minority of clients who would subscribe to a term deposit. The second AdaBoost model that utilized undersampling and grid search was the optimal model because of its superior ability to identify potential subscribers, achieving a test recall of 0.70. This marked a dramatic improvement over all other approaches, successfully flagging 70% of all clients who would subscribe. While this came at the cost of lower precision (0.25), this trade-off is acceptable for a lead-generation initiative where the cost of marketing to a non-converting client is low, but the opportunity cost of missing a potential subscriber is high.

The single most predictive feature was poutcome_success, indicating that clients who had a positive outcome in a previous marketing campaign are by far the most likely to subscribe again. Furthermore, key economic indicators like nr.employed (number of employees) and euribor3m (3-month Euribor rate) were highly influential, suggesting that campaigns are more effective during periods of economic uncertainty when clients may be seeking safer investments. These insights allow the bank to not only trust the model's predictions but also to form broader strategic initiatives based on why it is making them.

For future data science work, we have two recommendations. First, focus on improving the precision of the lead generation model without significantly sacrificing the gains in recall. This could involve exploring more advanced data balancing techniques like SMOTE or experimenting with different algorithms sensitive to class imbalance. Second, an effort should be made to engineer new features that may provide a stronger signal for identifying subscribers, potentially leading to a model that is both sensitive and precise.

If a short term model is necessary, we recommend that the bank deploys the second AdaBoost model to generate targeted marketing lists for its term deposit campaigns. By focusing outreach on the leads identified by this model, the bank can significantly increase the efficiency of its marketing spend, ensuring it reaches a high percentage of likely subscribers and maximizing the campaign's conversion rate.

References

Moro, S., P. Rita, and P. Cortez (2014). *Bank Marketing*. UCI Machine Learning Repository. DOI: https://doi.org/10.24432/C5K306.