Predicting Customer Responses to Term Deposit Campaigns: A Banking Case Study

BA509 Final Project - Group 6

Team Member: Songru Chang, Leo Li, Nancy Liu, Romee Sun, Hanlin Zhao

Part 1 - Introduction

1.1 Background

Marketing campaigns are essential for banks to attract new customers and promote financial products. However, many campaigns face inefficiencies, including low conversion rates and high costs due to outreach to uninterested customers. This project leverages machine learning to develop a predictive model that identifies customers most likely to subscribe to a term deposit. By analyzing historical marketing data, we can distinguish between potential subscribers and non-subscribers. Understanding these key factors helps the bank refine its outreach strategy, improve targeting accuracy, and maximize return on investment (ROI).

1.2 Motivation

Traditional bank marketing often results in wasted resources by targeting unlikely subscribers. This inefficiency raises marketing costs and reduces overall campaign effectiveness. This project aims to address these challenges by:

- Enhancing Marketing Efficiency – Identifying high-potential customers to reduce unnecessary outreach.
- Optimizing Campaign Costs – Allocating resources to maximize ROI.
- Driving Data-Driven Decisions – Using machine learning to generate actionable insights.

1.3 Business Question & Hypothetical Client

The key business question is: How can the bank effectively target customers for its marketing campaign to maximize subscription rates while minimizing costs? Our hypothetical client is a multinational financial institution conducting marketing campaigns for term deposits. The institution seeks a data-driven solution to enhance customer targeting, improve conversion rates, and optimize marketing investments. Its main objectives include predicting customer subscriptions by identifying the most likely subscribers, maximizing conversion rates by focusing on high-potential leads, and reducing marketing costs by minimizing outreach to unlikely subscribers.

1.4 Database Description

The dataset contains 45,211 rows and 17 columns and is derived from the Bank Marketing Dataset. It includes customer information, past marketing interactions, and subscription outcomes, providing a structured foundation for analyzing factors influencing term deposit subscriptions.

1.4.1 Features in the Dataset

1. Demographic Information: Age, job type, marital status, education level.
2. Financial Details: Account balance, presence of a housing loan, personal loan status.
3. Contact Information: Contact method (cellular, telephone), last contact duration.
4. Campaign Details: Number of contacts during the campaign, previous campaign outcome.
5. Target Variable: "Subscribed" (Yes/No) - whether the customer subscribed to the term deposit.

1.4.3 Data Loading and Splitting

The following steps were taken to prepare to look closer to the data for analysis:

1. Loading Data: The dataset was loaded using Pandas, and the target column was renamed from `y` to `subscribed` for clarity.
2. Initial Split: To facilitate exploratory data analysis, the dataset was split into a training set (60%) and a test set (40%), with stratified sampling to preserve the balance of the target variable.

Part 2 - EDA & Data Insights

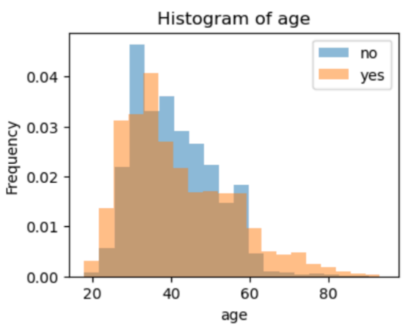
2.1 Exploratory Data Analysis (EDA)

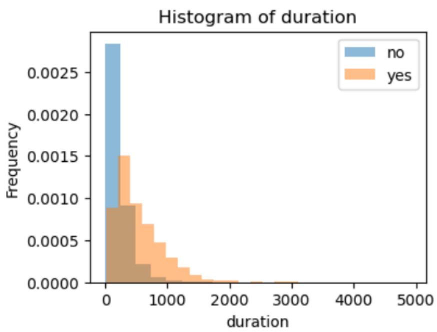
The training dataset consists of 27,126 rows and 17 columns. A statistical summary was generated using the .describe() method to analyze both numerical and categorical variables. Numerical data reveals that age ranges from 18 to 94, with an average of 40.93 and a standard deviation of 10.63, clustering mostly between the late 20s and late 40s. Balance varies from -8,019 to 102,127, with a mean of 1,359.85 and a median of 443.50, indicating a strong right skew. Campaign contact counts range from 1 to 63, but most fall between 1 and 3, with higher values being rare. Call duration spans from 1 to 4,918 seconds, also showing a strong right skew, meaning shorter calls are more frequent. Among categorical variables, "blue-collar" is the most common job, "married" is the predominant marital status, and "secondary" education appears most frequently. The "default" variable is typically "no," while housing loans ("housing") are usually "yes." Additionally, the "poutcome" category is often recorded as "unknown," suggesting no documented outcome for many customers. The dataset was checked for missing values. Fortunately, no missing data was found.

2.2 Distribution of Numerical Columns

Histograms were plotted for each numerical column to understand the distribution of numerical variables across the target classes (subscribed).

The distribution of numerical columns grouped by subscription status reveals several important insights. Call duration emerges as the most critical factor influencing subscriptions. Customers who subscribed had significantly longer call durations, while non-subscribers had shorter calls. This suggests that engaging conversations during marketing calls are essential for increasing conversion rates. Additionally, age shows that subscribers are concentrated in the 30–40 age range, making middle-aged customers a promising target for future campaigns. The account balance also plays a role, as subscribers tend to have higher balances compared to non-subscribers, indicating that tailoring offers to customers with substantial savings might yield better results.





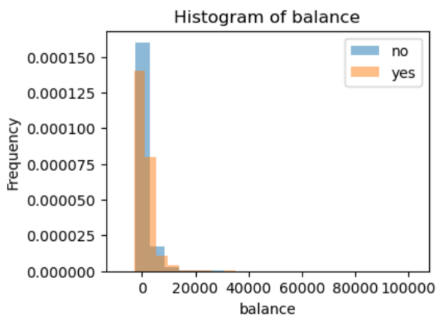
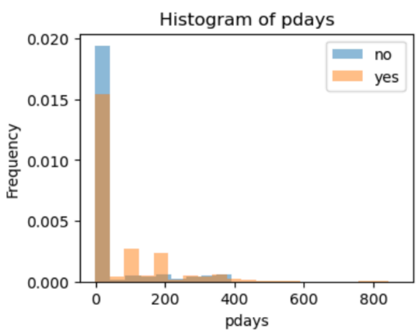
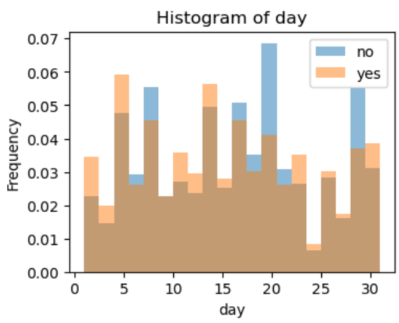


Fig 1. Histogram of age *Fig 2. Histogram of duration* *Fig 3. Histogram of balance*

Other factors, such as pdays (days since the last contact) and previous contacts, highlight the importance of recent and past interactions. When pdays is very low (close to 0), the plot shows more “yes” responses. This suggests customers reach again shortly after a prior contact are somewhat more likely to subscribe. On the other hand, the day of the month on which customers were contacted does not show a clear pattern, suggesting that timing within the month might not significantly affect outcomes. Lastly, the distribution of the campaign (number of contacts during the campaign) indicates diminishing returns from excessive follow-ups, as subscribers generally require fewer contacts than non-subscribers. These findings provide actionable insights for designing more efficient and effective marketing strategies.





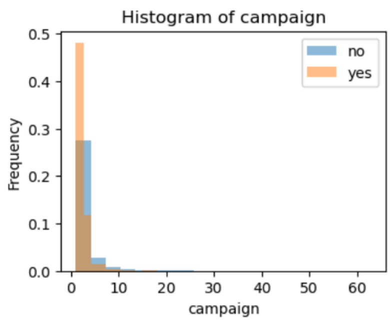


Fig 4. Histogram of pdays *Fig 5. Histogram of day* *Fig 6. Histogram of campaign*

2.3 Target Variable Distribution

Visualizing the target variable (subscribed) reveals a significant class imbalance, with most customers classified as "no" (did not subscribe) and only a small proportion as "yes" (subscribed). This imbalance affects both model performance and business strategy. From a modeling perspective, the model may favor the majority class, reducing its ability to identify potential subscribers. To address this, techniques such as class weighting, oversampling (e.g., SMOTE), and balanced evaluation metrics (precision, recall, and F-score) are necessary to improve predictive accuracy. From a business standpoint, this imbalance highlights inefficiencies in the bank's marketing strategy, as only a small fraction of contacted customers subscribe. A more targeted and personalized approach could enhance engagement and improve conversion rates. These insights are crucial for guiding both model development and marketing optimization.

2.4 Unique Values in Categorical Columns

Analyzing key categorical features (*job, marital status, education*) provides insights into customer demographics. These variables help identify distinct customer groups, influencing their likelihood of subscribing. The diverse distribution of these values suggests that features like job and education may play a crucial role in predicting customer subscriptions, supporting more refined marketing segmentation.

Part 3 - Model Selection & Scoring Metrics

This section outlines the machine learning models implemented in our project, the feature engineering techniques applied, and the scoring metrics used to evaluate model performance. The goal is to develop a robust predictive model that effectively addresses our business question while ensuring interpretability and optimal classification performance.

3.1 Data Preprocessing & Feature Engineering

The preprocessing phase focused on ensuring data quality and optimizing model performance through cleaning, transformation, and feature selection. Missing values were checked and removed if present, though none were found in this dataset. Several features were dropped due to limited predictive value or potential data leakage. The *poutcome* column was removed as past campaign results may not reliably predict current behavior. The *contact* column was excluded since the method of contact (mobile or landline) had little influence on subscription likelihood. The *duration* column was also dropped to prevent data leakage, as it is only known after the call and directly affects the outcome. Additionally, the *day* and *month* columns were removed to avoid redundancy. After these adjustments, the cleaned dataset consisted of 45,211 rows and 14 columns.

Feature engineering was performed to enhance predictive power by categorizing variables appropriately. Numerical features included age, balance, campaign, pdays, and previous, while categorical features such as job and marital status were one-hot encoded to prevent ordinal bias. The education variable was ordinally encoded, maintaining the natural ranking from primary to tertiary. Binary features, including default, housing, and loan, were converted into 0s and 1s for efficient processing without increasing dimensionality. To standardize numerical features, `StandardScaler()` was applied, ensuring consistent value ranges and improving model performance.

A transformation pipeline using `make_column_transformer` was implemented to systematically apply encoding and scaling. The pipeline was fitted on the training dataset and then applied to the test dataset to maintain consistency. After preprocessing, the dataset was refined to 24 features, ensuring a structured and optimized input for accurate and effective model training.

3.2 Model Selection & Hyperparameter Tuning

To develop a well-performing predictive model, three different supervised learning algorithms were implemented: k-Nearest Neighbors (KNN), Logistic Regression, and Random Forest. These models were selected based on their distinct advantages and applicability to the given classification problem. To optimize model performance, hyperparameter tuning was conducted using GridSearchCV, ensuring the best parameter selection for each model.

The KNN model was included due to its ability to capture local patterns in data without making strong assumptions about the underlying distribution. A pipeline was constructed to integrate preprocessing with `KNeighborsClassifier`, and grid search optimization was performed over different values of `n_neighbors` to determine the optimal number of neighbors for classification.

Logistic Regression was selected for its interpretability and efficiency in handling binary classification problems. It was configured with `class_weight={'no':1, 'yes':10}` to counteract the effects of class imbalance, ensuring that minority class predictions were given appropriate weight. A range of values for the regularization parameter `C` was explored using grid search to prevent overfitting while maintaining predictive accuracy.

Random Forest, an ensemble learning method, was included due to its robustness in handling high-dimensional datasets and its ability to model complex relationships between features. The classifier was trained with `class_weight={'no':1, 'yes':10}` and fine-tuned through grid search, optimizing hyperparameters such as the number of estimators (`n_estimators` ranging from 100 to 500) and tree depth (`max_depth` ranging from 3 to None). To ensure fair model evaluation, all models were trained using `GridSearchCV`, with **F-beta Score (β=5)** as the primary scoring metric.

3.3 Scoring Metrics & Evaluation

To assess model performance, multiple evaluation metrics were utilized, each providing insights into different aspects of classification quality. Accuracy, while a widely used metric, was insufficient in this case due to class imbalance, as it does not differentiate between false positives and false negatives effectively.

The **F-beta Score (β=5)** was the primary metric chosen for model optimization, as it balances precision and recall with an emphasis on recall. The F-beta score is calculated as follows:

$$F_{\beta} = (1 + \beta^2) * (Precision \times Recall) / (\beta^2 * Precision + Recall)$$

Unlike the F1 score (where β=1), which treats precision and recall equally, the F-beta score allows for an adjustable trade-off. In this case, β=5 was chosen to prioritize recall significantly over precision. This decision aligns with the business objective of minimizing False Negatives, as missing potential customers (False Negatives) in a subscription campaign is more costly than mistakenly targeting uninterested customers (False Positives). While False Positives may result in wasted marketing resources, False Negatives directly lead to missed revenue opportunities. By optimizing for recall, the model ensures that more potential customers are correctly identified.

The ROC Curve (Fig 8) further illustrates the model's ability to distinguish between positive and negative classes. An AUC-ROC score of 0.72 indicates a good guess, meaning the model performs better than random guessing and has moderate predictive power. With an AUC score of 0.72, the model demonstrates a moderate level of predictive capability.

Additionally, the Confusion Matrix (Fig 7) provides further insight into the model's classification performance. As seen below, while the model captures a significant number of True Positives, it also produces a high number of False Positives, reinforcing the importance of balancing precision and recall.

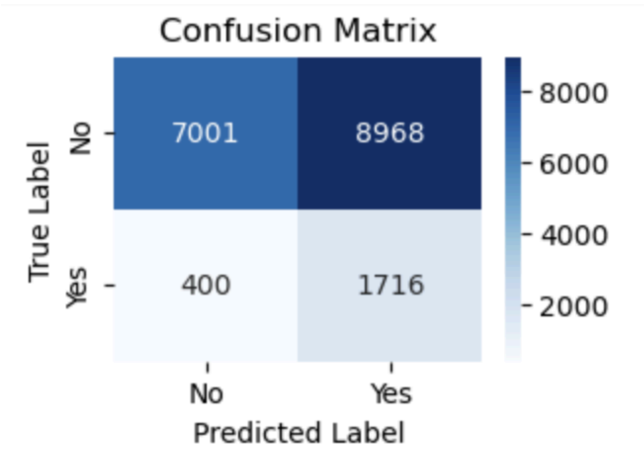


Fig 7. Confusion Matrix

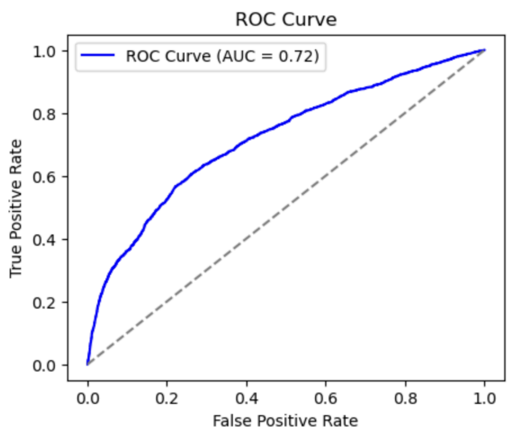


Fig 8. ROC Curve

3.4 Results & Business Interpretation

The final model selection was based on evaluation metrics derived from the Random Forest model, which outperformed the other models in terms of recall and overall predictive performance. The following table summarizes the results:

Metric	Value	Interpretation
Accuracy	0.473	The model correctly classifies ~47.3% of instances. However, accuracy can be misleading due to class imbalance.
F-beta Score (β=5)	0.702	Emphasizes recall, meaning the model effectively identifies 'Yes' cases, which is important for minimizing missed opportunities.
AUC-ROC Score	0.720	Indicates a moderate ability to separate 'Yes' and 'No' classes. The model performs reasonably well across different classification thresholds.
AUC-PR Score	0.301	Reflects the model's ability to distinguish 'Yes' cases in an imbalanced dataset. A lower score suggests challenges in precision-recall trade-offs.

Table 1. Random Forest Model Performance Summary

The Random Forest model demonstrated superior performance, particularly in terms of F-beta score, which confirms its strength in prioritizing recall. The KNN model exhibited overfitting, with high training accuracy but poor test performance, making it unsuitable for deployment. Logistic Regression performed well in terms of AUC-ROC but lacked the predictive power needed for recall-focused optimization.

3.5 Conclusion & Future Considerations

The Random Forest model demonstrated strong predictive capabilities, prioritizing Recall to minimize False Negatives, making it well-suited for business applications. However, further improvements could enhance performance and interpretability.

Neural Networks, such as Multilayer Perceptrons (MLP) or Recurrent Neural Networks (RNNs), could capture more complex, non-linear customer behavior patterns. Additionally, cost-sensitive learning methods that penalize False Negatives more heavily could improve targeting efficiency. For better model transparency, techniques like SHAP values, feature importance visualization, and LIME could provide deeper insights into decision-making, increasing trust and interpretability. Refining these aspects can further strengthen the model's role in supporting business decisions and optimizing targeted marketing strategies.

Part 4 - Feature Importance: Business and Statistical Perspective

Understanding the key factors that influence customer decisions is essential for optimizing direct marketing campaigns. This study focuses on two key questions: How can the bank effectively target customers to maximize subscription rates while minimizing costs? and What are the top three features driving term deposit subscriptions? To answer these, we used two feature importance techniques—Random Forest Feature Importance and Permutation Importance—to identify the most influential predictors.

4.1 Comparing Feature Importance Methods

Random Forest Feature Importance evaluates each feature's contribution by measuring its impact on impurity reduction in decision trees. This method is computationally efficient and widely used but can be biased toward high-cardinality features and may struggle to capture complex feature interactions.

To mitigate these limitations, we also applied Permutation Importance, which assesses feature relevance by measuring how random shuffling affects model performance. Unlike Random Forest, this method provides a more generalizable measure of feature impact, avoiding biases associated with tree-based models and offering a clearer understanding of how each factor influences subscription predictions.

4.2 Top 3 Features from Each Method and Business Implications

The feature importance analysis highlights key factors influencing term deposit subscriptions. In Random Forest, pdays (0.2209) emerged as the strongest predictor, showing that recent contact significantly increases subscription likelihood, reinforcing the importance of timely follow-ups in marketing. Housing loan status (0.1947) suggests that clients without a housing loan may have greater financial flexibility, making them more likely to invest. This insight indicates that banks should prioritize financially flexible customers when promoting term deposits. Previous contacts (0.1779) also play a major role, demonstrating that multiple interactions improve conversion rates. This supports the effectiveness of multi-touch campaigns, where structured follow-ups help nurture customer interest over time.

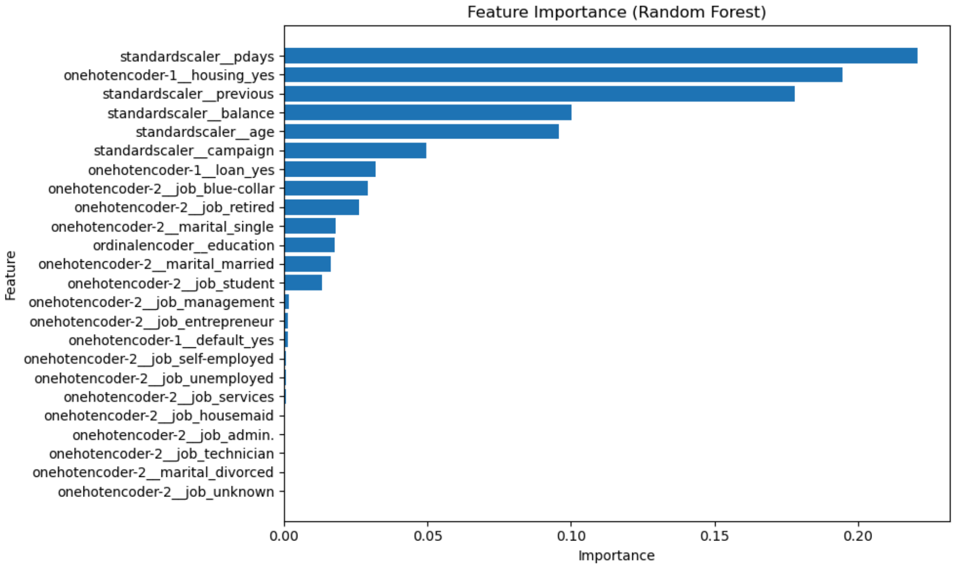


Fig 9. Feature Importance in the Random Forest Model

For permutation Importance, pdays (0.0476) remains the most influential feature but with a lower importance score, suggesting that while recent contact matters, other factors should also be considered in marketing decisions. Previous contacts (0.0467) continue to validate the importance of repeat engagement in building trust and improving conversions. Housing loan status (0.0134) appears far less influential than in the Random Forest method, indicating that financial commitments alone do not strongly predict subscription likelihood. Instead, banks should consider broader financial indicators, such as account balance and transaction history, to refine targeting strategies.

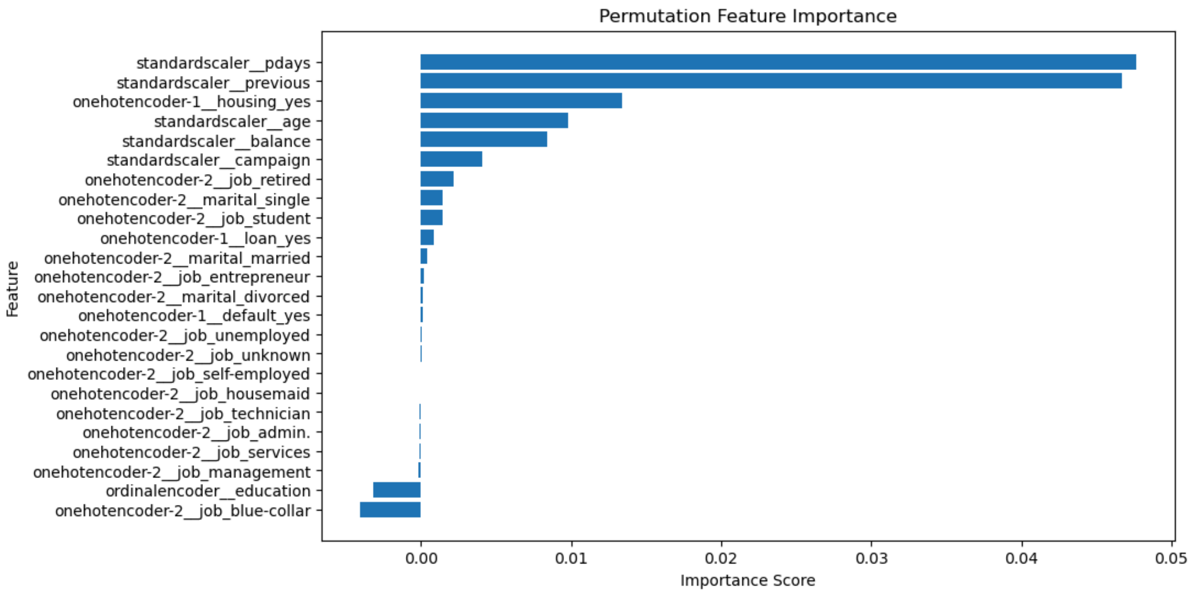


Fig 10. Permutation Feature Importance

4.3 Why Permutation Importance is More Reliable

While both methods identify key predictors, Permutation Importance provides a more accurate feature ranking by avoiding bias toward high-cardinality features and directly measuring their effect on model performance. This method confirms pdays and previous contacts as consistently strong predictors while showing that housing loan status may have been overestimated in Random Forest. By prioritizing insights from Permutation Importance, banks can develop more precise and effective marketing strategies.

4.4 How These Insights Guide Marketing Strategies

The feature importance analysis provides valuable insights for optimizing the bank's marketing strategies.

- Prioritizing Recent Contacts: Since *pdays* (days since last contact) is the most influential factor, marketing campaigns should focus on timely follow-ups. Clients who were contacted recently are more likely to subscribe, emphasizing the need for structured re-engagement plans to improve conversion rates.
- Multi-Touch Campaigns Matter: The significance of *previous contacts* highlights the importance of repeat interactions. Instead of relying on a single outreach attempt, banks should implement multi-touch campaigns that gradually build customer interest and trust over time.
- Financial Flexibility as a Secondary Factor: While features like *housing loan* influence decisions, they are less predictive than customer engagement metrics. Instead of targeting clients solely based on financial indicators, the bank should combine behavioral and financial data to refine its segmentation and outreach efforts.

Part 5 - Conclusion

This project demonstrated how machine learning can improve bank marketing by predicting which customers are most likely to subscribe to term deposits. Instead of relying on broad and inefficient outreach, banks can now use data-driven insights to focus on high-potential customers, reducing wasted marketing efforts and improving conversion rates.

Our analysis found that recent contact history, past interactions, and financial traits play a crucial role in predicting customer subscriptions. The Random Forest model provided the best balance between accuracy and recall, ensuring that potential subscribers were effectively identified. By implementing these insights, banks can refine their marketing strategies, enhance customer engagement, and increase return on investment (ROI). Moving forward, expanding the use of predictive analytics can further optimize customer targeting and campaign efficiency.