2025

Driving Term Deposit Success: Predictive Analytics and Strategic Insights for Optimized Marketing

Actionable insights from advanced modeling for banking competitiveness
1. Jerin George Thomas
2. Mohamed Thalha Ahamed Ali
3. Tereena Perumpally Raphy
4. Vishnu Lal Raveendran Nair

Centennial College | Toronto, ON

Contents

Executive Summary

In today's highly competitive banking industry, efficient and effective marketing strategies are essential for sustainable growth and profitability. Among retail banking products, term deposits are critical for ensuring liquidity and customer retention. However, traditional marketing efforts aimed at term deposits often result in high costs and low returns due to broad targeting and low personalization.

This report leverages predictive analytics on historical marketing campaign data from a Portuguese bank to uncover key drivers of customer subscription behavior. Using advanced modeling techniques Decision Trees, Logistic Regression, and Neural Networks, robust models are built to identify the most promising customer segments and outreach strategies.

The analysis highlights call duration, contact method, financial profile, and prior engagement as key predictors of customer subscription behavior. Strategic recommendations include focusing on high-quality interactions, leveraging real-time scoring within CRM systems, and aligning channel and messaging strategies with customer behavior. Implementing these findings is expected to increase campaign conversion rates, reduce cost-per-acquisition, and strengthen the bank's market competitiveness.

1. Introduction & Industry Context

The financial services industry is navigating a dynamic landscape, facing mounting pressure to optimize operational efficiency while delivering highly personalized, customer-centric experiences. Term deposits, a cornerstone of retail banking for ensuring liquidity and fostering customer loyalty, are increasingly challenged by customer inertia, shifting preferences, and fierce competition from high-yield investment alternatives such as mutual funds, fintech savings platforms, and cryptocurrency products. Marketing teams are tasked with the complex challenge of engaging the right customers at the optimal time through the most effective channels a task made increasingly difficult by outdated, generic targeting approaches that fail to account for individual customer behaviors and preferences.

Industry benchmarks underscore the limitations of traditional strategies, revealing that mass-marketed term deposit campaigns typically achieve conversion rates of just 10–12%, often falling short of justifying their high operational costs. In contrast, forward-thinking banks leveraging predictive analytics have reported significant improvements, including up to a 30% lift in response rates, enhanced customer retention, and substantial reductions in marketing waste. These data-driven approaches enable institutions to move beyond broad, inefficient campaigns toward precise, impactful strategies that resonate with customers.

This report adopts a data-driven approach, harnessing advanced machine learning techniques such as Decision Trees, Logistic Regression, and Neural Networks, applied to real-world campaign data from

a Portuguese bank. The objective is to derive actionable insights that transform decision-making processes, addressing critical questions: who to contact, when to reach out, how to engage effectively, and what motivates a customer to commit to a term deposit. By uncovering the key drivers of subscription behavior, this analysis aims to empower the bank to optimize its marketing efforts, boost conversions, and strengthen its competitive position in a rapidly evolving financial landscape.

2. Comprehensive Dataset Overview

The analysis utilizes the Bank Marketing dataset from the UCI Machine Learning Repository, comprising 45,211 records from direct telephonic marketing campaigns conducted by a Portuguese bank. The target variable is binary, indicating whether a client subscribed to a term deposit ("yes" or "no").

The dataset includes 16 input variables covering client demographics (e.g., age, job, marital status), financial data (e.g., account balance, housing/loan status), and campaign details (e.g., number of contacts, communication type, call duration). The campaign success rate is 11.7%, reflecting class imbalance, a common issue in real-world marketing data that requires careful handling for robust model performance.

Note on Variables: The variable campaign is numeric, indicating the number of contacts with a client during the current campaign, and treated as an interval variable. The balance variable represents the average account balance, with negative values included to reflect overdraft scenarios, as validated by domain expertise.

This dataset provides a solid foundation for developing predictive models to uncover actionable insights into customer subscription behavior.

3. Data Quality Assessment and Variable Rejection Rationale

While the dataset was structurally clean. Free from missing values and inconsistencies, several variables required evaluation for relevance and analytical contribution.

Rejected Variables & Rationale

Variable	Correlation with Target	Reason for Rejection
day	-0.028	Minimal predictive power; no actionable business logic.
pdays	Near zero	95%+ of values were '999' (never contacted); uninformative.

previous	0.093	Somewhat predictive, but redundant. The campaign variable and the newly engineered Contacted feature captured engagement more effectively.

Additionally, extreme outliers (e.g., large balances, abnormally high campaign calls) were addressed using cap-and-floor techniques to prevent distortion in model training. Skewed variables like balance, campaign, and duration were transformed using logarithmic functions to improve distribution symmetry.

The variable transformation process ensured improved model stability, accuracy, and interpretability.

4. Feature Engineering and Transformation

To enrich the model's learning capability, several features were engineered and transformed existing ones:

- Day_of_Month: Replaced the day variable to better capture periodic patterns.

- Contacted: New categorical field based on pdays. Values such as 'Previously Contacted' and 'Never Contacted' helped group clients effectively.

- Log Transforms: Applied to balance, campaign, and duration to reduce skewness and improve model performance.

- Grouped Categories:

    o Job: Clustered into Admin & Tech, Management & Professional, and Low-Income Groups.

    o Education: Consolidated into 'School', 'College', and 'Unknown'.

    o Month: Grouped by quarter to detect seasonality effects.

These engineered features were crucial in driving the models' ability to learn complex customer behavior patterns and enhance interpretability for business use.

5. Predictive Modelling Approaches and Results

5.1 Decision Trees

The Decision Tree model offered high interpretability. The best-performing tree was pruned to 3 splits for clarity. It revealed:

- Short Calls (<214.5 seconds / ~3.6 minutes): Less than 3% conversion.

- Medium Calls (214.5–628.5 seconds / ~3.6–10.5 minutes): Showed moderate success (16–20%), especially when the previous campaign was successful.

- Long Calls (>628.5 seconds / ~10.5 minutes): Conversions peaked at 58% when duration exceeded 827.5 seconds (~13.8 minutes).

This model provides actionable rules that can be used to train call center staff or set up automated call scoring.

5.2 Logistic Regression

Regression offered valuable odds-based interpretations:

Feature	Odds Ratio	Insight
Long Call Duration	6.57	Clients are 557% more likely to subscribe.
Contacted via Telephone	4.16	Higher odds compared to unknown contact methods.
High Account Balance (log)	1.40	Wealthier clients are more receptive.
No Loan/Housing	2.32/1.58	Clients without obligations are more open to deposits.

The regression model had a validation ASE of ~0.0723 and was used for feature importance and scoring.

5.3 Neural Networks

Several neural network configurations were tested. The best performing was trained on the Cap & Floor-adjusted dataset:

- ASE: 0.06477 (lowest of all models).

- ROC AUC: ~0.93.

- Use Case: Excellent for CRM integration; real-time scoring capability.

While less interpretable, the Neural Network was the most accurate and scalable model, ideal for live deployment.

6. Actionable Strategic Recommendations

1. Prioritize Long and Meaningful Calls

- Incentivize agents to sustain calls for 10+ minutes.

- Use decision tree rules to flag promising customers early in the call.

- Develop talking scripts tailored to income groups, age bands, and loan status.

2. Use Contact History to Segment Leads

- Clients with previous successful engagement have >70% conversion likelihood in medium-duration calls.

- Re-target these segments with refined offers and personalized follow-ups.

3. Embed Scoring into CRM Workflows

- Use regression outputs and neural networks to generate real-time "Likelihood to Subscribe" scores.

- Set automated thresholds to trigger sales calls, emails, or promotions based on score tiers.

4. Optimize Channels and Messaging

- Prefer telephone and cellular methods.

- Avoid unknown contact methods.

- Develop scripts focused on value (safety, returns, convenience).

5. Regular Model Refresh and Feedback Loop

- Retrain models quarterly using recent campaign outcomes.

- Integrate agent feedback into training scripts and targeting rules.

7. Implementation Roadmap

Phase	Timeline	Activities
Phase 1	0–3 Months	Deploy models in CRM, train staff, define KPIs
Phase 2	3–6 Months	Launch dashboards, monitor performance, adjust scripts
Phase 3	6–12 Months	Retrain models, refine rules, expand into other banking products

Each phase includes cross-functional collaboration between marketing, IT, and analytics teams.

8. Conclusion and Strategic Impact

This project delivers a comprehensive, data-driven strategic toolkit designed to revolutionize the bank's term deposit marketing efforts. By leveraging advanced predictive modeling techniques: Decision Trees, Logistic Regression, and Neural Networks, applied to a robust dataset from a Portuguese bank's marketing campaigns, the analysis has identified critical drivers of customer subscription behavior. These drivers include engagement duration, financial characteristics such as

account balance, contact history, and optimized channel strategies, providing a clear roadmap for targeting high-potential customers with precision.

The proposed solutions are both statistically robust and operationally feasible, ensuring practical implementation within existing banking workflows. By adopting these recommendations, the bank can expect to achieve the following transformative results:

- **Improve Targeting Accuracy:** By utilizing real-time scoring from neural networks and regression-based insights, marketing teams can prioritize high-likelihood prospects, reducing wasted efforts on low-conversion segments.
- **Increase Conversion Rates:** Insights from decision tree analysis, such as the significant impact of longer call durations (e.g., >10.5 minutes yielding up to 58% conversion rates), enable tailored engagement strategies that resonate with customers.
- **Reduce Campaign Costs:** Optimized channel selection (favoring telephone and cellular methods) and refined lead segmentation based on contact history minimize resource expenditure, ensuring higher returns on marketing investments.
- **Strengthen Customer Relationships:** Personalized outreach, informed by predictive models, fosters trust and loyalty, aligning offerings with customer needs and preferences to enhance long-term engagement.

In today's rapidly evolving financial landscape, where competition from alternative investment products and fintech disruptors is intensifying, data-driven strategies are no longer optional, they are imperative for sustained success. The actionable insights and implementation roadmap provided in this report empower the bank to differentiate itself as a leader in retail banking. By embracing predictive analytics, the bank can act with intelligence, agility, and measurable impact, achieving higher conversion rates, lower acquisition costs, and stronger customer relationships. This strategic approach not only positions the bank for immediate marketing success but also lays a scalable foundation for future innovation across other banking products and services, ensuring long-term competitiveness in a dynamic market.

Appendix

Model Variations and Exploration

Multiple configurations were tested to refine the predictive models, ensuring robust performance.
Below are summaries of key approaches based on the analysis:

Table 1: ASE for different models

Model	Average Squared Error (ASE)
Neural Network	0.0648
Random Forest	0.067536
Decision Trees	0.065974
Logistic Regression	0.068

Table 2: Decision Trees

Model	Leaves	ASE
Maximal Tree	46	0.070512
ASE Tree	22	0.070082
ASE Tree-3 Splits	29	0.065974
Misclassification Tree	17	0.072554

Table 3: Logistic Regression

Model	ASE
Full Regression	0.072347
Forward Regression	0.072348
Backward Regression	0.072347
Stepwise Regression	0.072348

Table 4: Regression Summary

Parameter	Comparison	Odds Ratio	Likelihood Change (%)

Contacted	No vs. Yes	0.651	-65.0%
LOG_REP_balance	Higher vs. Lower	1.406	+40.0%
LOG_REP_campaign	Higher vs. Lower	0.662	-66.0%
LOG_REP_duration	Longer vs. Shorter	6.574	+557.4%
REP_education (College)	College vs. Unknown	1.244	+24.4%
REP_education (School)	School vs. Unknown	0.876	-87.6%
REP_job (Admin & Tech)	Admin & Tech vs. Mgmt & Prof	0.978	-97.8%
REP_job (Low wages & Others)	Low wages & Others vs. Mgmt & Prof	1.351	+35.1%
contact (cellular)	Cellular vs. Unknown	3.621	+262.1%
contact (telephone)	Telephone vs. Unknown	4.156	+315.6%
housing	No vs. Yes	2.325	+132.5%
loan	No vs. Yes	1.586	+58.6%

Note: Odds ratios reflect the change in likelihood of subscribing to a term deposit. Values below 1 indicate a decreased likelihood (negative percentage), while values above 1 indicate an increased likelihood (positive percentage).

Table 5: Neural Networks

Model	Iteration	ASE Value	Converged
NN Cap and Floor	86	0.064776953	104
NN Partition	160	0.065488481	184
NN Transform	33	0.065928802	70
NN Forward Reg 4H	32	0.067886562	56
NN Forward Reg 5H	99	0.06807946	114
NN Recode	59	0.068113586	73
NN Forward Reg 7H	50	0.068269747	132
NN Forward Reg 3H	80	0.068372297	84
NN Forward Reg 9H	47	0.068396078	189
NN Forward Reg 6H	83	0.068511231	110
NN Forward Reg 8H	81	0.068796695	128

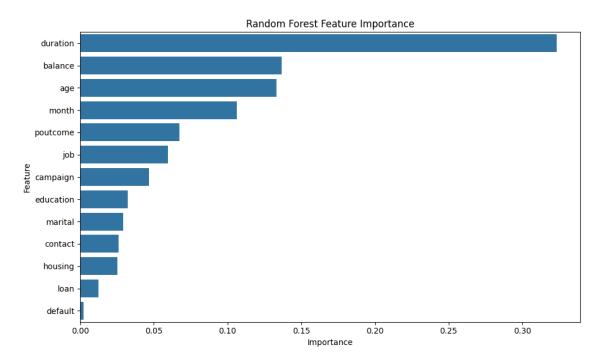**Figure 1: Feature importance in Random Forest**

Random Forest Feature Importance

Figure 2: Heatmap including dropped variables

Correlation Heatmap - Bank Marketing Dataset