International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | Mar – 2026 DOI: 10.55041/ISJEM05633
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Predictive Modelling Framework Using Machine Learning for Bank Marketing Campaign Optimization

Prasad Kamble[1]

Student, MBA Department
Dhole Patil College of Engineering, Pune

Prof.Kanifnath S Satav[2]

Professor, MBA Department
Dhole Patil College of Engineering, Pune

---***---

Abstract

Marketing campaigns are widely used by banks to promote financial products such as term deposits, loans, and investment services. However, traditional marketing approaches often involve contacting a large number of customers without accurately identifying potential subscribers, leading to low response rates and increased operational costs. With the rapid growth of data availability and advancements in machine learning, financial institutions can utilize predictive analytics to improve customer targeting and marketing efficiency.

This study develops a predictive modelling framework using machine learning techniques for bank marketing campaign optimization. The research utilizes the Bank Marketing dataset containing 41,188 customer records with demographic attributes, campaign interaction history, and macroeconomic indicators. The modelling process includes exploratory data analysis, domain-driven feature engineering, statistical feature selection, and a structured preprocessing pipeline using scaling and one-hot encoding.

Several machine learning algorithms including Logistic Regression, Random Forest, XGBoost, LightGBM, and K-Nearest Neighbors were evaluated using performance metrics such as F1-score, ROC-AUC, Average Precision, and cross-validation performance. To address class imbalance in the dataset, class-weight balancing techniques were incorporated into the modelling process. Experimental results indicate that the Balanced LightGBM model achieved the best overall performance, providing the highest composite evaluation across F1-score, ROC-AUC, and Average Precision metrics. The proposed framework demonstrates how machine learning-based predictive analytics can significantly improve marketing efficiency, reduce campaign costs, and enhance customer targeting strategies for banking institutions.

Key Words**:** Machine Learning, Predictive Analytics, Bank Marketing, Business Analytics.

I. INTRODUCTION

The banking industry relies heavily on marketing campaigns to promote financial products and services such as term deposits, credit cards, insurance policies, and investment plans. Direct marketing methods, including telemarketing and digital communication, are commonly used to reach potential customers. However, traditional marketing strategies often involve contacting a large group of customers without accurate targeting, which leads to low conversion rates and inefficient use of marketing resources.

With the advancement of information technology and data analytics, financial institutions now generate large volumes of customer data through various interactions such as transactions, online banking, and marketing campaigns.

Machine learning (ML) has emerged as a powerful tool for analyzing large datasets and identifying patterns that may not be easily detected using traditional statistical methods. Machine learning algorithms can learn from historical data and generate predictive models that estimate the likelihood of future events [1]. In the context of bank marketing, predictive models can help identify customers who are more likely to subscribe to financial products.

Predictive analytics allows organizations to shift from traditional mass marketing strategies to data-driven targeted marketing approaches. By analyzing historical campaign outcomes, banks can develop models that predict customer responses and focus their marketing efforts on high-potential customers.

Several studies have demonstrated the effectiveness of machine learning techniques in customer behavior prediction and marketing optimization [2]. Ensemble learning algorithms such as Random Forest and Gradient Boosting have shown strong predictive performance due to their ability to combine multiple decision trees and reduce overfitting [3].

The objective of this research is to develop a predictive modeling framework using machine learning techniques to optimize bank marketing campaigns. The study uses the Bank Marketing dataset to analyze customer attributes and campaign outcomes. Multiple classification algorithms are implemented and evaluated to determine the most suitable model for predicting term deposit subscriptions.

The results of this study aim to demonstrate how machine learning can improve marketing efficiency, reduce campaign costs, and support data-driven decision-making in the banking sector.

II. RELATED WORK

The application of machine learning in banking and financial marketing has gained significant attention in recent years. Researchers have explored various predictive modeling techniques to improve customer targeting and campaign performance.

Moro et al. [4] conducted a well-known study using the Bank Marketing dataset to predict the success of telemarketing campaigns. Their research demonstrated that data mining techniques can effectively analyze customer characteristics and campaign attributes to improve marketing outcomes.

Burez and Van den Poel [5] investigated predictive analytics methods for marketing response modeling. The study showed that predictive models can significantly improve campaign efficiency by identifying customers who are more likely to respond positively.

Verbeke et al. [6] examined the use of machine learning algorithms for marketing decision support systems. Their findings indicated that classification models such as decision trees and ensemble learning techniques can enhance prediction accuracy in marketing applications.

Baesens et al. [7] explored the role of predictive analytics in financial services and highlighted how machine learning models can support strategic decision-making by analyzing large volumes of customer data.

Recent research has also emphasized the importance of feature engineering and data preprocessing in predictive modeling. Proper handling of categorical variables, missing values, and class imbalance can significantly improve model performance [8].

Despite the growing use of machine learning in marketing analytics, many organizations still rely on traditional marketing strategies due to the absence of structured predictive frameworks. Therefore, there is a need for comprehensive machine learning pipelines that integrate data preprocessing, model comparison, and evaluation for practical marketing optimization.

III. RESEARCH GAP

Although several studies have applied machine learning techniques to banking datasets, many financial institutions still rely on traditional marketing strategies that involve broad customer targeting. These approaches often fail to accurately identify customers who are most likely to respond to marketing campaigns.

Existing research primarily focuses on algorithm development rather than building comprehensive predictive frameworks that combine data preprocessing, feature engineering, model comparison, and evaluation.

Furthermore, practical decision-support systems that integrate predictive analytics into marketing strategy development are still limited. This research aims to address this gap by developing a machine learning-based predictive framework that can assist banks in identifying potential subscribers and optimizing marketing campaigns.

IV. OBJECTIVE OF THE STUDY

The primary objective of this research is to develop a predictive modeling framework using machine learning techniques to identify customers who are likely to subscribe to a term deposit.

Another objective is to evaluate multiple classification algorithms and compare their predictive performance using suitable evaluation metrics.

The study also aims to analyze the impact of customer demographic characteristics and campaign-related attributes on subscription behavior.

V. RESEARCH METHODOLOGY

This research follows a quantitative analytical approach based on machine learning and predictive analytics techniques. The dataset used in this study is the publicly available Bank Marketing dataset containing 41,188 observations and 21 features representing customer demographics, campaign interactions, and macroeconomic indicators. The target variable indicates whether a client subscribed to a term deposit.

The modelling process began with exploratory data analysis (EDA) to understand the dataset structure, distribution of variables, and potential relationships between features and the target variable. Initial analysis showed that the dataset contains both numerical and categorical features and exhibits significant class imbalance, with a much higher number of non-subscribers compared to subscribers.

During the data preprocessing stage, the duration variable was removed based on domain knowledge because it

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | Mar – 2026 DOI: 10.55041/ISJEM05633
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

represents the call duration after the interaction and would introduce data leakage if used as a predictor.

Unknown values in categorical variables were replaced using the most frequent category to ensure consistency in the dataset.

To improve predictive performance, feature engineering techniques were applied. Two new features were introduced:

• was_contacted: indicating whether the client had been contacted in previous campaigns

• contact_frequency: representing the level of past campaign interactions

Statistical feature selection was then performed to identify the most relevant predictors. Numerical features were evaluated using point-biserial correlation, while categorical variables were assessed using the Chi-square test of independence. Features with statistically significant relationships with the target variable were retained for modelling.

After feature selection, the dataset was split into training and testing sets using stratified sampling to preserve the class distribution.

A preprocessing pipeline was implemented using ColumnTransformer, combining two separate pipelines:

Numerical features were processed using:

• Median imputation for missing values

• StandardScaler for feature scaling

Categorical features were processed using:

• Most frequent value imputation

• OneHotEncoder to convert categorical variables into numerical format.

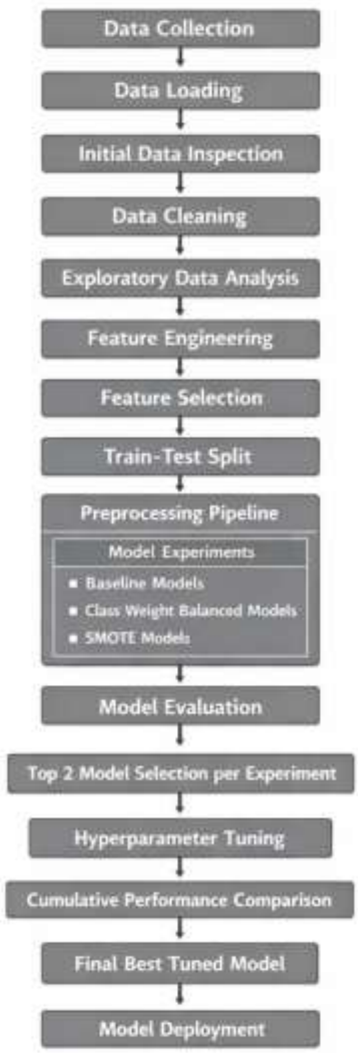
This structured preprocessing pipeline ensured that all transformations were applied consistently during model training and evaluation.

Multiple machine learning models were evaluated including Logistic Regression, Random Forest, XGBoost, LightGBM, and K-Nearest Neighbors. Model performance was evaluated using several metrics including F1-score, ROC-AUC, Average Precision, and cross-validation F1 scores.

To address the class imbalance problem in the dataset, class-weight balancing techniques were incorporated. Tree-based algorithms such as LightGBM and Random Forest were configured with class_weight = "balanced", while XGBoost utilized the scale_pos_weight parameter.

Hyperparameter optimization was performed using RandomizedSearchCV to further improve model performance.

Fig 1: Machine Learning Pipeline Architecture for Bank Marketing Prediction



VI. MODEL IMPLEMENTATION AND RESULTS

To evaluate the predictive performance of machine learning algorithms for bank marketing campaign optimization, multiple classification models were trained and compared.

The evaluated models included Logistic Regression, Random Forest, XGBoost, LightGBM, and K-Nearest Neighbors. Each model was integrated into the preprocessing pipeline to ensure consistent data transformation and evaluation.

The models were trained using the training dataset and evaluated using the test dataset. Performance was assessed using several evaluation metrics including:

• F1-Score

• ROC-AUC

• Average Precision

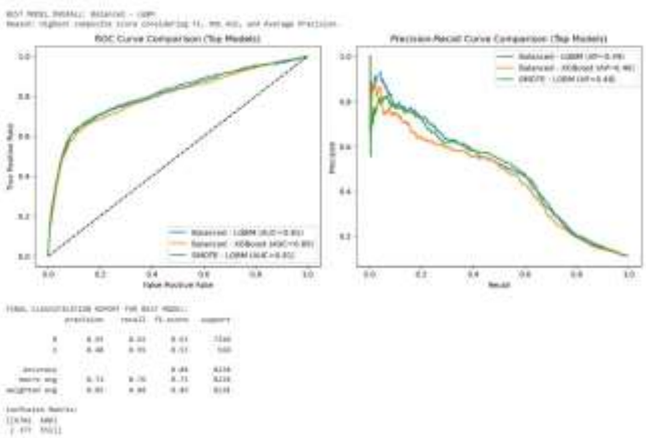
• Cross-Validation F1 Mean

The comparative results of the evaluated models are shown in Table 1.

Table 1: Comparison of Machine Learning Models

Model	F1 Score	ROC AUC	Avg Precision	CV F1 Mean
XGBoost	0.4129	0.8005	0.4763	0.3686
LightGBM	0.393	0.8103	0.4938	0.3739
Random Forest	0.3867	0.78	0.4186	0.3589
K-Nearest Neighbors	0.38	0.7434	0.3339	0.3603
Logistic Regression	0.3374	0.801	0.466	0.3368

Fig 2: ROC-AUC Curve & Confusion Matrix



Discussion of Results

Although XGBoost achieved the highest F1-score in the initial evaluation, further experiments were conducted using class-weight balancing strategies to address dataset imbalance.

After applying class balancing techniques, LightGBM with balanced class weights produced the highest composite performance across F1-score, ROC-AUC, and Average Precision, making it the most effective model for this predictive task.

The Balanced LightGBM model demonstrated strong capability in identifying positive subscription cases while maintaining stable performance across cross-validation folds.

These results highlight the importance of addressing class imbalance when developing predictive models for marketing response prediction.

VII. MAJOR FINDINGS

The experimental results highlight several important insights regarding the application of machine learning in bank marketing campaign optimization.

First, the study demonstrates that predictive analytics can significantly improve marketing efficiency by identifying customers who are more likely to subscribe to financial products.

Second, feature engineering and statistical feature selection played an important role in improving model performance by identifying relevant predictors from customer demographic and campaign interaction data.

Third, handling class imbalance using class-weight balancing significantly improved the ability of the models to correctly identify subscribers.

Finally, the Balanced LightGBM model achieved the best overall performance across multiple evaluation metrics, indicating its suitability for predictive marketing analytics applications in the banking sector.

VIII. CONCLUSION

This study developed a machine learning-based predictive modelling framework for optimizing bank marketing campaigns. Using the Bank Marketing dataset, several classification algorithms were implemented and evaluated to determine the most effective model for predicting customer subscription behavior.

The modelling pipeline included exploratory data analysis, feature engineering, statistical feature selection, preprocessing using scaling and one-hot encoding, model training, cross-validation, and hyperparameter optimization.

Experimental results demonstrate that machine learning techniques can significantly enhance marketing decision-making by accurately predicting customer responses to marketing campaigns.

Among all evaluated models, the Balanced LightGBM model achieved the best overall performance, providing the highest composite evaluation across F1-score, ROC-AUC, and Average Precision metrics.

The proposed predictive modelling framework can help banks improve marketing campaign efficiency, reduce operational costs, and implement data-driven customer targeting strategies.

Future research may explore deep learning approaches, additional feature engineering techniques, and real-time predictive systems to further enhance marketing optimization.

ISSN: 2583-6129
DOI: 10.55041/ISJEM05633

REFERENCES

[1] Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques.

[2] Provost, F., & Fawcett, T. (2013). Data Science for Business.

[3] Breiman, L. (2001). Random Forests. Machine Learning.

[4] Moro, S., Cortez, P., & Rita, P. (2014). A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems.

[5] Burez, J., & Van den Poel, D. (2009). Handling Class Imbalance in Customer Churn Prediction.

[6] Verbeke, W., Martens, D., & Baesens, B. (2012). Predictive Modeling for Customer Analytics.

[7] Baesens, B. (2014). Analytics in a Big Data World.

[8] Kuhn, M., & Johnson, K. (2013). Applied Predictive Modeling.