American Journal of Business and Operations Research

ISSN: 2692-2967 Vol: 11, No. 01, 2025 PP. 65–72

Article DOI: https://doi.org/10.54216/AJBOR.110110

FULL LENGTH ARTICLE

Predictive Analytics and Machine Learning in Direct Marketing for Anticipating Bank Term Deposit Subscriptions

Ahmed Mohamed Zaki[1] **Nima Khodadadi**[2] **Wei Hong Lim**[3] **S. K. Towfek**[1,*]

[1] Computer Science and Intelligent Systems Research Center, Blacksburg 24060, Virginia, USA

[2] Department of Civil and Architectural Engineering, University of Miami, Coral Gables, FL, USA

[3] Faculty of Engineering, Technology and Built Environment, UCSI University, Kuala Lumpur 56000, Malaysia

Emails: azaki@jcsis.org · nima.khodadadi@miami.edu · limwh@ucsiuniverisity.edu.my · sktowfek@jcsis.org

Received: February 03, 2023 Revised: March 24, 2023 Accepted: May 12, 2023 ⋆ Corresponding author

ABSTRACT

Direct marketing strategies in the banking sector have undergone evolution with the integration of predictive analytics and machine learning techniques. The focus of this study is on the utilization of these technologies to foresee bank term deposit subscriptions. The methodology encompasses data exploration, visualization, and the implementation of machine learning models. Datasets from Kaggle are employed, relationships within the data are explored through crosstabulations and heat maps, and feature engineering and preprocessing techniques are applied. The study individually implements models such as SGD Classifier, k-nearest neighbor Classifier, and Random Forest Classifier. The results indicate that the best performance among the evaluated models was exhibited by the Random Forest Classifier, achieving an accuracy of 87.5%, a negative predictive value (NPV) of 92.9972%, and a positive predictive value (PPV) of 87.8307%. These findings provide valuable insights for banks seeking to optimize their marketing strategies within the dynamic landscape of the financial industry.

Keywords: Direct Marketing ▪ Predictive Analytics ▪ Machine Learning ▪ Bank Term Deposit Subscriptions ▪ Data Exploration ▪ Feature Engineering

1. INTRODUCTION

The role of direct marketing in the banking sector is foundational, serving as a strategic conduit for personalized and targeted communication with clients. In the contemporary financial landscape, predictive analytics and machine learning emerge as transformative forces, introducing unprecedented opportunities to refine and elevate direct marketing strategies. This research thoroughly explores the intricate intersection of advanced analytics and the anticipation of bank term deposit subscriptions. The primary objective is to scrutinize the application of sophisticated analytical models that transcend traditional approaches to significantly enhance the precision and efficacy of direct marketing endeavors within the banking domain [1]. The swift evolution of technology and the exponential expansion of data have fostered an environment where traditional marketing methods are increasingly supplemented, if not supplanted, by advanced analytical tools. Rooted in predictive analytics and optimization algorithms, these tools offer a nuanced understanding of customer behavior, enabling financial institutions to tailor their marketing efforts with unprecedented granularity. The research embarks on unraveling the potential impact of these tools on the banking industry, as shown in Figure 1, where personalized interactions and targeted campaigns are pivotal for success [2].

Anticipating bank term deposit subscriptions assumes paramount importance in a competitive market. By proactively identifying potential subscribers, banks can optimize resource allocation, personalize customer engagement, and

© 2024 The Author(s). Published by ASPG. This article is licensed under CC BY 4.0.

enhance conversion rates. The essence of this research lies in navigating the intricate landscape of direct marketing, where data-driven insights become the compass guiding strategic decisions. As the financial sector undergoes digital transformation, integrating predictive analytics and machine learning emerges as a strategic imperative for banks aiming to remain competitive and relevant.

In the subsequent sections, this paper unfolds a comprehensive exploration, commencing with the intricacies of data collection and exploration [3]. Acknowledging that data constitutes the lifeblood of predictive modeling, the research meticulously delves into relevant sources, assimilating diverse information, including client demographics, financial history, and interactions with previous marketing campaigns. This robust dataset forms the foundation for subsequent analysis, reflecting the richness of information necessary for training and validating advanced machine learning models.

The authors then seamlessly transition to exploratory data analysis (EDA), a pivotal step in unraveling the hidden patterns and relationships within the dataset. Through counterplots, distribution analyses, and crosstabulation heatmaps, the research unveils insights that can considerably influence the success of direct marketing campaigns. The significance of EDA lies in uncovering patterns and informing subsequent feature selection and model development, ensuring that the predictive models are grounded in a thorough understanding of the underlying data [4]. The research incorporates feature engineering and preprocessing techniques to prepare the data for the intricacies of machine learning models. Categorical variables undergo encoding, and numeric features are subjected to scaling for uniformity. Pair plots and statistical analyses visually articulate the impact of these transformations on the dataset, setting the stage for implementing a diverse set of machine learning models. As we progress, the core focus of our research revolves around three pivotal research questions:

1. How can predictive analytics and machine learning contribute to the refinement and precision of direct marketing strategies in the banking sector?
2. What insights can be gleaned from the comprehensive exploration of client demographics, financial history, and interactions with previous marketing campaigns, and how can these insights inform the anticipation of bank term deposit subscriptions?
3. In what ways do the implemented machine learning models, including Logistic Regression, Random Forest, and Support Vector Machines, enhance the effectiveness of direct marketing campaigns, and how can their performance be quantified and compared using key metrics such as accuracy, precision, and F1 score?

After the introduction, our inquiry unfolds through a structured sequence of sections. We initiate with a meticulous literature review, offering crucial contextualization, followed by the proposed methodology. This phase encompasses intricate stages such as Data Collection, Exploratory Data Analysis (EDA), Feature Engineering, Preprocessing, and Model Implementation. The ensuing segments rigorously analyze

Figure 1. Integration of AI in bank term deposit subscriptions.

Results and Evaluation, leading to a reasonable synthesis in the conclusion. This systematic framework ensures a comprehensive exploration, rendering valuable contributions to the discourse on incorporating predictive analytics and machine learning in direct marketing for bank term deposit subscriptions.

2. LITERATURE REVIEW

Amidst the dynamic landscape of banking and financial institutions, the strategic use of advanced technologies, particularly machine learning and data mining techniques, has become essential for predicting and improving different operations. This comprehensive review, which comprises eight distinct studies, examines various banking subjects, including the stability of time deposits and the difficulty of telemarketing forecasts. Each research introduces innovative methodologies, algorithms, and models for industrial issues.

Because of their stability and low cost, time deposits are a good funding source for banks. Yan et al. [5] predict the success rate of fixed deposits in bank telephone marketing using the S_Kohonen network. This forecast is situational. Including an output layer improves the Kohonen network and turns it into a supervised learning S_Kohonen network. The Improved Whale Optimisation Algorithm (IWOA) optimizes input and WOA weights. Despite inherent volatility and low forecast accuracy, difficulties remain. A more complicated algorithm is constructed by combining the inertia weight and Levy flight pattern. The algorithm would have been different. S_Kohonen networks optimized using GA, WOA, and LWOA algorithms exhibit lower classification accuracy. According to empirical findings, this is caused by the improved representative capacity of the strengthened S_Kohonen network.

Ładyżyński et al. [6] used bank time-series data to predict personal loan acceptance inclination. A pivotal facet of banks and other financial institutions is the availability of credit products. The recommended testing procedure involves using a moving window. This window shows complicated, sequential, time-based dependencies within transactions while eliminating noise from moving-window dependencies. This research presents a system for identifying potential credit product customers that incorporates classification, random

forests, and deep neural networks. Empirical studies that provide promising results demonstrate the system's ability to identify critical trends in historical customer data and predict credit purchase likelihood. Testing is expanded beyond the banking sector, establishing a broad system for direct marketing campaigns. This plan was a success.

Feng et al. [7] introduce META-DES-AAP, a dynamic ensemble selection method created exclusively to predict the success of time deposit sales in bank telemarketing campaigns. It was designed for financial institutions. META-DES-AAP offers a unique perspective on sales prediction compared to traditional machine-learning-based methods concentrating only on accuracy. A multi-objective optimization algorithm is designed to improve accuracy and average profit while selecting base classifiers within the framework of dynamic ensemble selection using meta-training. Base classifiers are included via a dynamic mechanism customized for each telemarketing campaign. META-DES-AAP outperforms other state-of-the-art machine learning methods in accuracy and average profit. The empirical findings come from bank telemarketing data. The factors influencing telemarketing success and the average profit META-DES-AAP produces are also examined.

Telemarketing is still a prevalent avenue for banks to secure deposits because of its cheap cost and easy implementation. Xie et al. [8] construct bank telemarketing prediction models and identify the most effective one using three machine learning methods: Random Subspace, Multi-Boosting, and RS-MB. Interpretability analysis provides the information needed to create an impactful marketing plan. Machine learning is used to evaluate the importance of the original independent variables. Prediction models are then constructed using these factors. The most reliable predictor is the RS-MB method, which uses a subset of independent factors. The research emphasizes the importance of telemarketing and avoiding false negatives in predictions compared to false positives. The kind of work, connection month, and contact day of the week are crucial for telemarketing campaigns.

A sophisticated machine learning approach may address the marketing challenge in banking systems [9]. Classic models such as linear regression, logistic regression, decision tree, and KNN are favorable regarding ease of explanation but are not as effective. ANNs and other complex models are accurate but complex black-box models with low interpretability. This research suggests that banks use a simple AI-based technique to predict potential customers for term deposits. The research uses a dataset from Portuguese banks as a benchmark to achieve its target recall score of 0.99. The objective model evaluation employs the Recall score, FPR value, and AUC score. The research emphasizes predictability and interpretability while addressing practical issues in predicting and converting potential customers.

Extracting patterns from customer datasets using machine learning techniques is invaluable [10]. ML models' efficiency requires enhancement despite often providing good results. The pivotal feature-selection task in machine learning is realistically addressed by genetic algorithms, which are heuristic methods for simulating processes. To solve optimization problems, genetic algorithms, also known as evolutionary algorithms, use mutation methods, crossover methods, and algorithms. This research introduces a genetic-algorithm-based feature-selection approach to augment the efficiency of machine learning techniques for customer-related datasets. The proposed model produces promising results when applied to two consumer information datasets from the UCI repository. Applications employ Python's extensive packages to perform machine learning tasks throughout the implementation process.

Bank market prediction is a crucial domain in data mining research due to the abundance of data from several banking institutions. Despite the wealth of information accessible, extracting valuable insights is always a challenge. Ghosh et al. [11] analyze and apply vital data-mining techniques such as MLPNN, DT, and SVM. Specifically, MLPNN is a Multilayer Perceptron Neural Network. The objective of this dataset is to evaluate the performance of MLPNN, DT, and SVM in predicting success for bank deposit subscriptions. Accuracy, root-mean-square error, Precision, Recall, F-measure, and ROC Area values are used to evaluate the efficacy of these models. According to the trials' outcomes, these models are effective with higher accuracies.

The success of direct marketing hinges on customer responses; hence, this is a focal point of research. To identify target customers for banking campaigns, Miguéis et al. [12] provide a data-mining response model that uses random forests. The prevalent issue of class imbalance in telemarketing is addressed via a comparative exploration of undersampling methods (using the EasyEnsemble Algorithm) and oversampling methods (using the Synthetic Minority Oversampling Technique) in the banking context. This research contributes insights into the impact of attribute features on response-model performance. The inclusion of demographic, contact, and socioeconomic data enables these insights. Undersampling methods help random forests predict better than previously explored techniques, surpassing them in terms of performance.

This comprehensive review explores innovative methodologies and approaches to propel the banking sector toward enhanced decision-making and strategic planning. Each research contributes valuable insights by providing a nuanced understanding of how advanced technologies might be harnessed to address financial landscape concerns.

3. PROPOSED METHODOLOGY

3.1 Data Collection and Exploration

In the initial phase of our research methodology, we meticulously undertake the critical tasks of data collection and exploration. These foundational steps are pivotal in acquiring a profound understanding of the dataset's intricacies and characteristics. Our primary data source, Kaggle [13], is a renowned platform for its diverse datasets. In this context, it is the bedrock for our investigation into bank term deposit subscriptions.

Moving seamlessly into the exploration phase, we employ various analytical techniques to unravel patterns and insights inherent in the dataset. Visualization tools, including bar charts and counterplots, as shown in Figure 2, become instrumental in depicting the distribution of critical features. By scrutinizing the dataset's nuances, we gain a comprehensive overview, setting the stage for subsequent phases of our

research. Acknowledging Kaggle as our data source is not merely a formality; it underscores transparency and adherence to scholarly standards. Clearly stating our dataset's origin enhances our research's credibility, fostering an environment of trust and reproducibility within the academic community. This conscientious approach within the data collection and exploration phase exemplifies the rigor and scholarly integrity of our research endeavor.

3.2 Exploratory Data Analysis (EDA)

Advancing into the subsequent phase of our research, the focus now shifts to the intricate domain of Exploratory Data Analysis [14, 15]. A meticulous examination of the dataset is undertaken, with advanced techniques employed to reveal deeper layers of insight. The dataset, sourced from Kaggle, is the foundation for our analytical endeavors. The comprehensive exploration of data is conducted through a sophisticated array of methodologies. Crosstabulation (crosstab), as exemplified in Figure 3, is utilized as a pivotal technique. This systematic approach examines relationships between categorical variables, providing a structured framework for analyzing their joint distribution. The utilization of crosstabulation proves invaluable in unraveling the intricacies of variables that influence the dynamics of bank term deposit subscriptions. This technique systematically explores the complex landscape of dependencies, ensuring a nuanced comprehension of categorical intricacies within the dataset.

In addition to employing crosstabulation for a comprehensive exploration of data dynamics, Figure 4 introduces the strategic integration of heatmap visualization as an instrumental tool. This visually compelling technique serves as an invaluable asset in unraveling the intricate correlation dynamics that exist between various variables within our dataset. The utilization of color gradients on the heatmap provides an intuitive guide, allowing us to discern not only the strength but also the direction of correlations.

With its rich color spectrum, the heatmap serves as a visual aid, enhancing our analytical capacity by offering a nuanced representation of the relationships between different features. This visualization technique goes beyond the limitations of traditional statistical analyses, providing a more intuitive and insightful understanding of subtle patterns and dependencies inherent in the data. The incorporation of heatmaps into our EDA represents a commitment to a meticulous and sophisticated analytical approach, reinforcing our dedication to extracting meaningful insights from the dataset. This strategic combination of crosstabulation and heatmap visualization contributes to a holistic exploration of the dataset, ensuring a comprehensive understanding of its underlying patterns and intricacies.

EDA delves into bid behaviors, exploring correlations between bidding patterns and auction outcomes. By discerning how bid amounts, frequencies, and temporal factors interplay, we aim to uncover influential factors that shape successful auction results. This bid-behavior correlation is integral to identifying strategies that contribute to favorable outcomes for buyers and sellers; the EDA serves as a crucial bridge between data collection and subsequent modeling. It offers nuanced insights that pave the way for a comprehensive understanding of eBay auction dynamics, setting the stage for the subsequent phases of our research.

3.3 Feature Engineering and Preprocessing

Feature Engineering is a systematic process that unfolds and is integral to the preparatory stages of our predictive analytics and machine learning venture. This pivotal phase involves the deliberate and systematic refinement of the dataset, positioning it strategically for subsequent utilization by a diverse array of machine learning models. The cornerstone of this process is feature engineering, a sophisticated maneuver that involves the nuanced transformation and creation of variables. The primary aim is to distill valuable insights from the data, enhancing the overall performance of the machine learning models [16]. This tailored approach ensures that each feature within the dataset harmoniously aligns with the overarching goal of predicting bank term deposit subscriptions. A critical facet of feature engineering is the conversion of categorical variables into numerical formats, a prerequisite for their seamless integration into machine learning algorithms.

Simultaneously, preprocessing techniques come into play, standardizing and normalizing numerical features. This intervention is instrumental in establishing uniformity across the dataset, effectively mitigating biases that could arise from variations in scale. The careful standardization ensures that each feature contributes proportionally to the model, preventing the undue influence of variables merely based on their scale. To provide a comprehensive understanding of this transformative journey, the impact of these measures is meticulously visualized through pair plots and subjected to statistical analyses, presenting a thorough portrait of the dataset's evolution [17]. The orchestrated interplay of feature engineering and preprocessing emerges as a crucial prelude to implementing machine learning models, substantially enhancing their efficacy in predicting bank term deposit subscriptions. As this phase is navigated, the symbiotic relationship between feature engineering and preprocessing becomes the bedrock for a dataset meticulously sculpted for optimal predictive modeling. This approach aligns seamlessly with the exacting standards inherent in rigorous academic research, underscoring the precision and thoroughness characterizing our methodology [18, 19, 20].

3.4 Model Implementation

In model implementation, the focus of our analytical exploration shifts toward utilizing diverse machine learning models, each employed individually without resorting to ensemble techniques. This meticulous approach permits a detailed examination of each model's intrinsic capabilities and intricacies. The suite of models enlisted for this empirical investigation includes the following:

1. The **Stochastic Gradient Descent (SGD) Classifier**, renowned for its efficiency and scalability, is harnessed to navigate the intricate landscape of predicting bank term deposit subscriptions. Its iterative optimization process and adaptability to large datasets position it as a potent contender within our model repertoire.

2. The **K-Nearest Neighbors (KNN) Classifier**, a non-parametric and instance-based algorithm, is deployed to discern patterns in the data. By leveraging the proximity

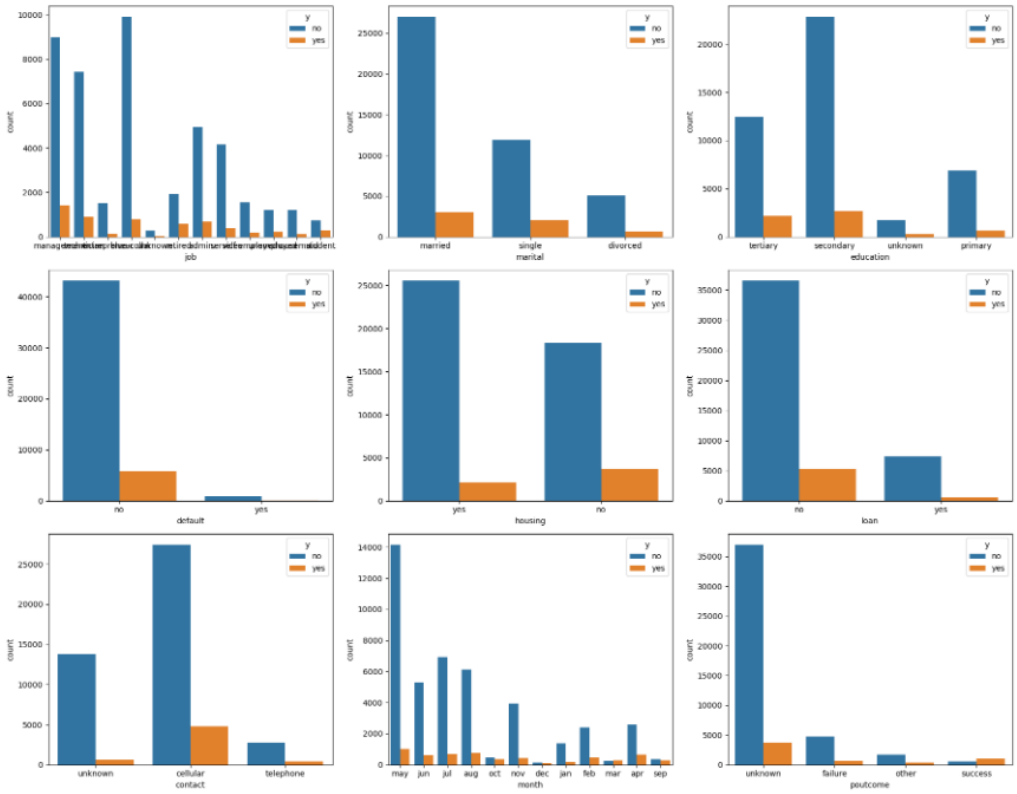

Figure 2. Dataset exploration.

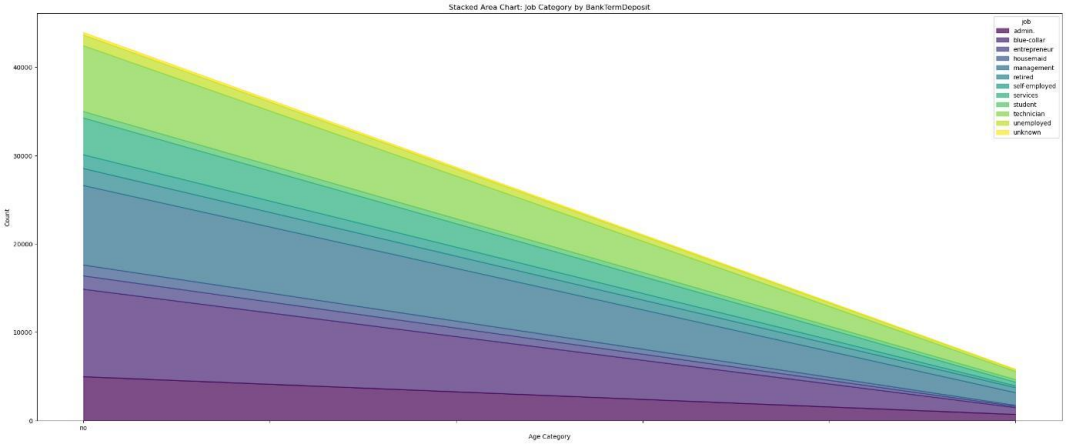

Figure 3. Crosstab as a visualization technique.

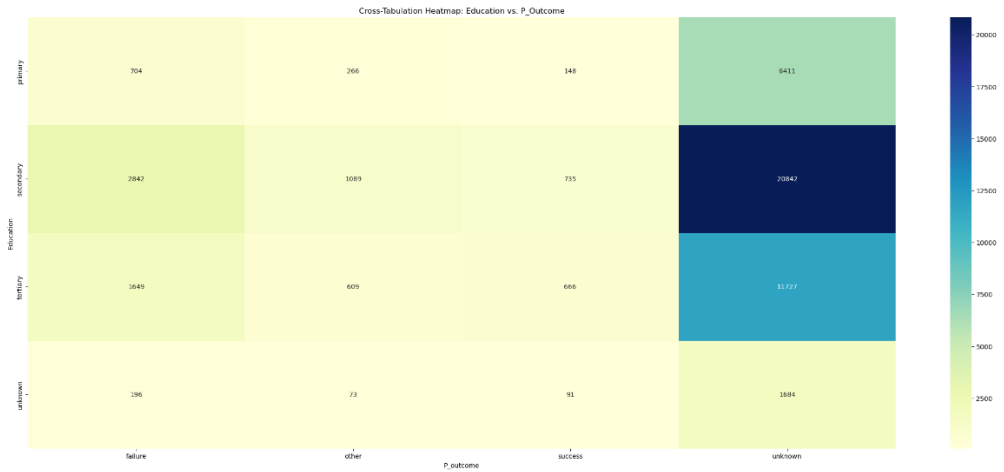

Figure 4. Heat map as a visualization technique.

of data points in the feature space, KNN offers an intuitive approach to predicting outcomes, making it a valuable asset in our predictive modeling ensemble.

3. **Logistic Regression**, a stalwart in binary classification scenarios, takes center stage in our model suite. Renowned for its interpretability and simplicity, Logistic Regression is a benchmark for assessing the predictive prowess of more complex models in our research.

4. The **Gaussian Naive Bayes (GNB) Classifier**, rooted in Bayesian probability, contributes its probabilistic framework to our predictive analytics arsenal. Particularly adept at handling numerical and categorical features, GNB provides a probabilistic foundation for predicting bank term deposit subscriptions.

5. The **Decision Tree Classifier**, characterized by its tree-like structure, is employed to decipher the intricate decision-making process within our dataset. Decision trees offer transparency in model decision paths, facilitating a deeper understanding of the underlying mechanisms influencing predictions.

6. The **Random Forest Classifier**, an ensemble of decision trees, adds a layer of sophistication to our model spectrum. By aggregating the insights from multiple decision trees, Random Forest enhances predictive accuracy and robustness, contributing a versatile dimension to our analytical toolkit.

Each classifier undergoes rigorous training on the data. The subsequent evaluation phase meticulously scrutinizes their performance using key metrics such as accuracy, precision, and F1 score. This discerning approach not only delineates the individual strengths and weaknesses of each model but also lays the groundwork for a nuanced comparative analysis essential for deriving substantive insights.

4. RESULTS AND EVALUATION

As elucidated in Table 1, each model's classification results yield valuable insights into their respective performances. The P-value, a statistical measure, indicates the significance of the model's predictions. This measure aids in understanding the reliability of the observed results, helping to discern whether the projections are statistically meaningful. The Positive Predictive Value (PPV) is a crucial metric that denotes the proportion of accurate optimistic predictions out of the total optimistic predictions. It measures the model's accuracy in correctly identifying instances of bank term deposit subscriptions. A higher PPV signifies a more reliable prediction of positive outcomes, which is vital for the effectiveness of direct marketing campaigns.

Conversely, the Negative Predictive Value (NPV), also depicted in Table 1, represents the model's accuracy in identifying instances where bank term deposit subscriptions are not likely to occur. A higher NPV indicates a better ability to correctly identify negatives, contributing to the overall precision of the model. The F1-Score, a composite metric of precision and recall, provides a balanced assessment of the model's overall performance. It is beneficial when there is an uneven class distribution. A higher F1-Score suggests a model maintaining a harmonious equilibrium between precision and recall, striking an optimal balance in predicting positive and negative outcomes.

Accuracy, as a fundamental metric, gauges the correctness of the model's predictions. It considers both accurate optimistic and accurate pessimistic forecasts about the total predictions. A higher accuracy score signifies a more reliable and effective predictive model. Sensitivity, also known as Recall or True Positive Rate (TPR), measures the ability of the model to identify positive instances out of all actual positives correctly. It is essential in scenarios where the cost of false negatives is high, such as direct marketing for bank term deposit subscriptions. The Specificity metric, also showcased in Table 1, evaluates the model's ability to correctly identify negative instances out of all negatives. It is crucial in situations where minimizing false positives is essential, ensuring that resources are not spent on individuals unlikely to subscribe to bank term deposits.

Table 2 presents the detailed criteria used for evaluating the classification results. Accuracy, Sensitivity (Recall), Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), and F1 Score collectively provide a nuanced and comprehensive analysis of the models' strengths and areas for improvement.

The analysis reveals that the Random Forest Classifier exhibited the best performance among the models evaluated. This highlights its effectiveness in anticipating bank term deposit subscriptions within direct marketing campaigns. These insights serve as valuable guides for refining and optimizing future direct marketing strategies within the banking sector.

5. CONCLUSION

The research's culmination emphasizes the transformative impact of predictive analytics and machine learning on the direct marketing landscape within the banking sector. Integrating sophisticated analytical models has been a game-changer, offering novel avenues for anticipating bank term deposit subscriptions. This conclusion is drawn from the comprehensive exploration of data, rigorous exploratory data analysis (EDA), meticulous feature engineering, and the reasonable implementation of machine learning models.

The intricate relationships within the dataset were unraveled throughout this investigation, providing a foundation for informed decision-making in direct marketing. The study delved into feature-engineering and preprocessing complexities, ensuring the data was finely tuned to extract maximum predictive power from the machine learning models. A suite of models, including the SGD Classifier, k-nearest neighbor Classifier, Logistic Regression, Gaussian NB, Decision Tree Classifier, and Random Forest Classifier, were applied in the realm of model implementation. Each model operated independently, avoiding ensemble techniques to evaluate individual model performance.

The results and evaluation phase furnished a rich tapestry of insights, providing a nuanced understanding of the strengths and weaknesses of each model. The Random Forest Classifier emerged as the frontrunner, showcasing superior predictive power in identifying bank term deposit subscriptions. The criteria for evaluating the models served as a comprehensive

Table 1. Classification result.

Models	P-value PPV	P-value NPV	F1-Score	Accuracy	Sensitivity (TPR)	Specificity (TNP)
SGDClassifier	0.868421	0.540984	0.177419	0.440	0.392857	0.687500
KNeighborsClassifier	0.844560	0.903047	0.285714	0.825	0.970238	0.062500
LogisticRegression	0.852041	0.917582	0.750000	0.850	0.994048	0.093750
GaussianNB	0.896552	0.912281	0.538462	0.850	0.928571	0.437500
DecisionTreeClassifier	0.892655	0.915942	0.565217	0.855	0.940476	0.406250
RandomForestClassifier	0.878307	0.929972	0.818182	0.875	0.988095	0.281250

Table 2. Criteria for evaluating classification results.

Evaluation Criterion	Value
Accuracy	$\frac{\text{Number of Correct Predictions}}{\text{Total Predictions}}$
Sensitivity (Recall)	$\frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$
Specificity	$\frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$
Positive Predictive Value	$\frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$
Negative Predictive Value	$\frac{\text{True Negatives}}{\text{True Negatives} + \text{False Negatives}}$
F1 Score	$\frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}$

yardstick for assessing their efficacy. Accuracy, Sensitivity, Specificity, PPV, NPV, and F1 Score collectively depicted the models' performance. These metrics contribute to a retrospective evaluation and lay the groundwork for future refinements and optimizations in direct marketing strategies.

As we progress, the core focus of our study revolves around the broader narrative of leveraging advanced analytics for strategic decision-making in the banking industry. The findings of this study not only illuminate the current landscape but also serve as a beacon guiding future endeavors in the dynamic intersection of banking, direct marketing, and data-driven insights. In a rapidly evolving financial landscape, the fusion of technology and analytics becomes imperative for staying ahead of the curve and delivering value to financial institutions and their clientele.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

[1] S. Moro, P. Cortez, and P. Rita, "Using customer lifetime value and neural networks to improve the prediction of bank deposit subscription in telemarketing campaigns," *Neural Computing and Applications*, vol. 26, no. 1, pp. 131–139, 2015.

[2] D. Khafaga, A. Alhussan, E.-S. M. El-Kenawy, A. E. Takieldeen, T. M. Hassan, E. Hegazy, E. Abdel, A. Ibrahim, and A. A. Abdelhamid, "Meta-heuristics for feature selection and classification in diagnostic breast cancer," *Computers, Materials & Continua*, vol. 73, no. 1, pp. 749–765, 2022.

[3] S. Hosseini, "A decision support system based on machine learned Bayesian network for predicting successful direct sales marketing," *Journal of Management Analytics*, vol. 8, no. 2, pp. 295–315, 2021.

[4] E. M. Hassib, A. E.-S. I. El-Desouky, E.-S. M. El-Kenawy, and S. M. El-Ghamrawy, "An imbalanced big data mining framework for improving optimization algorithms performance," *IEEE Access*, vol. 7, pp. 170 774–170 795, 2019.

[5] C. Yan, M. Li, and W. Liu, "Prediction of bank telephone marketing results based on improved whale algorithms optimizing S_Kohonen network," *Applied Soft Computing*, vol. 92, p. 106259, 2020.

[6] P. Ładyżyński, K. Żbikowski, and P. Gawrysiak, "Direct marketing campaigns in retail banking with the use of deep learning and random forests," *Expert Systems with Applications*, vol. 134, pp. 28–35, 2019.

[7] Y. Feng, Y. Yin, D. Wang, and L. Dhamotharan, "A dynamic ensemble selection method for bank telemarketing sales prediction," *Journal of Business Research*, vol. 139, pp. 368–382, 2022.

[8] C. Xie, J.-L. Zhang, Y. Zhu, B. Xiong, and G.-J. Wang, "How to improve the success of bank telemarketing? prediction and interpretability analysis based on machine learning," *Computers & Industrial Engineering*, vol. 175, p. 108874, 2023.

[9] M. Z. Khan, S. Munquad, and T. S. M. Rao, "A study on improving banking process for predicting prospective customers of term deposits using explainable machine learning models," in *Proceedings of International Conference on Computational Science and Applications*, S. Bhalla, M. Bedekar, R. Phalnikar, and S. Sirsikar, Eds. Springer Nature, 2022, pp. 93–103.

[10] B. Ajay Ram, D. J. Santosh Kumar, and A. Lakshmanarao, "Improving efficiency of machine learning model for bank customer data using genetic algorithm approach," in *International Conference on Innovative Computing and Communications*, A. Khanna, D. Gupta, S. Bhattacharyya, A. E. Hassanien, S. Anand, and A. Jaiswal, Eds. Springer, 2022, pp. 649–657.

[11] S. Ghosh, A. Hazra, B. Choudhury, P. Biswas, and A. Nag, "A comparative study to the bank market prediction," in *Machine Learning and Data Mining in Pattern Recognition*, P. Perner, Ed. Springer International Publishing, 2018, pp. 259–268.

[12] V. L. Miguéis, A. S. Camanho, and J. Borges, "Predicting direct marketing response in banking: Comparison of class imbalance methods," *Service Business*, vol. 11, no. 4, pp. 831–849, 2017.

[13] The Devastator, "Bank term deposit predictions," Kaggle dataset, 2023, retrieved December 14, 2023.

[14] A. S. Arora, H. Rajput, and R. Changotra, "Current perspective of COVID-19 spread across south korea: Exploratory data analysis and containment of the pandemic," *Environment, Development and Sustainability*, vol. 23, no. 5, pp. 6553–6563, 2021.

[15] M. M. Eid, E.-S. M. El-Kenawy, N. Khodadadi, S. Mirjalili, E. Khodadadi, M. Abotaleb, A. H. Alharbi, A. A. Abdelhamid, A. Ibrahim, G. M. Amer, A. Kadi, and D. S. Khafaga, "Meta-heuristic optimization of LSTM-based deep network for boosting the prediction of monkeypox cases," *Mathematics*, vol. 10, no. 20, p. 3845, 2022.

[16] A. M. U. D. Khanday, S. T. Rabani, Q. R. Khan, N. Rouf, and M. Mohi Ud Din, "Machine learning based approaches for detecting COVID-19 using clinical text data," *International Journal of Information Technology*, vol. 12, no. 3, pp. 731–739, 2020.

[17] A. Correa Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Feature engineering strategies for credit card fraud detection," *Expert Systems with Applications*, vol. 51, pp. 134–142, 2016.

[18] N. Samee, E.-S. M. El-Kenawy, G. Atteia, M. Jamjoom, A. Ibrahim, A. A. Abdelhamid, N. El-Attar, T. Gaber, A. Slowik, and M. Shams, "Metaheuristic optimization through deep learning classification of COVID-19 in chest X-ray images," *Computers, Materials & Continua*, vol. 73, no. 2, pp. 4193–4210, 2022.

[19] T. Verdonck, B. Baesens, M. Óskarsdóttir, and S. vanden Broucke, "Special issue on feature engineering editorial," *Machine Learning*, 2021.

[20] V. Cerqueira, N. Moniz, and C. Soares, "VEST: Automatic feature engineering for forecasting," *Machine Learning*, 2021.