FOREX Publication
Open Access | Rapid and quality publishing

International Journal of Computer Science Research (IJCSR)
Survey Article | Volume 2, Issue 4 | Pages 118-122 | ISSN XXXX-XXXX

Data Mining Solutions towards Bank Telemarketing Predicting

Ahmed Adeeb Jalal[1*], Wasseem N. Ibrahem Al-Obaydy[2] and Abdulhakeem Qusay Albayati[3]

[1]*Computer Engineering Department, College of Engineering, Al-Iraqia University, Baghdad, Iraq, ahmedadeeb@aliraqia.edu.iq*
[2]*Computer Engineering Department, College of Engineering, Al-Iraqia University, Baghdad, Iraq, wasseem.nahi@aliraqia.edu.iq*
[3]*Department of Computer Engineering, University of Technology-Iraq, Baghdad, Iraq, Abdulhakeem.Q.Ali@uotechnology.edu.iq*

***Correspondence:** *ahmedadeeb@aliraqia.edu.iq*

ABSTRACT- Lending in international markets has become more restricted, and the focus of funding has shifted to domestic customers and their deposits. This desire has created a demand for knowledge about customer deposit behavior, especially customer reactions to telemarketing campaigns. Therefore, promotional transaction banking strategies shifted from traditional methods to advanced data analytics and machine learning techniques to keep pace with the growth of the banking industry. In a telemarketing business, the ability to select potential customers for purchase is important, because it reduces processing time and operational costs. This paper presents a data mining solution using machine learning techniques to predict customer response to a bank's telemarketing campaign through clustering similar groups by the data analysis. Whereas, the dataset groups are pre-processed and using three types of algorithms: k-nearest neighbor (KNN), support vector machines (SVM), and decision trees (DTs), for modeling. The tests were conducted on a real dataset that resembles the actual information associated with the bank client, telemarketing calls, and the results of the customer's bank time deposits as a result of the call. The experimental results show that DTs has superior performance in some settings with a true positive rate of over 90%, outperforming their equivalent methods.

Keywords: Data mining; Classification; Clustering; Machine Learning; Telemarketing.

ARTICLE INFORMATION

Author(s): Ahmed Adeeb Jalal, Wasseem N. Ibrahem Al-Obaydy and Abdulhakeem Qusay Albayati;

Received: 17/11/2023; **Accepted**: 06/12/2023; **Published**: 30/12/2023;
e-ISSN: XXXX-XXXX;
Paper Id: IJCSR-020404;
Citation: 10.37391/IJCSR.020404

Publisher's Note: FOREX Publication stays neutral with regard to Jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

Banks have traditionally considered that telemarketing to be the most effective and cheapest marketing strategy for selling bank deposits. However, the drawbacks of traditional telemarketing may be reflected in the inconvenience and waste of bank resources for many customers by access the telephone calling. Early prediction of which customers are likely to purchase bank deposits is a key step in achieving accurate marketing and increasing the success rate of banks' telemarketing efforts [1], [2]. Additionally, choosing the most recommended clients for your telemarketing efforts is essential to reduce processing time and operational costs. However, customers may resent receiving telemarketing if your product or service does not meet their interests or needs. Therefore, being able to select the most likely customers is an urgent requirement [3], [4]. Therefore, accurate prediction is beneficial for both the customer and the company.

Data mining is the key to extracting deep knowledge about problems where large amounts of information exist. It provides a systematic approach to obtain valuable information against the information explosion in modern industries [5], [6]. There are many goals that companies set for decision support making, and data mining can be used to achieve these goals. Classification is one of the most common data mining operations that employed in marketing to predict customer behavior [7], [8].

Decision-making techniques play an important role in responding to the rapid developments facing challenging business environments [9]. Specifically, customer targeting based on identifying potential buyers of a product or subscribers of a service offering in the context of a marketing campaign in order to better identify and predict reality [10]. This allows us to conclude that for bank customers interested in fixed deposits, it is possible to predict customer behavior with high predictability [11]. By integrating classification models into decision support systems, decision makers can identify which data mining models are most useful for business improvement [12], [13]. Therefore, due to the large number of datasets available within enterprises, data mining has become a useful method for searching and discovering information that managers can use to make the decisions are better [14], [15].

Banks are currently investing heavily in marketing campaigns, so related studies are widely known in the scientific literature. Moro et al. [16] proposed a personalized intelligent decision support system that can automatically predict the outcome of long-term deposit sales calls using a data mining approach. This system helps managers prioritize and select the next customers to contact during the bank's marketing campaigns. They compared a total of four data mining models (neural networks,

Survey Article | Volume 2, Issue 4 | Pages 118-122 | ISSN XXXX-XXXX

logistic regression, decision trees, and support vector machines) and created a model that is robust and valuable to telemarketing campaign managers.

Kougétio and Toulni [17] presented a method to improve the prediction of telemarketing using an improved KNN model. This method aimed to selling long-term bank deposits in smart cities. It also aims to speed up the proposed KNN model based on preprocessing, filtering and regularization of important features. Therefore, the contributions can be summarized as follows. (1) We introduce a high-throughput preprocessing method for each feature type separately to impute missing feature values; feature normalization automates (2) the selection of the most realistic and important features known before recall, and (3) the combination of the optimal K model with a similar model that provides the best performance.

Unbalanced datasets pose special challenges in data mining modeling, as most algorithms are unable to capture properties of classes not represented in the dataset [18], [19]. This issue is especially important for online marketing campaigns. In online marketing campaigns, this occurs when the number of customers in these databases who purchase products or services is lower than the number of customers who do not. Rogi´c et al. [20] proposed a method to create a balanced SVM data preprocessor (B-SVM) based on a classifier that combines random classification and SVM classification to be applied to unbalanced datasets. The B-SVM approach effectively processes datasets to correct for noise and imbalance within classes. Therefore, companies can use this approach to predict which customers are most likely to respond effectively to their campaigns.

Safarkhani and Moro [12] developed the accuracy of classification in order to predict which of customers will sign up for long-term deposits that offered by the bank. Therefore, they focus on combining resampling to reduce scattered and unbalanced data, reducing the complexity of data computation using feature selection, and reducing the dimensionality of inefficient data modeling. The implemented process improved the performance of the classification accuracy in selecting potential customers. Additionally, other data mining and machine learning algorithms such as [21], [22], [23] have been used to predict bank telemarketing sales.

The objectives of this article is to predict the success of banks' telemarketing activities and propose a heuristic map that helps telemarketers to selecting potential customers by applying algorithmic weighting methods. Additionally, this paper aims to present a framework that combines clustering and classification for bank telemarketing prediction. The rest of this document is organized as follows. *Section 2* introduces the methodology by describing the dataset, its assembly, and subsequent processing for applying machine learning algorithms. *Section 3* presents experimental results of the presented approach to dataset clustering and classification. Finally, the conclusions are reviewed in *Section 4.*

2. MATERIALS AND METHODS

In this section, we introduce our proposed approach, which consists of three stages: clustering, preprocessing, and classification. Preprocessing reduces processing time and improves classification performance by converting different types of features into specific formats and processing each feature individually, while maintaining high robustness to performance variations.

2.1 Dataset

The direct marketing dataset was collected during a banking institution's marketing campaign that based on phone calls. This database consists of information about clients participating in this promotion. This dataset is also widely used for predictive modeling in direct marketing to determine and rank which clients are respond positively for marketing campaigns. We use a direct marketing dataset that containing 21 attributes of 41,188 Portuguese bank customers who were contacted by phone between 2008 and 2013 and offered to open long-term savings accounts at attractive interest rates [14], [24]. The dataset consist of 20 input variables and 1 output variable as target. Additionally, the information about the categories of the input variables is unbalanced, with less than 12% receiving a “yes” rating and the rest receiving a “no” rating [14].

When we seek to understand customer behavior [25], we first study the history of customer interactions with a company, attributes, metrics, ratings, and marketing feedback. The dataset includes attributes such as age, marital status, current loan status, education level, employment, job, and average annual balance, as well as class labels that indicate whether a customer has been accepted. The next step was to categorize the clients into groups. Therefore, we aim to classify the dataset into four groups based on marital status attributes. The process of data clustering divides dataset into groups or categories based on data similarity. As shown in *figure 1*, the 41,188 clients in the dataset were distributed into four clusters according to marital status attributes, which are: married, single, divorced, and unknown. Each of these categories is classified according to the type of data. Data aggregation allows businesses to discover new groups within their customer database. It is also possible to classify the existing customer base into distinct groups based on attribute patterns [26], [27].

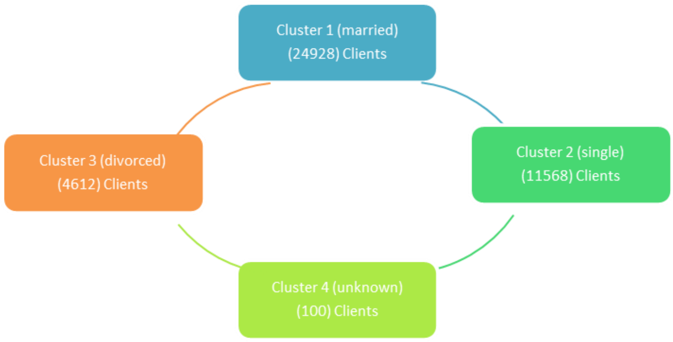

Figure 1. Clients’ clustering according to the marital status attributes

2.2. Data Preprocessing

Survey Article | Volume 2, Issue 4 | Pages 118-122 | ISSN XXXX-XXXX

The number of clients' instance that records in the dataset contains a total of 41,188 customer records, of which 4,640 customers successfully subscribed for bank term deposit, representing 11.27% of the total. It turns out that traditional telemarketing clients have very low response rates, which leads to most of being intrusion on them. Data preprocessing is one of the very important steps in the predictive modeling as performance closely depends on it [28]. Here, we propose a special preprocessing based on data clustering followed by replacing all feature information with average frequency values. This function calculates the average value of all values Vij of the feature Vj if they are scaling or numeric variables, or if they are Boolean or nominal variables in a given dataset, as in (1). In this phase, data is prepared according to data mining models [29], [30].

$$Mean\ (V_{ij}) \leftarrow \frac{fV_{j,N}}{N_i} \qquad (1)$$

Where, fV_j, N denotes to the frequency of a feature V of variable j in the dataset N, and N_i denotse to the total number of classes i in the dataset.

3. RESULTS AND DISCUSSIONS

This study aims to improve the sale of long-term deposits promoted via telephone through banks' telemarketing campaigns in order to predict the likely outcome for each client as well as the various contacts. In this section, we review the results obtained by applying the classifications method to the banking marketing dataset. *Table 1* shows the predictive performance of all tested classifiers: KNN, SVM, and DT, before clustering and data preprocessing.

Table 1. Predictive performance of classification algorithms without clustering and data preprocessing.

Classifiers		KNN	SVM	DT
Testing dataset	10%	88.27	88.88	55.96
	20%	87.97	88.72	55.91
	30%	87.51	88.6	55.8
	40%	87.15	88.59	55.66
	50%	86.76	88.56	55.96
	60%	86.75	88.54	55.74
	70%	86.63	88.32	69.95
	80%	86.37	87.98	69.89
	90%	86.38	87.72	69.99

Each classifier was tested on the original training dataset and applied to the test dataset. *Table 1* show the results obtained on the test dataset. For example, the scores of SVM and KNN were less than 90% in all test samples, and DT was also less than 70%, giving unsatisfactory results. The accuracy of all independent models is high because the models are biased toward the positive class. Therefore, considering the importance of accurately identifying customers who will respond to direct marketing campaigns, it will lead to satisfactory results. This work focuses on clustering and data preprocessing. *Table 2* shows the predictive performance of all tested classifiers: KNN, SVM, and DT, after dataset clustering and data preprocessing. We also note that the dataset is divided into four groups based on marital status attributes, where all cases available in the dataset were tested to find out which of them are most worthy for sampling.

Table 2. Predictive performance of classification algorithms with clustering and data preprocessing.

		Cluster 1			Cluster 2			Cluster 3			Cluster 4		
Classifiers		KNN	SVM	DT	KNN	SVM	DT	KNN	SVM	DT	KNN	SVM	DT
Testing dataset	10%	88.95	90.2	90.12	87.85	89.37	89.37	78.67	85.6	85.24	71.43	57.15	99
	20%	88.31	89.7	89.8	87.8	88.67	89.2	78.54	86.08	86.17	61.91	61.91	95.24
	30%	88.34	89.56	89.69	88.24	89.87	90.23	76.58	85.73	86	55.17	55.17	89.66
	40%	88.42	89.76	89.93	86.74	89.56	90.1	76	85.39	85.64	51.28	51.28	84.62
	50%	87.79	89.64	89.77	85.69	89.62	90.18	79.39	85.54	85.99	39.58	50	83.33
	60%	87.9	89.71	89.78	85.94	88.89	89.67	74.27	84.55	86.15	29.83	19.3	84.2
	70%	87.58	89.26	89.87	84.78	88.84	89.55	74.34	84.35	85.98	18.84	21.74	85.51
	80%	86.35	87.79	89.85	84.41	88.22	89.6	72.91	85.29	86	14.29	15.71	85.71
	90%	83.42	86.83	89.85	82.44	88.29	89.71	72.25	82.73	85.96	14.08	16.9	85.92

The performance of the proposed approach was evaluated using three common validation metrics, as shown in *table 2*. We can see that DT achieved the best performance in over 90% in most clusters. This result indicates that future campaigns may be more profitable because the company can precisely target customer groups that are more likely to respond. His SVM results of around 90% were also achieved on the first and his second set of the same test samples. Additionally, speed and time factors need to be considered to obtain the effectiveness of the proposed approach. The partitioned dataset makes it easier

Survey Article | Volume 2, Issue 4 | Pages 118-122 | ISSN XXXX-XXXX

to sort customers and compare them to similar categories, and also speeds up the process of displaying results. Moreover, when we calculate the mean scores of all tested samples, the DT classifier shows better performance for all clusters compared to the non-clustering dataset, as shown in *figure 2*.

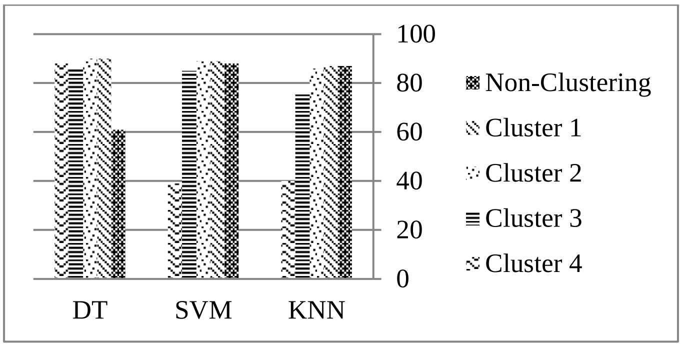

Figure 2. Distribution of samples test score averages for clustering and non-clustering datasets

4. CONCLUSIONS

The result of this research is a classifier that can predict whether a customer will subscribe for a long-term deposit. A combination of resampling and feature selection was applied to 10-90% of the entire dataset (41,188). The proposed method deals with clustering and data preprocessing. Three classifiers: KNN, SVM, and DT, were computed to evaluate the performance of the classifier. Comparing the metrics, it can be proven that clustering and data preprocessing increased the accuracy value of DT classifier over 90%. The purpose of this study was to divide the data into groups and present results to predict the success of selling long-term deposits through telemarketing.

REFERENCES

[1] Che, J., Zhao, S., Li, Y., and Li, K., (2020). Bank Telemarketing Forecasting Model Based on t-SNE-SVM. Journal of Service Science and Management, 13(3), 435-448. https://doi.org/10.4236/jssm.2020.133029.

[2] Sagala, N. T. M., and Permai, S. D., (2021). Predictive Model using SVM to Improve the Effectiveness of Direct Marketing. 2021 2nd International Conference on Artificial Intelligence and Data Sciences (AiDAS), 1-6. https://doi.org/10.1109/AiDAS53897.2021.9574440.

[3] Prompreing, K., and Prompreing, T., (2021). A Telemarketing Guidance in Selling Banking Services: A Data Mining Approach. Indonesian Journal of Business Analytics, 1(1), 1-16. https://doi.org/10.55927/ijba.v1i1.1.

[4] Saeed, S. E., Hammad, M., and Alqaddoumi, A., (2022). Predicting Customer's Subscription Response to Bank Telemarketing Campaign Based on Machine learning Algorithms. 2022 International Conference on Decision Aid Sciences and Applications (DASA), 1474-1478. https://doi.org/10.1109/DASA54658.2022.9765152.

[5] Liu, C., Fakharizadi, E., Xu, T., and Yu, P. S., (2023). Recent advances in domain-driven data mining. International Journal of Data Science and Analytics, 15, 1–7. https://doi.org/10.1007/s41060-022-00378-1.

[6] Chang, Q., and Hu, J., (2022). Research and Application of the Data Mining Technology in Economic Intelligence System. Computational Intelligence and Neuroscience, 2022, 1–11. https://doi.org/10.1155/2022/6439315.

[7] Tope-Oke, A., Afolalu, C. A., and Omofade, O., (2019). A Data Mining Based Approach to Customer Behaviour in an Electronic Settings. Journal of Computer and Communications, 7(5), 42–53. https://doi.org/10.4236/jcc.2019.75004.

[8] Hassan, D., Rodan, A., Salem, M., and Mohammad, M., (2019). Comparative Study of using Data Mining Techniques for Bank Telemarketing Data. 2019 Sixth HCT Information Technology Trends (ITT), 177-181. https://doi.org/10.1109/ITT48889.2019.9075123.

[9] Ghatasheh, N., et al., (2020). Business Analytics in Telemarketing: Cost-Sensitive Analysis of Bank Campaigns Using Artificial Neural Networks. Applied Sciences, 10(7), 1–15. https://doi.org/10.3390/app10072581.

[10] Barraza, N., Moro, S., Ferreyra, M., and De La Peña, A., (2019). Mutual information and sensitivity analysis for feature selection in customer targeting: A comparative study. Journal of Information Science, 45(1), 53-67. https://doi.org/10.1177/0165551518770967.

[11] Amzile, K., and Rajaa, A., (2021). Using SVM for Smart Direct Marketing (SDM): A case of predicting bank customers interested in the Term Deposits. International Journal of Accounting, Finance, Auditing, Management and Economics, 2(5), 525-537. https://doi.org/10.5281/zenodo.5527861.

[12] Safarkhani, F., and Moro, S., (2021). Improving the Accuracy of Predicting Bank Depositor's Behavior Using a Decision Tree. Applied Sciences, 11(19), 1–113. https://doi.org/10.3390/app11199016.

[13] Farooqi, Md. R., and Iqbal, N., (2019). Performance Evaluation for Competency of Bank Telemarketing Prediction using Data Mining Techniques. International Journal of Recent Technology and Engineering, 8(2), 5766-5774. https://doi.org/10.35940/ijrte.A1269.078219.

[14] Vitório, A., and Marques, G., (2021). Impact of Imbalanced Data on Bank Telemarketing Calls Outcome Forecasting using Machine Learning. 2021 International Conference on Data Analytics for Business and Industry (ICDABI), 380-384. https://doi.org/10.1109/ICDABI53623.2021.9655824.

[15] Khairi, T. W. A., Zaki, R. M., and Mahmood, W. A., (2019). Stock Price Prediction using Technical, Fundamental and News based Approach. Proceedings of the 2019 2nd Scientific Conference of Computer Sciences (SCCS), 177-181. https://doi.org/10.1109/SCCS.2019.8852599.

[16] Moro, S., Cortez, P., and Rita, P., (2014). A data-driven approach to predict the success of bank telemarketing. Decision Support Systems, 62, 22-31. https://doi.org/10.1016/j.dss.2014.03.001.

[17] Koumétio, S. C. T., and Toulni, H., (2021). Improving KNN Model for Direct Marketing Prediction in Smart Cities. Machine Intelligence and Data Analytics for Sustainable Future Smart Cities. Studies in Computational Intelligence, 971. https://doi.org/10.1007/978-3-030-72065-0_7.

[18] Marinakos, G., and Daskalaki, S., (2017). Imbalanced customer classification for bank direct marketing. Journal of Marketing Analytics, 5, 14-30. https://doi.org/10.1057/s41270-017-0013-7.

[19] Hadi, R. M., Al-khalisy, S. H. J., and Hamza, N. A., (2019). Prediction Model for Financial Distress Using Proposed Data Mining Approach. Journal of Al-Qadisiyah for Computer Science and Mathematics, 11(2), 37-44. https://doi.org/10.29304/jqcm.2019.11.2.570.

[20] Rogi´c, S., Kaš´celan, L., and Bach, M. P., (2022). Customer Response Model in Direct Marketing: Solving the Problem of Unbalanced Dataset with a Balanced Support Vector Machine. Journal of Theoretical and Applied Electronic Commerce Research, 17(3). https://doi.org/10.3390/jtaer17030051.

[21] Tékouabou, S. C. K., et al., (2022). A Machine Learning Framework towards Bank Telemarketing Prediction. Journal of Risk and Financial Management, 15(6), 1-19. https://doi.org/10.3390/jrfm15060269.

[22] Fawei, T., and Ludera, D. T. J., (2022). Data Mining Solutions for Direct Marketing Campaign. Intelligent Systems and Applications. IntelliSys 2020. Advances in Intelligent Systems and Computing, 1251, 1-15. https://doi.org/10.1007/978-3-030-55187-2_46.

International Journal of
Computer Science Research (IJCSR)
Survey Article | Volume 2, Issue 4 | Pages 118-122 | ISSN XXXX-XXXX

[23] Parlar, T., and Acaravci, S. K., (2017). Using Data Mining Techniques for Detecting the Important Features of the Bank Direct Marketing Data. International Journal of Economics and Financial Issues, 7(2), 692-696.

[24] Selma, M., (2020). Predicting the Success of Bank Telemarketing Using Artificial Neural Network. World Academy of Science, Engineering and Technology. International Journal of Economics and Management Engineering, 14(1), 1-4. https://doi.org/10.5281/zenodo.3607876.

[25] Mohammed, Z. N., and Kadhem, S. M., (2021). A Study about E-Commerce Based on Customer Behaviors. Engineering and Technology Journal, 39(7), 1060-1068. https://doi.org/10.30684/etj.v39i7.1631.

[26] Atre, O., Modhave, S., Torane, P., and Jadhav, S. B., (2022). Customer Segmentation for Banking Strategy using Machine Learning. International Journal of Creative Research Thoughts, 10(12), 428-433.

[27] Raiter, O., (2021). Segmentation of Bank Consumers for Artificial Intelligence Marketing. International Journal of Contemporary Financial Issues, 1(1), 39-54. https://doi.org/10.17613/q0h8-m266.

[28] Alsolami, F. J., Saleem, F., and AL-Ghamdi, A. A., (2020). Predicting the Accuracy for Telemarketing Process in Banks Using Data Mining. JKAU: Comp. IT. Sci., 9(2), 69–83.

[29] Tekouabou, S. C. K., Cherif, W., and Silkan, H., (2019). A data modeling approach for classification problems: application to bank telemarketing prediction. Proceedings of the 2nd International Conference on Networking, Information Systems & Security, 56, 1-7. https://doi.org/10.1145/3320326.3320389.

[30] Sun, Y., et al., (2019). Prediction of Bank Customer Telemarketing Success Using Data Mining Models. DRRI Journal of Engineering and Technology, 4(5), 18–32. https://doi.org/10.55058/adrrijet.v4i5.461.

[31] Borugadda, P., Nandru, P., and Madhavaiah, C., (2021). Predicting the Success of Bank Telemarketing for Selling Long-term Deposits: An Application of Machine Learning Algorithms. St. Theresa Journal of Humanities and Social Sciences, 7(1), 91–108.

[32] Jin, W., and He, Y., (2019). Three data mining models to predict bank telemarketing. IOP Conference Series: Materials Science and Engineering, 490(6), 1–3. https://doi.org/10.1088/1757-899X/490/6/062075.

[33] Palaniappan, S., Mustapha, A., Foozy, C. F. M., and Atan, R., (2017). Customer Profiling using Classification Approach for Bank Telemarketing. International Journal on Informatics Visualization, 1(4-2), 214–217. https://doi.org/10.1088/1757-899X/490/6/062075.