SYSTEMATIC LITERATURE REVIEW OF MACHINE LEARNING MODELS USED IN THE ANALYSIS OF BANK DATA FOR BETTER DECISION MAKING

Sufaira Shamsudeen*[1], K.Ranjith Singh[2]

Abstract

Machine Learning (ML) Classification algorithms used in the Banking data set are reviewed in this paper. The algorithms performance and results may differ. The evaluation metrics used in the survey papers are measured and summarize each machine learning algorithm's performance. Therefore, in order to choose the best approach, a comparative study is needed to determine the most appropriate approach for a given domain. As we know, nowadays the banking business is based on the customer, so knowing profitable customer is essential for the better decision making in the banking business. The real-world of Bank is changed to customer-centricfromtheproduct-orientedones.In this paper, a study is presented between various classification methods, namely the linear machine learning algorithms and ensemble algorithms. The reviewed articles include various data set of different records and attributes consisting of demographic, behavioral and credit information of Bank from the primary sources and secondary sources such as Kaggle, UCI repositories etc. The paper concluded with the evaluation and comparative study of algorithms based on their classification reports. In the classification reports, the metrics of the algorithms has been computed. And also examined the confusion matrix of each machine learning algorithms.

Keywords: Machine Learning; EnsembleAlgorithms; Confusion Matrix;F-measure;Recall; Precision.

Department of Computer Science
Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India
*Corresponding Author

I. INTRODUCTION

The banking sector has undergone tremendous transformations in the past ten years when compared to the past few decades. The adoption and implementation of technologies and technology enabled solutions have brought in visible benefits in terms of deriving business value, efficiency and profitability. Multiple studies have shown that the use of artificial intelligence (AI) in corporate organizations has provided significant advantages in terms of cost savings, efficient resource use, and in many other ways. Banking is no exception. The Indian banking sector has also embraced AI applications in their systems. AI applications not only help the banks in cost reduction, anti-money laundering or managing risk, but can be used for many other purposes. AI offers unique competence in deriving business value while operating in a highly complex business environment.

Due to significant changes in the economy, work volume, consumer preferences, and the need for a strong access management system and a secure banking environment for transactions, the banking sector has started utilizing artificial intelligence to digitize the time-consuming manual tasks.

In the article, author reviewed many research papers of classifying customers in the banking sector using Machine Learning (ML) algorithms for better decision making and identifying theprofitable customers. Knowing your customers better and target them in the most relevant way is the success of the business. Profiling profitable customers into various segments to customize product offerings is a great task in the banking business world. Nowadays, the business is shifted to customer-centric from the product-oriented ones. Focus is on customers, knowing the customer

behavior, demographic data, transactional data, and credit information the bank can classify the customers into multiclass. Based on the classification, the bank can decide which customer is a brand ambassador of the bank.

II. LITERATURE REVIEW

The analysis of customer bank data is a tedious task. The big data of the bank may consist of demographic data, behavioral data, transactional data and credit information of the customers. From this, we have to classify the customers into multiclass based on the attributes defined. In one of the review focus on [1] the Support Vector Machine and Naive Bayesclassification approaches from the Machine Learning Library of the Apache Spark Data processing System are compared. Later using the two classifiers on a dataset made up of customer personal and behavioral data from Spain's Santander Bank, a comparison was done. The outcomes of the studies show that Nave Bayes performs better than Support Vector Machine in terms of some measures. The review of one research article,[2] focus on several Machine Learning (ML) classifiers for customer analysis, including K-NN, C4.5, Random Forest, Random Tree, LR, MLP, and NB. The results show that the C4.5 model outperforms other ML classifiers in terms of prediction accuracy (89.87%), and that it also outperforms the NB model in terms of time complexity (0.16 seconds), the best time complexity.

In another study, [3] machine learning algorithms have presented good performance in a variety of classification tasks, with character identification, credit default and fraud prediction, as well as forecasting. Customer segmentation is a critical step in creating a successful organization in the private banking sector since it enables financial institutions to target certain consumer groups with their products and services. This study discusses the performance of neural networks and support vector machines.

[4] In order to develop performance models that forecast the future performance of the Nigerian Commercial Bank Industry based on human resources outcomes (employee skills, attitude, and behavior), various machine learning algorithms were applied. K-Nearest Neighbour, Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Machine, Deep Learning, and Naive Bayes were some of the methods used.The study's conclusions unequivocally show how crucial human resource results are to achieving organizational effectiveness.The estimates of the bank performance made by machine learning algorithms are 74–81% accurate. The efficiency of the models is adequately demonstrated by the good results of metrics, which were used to evaluate the performance of the classifier algorithms.Gradient boosting has the best performance results out of the eight classifier models, with an accuracy rating of 81.7%.

The study focused on [5] the KNN, SVM, Decision Tree, and Random Forest classifiers. The performance is computed on the dataset and figure out the higher predictable model.The result concluded that the Random Forest model is better compared to other models in terms of accuracy (95.74%).

The researcher focuses [6] on banks to use a classification algorithm to improve the accuracy of their consumer profiling. The technology used here are NB, DT, and RF. Classification is helpful for forecasting consumer profiles and boosting telemarketing sales, according to measurements of accuracy percentage, precision, and recall rates. According to the data, Naive Bayes had the worst classification accuracy, scoring 86.27%. The highest accuracy is achieved by Decision Tree (90.68%), while the average accuracy of Random Forest is 88.81%.

According to the author, [7] the study focus on the Logit Leaf Model (LLM), a novel hybrid method, is suggested in this research to more accurately identify data. LLM provides models that are more accurate when compared to using Logistic Regression and Decision Tree as independent classification techniques.

[8] Here models like Random Forest, XGBoost, Adaboost, Lightgbm, Decision tree, and K- Nearest Neighbor were used. Different standard metrics are used here and compared for each algorithm. In the midst of these, the highest accuracy score on Logistic Regression (92%), which is the best model. The metrics performed well better than other machine learning methods in relation to F1-Score, which is 96%. [9] Here the recommended system is to utilize various machine learning approaches to better profile bank customer behavior. By several metrics, the artificial neural network demonstrated the highest accuracy. The study, [10] concentrate proactively on utilizing AI with the goal of delighting customers. AI is useful for operations, decision-making, and improving the efficacy of plan implementation in many important areas of banks. [11] Identifies the characteristics that predict the digitalization of customers at banks, depicts the order in which customers make decisions, and investigates the existence in the digitalization process. Random forests, Conditional inference trees, and Causal forests are the technologies employed here. Based on the review of this paper, [12] the banks are embracing the power of AI to deliver new customer experiences through a variety of solutions and are establishing new standards for the Indian banking ecosystem. [13] In this study, supervised artificial neural network algorithm is executed for classification purpose. For categorization, here employed two datasets: bank customer data and germen credit data.For datasets 1 and 2, this method provides accuracy of 72% and 98%, respectively. Results indicate that the created model performs well for two datasets.

III. SUMMARY

Authors	Model	Accuracy
WaelEtaiwietal.(2017)[1]	SVM	85.68%
R.SivaSubramanianetal. (2021)[2]	C4.5 Model	89.87%
SulaimanO etal.(2021)[4]	GradientBoosting	81.70%
ManasRahmanetal.(2020)[5]	RandomForest	95.74%
ShamalaPalaniappanetal. (2018)[6]	DecisionTree	90.68%
MirazAlMamunetal.(2018)[7]	LogisticRegression	92%

IV. CONCLUSION

The purpose of the reviewing the research papers are to sort the intelligent Machine Learning (ML) classifier incustomeranalysisandtoanalyzecustomerdatausingvarious MLclassifiers.Better customer data analyses allow for a more insightful understanding of consumer behavior information, and depending on the analysis, both customer satisfaction and corporate income can be raised. In order to comprehend customer data, a variety of ML classifiers are employed and investigated. These studies apply various machine learning classifiers and the results of the experiments demonstrate that Random Forest has a higher accuracy score than the competition.

REFERENCES

[1] Evaluation of classification algorithms for banking customer's behavior under Apache Spark Data Processing System WaelEtaiwi, MariamBiltawi, Ghazi Naymat, https://doi.org/10.1016/j.procs.2017.08.280.

[2] Customer Analysis Using Machine Learning Algorithms: A Case Study Using Banking Consumer Dataset R.SivaSubramanian, D.Prabhab, B.Maheswari, J.Aswini Recent Trends in Intensive Computing M.Rajesh et al. (Eds.) © 2021 The authors and IOS Press. This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-CommercialLicense4.0(CCBY-NC4.0).doi:10.3233/APC210263

[3] Customer Segmentation in Private Banking Sector Using Machine Learning Techniques. I on Smeureanu, Gheorghe Ruxanda, Laura Maria Badea, Journal of Business Economics and Management, ISSN 1611-1699 print / ISSN 2029-4433 online 2013 Volume 14(5): 923–939.doi:10.3846/16111699.2012.749807.

[4] Machine Learning Classification Techniques for

Detecting th Impact of Human Resources Out comeson Commercial Banks Performance. Sulaiman O. Atiku and Ibidun C.Obagbuwa, Applied Computational Intelligence and Soft Computing Open Access Volume 2021 | Article ID7747907 |https://doi.org/10.1155/2021/7747907.

[5] Machine Learning Based Customer Churn Prediction in Banking, November 2020, DOI:10.1109/ ICECA 49313.2020.9297529, Conference:2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA). ManasRahman, Central University of Kerala, Department of Computer Science, Master of Computer Science,Ph.D. Scholar at Central University of Kerala, VKumar Department of Computer Science Central University of Kerala Periye, Kasaragod.

[6] Customer Profilingusing Classification Approach for Bank Telemarketing, ShamalaPalaniappan, AidaMustapha, CikFeresaMohdFoozyRodziahAtan, International Journal On Informatics Visualization (JOIV), VOL 1 (2017) NO 4 – 2, e-ISSN : 2549-9904, ISSN : 2549-9610.DOI:10.30630/joiv.1.4-2.68

[7] DeCaigny,A., Coussement,K., & DeBock, K.W.(2018). A new hybrid classification algorithm for customer churnprediction based onlogistic regression and decision trees. European Journal of Operational Research, 269(2), 760–772. Volume 269, Issue 2, 1 September 2018, Pages 760-772,https://doi.org/10.1016/j.ejor.2018.02.009.

[8] Predicting Bank Loan Eligibility Using Machine Learning Models and Comparison Analysis. Miraz Al Mamun, Afia Farjana and Muntasir Mamun. 7th North American International Conference on Industrial Engineering and Operations Management, Orlando, Florida, USA, June12-14, 2022.

[9] Improve profiling bank customer's behavior using machine learning, EMAD ABD ELAZIZ DAWOOD, Department of Information systems, Arabacademy for science, technology, and maritime transport,Cairo,Egypt ESSAMELFAKHRANY&FAHIMAA.MAGHRABY, Department of computer science, Arab academy for science, technology, and maritime transport, Cairo, Egypt.,DOI10.1109/ACCESS.2019.2934644,IEEE Access.

[10] Application of Artificial Intelligence for Successful Strategy Implementation in India's Banking Sector, Ms. Bhavna Agarwal, Dr.Himanshu Agarwal and Dr. Parvez Talib International Journal of Advanced Research (IJAR), ISSN: 2320-5407 Int. J. Adv. Res. 7(11), 157-166.

[11] Amachinel earning approachtothe digitalization of bank customers: Evidence fromr and om and causal for ests, Santiago Carbo-Valverde,Pedro Cuadros-Solas, Francisco Rodríguez-Fernández, Published: October 28, 2020, https://doi.org/10.1371/ journal. pone.0240362, Research Article in Plo SONE.

[12] Impact and Challenges of Artificial Intelligence in Banking, Dr. K. Suresh Kumar, Aishwarya Lakshmi S, Akalya. A, Journal of Information and Computational Science, Volume10Issue2– 2020, ISSN:1548-7741.

[13] Analysis of Banking Data Using Machine Learning, Priyanka S. Patil & Nagaraj V.Dharwadkar, International conference on I-SMAC (IoTin Social, Mobile, Analyticsand Cloud) (I-SMAC2017). DOI:978-1-5090-3243-3/17/$31.00 ©2017 IEEE