The Seybold REPORT

ISSN 1533-9211

DOI 10.17605/OSF.IO/UTD3B

MACHINE LEARNING BASED CUSTOMER CHURN PREDICTION IN BANKING

JOBY JACOB

Research Scholar, Karpagam Academy of Higher Education (Deemed to be University), Coimbatore, India. Email: jobyjacob1975@gmail.com

Dr. V. SANGEETHA

Research Guide, Karpagam Academy of Higher Education (Deemed to be University), Coimbatore, India. Email: drsangeetha.v@kahedu.edu.in

Abstract

Customer connections are of the utmost significance to any bank in the highly competitive industry of banking. The banks view every customer as a lifelong customer. The state in which a client or subscriber ceases engaging in business transactions with a firm or a service provider is referred to as "customer churn". Customer relationship management and business value are improved by customer churn study and forecast. One of the most important aspects of a country's development is the use of AI in the banking industry. The banking industry's growth is largely dependent on its important clients. In order to identify whether a client is worth keeping or at risk of leaving, customer churn analysis is required. From an organizational perspective, acquiring new clients is typically harder or more expensive than keeping the ones you already have. Therefore, predicting customer churn has become common in the banking sector. Commercial banks increase their core competitiveness among rivals while also increasing earnings by lowering client attrition or churn. Although numerous researchers put forth numerous single prediction models and some hybrid models, accuracy is still poor and some algorithms' computation times are still lengthy.

INTRODUCTION

The market of the banking industry in the present day shows many of the competitive performances of the banks. The increase if the services providers of the bank make the customers mean the public of the country is confused in the process of investing their money in the bank.

This leads to the distraction of them in the making decision-making process. This makes the banks face the issue of less interaction with the customers in the bank. So the revenue in the banking organization is also decreasing.

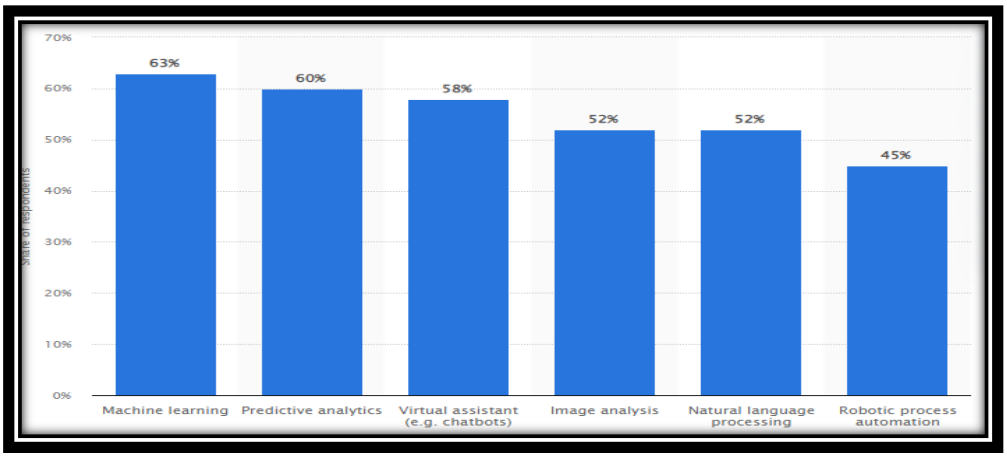
As per the view Khanal et al. (2020), in this work, the introduction of machine learning in the banking processes of prediction of customer churn is shown as per the proposal. This includes the application of digital methods and technologies in the prediction of the customers' minds.

This work shows the analysis of machine learning in the prediction of banking customers.

ISSN 1533-9211

DOI 10.17605/OSF.IO/UTD3B

Figure 1: Investment banks worldwide using artificial intelligence



(Source: Statista, 2022)

A lot of banks in the world use digitalization in their banking process of them. This makes them easier the examination the mind-set of the customers of them in making the prediction of the customer to invest in the banks. According to the views of Sungheetha & Sharma (2021), in the current market of competition, the banks also face a lack of customer interaction in the banking process. This shows the performance of the banks in the implementation of the digitalization of the banking process. The recent economical examination of banking performance also shows the performance of the banking organization through machine learning.

LITERATURE REVIEW

Machine learning in the baking process

The changes in the current scenario, the world is making the increase of the diversification of the mindsets of the customer. To attract the maximum number of the customer the banks in the market makes the introduction of new features in their banking facilities of them. As per the views of Sarker et al. (2020), this leads to an increase in competition among baking service providers. One of the banking process factors that make the customer more attracted to the facility of banking is the service of digital making. This makes the customer get the actual and effective solution for their banking transactions. On the other hand Tahsien et al. (2020) argued that the task of the banks in making the prediction of the customer interaction in them becomes a hard task to be performed. Therefore they use the digital medium and the interface to analyze the mindset and their future interaction with the customers in them.

Impact of machine learning on the Baking Process

The banks in the market make the examination of needs of the customer in the market. They provide and try to provide the best facilities to the customers of the market. According to the views of Rahman & Kumar, (2020), this makes them gain maximum competitive advantages

ISSN 1533-9211

DOI 10.17605/OSF.IO/UTD3B

over the other producers of the services of banking. The digitalization of the facilities of the banking processes makes the customer feel comfortable in the making of the transactions. On the other hand Tran et al. (2021) argued that the customer easily performs in the time making transactions from any location without performing the task from the banking institution. The banks also get the record of the transaction in the digital process of strong without any extra effort.

Issues of machine learning in the banking process

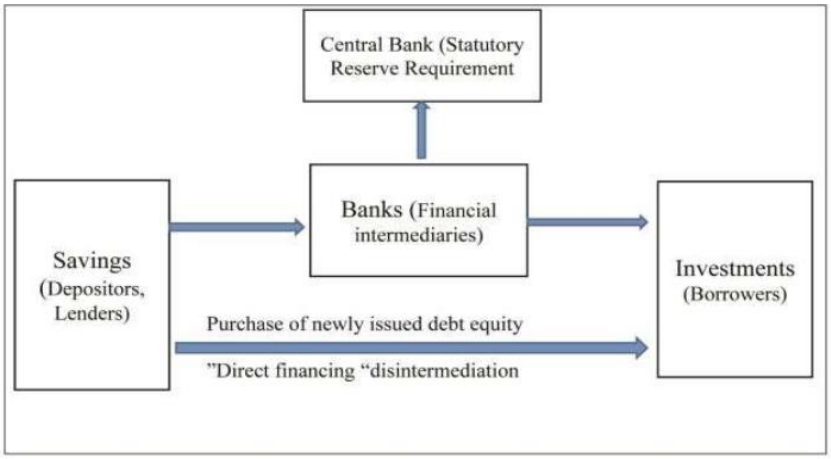
Despite being the most customer-attracted feature of the banking process the banking process of digitalization makes customers and the baking authorities face many issues in the performance of the digitalization of the banking process. Baszed on the views of Domingos et al. (2021), those issues are as the prediction of those customers of the banks are done through the digital learning process so the effectiveness of the methods does not first accurately. On the other hand Leo et al. (2019) argued that the information provided by the digital medium is not reliable. In the case of the customer, it is not possible for there a digital interface to represent the expectation of humans.

Mitigation of the Issues

This work makes analysis all the issues generated in the ongoing process of the banking organization and provides all the information that is required to solve the issue quickly. Based on the views of Guerra & Castelli (2021), this includes the strategic methods that increase the effectiveness of digital banking and machine learning. On the other hand Fati (2021) argued that the study shows all the issues generated by machine learning in the banking process of predicting customers. The strategies and the methods provided in the work increase the impact of the machine learning of the banks in making the prediction of the banks.

Financial intermediation theory

Figure 2: Financial Intermediation Theory



(Source: Rashid et al.2020)

The Seybold REPORT
ISSN 1533-9211
DOI 10.17605/OSF.IO/UTD3B

Traditional theories of intermediation are based on transaction costs and situational information. They are designed to account for the baking institutions which take deposits from the customers or issue insurance policies and channel funds to firms or other working organizations (Rashid et al.2020). However, in recent years there were significant changes in that process. These theories make the analysis of the customers mind-set easily. The more quickly the banking organisations become in the analysis of the situation in the working environment of the bank more effectively, the banks perform in making the prediction of the customers of the banks. He research makes analyses of the current issues generated in the environment of the banks. This also makes the analysis of the previous work and provides all the methodologies that can be applied in the current environment of banks performing digital learning for customer prediction (Daluwathumullagamage & Sims 2021). This provides all the missing information from the previous work and provides effective solutions**.**

ARCHITECTURE DESIGN

The proposed model is detailed in the following sections.

Data Collection

Table 1: Dataset and Description

Attribute	Description
Row Number	Number of customers
Customer ID	ID of customer
Surname	Customer name
Credit Score	Score of credit card usage
Geography	Location of customer
Gender	Customer gender
Age	Age of Customer
Tenure	The period of having the account in months
Balance	Customer main balance
NumOfProducts	No of products used by customer
HasCrCard	If the customer has a credit card or not
IsActiveMember	Customer account is active or not
Estimated Salary	Estimated salary of the customer.
Churn	Indicates customer leaved or not

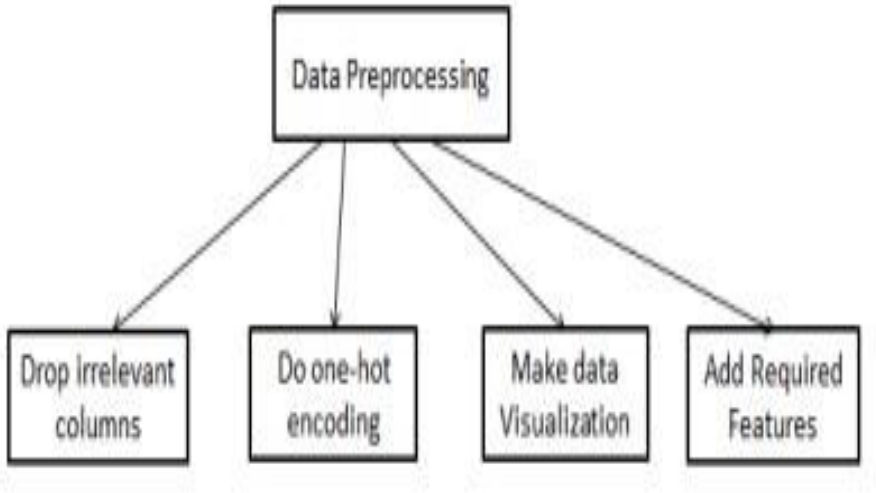
This dataset includes 10k bank customer data records with 14 attributes including sociodemographic attributes, account level and behavioural attributes shown in Table 1.

ISSN 1533-9211

DOI 10.17605/OSF.IO/UTD3B

Data pre-processing

Figure 3: Data Pre-Processing



Pre-processing data is divided into four sections, as depicted in Figure 1. First, preprocessing sample data involves removing irrelevant attributes (such as Row Number, Customer ID, and Surname) that have no bearing on churn prediction. Because K-means can only understand numerical attributes, the strings or categorical attributes were then converted using one-hot encoding to numerical values.

Additionally, visualize the data to determine which characteristics are most frequently linked to churn. Add a few features based on the visualization's findings to improve the churn categorization. The pre-processed data is then split into 20% for testing and 80% for training.

RESULT AND ANALYSIS

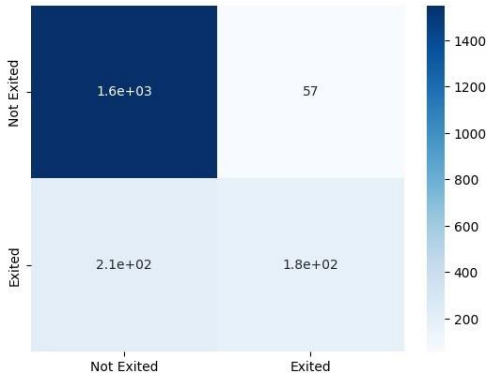
Confusion Matrix

The framework contains genuine positive (TP: assault information accurately delegated positive), bogus negative (FN: assault information wrongly named negative), bogus positive (FP: typical information wrongly named positive), and genuine negative (TN: ordinary information effectively named negative).

ISSN 1533-9211

DOI 10.17605/OSF.IO/UTD3B

Figure 4: Confusion matrix for target class of testing



Performance Comparison

Table 2: Result Comparison

Method	Accuracy	F1 score	Recall
Svm	79	82	78
Random forest	86	86	84
XGBoost	88	88	87

CONCLUSION

Customer churn analysis has become a major concern in almost every industry that offers products and services. The model developed will help banks identify clients who are likely to be churners and develop appropriate marketing actions to retain their valuable clients. And this model also supports information about similar customer group to consider which marketing reactions are to be provided. Thus, due to existing customers are retained, it will provide banks with increased profits and revenues.

References

1) Chatterjee, A., Bhowmick, H., & Sen, J. (2021, October). Stock price prediction using time series, econometric, machine learning, and deep learning models. In 2021 IEEE Mysore Sub Section International Conference (MysuruCon) (pp. 289-296). IEEE. Retrived on: 22nd march,2023. From: https://arxiv.org/pdf/2111.01137

2) Daluwathumullagamage, D. J., & Sims, A. (2021). Fantastic beasts: Blockchain based banking. Journal of risk and financial management, 14(4), 1-43.Retrived on: 22nd march,2023. From: https://www.econstor.eu/bitstream/10419/239586/1/1759808202.pdf

3) Domingos, E., Ojeme, B., & Daramola, O. (2021). Experimental analysis of hyperparameters for deep learning-based churn prediction in the banking sector. Computation, 9(3), 34. Retrived on: 22nd march,2023. From: https://www.mdpi.com/2079-3197/9/3/34/pdf

4) Fati, S. M. (2021). Machine learning-based prediction model for loan status approval. Journal of Hunan University Natural Sciences, 48(10). Retrived on: 22nd march,2023. From: http://jonuns.com/index.php/journal/article/download/783/779

The Seybold REPORT
ISSN 1533-9211
DOI 10.17605/OSF.IO/UTD3B

5) Guerra, P., & Castelli, M. (2021). Machine learning applied to banking supervision a literature review. Risks, 9(7), 136.Retrived on: 22nd march,2023. From: https://www.mdpi.com/2227-9091/9/7/136/pdf

6) Hameed, M. M., Ahmad, R., Kiah, M. L. M., & Murtaza, G. (2021). Machine learning-based offline signature verification systems: A systematic review. Signal Processing: Image Communication, 93, 116139.Retrived on: 22nd march,2023. From: https://www.researchgate.net/profile/Ghulam_Murtaza58/publication/348475222_Machine_learning-based_offline_signature_verification_systems_A_systematic_review/links/60fbb274169a1a0103b20ba4/Machine-learning-based-offline-signature-verification-systems-A-systematic-review.pdf

7) Kawa, D., Punyani, S., Nayak, P., Karkera, A., & Jyotinagar, V. (2019). Credit risk assessment from combined bank records using federated learning. International Research Journal of Engineering and Technology (IRJET), 6(4), 1355-1358. Retrived on: 22nd march,2023. From: https://www.academia.edu/download/59805474/IRJET-V6I428820190620-14769-hvp12k.pdf

8) Khanal, S. S., Prasad, P. W. C., Alsadoon, A., & Maag, A. (2020). A systematic review: machine learning based recommendation systems for e-learning. Education and Information Technologies, 25, 2635-2664.Retrived on: 22nd march,2023. From: https://researchoutput.csu.edu.au/ws/portalfiles/portal/56945122/37059927_Published_article.pdf

9) Leo, M., Sharma, S., & Maddulety, K. (2019). Machine learning in banking risk management: A literature review. Risks, 7(1), 29. Retrived on: 22nd march,2023. From: https://www.mdpi.com/2227-9091/7/1/29/pdf

10) Rahman, M., & Kumar, V. (2020, November). Machine learning based customer churn prediction in banking. In 2020 4th international conference on electronics, communication and aerospace technology (ICECA) (pp. 1196-1201). IEEE.Retrived on: 22nd march,2023. From: https://www.researchgate.net/profile/Manas-Rahman/publication/348094541_Machine_Learning_Based_Customer_Churn_Prediction_In_Banking/links/615fe96a0bf51d48175149de/Machine-Learning-Based-Customer-Churn-Prediction-In-Banking.pdf

11) Rashid, J., Mahmood, T., Nisar, M. W., & Nazir, T. (2020, November). Phishing detection using machine learning technique. In 2020 first international conference of smart systems and emerging technologies (SMARTTECH) (pp. 43-46). IEEE. Retrived on: 22nd march,2023. From: https://www.researchgate.net/profile/Tahira-Nazir-2/publication/347308591_Phishing_Detection_Using_Machine_Learning_Technique/links/604b04fb92851c1bd4e291c5/Phishing-Detection-Using-Machine-Learning-Technique.pdf

12) Salahdine, F., El Mrabet, Z., & Kaabouch, N. (2021, December). Phishing Attacks Detection A Machine Learning-Based Approach. In 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) (pp. 0250-0255). IEEE.Retrived on: 22nd march,2023. From: https://arxiv.org/pdf/2201.10752

13) Sarker, I. H., Abushark, Y. B., Alsolami, F., & Khan, A. I. (2020). Intrudtree: a machine learning based cyber security intrusion detection model. Symmetry, 12(5), 754. Retrived on: 22nd march,2023. From: https://www.mdpi.com/2073-8994/12/5/754/pdf

14) Statista, 2022 Available.https://www.statista.com/statistics/1246874/ai-used-in-investment-banking-worldwide-2020/ Accesed on: 22nd march,2023. Available at:

15) Sungheetha, A., & Sharma, R. (2021). 3D image processing using machine learning based input processing for man-machine interaction. Journal of Innovative Image Processing (JIIP), 3(01), 1-6. Retrived on: 22nd march,2023. From: https://irojournals.com/iroiip/V3/I1/01.pdf

The Seybold REPORT
ISSN 1533-9211
DOI 10.17605/OSF.IO/UTD3B

16) Tahsien, S. M., Karimipour, H., & Spachos, P. (2020). Machine learning based solutions for security of Internet of Things (IoT): A survey. Journal of Network and Computer Applications, 161, 102630.Retrived on: 22nd march,2023. From: https://arxiv.org/pdf/2004.05289

17) Tékouabou, S. C. K., Gherghina, Ş. C., Toulni, H., Neves Mata, P., Mata, M. N., & Martins, J. M. (2022). A machine learning framework towards bank telemarketing prediction. Journal of Risk and Financial Management, 15(6), 269. Retrived on: 22nd march,2023. From: https://www.mdpi.com/1911-8074/15/6/269/pdf

18) Tran, K. Q., Duong, B. V., Tran, L. Q., Tran, A. L. H., Nguyen, A. T., & Nguyen, K. V. (2021). Machine learning-based empirical investigation for credit scoring in Vietnam's banking. In Advances and Trends in Artificial Intelligence. From Theory to Practice: 34th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, IEA/AIE 2021, Kuala Lumpur, Malaysia, July 26–29, 2021, Proceedings, Part II 34 (pp. 564-574). Springer International Publishing.Retrived on: 22nd march,2023. From: https://www.researchgate.net/profile/Tran-Khanh-8/publication/353329117_Machine_Learning-Based_Empirical_Investigation_for_Credit_Scoring_in_Vietnam's_Banking/links/60f6c23cfb568a7098c05e2d/Machine-Learning-Based-Empirical-Investigation-for-Credit-Scoring-in-Vietnams-Banking.pdf

19) Ustun, T. S., Hussain, S. S., Ulutas, A., Onen, A., Roomi, M. M., & Mashima, D. (2021). Machine learning-based intrusion detection for achieving cybersecurity in smart grids using IEC 61850 GOOSE messages. Symmetry, 13(5), 826.Retrived on: 22nd march,2023. From: https://www.mdpi.com/2073-8994/13/5/826/pdf