THE ASIAN BULLETIN OF BIG DATA MANAGMENT
Vol. 4. Issue 3 (2024)
https://doi.org/ 10.62019/abbdm. v4i3.206

ASIAN BULLETIN OF BIG DATA MANAGEMENT
ISSN (Print): 2959-0795
http://abbdm.com/
ISSN (online): 2959-0809

From Data to Decisions: Predictive Machine Learning Models for Customer Retention in Banking

Abdul Khaliq*, Sophia Ajaz, Asif Ali, Daniyal Shakir, Kashif Baig

Chronicle

Article history

Received: Aug 25, 2024
Received in the revised format: Aug 27, 2024
Accepted: Aug 31, 2024
Available online: Sept 1, 2024

Abdul Khaliq is currently affiliated with Department of Computer Science Institute of Business Management, Karachi, Pakistan.
Email: Sm1k0ol28@gmail.com

Sophia Ajaz & Asif Ali are currently affiliated with Department of Computer Science, Iqra University, Karachi, Pakistan.
Email: sophia.ajaz@iqra.edu.pk
Email: asifali@iqra.edu.pk

Daniyal Shakir is currently affiliated with S & P Global, Karachi, Pakistan.
Email: daniyalshakir@gmail.com

Kashif Baig is currently affiliated with FEST, iqra university, Pakistan.
Email: kashifbaig@iqra.edu.pk

Corresponding Author*

Abstract

This research proposes machine learning—a subfield of artificial intelligence-based approaches to forecast bank client attrition. The study promotes investigating the possibility of churn through the examination of consumer behaviour. In this paper, a method to predict customer churn in a bank using machine learning techniques, which is a branch of artificial intelligence, is proposed. The research encourages the exploration of the likelihood of churn by analysing customer behaviour. The number of service providers is increasing very rapidly in every business these days. As a result, customer churn and engagement have become one of the top issues for most of the banks. In this work, the Random Forest, SVC, XGB, LGBM, and Logistic Regression classifiers are employed. A few feature selection strategies have also been applied to validate system performance and determine which features are more relevant. The test was conducted using the churn modelling dataset from IEEE dataset. The results are compared to find an appropriate model with increased precision and predictability. As a result, when utilised after oversampling, the Random Forest model outperforms other models in terms of accuracy. The Random Forest model showed a significant improvement in accuracy compared to the other classifiers, making it the preferred choice for predicting customer churn in the banking sector. The feature selection strategies helped identify the most relevant features that contribute to accurate predictions. Overall, this study highlights the importance of implementing advanced machine learning techniques and feature selection methods to enhance customer retention strategies in the banking industry.

Keywords: Bank Churn System, XGB, Random Forest.

INTRODUCTION

These days, the market is extremely competitive and dynamic. The availability of numerous service providers is the reason behind this. The evolving customer landscape and their growing expectations provide issues for service providers. The evolving expectations of today's consumers, coupled with their varied needs for connectivity and creative, customized solutions, set them apart from earlier consumer generations. To stay ahead in this competitive market, service providers must continuously adapt and innovate to meet the changing needs of their customers. This requires a deep understanding of consumer behavior, market trends, and technological advancements. By staying proactive and responsive to customer feedback, service providers can ensure they are delivering the high-quality, personalized services that today's consumers demand. Additionally, fostering strong relationships with customers through exceptional customer service and support can help service providers differentiate themselves from their competitors and build brand loyalty. They have more education and are more up-to-date on new strategies. Their buying habits have altered as a result of this enhanced information, leading to a tendency of "analysis-paralysis"—an excessive amount of analysis of the selling and purchase scenario—which eventually aids in the improvement of their purchase selections. As a result, it will be difficult for the new generation of service providers to come up with creative ways to satisfy and enhance the value for their clients. Businesses must identify their customers. By suggesting that firms are under increased pressure from the competition to create novel marketing strategies to match customer expectations and improve loyalty and retention, Liu and Shih (Fei, D. et al., 201) support this claim. According to Canning (Keshavarz, H., et al., 2021), providing more to everyone is no longer a practical sales approach, and in an increasingly competitive market, a plan that prioritizes the maximum efficient usage of promotion resources is required. Businesses have employed technology to help them maintain a competitive edge (Lottu, O. et al., 2023)

One popular information technology for extracting marketing knowledge and additional direction for company decisions is data mining techniques (Mengjing Hao. Et al., 2024). Customers can easily move from one organization (Bank) to another in search of better rates or service quality. Businesses firmly believe that acquiring new clients is significantly more difficult and costly than retaining current ones (N.F. Mansor, et al., 2020). However, another big problem for them is providing consumers with dependable facilities while under economical while still having a positive occupied relationship with them. To address these challenges, companies often turn to data mining techniques to analyze customer behavior, preferences, and trends. By leveraging data mining, businesses can gain valuable insights that help them improve customer retention, target marketing efforts more effectively, and streamline operations (Routh, P., et al., 2021). Additionally, data mining can also help companies identify potential issues before they escalate, allowing them to proactively address customer concerns and maintain positive relationships. Overall, data mining plays a crucial role in helping businesses stay competitive in today's rapidly evolving market landscape.

To overcome these obstacles, they must take into account the needs of their customers. They will place a lot of focus on client turnover among these. Customer churn is the term used to describe when clients or subscribers discontinue utilizing a company or service. When customers or subscribers stop using a business or service, this is known as customer churn. Gaining new business for every company requires utilizing its sales and marketing resources to go through the sales funnel (Saha, et al., 2023). However, since they already have the trust and loyalty of existing clients, customer retention is typically more cost-effective. Thus, for any firm, having a structure that canister accurately forecast client attrition in its early phases is crucial. The goal of this research is to develop a system that uses some machine learning approaches to forecast client attrition in the banking industry (Tirkkonen, T. et al. 2021). To achieve this, we will gather historical data on customer behavior and preferences, as well as demographic information, transaction history, and customer feedback. By analyzing this data using machine learning algorithms such as decision trees, random forests, and neural networks, we aim to identify patterns and trends that indicate a high likelihood of customer churn. This predictive model will enable banks to proactively reach out to at-risk customers and offer targeted retention strategies, ultimately reducing customer attrition rates and increasing overall profitability. Additionally, by understanding the factors that contribute to churn, banks can make strategic decisions to improve customer satisfaction and loyalty, ultimately driving long-term success in the competitive banking industry.

LITERATURE REVIEW

The examination of customer attrition in the banking industry is a very vast field. In one of these studies, (Lalwani, P., et al., 2022), client attrition in commercial banks is predicted using the SVM model. For this study, a consumer dataset from a Chinese commercial bank with 50,000 records was used. 46,406 legal statistics records remain afterward preprocessing records. The linear SVM and the SVM with radial basis kernel function are the two types of SVM models that are chosen (Manzoor, A., et al., 2024). The use of under-sampling considerably increased the classification models' predictive power. Uniform the overall valuation parameters are unable to compute the predictive capability of the SVM model because of the asymmetric characteristics included in the real profitable bank customer churn dataset, which prevents the typical from precisely predicting churners (Johnson, J. et al., 2022). The results demonstrate how the random sampling approach may be integrated with the SVM typical to knowingly improve systematic ability and aid commercial banks in making more accurate churner predictions. However, the study's proportion of churners to non-churners was 1:10. In a 1:1 ratio, the maximum result is 80.84% (Lottu, O. A., et al., 2024). This is the work's primary flaw, another article provides a scientific investigation of the use of data mining to get information from repositories in the banking sector (Jain H, et al., 2020).

Affording to the research, clients whom use additional banking facilities and harvests appear to be extra devoted. As an outcome, the bank must focus on selling items that meet the demands of its clients who use fewer than three goods. Records on 1866 clients at the time of the study are contained in the database that was used. The study is based on a neural network-based approach to churn prediction from the Alyuda Neuro Intelligence software suite (Z.A. Abas, et al., 2022). It separates Data into three sets: sets for testing, authentication, and training. Three categories of characteristics are provided in the data analysis process: In the data analysis step, three types of characteristics are described: the qualities to be measured, the traits to be refused, and the characteristics that are needed. Several unseen layers in the network plan procedure are selected by the model. The results of the exercise of the network display that the authentication's CCR% is 93,959732.

The analysis found that the bank provides very well-tailored programs for retirees and that there is very little chance of competition due to the high percentage of pensioners in the overall number of patrons (691/1886) (Keshavarz, H., et al., 2021). It also highlighted the shortcomings of previous studies and suggested an ML-based purchaser churn guess in the banking system using the dataset "churn modeling data" to address these shortcomings. A churn analysis model was proposed in the study (J. Raj et al., 2019) to back telecom operators in forecasting which customers are greatest likely to leave. The system takes advantage of big data platforms and machine learning techniques. The model's efficiency is evaluated using the conventional quantity known as the (AUC). The telecom operator Syriatel contributed to the dataset that was used in the study. Decision trees, random forests, gradient-boosted machine trees (GBM) (Wang, K., et al., 2022), and extreme gradient boosting (XGBOOST) (Hancock, J. et al., 2021) are the four approaches that the model has been used with. The big data stage that was chosen was the Hortonworks Data Platform (HDP) (Usman-Hamza, et al., 2024). Almost every stage of the product's expansion, counting data examination, function creation, software testing, and training, involved the use of spark engines. K-fold cross-validation was used to enhance the algorithm's hyper-parameters (Gaur A, et al., 2018). The learning model is rebalanced by taking a sample of data to stability the two lessons because the mark class is rough. To fit the churn class with the other class, the investigation started with oversampling (Dhanawade, A., et al., 2023).

To reduce the taster size of the wide-ranging class that needs to be associated with the additional class, a chance under-sampling technique was also practical. The Decision Tree algorithm was the first to be trained, with the hyper-parameter complexity and supreme number of nodes being improved. The best findings indicate that 200 trees were the ideal number of trees in both Random Forest and GBM. Also, GBM outperformed DT and RF in their results. The best AUC value, according to the results, was 93.301% for XGBOOST on 180 trees. By installing a new dataset for different times to test the models, XGBOOST also produced the best results (89% AUC) when no positive marketing intervention was used. According to the study's hypothesis, the phenomena of non-stationary data models may be to blame for the subsequent decline, necessitating further model training.

METHODOLOGY

This work aims to predict customer churn in a commercial bank as early as possible using efficient data mining methods. This study aims to comprehend and forecast bank client attrition. To be more precise, we will use exploratory data analysis (EDA) (Dhanawade, A., et al., 2023) at the beginning to pinpoint and illustrate the elements that lead to client attrition. In the future, this data will assist us in developing machine learning models that will enable us to forecast client attrition. This is an example of a common categorization task (Fei, D. et al., 2021). Which performance indicator we should utilize to optimize our machine-learning models is not specified in the job. Accurately distinguishing members of the positive class (customers who would churn) is more important for the bank (Geiler, L., et al., 2022).

1. Dataset Description

The dataset was gathered from Kaggle and used to model churns. The dataset contains information on 10,000 bank clients, with the goal parameter being a binary variable that shows whether a customer has left the bank or is still a customer. Of these samples, 7963 were positive class (kept), while 2037 were negative class (exited). The target variable displays the binary flag 1 when a customer's bank account is closed and the binary flag 0 when the client is retained. Thirteen feature vectors (predictors) total from customer data and customer-processed transactions are included in the dataset (Mengjing Hao. Et al, 2024).

2. Exploratory Data Analysis

- Target Variable: Exited

As previously stated, the target variable is pre-encoded and has two potential values (as seen in Figure 1): zero (0) for a client who has not churned and one (1) for a customer who has.

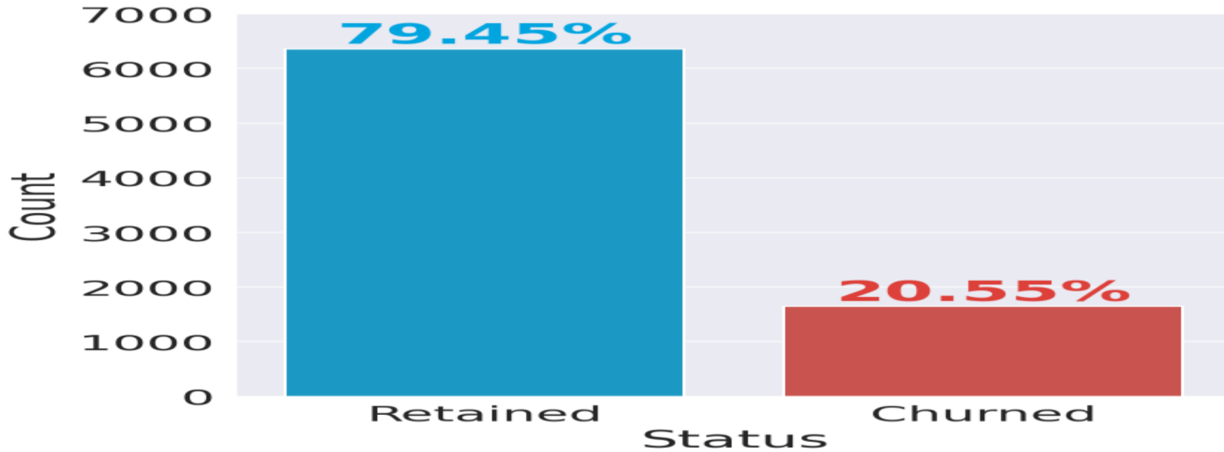

Figure 1.
Minority and majority class of target with imbalanced data.

80% of the bank's customers were retained. The fact that there are significantly more examples in the "Retained" class than in the "Churned" class indicates that our dataset is biased or lopsided (Ullah I, et al., 2019), (Zhao, M., et al., 2021). As a result, precision is most likely not the ideal indicator of model performance. Differentiating between continuous and categorical data and looking at them individually is beneficial since various visualization approaches apply to different types of variables (E V, Ravikumar et al., 2019). Figure 2 shows the different visualization of multiple attributes related to the exited target column.

Figure 2.
Different features related to the target.

In the above figure 2 it is shown that Age has a modest tail-heavy distribution, meaning that it extends more to the right than the left of the median. In a similar vein, most "Credit Score" scores are higher than 600. The distribution of "Balance" is quite regular if the first bin is ignored, but the distribution of "Estimated Salary" is rather consistent and offers little insight. We can calculate the regular correlation constant among every pair of (uninterrupted) features (Dahouda, M. et al., 2021). There is no significant intercorrelation between our features, so we do not have to worry about multicollinearity. Figure 3 shows the Age attribute relation with churn and retained with specific count number.

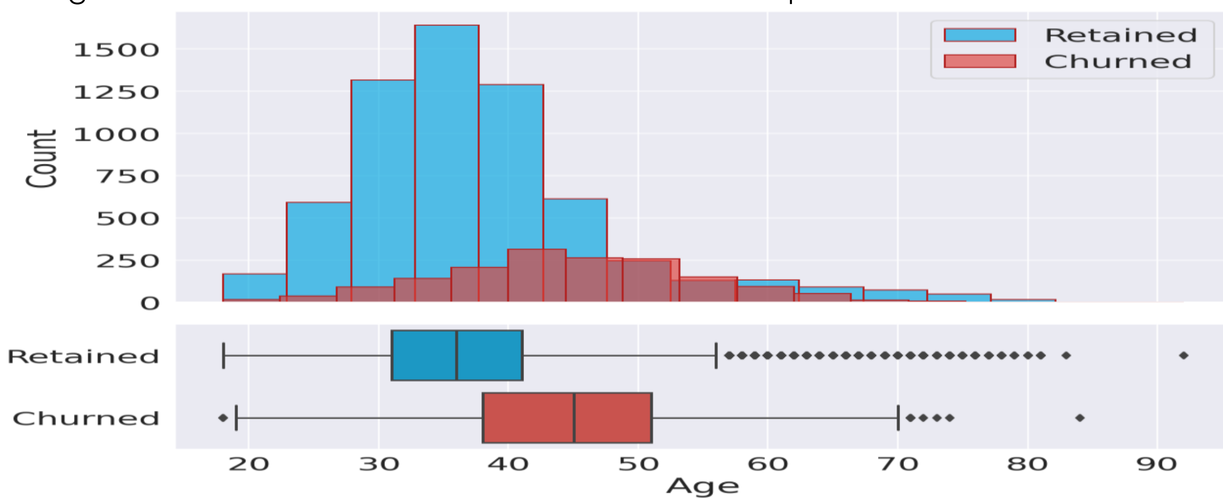

Figure 3.
Retained and churn w.r.t Age attribute.

It's interesting to see that there is a discernible variation in the age groupings since older clients are more prone to leave. This finding may suggest that tastes vary with age and that the bank has not modified its approach to accommodate senior citizens' needs (A. K. Ahmad et al., 2019). Regarding credit ratings, there is no discernible difference between clients who are maintained and those who are lost. The relationship between several parameters and the target column is depicted in Figure 4.

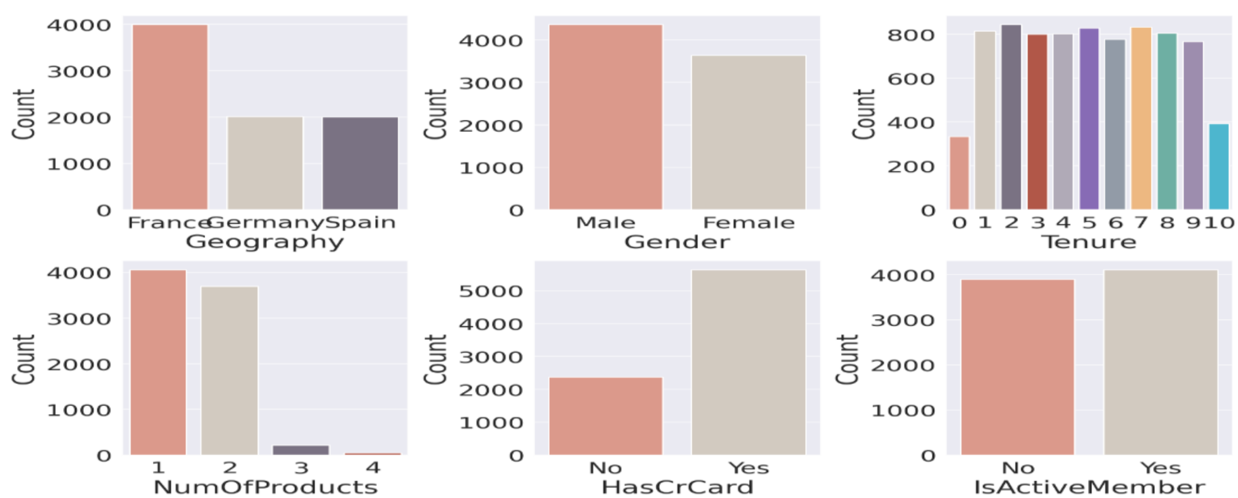

Figure 4.
Frequency of different features.

Important points

- The bank has clients in 3 countries (Spain, France, and Germany). Maximum clients are in France. As shown in figure 5 the account ratio of different bank customers in different countries.
- There are supplementary man clients than women.
- Only a small fraction leaves within the primary year. The count of customers in tenure years between 1 and 9 is virtually the same.
- Most of the customers have purchased 1 or 2 goods, while a small portion has purchased 3 or 4.
- A important mainstream of clients have a credit card, and nearly 50% of clients are not active.

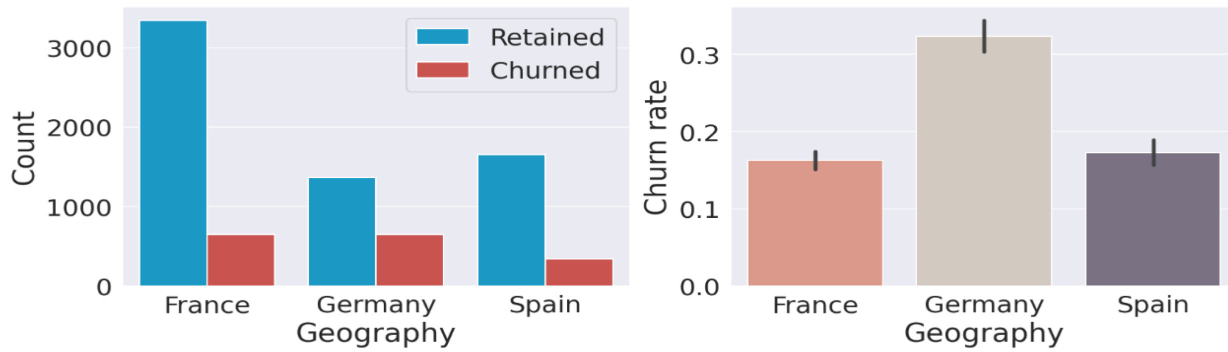

Figure 5.
Churn rate ratio concerning different countries.

Germany has a higher customer turnover rate than the other two nations (it is about twice as high in Germany as it is in Spain and France). This conclusion might be explained by a variety of factors, including more competition or distinct tastes among German consumers (Agha, A et al., 2021), (Ahmed AAQ et al., 2017). Similarly, figure 6 shows the male and female ratio of churn rate and retained rate, where it is observed that Female customers are more likely to churn.

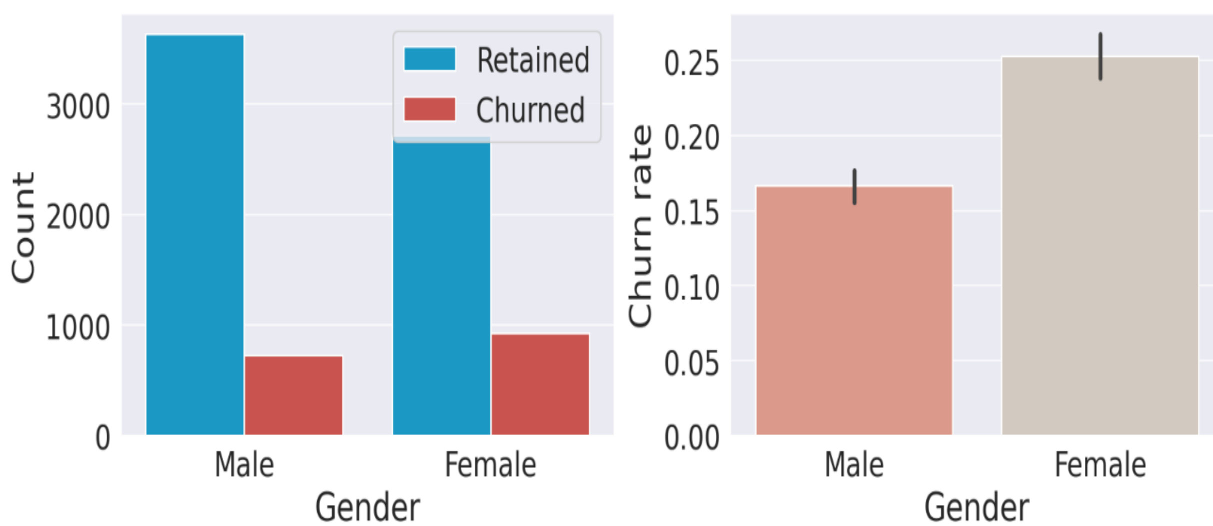

Figure 6.
Churn rate ratio w.r.t gender.

DATA PREPROCESSING

Encoding Categorical Features

Most machine learning methods require numerical characteristics for all input and output. Therefore, before creating models, categorical characteristics must be translated (encoded) into numerical values. Two characteristics in our dataset need to be encoded. 'Gender' will be represented by (Male --> 1 and Female --> 0). We will manually map values for "Geography" such that consumers in Germany are valued at one (1) while customers in France and Spain are valued at zero (0). Since the customer turnover rate in the other two nations is almost equal and significantly lower than in Germany, we decided to use this strategy (Pamina J, et al., 2019). Encoding this feature to distinguish between German and non-German clients makes sense as a result.

Scaling

One method for normalizing the range of features in a dataset is feature scaling. While certain algorithms, like Random Forests, are invariant, others, like Support Vector Machines, are sensitive to feature scaling. A common issue in many real-world jobs is class imbalance. Machine learning algorithms that utilize uneven data for classification are likely to produce models that only predict the most common class (N.F. Mansor et al., 2020). This is because classification using imbalanced data is biased in favor of the majority class. Furthermore, when dealing with class-imbalanced data, typical metrics might be deceptive (e.g., a classifier that always predicts 0 will have 99.9% accuracy if a dataset has 99.9% 0s and 0.01% 1s (Nadeem, G., et al., 2024)

Handling data imbalance

Several tactics can deal with this issue. We employ an algorithm called SMOTE ('Synthetic Minority Oversampling Technique,' which locates a record that is similar to the unsampled record and builds a synthetic record (Dhanawade, A., et al., 2023), which is a randomly weighted average of the original record and its neighbor, with each predictor's weight being generated independently (Hasan, B., et al., 2024).

Baseline Models

To begin this part, we first build two basic models to estimate the training set's baseline performance. We will specifically employ Logistic Regression (Jain H, et al., 2020) and Gaussian Naïve Bayes. Using k-fold cross-validation, we will assess their (mean) recall using their default parameters. The basic concept of k-fold cross-validation is as follows: the (training) set is divided into k subsets/folds (Nadeem, G., Et al., 2023), the models are trained on k-1 folds, and the model is evaluated on the remaining one-fold. Until each fold is tested once, these steps are repeated. By using various machine learning models for analysis (Rahman, M. M. et al., 2023), Figure 7 displays the properties that are most pertinent to the goal.

Attributes selection using multiple ML models.

For all classifiers, "Age" and "Num of Products" appear to be the most helpful characteristics, followed by "Is Active Member" and "Balance." Conversely, with the exception of LGBM, "Credit Score" is the least significant attribute, having a minuscule value that is almost equal to zero

Figure 7.
Experimental Result

Performance Assessment

First, we can assess how well our classifiers perform using four different metrics: area under the ROC curve, or simply AUC; accuracy, precision, recall, and recall. as displayed in Table 1.

Table 1.
Shows the accuracy of metrics.

Models	Accuracy	Precision	Recall	AUC
LR	0.68	0.66	0.70	0.74
RF	0.95	0.98	0.87	0.97
GBC	0.78	0.80	0.75	0.85
XGB	0.79	0.80	0.71	0.86
LGBM	0.80	0.82	0.70	0.88
SVC	0.79	0.88	0.78	0.86

Recall for the remaining classifiers is greater than 70% (baseline performance). The model with the greatest recall percentage (78.5%) is XGB. But in terms of overall performance, accuracy, precision, and AUC, the LGBM classifier excels.

CONCLUSION

Like any other firm, the banking industry is taken into consideration, but one of the main issues these days is client involvement. Banks must act swiftly to identify potential consumer churn to fix this situation. Numerous studies on the prediction of banking churn are now underway. Various organizations use various bits of data or information to calculate the customer turnover rate. It is crucial to have a system in place that can predict customer attrition in banks in a broad sense during the initial phases. Fixed and prospective data sources that are not dependent on any one service provider must be used by the system. Furthermore, the model needs to be designed so that it can produce predictions with the highest throughput while requiring the least amount of input. The goal of this study is to meet these demands. The goal of this research is to develop the best model possible for early client churn prediction in a bank. The study was extremely skewed and only employed 10,000 samples, which is a small quantity of data. Real commercial bank data, however, would have considerably more. Both of these headaches can be partially alleviated by oversampling. RF, SVC, LR, XGB, LGBMC, and GBC were all evaluated by the model under various study settings. When oversampling is combined with the RF classifier, a higher percentage of success is obtained (95.74%). The outcome shows that feature selection, or feature reduction, is lowering the tree classifiers' prediction score. Another finding is that oversampling in SVC lowers the score, in contrast to other classifiers. This is due to the asymmetry of the Bank dataset. As a result, SVC is unable to handle the data sufficiently.

DECLARATIONS

Acknowledgement: We appreciate the generous support from all the contributor of research and their different affiliations.
Funding: No funding body in the public, private, or nonprofit sectors provided a particular grant for this research.
Availability of data and material: In the approach, the data sources for the variables are stated.
Authors' contributions: Each author participated equally to the creation of this work.
Conflicts of Interests: The authors declare no conflict of interest.
Consent to Participate: Yes
Consent for publication and Ethical approval: Because this study does not include human or animal data, ethical approval is not required for publication. All authors have given their consent.

REFERENCES

A. K. Ahmad, A. Jafar, and K. Aljoumaa, "Customer churn prediction in telecom using machine learning in big data platform," Journal of Big Data, vol. 6, no. 1, p. 28, 2019.
Agha, A. A., Rashid, A., Rasheed, R., Khan, S., & Khan, U. (2021). Antecedents of customer loyalty in the telecom sector. Turkish Online Journal of Qualitative Inquiry,12(9).
Ahmed AAQ, Maheswari D. Churn prediction on huge telecom data using hybrid firefly based classification. Egypt Inform J 2017;18(3):215–20. https://doi.org/ 10.1016/j.eij.2017.02.002.
Berger P, Kompan M. User modeling for churn prediction in e-commerce. IEEE Intell Syst 2019;34(2):44–52. https://doi.org/10.1109/MIS.2019.2895788. March-April.
Dahouda, M. K., & Joe, I. (2021). A deep-learned embedding technique for categorical feature encoding.IEEE Access,9, 114381-114391.
Dhanawade, A., Mahapatra, B., & Bhatt, A. (2023, November). A smote-based churn prediction system using machine learning techniques. In2023 1st DMIHER International Conference on Artificial Intelligence in Education and Industry 4.0 (IDICAIEI)(Vol. 1, pp. 1-7). IEEE.
E V, Ravikumar P, S C, M SK. An efficient technique for feature selection to predict customer churn in telecom industry. In: Proceedings of the 1st International Conference on Advances in Information Technology (ICAIT); 2019. p. 174–9. https://doi.org/10.1109/ICAIT47043.2019.8987317.
Fei, D. (2021). Chinese telecommunications companies in Ethiopia: The influences of host government intervention and inter-firm competition.The China Quarterly,245, 186-207.

Gaur A, Dubey R. Predicting customer churn prediction in telecom sector using various machine learning techniques. In: Proceedings of the international conference on advanced computation and telecommunication (ICACAT); 2018. p. 1–5. https://doi.org/10.1109/ICACAT.2018.8933783.
Geiler, L., Affeldt, S., & Nadif, M. (2022). A survey on machine learning methods for churn prediction.International Journal of Data Science and Analytics,14(3), 217-242.
Hancock, J. T., & Khoshgoftaar, T. M. (2021). Gradient boosted decision tree algorithms for medicarefraud detection.SN Computer Science,2(4), 268.
Hasan, B., Shaikh, S. A., Khaliq, A., & Nadeem, G. (2024). Data-Driven Decision-Making: Accurate Customer Churn Prediction with Cat-Boost. The Asian Bulletin of Big Data Management, 4(02), Science-4.
J. Raj and V. Ananthi, "Recurrent neural networks and nonlinear prediction in support vector machines," Journal of Soft Computing Paradigm, vol. 2019, pp. 33–40, 09 2019.
Jain H, Khunteta A, Srivastava S. Churn prediction in telecommunication using logistic regression and logit boost. Procedia Comput Sci 2020;167:101–12. https://doi.org/10.1016/j.procs.2020.03.187.
Johnson, J. M., & Khoshgoftaar, T. M. (2022). Encoding high-dimensional procedure codes for healthcare fraud detection.SN Computer Science,3(5), 362.
Keshavarz, H., Mahdzir, A. M., Talebian, H., Jalaliyoon, N., & Ohshima, N. (2021). The value of big data analytics pillars in the telecommunication industry.Sustainability,13(13), 7160.
Lalwani, P., Mishra, M. K., Chadha, J. S., & Sethi, P. (2022). Customer churn prediction system: a machine learning approach.Computing,104(2), 271-294.
Li, J., Bai, X., Xu, Q., & Yang, D. (2023). Identification of Customer Churn Considering Difficult Case Mining.Systems,11(7), 325.
Lottu, O. A., Ezeigweneme, C. A., Olorunsogo, T., & Adegbola, A. (2024). Telecom data analytics: Informed decision-making: A review across Africa and the USA.World Journal of Advanced Research and Reviews,21(1), 1272-1287.
Louk, M. H. L., & Tama, B. A. (2023). Dual-IDS: A bagging-based gradient boosting decision tree model for network anomaly intrusion detection system.Expert Systems with Applications,213, 119030.
Manzoor, A., Qureshi, M. A., Kidney, E., & Longo, L. (2024). A Review of Machine Learning Methods for Customer Churn Prediction and Recommendations for Business Practitioners.IEEE Access.
Mengjing Hao. (2024). WA_Fn-UseC_-Telco-Customer-Churn. IEEE Dataport. https://dx.doi.org/10.21227/0q5y-3529
N.F. Mansor, A.S. Shibghatullah, A.A. Hussin, S.M.Chit, S.M., T. Eldabi, "Multi Agent System (MAS) for Bus Driver Duty Reassignment in the Event of Late for Second Piece of Work (LSPW)", Journal of Physics: Conference Series, Vol. 1529, No. 4, pp. 1-10, 2020.
Nadeem, G., & Anis, M. I. (2024). Investigation of bovine disease and events through machine learning models.Pakistan Journal of Agricultural Research,37(2), 102-114.
Nadeem, G., Ali, S., Hasan, B., Yasin, S., & Hasan, B. (2024). Decoding Bovine Behavior: A Machine Learning Analysis of Disease and Event Detection. The Asian Bulletin of Big Data Management, 4(1), Science-4.
Nadeem, G., Rehman, Y., Khaliq, A., Khalid, H., & Anis, M. I. (2023, March). Artificial Intelligence-basedprediction system for General Medicine. In2023 4th International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)(pp. 1-6). IEEE.
Pamina J, Beschi R, SathyaBama S, Soundarya S, Sruthi MS, Kiruthika S, Aiswaryadevi VJ, Priyanka G. An effective classifier for predicting churn in telecommunication. J Adv Res Dyn Control Syst 2019;11(01–Special Issue). https://ssrn.com/abstract=3399937.
Prabadevi, B., Shalini, R., & Kavitha, B. R. (2023). Customer churning analysis using machine learning algorithms.International Journal of Intelligent Networks,4, 145-154.

Rahman, M. M. (2023). The effect of business intelligence on bank operational efficiency and perceptions of profitability.FinTech,2(1), 99-119.Rahman, K., Ghani, A., Alzahrani, A., Tariq, M. U., & Rahman, A. U. (2023). Pre-Trained Model-Based NFR Classification: Overcoming Limited Data Challenges.IEEE Access.

Routh, P., Roy, A., & Meyer, J. (2021). Estimating customer churn under competing risks.Journal of the Operational Research Society,72(5), 1138-1155.

Saha, L., Tripathy, H. K., Gaber, T., El-Gohary, H., & El-kenawy, E. S. M. (2023). Deep churn prediction method for the telecommunication industry.Sustainability,15(5), 4543.

T. J. Swamidason, "Survey of data mining algorithm's for intelligent computing system," Journal of Trends in Computer Science and Smart Technology, vol. 01, pp. 14–23, 09 2019.

Tirkkonen, T. (2021). Improving Customer Experience with Artificial Intelligence, Data Analysis, and Automation in Azets.

Ullah I, Raza B, Malik AK, Imran M, Islam SU, Kim SW. A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector. IEEE Access 2019;7:60134–49. https://doi.org/10.1109/ACCESS.2019.2914999.

Usman-Hamza, F. E., Balogun, A. O., Nasiru, S. K., Capretz, L. F., Mojeed, H. A., Salihu, S. A., ... & Awotunde, J. B. (2024). Empirical analysis of tree-based classification models for customer churn prediction.Scientific African,23, e02054.

Wang, K., Lu, J., Liu, A., Song, Y., Xiong, L., & Zhang, G. (2022). Elastic gradient boosting decision tree with adaptive iterations for concept drift adaptation.Neurocomputing,491, 288-304.

Z.A. Abas Z.A. Abas, Z.Z. Abidin, A.F.N.A. Rahman, H. Rahmalan, G. Pramudya & M.H.A. Hamid, "Internet of Things and Healthcare Analytics for Better Healthcare Solution: Applications and Challenges", International Journal of Advanced Computer Science and Applications(IJACSA), Vol. 9, No. 9, pp.446-450, 2018.

Z.A. Abas, M.N. Norizan, Z.Z. Abidin, A.F.N.A Rahman, H. Rahmalan, I.H.A. Tharbe, W.F.W.W. Fakhruddin, N.H.M. Zaki, S.A. Sobri, "Modeling Physical Interaction and Understanding Peer Group Learning Dynamics: Graph Analytics Approach Perspective", Mathematics, Vol. 10, No. 9, pp.1-18, 2022.

Zhao, M., Zeng, Q., Chang, M., Tong, Q., & Su, J. (2021). A prediction model of customer churn considering customer value: an empirical research of telecom industry in China.Discrete Dynamics in Nature and Society,2021, 1-12.