Journal of Information Systems Engineering and Management
2025, 10(31s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Optimizing Customer Retention in Banking Through Advanced AI Technologies

Joby Jacob[1], Prof. Dr. E. J Thomson Fredrik[2]
[1] *Research Scholar, Karpagam Academy of Higher Education (Deemed to be University), Coimbatore, India.*
[2] *Research Guide, Karpagam Academy of Higher Education (Deemed to be University), Coimbatore, India.*
[1] *jobyjacob1975@gmail.com*, [2] *thomson500@gmail.com*

ARTICLE INFO

Received: 31 Dec 2024

Revised: 20 Feb 2025

Accepted: 28 Feb 2025

ABSTRACT

This research presents this project, which explores how advanced AI technologies can help optimize customer retention in banking. Customer churn is predicted using historical data using machine learning models such as Random Forest, XGBoost, SVM, Gradient Boosting, KNN, and Naive Bayes respectively. Unlike typical methods of segmentation, in K-means clustering the data is partitioned based on customer behaviour and thus personalized retention strategies are possible. The metrics such as accuracy, precision-recall and F1 score demonstrate that Gradient Boosting performs better than other algorithms. This finding underscores its importance to predictive analytics and to customer segmentation in the banking sector to improve loyalty and reduce churn.

Keywords: Machine Learning, AI, Customer Retention, Churn, SVM, KNN, Naive Bayes, Random Forest, K-means Clustering, XGBoost, Gradient Boosting, Tenure, Monthly Charges, Total Charges.

1. INTRODUCTION

Background

In the highly competitive and highly customer acquisition cost banking industry, customer retention is a critical factor. The bank can retain its brand loyalty and long-term profitability and maintain a stable revenue stream by gaining existing customers. Since customer interactions are diverse and dynamic, it is challenging to predict customer behaviour and identify at-risk customers. These challenges have become excellent problems to be solved with Machine Learning and/or a Data Driven approach (Yalamati, S., 2023) [1]. AI enables banks to analyse big data, find patterns, predict churn, and then personalize retention strategies. Through Predictive Modelling and Customer Segmentation, financial institutions are able to proactively intervene and stop customers from churning.

Machine learning models such as Support Vector Machines (SVM), Random Forest, XGBoost, Gradient Boosting etc. provide robust answers in churn prediction and give actionable insights into customer behaviour; which helps to identify customer behaviour for targeted marketing and service improvements. In this study (Babakhanian, M.R., Amin Mousavi, S.A., Soltani, R. and Vakilifar, H.R., 2023) [2], the author used AI technologies to optimize customer retention in banking. The research integrates segmentation and predictive analytics to provide banks with better decision-making opportunities, cut down bank churn rates, and make them valuable and competitive in a fast-evolving financial environment.

Aim and Objectives

Aim

The aim of this study is to enhance customer loyalty in banking with the help of predictive analytics models together with customer segmentation.

Objectives

- To generate a machine learning framework for customer churn (Support Vector Machine, Random Forest, XGBoost) to help the banks reveal at-risk customers from historical data.

- To utilize machinery such as the K-means algorithm for clustering customers into segments based on their activity and further pursue retention techniques that focus on the behaviour and interests of individual segments.
- To compare the features of the developed predictive models and clustering techniques in forecasting customer churn and preferences and better comprehension of customers.
- To give recommendations and focused retention models on the basis of the studies, enabling the banking institutions to apply effective retention approaches for enhancing customer loyalty and minimizing churn rate.

Rationale

Banking profitability depends on customer retention, which is cheaper than acquisition. Using advanced AI technologies such as predictive modelling and clustering, banks can identify churn risks and personalize strategies to retain customers, enhancing customer loyalty, lowering churn rates and strengthening competitive advantage in the sphere of banking.

II. LITERATURE REVIEW

2.1: Applications of AI in customer behavior analysis and predictive modeling for retention

Artificial intelligence (AI) represents a breakthrough technology in customer behavior analytics which delivers businesses enhanced tools for retention strategies. The massive amount of data which AI-driven analysis handles allows businesses to identify buying tendencies that traditional customer behavior research would fail to detect. Machine learning algorithms allow businesses to form more nuanced customer need comprehension as they analyze customer behaviors in relationship to preferences and historical transactions to develop clear consumer segments.

The principle that predictive modeling is the foundation in customer retention business operations is what makes FPS run seamlessly. Past data records are used in the analytical process to create predictions for future steps like customer dissolution rates. Decision trees, neural networks and support vector machines are needed, machine learning models, to get accurate predictions of behaviour of the customer. Businesses can use predictive models to send solutions that prevent client attrition through personalization by knocking down churn signals before customers decide to leave.

By combining natural language processing (NLP) with sentiment analysis, this breaks down yet further to better analyze customer behavior by extracting sentiments from the insight gained from unstructured data sources – namely customer reviews, customer emails and social media engagements. In real time, these tools show business pain points with customer satisfaction measurements, hence tracking customer satisfaction on real time. By merging data analysis of consumer purchase behavior with predictions of future-behavior, businesses can through AI platforms optimize their pricing models and marketing approach.

Information processing in the moment allows businesses to quickly and confidently make decisions. With AI dashboards and analytics tools, companies can immediately track their key customer metrics and see where they might have a shot at retaining customers. AI recommendation solutions lead to personalized product suggestion systems that improve customer satisfaction levels and improve customer relationship with businesses.

2.2: Role of advanced algorithms in real-time churn detection and risk assessment

Advanced algorithms are of vital importance for real time churn detection and risk assessment, which means that such businesses can actually find and address customer attrition before it takes place. Widespread customer data analysis algorithm using machine learning (ML) and artificial intelligence (AI), with functions to merge together to build patterns revealing negative acts indicative of future customer churn. Transaction history analysis and behavior data assessment allow organizations to identify the existing customer signs and then apply prevention steps.

Supervised learning techniques as well as logistic regression and gradient boosting enable predictability of churn indicators after analyzing past evidence. The algorithms create risk scores for individual customers to inform business priorities for retention measures. Unsupervised learning clustering methods split risk-profiled customer segments for targeted preventive care measures.

Recipients benefit from real-time data processing features that trigger immediate operational responses from observing present customer behaviors. The AI platforms use embedded algorithms to track client interactions that span different channels which include mobile applications websites and customer service documentation. Organizations gain the ability to detect rapid technological changes through usage frequency declines or adverse feedback in support tickets together with rapid response capabilities (Maseke, B.F., 2024) [4].

Risk assessment models implement survival analysis and reinforcement learning to forecast customer lifetime value and measure how exit events affect organizations. The tools direct organizations to implement budget-friendly retention systems while devoting their resources to profitable clientele.

Organizations now leverage advanced algorithms to detect risks of customer churn and simultaneously reveal its fundamental causes. The predictions generated through these models allow businesses to create customized retention measures which include specific discounts alongside loyalty programs and improved customer assistance. Real-time evaluation of customer churn combined with advanced algorithms drives essential business performance in competitive markets to maintain valuable customer relationships and boost long-term success.

2.3: Integration of AI-driven personalization in banking services to improve customer engagement

Integration of AI driven personalization in banking services has reduced customer engagement significantly and increased it rather through delivering more personal experience to the customer on the basis of their preference and need. Having large volumes of customer data, transactional history, spending patterns, online interaction at its disposal, AI powered systems process and analyze vast amounts of customer data to provide extremely customized banking solutions.

A hallmark of AI in banking is personalized recommendations. With the help of machine learning algorithms, banks can often suggest financial products (such as credit cards, loans or investment plans) that correspond with a customer's financial aspirations. For example, recommendation engines recommend appropriate savings plans for customers based on their income and expenditure and only advise what is relevant and actionable.

Natural language processing (NLP) has made banking customer service revolutionized. Instant personalized responses to customer inquiries improve service and access for customers, and by bringing speed of service and access to the myriad of human customer issues, these AI tools improve customer satisfaction (Madasamy, S. and Aquilanz, L.L.C., 2023) [5]. However, by using conversational AI, customers receive real time financial/advice, account updates as well as support on transactions, making them feel they are getting individual attention.

Personalization doesn't stop there: AI driven marketing campaigns segment customers by targeting specific types of promotions. Predictive analytics helps identify what channels, when and what content to run marketing messages on, making sure the content is relevant to the audience they are targeting. For instance, the more relevant the communication is for a specific user, the more the communications would be personalized, e.g., personalized notifications about loan approvals or interest rate changes.

Banks with the help of AI, can able to predict their customers' behavioral analysis. Say an anomaly detection algorithm sees those unusual transactions, then it will send up security alerts and subsequently increase that trust. Like financial planning, AI systems can also predict life events such as marriage or retirement and tailor your own financial planning before the big day.

AI based personalization in banking turns customer engagement to bespoke services and experience in a nutshell. In addition to this, this technology is more than improving customer service – it fosters long term loyalty, and informs banks to be perceived as trusted financial partners.

2.4: Evaluation of AI-based customer segmentation for tailored retention strategies

Customer segmentation is an AI based tool that is very important in crafting tailored retention strategies for better customer satisfaction and customer loyalty. In contrast to traditional segmentation, traditional uses of traditional methods, such as demographic and behavioral data, drive a segmentation into self-evident groups.

There are algorithms such as k-means, hierarchical clustering and density based spatial clustering of applications with noise (DBSCAN), which help businesses cluster customers based on their shared characteristics. These two models allow businesses to unveil hidden patterns of customer behavior while absorbing spending patterns, product preference, and customer engagement frequency.

Even more so, deep learning models take that given segmentation and then search data that is unstructured in some way, which is often social media interactions, customer reviews, or web activity. With these advanced techniques we can learn what customers prefer, why they prefer this or that, and what motivates them in order to shape retention on the basis of customers' needs (Adeniran, I.A., Efunniyi, C.P., Osundare, O.S., Abhulimen, A.O. and OneAdvanced, U., 2024) [6].

Through supervised learning models, predictive segmentation is used to predict customer behavior and predicted churn risk across each segment. Integrating historical data with predictive analytics allows the business to identify segments that need immediate attention, and put resources where it matters the most. For example, customers on the receiving end of exclusive offers or loyalty programs will relate towards those as high value customers, creating a bit of dissatisfaction towards the company to begin with.

AI based segmentation improves the efficacy of both marketing and communication strategies. Messaging is created on the basis of tailored campaigns and customized messages to arrive at something that would click with specific demographics. In turn, retention across demographics hinges on patient vs. always-on pricing strategies like premium services to high income groups or promotional discounts to price sensitive customers.

III. METHODOLOGY

Research Philosophy

The positivist research philosophy is followed in this study which mainly looks at data-driven analysis and objective evaluation. The research uses quantitative methods, such as machine learning algorithms and clustering techniques to derive patterns in historical customer data to predict churn as well as optimize the retention strategy to have reliable, reproducible and actionable insight for the banking sector.

Research Approach and Design

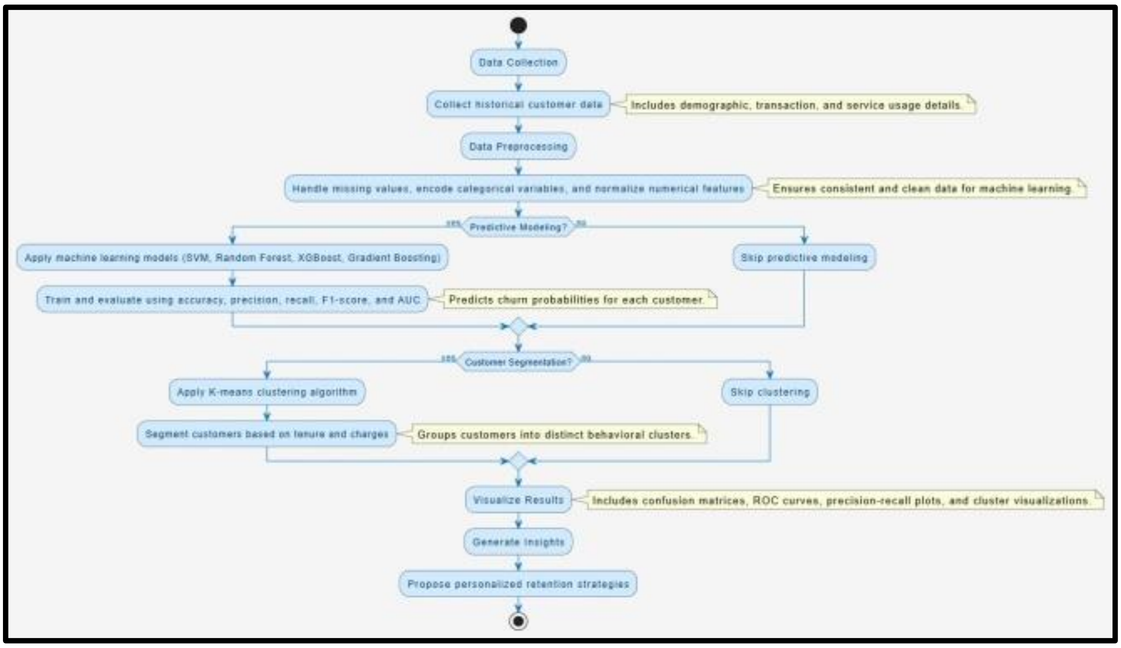

Fig. 1: Design Flowchart

The underlying research in this study employs a quantitative research approach utilizing data-driven techniques to analyse customer behaviour for churn (Elias, O., Esebre, S.D., Abijo, I., Timothy, A.M., Babayemi, T.D., Makinde, E.O., Oladepo, O.I. and Fatoki, I.E., 2024) [7]. For churn prediction, machine learning algorithms such as SVM, Random Forest, XGBoost, KNN, Naive Bayes and Gradient Boosting are used and for segmentation, K-means clustering is used. The research follows a structured process: model development, evaluation, visualization, and data preprocessing. Categorical features undergo encoding numerical values get normalized and the history of customer data is prepared. Metrics such as accuracy, precision, recall, F1 score, etc. are used to train and evaluate the models. Targeted personalized customer retention strategies based on clustering results are informed.

Data Analysis and Collection Techniques

The research uses secondary data from historical customer records at banking, including demographic, transactional and service use information from the following sources such as Kaggle, Yahoo, etc. In order to build predictive models and cluster customer segments, features such as tenure, monthly charges, contract type and churn status are mainly focused here (Reddy, S.R.B., 2021) [8]. Preprocessing for data is the first step in the analysis, consisting of dealing with missing values, encoding categorical variables, and normalizing numerical features to obtain consistency. Some machine learning algorithms (SVM, Random Forest, XGBoost, gradient boosting, KNN, Naive Bayes) to predict customer churn are implemented. Model performance is assessed using evaluation metrics, such as accuracy, precision, recall, and F1-score (Al Faisal, N., Nahar, J., Waliullah, M. and Borna, R.S., 2024) [9]. The K-means clustering algorithm is applied to segment the customers according to behavioural attributes such as tenure and monthly charges for customer segmentation. Model performance and clustering insights are highlighted with visualizations such as distribution plots, ROC curves, precision-recall plots, and confusion matrices. Banks use the techniques to identify at-risk customers and build data-driven retention strategies.

Model Implementation

In this study, a structured workflow of implementing machine learning models is followed to predict customer churn and achieve better retention strategies. In the first instance, the preprocessed dataset is split into training and testing sets to avoid biased model evaluation. Previous work includes encoded categorical variables and normalized numerical features (Adewumi, A., Ewim, S.E., Sam-Bulya, N.J. and Ajani, O.B., 2024) [10]. Support Vector Machines (SVM), Random Forest, XGBoost and Gradient Boosting are churn prediction models. Each model is trained with the training data to detect at-risk customers.

The Kmeans clustering algorithm is used to segment customers according to attributes such as tenure, monthly charges and total charges (Adeoye, O.B., Okoye, C.C., Ofodile, O.C., Odeyemi, O., Addy, W.A. and Ajayi-Nifise, A.O., 2024) [11]. Behavioural patterns are visualised in clusters to provide an opportunity for targeted retention. For implementing the model, libraries such as Scikit-learn and XGBoost for machine learning, Matplotlib and Seaborn for data visualization have been used. Model performance is optimised via automated hyperparameter tuning. The study combines churn prediction and clustering to finally achieve the integrated development of customer-centric personalized retention strategies to reduce churn rates and increase banking profitability.

Ethical Consideration

This study ensures ethical compliance by maintaining data privacy and confidentiality. The data about the customers is anonymized and the analysis is pursued following the applicable data protection regulations (GDPR, etc.). It does not use any personally identifiable information (PII). The research is only concerned with patterns and behaviours and pays no mind to bias in predictive modelling and segmentation.

IV. RESULTS AND DISCUSSION

Introduction

The results and discussion section includes the performance of the machine learning models such as SVM, Random Forest, XGBoost, Gradient Boosting, KNN and Naive Bayes for predicting customer churn. Metrics such as accuracy, recall, precision, F1 score and Confusion matrix visualisation are used to compare the predictive effectiveness and segmentation insight using sectional.

Result Analysis

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechSupport
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	No	No
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	Yes	No
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	No	No
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	Yes	Yes
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	No	No

5 rows × 21 columns

Fig. 2: First Few Rows of The Dataset

Essential customer data such as their gender, seniority, partner status, tenure, phone service, and other details are shown in this figure from the initial rows of the dataset. This data is used as the basis for further analysis to explain the demographics of the customers and attributes of the service before preprocessing.

	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport
0	0	0	1	0	-1.277445	0	1	0	0	2	0	0
1	1	0	0	0	0.066327	1	0	0	2	0	2	0
2	1	0	0	0	-1.236724	1	0	0	2	2	0	0
3	1	0	0	0	0.514251	0	1	0	2	0	2	2
4	0	0	0	0	-1.236724	1	0	1	0	0	0	0

Fig. 3: Pre-processed Customer Churn Data

Once the preprocessing is complete, the pre-processed data is shown which consists of categorical features encoded into numerical values and the data is ready for machine learning. Both the 'tenure' and service-related columns are normalized for consistent model input (Umamaheswari, S. and Valarmathi, A., 2023) [12]. This step is critical, to avoid inconsistencies in the data, and to ensure that the predictions are as accurate as possible.

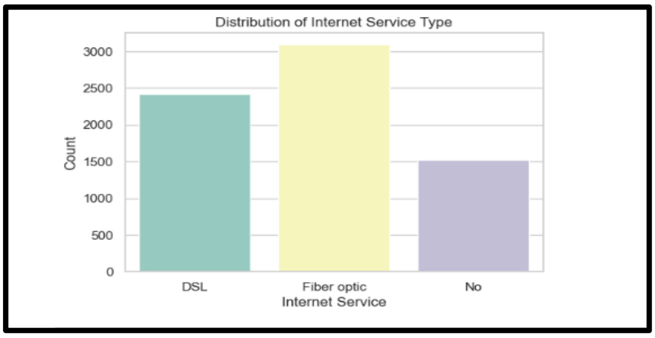

Fig. 4: Distribution of Internet Service Type

The distribution of internet service types amongst customers using this method is represented in this figure. However, most customers are using fibre optic services and DS is second. Some customers don't have internet service. This is useful to understand the behaviour of a customer and to segment the customer based on Internet service preferences.

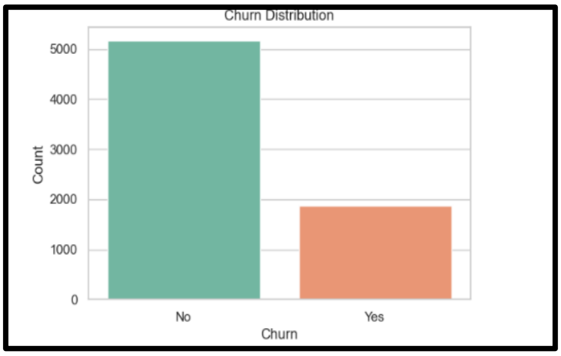

Fig. 5: Churn Distribution

A very unbalanced churn distribution, with more non-churning customers than churning ones. In churn prediction tasks, this imbalance is typical and needs oversampling or cost-sensitive learning to avoid performance degradation or overly extreme class discrimination.

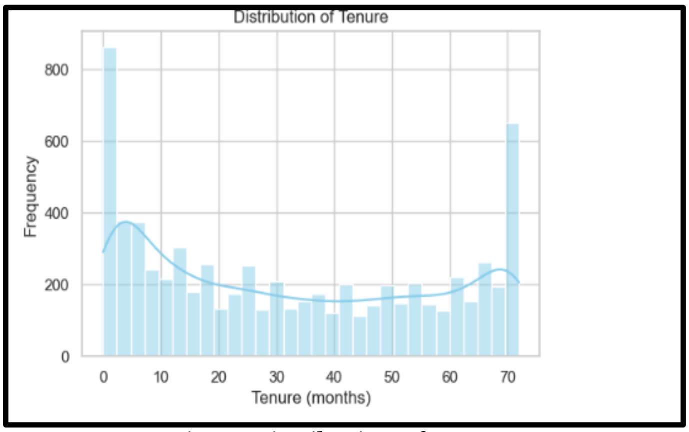

Fig. 6: Distribution of Tenure

The distribution of customer tenure in months is shown in this figure. Customers are highly concentrated (customer churners) or not (loyal customers) by tenure, and there are few customers with intermediate tenure among the two (Hassan, M., Aziz, L.A.R. and Andriansyah, Y., 2023) [13]. This distribution implies that retention must be focused on as much at the early churners as at the long-term customers.

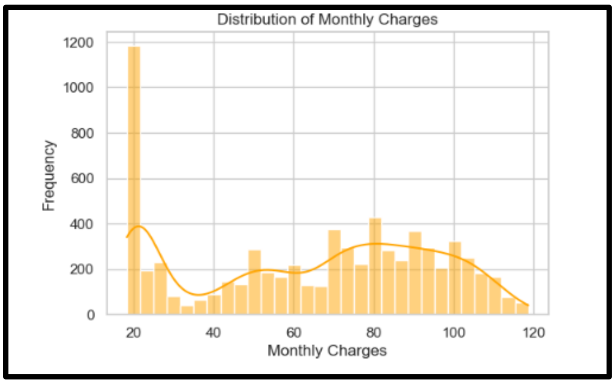

Fig. 7: Distribution of Monthly Charges

This is a visual representation of how monthly charges are distributed to customers. It is observed that there are peaks at smaller values along with charges spread at larger values. This distribution helps analyse customer spending patterns that are used for churn prediction and segmentation strategies to improve customer retention.

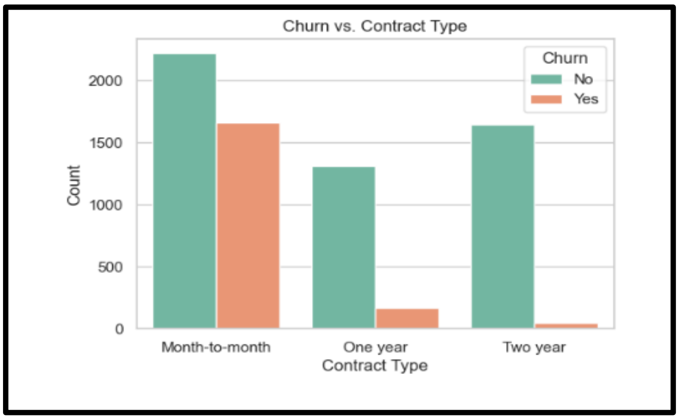

Fig. 8: Churn vs Contract Type Plot

This plot displays the way contract type changes customer churn. Churn is highest for month-to-month contracts and lowest for contracts of one year or longer (2 years). It implies that the banking business benefits from offering longer-term contracts given the competitive nature of the industry.

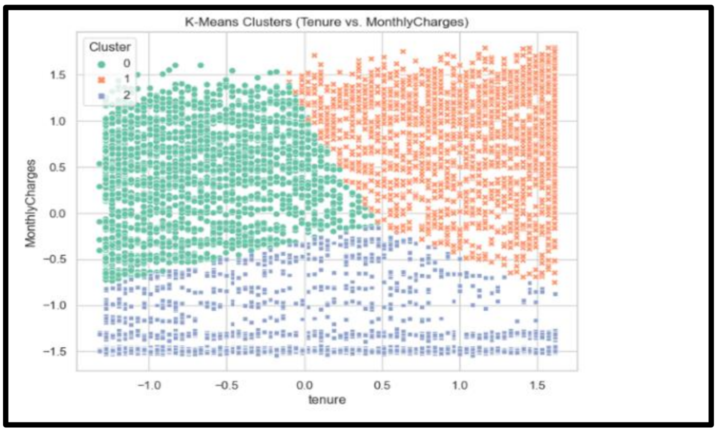

Fig. 9: Cluster Plot of Tenure vs Monthly Charges

K-means clustering results based on tenure and monthly charges for customer segments are shown in the figure. Three clusters are identified, revealing distinct groups: customers with lower charges and shorter tenure, those with higher charges and longer tenure, and another group with mixed characteristics (Tulcanaza-Prieto, A.B., Cortez-Ordoñez, A. and Lee, C.W., 2023) [14]. Personalized retention strategies are then aided through this segmentation.

	Model	Precision (Churn)	Recall (Churn)	F1-Score (Churn)	Accuracy
0	Random Forest	0.633898	0.500000	0.559043	0.790632
1	XGBoost	0.606250	0.518717	0.559078	0.782825
2	SVM	0.612308	0.532086	0.569385	0.786373

Fig. 10: Initial Models Evaluation Metrics

This figure shows the evaluation metrics (precision, recall, F1-score, and accuracy) for the initial models: Random Forest, XGBoost, and SVM. Finally, these metrics are used to compare how well models predict the probability of customer churn and Random Forest is satisfactory here with an accuracy of 79%. All models work well, however, precision and recall are crucial on imbalanced datasets such as all churn prediction tasks.

Cluster	tenure	MonthlyCharges	TotalCharges	Count
0	-0.777921	0.342727	-0.549953	2680
1	1.066919	0.827932	1.308572	2200
2	-0.121310	-1.266739	-0.649553	2163

Fig. 11: Clustering Summary

This figure illustrates the Kmeans clustering results; an average measure of tenure, monthly charges, and total charges for three clusters. It also provides the number of customers in each segment (Bhuiyan, M.S., 2024) [15]. The information about this clustering of customers is useful to understand how customers behave, and how retention strategies are moulded for each of the unique characteristics of the clusters.

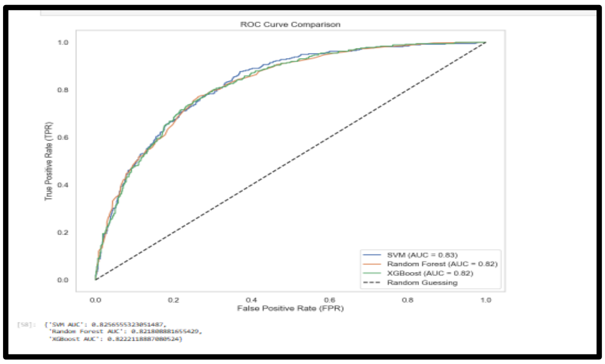

Fig. 12: ROC Curve Comparison of The Models

SVM, Random Forest and XGBoost models are compared using an ROC curve plot. Similar AUC scores (AUC = 0.82) are attained by XGBoost and Random Forest and just outperform SVM (AUC = 0.83). True positive rates are shown to be very strong and the models are demonstrated to be good at separating churning from non-churning customers.

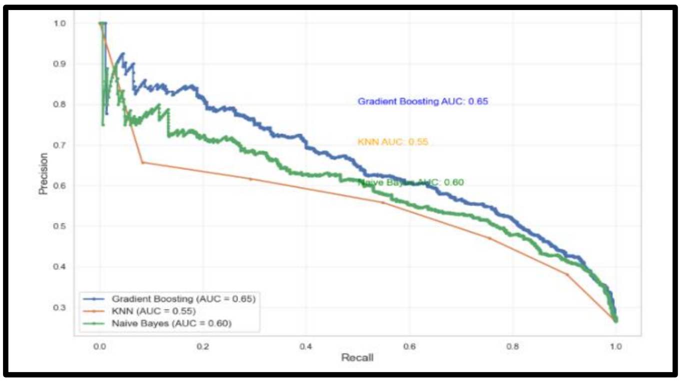

Fig. 13: Precision-Recall Curve of The Additional Models

Gradient Boosting, KNN and Naive Bayes models are evaluated by the Precision-Recall curve. Naive Bayes offers the next highest AUC of 0.60, followed by KNN (0.55), and Gradient Boosting surpassing them both with an AUC of 0.65. Gradient Boosting beats KNN in the curves in terms of precision-recall trade-off, and the performance of KNN plummets with increasing recall in figures, from which one can see that model sensitivity varies.

	Model	Precision (Churn)	Recall (Churn)	F1-Score (Churn)	Accuracy
0	Gradient Boosting	0.653302	0.493761	0.562437	0.796025
1	KNN	0.557971	0.549020	0.553459	0.764789
2	Naive Bayes	0.515971	0.748663	0.610909	0.746805

Fig. 14: Evaluation Metrics of Additional Models

The details about precision, recall, F1 score and accuracy for Gradient Boosting, KNN and Naive Bayes are shown in this table. At most metrics, Gradient Boosting reaches success with 79.60% accuracy (Ogundipe, D.O., Odejide, O.A. and Edunjobi, T.E., 2024) [16]. Accuracy and F1 scores respectively of KNN and Naive Bayes are comparable with each other, indicating a competitive performance between the two despite their difference in underlying methodology.

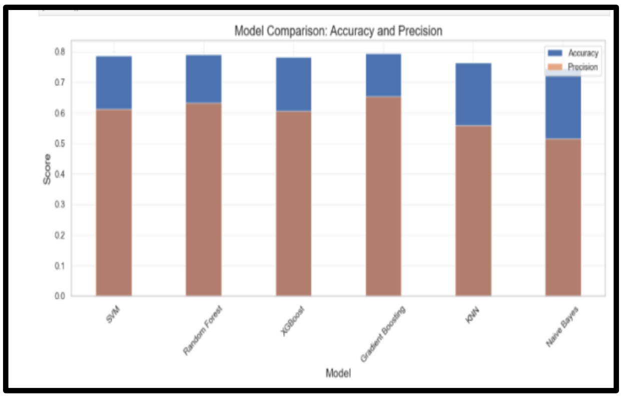

Fig. 15: Comparison of All Models

Accuracy and precision scores from all models (SVM, Random Forest, XGboost, Gradient boosting, KNN and Naive Bayes) are visualised in a bar plot. Accuracy scores are strong across all models again and all models do well on precision except the Gradient Boosting and the Random Forest, which score the highest precision.

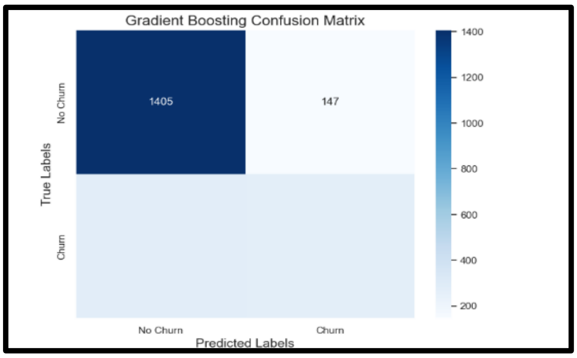

Fig. 16: Gradient Boosting Confusion Matrix

The evaluation of Gradient Boosting prediction is this confusion matrix with 1,405 non-churning and 147 churning customers. Hence the model is able to correctly classify 1,405 non-churning and 147 churning customers with less than 5% misclassification (Modak, C., Ghosh, S.K., Sarkar, M.A.I., Sharif, M.K., Arif, M., Bhuiyan, M., Ahmed, M.P., Pabel, M.A.H. and Devi, S., 2024) [17]. Gradient Boosting's good performance in separating classes is shown in the matrix and supports its very high accuracy and F1 score metrics in determining whether a customer is going to churn or not.

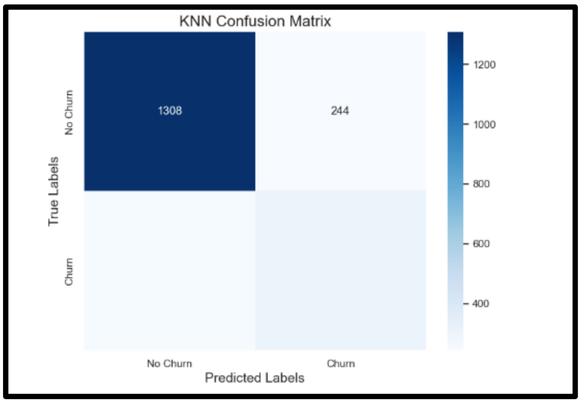

Fig. 17: KNN Confusion Matrix

KNN's confusion matrix discloses it classifies 244 churning and 1,308 non-churning customers correctly. Precision and recall are adversely influenced by higher misclassifications than Gradient Boosting. However, KNN performs reasonably well for churn; and as a result, it is a suitable model for simple implementations.

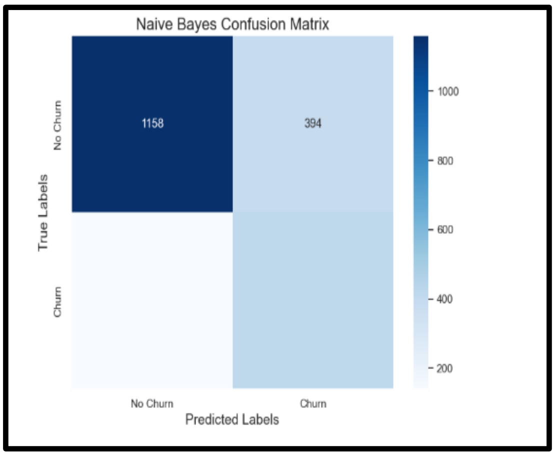

Fig. 18: Naive Bayes Confusion Matrix

The confusion matrix for Naive Bayes shows that most of them wrongly classified the highest 1,158 non-churning and 394 churning customers. The results demonstrate how Naive Bayes is prone to endorsing its assumptions since it deals with imbalanced datasets, yet with incredibly high recall it demonstrates its merit in finding churning customers.

Key Findings and Discussion

Model	Precision (Churn)	Recall (Churn)	F1-Score (Churn)	Accuracy
Random Forest	0.633898	0.500000	0.559043	0.790632
XGBoost	0.606250	0.518717	0.559078	0.782825
SVM	0.612308	0.532086	0.569385	0.786373
Gradient Boosting	0.653302	0.493761	0.562437	0.796025
KNN	0.557971	0.549020	0.553459	0.764789
Naive Bayes	0.515971	0.748663	0.610909	0.746805

Table 1: Model Evaluation Summary

It focuses on the simulations of underlying four machine learning models to provide significant insights into customer churn prediction. Random Forest lost to Gradient Boosting, which achieved the highest overall accuracy (79.60%) and precision (65.33%) over Random Forest, SVM and XGBoost. KNN and Naive Bayes perform well compared to many traditional classification methods based on recall, however, their lower precision demonstrates that these are unable to accurately predict non-churning customers (Gurung, N., Hasan, M.R., Gazi, M.S. and Chowdhury, F.R., 2024) [18]. Actionable insights are provided for personalized retention strategies based on three distinct customer segments uncovered by K-means clustering according to tenure and monthly charges. Churn rates are highest for month-to-month contract customers indicating the need to promote longer-term contracts to reduce churn rates.

The ROC curve and the Precision-Recall curve have validated Gradient Boosting's better predictive performance in responding to imbalanced datasets. Misclassification analyses using a confusion matrix have indicated that Naive Bayes has done properly in the recall, however, suffered in precision. In general, this project is evidence of the key role that advanced machine learning and clustering techniques play in developing effective customer retention strategies.

V. CONCLUSION

Summary of Findings

The effectiveness of the machine learning and clustering techniques in the prediction of customer churn and the recommendation for retention strategies in this study. SVM, Random Forest, and XGBoost are all outperformed with accuracy (79.60%) and precision by Gradient Boosting. Among tenure and charges, this created three distinct customers for which different strategies are created (Kolasani, S., 2023) [19]. Insights into what made customers churn revealed that month-to-month contract customers have the highest churn rates, and highlight the importance of long-term contracts. Accuracy, precision, recall, and AUC are used to evaluate models to provide robust predictions. The findings generally reflect the importance of advanced AI in banking customer retention.

Research Recommendation

Based on the results, the recommendation is made to expand customer retention in banking. The first is to prioritise banks' use of advanced machine learning models such as Gradient Boosting and XGBoost to predict their risk of churn in real-time. It is shown how these models are able to identify at-risk customers better than anyone else and intervene in time enough to reduce churn rates. Second, K means clustering of customer segmentation made out of behaviour patterns. Such insights must further allow banks to create individualized retention strategies for each segment including the provision of tailored incentives, loyalty programs, or promotional packages for these high-risk customers.

Third, the analysis shows that customers with month-to-month contracts are most likely to churn. The main point of this is efforts made to garner customer loyalty and retention as banks advertise for long-term contracts and extend discounts or extra benefits. Moreover, the combination of customer feedback with a predictive model enhances churn prediction (Nwosu, N.T., Babatunde, S.O. and Ijomah, T., 2024) [20]. Therefore, along with complementary qualitative data sources such as customer complaints and survey responses, machine learning insights are required. Another thing is when using such AI technologies to be sure that data is safe and private. Data protection regulations are combined (GDPR), and ethical issues are considered for the sake of customer trust. Future research investigates further expansion of deep learning models and advanced sentiment analysis algorithms to open further holds retention strategies can embrace.

Future Work

The future research is to combine deep learning models and neural networks to increase the churn prediction accuracy. In addition, real-time data streams, and advanced sentiment analysis are added to increase customer insights. Finally, retention strategies are proposed and validated and extended to other datasets and banking sectors.

VI. REFERENCES

[1] Yalamati, S., 2023. Revolutionizing Digital Banking: Unleashing the Power of Artificial Intelligence for Enhanced Customer Acquisition, Retention, and Engagement. International Journal of Management Education for Sustainable Development, 6(6), pp.1-20.

[2] Babakhanian, M.R., Amin Mousavi, S.A., Soltani, R. and Vakilifar, H.R., 2023. Survey the effect of fintech companies' profitability enhancement on winning customers' loyalty using an artificial intelligence-based optimization algorithm. International Journal of Nonlinear Analysis and Applications, 14(1), pp.2409-2423.

[3] Devan, M., Prakash, S. and Jangoan, S., 2023. Predictive maintenance in banking: leveraging AI for real-time data analytics. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(2), pp.483-490.

[4] Maseke, B.F., 2024. The transformative power of artificial intelligence in banking client service. South Asian Journal of Social Studies and Economics, 21(3), pp.93-105.

[5] Madasamy, S. and Aquilanz, L.L.C., 2023. THE EVOLUTION OF CHATBOTS: CLOUD AND AI SYNERGY IN BANKING CUSTOMER INTERACTIONS. Journal of Emerging Technologies and Innovative Research, 10.

[6] Adeniran, I.A., Efunniyi, C.P., Osundare, O.S., Abhulimen, A.O. and OneAdvanced, U., 2024. Implementing machine learning techniques for customer retention and churn prediction in telecommunications. Computer Science & IT Research Journal, 5(8).

[7] Elias, O., Esebre, S.D., Abijo, I., Timothy, A.M., Babayemi, T.D., Makinde, E.O., Oladepo, O.I. and Fatoki, I.E., 2024. Harnessing artificial intelligence to optimize financial technologies for achieving sustainable development goals. World J. Adv. Res. Rev, 23, pp.616-625.

[8] Reddy, S.R.B., 2021. Predictive Analytics in Customer Relationship Management: Utilizing Big Data and AI to Drive Personalized Marketing Strategies. Australian Journal of Machine Learning Research & Applications, 1(1), pp.1-12.

[9] Al Faisal, N., Nahar, J., Waliullah, M. and Borna, R.S., 2024. The Role Of Digital Banking Features In Bank Selection An Analysis Of Customer Preferences For Online And Mobile Banking. Frontiers in Applied Engineering and Technology, 1(01), pp.41-58.

[10] Adewumi, A., Ewim, S.E., Sam-Bulya, N.J. and Ajani, O.B., 2024. Advancing business performance through data-driven process automation: A case study of digital transformation in the banking sector. International Journal of Multidisciplinary Research Updates, 8(02).

[11] Adeoye, O.B., Okoye, C.C., Ofodile, O.C., Odeyemi, O., Addy, W.A. and Ajayi-Nifise, A.O., 2024. Integrating artificial intelligence in personalized insurance products: a pathway to enhanced customer engagement. International Journal of Management & Entrepreneurship Research, 6(3), pp.502-511.

[12] Umamaheswari, S. and Valarmathi, A., 2023. Role of artificial intelligence in the banking sector. Journal of Survey in Fisheries Sciences, 10(4S), pp.2841-2849.

[13] Hassan, M., Aziz, L.A.R. and Andriansyah, Y., 2023. The role artificial intelligence in modern banking: an exploration of AI-driven approaches for enhanced fraud prevention, risk management, and regulatory compliance. Reviews of Contemporary Business Analytics, 6(1), pp.110-132.

[14] Tulcanaza-Prieto, A.B., Cortez-Ordoñez, A. and Lee, C.W., 2023. Influence of customer perception factors on AI-enabled customer experience in the Ecuadorian banking environment. Sustainability, 15(16), p.12441.

[15] Bhuiyan, M.S., 2024. The role of AI-Enhanced personalization in customer experiences. Journal of Computer Science and Technology Studies, 6(1), pp.162-169.

[16] Ogundipe, D.O., Odejide, O.A. and Edunjobi, T.E., 2024. Agile methodologies in digital banking: Theoretical underpinnings and implications for customer satisfaction. Open Access Research Journal of Science and Technology, 10(2), pp.021-030.

[17] Modak, C., Ghosh, S.K., Sarkar, M.A.I., Sharif, M.K., Arif, M., Bhuiyan, M., Ahmed, M.P., Pabel, M.A.H. and Devi, S., 2024. Machine Learning Model in Digital Marketing Strategies for Customer Behavior:

Harnessing CNNs for Enhanced Customer Satisfaction and Strategic Decision-Making. Journal of Economics, Finance and Accounting Studies, 6(3), pp.178-186.

[18] Gurung, N., Hasan, M.R., Gazi, M.S. and Chowdhury, F.R., 2024. AI-Based Customer Churn Prediction Model for Business Markets in the USA: Exploring the Use of AI and Machine Learning Technologies in Preventing Customer Churn. Journal of Computer Science and Technology Studies, 6(2), pp.19-29.

[19] Kolasani, S., 2023. Optimizing natural language processing, large language models (LLMs) for efficient customer service, and hyper-personalization to enable sustainable growth and revenue. Transactions on Latest Trends in Artificial Intelligence, 4(4).

[20] Nwosu, N.T., Babatunde, S.O. and Ijomah, T., 2024. Enhancing customer experience and market penetration through advanced data analytics in the health industry. World Journal of Advanced Research and Reviews, 22(3),pp.1157-1170.