Predicting Customer Churn in the Banking Sector Using Machine Learning Classification Models

1st Ogweno Washington Odhiambo
Department of Computer Science
Makerere University
Kampala, Uganda
2400721955
2024/HD05/21955E
odhiambo64@yahoo.com

2nd Ochalo Deogratius
Makerere University
Department of Computer Science
Kampala, Uganda
2400726060
2023/HD05/20127U
deogoch@yahoo.com

***Abstract*—Customer churn poses a critical challenge for modern retail banking, due to increased competition, innovative products and delivery channels. Predicting churn early allows retail banks to take proactive steps in retaining their valuable customers, enhancing satisfaction and sustainable profitability. This study leverages Retail Banking Demo Data [1] by Petrocelli that have been stitched together from real-world Core banking data sources, to develop predictive machine learning model that can positively identify customers at risk of churn. Using feature engineering we compared performance of five popular predictive models: Support Vector Machine (SVM), logistic regression, decision trees, gradient boosting and random forest. Our results showed that random forest model outperformed others, achieving an impressive accuracy and precision of 99.79%, recall of 99.59%, F1-score of 99.69% and AUC-ROC of 100%. We then derived reports inform of Confusion matrix, Classification Report, ROC Curve and AUC, Precision-Recall Curve, F1 Score, Cross-Validation Scores and Learning Curve. The findings from this study offers actionable insights for banks aiming to implement sustainable data-driven customer retention strategies.**

***Index Terms*—customer churn, core banking, machine learning, feature engineering, random forest.**

I. Introduction

Customer churn prediction has become a vital area of focus for businesses in highly competitive service sector such as retail banking. Churn, defined as the loss of customers who stops using products or services, poses a significant challenge to business growth and profitability. Accurately predicting which customers are more likely to churn enables banks to implement target retention strategies, lowering customer acquisition costs, maintain steady revenue streams, and design innovative products and services enhancing profitability (Kassem et al., 2020). This project aims to explore customer churn prediction

through machine learning models, using dataset obtained from Retail Banking Demo Data [1]. The dataset includes a diverse set of attributes, such as customer demographics, accounts information, transaction behaviors, and historical interactions which provides valuable insights for assessing the likelihood of churn. Leveraging predictive analytics powered by Artificial

Intelligence (AI), this project analyzed historical customer transactional and interactions to forecast potential churners, using various machine learning models to deliver accurate and actionable predictions. By employing AI-driven approaches, this analysis enables banks to proactively identify churners, improve retention strategies positively enhancing competitive edge (Suh, 2023).

II. BACKGROUND AND MOTIVATION

A. Maintaining the Integrity of the Specifications

Customer retention is a cornerstone of success in the banking sector, where maintaining long-term relationships with customers significantly impacts profitability and competitiveness. In this context, customer churn, or the loss of clients to competitors, poses a major challenge. Studies indicate that acquiring a new customer can cost five times more than retaining an existing one (Dawes & Swailes, 1999), thus emphasizing on the critical importance of proactive churn management strategies.

The rise of digital banking, increased competition, and evolving customer expectations have further heightened the need for effective churn prevention strategies. According to Uganda bankers Association, 12.5% of the banking population have access to 23 commercial banks with multiple banking products and services, making it easier for customers to switch from one bank to another [5][6].

Traditional survey methods often fail to capture complex customer [7] patterns giving arise to adoption of machine learning models as a powerful solution. In this project, we focus on the application of machine learning-based classification models for customer churn prediction in the

Identify applicable funding agency here. If none, delete this.

retail banking sector, using core banking equivalent dataset. With these insights, banks can not only identify high-risk customers but also develop personalized interventions, improving customer satisfaction, and reduce overall churn rates.

The motivation for this study lies in the potential use of machine learning to transform customer relationship management in banking sector, increasing competitive edge in a fast-changing industry landscape (Kassem et al., 2020).

III. LITERATURE REVIEW

Several studies have explored the application of machine learning (ML) techniques in customer churn predictions with early focus been on telecom, later financial, Retail E-commerce and Subscription Services.

Telecom was among the first industries to adopt churn prediction due to high customer turnover and competitive markets. Models such as decision trees and regression models demonstrated the impact of ML algorithms on predicting customer churn with high accuracy, highlighting the ability in segmenting churn-prone customers [8][9].

Financial Services, transitioned from heuristic approaches customer surveys [7] to ML models to proactively predict customer churns and develop retention strategies, with studies like ”Explainable Ensemble Learning and Trustworthy Open AI for Customer Engagement Prediction” (S. Murindanyi et al., 2023) demonstrating how ensemble and explainable AI models could improve customer engagements and predict churns in a banking sector.

Retail E-commerce and subscription services adopted AI for product recommendations and cart abandonment analysis to improve customer lifetime value (CLV). Due to growing importance of software as a service (SaaS) and streaming platforms, studies such as “User Churn Prediction within Music Streaming Service Industry: A Comparison of Machine Learning Models.” showed XGBoost had the highest accuracy and random forest performed similarly well and was more computationally efficient (M. Matusevičius, 2021).

A. Identify the Research gaps in Literature.

- ***Real-time adaptability remains unexplored***
  Lack of focus on using dynamic, real-time data to adjust to ever changing customer behaviors, implying that decision-making becomes reactive rather than proactive [12].
- ***Limited Focus on Localized Data***
  Apparently, there very little information that’s specific to Uganda or even greater east Africa region, which easily lead to generalized strategies reducing ML adoption, impact and relevance.
- ***Understanding of Customer Segmentation*** Available dataset does not provide insight distinct group of customers, different customer groups may churn for different reasons. Segmentation allows businesses to develop tailored retention strategies for each segment.
- ***Churn Prediction Methodologies***
  The current available dataset lacks systematic approaches to identify customers who are likely to stop using a product or service. This information is not available even in core-banking systems greatly limits adoption of ML within the retail banking sector.
- ***Ethical Considerations and Data Privacy***
  Churn analysis involves analyzing customer data to predict and reduce attrition, and can raise significant ethical and privacy concerns, therefore there is a great need to balance predictive analytics with respect for customer rights and privacy.

B. Summary of our term paper contributions

- ***Improved Revenue Optimization***
  Churn analysis plays a pivotal role in revenue optimization by enabling businesses to retain customers, reduce acquisition costs, and increase customer lifetime value (CLV).
- ***Enhanced Customer Experience***
  By leveraging insights gained from churn analysis retail banking can offer personalized services, develop targeted retention strategies, simplified customer journey, proactively resolve a problem and continuously improve their products and services.
- ***Reduction in Churn Rates***
  Reducing churn rates is a critical goal for any organization aiming for sustainable growth and profitability. By leveraging churn analysis, businesses can proactively identify at-risk customers, implement targeted interventions, and build lasting relationships.
- ***Ethical Frameworks***
  Ethical frameworks are essential for ensuring that churn analysis is conducted responsibly, respecting customer rights, fostering trust, and aligning with societal values. These can be done through enabled transparent data collection, gaining customer consent while adhering to data privacy and security regulatory requirements.

IV. METHODOLOGY

Problem Being Investigated

The core problem being investigated in this study is customer churn analysis, scenario where customers stop using retail banking products and services, which directly impacts organizational growth, profitability, and sustainability. This study sought to develop accurate Predictive Models to enhance customer understanding and optimize revenue through strategies that can be personalized, scalable, and ethically sound.

Significance and Scope

Churn analysis is a critical component of business strategy, especially in banking sector where customer retention directly affects profitability and long-term growth. By identifying patterns and predicting customer behavior, banks can take proactive measures to minimize churn, enhance customer satisfaction, and drive sustainable revenue. Churn analysis is an innovation approach whose scope extends across industries such as telecommunications, retail e-commerce, and subscription services, making it an invaluable tool for modern organizations. Its versatile application empowers businesses to proactively address customer attrition, enhance customer experiences, and drive sustainable growth.

AI Approach

Artificial Intelligence (AI) plays a transformative role in churn analysis by enabling businesses to understand, predict and mitigate customer churn more effectively and efficiently. We used advanced machine learning predictive models such as Support Vector Machine, Logistic Regression, Decision Trees, Gradient Boosting and Random Forests to transform core banking data into actionable insights.

Churn Prediction Workflow

1. Data Collection: We selected Retail Banking Demo Data core baking equivalent dataset from and Oracle database to extract final data which we later used to perform EDA.

2. Feature Engineering: We derived meaningful data features from active years, CRM events, transactions counts, total balance and total payments to map a potential churn customer using binary keys.

3. Model Training: We trained all the 5 machine learning models, compared results from each model and performed Cross-Validation to estimate the generalization performance Random Forest.

4. Classification Report: Our report provides detailed metrics that evaluate how well the Our model was able predicts whether a customer will churn or not.

A. *DATASET DESCRIPTION*

We used the Retail Banking Demo Data by Petrocelli of fictional Eagle National Bank dataset, a comprehensive collection of customer and banking data, stitched together from real-world data sources, including CRM and core banking systems. The dataset was initially prepared by an intern for a software demo at Cambridge Semantics is built upon the 1999 Czech banking dataset and has been fully translated, modified and augmented. The dataset is presented in 12 .CSV files, covering six categories of information such as customer data, account information, transactional data, loan data, and credit card data, offering a comprehensive view of customers' activities. From the dataset, Appaneni (2023) presented an Entity-Relationship Diagram (ERD) below, in his study titled ”Retail Banking Data - CRM”, which evolves around the recreation of a functional core-banking application. That can be used to store and retrieve information about the customers, accessed via a web interface to check customer balance, transactions, addresses, card details and customer support incidents.

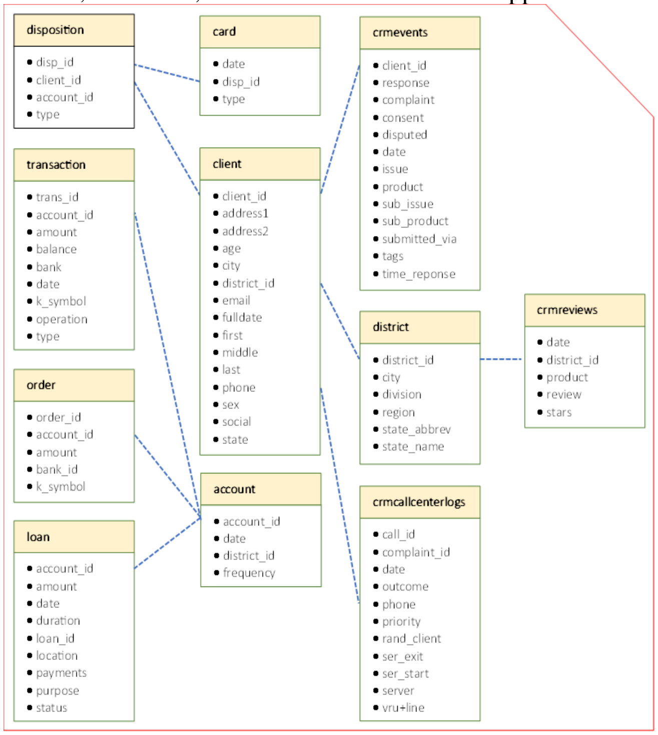
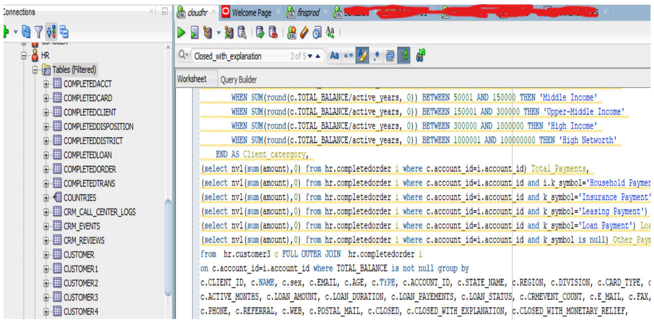


Fig1: Simplified Entity Relationship Diagram

To apply the machine learning knowledge acquired, we selected this dataset because it aligns well in terms of quality, coverage, relevance, and completeness in relations to real-world core banking applications such as Finacle, Flexcube, and T24, which are within our area of expertise. Due to the sensitivity and strict confidentiality requirements of financial data, we opted not to use an actual core banking dataset. We utilized an Oracle database to upload all the .csv files into relevant tables then employed SQL queries to extract the data from these tables, which we then used for our predictions.



```
        WHEN SUM(round(c.TOTAL_BALANCE/active_years, 0)) BETWEEN 50001 AND 150000 THEN 'Middle Income'
        WHEN SUM(round(c.TOTAL_BALANCE/active_years, 0)) BETWEEN 150001 AND 300000 THEN 'Upper-Middle Income'
        WHEN SUM(round(c.TOTAL_BALANCE/active_years, 0)) BETWEEN 300000 AND 1000000 THEN 'High Income'
        WHEN SUM(round(c.TOTAL_BALANCE/active_years, 0)) BETWEEN 1000001 AND 100000000 THEN 'High Networth'
    END AS Client_catergory,
(select nvl(sum(amount),0) from hr.completedorder i where c.account_id=i.account_id) Total_Payments,
(select nvl(sum(amount),0) from hr.completedorder i where c.account_id=i.account_id and i.k_symbol='Household Paymer
(select nvl(sum(amount),0) from hr.completedorder i where c.account_id=i.account_id and k_symbol='Insurance Payment'
(select nvl(sum(amount),0) from hr.completedorder i where c.account_id=i.account_id and k_symbol='Leasing Payment')
(select nvl(sum(amount),0) from hr.completedorder i where c.account_id=i.account_id and k_symbol='Loan Payment') Loa
(select nvl(sum(amount),0) from hr.completedorder i where c.account_id=i.account_id and k_symbol is null) Other_Paym
from  hr.customer3 c FULL OUTER JOIN  hr.completedorder i
on c.account_id=i.account_id where TOTAL_BALANCE is not null group by
c.CLIENT_ID, c.NAME, c.sex, c.EMAIL, c.AGE, c.TYPE, c.ACCOUNT_ID, c.STATE_NAME, c.REGION, c.DIVISION, c.CARD_TYPE, c
c.ACTIVE_MONTHS, c.LOAN_AMOUNT, c.LOAN_DURATION, c.LOAN_PAYEMENTS, c.LOAN_STATUS, c.CRMEVENT_COUNT, c.E_MAIL, c.FAX,
c.PHONE, c.REFERRAL, c.WEB, c.POSTAL_MAIL, c.CLOSED, c.CLOSED_WITH_EXPLANATION, c.CLOSED_WITH_MONETARY_RELIEF,
```

Fig2: Oracle Database snippet

Alternative datasets exist in Kaggle notebook ”Bank Churn Data Exploration and Churn Prediction”, although useful for analyzing and predicting customer churn with pre-engineered features, its greatly lacking transparent explanations, rendering it less suitable for real-world banking applications [14].

B. Data Preparation and Exploratory Data Analysis

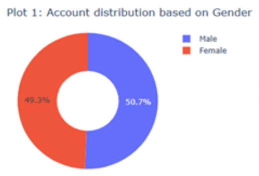
Data Preparation

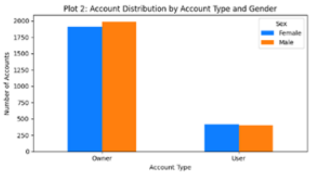
1. The extracted dataset was imported using Python's Pandas library in Google Collab.

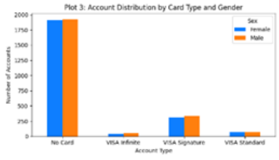
2. The dataset uploaded comprised of 4,708 rows and 41 columns, categorizing customer information, loan details, transactional and financial data, complaint tracking, and payment records, with no missing values.

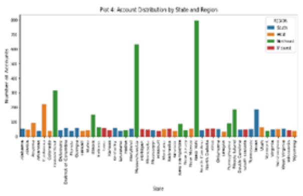
Exploratory Data Analysis (EDA)

Univariate Analysis:



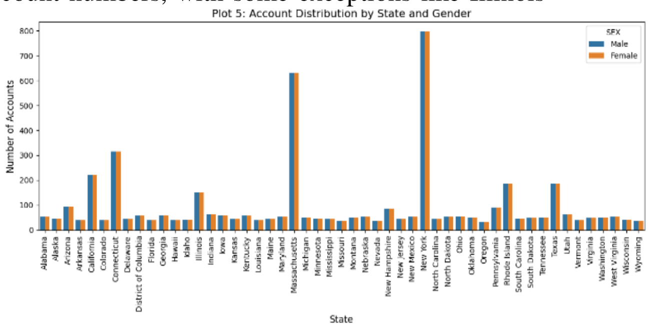




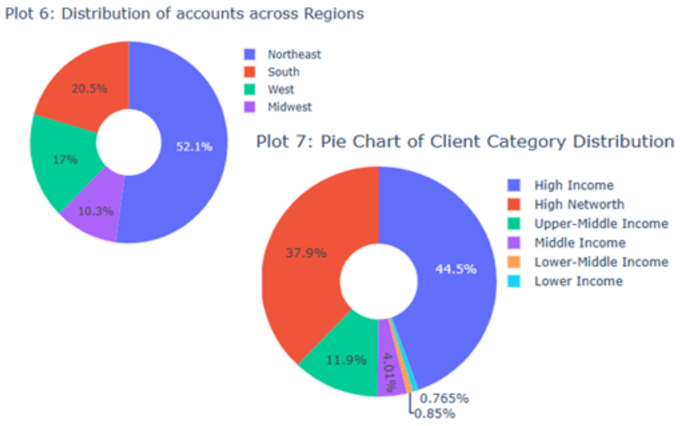


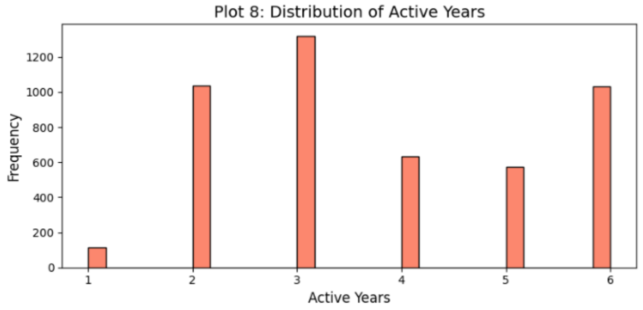
Plot 4: State-Level: New York has the highest number of accounts (over 800), which is significantly higher than other states followed by States like California, Texas, and Massachusetts.

Regional Distribution: Northeast (green bars) dominates, with New York, Massachusetts, and Pennsylvania contributing substantially, South (blue bars) shows moderate contributions, with states like Texas and Florida having higher customers, West (orange bars) features prominent contributions from California and Midwest (red bars) generally shows lower account numbers, with some exceptions like Illinois

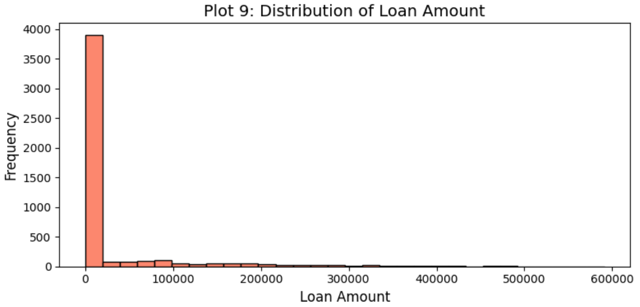


Plot 5: Indicates a balanced gender distribution in account ownership across states. New York has the highest number of accounts for both genders, with more than 800 total accounts, followed by California, Massachusetts, and Texas with some states, like Wyoming, Vermont, and North Dakota, have very low account numbers for both genders.

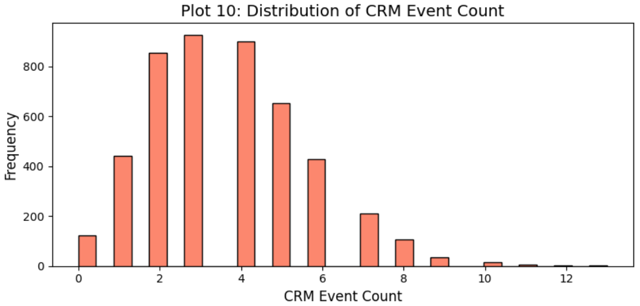




Plot 8: The distribution peaks at 3 years and tapers off as the number of active years increases, except for a minor rise at 6 years. This suggests a typical lifecycle with 6 years been for long-term customers.

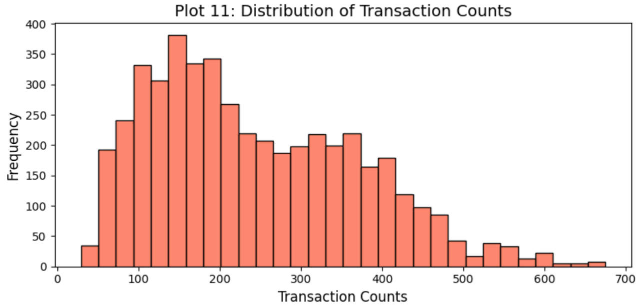


Plot 9: Suggests that most loans issued are of relatively small amounts, with majority of client not having any active loans.

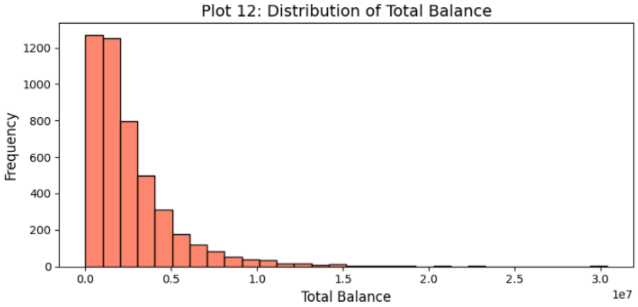


Plot 10: The majority of CRM events cluster around the range of 2 to 5 events, with the highest frequencies observed at 3 and 4 events. which could represent a typical workflow for customer interactions or issue resolution. By the time the event count exceeds 8 to 10, the frequencies are minimal, indicating

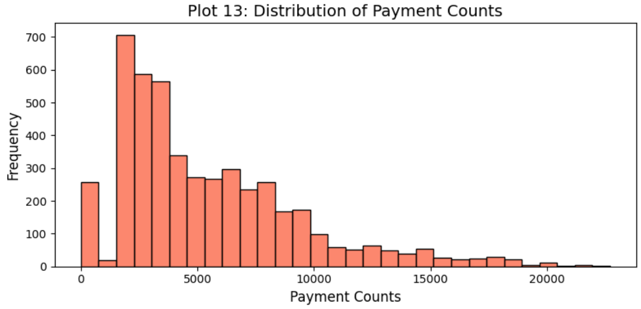
that such high levels of engagement are uncommon.



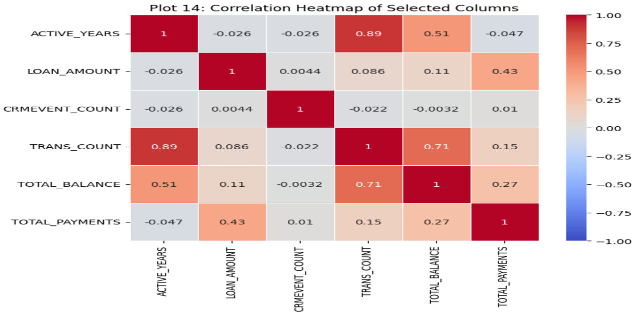
Plot 11: Most transactions are concentrated in the lower range between 100 and 400, while fewer occur at higher transaction counts. The distribution appears to be right skewed indicating a small number of heavy users or high-volume transactions.



Plot 12: The histogram shows that majority of customers have relatively low account balances while few maintain high-balance accounts.



Plot 13: The histogram provides an overview of payment counts of customers, indicating that most customers have lower payment activity, with a small proportion showing high and very high activity.



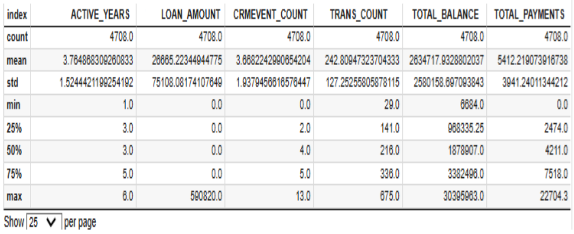
Plot 14: Strong Positive Correlations Customers with a higher number of active years tend to have a higher transaction count. A high transaction count correlates with a higher total balance.

Moderate Positive Correlations Total balance and active years suggest that customers who remain longer tend to accumulate a higher balance Customers with higher loan amounts tend to make higher total payments.

Weak Positive Correlations Loan amount doesn't significantly affect the balance in accounts. CRM event counts may be independent of other financial records.

Feature Engineering

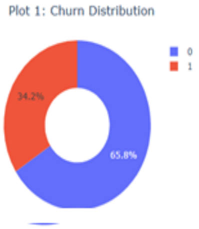
We selected and explored the relevant data columns (ACTIVE_YEARS, LOAN_AMOUNT, CRMEVENT_COUNT, TRANS_COUNT, TOTAL_BALANCE and TOTAL_PAYMENTS), which are crucial in helping to identify patterns used to predict customer churn effectively.

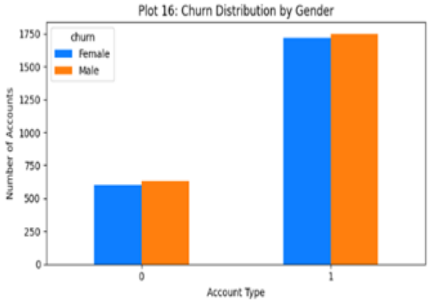


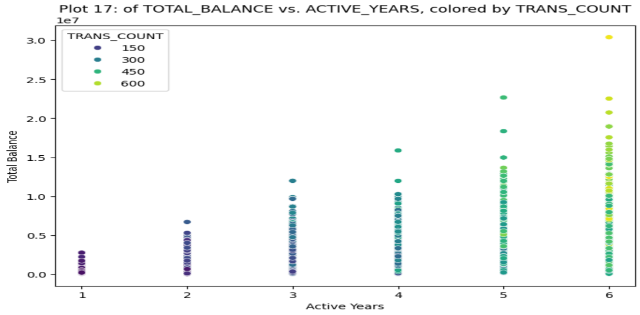
index	ACTIVE_YEARS	LOAN_AMOUNT	CRMEVENT_COUNT	TRANS_COUNT	TOTAL_BALANCE	TOTAL_PAYMENTS
count	4708.0	4708.0	4708.0	4708.0	4708.0	4708.0
mean	3.764868309260833	26665.22344944775	3.6682242990654204	242.80947323704333	2634717.9328802037	5412.219073916738
std	1.5244421199254192	75108.08174107649	1.9379456616576447	127.25255805878115	2580158.697093843	3941.24011344212
min	1.0	0.0	0.0	29.0	6684.0	0.0
25%	3.0	0.0	2.0	141.0	968335.25	2474.0
50%	3.0	0.0	4.0	216.0	1878907.0	4211.0
75%	5.0	0.0	5.0	336.0	3382496.0	7518.0
max	6.0	590820.0	13.0	675.0	30395963.0	22704.3

Show 25 per page

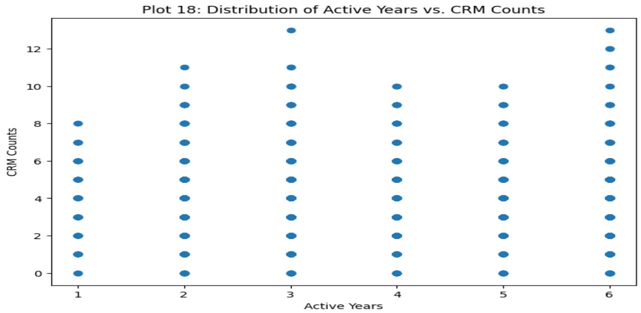
Table 1: Summary descriptions of selected columns



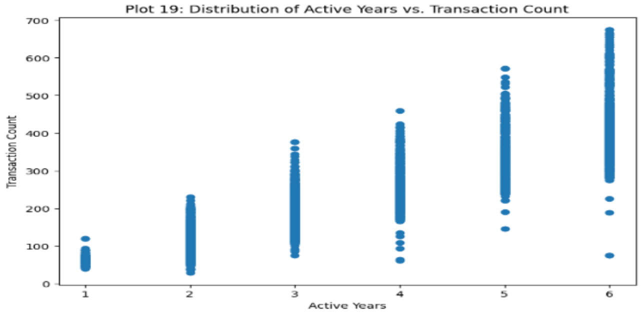




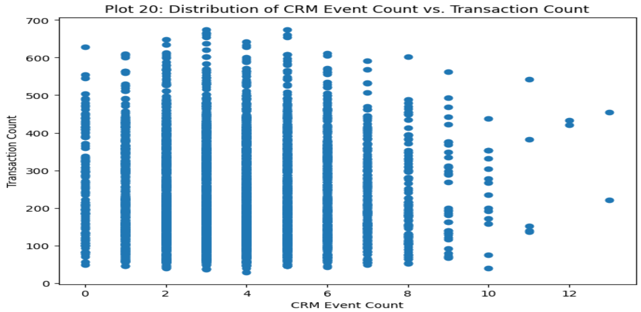
Plot 17: Early years (1 to 2) are associated with lower balances and transaction counts. However few customers show high balances or transaction counts. Longer active years are associated with both higher balances and higher transaction counts.



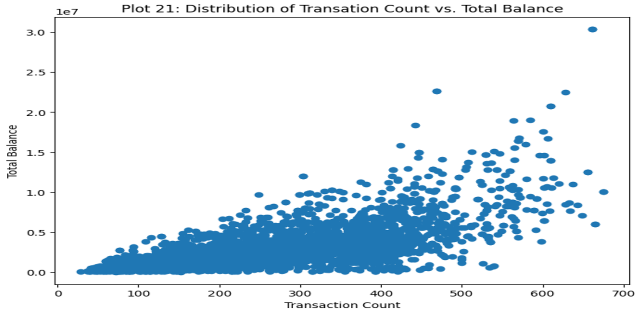
Plot 18: CRM counts range from 0 to around 12, indicating variability in how often customers interact with CRM systems and are fairly distributed across all active years, from 1 to 6 years.



Plot 19: Show clear trend where the transaction count tends to increase with the number of active years. Customers with more years of activity (5-6 years) display higher transaction counts, with some outliers reaching as high as 600+ transactions. Customers with low transaction counts despite having long active years (e.g., year 6 with under 100 transactions) may be at risk of churn due to reduced engagement.



Plot 20: This plot shows Several outliers exist in the plot, such as a few customers with over 10 CRM events but low transaction counts, indicating that excessive CRM activity does not necessarily correlate with engagement.



Plot 21: Most data points are clustered in the lower ranges of Transaction Count (0–300) and low Total Balance of approximately 1 million. Fewer data points at higher values, but some individuals reach transaction counts to 700 and total balances over 3 million representing highly active and high-balance accounts.

C. ML model selection and optimization.

We trained multiple machine learning models to predict customer churn and evaluated their performance.

1. Splitting the Data

Input Features (x): The code selects a subset of columns (ACTIVE_YEARS, CRMEVENT_COUNT, TRANS_COUNT, TOTAL_BALANCE, and TOTAL_PAYMENTS) as features for the model.

Target (y): The churn column is the target variable.

Train-Test Split: The dataset was split into training and testing sets, with 70% for training and 30% for testing. The stratify by ensures the proportion of churned and non-churned customers is preserved in both splits.

2. Initializing Models

A dictionary models was created, containing five popular machine learning classifiers:

- Support Vector Machine (SVM)
- Logistic Regression
- Decision Tree
- Gradient Boosting
- Random Forest

3. Model Training

Each model was trained on the training data (x_train, y_train).

The evaluation metrics for each model were then appended to a list as dictionaries and each metric formatted to 4 decimal places and left-aligned for better readability in the printed output.

4. Cross-Validation To estimate the generalization performance of our model, we then performed the following evaluations.

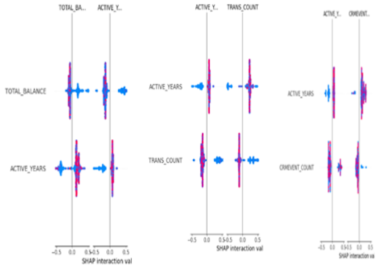
- ***k-Fold Cross-Validation:*** To systematically splits the dataset into k equally sized subsets (folds) and ensures that each subset is used as a testing set exactly once while the remaining folds are used for training.
- ***Stratified K-Fold Cross-Validation:*** To split data in such a way that each fold contains approximately the same percentage of samples for each target class, leading to more reliable model evaluation.
- ***Shuffle-Split Cross-Validation:*** To split data randomly shuffled and split into training and testing sets multiple times, meaning the training and testing sets are reshuffled each time, and it does not necessarily result in equal-size splits.

D. ML model selection Accountability.

AI accountability refers to the idea that AI systems must be transparent, explainable, and responsible for their actions and outcomes, ensuring that AI models and algorithms are used ethically.

In this study we ensured that the model is not only perform well but that its predictions and decisions are understandable, fair, and can be trusted by users in the banking sector. Our engineered features are dynamic and easy to understand for any banking stakeholders.

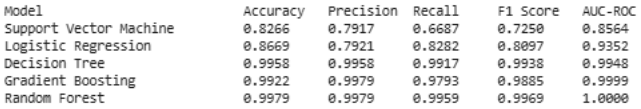
For this project, we can use one of the most common Explainable AI (XAI) techniques: SHAP (SHapley Additive exPlanations) to understand how some features are influenced in the Random Forest model's predictions.



Insight from distribution:

The dispersion and location of points around 0 reveal the strength and direction of the interaction effect. A strong clustering near zero means the interaction has little effect, while wider dispersion shows a stronger effect.

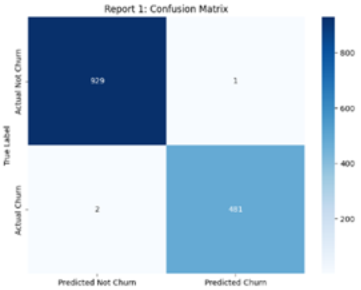
Results and Discussion

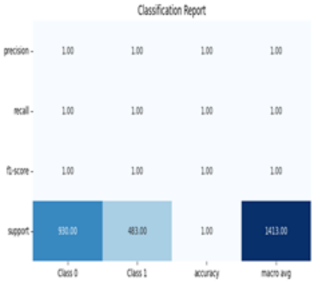


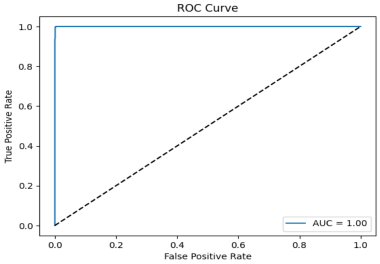
```
Model                        Accuracy   Precision  Recall     F1 Score   AUC-ROC
Support Vector Machine       0.8266     0.7917     0.6687     0.7250     0.8564
Logistic Regression          0.8669     0.7921     0.8282     0.8097     0.9352
Decision Tree                0.9958     0.9958     0.9917     0.9938     0.9948
Gradient Boosting            0.9922     0.9979     0.9793     0.9885     0.9999
Random Forest                0.9979     0.9979     0.9959     0.9969     1.0000
```

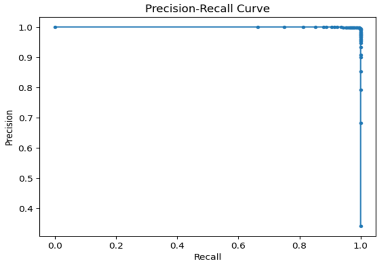
Top Performers: The Random Forest and Gradient Boosting models are the strongest performers across all metrics, with high accuracy, F1 Score, and AUC-ROC. Random Forest edges out slightly with its near-perfect AUC-ROC.

Classification Report









The model performs exceptionally well with no misclassifications, achieving perfect scores across all metrics. This might suggest a near-ideal fit, but further validation to ruled out overfitting.

Conclusion and Future works

In the context of the banking industry, the accurate prediction of customer churn is of utmost importance to maintain a strong customer base and ensure long-term profitability. This paper has explored the application of various machine learning classification models to address this critical challenge,

Future works

1. Incorporate Modern and Diverse Datasets

Collaboration with banks to access anonymized localized real-world data, ensuring a balance between data privacy and model accuracy, can greatly extend ML study.

2. Real-Time Churn Prediction Systems

Develop real-time prediction systems to assess customer churn dynamically and enable banks to implement proactive retention strategies as well as integration predictive models with streaming data pipelines for continuous updates.

3. Enhanced Explainability Techniques

Expand the use of Explainable AI (XAI) tools by incorporating SHAP, counterfactual explanations, and Partial Dependence Plots (PDPs) with user-friendly dashboards for non-technical stakeholders to understand the key drivers of churn predictions.

4. Integration with Banking Systems

Develop APIs and software integrations to embed predictive models into customer relationship management (CRM) and business intelligence (BI) systems and testing the effectiveness of these integrations in live banking environments.

Acknowledgment

We extend our gratitude to the following individuals for their contributions and support: Academic Lecturer and Assistant: Special thanks to our Lecturer Dr. Joyce Nabende and Mr. Sudi in the Machine Learning course for their constructive feedback and expert advice.

References

series=listWWNumxvi,label=0.,ref=0

1) L. Petrocelli, Retail Banking Demo Data, *data.world*. [Online]. Available: https://data.world/lpetrocelli/retail-banking-demo-data.
2) Kassem, Essam & Ali, Shereen & Mostafa, Alaa & Kamal Alsheref, Fahad. (2020). Customer Churn Prediction Model and Identifying Features to Increase Customer Retention based on User Generated Content. International Journal of Advanced Computer Science and Applications. 11. 10.14569/IJACSA.2020.0110567.
3) Suh, Youngjung. (2023). Machine learning based customer churn prediction in home appliance rental business. Journal of Big Data. 10. 10.1186/s40537-023-00721-8.
4) Dawes, J., & Swailes, S. (1999). Retention sans frontieres: Issues for financial service retailers. International Journal of Bank Marketing, 17(1), 36–43. https://doi.org/10.1108/02652329910254037
5) Uganda Bankers Association, Uganda's Banking Sector Report for the Year 2023, *Uganda Bankers Association*, Kampala, Uganda, 2023. [Online]. Available: https://ugandabankers.org/Uganda's
6) Wikipedia contributors, List of banks in Uganda, *Wikipedia, The Free Encyclopedia*. [Online]. Available: https://en.wikipedia.org/wiki/List_of_banks_in_Uganda.
7) Checkbox, How to Run Banking Surveys to Better Understand Your Bank Clients, *Checkbox Blog*. [Online]. Available: https://www.checkbox.com/blog/how-to-run-banking-surveys-to-better-understand-your-bank-clients.
8) V. Kavitha, G. Hemanth Kumar, S. V. Mohan Kumar, and M. Harish, Churn Prediction of Customer in Telecom Industry using Machine Learning Algorithms, *Int. J. Eng. Res. Technol.* (IJERT), vol. 9, no. 5, pp. 181–184, May 2020. Available: https://www.ijert.org.
9) S. K. Wagh, A. A. Andhale, K. S. Wagh, J. R. Pansare, S. P. Ambadekar, and S. H. Gawande, Customer churn prediction in telecom sector using machine learning techniques, *Results in Control and Optimization*, vol. 14, 100342, 2024. Available: https://doi.org/10.1016/j.rico.2023.100342.
10) S. Murindanyi, B. W. Mugalu, J. Nakatumba-Nabende, and G. Marvin, Interpretable Machine Learning for Predicting Customer Churn in Retail Banking, in *Proceedings of the 7th International Conference on Trends in Electronics and Informatics (ICOEI)*, 2023, pp. 967–974. DOI: 10.1109/ICOEI56765.2023.10125859.
11) M. Matusevičius, User Churn Prediction within Music Streaming Service Industry: A Comparison of Machine Learning Models, 2021. [Online]. Available: http://arno.uvt.nl/show.cgi?fid=157849.
12) Lalwani, P., Mishra, M.K., Chadha, J.S. *et al.* Customer churn prediction system: a machine learning approach. *Computing* **104**, 271–294 (2022). https://doi.org/10.1007/s00607-021-00908-y
13) M. Appaneni, Retail Banking Application Project Report, GitHub, 2023. [Online]. Available: https://github.com/madhavappaneni/Retail-Banking-Application/blob/main/Project
14) T. Konstantin, Bank Churn Data Exploration and Churn Prediction, *Kaggle*. [Online]. Available: https://www.kaggle.com/code/thomaskonstantin/bank-churn-data-exploration-and-churn-prediction/notebook.