Banking Customer Churn Prediction:

Leveraging Microsoft Azure ML to retain customers

Priyal Rawat

DeVos Graduate School, Northwood University

MIS 658: Data Management, Organization and Visualization

Dr. Itauma Itauma

03/06/2025

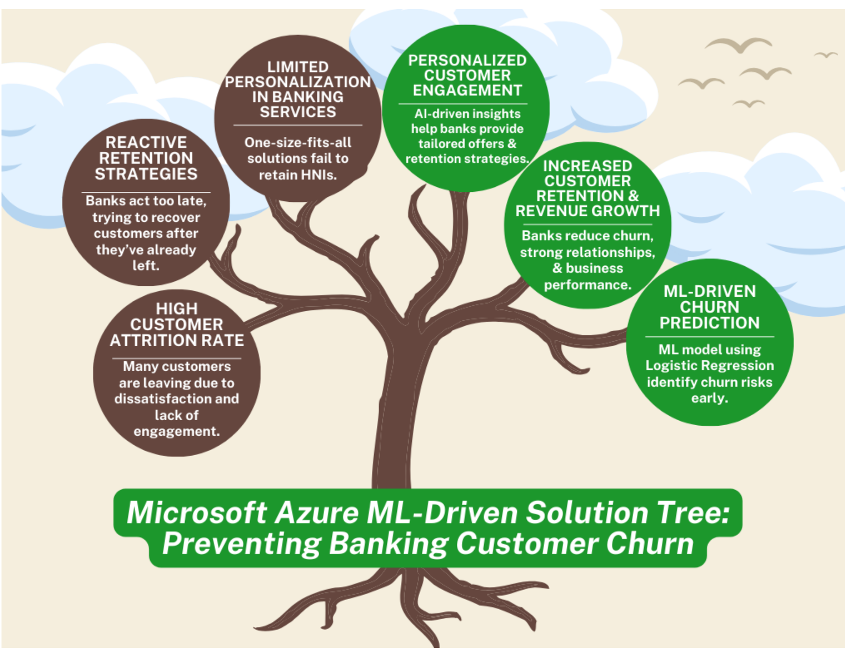
Abstract

Customer Churn is one of the major challenges in the banking industry, which affects revenue and customer engagement with the bank. This paper studies the application of logistic regression, a machine learning model using Microsoft Azure to predict banking customer churn prediction which ensures banks take proactive measures to retain customers timely. To further enhance the model and improve accuracy, other ML techniques like Decision Tree, Random Forest and Gradient Boosting have been included. The comparative analysis between four different models provided insights into the most effective approach to identifying customers at risk. The study has used a dataset which includes customers' demographics, transaction history and their engagement with the bank and I have preprocessed and developed a predictive model. Additionally, I have tested various hypotheses to understand the relationship between customer attributes and churn probability which included studying whether lower account balance, decreased transaction frequency, and reduced engagement with banking services significantly contribute to the customer churn. The findings from the model suggested that my ML prediction model can provide some insight regarding the customer which can be used for improving customer engagement and retention strategies. I feel this research could be very useful for financial services or banks which can use artificial intelligence for customer churn prediction.

Introduction

Customer churn or attrition is one of the significant challenges in the banking industry which affects revenue, customer engagement and brand reputation. During the last few years, the banking industry has become extremely competitive and customer retention has also become the key strategies for the bank. Studies show that acquiring a new customer cost significantly more than making an effort and retaining an existing one, which makes customer retention a key driver for business success (Datrics AI, 2023). Even with digital banking and other technological advancements, customers tend to leave banks because of service dissatisfaction, better interest rates with the competitors or many other reasons. Traditional retention strategies rely on reactive strategies, means identifying customer churn once they have disengaging with the bank. This delay in retention reduces the chance of making the high-net-worth customer engage again with the bank.

I have worked in a banking industry for five years as a relationship manager for High-Net-Worth clients and I have experienced inefficiencies in traditional churn management. I used to receive an Excel file every month containing customer details such as their user ID and net worth or holdings with us over the past 5-7 years. I would then have to research each customer across three different bank's software platforms to identify the potential issue and prepare a personalized solution even before scheduling a meeting with the client to discuss. The entire process would take around one hour per customer, and sometimes, the issue I identified wouldn't even be the actual problem. Also, by the time banks or relationship managers make efforts or initiate recovery process, customers have already shifted significant funds to the competitor bank. Banks need a more proactive and data-driven approach



to predict customer churn before it has happened so that there could be timely intervention and personalized retention strategies can be offered to customers.

Machine learning solution by Microsoft Azure provides the best solution to predict customer churn in a bank within time by analyzing large amounts of data like customer transactions, engagement metrics and other behavioral patterns. Unlike traditional methods or humans analyzing the data, ML, a logistic regression solution, can even identify slight changes in the pattern and can raise a flag for banks to work on it, before there is a significant loss to the bank. Research suggests that AI-driven churn prediction models outperform conventional methods by providing accurate real-time insights into customer disengagement trends (Pahul Preet Singh & Fahim Islam Anik, 2023). By leveraging this model, banks can proactively engage with the customers at risk, provide them with solutions and prevent revenue loss which will ultimately result in customer long term growth and loyalty.

This paper aims to develop a Logistic Regression-based churn prediction model within Microsoft Azure. To further enhance the model and improve accuracy, other ML techniques like Decision Tree, Random Forest and Gradient Boosting have been included. The comparative analysis between four different models provided insights into the most effective approach, which will give banks to work on their proactive strategies and it will eventually result in enhance customer retention efforts, reduce attrition rates, and optimize resource allocation for relationship managers.

Literature Review

Customer churn is one of the major challenges for banks and financial institutions that can significantly affect revenue and more importantly, long-term business sustainability. Churn happens when customers withdraw their business or funds from a bank either partially or entirely. It usually happens due to dissatisfaction, better competitive offers, or change in financial needs. The expense involved in acquiring new customers is typically much higher than that of retaining existing ones, thus churn prediction has become a matter of strategic importance for banks (Datrics AI, 2023). Traditional banking systems rely on manual monitoring and reactive approaches, which are usually very tedious and result in delayed intervention (Pahul Preet Singh & Fahim Islam Anik, 2023).

Research conducted by (Datrics AI, 2023) highlighted some of the key churn drivers, including high service fees, inadequate customer service, and adoption of digital financial solutions. It is also said that high-net-worth individuals are more inclined to leave a bank, as they expect premium and personalized services (Hoang Tran, 2023). And that enrolling with other banks or financial institutions these days is easier because of digital banking or some alternatives, this is making demands greater and also ensuing a scenario in which proactive churn forecasting becomes a necessity (Michael, 2024).

The tremendous advancements in machine learning, or ML, have given churn prediction a completely new look enabling banks to analyze customer behavior, monitor early warning signs, and take preemptive action. Several ML models, including random forest, gradient boosting, support vector machines (SVM), and logistic regression, have been explored for predicting customer churn in the financial sector.

A research study by Keldine Malit (2018,) has made comparisons between various ML models for customer churn prediction which concluded that methods such as random forest and gradient boosting, have a better predictive accuracy than traditional statistical models. However, they demand a significantly higher number of resources, and it reduces their practicability with respect to real-time applications in banking settings. (Derek Papierski, 2023) added that although complex ML models hold higher accuracy, they are considered difficult to interpret and their application is not that feasible in the banking sector where transparency is the major concern.

Despite these advanced ML techniques available today, logistic regression is one of the simplest but mostly used in churn prediction in the banking sector because of its simplicity, interpretability, and minimum computation effort. In other words, logistic regression gives clear outputs of churn probability, which makes it friendly and easy for relationship managers and decision-makers to interpret and act upon real churn risks. (Hoang Tran, 2023). However, other ML techniques like decision tree, random forest and gradient boosting can improve the accuracy of the model which is extremely important to accurately identifying the customer churn.

According to (Pahul Preet Singh & Fahim Islam Anik, 2023) logistic regression deals well with structured banking data like customer demographics, transaction frequency, and account balance. Logistic regression is relatively simple and does not require complex hyperparameters to be tuned and it performs quite well even for small datasets. According to Salesforce, financial institutions would prefer a model that is partially aligned to the regulatory requirements of compliance through explainability and transparency.

Another reason behind the selection of Logistic Regression is its power to manage imbalanced data. Customer churn data typically have less cases of churners than non-churners. This imbalance class will mean that any predictive learning algorithms will tend to favor the frequent class. The other ML techniques like Random Forest and Neural Networks require resampling to adjust this imbalance. Logistic regression can implement the changes needed and can deal with biasness naturally (Joao B. G. Brito, 2024), but these ML techniques like decision tree, random forest and gradient boosting can improve the accuracy of the model.

Additionally, (Totango) further notes how banks use ML models with their existing risk assessment framework. Logistic regression is also interpretable in such a way that it can provide coefficient, which means it tells you about the impact of each feature. This makes it easy for the banks to process the factors that contribute most to churn, thus helping them design an effective retention strategy. This is in line with research by (Ke Peng & Yan Peng, 2023) in which models with higher interpretability improve stakeholder trust and adoption in financial institutions.

While ML has contributed significantly to churn prediction, several gaps and problems remain. First, most studies focus on performance metrics of models, such as accuracy and F1-score, which do not address implementation problems associated with deploying them in actual banking environments (Michael, 2024). Another relevant limitation is the lack of real-time customer retention prediction frameworks that would integrate with banking systems to provide actionable insights.

Second, the majority of studies do not address sentiment analysis together with the structured banking data. Given that finance decisions strongly rely on emotional judgment customers made by collecting transactional data along with their feedback through emails, surveys, and tickets, the churn prediction models might be improved (Pahul Preet Singh & Fahim Islam Anik, 2023).

Finally, there are few studies that discuss the long-term effects of AI-based churn rescue mechanisms. Most of the studies illustrate how ML models predict churn, with very low studies having observed the effects of such predictions in real banking situations where the churn rate was actually reduced (Hoang Tran, 2023).

The literature review indicates that customer churn prediction is an integral part of banking strategy, and ML has transformed banking strategies for retention. Some such models like Random Forest and Gradient Boosting have very high accuracy; however, they are complex, computationally expensive and most importantly lack interpretability, which makes them relatively non-applicable in banking where product transparency is of topmost priority. Logistic regression is still preferred due to its explainability, brainless implementation, and adherence to standards by the banking regulator. This research provides an answer to existing gaps by focusing on building a practical, basic, scalable, and interpretable churn prediction model through ML Logistic Regression in Microsoft Azure, plus doing a model comparison between other ML techniques like decision tree, random forest and gradient boosting for identifying the most effective and accurate approach for banking customer churn prediction.

Methodology

Problem Statement

Customer retention is among the biggest challenges in the banking industry. Losing customers means losing revenue, both in the short and long run. Determining why customers leave is a top priority for banks if they want to act on retention strategies in a timely manner. This study aims to develop a machine-learning tool using Microsoft Azure to facilitate churn prediction by working on customer attributes, financial behavioral aspects, and banking engagement metrics. The patterns of customer transactions and banking habits are tracked through the model to take preventive measures before a customer has an intention to leave. The aim is also to build a strong machine-learning model that would predict churn while providing insights to build better customer retention strategies and to do a model comparison to find out the model effective model for customer churn predictions.

Formulation of Hypotheses

I have formulated four hypotheses based on my literature review for investigating the factors influencing customer churn. The first hypothesis explains that customers with higher balance rates on their accounts would be less likely to churn, as customers with high balances would be expected to maintain a closer financial relationship with the bank, leading to a lower chance of leaving. The second hypothesis suggested that customers using multiple products will have a lower rate of churn, because using more financial services mutes a customer with more than one product, like a loan, a savings account, and a credit card, giving this customer some kind of dependency on the bank. The third studies the relationship of active membership status with lower churning rates, assuming that such customers will remain in close contact with their banks and will be less likely to go elsewhere. The fourth and last hypothesis finds out whether customers with higher credit scores would have lower churn, assuming a good credit rating reflects both financial stability and a long-term banking relationship. These hypotheses guided the selection of features for model building and evaluation.

Data Collection and Overview

Figure 1: Dataset Overview

	customer_id	credit_score	country	gender	age	tenure	balance	products_number	credit_card	active_member	estimated_salary	churn
0	15634602	619	France	Female	42	2	0.00	1	1	1	101348.88	1
1	15647311	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
2	15619304	502	France	Female	42	8	159660.80	3	1	0	113931.57	1
3	15701354	699	France	Female	39	1	0.00	2	0	0	93826.63	0
4	15737888	850	Spain	Female	43	2	125510.82	1	1	1	79084.10	0

The data set used in this study was downloaded from Kaggle and is Bank Customer Churn Dataset of 10,000 customers who are account holders at ABC Multistate Bank. The dataset includes 12 features which represent demographic, financial, and customer engagement. Some key features included credit score, age, balance, number of products used, active membership, country, gender, tenure, credit card ownership, and estimated salary. The target variable of this study is churn, represented by a binary variable where a value of 1 means the customer left the bank and 0 means they did not leave. The dataset gave a complete picture of customer banking behavior which would facilitate predictive churn modeling. (Figure 1)

Data Pre-Processing

Before the training of the machine learning model, data pre-processing was done to maintain data quality and consistency of the data. It involves dealing with empty values which had missing values, converting categorical list fields like gender and country into numerical values. (Figure 2). It is one of the major steps as it would affect the accuracy of the model.

Figure 2: Conversion of Categorical into numerical

	customer_id	credit_score	country	gender	age	tenure	balance	products_number	credit_card	active_member	estimated_salary	churn
0	15634602	619	0	1	42	2	0.00	1	1	1	101348.88	1
1	15647311	608	2	1	41	1	83807.86	1	0	1	112542.58	0
2	15619304	502	0	1	42	8	159660.80	3	1	0	113931.57	1
3	15701354	699	0	1	39	1	0.00	2	0	0	93826.63	0
4	15737888	850	2	1	43	2	125510.82	1	1	1	79084.10	0

Another step is features selection (Figure 3) was carried out to align the data with the hypotheses, out of which five essential features such as credit score, age, balance, number of products used, and active membership were used for further study. Those features were selected

Figure 3: Feature Selection

	credit_score	age	balance	products_number	active_member
0	619	42	0.00	1	1
1	608	41	83807.86	1	1
2	502	42	159660.80	3	0
3	699	39	0.00	2	0
4	850	43	125510.82	1	1

due to their strong influence over churn and to provide the capacity for model interpretability and accuracy. The numerical features were standardized to ensure consistency in numerical data, while categorical features were converted into numerical forms applying one-hot encoding. Once all necessary pre-processing steps were completed, the data was set for training the model.

Model Development

For this study, logistic regression was selected as the classification model because of its efficiency in classification problems such as churn prediction. Logistic regression was also best

Figure 4: Model Training: Training Data (80%) & Testing Data (20%)

Step 5: Split Data into Training & Testing Sets To evaluate our model, we divide our dataset into:

- Training Set (80%): Used to train the machine learning model.
- Testing Set (20%): Used to check the model's accuracy. This helps in understanding how well the model performs on unseen data.

```
from sklearn.model_selection import train_test_split

# Split data into training (80%) and testing (20%)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Display dataset shapes
print(f"Training data: {X_train.shape}, Testing data: {X_test.shape}")
```

[16]

```
Training data: (8000, 5), Testing data: (2000, 5)
```

fitted for this analysis because it provides probabilities of predicting churn, it required less computational effort, and it was effective to use standardized numerical data. Feature engineering was performed by removing such irrelevant columns like customer id, tenure, credit card ownership, and country to reduce any issue. This way the final dataset included five independent variables and one dependent variable (churn). The dataset was split into 80:20 (Figure 4) ratio which is 80% of training data and 20% of testing data which ensured balance representation for model to be evaluated. As shown in Figure 4, the dataset is of10,000 customers so 8000 have been used as training data and 2000 for testing data.

It is this structured process that helps guarantee that the model was properly trained and optimized for churn prediction.

Model Evaluation (Logistic Regression Model)

Evaluation of the model is extremely important to test the reliability of the model. Metrics for model evaluation were accuracy, precision, recall, and the F1-score. The Logistic Regression model was trained and considered for performance evaluation through the assessment of key classification metrics. The accuracy of the model is 81%, which indicates that the model accurately predicts customer churn in 81 out of 100 times, giving the model a good rating for customer churn prediction.

The logistic regression model's classification report (Figure 5) shows the model's performance with respect to both churned and non-churned customers. A precision of 0.83 and recall of 0.96 lead to an F1- score of 0.89 for non-churned customers indicates a good model, whereas it struggles for the churned customers showing a precision of 0.55, recall of 0.20, and F1-score of 0.29. This performance disparity also indicates the effect of class imbalance, as the dataset has more non-churned customers than churned which is causing a model to favor predicting the non-churned cases. Hence, overall accuracy was 81%, and the very low recall of churners denotes that it misses out on a considerable number of the actual churners. Macro-average metrics (precision: 0.69, recall: 0.58, F1-score: 0.59) and weighted average F1-

Figure 5: Logistic Regression Model Classification Report

Logistic Regression Report:

	precision	recall	f1-score	support
0	0.83	0.96	0.89	1607
1	0.55	0.20	0.29	393
accuracy			0.81	2000
macro avg	0.69	0.58	0.59	2000
weighted avg	0.78	0.81	0.77	2000

score of 0.77 further point out that the model should improve on accurately predicting churned customers.

Further Model Testing

To ensure the validity and effectiveness of the model, additional testing was done to predict customer churn. Results were displayed in tabular form for closer manual inspection of customers. The idea was to first put customer IDs through but found difficulty in pre-processing and then use credit score or age as unique identifiers. This showed the effectiveness of the model to make realistic predictions. Before generating the final predictions, the raw test data were prepared correctly so that it matched the final form of the training data. It was made sure that the test dataset included identical feature settings used during training and manually aligned columns when discrepancies existed to ensure both datasets were on the same scale. Initially, errors due to feature mismatches were corrected, such that extra features were removed in order to only keep those that were employed during model training. The test data were then scaled to fit on the training data, so the distributions of the features developed similarly. Following the preparation of the test data, we predicted churn probability for each customer using the Logistic Regression model that we trained. The final table (Figure 6) showed clearly the justified churn, adding to the trust in our model.

Figure 6: Further Model Testing

✅ Predicted Churn Results:

	Credit Score	Age	Predicted Churn
0	751.0	36.0	0
1	581.0	34.0	0
2	735.0	43.0	0
3	661.0	35.0	0
4	675.0	21.0	0
5	738.0	58.0	1
6	813.0	29.0	0

Model Comparison

Model Comparison between logistic regression models and other model techniques like decision tree, random forest and gradient boosting have been done to have an insight into the best model in terms of accuracy and effectiveness for further implementation in the banking industry.

Within the four tested models—the Decision Tree, Random Forest, Logistic Regression, and Gradient Boosting—the Random Forest and Gradient Boosting models were significantly better than the Decision Tree and Logistic Regression ones. The Decision Tree has 78% accuracy, with a recall of only 50% for the churn class (1), which is rather low, meaning it could not find customers who were likely to churn well. Logistic Regression has 81% accuracy, but its ability to detect churners is below par, earning it just 20% recall, which is no good for real-life churn prediction because it cannot detect customers that are in danger of churning.

On the other hand, both Random Forest and Gradient Boosting models outperforms the other two—the two approaches that provide 87% accuracy. Also, they have higher recall values for churn class (Random Forest 47%, and Gradient Boosting this year 49%), so they are better suited to catch customers at risk of leaving. Among the two, the Gradient Boosting appears to be the best model that provides a balance between precision and recall with a balanced f1-score, assuring high accuracy along with better identification of the churn. Owing to its very predictive strong capability, this could be the preferred method for predicting customer churn in real banking situations.

Decision Tree Classification Report

The Decision Tree classification report (Figure 7) explains how effectively the model can identify whether or not a banking customer has churned. The model was shown to achieve 87% precision for non-churners, which means that when it states a customer won’t churn, there is an

87% chance that such a customer was correct. In addition, a recall of 85% means that the model found that 85% of those customers who haven't actually churned have been properly identified. Thereafter, the F1-score considers an above-average value of 0.86, which indicates that the model has struck a good balance in predicting customers that stay with the bank. In contrast, for the churned customers, precision stands at a very low 45%, which implies identification with high inaccuracy - that over half of those customers predicted by the model to churn end up not doing so. A recall of 50% indicates only half of the real churned customers were correctly detected, thereby an F1-score of 0.47, showing the model does have an issue in its capabilities to detect churn.

Overall, this model, in general, has an accuracy of 78%, which means that it correctly classifies the customers in 78% of cases. However, the macro-average of 0.66 precision, 0.68 recall, and 0.67 F1-score indicates the model does not give equal performance for churned and non-churned customers. The weighted average of precision, recall, and F1 would be 0.79, 0.78, and 0.79, respectively; it considers class imbalance and reflects better performance for the majority class (non-churned customers). The low Recall for churned customers indicates the model is not fully capturing the true churn risk, meaning that many customers who are at risk of leaving might go unnoticed.

Figure 7: Decision Tree Model Classification Report

Decision Tree Report:

	precision	recall	f1-score	support
0	0.87	0.85	0.86	1607
1	0.45	0.50	0.47	393
accuracy			0.78	2000
macro avg	0.66	0.68	0.67	2000
weighted avg	0.79	0.78	0.79	2000

Random Forest Classification Report

The Random Forest classification report (Figure 8) suggests that this model performs with an accuracy of 87%. The predictive power of the model is very strong for non-churned customers (class 0), with precision equal to 88% and recall equal to 96%, giving an F1-score of 0.92, signifying the model's capability of effectively spotting customers who no longer plan to continue their relationship with the bank. For churned customers (class 1), on the other hand, precision is 76%, meaning the model is right 76% of the time when it predicts a customer will churn. And yet, recall is only 47% in this respect; thus, more than half of the customers who actually churn is incorrectly classified as non-churned. The computed F1-score for churned customers (0.58) highlights this weakness, finding that the identification of the customers at risk of churning must be improved.

Figure 8: Random Forest Model Classification Report

Random Forest Report:

	precision	recall	f1-score	support
0	0.88	0.96	0.92	1607
1	0.76	0.47	0.58	393
accuracy			0.87	2000
macro avg	0.82	0.72	0.75	2000
weighted avg	0.86	0.87	0.85	2000

The macro average of 0.82 - precision, 0.72 - recall and 0.75 - F1 represents poor performance by the model in identifying churn customers, while floating above macro average are better F1 scores of 0.86 – precision, 0.87 – recall and 0.85 - F1. Random Forest model validated and bettered all other key metrics with recall for churn customers in particular, although it still does not reach an ideal classification.

Gradient Boosting Classification Report

The Gradient Boosting classification report (Figure 9) tells us that this model earned an accuracy score of 87%. The model achieved a precision of 88% for non-churning customers, meaning it is able to correctly flag almost all customers who will not leave their banks. Furthermore, its ability to recall non-churning customers is 96%, implying identification of nearly all customers likely to stick with the bank; thus, earning an F1 score of 0.92. This means that the model is quite effective in identifying customers who are not likely to churn. The precision for churned customers is equal to 75%, which is great. In other words, whenever a model says a customer will churn, that prediction is right 75% of the time. However, recall for churned customers is only 49%, meaning that over half of the churned customers were wrongly classified as non-churned. This gave it an F1-score of 0.59 and proves that there is a need for further room for improvement, specifically with regard to churned customers.

Figure 9: Gradient Boosting Model Classification Report

Gradient Boosting Report:

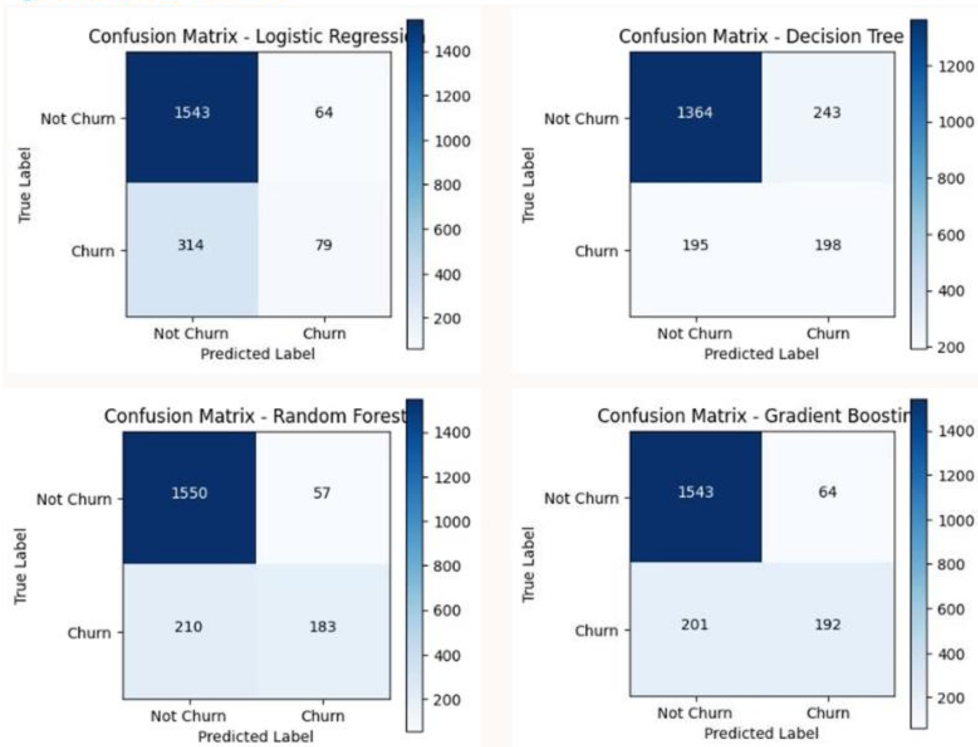
	precision	recall	f1-score	support
0	0.88	0.96	0.92	1607
1	0.75	0.49	0.59	393
accuracy			0.87	2000
macro avg	0.82	0.72	0.76	2000
weighted avg	0.86	0.87	0.86	2000

In terms of macro averages, with a precision of 0.82, a recall of 0.72, and an F1-score of 0.76, it indicates a disparity between the prediction quality of churned and non-churned classes whereby the model is quite biased towards the majority class (non-churned customers). Reversed, the weighted average indicates that the model performs well, 0.86 for precision, 0.87 for recall, and 0.86 for F1, especially when taking into consideration the data imbalance. Gradient Boosting possesses features with higher precision and recall showing the model's power in customer churn prediction.

Confusion Matrix for all four Models

The confusion matrix of each model shows an efficiency in the classification of customer churn. In the case of the Decision Tree model, the system does quite well in the identification of non-churned customers but instead confuses a large number of actual churners: 1364 against 195.

Figure 10: Confusion Matrix



It has relatively few similarities to Logistic Regression otherwise, but the number of false negatives is much higher in Logistic Regression: it confused 314 churners as non-churners. Therefore, it has difficulties in the prediction of churn. The Random Forest model has a little better performance than the previous ones: it managed to classify 1550 non-churned customers correctly and also predicted 183 churners correctly. It appears that this model balances between precision and recall better than both the Decision Tree and Logistic Regression models.

Gradient Boosting is the most effective model, with 1543 true negatives and 192 true positives classified as churners, reducing false positives and also false negatives. This shows that Gradient Boosting is more reliable at capturing churn patterns and is, nonetheless, stable overall. It offers the best trade-off between precision and recall among the tested models; thus, it will be the model of choice to predict churning clients in a banking environment. Because of its good performance, this model will allow banks to develop proactive retention strategies because they would be able to identify with high precision a small number of customers who are likely to leave and carry out interventions on time.

Testing and Interpretation of Results of Hypotheses

The hypothesis was tested against the customer churn to understand their validity. The first hypothesis was rejected suggesting that customers who have higher account balances had a lower probability of churning. Results showed that higher balances had higher chances of being churned, which is not a popular belief in traditional banking assumptions. The second hypothesis stated that customers using multiple banking products have lower churn rates, and it was confirmed as customers who are engaged in multiple banking products are less likely to churn as it’s going to be difficult for them to disengage with the bank and move their entire holding from one financial institution to another. The third hypothesis proposed by the model was validated which is that active membership reduces churn. It was shown that actively engaged customers are usually the ones with significantly lower churn rates. The fourth hypothesis which explained that lower churn rates are caused by higher credit scores, was rejected. The study indicated that credit scores have little effect on churn.

Challenges, Limitations, and Assumptions

There were several challenges during the model development. One of the challenges was the feature misalignment of the test dataset from the training dataset, which had insufficient features; such situations confuse the model and lead to prediction errors. This means that proper feature selection matters when there is a need to align the two datasets correctly. Another point of concern was that missing values resulted in problems, especially in categorical features.Further, according to initial predictions, the customer IDs were not included; this made tracking results per customer difficult. We decided to pinpoint either age or credit scores as the alternate unique identifiers for customers' choices.

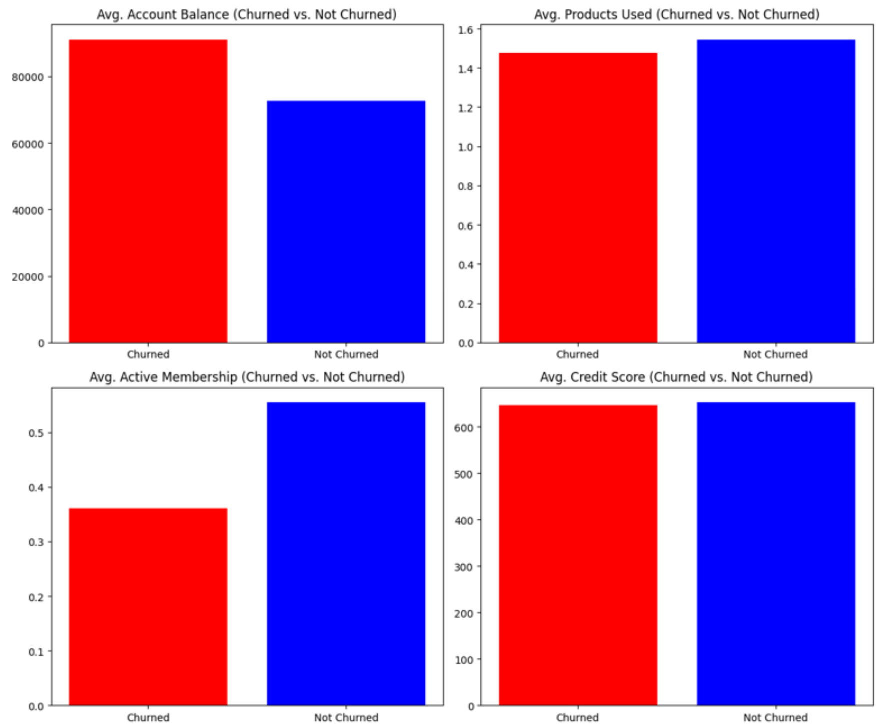
The most significant limitation was the existence of imbalanced classification in the dataset. More customers did not churn, which is only 20% of customers within the database. This made it impossible for the model to generalize patterns for making predictions in favor of the minority group of customers, hence yielding lower recall for predicting already churned customers. I further assumed that all numerical features had equal importance in predicting churn, when in reality it is very possible that some factors have a non-linear effect. The other important way to provide a reasonably informative baseline is through Logistic Regression. To improve baseline performance, more challenging types of models that can be looked into include Random Forest or Gradient Boosting. Finally, it was further assumed that the dataset was an adequate reflection of customer behavior in the real world; however, there were other possible variables that were not captured in the dataset, such as customer services offered, or competitor offers.

Results

These visualizations (Figure 11) give useful insights into customer churn behavior by comparing churned customers with non-churned customers across four hypotheses. The graph in the top-left shows that churned customers tend to have a slightly larger account balance on average than non-churned customers, meaning that account balance alone does not help stop the churn. Again, the average number of products used was practically the same for both groups, as illustrated in the top right graph, thereby making product variety not a strong predictor for churn.

The graph in the lower left shows that there is a very large distinction in active membership, with

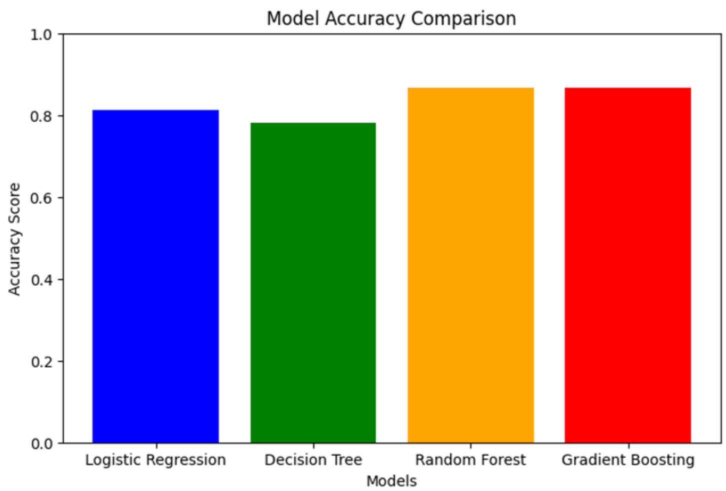
Figure 11: Hypothesis Testing Results



non-churned customers showing exceptional engagement, reiterating the importance of continued activity in customer retention. Finally, from the analysis, we see that both don't show any significant difference in credit scores observed across both the groups, exhibiting that financial credibility doesn't seem to play a strong role in impacted churn predictions. This further shows that customer engagement and activity levels take precedence in retaining customers rather than balance scorecard or credit scores.

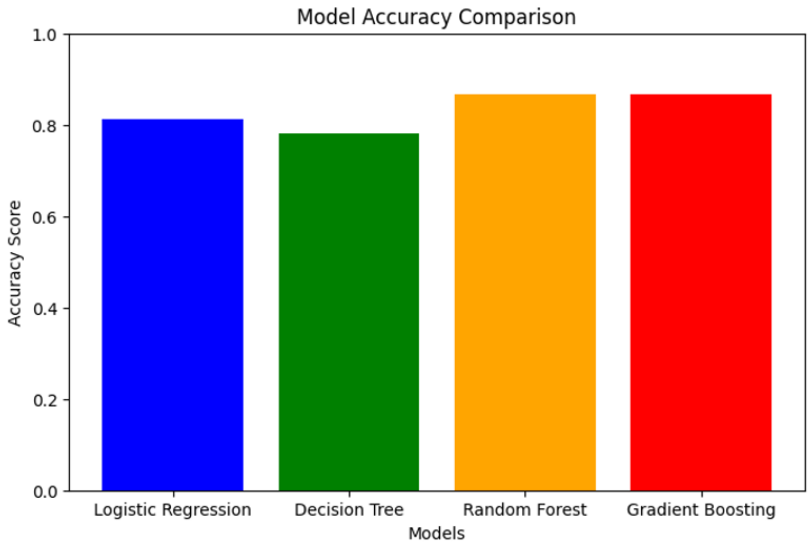
In a comparison of machine learning models for the prediction of customer churn in banking, ensemble methods scored higher than traditional. The Decision Tree model had the lowest accuracy with an accuracy of 78% which adds to its constraints of being unable to tackle complicated data patterns. The Logistic Regression performed a bit better scoring 81% which stands for its interpretability but had struggles with nonlinear relationships. Both Random Forest and Gradient Boosting show up with an impressive score of 87%, proving the fact that catching complex patterns in the data is definitely the forte of these two algorithms: one through multiple decision trees and the second by learning iteratively. Both these models matched in accuracy however, Gradient Boosting still has an upper hand as it helps in progressively reducing the errors which makes it suitable for churn prediction more exactly. Therefore, these findings seem to provide substantial support for banks to enhance customer churn prediction and reduction strategies through more advanced ensemble approaches.

Figure 12: Model Accuracy Comparison



Further the model performance comparison of four machine learning models: Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting is done based on Precision, Recall, and F1-Score. With a precision rate of around 0.76-0.75, both Random Forest and Gradient Boosting are good at predicting churned customers, which means that they manage to identify them with fewer false positives. In recall, Decision Tree and Gradient Boosting seem to

Figure 12: Model Accuracy Comparison



do a better job than other models, meaning they are able to more accurately predict more churn events. Logistics Regression is the weakest when it comes to recall (around 0.2), indicating its poor ability to catch churned customers. The F1-score is the highest for Gradient Boosting and Random Forest, averaging around 0.6, thereby assuring their reliable use in churn prediction. This shows that ensemble methods in churn predictions have outperformed the traditional models supporting the argument put forth by all key performance metrics in favor of ensemble methods.

From a business perspective, the machine learning provides a predictive nature for early intervention strategies. Knowing the at-risk customers, it will help banks introduce personalized retention plans by making targeted offers or increasing customer engagement and much more. The insight from the model highlights the key factors such as active membership and account can lead to long term growth of customer relationship and bank revenue. At the end of it all, this model equips banks with data-driven interventions to deal with churn and focuses their attention on retaining high-value customers to stem revenue loss and improve customer satisfaction.

Conclusion

Customer churn is one of the biggest challenges in the banking sector today, with negative implications for revenues, long-term growth and competitive advantage. The findings of this research present a very solid argument for employing machine learning models for banking consumer churn prediction, allowing banks to proactively manage customer retention efforts. Gradient Boosting and Random Forest Models outperformed their Logistic Regression and Decision Tree counterparts in precision, recall and F1 scores. Additional performance and accuracy strongly identify ensemble models as being ahead of the rest of the models, minimizing false positive and false negative predictions while catching profitable churn. These results hold substantial weight in supporting the assertion whereby advanced machine learning methods improve churn prediction accuracy over traditional statistical techniques.

The findings of this research are very significant for the banking sector. Predictive analytics has come can be extremely beneficial for banks to understand the customers who are most likely to leave, so that they can timely intervene through personalized offers, support, or loyalty programs for retention and minimum loss of revenue. The major focus on at-risk churners allows a channeling of resources into areas with reasonable customer satisfaction and saves on replacements of lost customers. The comparison between models further focuses on the necessity to utilize ensemble methods rather than regular ones so that the predictions are more dependable and actionable.

For future work, to improve model performance, several refinements might be integrated, for instance, sentiment analysis from customer reviews and interactions, deep learning to contribute complex pattern recognition, a real-time churn prediction system that interacts with banking CRMs, recurrent neural network and graph neural network. Methods of feature selection and explanation tools may help make a model interpretable for showing business stakeholders why AI-driven choices are made. Other venues for fine-tuning churn projection models in banking include constantly updating with long-term feedback and including together external economic inputs.

The paper shows the efficiency and advantages of using the tool that I have built for predicting customer churn in the banking sector. By using these tools, financial institutions would no longer have to depend on set traditional methods but could create personalized strategies for customer retention purposes. The model will require continued refinement and AI-driven customer insights will unlock its capability for sustainable customer engagement and organizational growth.

References

Datrics AI. (2023). *Bank churn prediction using ML to retain customers*. https://www.datrics.ai/articles/bank-churn-prediction-using-ml-to-retain-customers

Pahul Preet Singh & Fahim Islam Anik (2023), Investigating customer churn in banking: a machine learning approach and visualization app for data science and management https://www.sciencedirect.com/science/article/pii/S2666764923000401

Hoang Tran (2023), Customer Churn Prediction in the Banking Sector Using Machine Learning-Based Classification Models https://www.ijikm.org/Volume18/IJIKMv18p087-105Tran8783.pdf

Michael (2024), Bank Customer Churn Prediction Using Machine Learning https://www.analyticsvidhya.com/blog/2022/09/bank-customer-churn-prediction-using-machine-learning/

Keldine Malit (2018), Kaggle - Bank Customer Churn Prediction https://www.kaggle.com/code/kmalit/bank-customer-churn-prediction

Derek Papierski (2023), Investigating Customer Churn in the Banking Industry https://medium.com/@dpapcodes/investigating-why-customers-leave-a-bank-47b41278e36c

Salesforce, Use the Retail Banking Customer Churn Prediction Dashboards https://help.salesforce.com/s/articleView?id=ind.fsc_use_churn_prediction_retail_banking_dashboard.htm&type=5

Joao B. G. Brito (2024), A framework to improve churn prediction performance in retail banking
https://jfin-swufe.springeropen.com/articles/10.1186/s40854-023-00558-3

Totango, Customer retention & churn prediction
https://www.totango.com/demo/live-demo

Ke Peng & Yan Peng (2023), Research on customer churn prediction and model interpretability analysis
https://pmc.ncbi.nlm.nih.gov/articles/PMC10707658/pdf/pone.0289724.pdf

Generative Pre-trained Transformer (ChatGpt): For Model Coding

Canva (For introduction image generation)

GitHub Repository link

https://github.com/priyalrawat/BankingCustomerChurnPrediction/tree/main