Customer Churn Analysis and Prediction in Banking Sector

FINAL REPORT

Volkan Çamlı

ABSTRACT

This project investigates customer churn in the banking sector, a phenomenon that poses critical challenges to financial institutions worldwide. Customer churn, defined as the discontinuation of a customer's relationship with a bank, directly impacts revenue streams, operational efficiency, and long-term growth prospects. The study explores key drivers of churn, leveraging advanced data analytics and machine learning techniques to develop robust predictive models. A meticulously curated dataset enriched with feature engineering serves as the foundation for analysis, enabling the identification of demographic, behavioral, and transactional factors influencing churn. By employing state-of-the-art algorithms such as Logistic Regression, Decision Trees, Random Forest, Gradient Boosting Machines and Neural Networks, the project aims to achieve high prediction accuracy and actionable insights.

The study transcends mere prediction, focusing on the practical application of results to inform customer retention strategies. Insights derived from this research will assist banking institutions in tailoring interventions to specific customer segments, optimizing resource allocation, and enhancing overall customer satisfaction. Moreover, the integration of a Power BI dashboard provides a visual and interactive representation of findings, facilitating data-driven decision-making among stakeholders. This comprehensive approach not only addresses the immediate challenges of churn but also contributes to the broader goal of fostering sustainable growth, competitive advantage, and enhanced financial performance in the banking industry. The implications of this study extend to improving customer experience, reinforcing trust, and setting a benchmark for predictive analytics applications in the financial sector.

CONTENTS

1. INTRODUCTION

Customer attrition, also known as customer churn, is the phenomenon where customers terminate their relationship with a business or organization. In the context of banking, customer attrition occurs when customers close their accounts or discontinue utilizing services of a particular bank. Effectively understanding and managing customer attrition are crucial for banks to maintain financial stability and safeguard their reputation. The financial impact of customer attrition on banks can be significant, resulting in potential revenue loss across various banking services. Consequently, establishing and nurturing long-term customer relationships is highly valuable for banks. By gaining insight into attrition patterns, banks can identify customers at risk of leaving and implement strategies to retain them. This approach enhances overall customer lifetime value and bolsters bank profitability.

Moreover, customer attrition has repercussions on a bank’s reputation and brand perception. High churn rates often indicate underlying issues, such as poor customer experience, inefficient processes, or a lack of competitive products and features. Therefore, understanding and managing customer attrition are crucial for banks to address these challenges and enhance their overall customer experience. Within the competitive banking industry, monitoring and managing customer attrition can provide banks valuable insights into customer preferences, needs, and pain points. This knowledge can help banks develop targeted strategies to differentiate themselves from their competitors and enhance customer retention.

This project aims to analyze the factors influencing customer churn in the banking industry and provide actionable insights to address this issue effectively. The scope of the project includes identifying key predictors of churn using data analytics techniques, developing a predictive model to classify customers at risk of leaving, and proposing retention strategies based on the findings. By leveraging statistical and machine learning tools, this study seeks to enhance decision-making processes and support banks in minimizing customer attrition, ultimately contributing to improved customer satisfaction and financial performance.

2. PROJECT SCOPE & OBJECTIVES

The primary objective of this project is to analyze customer churn in the banking sector using data science techniques and to develop machine learning models capable of predicting churn in advance. In today's competitive financial landscape, retaining customers is increasingly difficult, and every instance of customer loss results in reduced revenue, diminished brand value, and increased operational costs for institutions.

Within this scope, the goals of the project are as follows:

- To analyze demographic, financial, and behavioral attributes of banking customers
- To identify key factors contributing to churn using data-driven insights
- To build predictive models using various machine learning algorithms
- To evaluate and compare models using metrics such as Accuracy, F1-score, and AUC
- To develop a visual dashboard based on the best-performing model for decision-makers
- To support banks in designing strategies aimed at reducing churn

The project follows a systematic approach involving data preprocessing, feature engineering, modeling, evaluation, and visualization. The final outcomes are presented through an interactive and user-friendly Power BI dashboard.

3. LITERATURE REVIEW

In recent years, customer churn analysis using machine learning has gained significant attention. Research shows that it is applied across many industries, including telecommunications, banking, e-commerce, insurance, retail, energy, and healthcare. While the telecommunications sector has been a primary focus, banking has also seen growing interest due to its competitive nature and customer retention challenges. For example, studies by Kawale et al. (2009) and Ahn et al. (2006) highlight the importance of understanding churn behavior in these fields.

This project focuses on customer churn analysis in the banking sector. Various factors influence customer loyalty and satisfaction in this area. According to Chakiso (2015), building strong relationships through trust, communication, and personalized services is crucial for customer retention in banks. Similarly, Ozatac et al. (2016) identified key factors like reliable service, personalized communication, and secure financial transactions as critical for enhancing customer satisfaction. Studies like Pasha and Waleed (2016) have also shown that pricing policies, service quality, and brand loyalty play major roles in determining customer loyalty.

In addition to relationship marketing, studies reveal that customer preferences are increasingly influenced by digital services. Technological advancements have led to the widespread use of internet banking, mobile apps, and ATMs, enabling customers to access banking services more conveniently. This digital transformation has heightened competition among banks, as customers can switch providers with greater ease to benefit from better technological features, lower fees, or higher interest rates. For example, Keramati et al. (2016) studied customer churn in electronic banking and found that dissatisfaction with digital services, combined with demographic factors such as age and employment status, significantly impacted customer retention. Addressing these issues with tailored technological solutions can reduce churn rates and improve customer satisfaction.

Machine learning models play a critical role in analyzing churn data and predicting which customers are likely to leave. Techniques like logistic regression, decision trees, and neural networks have been commonly used in the literature. Bilal (2016) employed neural network models using variables like gender, age, income, and customer engagement, showing that customers utilizing multiple banking products are less likely to churn. Similarly, Nie et al. (2011) compared logistic regression and decision tree models in predicting churn for credit card users, finding that logistic regression was more effective for large datasets. These studies underscore the value of leveraging machine learning to predict churn accurately and help banks develop personalized strategies to retain customers.

Predictive models also enable banks to design customer-specific retention campaigns. For example, Brânduşoiu et al. (2016) utilized a dataset with over 20 variables to create an advanced data mining model for predicting prepaid customer churn, achieving notable accuracy. Similarly, Rajamohamed and Manokaran (2018) compared algorithms like k-nearest neighbor, Random Forest, and Naive Bayes, concluding that Support Vector Machines provided the best performance in churn prediction for banking. These tools are not only valuable for prediction but also serve as a foundation for strategic decision-making. By identifying churn-prone customers and understanding the reasons behind their dissatisfaction, banks can implement proactive measures, such as offering personalized discounts, improving service accessibility, or addressing technical issues in their digital platforms.

As the banking industry evolves, customer expectations grow, making churn prediction models increasingly important. The competitive environment demands innovative approaches to retain customers and enhance their satisfaction. Machine learning has proven to be a transformative tool, enabling banks to analyze vast amounts of customer data and develop solutions that align with individual needs. This not only helps in retaining existing customers but also in attracting new ones by creating a reputation for superior service and reliability. Hence, predictive analytics will continue to play a vital role in shaping the future of customer relationship management in banking.

4. DATASET

4.1. Dataset Imprint Information

- **Dataset Name:** Bank Customer Churn Dataset
- **Source:** https://www.kaggle.com/datasets/rangalamahesh/bank-churn/data
- **Dataset Owner:** Rangala Mahesh
- **License:** **Apache 2.0**
- **Description:** This dataset was created to estimate a bank's customer churn.

4.2. Dataset Definition

In this study, an artificially created bank customer dataset was used. The dataset is publicly available on the Kaggle platform and can be accessed via the following link: https://www.kaggle.com/datasets/rangalamahesh/bank-churn/data.

The dataset contains critical information for customer profiling and churn analysis in the banking sector. Each new customer undergoes a "Know Your Customer" (KYC) process, ensuring that the collected information is complete and reliable. Customers are identified by a unique "CustomerID" and "Surname." Additionally, the dataset includes essential customer details such as credit scores, age, tenure with the bank, account balance, number of products used, and estimated salary. It also features categorical variables such as geography, gender, credit card ownership, active membership status, and a target variable, "Exited," which represents churn status. A value of 1 in the "Exited" column indicates that the customer has churned, while 0 signifies retention.

4.3. Feature Engineering

In this project, various preprocessing and enrichment tasks were performed to enhance the analytical value of the dataset and enable meaningful insights. These enhancements were added to the dataset as separate columns and are detailed below:

1. **Classification Based on Credit Scores:**
    - A new column was added to classify customers based on their credit scores into the following categories: "High Risk," "Moderate Risk," "Low Risk," "Good," and "Excellent." This classification aimed to understand customers' risk profiles based on their credit history.

2. **Age-Based Classification:**
    - Customers were categorized into three distinct groups based on their age: "Young," "Middle-Aged," and "Senior." This classification facilitated demographic analyses and the examination of age-related attributes.

3. **Income and Balance-Based Classification:**
    - Using estimated salaries and account balances, customers were categorized into "Low Income," "Middle Income," "Wealthy," and "Very Wealthy." This classification provided insights into the economic status of customers and supported income-based analyses.

4. **Segmentation and Clustering:**
    - Clustering algorithms (e.g., K-means) were employed to segment customers into five groups based on all available attributes. This segmentation enabled the identification of customer groups with similar characteristics and played a critical role in defining marketing strategies.

During the modeling process, all categorical variables were converted into numerical formats to prevent information loss. Variables deemed irrelevant to the analysis, such as "CustomerID" and "Surname," were removed from the dataset.

The primary objective of this study is to utilize this enriched dataset to understand the factors influencing customer churn and build a predictive model. This approach will enable the development of strategies to reduce customer churn effectively.

4.4. Dataset Description

The data set consists of two separate data sets named train and test. The train set consists of 165K rows and the test set consists of 110K rows of customers.

4.4.1. Train Set

- The dataset used for training the model.

CustomerId: Unique identifier for each customer.

CreditScore: Customer's credit score, indicating their credit history and financial reliability.

Geography: Country where the customer resides, used for geographic segmentation and customer behavior analysis.

Gender: Customer's gender, used for demographic analysis and customer segmentation.

Age: Customer's age, a significant factor for age-group and behavioral analysis.

Tenure: Duration of the customer's relationship with the bank, used to understand customer loyalty and bank relationship.

Balance: Customer's account balance, used to analyze their financial status and potential risks.

NumOfProducts: Number of products the customer has, indicating customer engagement and loyalty with the bank.

HasCrCard: Whether the customer has a credit card, used to determine loyalty to financial products and spending habits.

IsActiveMember: Whether the customer is an active member, helping to understand their level of engagement with the bank and customer loyalty.

EstimatedSalary: Estimated salary of the customer, used to determine their financial capacity and potential spending power.

Exited (churn): Target variable. Indicates whether the customer has left the bank; 1 means the customer has left, 0 means they have not.

Columns	Null Count	Unique Count
CustomerID	0	165.034
Surname	0	2.797
CreditScore	0	457
Geography	0	3
Gender	0	2
Age	0	71
Tenure	0	11
Balance	0	30.075
NumOfProducts	0	4
HasCrCard	0	2
IsActiveMember	0	2
EstimatedSalary	0	55.298
Exited	0	2

Table 3.1. Number of null and unique values by column name for Train Set

4.4.2. Test Set

- The dataset used to evaluate the performance of the model.

Columns	Null Count	Unique Count
CustomerID	0	110.023
Surname	0	2.708
CreditScore	0	454
Geography	0	3
Gender	0	2
Age	0	74
Tenure	0	11
Balance	0	22.513
NumOfProducts	0	4
HasCrCard	0	2
IsActiveMember	0	2
EstimatedSalary	0	41.670

Table 3.2. Number of null and unique values by column name for Test Set

4.5. Dataset Analysis

The dataset was preprocessed to effectively combine and visualize various input data parameters in a consistent manner.

4.5.1. Customer Churn Distribution

Churn distribution. The pie chart in Figure 3.1 shows the distribution of our dependent variable (churn) in the dataset. 78.8% of the records are for “non-churned” customers and 21.2% are for “churned” customers. That is, almost 1 in 5 customers are churned.

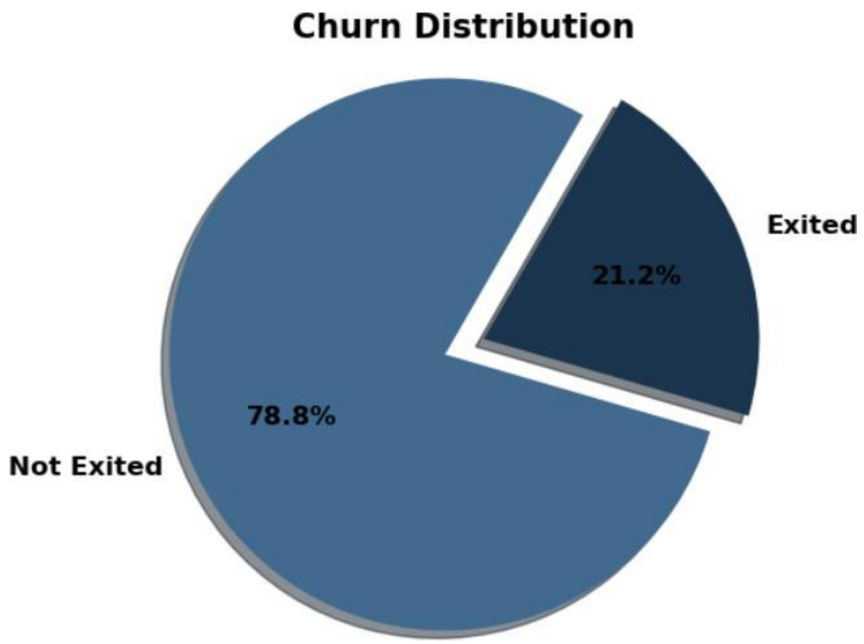

Figure 3.1. Customer churn distribution chart

4.5.2. Histogram Chart

Gender, active members, credit card ownership, and country-based analysis. Figure 3.2. provides valuable information on gender, active members, credit cards, and country-based analysis. We observed that 20,105 out of 71,884 female customers (~28%) were lost, while 14,816 out of 93,150 male customers (~16%) were lost. Furthermore, approximately 1/5 of all customers were lost regardless of whether they were credit card holders or not. Inactive customers are almost two and a half times more likely to leave than active customers (30% inactive customers vs. 12% active customers). The highest customer churn rate is found in Germany (37%), followed by Spain (17%) and France (16%).

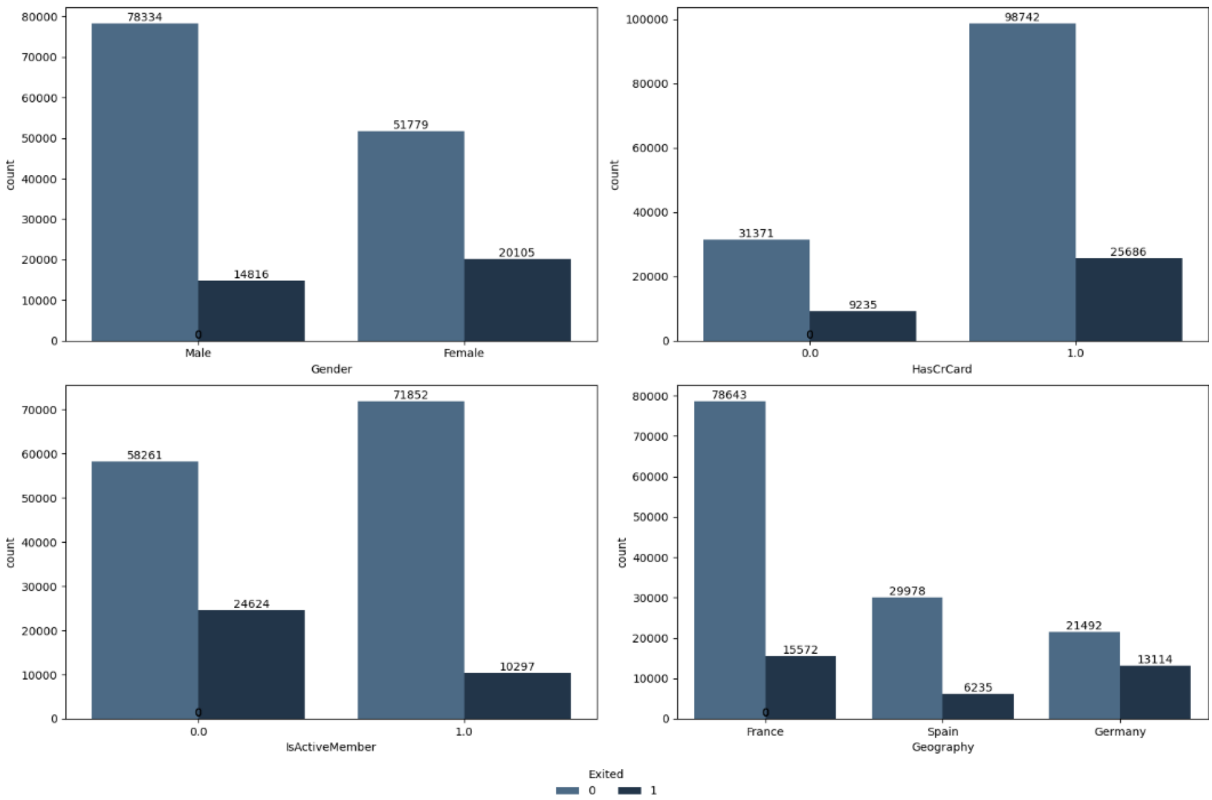

Figure 3.2. Histograms of the dataset. Gender vs. churn; customers having credit card vs. churn; active member vs. churn; country vs churn.

4.5.3. Density Plot

Balance, owned product quantity, credit score, and tenure-based analysis. Figure 3.3. illustrates a density plot to observe the balance, owned product quantity, credit score, and tenure-based analysis.

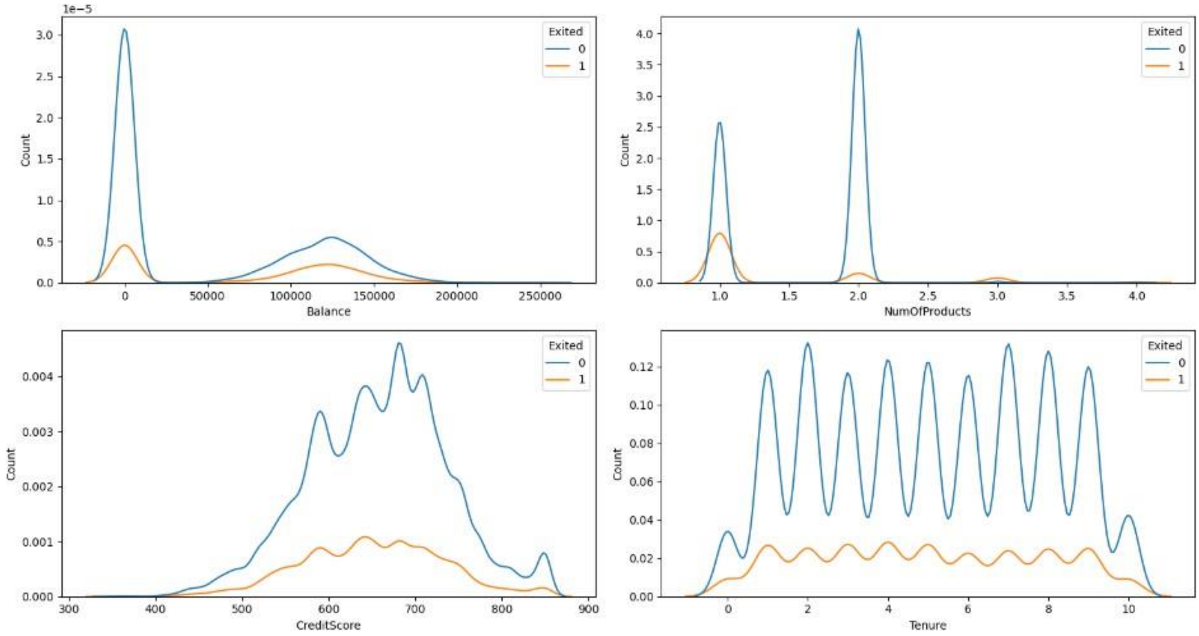

Figure 3.3. Density plots of observed balance, owned product quantity, credit score, and tenure.

- **Balance:** The analysis of account balances among customers who have churned reveals a higher propensity for attrition among those with lower balances. A notable peak in churn rate is observed around the 50,000 threshold.
- **Number of Products:** A negative correlation is evident between the number of products utilized and customer attrition. Customers with a more extensive product portfolio demonstrate a lower likelihood of churn. Notably, those with one or two products exhibit a significantly higher churn rate.
- **Credit Score:** A positive correlation exists between credit score and customer retention. This suggests that individuals with higher credit scores are more likely to remain loyal to the bank.
- **Tenure:** The duration of customer relationships demonstrates a negative correlation with churn. This finding aligns with the notion that long-standing customers exhibit a stronger loyalty to the bank. However, the relationship between tenure and churn is less pronounced compared to other variables.

4.5.4. Scatter Plot

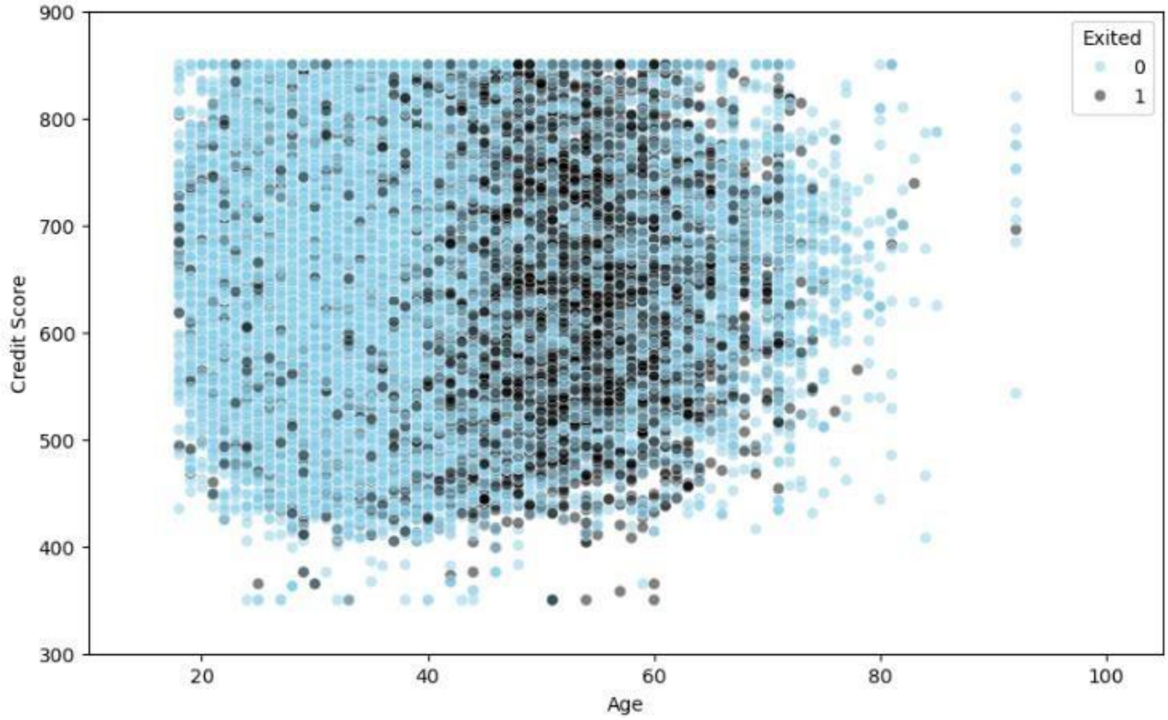

Figure 3.4. Distribution of customers based on credit score and age.

Figure 3.4. shows that although there is no clear concentration according to credit score, customer churn is concentrated in the 40-60 age group.

4.5.5. Correlation Matrix

A correlation matrix is a table that displays the degree and direction of the relationships between variables in a dataset. Each cell in the matrix represents the correlation coefficient between two variables. This coefficient ranges from -1 to +1, where positive values indicate a positive relationship, and negative values indicate an inverse relationship between the variables. The magnitude of the coefficient reflects the strength of the relationship: +1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no relationship. A correlation matrix is widely used to understand relationships between variables and to identify high collinearity during data analysis.

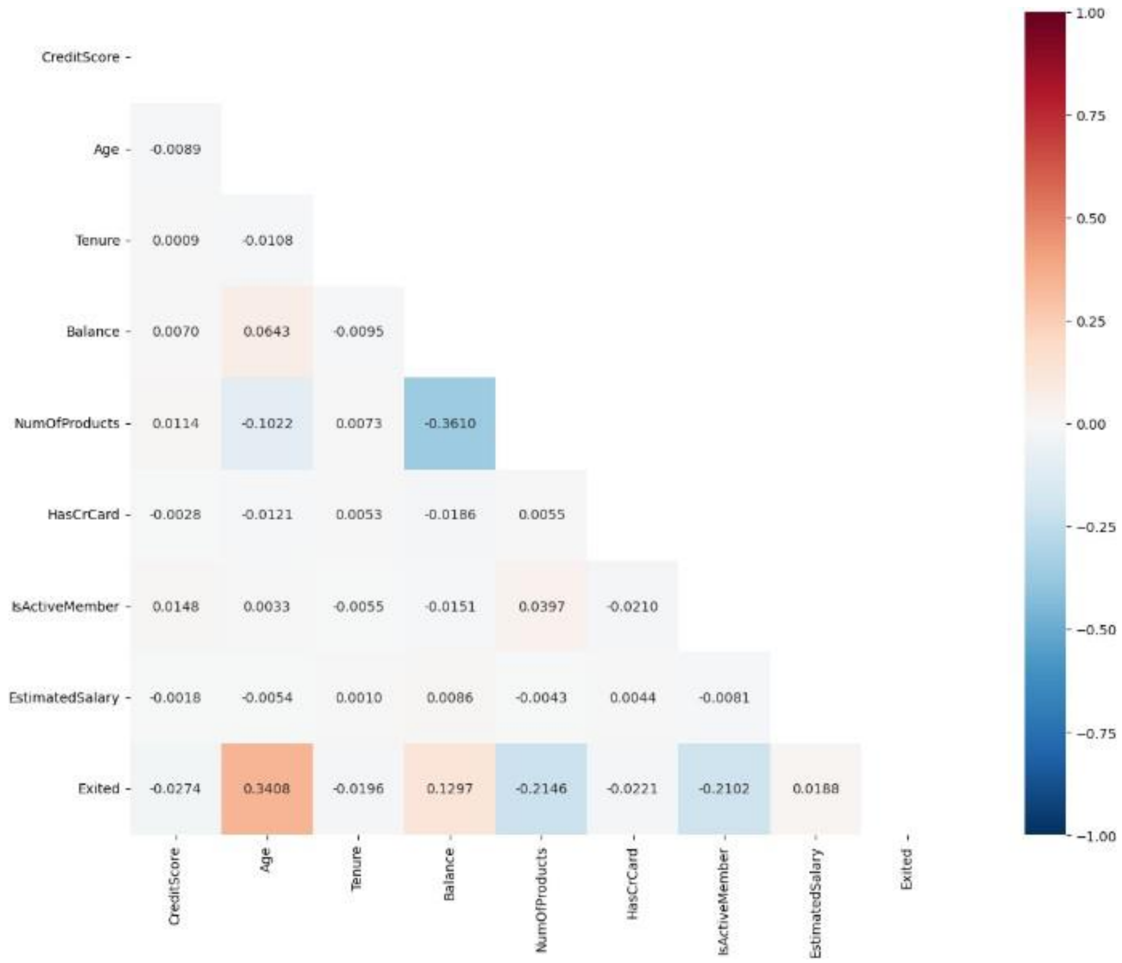

Figure 3.5. Correlation matrix of the dataset

A correlation matrix provides a visual depiction of the strength and direction of the linear relationships between pairs of variables in a dataset. Figure 3.5. reveals a positive correlation between the 'Exited' and 'Age' variables, indicating that older customers are more likely to churn. Conversely, a moderate negative correlation is evident between 'Exited' and both 'NumOfProducts' and 'HasCrCard', suggesting that customers with fewer products and without credit cards are more prone to churn. The remaining variables exhibit weak or negligible correlations with 'Exited'. These findings imply that age, product usage, and credit card ownership are significant factors influencing customer churn.

5. METHODS

5.1. Overview of Methods

This study employs a suite of machine learning algorithms to predict customer churn, each selected for its unique strengths in handling complex datasets and classification tasks. The methodologies include logistic regression, decision trees, Random Forest, Gradient Boosting Machines (GBM), and neural networks. Hyperparameter tuning, cross-validation, and rigorous evaluation metrics are integral to optimizing model performance and ensuring reliability.

5.2. Logistic Regression

Logistic regression serves as the baseline model due to its simplicity and interpretability. This statistical approach estimates the probability of a binary outcome, such as churn, based on independent variables like age, tenure, and account balance. The model uses the sigmoid function to map predicted values to probabilities, enabling a clear threshold for classification. To enhance its robustness, techniques such as feature scaling, multicollinearity checks, and regularization (L1 and L2 penalties) will be applied.

5.3. Decision Trees

Decision trees provide a non-parametric approach, partitioning the dataset into homogeneous subsets based on feature splits. Each node represents a decision rule, and leaves indicate outcomes. The algorithm's intuitive structure is valuable for interpreting churn determinants. Pruning techniques, such as cost complexity pruning, will be implemented to prevent overfitting. Metrics like Gini impurity or entropy will guide the selection of optimal splits.

5.4. Random Forest

Random Forest, an ensemble learning method, builds multiple decision trees and aggregates their predictions for improved accuracy and stability. Each tree is trained on a random subset of data and features, reducing variance and enhancing generalization. Feature importance scores derived from Random Forest provide actionable insights into key predictors of churn. Parameter tuning, such as adjusting the number of trees or maximum depth, ensures optimal performance.

5.5. Gradient Boosting Machines (GBM)

GBM algorithms, including XGBoost and LightGBM, iteratively build models to correct errors from previous iterations. By employing gradient descent to minimize loss functions, GBMs excel in capturing non-linear relationships and handling imbalanced datasets. Advanced techniques such as early stopping, learning rate adjustment, and tree-specific hyperparameters (e.g., max depth, min child weight) will be utilized to enhance model efficiency and accuracy.

5.6. Neural Networks

Neural networks offer unparalleled flexibility and power for modeling complex patterns in high-dimensional data. This study will implement a feedforward neural network comprising multiple hidden layers, with activation functions such as ReLU for non-linearity. Dropout regularization will mitigate overfitting, while optimization algorithms like Adam will ensure efficient convergence. Techniques such as weight initialization and batch normalization will further enhance model stability and performance.

5.7. Model Evaluation Metrics

Model performance assessed used a comprehensive set of metrics:

- **Accuracy:** Overall correctness of predictions.
- **Precision and Recall:** Evaluating false positives and false negatives.
- **F1-Score:** A harmonic mean of precision and recall, balancing the trade-off.
- **Area Under the ROC Curve (AUC-ROC):** Measuring the model's ability to distinguish between classes.
- **Confusion Matrix:** Providing a detailed breakdown of true positives, true negatives, false positives, and false negatives.

5.8. Model Comparison

Random Forest, an ensemble learning method, builds multiple decision trees and aggregates their predictions for improved accuracy and stability. Each tree is trained on a random subset of data and features, reducing variance and enhancing generalization. Feature importance scores derived from Random Forest provide actionable insights into key predictors of churn. Parameter tuning, such as adjusting the number of trees or maximum depth, ensures optimal performance.

Model	Accuracy	F1-Score (Churn)	AUC
Logistic Regression	0.83	0.46	≈ 0.74
Decision Tree	0.82	≈ 0.82	-
Random Forest	0.86	0.62	-
LightGBM	0.70	0.18	0.498
Neural Network	0.70	0.18	0.498

Table 5.1. Models and Scores

Among the models tested, Random Forest emerged as the most effective classifier, demonstrating strong performance in predicting the minority churn class and maintaining a high overall accuracy. Logistic Regression suffered from poor sensitivity, while LightGBM and Neural Network performed no better than random chance. Although Decision Tree showed strong results, it posed overfitting risks.

Therefore, Random Forest was selected as the final model for prediction and dashboard visualization.

6. VISUALIZATION

The results obtained from the machine learning model were visualized using Microsoft Power BI in a clear and accessible manner. The dashboard was designed to enable business users to explore churn patterns based on customer demographics and financial attributes.

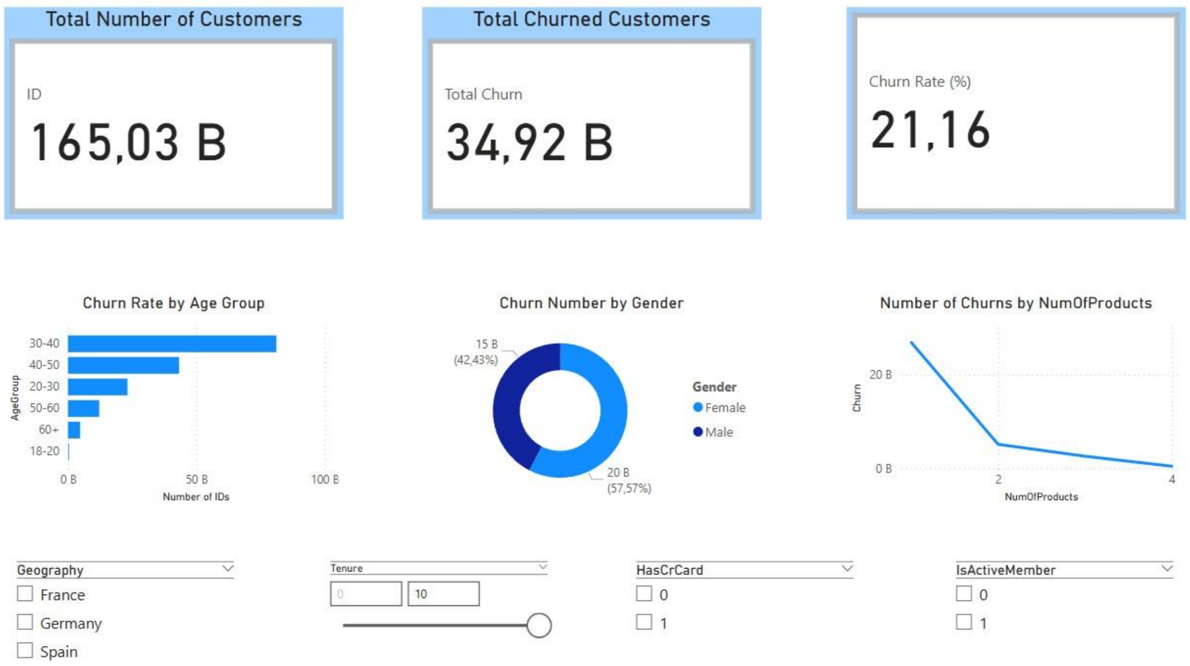

Figure 6.1. Power BI Dashboard

6.1. <u>Dashboard Content</u>

The data imported into Power BI includes prediction results generated by the Random Forest model, along with key customer attributes. The dashboard visuals were based on the following fields:

- Actual churn status (Exited)
- Model prediction (Prediction)
- Gender
- Geography

- Age
- Balance
- Estimated Salary
- Number of Products
- Active Member Status

6.2. Visual Elements and Analysis

The dashboard includes the following components:

- Churn Distribution Chart: Bar chart showing number of churned and retained customers
- Churn by Country: Analysis of churn across different countries
- Gender-wise Distribution: Comparison of churn rate between male and female customers
- Age vs. Churn: Scatter or distribution chart visualizing age-related churn trends
- Financial Charts: Visualization of Balance and Salary in relation to churn
- Filters: Interactive filters for Geography, Gender, and IsActiveMember enable dynamic analysis

6.3. Value for the End User

The dashboard allows stakeholders without a technical background to analyze churn behavior effectively. With its visual filters and simple structure, it supports the identification of high-risk customer segments and facilitates data-driven insights.

7. RESULTS AND INTERPRETATION

This project presents a data science and machine learning-based analytical system developed to predict customer churn in the banking sector. The modeling outcomes revealed that churn behavior is meaningfully influenced by several demographic and financial attributes.

Among all tested algorithms, the Random Forest model demonstrated the best performance in terms of both overall accuracy and churn-class F1 score. While Logistic Regression and Decision Tree also performed reasonably well, LightGBM and the Neural Network failed to deliver satisfactory results on this particular dataset.

Key outcomes of the project include:

- The trained models were able to predict churn with an accuracy of 86%.
- Variables such as age, gender, geography, income, and active membership status were found to be influential.
- The Power BI dashboard enabled these insights to be visualized clearly and accessibly.

Recommendations:

- Banks should develop tailored retention strategies for customer segments with high churn risk.
- Model outputs should be updated regularly by retraining with new customer data.
- The Power BI dashboard should be integrated into internal decision support systems.
- The model’s generalizability can be enhanced by using real-world operational data instead of synthetic datasets.

This study stands as a valuable example of how data-driven decision-making can be adopted in traditionally conservative sectors like banking.

8. FUTURE WORKS

The initial version of this project successfully demonstrated a machine learning-based analytical system for predicting customer churn in the banking industry. To further enhance its capabilities and practical applicability, the following future improvements are suggested:

8.1. Data Source Enhancement

The dataset used in this study is synthetic. Incorporating real operational data from a bank would significantly increase the model's credibility and generalizability.

Including time-series data (e.g., customer activity history) could enable more dynamic churn analysis.

8.2. Improving Model Performance

Class imbalance was a notable challenge. Applying balancing techniques like SMOTE or class weighting could help improve predictions for the minority class.

8.3. Deployment and Integration

Integrating the model into a web-based interface would allow field teams to make real-time decisions.

The Power BI dashboard can be connected to internal systems via API and scheduled for automated updates.

8.4. Strategic Use

The churn prediction model can be positioned at the core of customer loyalty programs by integrating it with personalized marketing strategies.

9. CONCLUSION

The comprehensive analysis and modeling efforts in this study provide valuable insights into customer churn dynamics within the banking sector. By leveraging machine learning algorithms and enriched datasets, the research identifies critical predictors of churn and enables precise classification of at-risk customers. The findings underscore the importance of demographic, behavioral, and transactional variables, emphasizing actionable strategies that banks can adopt to enhance retention rates.

The integration of a Power BI dashboard ensures that insights are presented in an accessible and interactive manner, enabling stakeholders to make data-driven decisions effectively. This tool bridges the gap between complex analytics and practical application, fostering a proactive approach to churn management. The study's methodology and results contribute to advancing predictive analytics practices in the financial industry, setting a benchmark for future research and applications.

In conclusion, this project equips banking institutions with the tools and knowledge to tackle customer churn strategically. By identifying at-risk customers and implementing targeted retention measures, banks can mitigate financial losses, improve customer satisfaction, and strengthen their competitive position in the market. This research not only addresses the immediate challenge of churn but also lays the foundation for sustainable growth and innovation in customer relationship management.

BIBLIOGRAPHY

AHN J. H., HAN S. P., & LEE Y.S. (2006). Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry.Telecommunications policy, 30(10-11), 552-568.

BILAL ZORIĆ A. (2016).Predicting customer churn in banking industry using neural networks. Interdisciplinary Description of Complex Systems:INDECS, 14(2), 116-124.

BRÂNDUŞOIU, I., TODEREAN, G., & BELEIU, H. (2016). Methods for churn prediction in the pre-paid mobile telecommunications industry. In 2016 International conference on communications (COMM) (pp. 97-100). IEEE.

CHAKISO C.B. (2015).The effect of relationship marketing on customers' loyalty (Evidence from Zemen Bank). EMAJ: Emerging Markets Journal, 5(2), 58-70.

Guliyev H., Yerdelen Tatoğlu F. (2021). Customer churn analysis in banking sector: Evidence from explainable machine learning model. Journal of Applied Microeconometrics, cilt.1, sa.2, ss.85-99, 2021 (Hakemli Dergi)

KAWALE, J., PAL, A., & SRIVASTAVA, J. (2009). Churn prediction in MMORPGs: A social influence based ap-proach. In 2009 international conference on computational science and engineering (Vol. 4, pp. 423-428). IEEE.

KERAMATI, A., GHANEEI, H., & MIRMOHAMMADI, S. M. (2016). Developing a prediction model for customer churn from electronic banking services using data mining. Financial Innovation, 2(1), 1-13.

NIE, G., ROWE, W., ZHANG, L., TIAN, Y., & SHI, Y. (2011). Credit card churn forecasting by logistic regression and decision tree. Expert Systems with Applications, 38(12), 15273-15285.

OZATAC, N., SANER, T., & SEN, Z.S. (2016). Customer satisfaction in the banking sector: the case of North Cy-prus. Procedia Economics and Finance, 39, 870-878.

Preet Singh P., Islam Anik F., Senapati R., Sinha A., Sakib N., Hossain E. (2024). Data Science and Management, volume 7, issue 1, pages 7-16.

RAJAMOHAMED, R., & MANOKARAN, J. (2018). Improved credit card churn prediction based on rough clustering and supervised learning techniques. Cluster Computing, 21(1), 65-77.

WALEED, A., PASHA, A., & AKHTAR, A. (2016). Exploring the impact of liquidity on profitability: Evidence from banking sector of Pakistan. Journal of Internet Banking and Commerce, 21(3).