Journal of Ecohumanism
2025
Volume: 4, No: 1, pp. 3478 – 3497
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i1.6188

AI-Driven Predictive Modeling for Banking Customer Churn: Insights for the US Financial Sector

Md. Sohel Rana[1], Anchala Chouksey[2], Saddam Hossain[3], Md Sumsuzoha[4], Proshanta Kumar Bhowmik[5], Miraz Hossain[6], Md Fazlay Rabby[7], Nisha Gurung[8], MD Abdul Fahim Zeeshan[9]

Abstract

The US banking sector is operating within a very dynamic and competitive environment, providing a wide array of services under the pressure of increasingly demanding customers. Customer churn in the context of financial institutions is defined as the phenomenon of customers terminating their relationship with a bank. The central tenet of this research project was to devise and develop predictive models of artificial intelligence that can help address the issue of customer churn from the banking perspective. The dataset of banking customer churn prediction used for this analysis comprises a comprehensive set of data about customers from a leading financial institution. It includes extensive customer records, each described by features representing different dimensions of customer behavior and demographics. The three most influential algorithms were selected for this study: Logistic Regression, Random Forest, and XG-Boost. Each model has different strengths that are quite appropriate for the intrinsic complexities of the customer churn forecast. Random Forest was the best in terms of accuracy among the models, with a relative accuracy, which may indicate that this algorithm fits the underlying pattern in the data best. The integration of AI-driven churn prediction models in the US financial sector has far-reaching implications for banks, enhancing their operational efficiency and customer relationship management. First and foremost, it can identify at-risk customers with a high degree of accuracy, thus helping the banks to implement focused retention strategies that can bring down the churn rate significantly.

Keywords: *Customer Churn, Banking Industry, Predictive Modelling, US Financial Sector, Artificial Intelligence, Machine Learning.*

Introduction

Context and Importance

According to Gurung et al. [9], the US banking industry is operating within a very dynamic and competitive environment, providing a wide array of services under the pressure of increasingly demanding customers. Customer churn in the context of financial institutions is defined as the phenomenon of customers terminating their relationship with a bank. Unlike other industries where churn could involve the loss of product sales, customer churn in banking is truly cascading because the process often equates to lost opportunities for cross-selling, reduced brand loyalty, and even a negative impact on long-term profitability.

Customer retention is undoubtedly paramount for the survival and growth of banks in the US. For instance, studies have shown that it costs five times more to acquire a new customer than to retain an existing one Al-Mansouri [4]. Besides, the lifetime value of a loyal customer is much greater than that of a transient customer. High churn rates not only erode the customer base but also inflate marketing and acquisition

[1] Doctorate in Business Administration and Management, University of Cumberlands, Email: mrana32911@ucumberlands.edu, (Corresponding Author)
[2] Masters of Science in Financial Mathematics, University of North Texas, Denton, Texas
[3] Master of Public Administration, Gannon University, Erie, PA.
[4] Master of Science in Business Analytics, Trine University.
[5] Department of Business Analytics, Trine University, Angola, IN, USA
[6] Master of Business Administration, Westcliff University
[7] MBA- Management Information System, American International University
[8] MBA Business Analytics, Gannon University, Erie, PA, USA
[9] Master of Arts in Strategic Communication, Gannon University, Erie, PA, USA

Journal of Ecohumanism
2025
Volume: 4, No: 1, pp. 3478 – 3497
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i1.6188

costs, impacting the bottom line. Despite these stakes, most banks poorly predict and forestall churn, leaving them vulnerable to revenue losses and competitive pressures Ahmad et al. [2].

As per Hayes & Mitchell, [7], the economic ramifications of churn are staggering: US banks are estimated to lose billions of dollars every year due to customer attrition. This loss is further increased by the increasing presence of digital-first competitors and fintech startups that promise and deliver personalized, seamless experiences to customers. To counter this, banks have to think out of the box and implement innovative solutions that go beyond traditional ways of performing churn analysis, using advanced technologies to reach the root of customer behavior.

Motivation

Buiya et al. [5], reported that AI-powered predictive modeling provides a compelling solution to employee churning challenges. Incumbent banks increasingly find themselves competing not only with their peers but also with the rising population of digital disruptors that have capitalized on their agility and customer-centric approaches. Indeed, the growth of neobanks and fintech platforms has shifted the tide of consumer expectations toward convenience, personalization, and transparency. The ability to change incumbent banks can only be made possible through their integration of more advanced technologies Islam et al., [10].

Traditional ways of dealing with churn often rely on historical data and reactive approaches. These methods cannot provide the predictive capabilities required to proactively identify customers at risk of leaving Daniel & Alexander [6]. Consequently, banks are often caught off guard, responding to churn after it has already happened. This not only reduces the effectiveness of retention efforts but also diminishes the overall customer experience. The answer to these challenges is very appealing with AI-driven predictive modeling Sumsuzoha et al. [23].

Rahman et al. [17], posited that banks can utilize machine learning algorithms coupled with big data analytics to analyze vast amounts of customer data on transaction histories, demographic information, and behavioral patterns for the subtlest signals of churn. These predictive models enable financial institutions to stay one step ahead in predicting churn and allow for the effective deployment of interventions that will lead to better customer satisfaction. AI in churn prediction is a paradigm shift in how banks address the challenge of retaining their customers.

Research Objective

The central tenet of this research project is devising and developing predictive models of artificial intelligence that can help address the issue of customer churn from the banking perspective. This research is also going to utilize machine learning in the modeling stage to help bring out the most eligible customer who plans to leave or will have different dissatisfaction issues related to their experience with the bank and possible ways and ways to eliminate their dissatisfaction so they do not get an urge of leaving such services. An overview of exactly how banks engage themselves in pragmatic efforts toward increased levels of satisfying customers to foster increased profitability will also form a key deliverable of the findings from this research.

By doing so, the research will also examine the specific attributes that influence customer loyalty in the banking context. Given such an understanding, financial institutions can then work to align their services and products to meet evolving customer needs. In turn, this would better position US banks with the concepts and knowledge relevant to sustaining long-term relationships with their customers through reduced customer loss and enhanced business performance.

Key Findings and Implications for US Banks

Expected results from this research are significant to the banking industry. The accurate predictive models developed based on the findings will allow the banks to accurately identify customers at risk and offer relevant interventions, which could improve retention rates. These can be in the form of personalized

Journal of Ecohumanism
2025
Volume: 4, No: 1, pp. 3478 – 3497
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i1.6188

offers, improvement of customer service, and products relevant to a particular need. Moreover, the insights derived from this research will assist banks in comprehending the economic impact of customer churn on their operations. In general, by providing the costs associated with churn and possible net gains in the future through improved retention strategies, banks will be well-positioned to make the right decisions as regards resource allocations and investments towards the implementation of any initiative in customer relationship management. As such, from this study, the expected finding could give a route that will eventually provide banks with competitive advantages in a seemingly hopeless environment.

In summation, this research project aims to bridge the gap between data analytics and practical application within the banking industry. Consequently, this paper intends to add significantly to AI-driven predictive modeling of customer churn in banks for the valuable benefit of helping US banks grapple with increasing complexities about customer retention amidst a constantly evolving financial scenario. The implications of this research certainly go beyond this theoretical exploration, since each finding has huge potential for translation into novelties regarding how banks organize their customer-centered operations while enhancing customer satisfaction and profitability.

Banking Customer Churn: A US Perspective

Industry Overview

Joy et al. [13], indicated that the US banking industry is considered among the most developed and competitive financial ecosystems in the world, consisting of a broad array of institutions such as commercial banks, credit unions, and investment banks. During the last decade, the industry has undergone significant changes threatened by rapid technological development, changing regulatory requirements, and shifting consumer behavior. Currently, US banks are being increasingly challenged to perform a balancing act between operational efficiency and offering personalized seamless customer experiences.

Market trends in the US banking sector have shown an increasing presence of digital channels. With increased mobile banking and digital wallets, customers want speed, safety, and ease of service delivery per Khan et al., [15]. Financial technology firms, popularly known as 'fintechs,' and neobanks have capitalized on such trends and have stolen some market share through their innovative solution offerings that appeal to technology-savvy consumers. This consequently places traditional banks under pressure to renew their operations and improve value propositions according to Kanade et al., [14].

Sanodia [20], argued that despite the exponential technological adoption by customers, consumer trust still seems to be one of the fundamentals of the banking relationship. Thereafter, perceived data security and transparency, backed by a great quality of service, are also considered key determinants influencing customer perception. These are precisely the reasons identified as causing dissatisfaction likely to lead, in turn, to churn because their expectations do not get met. Since the competition is so ruthless, customer retention is one crucial aspect through which banks would themselves need to show initiative, or else attrition risks being run.

Customer Churn Challenges

According to Yang et al. [24], customer churn in the US banking industry is propelled by a complex interplay of factors. Key causes include dissatisfaction with high fees, poor customer service, limited product offerings, and a lack of personalized experiences. Major causes are related to dissatisfaction with high fees, poor customer service, less variety of product offerings, and a lack of personalization of experiences. Given that most of the account opening processes are digital, switching banks has become very easy, which has lowered the barriers for people to search for better alternatives. This has been further heightened by the entry of fintech competitors who often provide cost-effective customer-centric solutions.

The impact of churn is all-encompassing. Apart from the loss in revenue, banks stand to suffer from reputational risk and loss in market share. High churn rates also raise acquisition costs because banks will have to spend heavily on promotional and marketing campaigns to capture new customers. Furthermore,

Volume: 4, No: 1, pp. 3478 – 3497
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i1.6188

churn upsets long-term customer relationships and reduces cross-selling opportunities, hence diminishing the lifetime value of customers as per Zarkesh, [25].

As per Zeeshan et al. [26], most of the solutions to reduce customer churn are usually not effective and sustainable. Traditional methods involve customer surveys and manual analysis, which give very limited insight into the actual reasons for churn. Besides, these methods are reactive; they focus on how to handle churn after it has happened, not how to prevent it. Although some of them have data analytics beginnings, these solutions are not sophisticated enough to delve deeply into complex and multi-dimensional data to predict their churn with a high degree of accuracy. It creates the necessity for advanced AI-enabled solutions that manage customer churn on a proactive basis.

Regulatory Environment

Shawon et al. [22], articulated that the regulatory environment of the US banking industry has a significant effect on customer retention strategies. Major regulations, such as the Dodd-Frank Act and the Gramm-Leach-Bliley Act (GLBA), stress transparency, data security, and consumer protection. For this reason, compliance with such regulations is crucial to gaining and maintaining the trust of customers, since even minor lapses can bring about reputational damage and other legal penalties. For instance, the GLBA was enacted into law to enforce customer data security and proper disclosure practices about sharing customer data by the banks. Similarly, the Dodd-Frank Act gives consumers greater protection from unfair and deceptive practices with the apparent intent of requiring financial institutions to operate with the best interests of their clients in mind.

These regulations not only define how banks must conduct business but also reinforce customer perceptions and loyalty. In retrospect, a vision toward what is over the horizon gives a much better perspective on challenges yet to be addressed in this book. Besides federal legislation, state laws such as the CCPA impose added responsibilities on banks in protecting consumer information and privacy as per Islam et al., [12]. With consumers increasingly aware of their rights, compliance with these regulations has turned out to be a serious factor in how banks maintain their customer base. Ingraining compliance within customer retention strategies can help engender greater trust and loyalty among customers, thereby minimizing the chances for customer churn according to Gurung [9].

Literature Review

Churn of Customers in the Banking Sector

According to Al-Mansouri [4], Customer churn, or the phenomenon where customers cease to renew their relationship with a financial institution, is the main problem affecting the banking industry. The gravity of customer churn in banking has been underlined by its direct influence on profitability, customer lifetime value, and the sustainability of a business. In an industry with low margins and high competition, retaining existing customers becomes ever more important. While banks typically spend significant resources on marketing and customer acquisition, the loss of a customer includes not only the current business of that customer but also the potential revenue that could have been generated by that customer in the future through continued transactions and referrals. Therefore, identifying the drivers of customer churn is crucial to banks for competitiveness and financial stability per Buiya et al., [5].

Daniel & Alexander [6], asserted that customer churn management is influenced by a variety of factors, leading to several complex challenges in its management. First, the banking sector has been going through very rapid changes in technology, which have completely altered customer behavior. Today, more financial products and services are offered than ever, with wide access to their choice, normally facilitated by fintech companies, in addition to convenience and efficiency. The convenience provided by mobile phones has, on the other hand, allowed customers to switch from one bank to another without any problems if their expectations are not well met, therefore increasing competition in customer loyalty. In addition, customer expectations have shifted; today's consumers are seeking a personalized experience, transparency, and efficiency in customer service. Hayes & Mitchell [7], contended that dissatisfaction will bring churn if banks

Journal of Ecohumanism
2025
Volume: 4, No: 1, pp. 3478 – 3497
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i1.6188

cannot meet such demands. Besides, digital banking has further complicated churn management. While it does have a lot of advantages, such as access to services at any time and the possibility to assess and compare different banking options with unprecedented ease, it puts banks in a position where they must be proactive to avoid driving away customers.

Traditional Methods

Ahmad et al. [2], articulated that traditional methods of customer churn prediction at banks are normally based on statistical techniques such as logistic regression analysis, decision trees, and survival analysis. These methods examine historical data for patterns and correlations that could indicate the likelihood of a customer churning. For instance, logistic regression can be applied to calculate the probability of a customer closing his or her account based on various demographic and behavioral factors. While these traditional approaches have provided some fundamental basic insights into customer behavior, they do have their own sets of limitations.

According to Al Montaser [3], another noteworthy limitation of the traditional methods of churn prediction is their reliance on the linearity assumption. Most of the traditional models assume linearity and additivity of the relationships between the variables, hence grossly simplifying complexities intrinsic to customer behavior. In reality, customers' decisions to leave a bank are influenced by several interrelated factors, including personal experiences, emotions, and external influences that cannot be captured by any linear model according to Kanade et al., [14]. Besides, traditional methods face serious challenges while dealing with high-dimensional datasets, where a large number of variables interact in nonlinear ways. This, in turn, could result in wrong predictions, late identification of risk customers, and failure to implement efficient intervention strategies on time. Conventional models usually require a great deal of feature engineering and domain expertise and are, therefore, generally resource-intensive to implement or burdensome Khan et al., [15].

With these limitations, Hasan et al. [8], argued that the demand for AI-driven approaches towards customer churn prediction is increasing. In this context, machine learning, as a subfield of artificial intelligence, is a strong alternative. It doesn't need much human intervention and can analyze a high volume of data to find complex patterns that traditional methods often fail to note. By applying advanced algorithms, banks can refine their predictive accuracy and get a deeper insight into the root causes behind customer satisfaction and disengagement per Joy et al. [13]. This transition to AI-powered methodologies will bring about a transformative opportunity for the banking industry by empowering institutions to become proactive in their churn management strategies and thus help retain more customers.

Machine Learning in Finance

Shawon et al. [22], postulated that Machine learning applications in finance have gained significant momentum and promise a whole new realm of possibilities in terms of decision-making processes and customer relationship management. The machine learning techniques themselves include a wide range of algorithms, including supervised learning, unsupervised learning, and reinforcement learning, which can be applied to many aspects of financial analysis. In the context of customer churn prediction, machine learning techniques have demonstrated the ability to analyze large and complex datasets, identifying patterns and relationships that can inform retention strategies according to Sumsuzoha et al., [23].

Rahman et al. [17], elucidated that one of the key advantages belonging exclusively to machine learning in financial prediction is its facility for adaptive learning. Unlike traditional financial models, which are usually based on fixed assumptions, machine learning algorithms learn continuously from new data, and their predictions keep refining over time. This makes them very adaptable, especially in the banking sector, where customer behavior is usually at the mercy of fluctuating market conditions and individual preferences. Machine learning algorithms can also handle high-dimensional data, enabling banks to expand the number of variables used in churn prediction studies. Among machine learning algorithms, techniques such as random forests, gradient boosting, and neural networks would be particularly capable of handling nonlinear

Journal of Ecohumanism
2025
Volume: 4, No: 1, pp. 3478 – 3497
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i1.6188

relationships or interactions among different variables that contribute to better accuracy in customer behavior predictions as per Islam et al., [10].

According to Zeeshan et al. [26], several distinctions come up when comparing machine learning techniques with traditional financial models. Traditional models often depend on naive assumptions and pre-defined features, which severely limit their predictive power. The machine learning algorithms, on the other hand, can pick up salient features themselves from the raw data, therefore requiring very limited feature engineering. Machine learning models can also input real-time data, thus enabling banks to quickly act on changing customer behavior and undertake proactive measures regarding the risk of churn. This is as opposed to traditional models, which tend to get outdated very fast and take a lot of time for frequent updates and recalibrations. All in all, machine learning in financial analysis introduces a paradigm shift in how banks tackle the issue of churn by making them capable of exploiting data-driven insights in devising better retention and satisfaction policies for their customers.

Key Indicators of Financial Health

According to Rana et al. [19], in customer churn prediction, the financial ratios and indicators will help in understanding the health of the customer relationship. The financial indicators usually applied are account balance, transaction frequencies, product utilization rates, and customer demographics. These metrics hold much valuable insight into customer involvement and satisfaction, thus serving as good potential predictors for churn. For instance, declines in the balance of accounts or the number of transactions can indicate dissatisfaction or disengagement and should therefore be addressed proactively. The more types of banking products customers use, the less likely they are to churn because of their more entrenched relationship with their financial institution per Sanodia, [20].

Joy et al. [13], reported that the use of different financial indicators in predicting customer churn has lately been the subject of extensive studies. Specific measures such as relationship duration, product diversification, and frequency of customer service contact can be used to indicate churners with a high degree of accuracy. For example, studies have proved that longer relationships in banking correspond to higher loyalty and lower churn rates. Building long-term relationships will require a personal touch in services and effective communication. Besides, it has been observed that the quality of customer service interactions has a great impact on the rate of churn; bad experiences will cause dissatisfaction and force customers to seek alternatives. Knowing these key indicators is important to understand how banks can build focused retention strategies as articulated by Mohaimin et al., [16].

As per Ahmad et al. [2], the predictive power of financial indicators could also be improved by integrating advanced analytics and machine learning. Using data-driven approaches, banks can identify which set of indicators best predicts their particular customer segments. For instance, machine learning algorithms can consider past churn data and determine which financial metrics most strongly predict churn behavior; thus, enabling banks to target at-risk customers. In doing so, it helps the financial institutions allocate their resources more appropriately, focusing on those interventions that are most likely to yield positive results. In the end, insights from the analysis of key financial indicators can serve as a starting point in building robust models for predicting churn, enabling guidance towards initiatives by banks to ensure better defect prevention and customer satisfaction Sizan et al., [21].

AI-Driven Predictive Modeling

Dataset Description

The dataset of banking customer churn prediction used for this analysis comprises a comprehensive set of data about customers from a leading financial institution. It includes 10,000 customer records, each described by 15 features representing different dimensions of customer behavior and demographics. The important attributes of the dataset are customer tenure, account balance, transaction frequency, product usage, and customer service interactions. It also includes demographic information, such as age, sex, and geographic location, in addition to a binary target variable that indicates whether the customer has that is,

Journal of Ecohumanism
2025
Volume: 4, No: 1, pp. 3478 – 3497
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i1.6188

closed their account within the last year. This dataset is very important in developing predictive models because it offers a great basis on which patterns and drivers of customer churn can be analyzed, hence enabling focused retention strategies that lead to enhanced customer satisfaction and loyalty.

Data Preprocessing

The Code fragments in Python performed steps for preprocessing data to prepare the data for machine learning. First, it started by dropping unnecessary columns. Secondly, it proceeded by splitting features from the target variable. Thirdly, the code facilitated the identification of categorical and numerical columns, each having different preprocessing pipelines. Fourth, a numerical one standardized by the Standard-Scaler and a categorical one, representing categorical features by the One-Hot-Encoder. Fifth, a Column-Transformer then combined these pipelines for efficient preprocessing. Finally, the code divided the data into training and testing sets, applied the preprocessing steps, and converted the preprocessed data back into Data Frames for further analysis or modeling.

Exploratory Data Analysis

Exploratory Data Analysis is a very important process in data analysis, consisting of the examination and visualization of data sets to find patterns, and anomalies, test hypotheses, and check assumptions. As opposed to confirmatory data analysis, which tests an already set hypothesis, EDA is more about exploring the underlying structure of the data and developing insights without preconceived notions. EDA is a collection of statistical and graphical methods that summarize the general features of the data, allowing the researcher to understand how data are distributed, relationships among the variable analyses, or violations in any of these that may affect subsequent analyses.

Distributed of Customer Churn (Exited)

This Python code snippet visually showcased the distribution of the target variable using matplotlib.pyplot and Seaborn. In this context, the target variable could be whether a customer has churned or not. The SNS. Counterplot () function called to create a count plot for customers who churned-1 and those who did not churn-0. The plot was customized with proper labels and a title, and then plot.show() was used to display the visualization. This plot is an effective overview of class imbalance in the dataset, which was developed when the predictive models for customer churn were built.

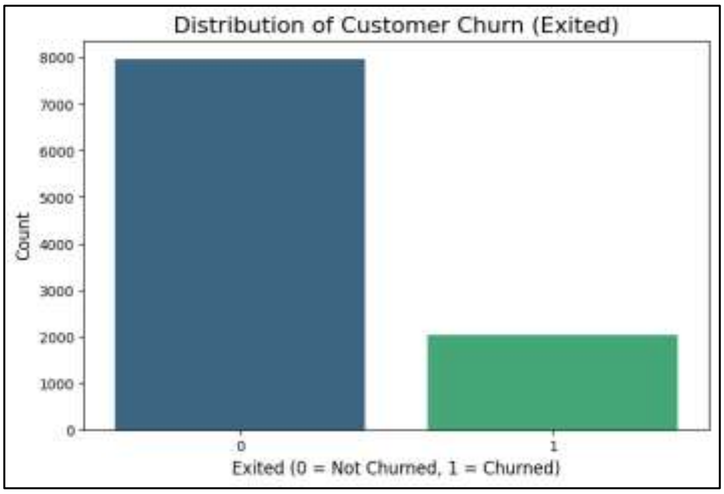

Figure 1. Distribution of Customer Churn (Exited)

The above histogram shows the distribution of customer churn. The count of exited customers is contrastingly different from the ones who stay in. There are almost 8,000 customers classified as "Not Churned" (0), showing very strong retention. The "Churned" category consists of only over 2,000

Journal of Ecohumanism
2025
Volume: 4, No: 1, pp. 3478 – 3497
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i1.6188

subscribers, while the rest highlights that about 25% of the total subscribers have chosen to leave. Such a huge gap shows that the rest of them remain subscribed, thus leaving about a quarter of the customers who have the potential to churn and need to be investigated to determine what caused their exit. Taken together, these statistics are a vital indication of the need to implement effective retention strategies aimed at improving customer loyalty and further reducing churn rates.

Customer Churn by Geography

Suitable Python code snippet that utilized the matplotlib. pyplot and seaborn libraries were adopted to create a count plot that shows how customer churn is related to geographic location. It first defines the figure size and creates the count plot using sns. Counterplot () with 'Geography' on the x-axis and 'Exited' as the hue, and further customizes this plot with a title, labels, and color palette. This visualization helps to understand whether there is a high variation in churn rates around different geographic regions, which would be useful for designing policies for retaining customers.

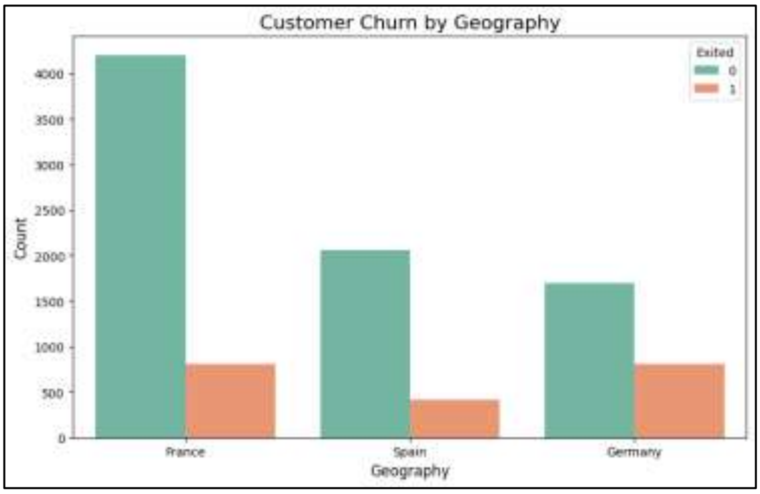

Figure 2. Customer Churn by Geography

The above histogram showing customer churn by geography is interesting, with quite considerable variations in customer retention amongst France, Spain, and Germany. France leads the lot by a good number beyond 4,000 customers, while having a decent churn rate as shown in the orange color, depicting excited customers, which is small in height compared to others. In Spain, we have 2,500, which is lower; the customers are relatively more pronounced for churn compared with France. Still, this reflects the manageable level of exits on the side of the businesses. For Germany, we have around 1,500 total customers, while for the country, the churn is higher, quite significantly bigger in proportion to the base of customers at hand. All in all, these statistics indicate that France has the most customers but also the lowest churn rate, while Spain and Germany may require specific retention policies due to an increase in this rate across both countries.

Customer Churn by Gender

Python script using matplotlib.pyplot and seaborn was computed for creating a counterplot to view the relationship of customer churn, putting into consideration one of the key factors being gender. Particularly, it provided the figure size and created a counterplot using sns. Counterplot (), with 'Gender' on the x-axis and 'Exited' as the hue, adds a title to the plot, and labels, and changes some colors. The visualization concluded whether there is a big difference in the churn rate between male and female subscribers, which may be the critical basis of targeted customer retention campaigns.

Journal of Ecohumanism
2025
Volume: 4, No: 1, pp. 3478 – 3497
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i1.6188

Output

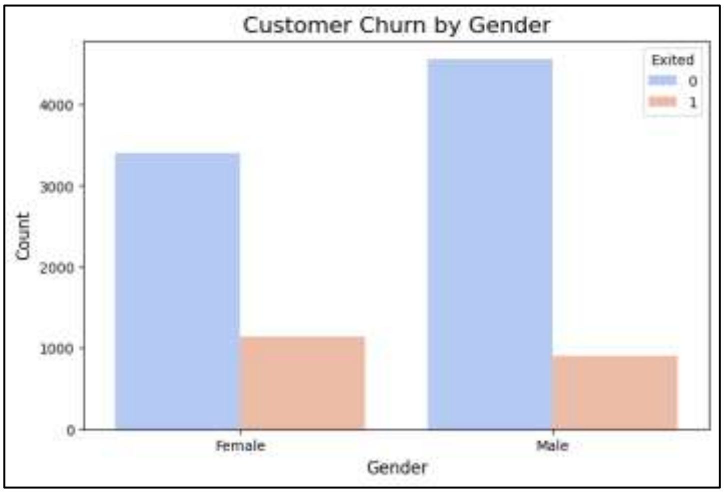

Figure 3. Displays Customer Churn by Gender

In the histogram, the customer churn by gender shows a large difference in customer retention between women and men. The total count for females is higher, about 3,500, with the relatively low churn rate indicated by the smaller orange segment which represents those who exited. In contrast, there are about 1,500 male customers, whose churn rate is also represented in orange but is much deeper than that of female customers. This indicates that females will be more attached to the service, while their churn rate for males is much higher, which may become an issue and perhaps requires some retention strategies aimed particularly at male customers to improve overall loyalty.

Pair plot of Selected Features

The code snippet in Python generated a pair plot using the Seaborn library. It visualized the relationships of the selected numerical features to the target variable ('Exited') in the dataset. The selected features include 'Credit Score', 'Age', 'Tenure', 'Balance', and 'Estimated Salary'. A pair plot creates a grid of scatterplots for each pair of features, with colors representing the 'Exited' value. Besides that, the diagonals show the Kernel Density Estimation of variable distribution. It is from this that good insight into possible relations of variables with the target feature, nonlinear relationships, or other patterns in the data could be obtained by visualization.

Output:

Volume: 4, No: 1, pp. 3478 – 3497
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i1.6188

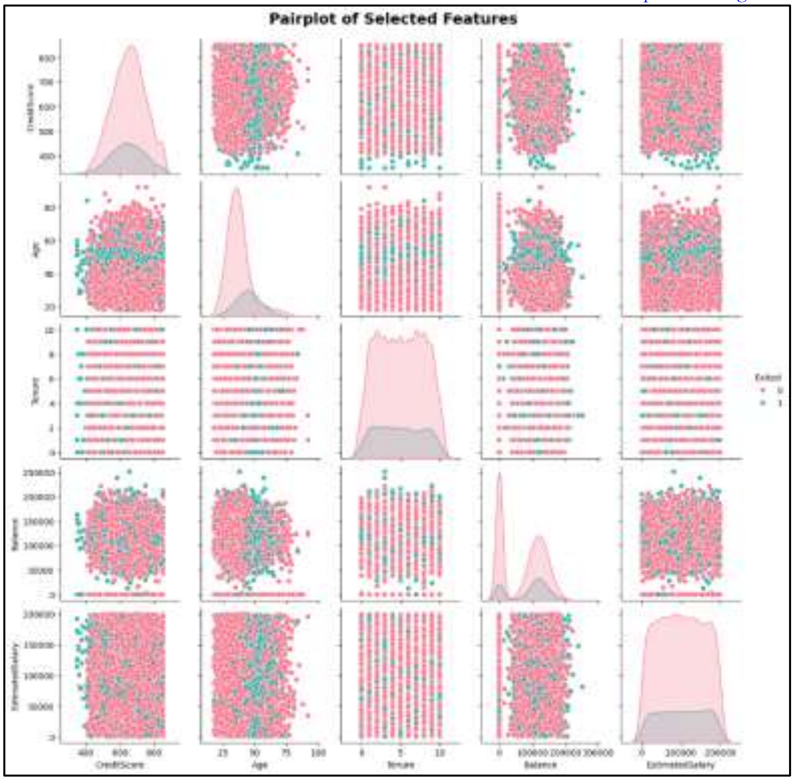

Figure 4. Pair plot of Selected Features

Above is the pair plot of selected features that show the complex relationship among variables: Credit Score, Age, Tenure, Balance, and Estimated Salary on Customer Churn. It is observed that there is a weak relationship between the variables of Age and credit score; the younger the customer, the lower the score. Balance Distribution: It can be observed from the distribution of Balance that the exited customers- are likely to have lower balances compared to the retained customers, especially in the lower ranges. Tenure also presents a very distinct pattern where longer tenures relate to a higher likelihood of retention. Besides, Estimated Salary has poor discrimination between churning and retaining customers and may indicate that salary itself is not such a strong factor in whether the customer stays or goes. A general overview is that a pair plot is quite informative concerning how these features interact, and it seems highly probable that balance and low-tenure customers are those to target for retention efforts.

Model Selection

In the quest for effective customer churn prediction within the banking industry, a careful selection of machine learning algorithms is paramount. The three most influential algorithms were selected for this study: Logistic Regression, Random Forest, and XG-Boost. Each model has different strengths that are quite appropriate for the intrinsic complexities of the customer churn forecast.

Logistic Regression is a base model because of its interpretability and simplicity of implementation. It works well for binary classification problems, such as whether a customer churns or not. One of the most salient features of Logistic Regression is that it produces probabilistic results, thus enabling banks to rank customers according to their likelihood of churning. Its simplicity allows the stakeholders to straightforwardly understand the relationship between the customer features and the likelihood of churn, thus the actionability of insights. Still, while Logistic Regression allows the identification of significant predictors, it can usually be poor in capturing nonlinear relationships and interactions among variables that are common in customer behavior.

Volume: 4, No: 1, pp. 3478 – 3497
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i1.6188

By contrast, Random Forest is a more advanced technique. As an ensemble learning model, Random Forest constructs a great number of decision trees during the process of training and then combines their outputs for more accurate and stable predictions. It is well suited for handling high-dimensional data and can, by itself, identify complex relationships between features, with little preprocessing of the features. The robustness of this Random Forest is beneficial in the banking context, where customer data might be noisy, with outliers. Besides, Random Forest returns feature importance scores that can help banks pinpoint the most influential factors driving customer churn; this interpretability of the model results is crucial for focused retention strategy implementation.

XG-Boost, or Extreme Gradient Boosting, was chosen due to its performance in various machine learning competitions and its ability to handle large datasets efficiently. It is an algorithm for a gradient boosting framework where, iteratively, model performance is optimized to learn from the mistakes of previous steps. XG-Boost works well with nonlinear relationships and interactions between different features, making it very effective in predicting customer churn. Its built-in regularization techniques also help to avoid overfitting, ensuring the generalization of the model on unseen data. The combination of speed, accuracy, and flexibility makes XG-Boost an ideal choice for any bank to enhance its predictive analytics capability.

Hyperparameter Tuning

Hyperparameter tuning was an important procedure in modeling, necessary to bring out the optimum performance of the selected machine learning models. In general, hyperparameters are the settings that control the training of machine learning algorithms, and their appropriate choice may have an immense effect on the performance of models. For this analysis, two major hyperparameter tuning approaches will be performed: Grid Search and Random Search.

Grid Search systematically explored a predefined set of hyperparameter values for each model, and among those are exhaustive evaluations to find the best combination. This procedure was helpful for models that have only a few hyperparameters because it can assure that all possible combinations will be tried. For example, for Random Forest, the hyperparameters considered included the number of trees, the maximum depth of the trees, and the minimum number of samples required to split any internal node. Since Grid Search gives an exhaustive search over specified hyperparameter combinations, it gives the best settings that can provide maximum predictive accuracy by a model. However, one notable disadvantage of Grid Search is its computational expense: increasing the number of hyperparameters or the number of values to check for each hyperparameter greatly increases the size of the search space, and thus greatly increases training time.

Model Evaluation and Comparison

Performance Metrics

Logistic Regression Modelling

Logistic Regression Modeling in Python for Binary Classification. The code snippet performed a Logistic Regression model in Python for binary classification. First of all, it imported the libraries that may be used during model building and evaluation. Then, it instantiated a Logistic Regression model, including a random state for the model reproducibility. Finally, it trained the model on the preprocessed training data and respective target variable y_train. The model performed predictions on the preprocessed test data X_test_preprocessed using the fitted model. Finally, using different metrics of performance that included accuracy, confusion matrix, and classification report, a check is provided about the performance on how well or accurately and precisely the model does in respect to recall or F1-score:

Journal of Ecohumanism
2025
Volume: 4, No: 1, pp. 3478 – 3497
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i1.6188

Output

Table 1. Portrays the Logistic Regression Results

```
Logistic Regression Results:
Accuracy: 0.811
Confusion Matrix:
 [[1543   64]
 [ 314   79]]
Classification 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
```

The Logistic Regression model yields an accuracy of 81.1%, which is a pretty good score in classifying the target variable. From the confusion matrix, it follows that the model correctly classified 1,543 true negatives and 79 true positives while misclassifying 314 actual positives as negatives and 64 actual negatives as positive. This gives a class 1 precision of 0.55, which means only 55% of its positive predictions are correct, hence casting doubt on its reliability for identifying positive cases. The recall for class 1 is low, with a value of 0.20; it shows that the model identifies only 20% of the actual positive cases, which is very low and indicates a high scope of improvement in capturing churn. Thus, class 1 has a score of 0.32 for the F1-score, showing poor balance between precision and recall. Overall, the classes do have a weaker performance depicted by the macro average-precision score and the recall score of 0.69 and 0.58, respectively; this is while the weighted averages of 0.78 for precision and 0.81 for recall might speak for somewhat better performance under consideration of class distribution. These results do suggest that while the performance of the model is quite adequate in classifying negatives, its ability to identify actual positives is considerably limited.

Random Forest Modelling

The Random Forest Classifier model in Python started by importing the necessary library Random Forest Classifier from sklearn. Ensemble. Afterward, it instantiated a Random Forest model with a random state for reproducibility. Then, the model was trained on preprocessed training data and the target variable. The trained model was used to make predictions on the preprocessed test data. In the end, the performance of the model was evaluated on various matrices such as accuracy, a confusion matrix, and classification reports, which can give information about the model's accuracy, precision, recall, and F1 score:

Journal of Ecohumanism
2025
Volume: 4, No: 1, pp. 3478 – 3497
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i1.6188

Output

Table 2. Random Forest Results

```
Random Forest Results:
Accuracy: 0.8635
Confusion Matrix:
 [[1545   62]
 [ 211  182]]
Classification Report:
              precision    recall  f1-score   support

           0       0.88      0.96      0.92      1607
           1       0.75      0.46      0.57       393

    accuracy                           0.86      2000
   macro avg       0.81      0.71      0.75      2000
weighted avg       0.85      0.86      0.85      2000
```

The Random Forest algorithm yielded a commendable accuracy of 86.35%, indicating its effectiveness in classifying the target variable. The confusion matrix shows that the correct classification of the model consists of 1,545 true negatives and 182 true positives while misclassifying 211 actual positives to negative and 62 actual negatives to positive. That means Class 1 precision of 0.75, wherein the model while predicting positivity, gets them right 75% of the time. However, class 1 recall is quite poor at 0.46, and that means that the model identifies only 46% of all actual positive cases, which might serve as a serious red flag concerning this model's effectiveness for effective churn detection. Class 1 reaches an F1 score of 0.57, which implies a real balance between precision and recall. Overall, the performance of the model for all classes is given by the macro average of 0.81 in terms of precision and 0.71 for recall, while for weighted average metrics, the values are 0.85 and 0.86, respectively, showing a better performance when class imbalances are considered.

XGB-Classifier Modelling

The Python code snippet implemented the XG-Boost Classifier model. The code began with importing the needed library XGB-Classifier from the xg-boost package. An instance of an XG-Boost model was created using the random state for reproducibility, the use-label-encoder parameter is set to False, and eval_metric to log-loss. Then it went ahead with the training on the preprocessed train data – X-train-preprocessed, with its respective target variable y-train. At last, the preprocessed test data, X_test_preprocessed, is used for prediction with the trained model. The performance of this model is evaluated using different metrics: accuracy, confusion matrix, and classification report, which allow us to gauge the performance based on accuracy, precision, recall, and F1-score.

Journal of Ecohumanism
2025
Volume: 4, No: 1, pp. 3478 – 3497
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i1.6188

Table 3. Portrays Random Forest Results

```
XGBoost Results:
Accuracy: 0.8625
Confusion Matrix:
 [[1523   84]
 [ 191  202]]
Classification Report:
              precision    recall  f1-score   support

           0       0.89      0.95      0.92      1607
           1       0.71      0.51      0.59       393

    accuracy                           0.86      2000
   macro avg       0.80      0.73      0.76      2000
weighted avg       0.85      0.86      0.85      2000
```

The XG-Boost algorithm yields a robust accuracy of 86.25%, reflecting its effectiveness in classifying the target variable. The model, from the confusion matrix, effectively identified 1,523 true negatives and 202 true positives while misclassifying 191 actual 'Positives' as 'Negatives', and vice-versa for 84. The class 1 precision values are 0.71, hence 71% of the Positives Predicted are indeed correct and highly improved rates across the different models built. However, for class 1 recall, the value is a little lower, at 0.50, indicating that the model identifies only 50% of the actual positive cases, hence some challenges in the effective detection of churn cases. The F1-score for class 1 is 0.59, hence there is a need for better balance between precision and recall. This is indicative of the overall good solid performance of the classes with the micro average of 0.80 regarding precision and 0.73 regarding recall, while weighted averages are 0.85 precision and 0.86 recall, which shows good performance concerning class imbalance. In this case, the XG-Boost model becomes promising, at least regarding accuracy, but it can still be worked upon to capture the positive cases more positively.

Comparison of All Models

The code in Python compared the performances of the Algorithms of Logistic Regression, Random Forest, and XG-Boost; it provided a dictionary that held the models' names and their corresponding accuracy results using the function accuracy score with the library sklearn. The dictionary fed the data into a pandas Data Frame for the sake of better visualization. Lastly, the code plotted a bar plot using seaborn for clear comparisons of the accuracy of each model. It was customized with a title and labels and colored appropriately for better readability.

Journal of Ecohumanism
2025
Volume: 4, No: 1, pp. 3478 – 3497
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i1.6188

Output

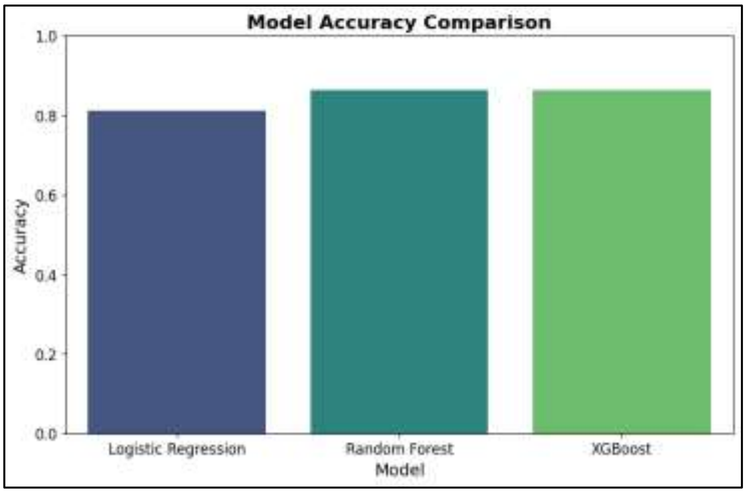

Figure 5. Depicts Model Accuracy Comparison

In the model accuracy comparison chart, it can be seen that the performance of three different algorithms is compared: Logistic Regression, Random Forest, and XG-Boost. Logistic Regression reaches an accuracy slightly above 0.80, which indicates decent classification capability. Random Forest is the best in terms of accuracy among the models, with an accuracy of more than 0.85, which may indicate that this algorithm fits the underlying pattern in the data best. XG-Boost, although a little lower than the Random Forest, maintains a strong accuracy of about 0.85 and hence is also effective. Overall, this chart shows that both Random Forest and XG-Boost are on the higher side in comparison with Logistic Regression, thereby underlining the potential for both to give more accurate predictions in this context.

Validation Techniques

Some of the most important techniques for checking the robustness and generalizability of machine learning models are validation techniques. K-fold cross-validation was considered the most helpful way to perform the verification procedure. In this protocol, the dataset was divided into k subsets or folds. The models were trained on k-1 folds and validated on the remaining one. This process was repeated k times; that is, trained and validated the model k times, each time on a different subset. The risk of overfitting decreases due to the model's performance was estimated across the more comprehensive segments. Cross-validation also used stratification techniques at the time of splitting to keep target class distribution balanced across different folds. This procedure helped separate cases of class imbalance. Depending on dataset size and particular requirements for modeling, the application of other validation methods, such as LOOCV or bootstrapping, enormously helped improve the reliability of results.

Evaluating algorithm consistency and stability across different datasets was paramount for ensuring that the model can perform well in real-world scenarios. This process was done through methods such as external validation, where the model was tested on an entirely separate dataset that had not been used either in training or in the initial validation. In this way, its performance was evaluated more objectively, helping to find potential biases and weaknesses. Besides, the performance metrics of the model, such as accuracy, precision, recall, and F1-score on different subsets or external datasets, indicated its robustness. Another way to check stability was by sensitivity analysis, which examines the change in predictions of the model with changes in input data. This holistic validation of the model not only ascertained the correctness of the model but also instilled confidence in its applicability across varied contexts, leading to better decision-making with improved outcomes in practical applications.

Journal of Ecohumanism
2025
Volume: 4, No: 1, pp. 3478 – 3497
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i1.6188

Interpretation of Model Outputs

Starting with the Logistic Regression model at an accuracy of 81.1%, the classification report shows a performance that is somewhat interesting across classes: in the case of Class 0 or Non-churners, the model is very good in both precision (0.83) and recall (0.96). Such a value suggests it is quite good at recognizing loyal customers who would not churn. However, it performed way worse on Class 1: churners- 0.55 precision and only 0.20 recall. This suggests that the model may mark a lot of non-churning customers as such with relatively high accuracy but has difficulty spotting those who may leave the bank. This is reflected in why the F1-Score for churners is the lowest at 0.29. Because of this insight, it's clear how banks need to enhance their current practices for the identification of at-risk customers. If they fail to do so, this could lead to very severe revenue loss.

By contrast, the Random Forest model had even better performance across classes, with a slightly higher accuracy of 86.35%. The model keeps a strong identification of loyal customers with a precision of 0.88 and a recall of 0.96 for non-churners. It still outperforms the churners with a precision of 0.75 and recall of 0.46, hence an F1-Score of 0.57. Thus, while it is observed that the Random Forest model is very good at classifying not-churners, on the other side, it was not that good at predicting the actual churners but was considerably better than its cousin Logistic Regression. Its higher precision means that once the model outputs a positive result when it predicts churn, it most definitely means that churn, reduces waste of retention activities on non-threatened customers.

The XG-Boost model also showed an accuracy of 86.25%, close to that of Random Forest but with different strengths. Precision for non-churners was 0.89 and recall was 0.95, similar to the Random Forest model. However, in the case of churners, the precision remained a bit behind at 0.71 while the recall only reached 0.51; thereby, the F1-Score accounted for 0.59 only. It will show that with XG-Boost performing very well while classifying as loyal, on the other side, the prediction for those vulnerable can be enhanced after more careful risk profiling. These overall performance metrics suggest that XG-Boost could be put to good use for targeted retention strategies, given its capability to model complex patterns of customer behavior.

Scenario Analysis for Different Customer Segments

Scenario analysis goes more profound into how these models can foretell, considering different customer segments based on demographic and behavioral aspects. As an example, customer grouping by age includes young adults between 18-30 years of age, middle-aged customers between 31 and 50 years of age, and seniors above 51. Each of these groups exhibits different behaviors and preferences that influence their likelihood to churn. For younger clients, who tend to have relatively lower account balances and are very likely to be more involved with fintech solutions, different predictions of models can be seen. The failure of the Logistic Regression model in correctly identifying churners may affect this segment of customers the most, who can easily switch banks due to attractive offers of more reasonable services or lower prices. On the other hand, the random forest and XG-Boost may be more effective in yielding better predictive insights that could enable banks to find young customers in a vulnerable state and reach out with more personalized offers or enhanced service options.

The middle-aged segment may also see different model runs, wherein people typically have more established banking relationships with higher account balances. In this segment, Random Forest and XG-Boost capture the subtlety of customer behavior pretty well, from which banks can devise specific retention strategies such as loyalty programs or special product offerings relevant for this particular demographic. In contrast, these models have better precision and recall for non-churners, hence banks can have greater confidence in targeting their retention strategies to customers who actually will leave the bank.

For senior clients, who may prioritize personal relationships and customer service with their banks, the insights derived from the models can inform banks about potential churn signals. XG-Boost is used in complex interactions; for this demographic, its strong performance in the case of non-churners would be able to permit banks to ameliorate customer service experiences. This, analyzed on senior-specific churn

Journal of Ecohumanism
2025
Volume: 4, No: 1, pp. 3478 – 3497
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i1.6188

predictors, such as frequency of service interactions and account management, can help banks take a proactive approach toward engaging customers to reduce their churn risks.

Background

In the banking competition, retaining customers is indispensable for long-term growth and profitability. Therefore, a well-known bank in the US, Citibank, faced some high customer churn challenges, mainly for its retail banking segment. This case describes how increasing competition, not only from traditional banks but also from fintech, urged the bank to recognize that determining the factors leading to customer attrition were important for establishing effective retention strategies. The bank's leadership recognized that the majority of their customers were leaving due to dissatisfaction with service quality, lack of personalized offerings, and better incentives from competitors.

To address these challenges, Citibank sought to exploit advanced analytics for insights into customer behavior and preference. The bank used AI-driven predictive modeling to identify customers likely to leave before they actually do, thus enabling proactive engagement strategies that improve customer satisfaction and loyalty. This adoption was not only about churn reduction but also about developing a better relationship with the customers through personalized services and communications.

Implementation

Implementation of AI-driven predictive modeling at involved a number of strategic steps aimed at effectively embedding advanced analytics into the core operations of the bank. First, there was data collection and analysis, where the bank tapped its rich repository of historical customer data, transaction records, and feedback mechanisms to create a robust profile of customer behaviors and preferences. This data formed the bedrock for developing predictive models capable of identifying patterns indicative of potential churn.

Subsequently, Citibank collaborated with data scientists to devise machine learning algorithms that could analyze data in real time. These algorithms were implemented to assess various factors contributing to customer dissatisfaction, such as transaction frequency, service interactions, and changes in account balances. It allowed identifying such customers showing signs of detachment: decreased frequency of transaction execution, sending in negative responses, among others. It prioritizes the outreach for customers at high risk of churn.

To operationalize these insights, U.S. Bank created a focused retention strategy implementation team, part of which was driven by the results of the predictive model. This includes customized communications to at-risk customers, including special incentives like lower fees or special financial products that would better serve their specific needs. Additionally, the bank integrated these predictive insights into its CRM systems to make sure proper follow-through is done with account managers.

Result

The results at Citibank from the implementation of AI-driven predictive modeling showed a high degree of impact in terms of customer retention rates. It reduced churn by 25% in the first year of deployment, as targeted interventions stemmed customer concerns before they could escalate into cancellations. By engaging those customers identified as being at risk of leaving, proactive efforts will help foster loyalty and enhance overall satisfaction.

Moreover, the customized retention strategies brought about increased retention and more cross-selling opportunities. Customers for whom personalized offers were made showed more inclination to use other banking products and services, thus increasing the bank's overall revenue. This successful initiative underlined how data-driven decisions could lead to stronger customer relationships. Furthermore, according to U.S. Bank, all that came with increased efficiency to drive those results: the partial automation of analysis and better management of communications with at-risk customers cut labor time spent on

Journal of Ecohumanism
2025
Volume: 4, No: 1, pp. 3478 – 3497
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i1.6188

manual outreach campaigns by about 30%, shifting valuable staff resources toward more complex personal customer interactions.

Discussion

Implications for the US Financial Sector

The integration of AI-driven churn prediction models in the US financial sector has far-reaching implications for banks, enhancing their operational efficiency and customer relationship management. First and foremost, it can identify at-risk customers with a high degree of accuracy, thus helping the banks to implement focused retention strategies that can bring down the churn rate significantly. The banks can make use of special machine learning algorithms that uncover styles and trends in huge data volumes, which would have otherwise been hidden from traditional methods. This capability not only supports proactive customer management but also helps the financial institution optimize resource utilization by concentrating on high-risk segments that deserve immediate attention. Furthermore, the ability to personalize offers and communications fosters deeper customer engagement, which can lead to improved customer satisfaction and loyalty over time.

However, AI-driven churn prediction will not be that straightforward to implement, with the likely hurdle being integrating the advanced AI models into the core banking systems and workflows. Most of the banks happen to operate on legacy systems that may not support the advanced data requirements of such AI applications. In such cases, financial institutions may have to invest in infrastructure and train their staff to use these new technologies. Besides, there is the challenge of ensuring the quality and consistency of data, since poor data leads to incorrect predictions and misguided strategy. Strong data governance frameworks will go a long way in mitigating such risks and ensuring the data used is accurate and current. By mitigating these challenges proactively, banks can maximize AI-driven churn prediction to further enhance their competitive advantage.

Ethical and Privacy Considerations

As banks in the USA increasingly tap customer data in a bid to drive the wheels of artificial intelligence for better churn prediction, it is with the highest hope that ethical and privacy issues will stand at the very fore of any such strategy. Above all, there are considerable concerns related to personal data privacy. Particularly in this day and age, this becomes a very important issue as more consumers become wary and concerned with the use of their information. The banks need to be more transparent in informing their customers how their data is being used to understand that their data will contribute to the process of predicting churn. Such transparency will not only help in building trust but also make the customers better equipped to make informed decisions regarding their data.

The other crucial ethical use of data in the financial sector is regulatory compliance. In the US, for example, regulations such as the General Data Protection Regulation and the California Consumer Privacy Act have set strict rules on how personal data can be collected, stored, and used. Banks must ensure that their data practices align with these regulations to avoid possible legal repercussions. Best practices that can help them include data anonymization, secure storage of data, and regular audits. Additionally, a clearly defined framework of ethics in the use of data, whereby the bank focuses on customer consent and protection of their data, would enhance reputation and help the bank build a long-term relationship with its customers.

Limitations

While promising, there are a few limitations that must be considered concerning data quality and model generalizability for the AI-driven churn prediction models. The algorithms are only as good as the data that goes in. Poor data quality in terms of inaccuracy, incompleteness, or being outdated leads to poor predictions, defeating the very reason for deploying such advanced analytics. Furthermore, biases in the data itself, from either historical inequities or methods of data collection, will also skew the model outputs

Journal of Ecohumanism
2025
Volume: 4, No: 1, pp. 3478 – 3497
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i1.6188

toward treating certain segments of customers unfairly. It is, therefore, incumbent upon banks to invest in an end-to-end cleaning and validation process that ensures sound data for training their models.

Another important limitation is the generalizability of the models across different customer segments and market conditions. While a model may perform well on the training dataset, its effectiveness diminishes when applied to new or diverse populations. This could be particularly problematic in a rapidly changing financial landscape where customer preferences and behaviors may shift due to economic factors or technological innovations. Future research should focus on developing more adaptive models that can learn and evolve with changing data inputs. Second, longer-term studies might detect changes in customer behavior over time and thus depict how churn predictors change. This allows the banks to fine-tune their retention policies. Overcoming some of these limitations will ensure that the financial sector enhances the robustness and effectiveness of AI-driven churn prediction in realizing better customer retention and improved customer satisfaction.

Conclusion

The central tenet of this research project is devising and developing predictive models of artificial intelligence that can help address the issue of customer churn from the banking perspective. The dataset of banking customer churn prediction used for this analysis comprises a comprehensive set of data about customers from a leading financial institution. It includes extensive customer records, each described by features representing different dimensions of customer behavior and demographics. The three most influential algorithms were selected for this study: Logistic Regression, Random Forest, and XG-Boost. Each model has different strengths that are quite appropriate for the intrinsic complexities of the customer churn forecast. Random Forest was the best in terms of accuracy among the models, with a relative accuracy, which may indicate that this algorithm fits the underlying pattern in the data best. The integration of AI-driven churn prediction models in the US financial sector has far-reaching implications for banks, enhancing their operational efficiency and customer relationship management. First and foremost, it can identify at-risk customers with a high degree of accuracy, thus helping the banks to implement focused retention strategies that can bring down the churn rate significantly.

References

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

Ahmad, M., Hasan, M. R., Olushola, Y. J., & Maqsood, M. (2024). AI-Infused Business Strategies: Spatial Insights for Market Success. In Empowering Entrepreneurial Mindsets With AI (pp. 101-118). IGI Global.

Al Montaser, M. A., Ghosh, B. P., Barua, A., Karim, F., Das, B. C., Shawon, R. E. R., & Chowdhury, M. S. R. (2025). Sentiment analysis of social media data: Business insights and consumer behavior trends in the USA. Edelweiss Applied Science and Technology, 9(1), 545-565.

Al-Mansouri, A. (2024). Anticipating Churn: AI-driven Insights for Sustaining Customer Loyalty in US Business Markets. Journal of Engineering and Technology, 6(1), 1-8.

Buiya, M. R., Laskar, A. N., Islam, M. R., Sawalmeh, S. K. S., Roy, M. S. R. C., Roy, R. E. R. S., & Sumsuzoha, M. (2024). Detecting IoT Cyberattacks: Advanced Machine Learning Models for Enhanced Security in Network Traffic. Journal of Computer Science and Technology Studies, 6(4), 142-152

Daniel, L., & Alexander, S. (2024). AI-Powered Client Turnover Prediction in US Business Markets. International Journal of Advanced Engineering Technologies and Innovations, 1(3), 116-131.

Hayes, J., & Mitchell, A. (2024). AI-Driven Solutions for Customer Retention and Churn Detection. Journal of Engineering and Technology, 6(2).

Hasan, M. R., Islam, M. R., & Rahman, M. A. (2025). Developing and implementing AI-driven models for demand forecasting in US supply chains: A comprehensive approach to enhancing predictive accuracy. Edelweiss Applied Science and Technology, 9(1), 1045-1068.

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

Islam, M. R., Nasiruddin, M., Karmakar, M., Akter, R., Khan, M. T., Sayeed, A. A., & Amin, A. (2024). Leveraging Advanced Machine Learning Algorithms for Enhanced Cyberattack Detection on US Business Networks. Journal of Business and Management Studies, 6(5), 213-224.

Islam, M. Z., Islam, M. S., Al Montaser, M. A., Rasel, M. A. B., Bhowmik, P. K., & Dalim, H. M. (2024). EVALUATING THE EFFECTIVENESS OF MACHINE LEARNING ALGORITHMS IN PREDICTING

Journal of Ecohumanism
2025
Volume: 4, No: 1, pp. 3478 – 3497
ISSN: 2752-6798 (Print) | ISSN 2752-6801 (Online)
https://ecohumanism.co.uk/joe/ecohumanism
DOI: https://doi.org/10.62754/joe.v4i1.6188
CRYPTOCURRENCY PRICES UNDER MARKET VOLATILITY: A STUDY BASED ON THE USA FINANCIAL MARKET. The American Journal of Management and Economics Innovations, 6(12), 15-38.

Joy, U. G., Hoque, K. E., Uddin, M. N., Chowdhury, L., & Park, S. B. (2024). A Big Data-Driven Hybrid Model for Enhancing Streaming Service Customer Retention through Churn Prediction Integrated with Explainable AI. IEEE Access.

Kanade, T. M., Batule, R. B., & Joseph, J. (2024). Leveraging Predictive Analytics for Success in Developing Economies: Integrating AI-Driven Technologies Into Service Marketing. In Integrating AI-Driven Technologies Into Service Marketing (pp. 451-476). IGI Global.

Khan, M. T., Akter, R., Dalim, H. M., Sayeed, A. A., Anonna, F. R., Mohaimin, M. R., & Karmakar, M. (2024). Predictive Modeling of US Stock Market and Commodities: Impact of Economic Indicators and Geopolitical Events Using Machine. Journal of Economics, Finance and Accounting Studies, 6(6), 17-33.

Mohaimin, M. R., Das, B. C., Akter, R., Anonna, F. R., Hasanuzzaman, M., Chowdhury, B. R., & Alam, S. (2025). Predictive Analytics for Telecom Customer Churn: Enhancing Retention Strategies in the US Market. Journal of Computer Science and Technology Studies, 7(1), 30-45.

Rahman, A., Debnath, P., Ahmed, A., Dalim, H. M., Karmakar, M., Sumon, M. F. I., & Khan, M. A. (2024). Machine learning and network analysis for financial crime detection: Mapping and identifying illicit transaction patterns in global black money transactions. Gulf Journal of Advance Business Research, 2(6), 250-272.

Rana, M. S., Chouksey, A., Das, B. C., Reza, S. A., Chowdhury, M. S. R., Sizan, M. M. H., & Shawon, R. E. R. (2023). Evaluating the Effectiveness of Different Machine Learning Models in Predicting Customer Churn in the USA. Journal of Business and Management Studies, 5(5), 267-281.

Sanodia, G. (2024). Enhancing CRM Systems with AI-Driven Data Analytics for Financial Services. Turkish Journal of Computer and Mathematics Education, 15(2), 247-265.

Sizan, M. M. H., Chouksey, A., Miah, M. N. I., Pant, L., Ridoy, M. H., Sayeed, A. A., & Khan, M. T. (2025). Bankruptcy Prediction for US Businesses: Leveraging Machine Learning for Financial Stability. Journal of Business and Management Studies, 7(1), 01-14.

Shawon, R. E. R., Dalim, H. M., Shil, S. K., Gurung, N., Hasanuzzaman, M., Hossain, S., & Rahman, T. (2024). Assessing Geopolitical Risks and Their Economic Impact on the USA Using Data Analytics. Journal of Economics, Finance and Accounting Studies, 6(6), 05-16.

Sumsuzoha, M., Rana, M. S., Islam, M. S., Rahman, M. K., Karmakar, M., Hossain, M. S., & Shawon, R. E. R. (2024). LEVERAGING MACHINE LEARNING FOR RESOURCE OPTIMIZATION IN USA DATA CENTERS: A FOCUS ON INCOMPLETE DATA AND BUSINESS DEVELOPMENT. The American Journal of Engineering and Technology, 6(12), 119-140.

Yang, T., Xin, Q., Zhan, X., Zhuang, S., & Li, H. (2024). ENHANCING FINANCIAL SERVICES THROUGH BIG DATA AND AI-DRIVEN CUSTOMER INSIGHTS AND RISK ANALYSIS. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 3(3), 53-62.

Zarkesh, B. (2023). Exploring the Impact of AI-driven Pricing on Customer Loyalty and Churn Rates in the Banking Industry (Master's thesis, NTNU).

Zeeshan, M. A. F., Sumsuzoha, M., Chowdhury, F. R., Buiya, M. R., Mohaimin, M. R., Pant, L., & Shawon, R. E. R. (2024). Artificial Intelligence in Socioeconomic Research: Identifying Key Drivers of Unemployment Inequality in the US. Journal of Economics, Finance and Accounting Studies, 6(5), 54-65