International Journal of Applied Mathematics

Volume 38 No. 4s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

PREDICTIVE ANALYTICS FOR CUSTOMER CHURN IN BANKING: A MACHINE LEARNING APPROACH TO RETENTION

Joby Jacob[1], Prof. Dr. E. J Thomson Fredrik[2]

[1]*Research Scholar, Karpagam Academy of Higher Education (Deemed to be University), Coimbatore, India.*

[2]*Research Guide, Karpagam Academy of Higher Education (Deemed to be University), Coimbatore, India.*

[1]jobyjacob1975@gmail.com, https://orcid.org/0009-0004-7501-2417,
[2]thomson500@gmail.com, https://orcid.org/0000-0003-1309-1936

Abstract:

Using cutting-edge AI methods focused on machine learning models like Random Forest, XGBoost, and Logistic Regression, this study investigates the prediction of customer attrition in the banking industry. The highest accuracy of 81.07% is achieved from Logistic Regression. XGBoost achieves 80.31% with similar results. The result of feature importances analysis shows that TotalCharges, MonthlyCharges and Contract have the highest influence on churn. Further clustering breaks customers down into an actionable group for targeted retention strategy. The results indicate that predictive analytics can be used to reduce churn, boost customer satisfaction, and improve the performance of the business, thus requiring data-based customer retention strategies in banking.

Keywords: Customer Churn Prediction, Banking Sector, Machine Learning, Decision Tree, Logistic Regression, XGBoost, Random Forest, Predictive Analytics, Feature Importance, Customer Retention, K-Means Clustering, Churn Drivers, MonthlyCharges, TotalCharges, Contract Type, AI Technologies, Model Evaluation, Data Analysis.

I. Introduction

Maintaining the customer base in the banking industry is another strategic importance since it is cheaper to serve existing clients than to source new ones. Using the data set analysis, the study shows that 26.54 per cent of customers churn, indicating a major problem to address in banking. The analysis establishes that churn is directly affected by such parameters as MonthlyCharges, TotalCharges, and Contracts. The more complex prediction options include Logistic Regression at 81.07% and XGBoost at 80.31% accuracy on churn prediction. Through the usage of Artificial Intelligence-based predictive modelling techniques coupled with feature

Received: August 11 , 2025

Volume 38 No. 4s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

importance analysis, banks can drive more accurate retention strategies, thus boosting customer satisfaction and reversing the high churn rate for better business performance.

Research Aim and Objectives

Aim

The purpose of this research is to use customer segmentation and predictive analytics models to increase client loyalty in banking.

Objectives

- Using historical data, create a machine learning framework for customer churn (Support Vector Machine, Random Forest, XGBoost) to assist banks in identifying at-risk clients.
- To further explore retention strategies that concentrate on the behaviour and interests of certain segments by using tools such as the K-means algorithm to cluster consumers into segments depending on their activity.
- In order to better understand customers and predict customer turnover and preferences, the features of the built prediction models and clustering algorithms will be compared.
- To offer suggestions and targeted retention models based on the research, allowing banking organizations to implement efficient retention strategies to boost client loyalty and reduce attrition.

II. LITERATURE REVIEW

2.1: AI in Enhancing Customer Retention through Predictive Insights

Predictive insights about the demands of the market and customer behaviour and preferences has no place like Artificial Intelligence to back your customer retention strategies with. AI powered tools analyze massive datasets to find patterns and trends that indicate ‘churn’ (customers leaving) so businesses can do something proactive when that happens. Seeing a customer off on a journey, predictive analytics, driven by AI, examines historical data, transactional behaviour and demographic information to forecast the customer’s next decision.

A common example is which AI models use machine learning algorithms to bring those customers that are at risk of churning. Learning from these insights enables companies to create targeted retention campaigns like personalized offers or loyalty programs that make sense to each customer (Kasem et al., 2024) [1]. For example, AI driven systems in any retail or subscription-based systems can be used to suggest products or services specific to a customer’s browsing or purchase history, improving customer satisfaction and loyalty.

Dynamic segmentation is another advantage of AI since customers are categorized differently according to behavioural patterns as opposed to static demographics. This in turn enables the business to deal with particular needs more effectively which in turn creates a stronger relationship with a customer. Real time also gives you the chance to solve issues quickly, by resolving grievances or offering incentives as the customer enters your interactions.

Received: August 11 , 2025

Volume 38 No. 4s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Furthermore, AI enabled tools fit right in with a Customer Relationship Management (CRM) system and provide you with data that is actionable, so that customer service teams can provide great experiences. There are also very important roles of chatbots and virtual assistants to address queries and keep the engagement.

Using predictive and proactive strategies, given AI, customer experience is improved, churn rates are lowered, and profitability is increased. By aligning their offerings with current customer expectations, businesses leveraging these advanced capabilities secure themselves a win because they are ensuring retention levels are sustainable.

2.2: Integration of Machine Learning Algorithms in Customer Engagement

Businesses are interacting with customers in more inventive and individualized ways thanks to machine learning (ML) algorithms. These algorithms can be used by a business to examine the volume of consumer data in order to identify trends, preferences, and behaviours that will enable it to better customize its engagement tactics.

Recommendation systems are a primary use of ML in customer engagement. Collaborative filtering and deep learning are models that are used by platforms like Netflix, Amazon and Spotify to suggest products or content that are aligned with user's preferences (Potla & Pottla, 2024) [2]. Because these personalized recommendations not only enhance the user experience, but also improve repeat interactions, they lead to loyalty.

Moreover, ML helps in providing predictive engagement to the customers. Algorithms are in a position to analyze past interactions and real time data to decide at what point and which channel to communicate with customers to drive the best response. For example, the predictive models might indicate that we should deliver promotional emails during certain hours when the customer will be more responsive.

ML also has an enormous role in the area of sentiment analysis. It means that businesses can measure the sentiment that an ordinary customer has towards them by going through the social media post, review or feedback to adjust their strategy accordingly. It allows us to point out to dissatisfied customer, and try fixing their issues before they do.

Chatbots and virtual assistants are also made ML powered requiring instant response, accuracy and contextually relevant response in customer support. The systems improve at learning from interactions to address customer needs eventually until they continue to get better and better. Similarly, companies may maintain a 360-degree view of client interactions and provide a smooth, personalized touch at every touchpoint by integrating ML algorithms with CRM systems. This allows you to offer your customers increasingly higher levels of engagement when they buy the same exact experience time and time again.

2.3: AI-Driven Automation in Customer Relationship Management (CRM)

Automation in Customer Relationship Management (CRM) is changing how Customer Relationship Management (CRM) is done using AI by creating a more productive and

Received: August 11 , 2025

Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

personalized CRM. AI will automate your simple tasks, gives you actionable insights, and supports your decisions and allow you to have better targeted customer relationships.

One of its principal applications is automating Data Management. AI tools can help you collect, process and organize massive amounts of customer data from several sources. It cuts manual work and keeps customer data accurate, which makes it possible for businesses to have detailed customer profiles (Singh, 2023) [3]. Understanding customer preferences, and tailoring engagement strategies is invaluable, and those profiles help.

Instant reply to queries is delivered by AI-powered chatbots and virtual assistants, a totally transforming customer support from its earlier boring days. Unlike traditional methods these systems run 24/7 so the availability is consistent. Typically, they learn from interactions how to better handle complex queries over time. Not only does it make for a better customer experience, but it reduces the burden on human support team.

Predictive analytics is another advantage to AI when it comes to CRM — it helps find out trends and predict customer needs. For instance, AI models can tell when a customer will probably need certain things, and this is when businesses can provide them, or advise them, according to the need. There's that proactive approach that helps strengthen customer loyalty.

Furthermore, AI simplifies lead scoring and prioritization. With the help of machine learning models, sales teams can allocate their efforts more strategically—to high value leads—with the use of historical data to inform the model. Sentiment analysis technologies also examine social media interactions and consumer comments to identify areas for improvement as well as positive and negative reviews.

Combining AI driven automation with CRM platforms brings the power of a complete system around personalization, process efficiency, or seamless communication across the channels. Moving to these technologies will allow businesses to provide amazing customer experiences, thus retaining and keeping them coming back for more.

2.4: Comparative Analysis of Traditional vs. AI-Based Customer Retention Strategies

Traditional customer retention strategies often utilized generalized approaches like loyalty programs, promotional campaigns and post purchase follow-ups. Though these methods showed promise on some level, they were not precise or scalable enough to yield significant impact in today's data driven world. First, however has revolutionized retention through hyper personalization (using AI based strategies) and predictive analytics.

Conventional approaches largely depended on static client data, such as demographics or purchases. Conversely, AI-based approaches examine dynamic data, including social media activity, real-time interactions, and behavioural trends. Businesses are better able to understand the needs of their customers and offer a customized experience as a result.

The other main difference is scalability. Manual effort is required in traditional approaches, which renders them resource intensive for unlimited scaling abilities. In contrast, AI powered systems take control of processes like data analysis, customer segmentation and

Received: August 11 , 2025

International Journal of Applied Mathematics

Volume 38 No. 4s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

communication thereby serving customer data without being inconsistent with large customer base.

AI also excels in predictive capabilities. Traditional methods often deal with churn after it happens, but AI models predict churn risk and suggest necessary interventions in time (Islam et al., 2023) [4]. For example, if we want to keep customers from leaving, machine learning algorithms can pick up on the slightest behavioural changes and put it to the business to target these customers with a retention strategy.

In addition, traditional strategies use surveys and feedback heavily to understand how customers are pleased. While useful, these methods can be time consuming and biased. With sentiment analysis, AI gives real time insights on customer sentiment so that businesses can react quickly (Sibanda et al., 2020) [5].

III. METHODOLOGY

Using advanced machine learning algorithms to understand and evaluate the data, this study shows a systematic approach to customer churn analysis. The complete process consists of several steps, including data pretreatment, exploratory data analysis (EDA), predictive modeling, and model evaluation and interpretability.

a. *Preprocessing Data*

Preparing the dataset for analysis is the initial step. Next, look for missing values in the ollection that have been imputed by the median for numerical.

b. *Analysis of exploratory data (EDA)*

Exploratory data analysis is used to find patterns and relationships in the dataset (Adeniran et al., 2024) [6]. The relationship between features and customer turnover is ascertained using feature distribution analysis, which is evaluated using visual aids including box plots, count plots, and histograms.

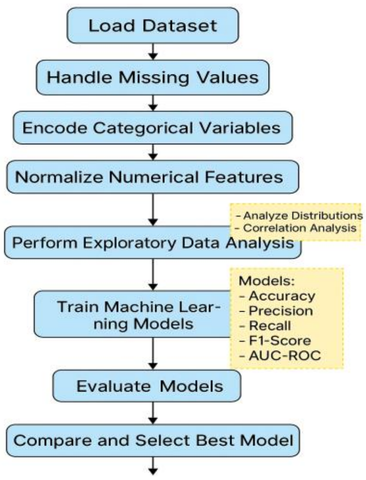

Fig 1: Methodology Flowchart for Customer Churn Prediction

Received: August 11 , 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

c. *Predictive Modeling*

Machine learning models like Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and XGBoost are used to forecast client attrition. Each model is trained using a processed training dataset, and to optimize performance, hyperparameters are corrected using a validation dataset (Agu et al., 2024) [7]. Because they can both learn non-linear relationships and produce feature importance scores, XGBoost and Logistic Regression are utilized, while Random Forest provides a baseline. To increase model reliability, the models are trained on balanced data, and their outputs are examined using a variety of performance criteria.

d. *Model Evaluation and Interpretability*

Metrics such as accuracy, precision, recall, F1-score, and AUC-ROC are used to evaluate the models. Classification performance is assessed over the churn and non-churn classes via the visualization of confusion matrices. ROC curves are plotted for each model and evaluate how well it can distinguish between classes. Random Forest and XGBoost features are analyzed to gain feature importance to churn predictions. Additionally, SHAP (SHapley Additive exPlains) values are computed to more comprehensively understand the contribution of features. The interpretation of the findings makes it possible for them to be actionable and in line with business objectives. The methodology adheres to a systematic process of churn analysis at a high level of accuracy and utilizes advanced machine learning methods to create meaningful insights (Segun-Falade et al., 2024) [8]. Combining this structured approach to balance prediction accuracy with interpretability allows stakeholders in this space to derive targeted strategies for the retention of customers (Rathod, 2023) [9].

Iv. Result And Discussion

Analysis

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
0	1	15634602	Hargrave	619	France	Female	42.0	2	0.00	1	1.0	1.0	101348.88	1
1	2	15647311	Hill	608	Spain	Female	41.0	1	83807.86	1	0.0	1.0	112542.58	0
2	3	15619304	Onio	502	France	Female	42.0	8	159660.80	3	1.0	0.0	113931.57	1
3	4	15701354	Boni	699	France	Female	39.0	1	0.00	2	0.0	0.0	93826.63	0
4	5	15737888	Mitchell	850	Spain	Female	43.0	2	125510.82	1	NaN	1.0	79084.10	0
5	6	15574012	Chu	645	Spain	Male	44.0	8	113755.78	2	1.0	0.0	149756.71	1
6	7	15592531	Bartlett	822	NaN	Male	50.0	7	0.00	2	1.0	1.0	10062.80	0
7	8	15656148	Obinna	376	Germany	Female	29.0	4	115046.74	4	1.0	0.0	119346.88	1
8	9	15792365	He	501	France	Male	44.0	4	142051.07	2	0.0	NaN	74940.50	0
9	10	15592389	H?	684	France	Male	NaN	2	134603.88	1	1.0	1.0	71725.73	0

Fig 2: Preview of the Datase

This figure features the first rows in the customer dataset, highlighting CreditScore, Geography, Age, Tenure, Balance and whether the customer churned (Challoumis, 2024a) [10]. It presents the organization of the data and what kinds of customer characteristics are included in the analysis.

Received: August 11 , 2025

International Journal of Applied Mathematics

Volume 38 No. 4s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

	RowNumber	CustomerId	CreditScore	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exite
count	10002.000000	1.000200e+04	10002.000000	10001.000000	10002.000000	10002.000000	10002.000000	10001.000000	10001.000000	10002.000000	10002.00000
mean	5001.499600	1.569093e+07	650.555089	38.922311	5.012498	76491.112875	1.530194	0.705529	0.514949	100083.331145	0.20375
std	2887.472338	7.193177e+04	96.661615	10.487200	2.891973	62393.474144	0.581639	0.455827	0.499801	57508.117802	0.40281
min	1.000000	1.556570e+07	350.000000	18.000000	0.000000	0.000000	1.000000	0.000000	0.000000	11.580000	0.00000
25%	2501.250000	1.562852e+07	584.000000	32.000000	3.000000	0.000000	1.000000	0.000000	0.000000	50983.750000	0.00000
50%	5001.500000	1.569073e+07	652.000000	37.000000	5.000000	97198.540000	1.000000	1.000000	1.000000	100185.240000	0.00000
75%	7501.750000	1.575323e+07	718.000000	44.000000	7.000000	127647.840000	2.000000	1.000000	1.000000	149383.652500	0.00000
max	10000.000000	1.581569e+07	850.000000	92.000000	10.000000	250898.090000	4.000000	1.000000	1.000000	199992.480000	1.00000

Fig. 3: Descriptive Statistics of Customer Churn

Central tendencies of features such as Age, Tenure, Balance and EstimatedSalary, are summarized using descriptive statistics (Dabo & Hosseinian-Far, 2023) [11]. It spotlights the common statistics of spread—mean, median and quartiles—to help find customer profiles and unusual cases that may lead to churn.

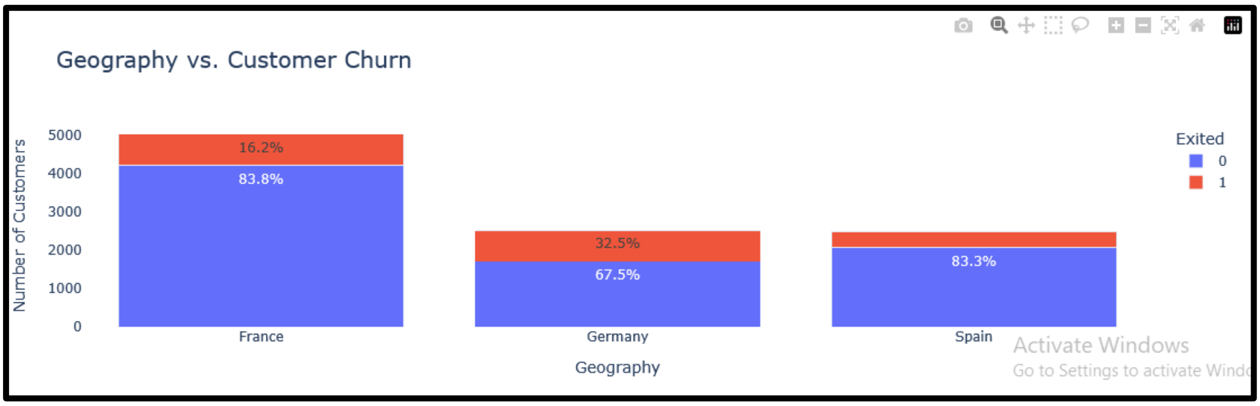

Fig. 4: Distribution of Customer Churn Based on Geographic Location

The graph shows churn rates grouped by areas across the business. This suggests that France holds the biggest customer base and also has less customer departure, while Germany has more customers leaving which suggests that area differences are key to how many customers (Bhat et al., 2023) [12].

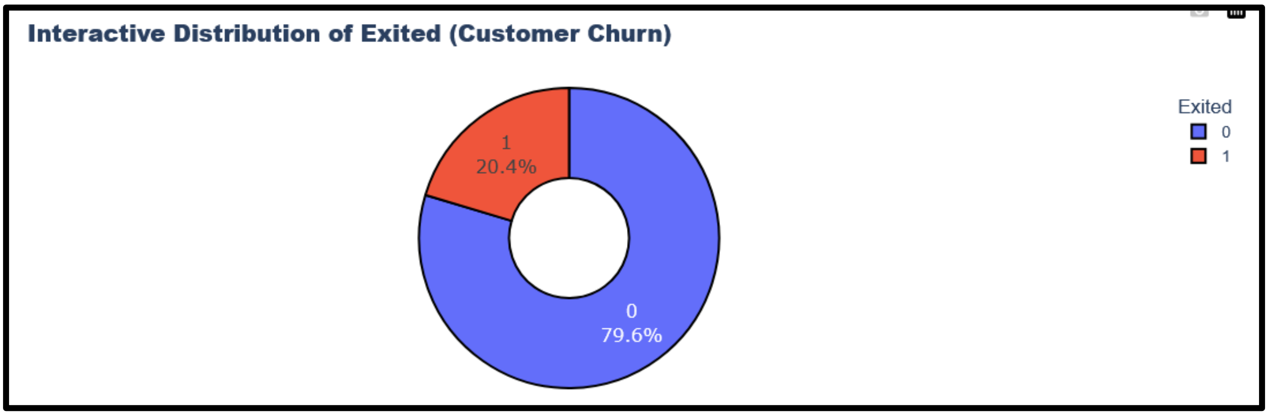

Fig. 5: Churn Distribution

Received: August 11 , 2025

International Journal of Applied Mathematics
Volume 38 No. 4s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Out of all customers, the interactive donut pie chart indicates that about 20.4% have churned. With this visual, the percentage of churn is quickly viewed, helping assess first steps to improve retention.

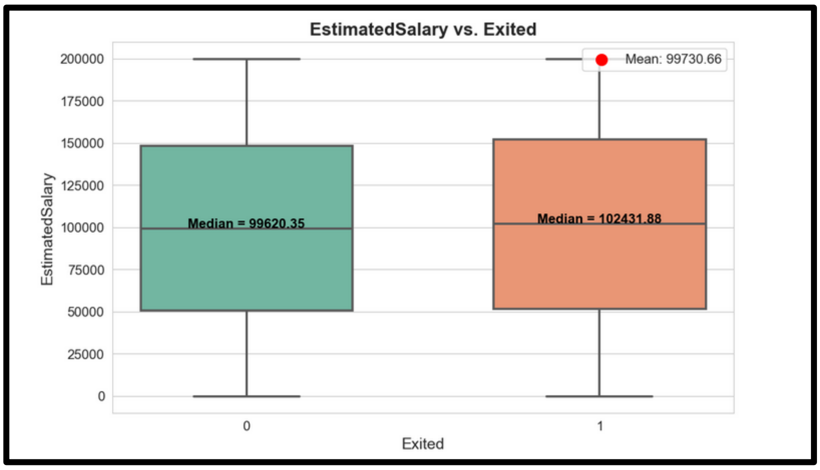

Fig. 6: Estimated Salary Based on Customer Churns

Estimated salaries for customers who chose either option are shown in this comparison (Challoumis, 2024b) [13]. Churned customers tend to earn a higher median income which may show that income is linked to banking churn behaviour.

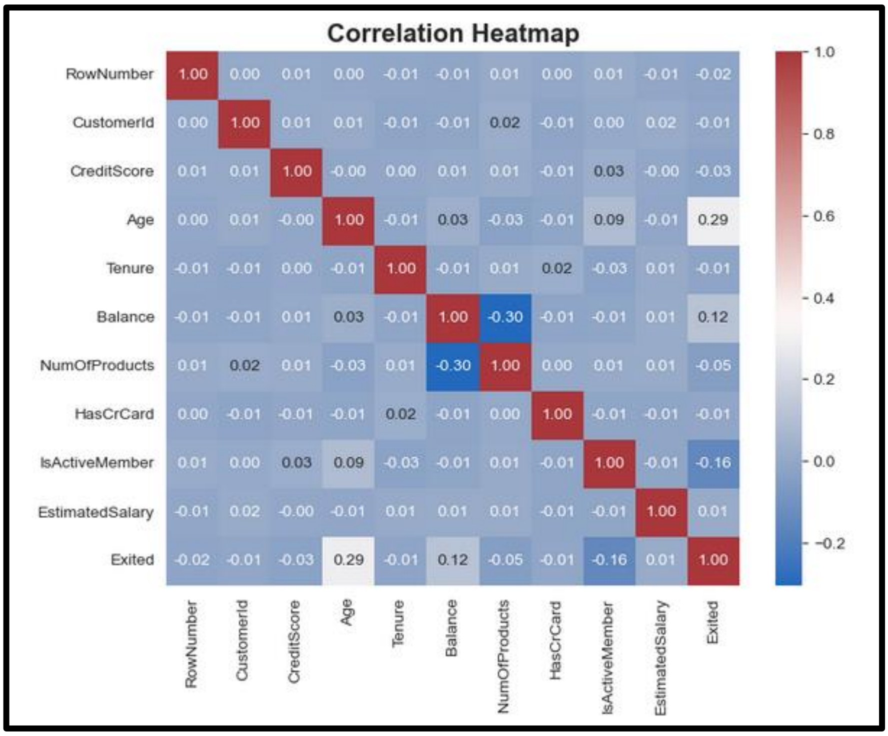

Fig. 7: Correlation Matrix

Received: August 11 , 2025

International Journal of Applied Mathematics

Volume 38 No. 4s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The heatmap shows the correlation between each of the numerical features. Moreover, customers with older accounts are slightly more likely to leave, however, features such as Balance and CreditScore have either weak or negative influences on whether a customer stays.

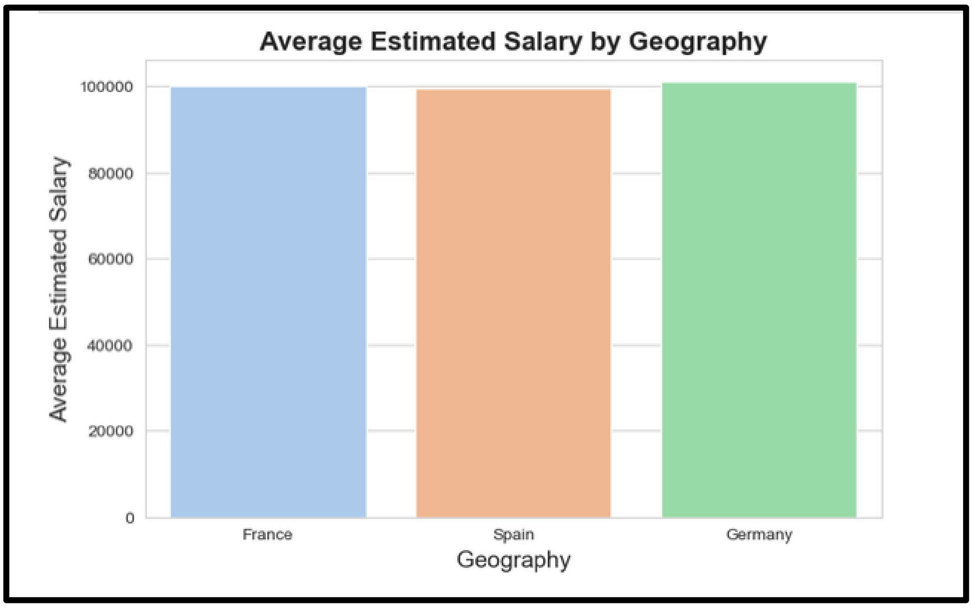

Fig. 8: Average Estimated Salary by Geographic Location

From the bar chart it is clear that average estimated salaries remain fairly similar in France, Spain and Germany. It appears that where customers are placed geographically does not strongly impact their salaries in this dataset.

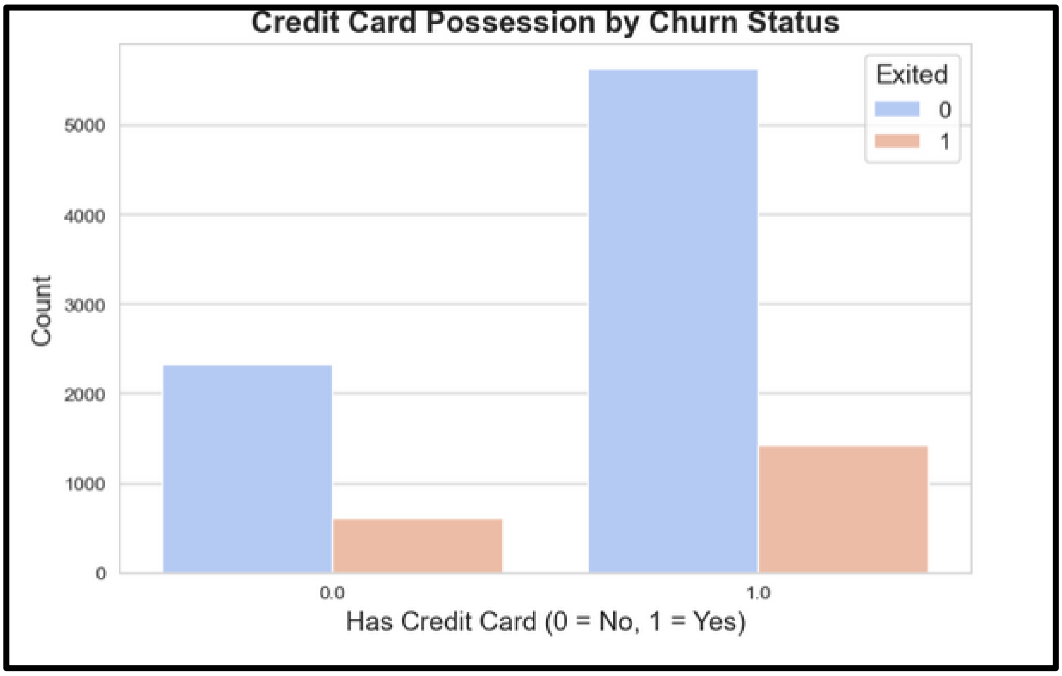

Fig. 9: Credit Card Possession by Churn Status

Received: August 11 , 2025

International Journal of Applied Mathematics

Volume 38 No. 4s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Most customers, whether they churn or not, seem to have credit cards (Karim et al., 2022) [14]. Yet, customers who don't possess a credit card have a slightly higher tendency to cancel their accounts.

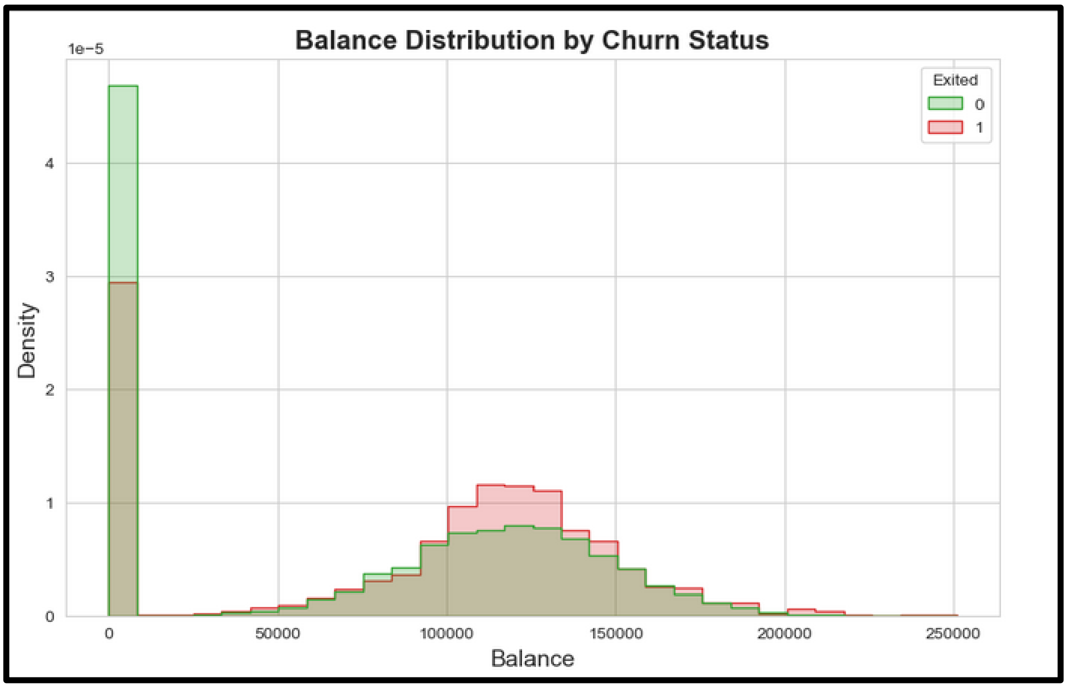

Fig. 10: Balance Distribution by Churn Status

Density plots help see how evenly spread the churned and retained customer groups are. Although both groups have similar money-in and money-out transactions, churned customers' balances tend to be somewhat higher, so balance might influence churn.

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
0	1	15634602	Hargrave	619	France	Female	42.0	2	0.00	1	1.0	1.0	101348.88	1
1	2	15647311	Hill	608	Spain	Female	41.0	1	83807.86	1	0.0	1.0	112542.58	0
2	3	15619304	Onio	502	France	Female	42.0	8	159660.80	3	1.0	0.0	113931.57	1
3	4	15701354	Boni	699	France	Female	39.0	1	0.00	2	0.0	0.0	93826.63	0
4	5	15737888	Mitchell	850	Spain	Female	43.0	2	125510.82	1	1.0	1.0	79084.10	0
5	6	15574012	Chu	645	Spain	Male	44.0	8	113755.78	2	1.0	0.0	149756.71	1
6	7	15592531	Bartlett	822	France	Male	50.0	7	0.00	2	1.0	1.0	10062.80	0
7	8	15656148	Obinna	376	Germany	Female	29.0	4	115046.74	4	1.0	0.0	119346.88	1
8	9	15792365	He	501	France	Male	44.0	4	142051.07	2	0.0	1.0	74940.50	0
9	10	15592389	H?	684	France	Male	37.0	2	134603.88	1	1.0	1.0	71725.73	0

Fig. 11: Preprocessed Data

The data captured in this picture is clean and organized, including all customer records and is ready for any type of modelling (Reddy & Nalla, 2024) [15]. It points out how making variables consistent leads to better predictions for churn.

Received: August 11 , 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Churn Prediction Model Results:

	Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
0	Logistic Regression	0.810696	0.681250	0.569686	0.620493	0.858096
1	Decision Tree	0.734027	0.510949	0.487805	0.499109	0.657273
2	Random Forest	0.797918	0.675418	0.493031	0.569990	0.837027
3	SVM	0.807383	0.712468	0.487805	0.579111	0.808359
4	XGBoost	0.803124	0.674779	0.531359	0.594542	0.853106

Fig 11: Churn Prediction Model Results Table

This figure provides a summary of the models' performance parameters. An impressive performance is provided by Logistic Regression which yielded an accuracy of 81.07% and an AUC-ROC of 85.8. SVM has the lowest values of precision with 71.24% hence reducing the possibility of having false positives. Decision Tree – has the lowest accuracy and AUC-ROC score of 65.7% (Chang et al., 2024) [16]. The table offers a snapshot of every model concerning its advantages and limitations and can be used to make informed choices regarding customer retention strategies for the selection and usage of models.

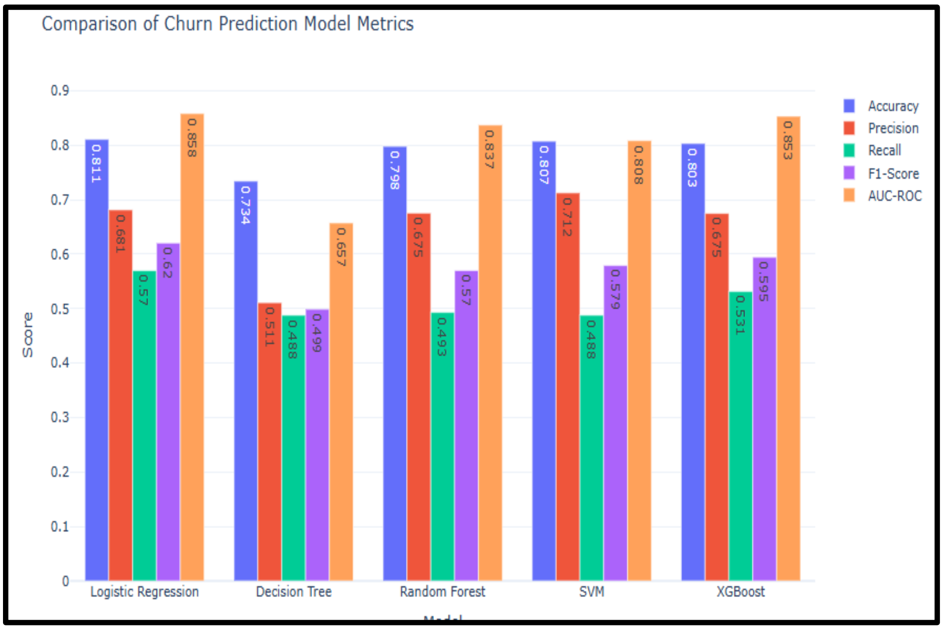

Fig 12: Comparison of Churn Prediction Model Metrics

Received: August 11 , 2025

International Journal of Applied Mathematics

Volume 38 No. 4s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

This grouped bar chart compares values such as accuracy, precision, recall, and F1-score in various models. The Logistic Regression and the XGBoost maintain a high average accuracy, while there is low accuracy in the Decision Tree. (Yaqoob et al., 2023) [17] This is why it emerges as strong in precision to offer reliability in the elimination of false positives. The tradeoff between metrics appears in the visual format to determine the appropriate business model depending on business needs, like reducing churn or retaining valuable customers.

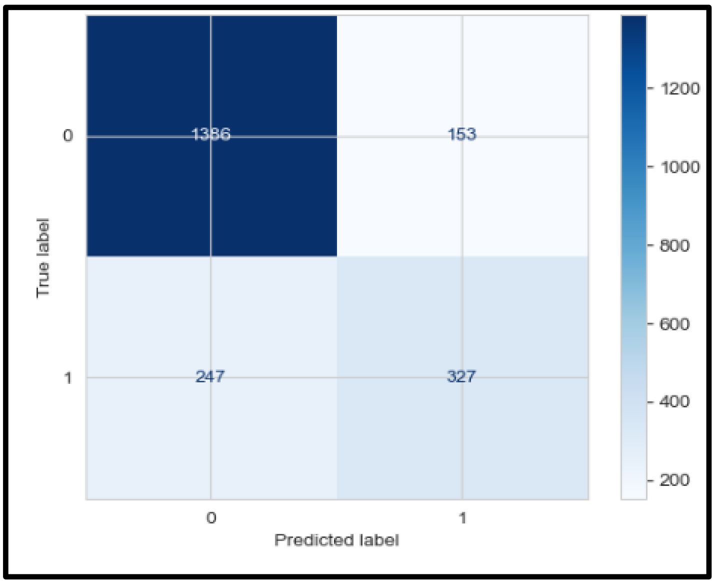

Fig 13: Confusion Matrix for Logistic Regression

In the confusion matrix, the true negatives are 1386, true positives are 327, false positives are 153 and the false negatives are 247 (Challoumis, 2024c) [18]. Hence, while Logistic Regression offers the best way to accurately identify customers who may not churn, it has a major drawback of a high False Negative Rat,e which affects recall.

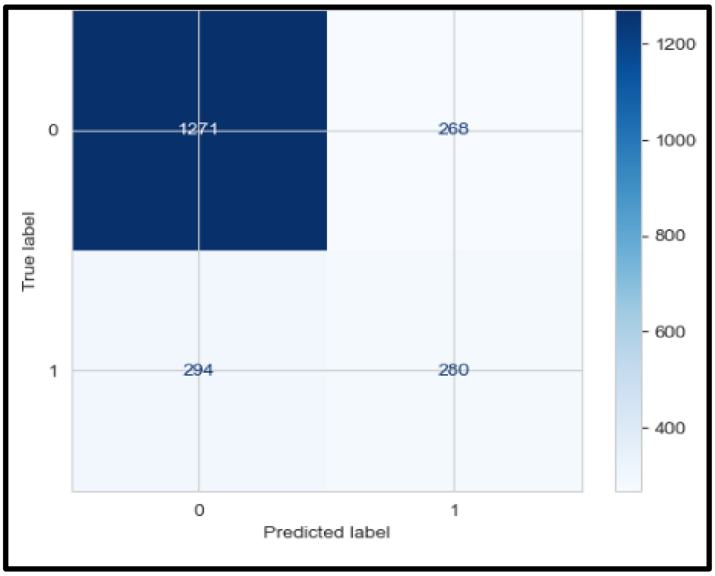

Fig 14: Confusion Matrix for Decision Tree

Received: August 11 , 2025

International Journal of Applied Mathematics

Volume 38 No. 4s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The Decision Tree confusion matrix shows its weakness: the true negatives equal 1,271, true positives equal 280, false positives equal 268, and false negatives equal 294. However, the model has higher false positives and false negatives which lower the reliability of the model. This figure highlights the problem of sheer versus accuracy since the performance of the decision Tree model is far lower than complex models like XGBoost and SVM.

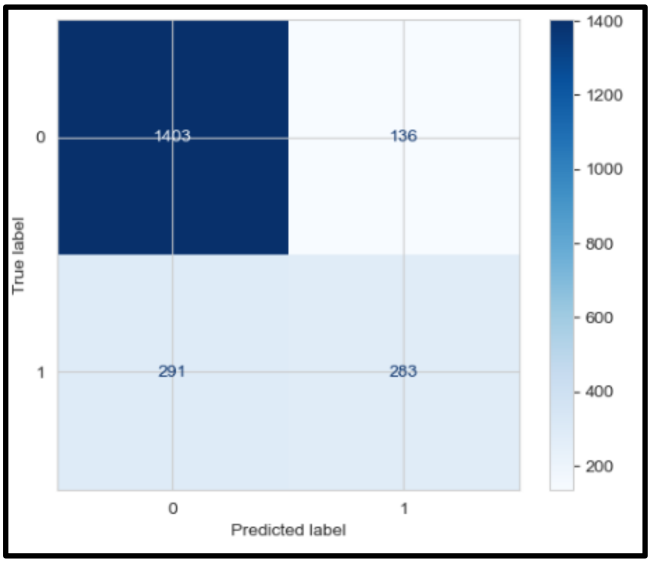

Fig 15: Confusion Matrix for Random Forest

Random Forest becomes 1,403 true negatives, 283 true positives, 136 false positives, but 291 false negatives. This means that the model performs well in identifying the loyal customers while struggling to identify the churners (Thowfeek et al., 2020) [19]. That figure gives an equal representation of the model and seems to bring out areas of concern as far as the recall is concerned.

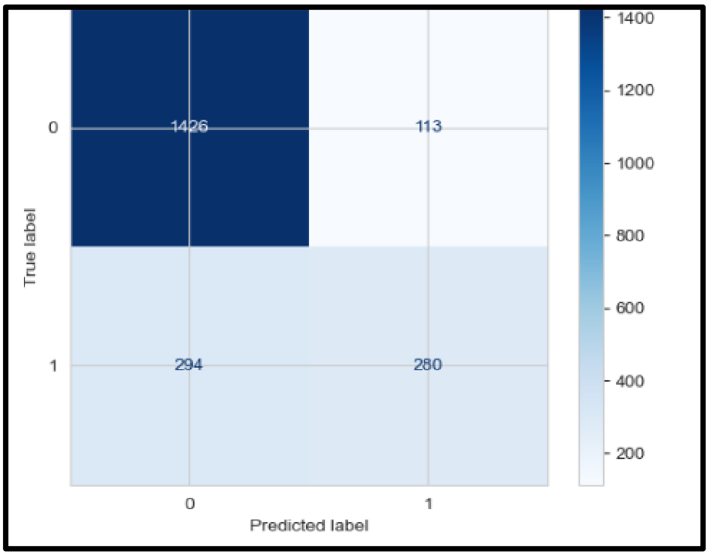

Fig 16: Confusion Matrix for SVM

Received: August 11 , 2025

International Journal of Applied Mathematics

Volume 38 No. 4s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The confusion matrix of the SVM gives 1,426 TNs, 280 TP, 113 FPs, and 294 FNs. Since SVM offers a high precision of 71.24%, it has a low false positive rate, thus making the attainment of other retention strategies economical.

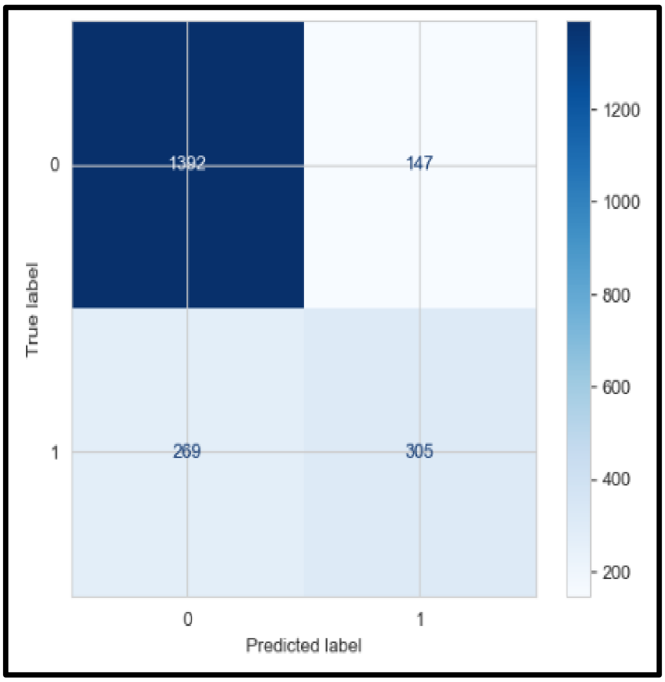

Fig 17: Confusion Matrix for XGBoost

XGBoost classifies 1392 true negative cases and 305 true positive cases with 147 false positive cases and 269 false negative cases. The confusion matrix reveals the kind of classification it can deliver for both churned and loyal customers with a comparatively lower false positive score than the Decision Tree model.

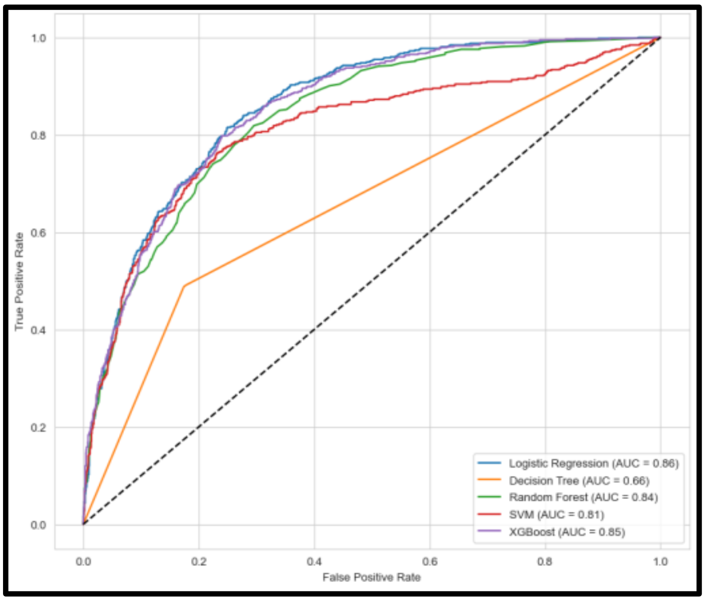

Fig 18: ROC AUC Curves for All Models

Received: August 11 , 2025

International Journal of Applied Mathematics

Volume 38 No. 4s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The ROC curves compare the discriminatory power of all models. The model with the highest AUC is Logistic Regression at 0.858 meaning Logistic Regression can separate customers that churned from loyal customers best. XGBoost also performs equally well with slightly better AUC than the previous one; that is, 0.853 (Patel & Trivedi, 2020) [20]. The chart also shows that the Decision Tree has a low AUC of 0.657.

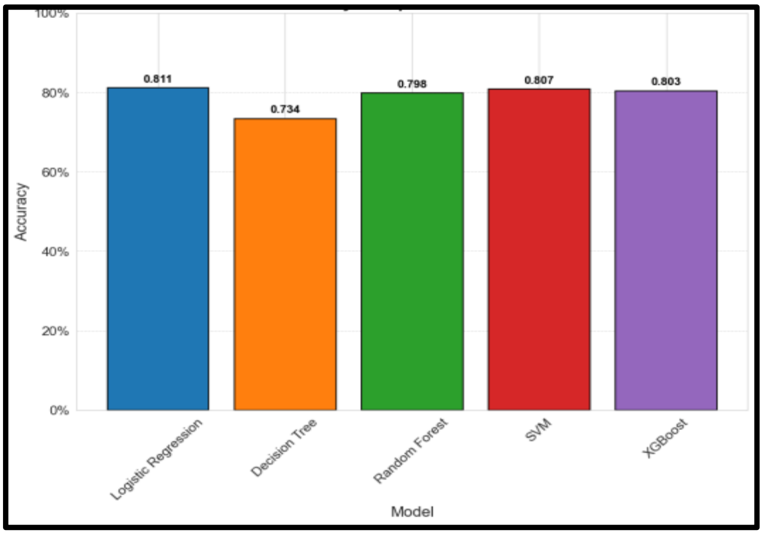

Fig 19: Accuracy Comparison of Models

This bar chart is highly accurate and geared more toward Logistic Regression (81.07 %) and SVM (80.74%). It shows the weakness of the Decision Tree, having a rate of only 73.4%. In the churn prediction, the chart reveals that the features of the dataset require the application of more complex models such as XGBoost and Logistic Regression. Cross-validation allows one to find out the available models that are most accurate in determining the balance between precision and reliability.

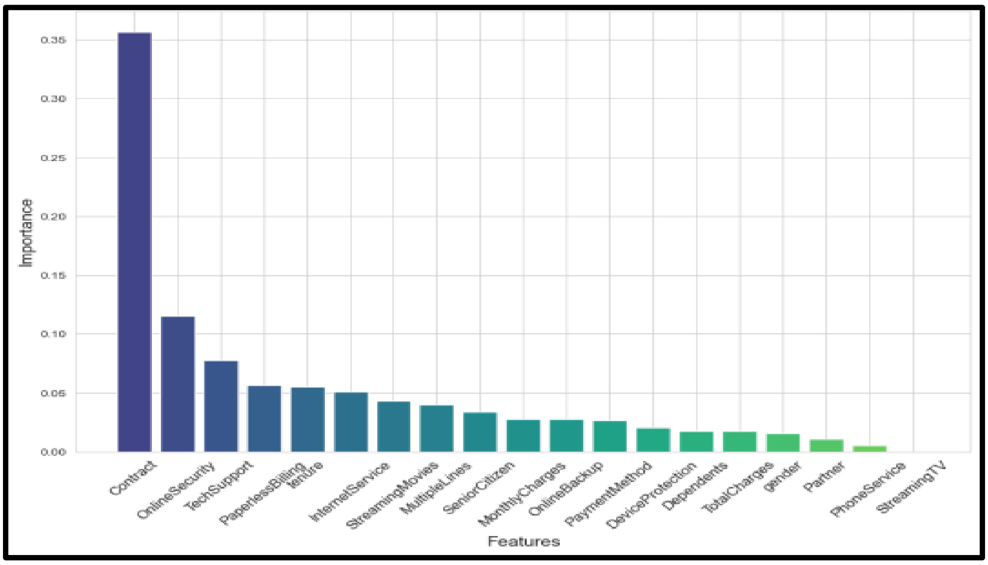

Fig 20: Feature Importance for XGBoost

Received: August 11 , 2025

Volume 38 No. 4s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

This bar chart sorts features by their relevance to the XGBoost algorithm. The most frequent feature observed in the page is the word “Contract” followed by “Online Security” and “Tech Support”. These observations point to contract type as the main determinant of churn behaviour to inform specific interventions. Products such as ‘Phone Service’ and ‘Streaming TV’ contribute little suggesting areas in which less may have to be invested in retention endeavours.

Discussion

The analysis shows that financial factors, such as TotalCharges, MonthlyCharges, and Contract, are the primary drivers of customer churn. The best performance in terms of accuracy and AUC-ROC is achieved by Logistic Regression and XGBoost models. Random Forest and feature importance analysis highlight actionable insights for retention strategies. Clustering identifies specific customer groups that can be targeted for interventions.

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**AUC-ROC**
Logistic Regression	0.810	0.681	0.570	0.620	0.858
Decision Tree	0.734	0.511	0.488	0.499	0.657
Random Forest	0.798	0.675	0.493	0.570	0.837
SVM	0.807	0.712	0.488	0.579	0.808
XGBoost	0.803	0.675	0.531	0.595	0.853

Table 1: Model Performance Summary

V. Conclusion

This study presents machine learning algorithms for predicting customer attrition, with XGBoost boasting an accuracy of 80.31% and logistic regression recording the best accuracy of 81.07%. In the feature importance analysis, churn is most dependent on TotalCharges, MonthlyCharges, and Contract. There is a possibility of clustering analysis to bring out a finer division of customers concerning retention. Risk and compliance and other analyses from models and the data visualization underlie concerns about loss of revenues in addition to offering services for high-risk clients. This can be done by predicting customer behaviour accurately and minimizing churn rates, improving the satisfaction level of customers besides achieving the twin objectives of minimizing costs and improving customer portfolio.

Received: August 11 , 2025

Volume 38 No. 4s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

References

[1] Kasem, M.S., Hamada, M. and Taj-Eddin, I., 2024. Customer profiling, segmentation, and sales prediction using AI in direct marketing. Neural Computing and Applications, 36(9), pp.4995-5005.

[2] Potla, R.T. and Pottla, V.K., 2024. Artificial Intelligence and Machine Learning in CRM: Leveraging Data for Predictive Analytics. Journal of Artificial Intelligence Research, 4(2), pp.31-50.

[3] Singh, D., 2023. A new approach to revolutionise organisations. International Journal of New Media Studies (IJNMS), 10(1), pp.57-63.

[4] Islam, M.R., Shawon, R.E.R. and Sumsuzoha, M., 2023. Personalized marketing strategies in the US retail industry: leveraging machine learning for better customer engagement. International Journal of Machine Learning Research in Cybersecurity and Artificial Intelligence, 14(1), pp.750-774.

[5] Sibanda, W., Ndiweni, E., Boulkeroua, M., Echchabi, A. and Ndlovu, T., 2020. Digital technology disruption on bank business models. International Journal of Business Performance Management, 21(1-2), pp.184-213.

[6] Adeniran, I.A., Abhulimen, A.O., Obiki-Osafiele, A.N., Osundare, O.S., Agu, E.E., Efunniyi, C.P. and OneAdvanced, U.K., 2024. Data-Driven approaches to improve customer experience in banking: Techniques and outcomes. International Journal of Management & Entrepreneurship Research, 6(8), pp.2797-2818.

[7] Agu, E.E., Chiekezie, N.R., Abhulimen, A.O. and Obiki-Osafiele, A.N., 2024. Optimizing supply chains in emerging markets: Addressing key challenges in the financial sector. World Journal of Advanced Science and Technology, 6(01), pp.035-045.

[8] Segun-Falade, O.D., Osundare, O.S., Kedi, W.E., Okeleke, P.A., Ijomah, T.I. and Abdul-Azeez, O.Y., 2024. Developing innovative software solutions for effective energy management systems in industry. Engineering Science & Technology Journal, 5(8), pp.2649-2669.

[9] Rathod, V.H., 2023. Marketing intelligence redefined: Leveraging the power of AI for smarter business growth. Towards Excellence, 15(2), pp.378-386.

[10] Challoumis, C., 2024. UNDERSTANDING THE CYCLE OF MONEY-HOW AI IS TRANSFORMING ENTERPRISES. In XIV International Scientific Conference (pp. 296-324).

[11] Dabo, A.A.A. and Hosseinian-Far, A., 2023. An Integrated Methodology for Enhancing Reverse Logistics Flows and Networks in Industry 5.0. Logistics, 7(4), p.97.

Received: August 11 , 2025

Volume 38 No. 4s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[12] Bhat, J.R., AlQahtani, S.A. and Nekovee, M., 2023. FinTech enablers, use cases, and role of future internet of things. Journal of King Saud University-Computer and Information Sciences, 35(1), pp.87-101.

[13] Challoumis, C., 2024. HOW AI INSIGHTS ARE REVOLUTIONIZING FINANCIAL STRATEGIES FOR ENTERPRISES. In XIV International Scientific Conference (pp. 108-140).

[14] Karim, R.A., Sobhani, F.A., Rabiul, M.K., Lepee, N.J., Kabir, M.R. and Chowdhury, M.A.M., 2022. Linking fintech payment services and customer loyalty intention in the hospitality industry: The mediating role of customer experience and attitude. Sustainability, 14(24), p.16481.

[15] Reddy, V.M. and Nalla, L.N., 2024. Personalization in E-Commerce Marketing: Leveraging Big Data for Tailored Consumer Engagement. Revista de Inteligencia Artificial en Medicina, 15, pp.691-725.

[16] Chang, V., Hahm, N., Xu, Q.A., Vijayakumar, P. and Liu, L., 2024. Towards data and analytics driven B2B-banking for green finance: A cross-selling use case study. Technological Forecasting and Social Change, 206, p.123542.

[17] Yaqoob, I., Salah, K., Jayaraman, R. and Omar, M., 2023. Metaverse applications in smart cities: Enabling technologies, opportunities, challenges, and future directions. Internet of Things, p.100884.

[18] Challoumis, C., 2024. THE FUTURE OF BUSINESS-INTEGRATING AI INTO THE FINANCIAL CYCLE. In XIV International Scientific Conference (pp. 44-78).

[19] Thowfeek, M.H., Samsudeen, S.N. and Sanjeetha, M.B.F., 2020. Drivers of artificial intelligence in banking service sectors. Solid State Technology, 63(5), pp.6400-6411.

[20] Patel, N. and Trivedi, S., 2020. Leveraging predictive modeling, machine learning personalization, NLP customer support, and AI chatbots to increase customer loyalty. Empirical Quests for Management Essences, 3(3), pp.1-24.

Received: August 11 , 2025