International Academic Institute for Science and Technology

International Academic Journal of Science and Engineering
Vol. 12, No. 3, 2025, pp. 205-220.

ISSN 2454-3896

International Academic Journal of Science and Engineering

www.iaiest.com

Leveraging Artificial Intelligence for Improved Customer Retention in the Banking Industry

Joby Jacob [a], E.J. Thomson Fredrik [b]

[a] *Research Scholar, Karpagam Academy of Higher Education (Deemed to be University), Coimbatore, India.*
E-mail: jobyjacob2275@gmail.com, https://orcid.org/0009-0004-7501-2417
[b] *Research Guide, Karpagam Academy of Higher Education (Deemed to be University), Coimbatore, India.*
E-mail: thomson500@gmail.com, https://orcid.org/0000-0003-1309-1936

Abstract

This paper examines the use of artificial intelligence technologies to promote customer loyalty in the banking industry. An analysis of customer exit patterns shows that three characteristics - age, projected income, and credit rating - are consistently among the most important predictors of attrition. Several different classification algorithms were benchmarked including Logistic Regression, Random Forest, AdaBoost, and Gradient Boosting, and these algorithms were assessed in terms of accuracy, precision, recall, and F1-score. Experimental results show that ensemble methods, especially Random Forest and Gradient Boosting, are more effective than baseline models for predicting client departure and more focused and proactive retention strategies.

Keywords: Customer Retention, Machine Learning, Churn Prediction, Artificial Intelligence (AI), Random Forest.

DOI: 10.71086/IAJSE/V12I3/IAJSE1259
Received: 04 July 2025/Revised: 11 August 2025/Accepted: 08 September 2025/Published: 30 September 2025

Introduction

Background

In an increasingly crowded financial marketplace, banking institutions have one of the most strategically important concerns in keeping an existing customer base. It requires numerous resources to acquire a new client than it does to retain one so client retention is one of the most important determinants of long run stability in revenue and brand equity. It is not that simple to forecast which customers will be highly likely to cease their interaction because the behaviour of clients is dynamic and influenced by an extensive range of economic, individual and institutional factors. Such complexities are addressed by complicated and powerful data-driven frameworks capable of analysing large amounts of behavioural information (Shah, 2023). With the help of AI-based tools, financial organisations can process large volumes of data, identify patterns and trends that would not be readily visible at first and act proactively in advance of losing their treasured customers (Nwosu et al., 2024). The strategies of ensemble machine learning like the Random Forests and Gradient Boosting and AdaBoost have emerged as highly useful tools in simplifying the process of credible churn prediction as well as presenting practical behavioural insight (Rakesh et al., 2024). With the results of predictive modelling and the segmentation of intelligence services offers, banks are able to create customised service strategies that will enhance the engagement rate, decrease the rate of attrition and enhance their competitive powers. The investigation has integrated the clustering strategy and predictive analysis and promotes the decision-making process on a changing financial environment of fast paces.

Aim and Objectives

Aim

This research aims at developing the organization client loyalty within the monetary services sector through the use of customer division models and machine learning based anticipating models.

Objectives

- To develop machine learning driven churn prediction systems - Random Forest, Gradient Boosting, and AdaBoost - will assist in determining vulnerable customers by the financial institutions based on historical data of customer interaction.
- To apply unsupervised methods of clustering, i.e. K-means, to determine the customers to behavioural cohort such that retention strategies may be created based on the unique requirements and trend of each group.
- To compare and contrast the diagnostics performance of these predictive and clustering models both regarding the churn forecasting performance and the extraction of meaningful customer preference profile.
- To develop the evidence-based retention recommendations that can help banking institutions to design retention strategies that can reduce churn rates and foster customer loyalty.

Literature Review

Artificial Intelligence in Behavioural Analytics and Retention Modelling

Customer behaviour analytics has been transformed by artificial intelligence, and it has enabled organisations to create more sophisticated and progressive systems of retention. Such power of AI to think and comprehend great amounts of data allows businesses to expose the trend of purchasing and engagement practices that are often overlooked by conventional analytical applications (Jacob & Thomson Fredrik, 2025; Sethi & Kapoor, 2024). By synthesising transaction history, product preference and time engagement data, financial institutions can construct complex models of the behaviour of their clients using the application of machine learning techniques (Kolasani, 2023; Nwosu et al., 2024). In industry (Babakhanian et al., 2023), retention operations have been developed based on the future behaviour of the customer prediction on past data analysis. Other models such as decision trees help organisations to imagine the moment when the occurrence of attrition will be predicted before it takes place. These prediction tools can also help the financial institution track intervention strategies to respond to the risk profile of a particular client by setting early warning of the looming churn related behaviours so as to create products or services that react to the particular client risk profile. It is also created to the next level by incorporating natural language processing (NLP) features to sentiment analytics systems to incorporate also the analysis of unstructured communication means written feedback, electronic correspondence and social media activity. The features provide the opportunity to follow the indicators of customer satisfaction in real-time and can identify service failures within a short time. In addition to behavioural

forecasting, then there may also be additional customisation of pricing and marketing reach by institutions with the assistance of AI-enhanced decision support systems.

Application of the natural language processing (NLP) capabilities on sentiment analytics software brings this intelligence a step further to cover analysis of unstructured vehicles of communication - written feedback, electronic correspondence and use of social media. The capabilities provide customer satisfaction measures that are real time based, and real time identification of service failures. These solutions are fully connected with behavioural forecasting, and thus give the institution the ability to adjust the pricing and marketing accessibility through a system of AI-enhanced decision support. Organisations that use AI-equipped dashboards get the power to monitor important customer engagement metrics in real time and make timely and informed decisions. Recommendation systems within AI platforms also make the customer experience more personalised by creating product or service recommendations that match exactly what the customer prefers, which further strengthens customer-institution relationships.

Advanced Algorithmic Approaches to Real-Time Churn Detection

Detecting client disengagement in real time and quantifying the risk that comes with it requires sophisticated algorithm infrastructures that can process dynamic data streams of customers. Both ML and AI algorithms are analysing the history of interactions systematically to bring to light negative behavioural trajectories that may predict future attrition. The ongoing monitoring of transaction patterns and engagement data helps institutions to identify the pre-exit signals and corrective strategies in a timely manner.

Supervised learning methods - such as logistic regression and classifier trained with gradient - use existing labelled data to learn what are emerging churn indicators. These models provide individualised risk scores to customers which are used to direct the allocation of resources for targeted retention efforts. Complementary unsupervised methods separate customers into risk-stratified segments to conduct preventive engagement campaigns specifically to target particular risk groups.

Real-time processing infrastructure enables institutions to monitor customer activity via multiple interaction channels - mobile platforms, digital portals and physical service records - automatically, in real-time (Dhanapal et al., 2023). Deterioration in the engagement frequency, the accumulation of negative service interactions and unresolved support requests play together as diagnostic signals that trigger timelines in the institutional responses.

Risk estimation Survival analysis and reinforcement learning systems of risk estimation give organisations the necessary tools to obtain the estimates of the customer lifetime value and the probabilistic effects of the exit events. These tools assist the institutions to target retention investment to beneficial groups of clients and costs of interventions embraced in low-risk groups are kept at a minimum. Churn analytics fuelled by AI thus has the impact of giving organisations the gastrointuition of prediction that must win their customer worth and competitive success in the long run.

AI-Driven Customer Segmentation for Customised Retention Planning

The profound customer segmentation through AI has transformed into a helpful tool in the development of variousiated retention strategies that in reality make the depth of a bank client base. In comparison to traditional segments that have been developed with regard to demographics, AI-based segmentation is developed using more detailed behavioural data to create dynamic and multidimensional customer segmentation (Devan et al., 2023). The clustering algorithms, including the K-means partitioning, hierarchical grouping algorithms and DBSCAN are making it possible to find clusters of clients with regard to the similarity in their behaviour. Such are analytic instruments whereby they help the institutions to unravel concealed habits of interaction which consider both the expenditure patterns, the rate of utilisation of the product and the time patterns of interaction. The segmentation faithfulness of the deep learning execution advances a phase further, the learning knowledge is obtained based on the anarchic sources of data including social media interactions, product reviews and web browsing history. Such advanced procedures provide some context on the subjective motivational nature of the customer behaviour that can cause the institutions to develop retention strategies that simultaneously align with the client expectations.

Predictive segmentation models complement this feature by projecting forward-looking churn risk scores for each identified segment so that institutions can focus on retaining their resources on the most vulnerable high-value cohorts. Exclusive incentive programmes and loyalty schemes can therefore be targeted to exactly the customers for whom such interventions are likely to be most successful.

The results of AI-powered segmentation also give insights on how to calibrate the pricing models and communication strategies. Personalised pricing tiers to offer premium service packages for wealthy segments and/or discount-oriented

plans for cost-sensitive clients allows banks to simultaneously maximise revenue extraction and client satisfaction on their entire client portfolio.

Methodology

Research Philosophy

This investigation is based on a positivist epistemological approach which prioritises objective data-driven investigation over interpretive or speculative reasoning. A quantitative research paradigm is adopted in order to permit the systematic recognition of patterns in historical client records, the statistical modelling of attrition probability and the empirical comparison of competing algorithmic strategies.

Research Approach and Design

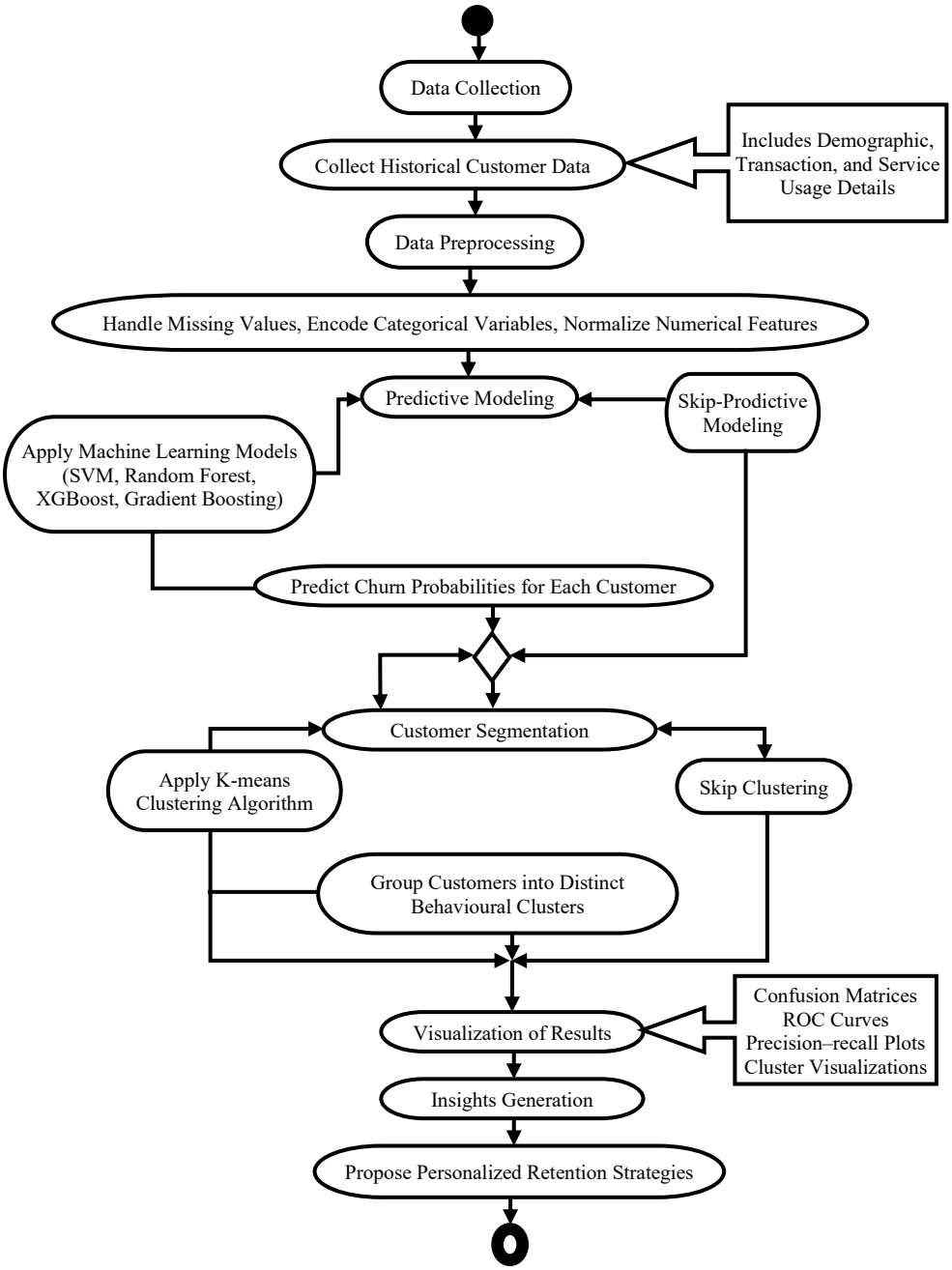


Fig. 1: Design Flowchart

An attempt to conduct a systematic analysis of the customer attrition behaviour was undertaken with a quantitative research design approach involving a data-driven approach (Maseke, 2024). The primary method of grouping customers was K-means clustering and the main predictors of churn were Random Forest, Gradient Boosting and AdaBoost classifiers. The four key steps that are included in the research pipeline were data preparation, model construction, performance evaluation, and results visualisation. Preprocessing operations were done such as normalisation of features, encoding and structuring of client records on the longitudinal direction which were categorical. Model performance was evaluated using the standardised measures accuracy, precision, recall and F1-score - performance was then converted into retention strategy recommendations. General planning is as illustrated in fig. 1.

Data Collection and Analytical Techniques

The empirical background of this study was constituted by secondary datasets, which are acquired in the repositories that are publicly accessible such as Kaggle. The data-sets would include the demographics of customers, their transaction and service records (Madasamy & Aquilanz, 2023). The length of the contract, amount of monthly payment, nature of agreement and attrition were the significant variables that were used in developing the prediction model and client segmentation. A pipeline in the form of analysis was already being implemented that began with a preprocessing phase that involved filling-in of missing values, ordinal translation of nominal variables and normalisation of numerical variables. Three machine learning classifiers which include. Random Forest, Gradient Boosting and AdaBoost were trained on the task of predicting customer attrition and their performance determined in comparison to a series of classification metrics (Adeniran et al., 2024). K-means clustering was then employed to stratify the customers based on their behavioural attributes like time spent in the system and monthly bill. The visualisation of the outputs was aided by the help of the confusion matrices, ROC curves, and visualisation of the clusters that made it possible to perceive the models as a whole and to see the risk segments.

Model Implementation

The training of churn prediction and segmentation models was done under an organized and gradual implementation plan. The processed data was divided into training and validation sets so as to have the sound generalisation of the model. Continuous variables are normalised and categorical variables are coded before starting to fit any model activities (Elias et al., 2024). The customers who were at high risk to leave the company were targeted by constructing three ensemble classifiers namely Random Forest, Gradient Boosting and AdaBoost. The K-means clustering was configured in such a way that the customers were clustered in the cluster in three behavioral characteristics which include account tenure, average monthly outgoings and cumulative payment history (Reddy, 2021). Resultant clusters were interpreted in order to identify clear behavioural archetypes that would favour targeted retention planning. The modelling infrastructure was provided by the scikit-learn library, where the standard implementation of Gradient Boosting was used, and Matplotlib and Seaborn was used to produce all of the visualisation outputs. The combination of churn forecasting and behavioural clustering is helpful in supporting highly personalised service design and measurable protection of revenue.

Ethical Consideration

Customer records were anonymised before were analysed and all data handling complied with relevant regulatory requirements such as GDPR. No personally identifiable information was accessed or retained for longer than it was necessary for analytical purposes. The modelling process was designed to be kept free of demographic bias, as a result churn predictions and segmentation outcomes are only based on objectively observable behavioural signals.

Results and Discussion

Introduction

This section provides a systematic evaluation of three ensemble classifiers namely Random Forest, Gradient Boosting and AdaBoost deployed for the churn prediction of customers. The comparative analysis is based on the classification performance metrics such as accuracy, precision, recall, and F1-score respectively, with visualisations of confusion matrix and segmentation knowledge extracted from the K-means clustering.

Result Analysis

Fig. 2 displays the first records of the analytical dataset that presents the structure of the available customer features. Variables included for the analysis include Credit Score, geographical location, age, tenure, account balance and the churn outcome variable which is a binary variable. This is a structured view which is used to confirm the completeness and organisation of input features before modelling.

	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: First Few Rows of the Dataset

Fig. 3 shows a summary of some of the statistical data that are pertinent to key continuous variables - age, account tenure, account balance, and estimated salary (Al Faisal et al., 2024). The mean, median, and interquartile ranges report and other measures of central tendency and dispersion are significant to put in perspective the interpretation of the distributional properties of churn behaviour-relevant features.

	RowNumber	CustomerId	CreditScore	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
count	10002.000000	1.00200e+04	10002.000000	10001.000000	10002.00...	10002.00..	10002.000000	10001.000...	10001.000000	10002.000000	1002.000000
mean	5001.499600	1.56993e+07	650.555009	38.922311	5.012498	76491.1128r75	1.530194	0.705529	0.514949	100083.33114.5	0.203743
std	2887.472338	7.93177e+04	96.661615	10.487200	2.891973	62393.4741144	0.581639	0.455827	0.499801	57508.117802	0.402814
min	1.000000	1.55657e+07	350.000000	18.000000	0.000000	0.000000	1.000000	0.000000	0.000000	11.580000	0.000000
25%	2501.250000	1.56282e+07	584.000000	32.000000	3.000000	0.000000	1.000000	0.000000	0.000000	50983.750000	0.000000
50%	5001.500000	1.56073e+07	652.000000	37.000000	5.000000	97198.540000	1.000000	1.000000	1.000000	100185.240000	0.000000
75%	7501.750000	1.57532e+07	718.000000	44.000000	7.000000	127647.84000	2.000000	1.000000	1.000000	149383.652500	0.000000
max	10000.000000	1.58156e+07	850.000000	92.000000	10.000000	250898.09000	4.000000	1.000000	1.000000	199992.480000	1.000000

Fig. 3: Descriptive Statistics of Customer Churn

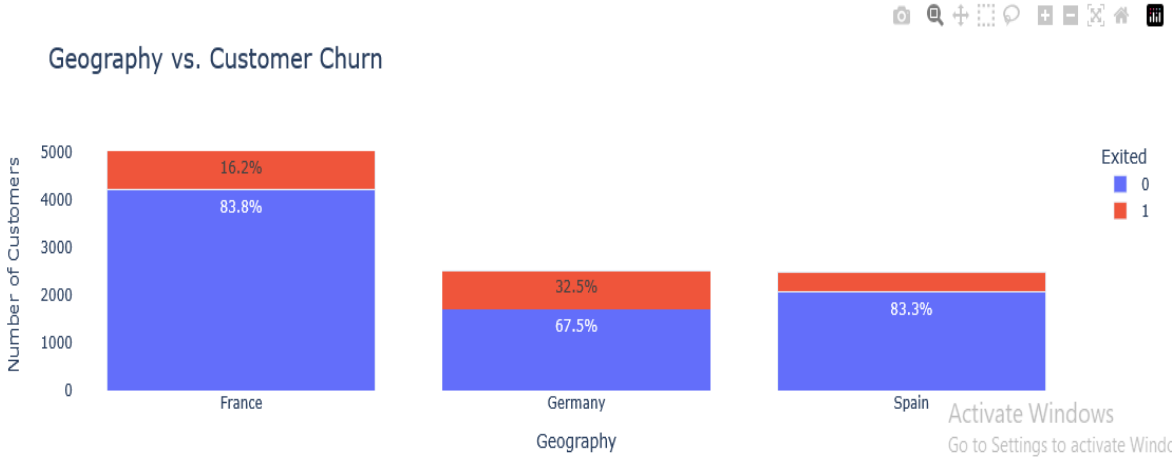


Fig. 4: Distribution of Customer Churn Based on Geographic Location

Fig. 4 provides the regional change of the churn rates of the three geographical markets included in the dataset. France is the most populated with the highest number of customers, and least attrition. On the other hand, the German customers are far much higher in account closure which again means that geographic origin is a good predictor of customer stickiness.

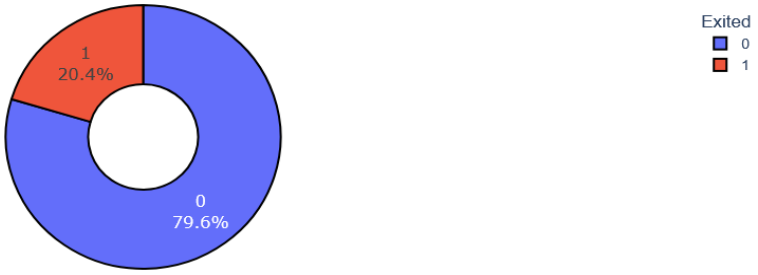


Fig. 5: Churn Distribution

Fig. 5 shows that the overall churn rate of the sampled is supported by the doughnut chart that indicates 20.4%. The visual representation makes it easy to get a grasp on the class imbalance that exists in the data set, which in turn directly can inform choices about how to preprocess the data but also what evaluation metric should be used when comparing models.

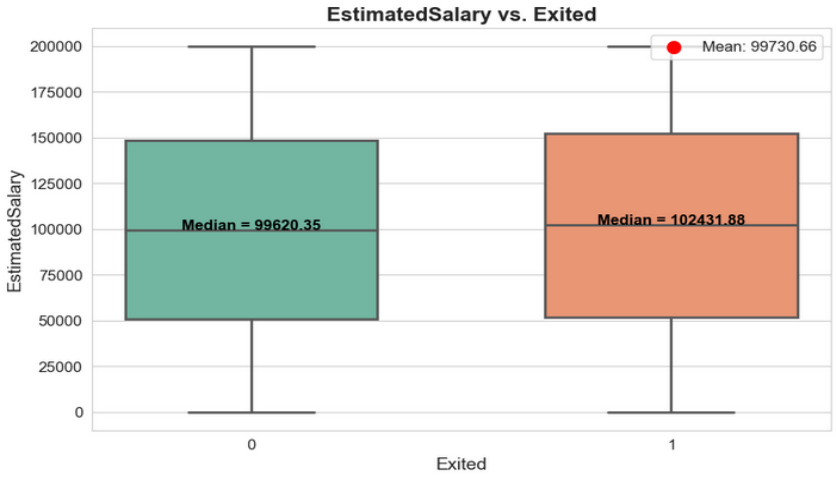


Fig. 6: Estimated Salary Based on Customer Churn

Fig. 6 is a comparison of churned and retained customers income distributions (Adewumi et al., 2024). The data shows that customers that left the bank have reported slightly higher median estimated salaries and it could be that income level had some discriminative signal for churn, although the relationship was not very deterministic.

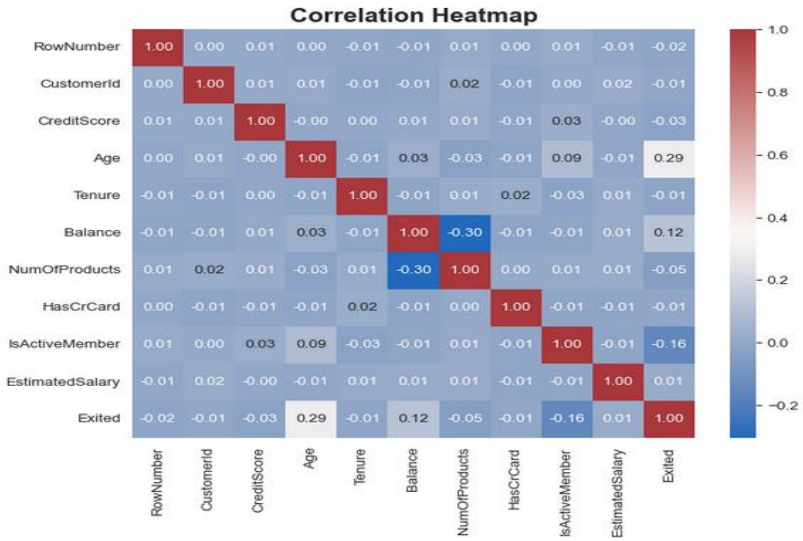


Fig. 7: Correlation Matrix

The pairwise correlation heatmap of all the numerical features is provided in fig. 7. The analysis suggests a poor linear relationship between the account balance and credit score and churn status. A small positive relationship between customer age and probability of attrition is found, which means the longer a customer has been with the company, the older are, the more likely are to leave.

As visualised in fig. 8, the average estimated salary is being broadly consistent across the three national markets. This pattern implies that there is no meaningful variation of income level with geography in this data set and that the effects of geography on churn from a market perspective are likely to be due to factors other than remuneration. It can be seen in the picture (Fig 9) that the distribution of churners and non-churners appears to be equal among credit card holders and non-holders (Adeoye et al., 2024). Customers who lack credit card products are a slightly higher churn predictor, however, again the distinction is not immense, proving that credit card status is not the definitive churn predictor.

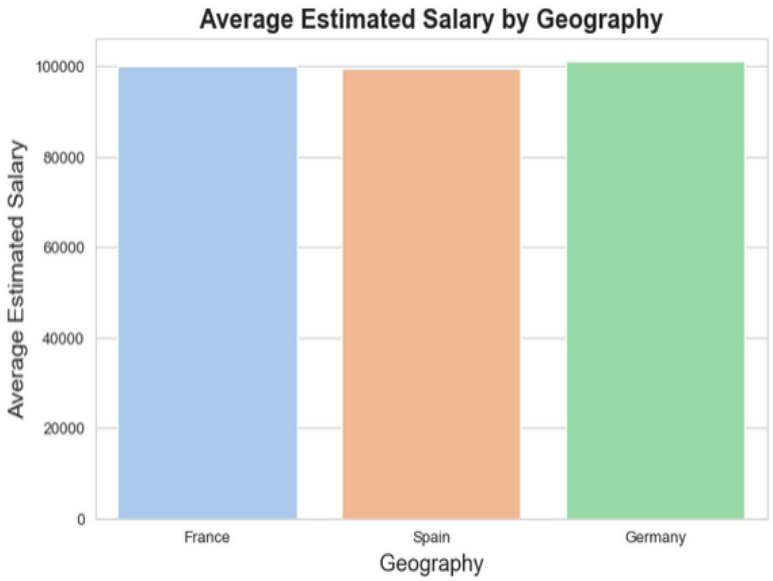


Fig. 8: Average Estimated Salary by Geographic Location

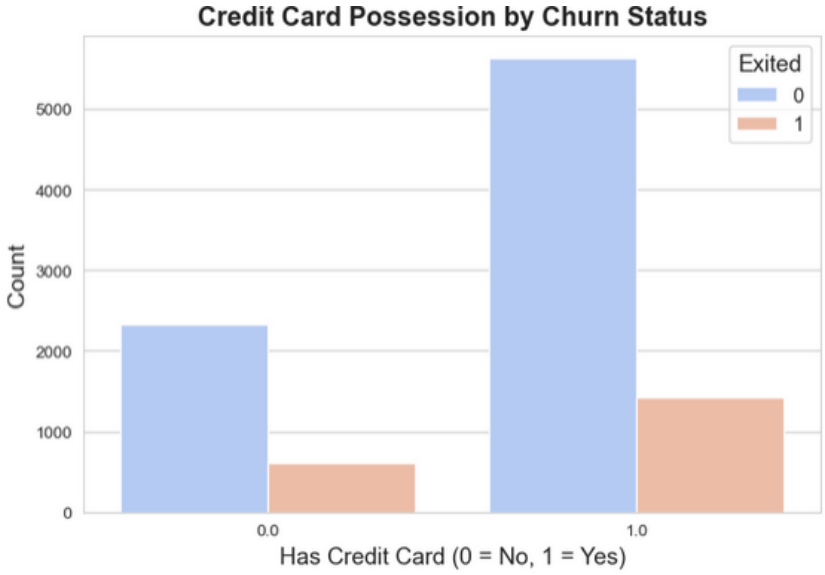


Fig. 9: Credit Card Possession by Churn Status

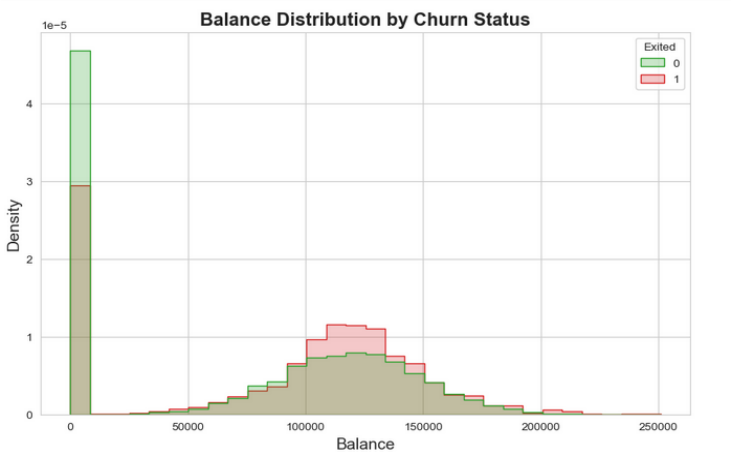


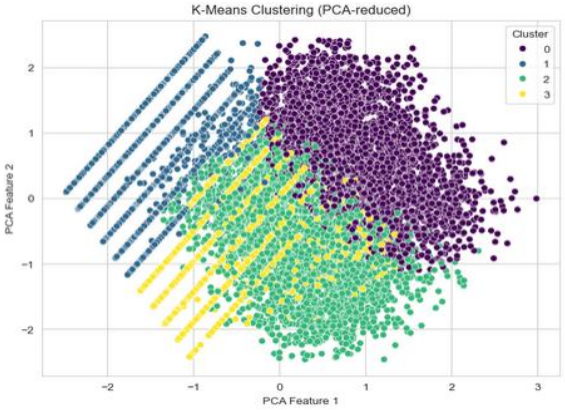
Fig. 10: Balance Distribution by Churn Status

In the cases where the volumes of transactions are similar churned customers have marginally higher average account balances compared to retained clients as the density distributions in fig. 10 indicate. This negative correlation implies that account balance is a moderating and not a leading determinant of churn and could be combined with other attributes such as product engagement or tenure.

	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

Fig. 11 (Umamaheswari & Valarmathi, 2023) depicts the structural integrity of the prepared dataset following the preprocessing step. All recordings are clean, in a regular format and can be taken in by the algorithm. Each of the features has a harmonisation of the variable scales and encodings such that when features are used in models, then the output will not be influenced by anomalies in input data.




```
Cluster Summary:
            Tenure  EstimatedSalary   Balance
Cluster
0         0.074725         0.899084  0.750015
1         0.913078        -0.119171 -1.094575
2        -0.101010        -0.873997  0.762215
3        -0.874573         0.074511 -1.092169
```

Fig. 12: K-Means Clustering Summary

The visualisation of the four clusters of customers by K-means are provided in fig. 12 as the reduced scatter plot (PCA). There is a differentiation in grouping based on account tenure, the anticipated earnings and the balance profile. Every cluster constitutes a consistent behavioural archetype that gives data to the design of targeted retention programmes.

Fig. 13 shows the comparison of the discriminative performance between Gradient Boosting, Random Forest and AdaBoost (Hassan et al., 2023) which are the ROC curves. Gradient Boosting has the best Area Under the Curve or AUC = 0.86; this indicates that it is more likely to discriminate between churning and non-churning customers particularly where there is class imbalance.

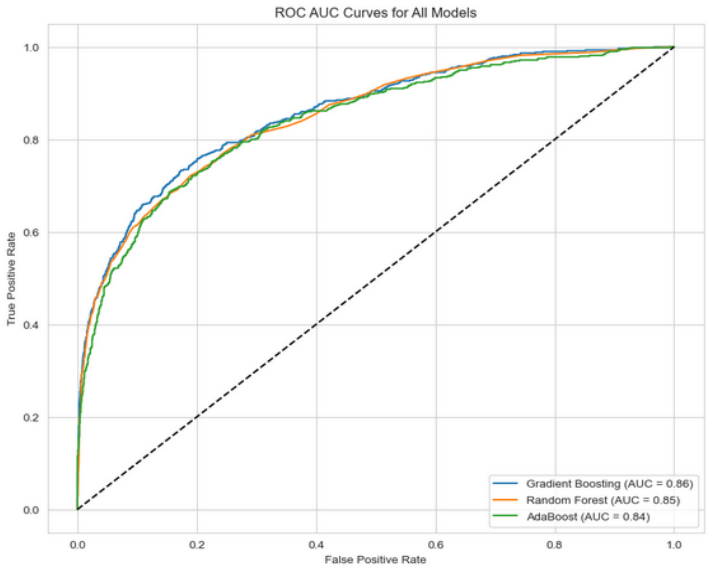
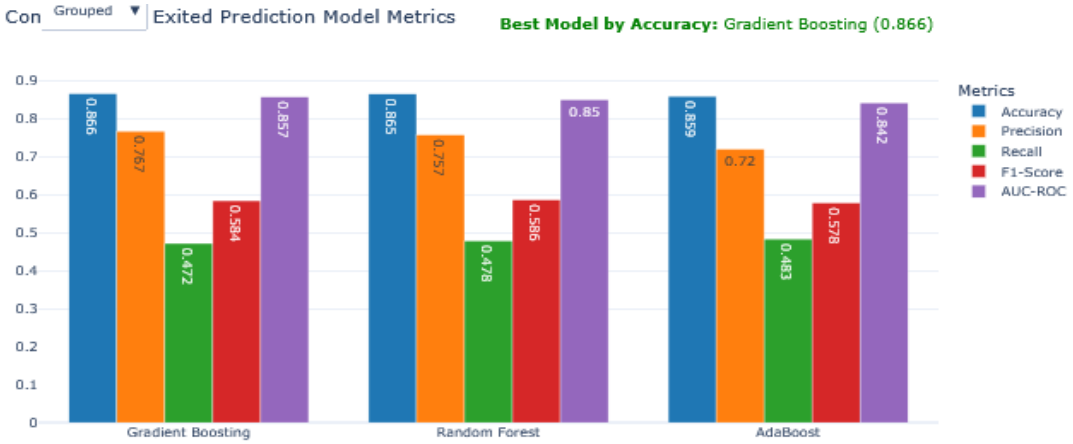


Fig. 13: ROC-Curve for All Models



Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Gradient Boosting	0.866	0.767	0.472	0.584	0.857
Random Forest	0.865	0.757	0.478	0.586	0.85
AdaBoost	0.859	0.72	0.483	0.578	0.842

Fig. 14: Comparison of All Models

Fig. 14 below provides break down of assessment measures of all the three classifiers side-by-side. The composition accuracy and the highest precision and the recall of Random Forest is the highest. These competing strengths indicate that there are substantial trade-offs: the deployment of Gradient Boosting based applications where accuracy is important and Random Forest based applications where recall sensitivity (missing churner is extremely costly to application) is important.

The performance output of the confusion matrix scheme used on Gradient Boosting in fig. 15 is good performance on the non churn class (Tulcanaza-Prieto et al., 2023). The fact that the false negative rate of the churn class is relatively high however highlights the challenge of making predictions of outcomes of minority classes on an imbalanced data. This limitation has direct implications on the deployment of the model in real-time early warning systems.

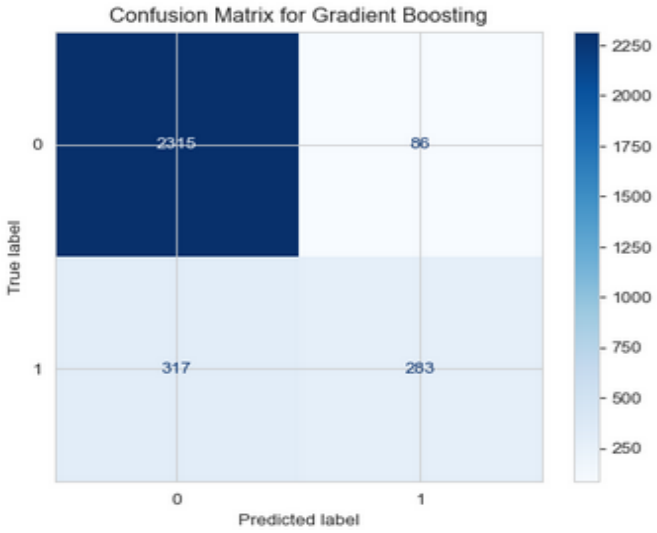


Fig. 15: Gradient Boosting Confusion Matrix

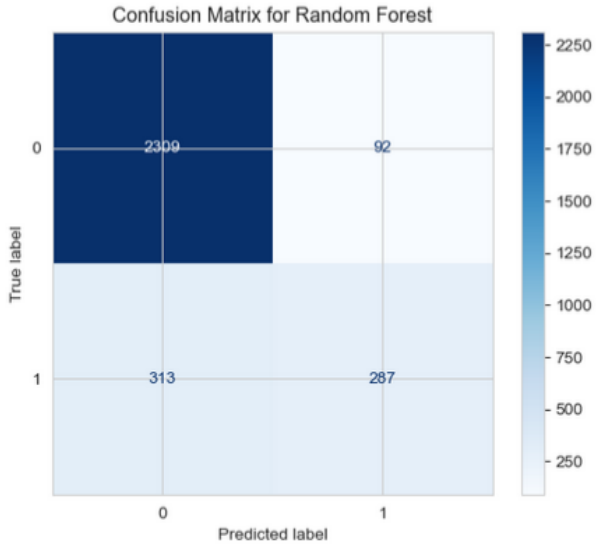


Fig. 16: Random Forest Confusion Matrix

Random Forest confusion matrix in fig. 16 hits a similar pattern, and there is a high true negative identification, but poor recall for churned customer class. This characteristic is indicative of the structural problem of ensemble-based classification on imbalanced datasets and propose cost sensitive learning strategies as a path for further consideration in future versions of the model.

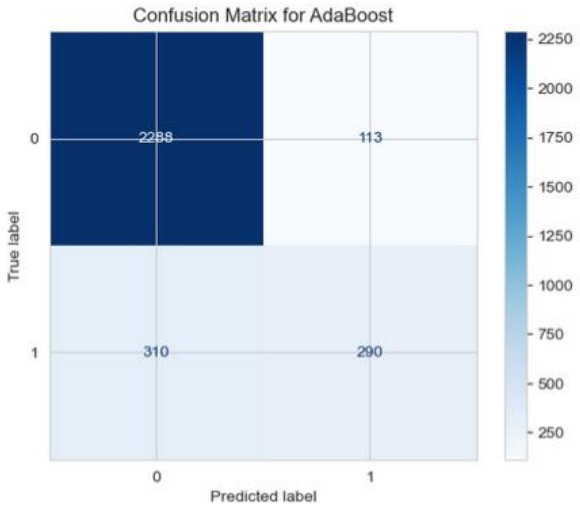


Fig. 17: Adaboost Confusion Matrix

Fig. 17 shows that AdaBoost has a better ability to correctly classify churned customers compared to the other two models. It has a relatively higher recall performance on the positive class, making it a candidate model for the situation where the cost of undetected churn is a particularly high cost, despite some reduction in precision.

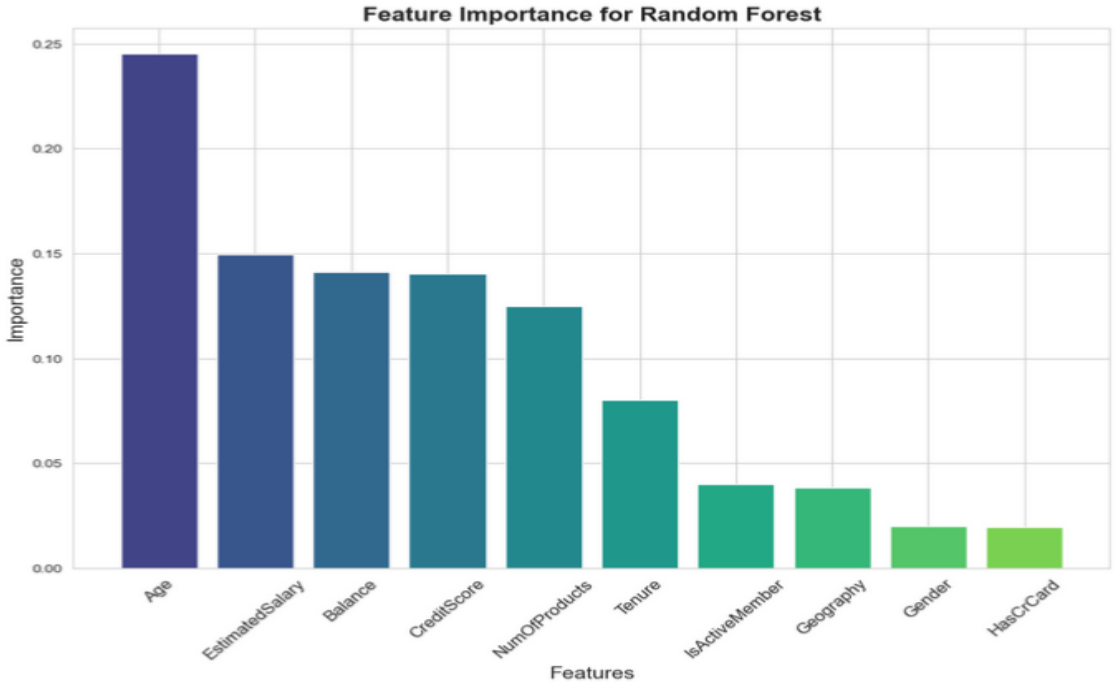


Fig. 18: Feature Importance of Random Forest Model

The feature importance visualisation for Random Forest in fig. 18 shows that customer age and estimated salary are the most important features to predict customer churn in the context of banking. Account balance and credit score follow in relative importance, providing a good ranking of actionable variables that should inform the design of retention interventions.

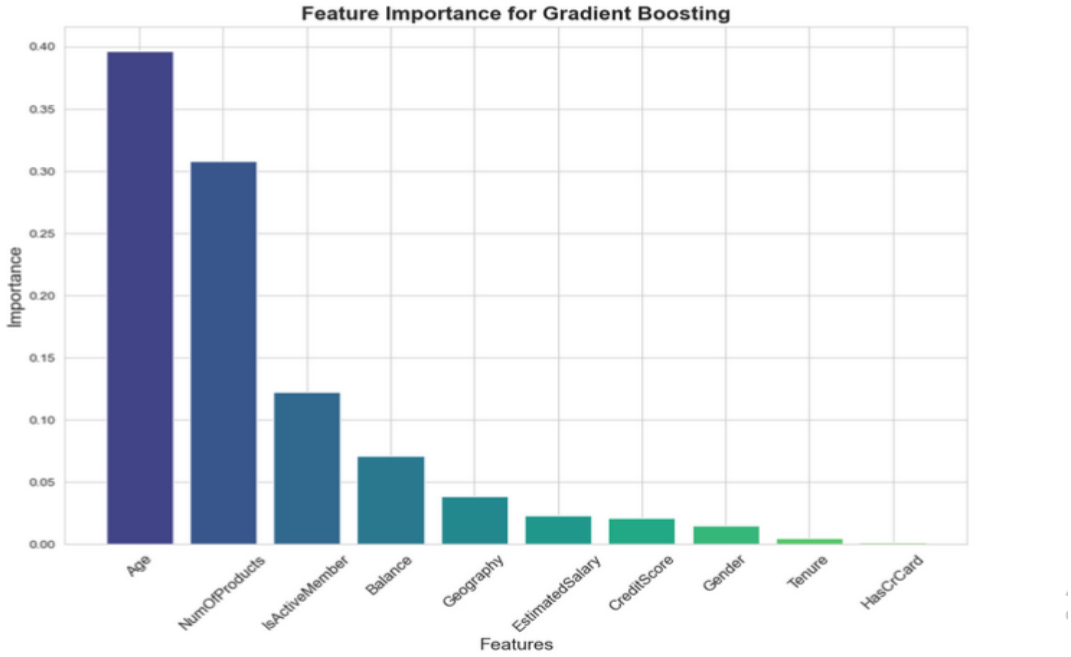


Fig. 19: Feature Importance of Gradient Boosting Model

Again, Gradient Boosting's feature importance profile in fig. 19 suggests that age is the most important predictive variable, followed by product holdings and active membership status. The prominence of these structural engagement variables in the Gradient Boosting ranking suggests that product depth and service activity may have a particular diagnostic value when using this classifier to run targeted retention campaigns.

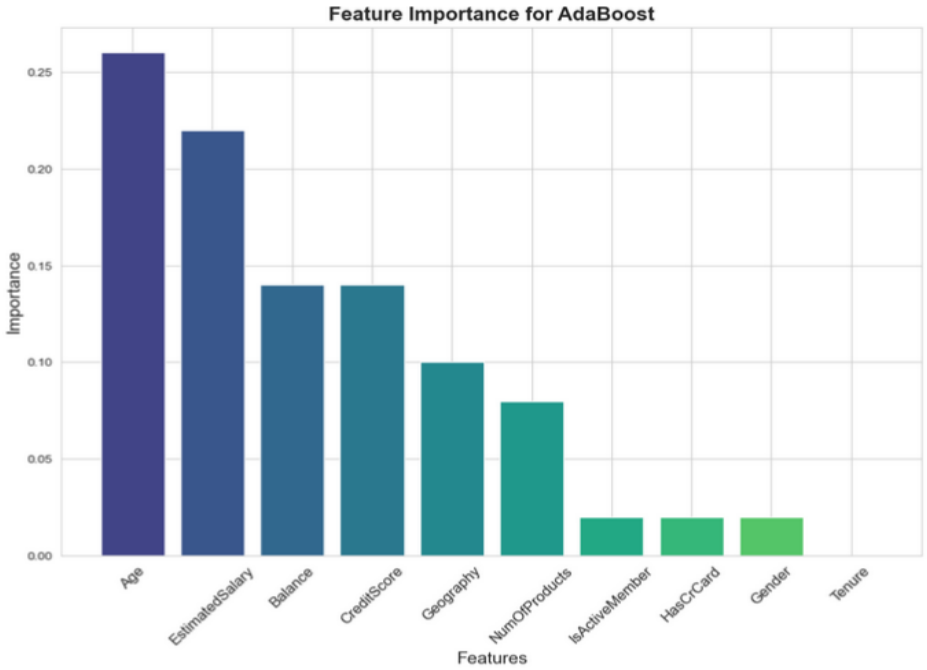


Fig. 20: Feature Importance of The Adaboost Model

AdaBoost feature ranking shown in fig. 20 is in good agreement with those of the other two models in recognizing age and estimated salary as primary predictors. Secondary contributors of balance, credit score and geographic region refine the prediction surface and confirm that churn risk is a Multi-dimensional phenomenon that benefits from a Composite Multi-feature modelling approach.

Key Findings and Discussion

Table 1: Model Evaluation Summary

Model	Precision (Churn)	Recall (Churn)	F1-Score (Churn)	Accuracy
Logistic Regression	0.767	0.472	0.584	0.866
Random Forest	0.757	0.478	0.586	0.865
AdaBoost	0.720	0.483	0.578	0.842
Gradient Boosting	0.767	0.472	0.584	0.866

The parallel analysis of four machine learning classifiers can offer some useful information that can be not called the assessment of the decision to use the theoretical model only, but the issues of deployment also. The most accurate (86.60), and most precise (76.70) one is Gradient Boosting, and this fact proves the highest discriminative power of the algorithm in the case when the false-positive issue is significant (Table 1) (Bhuiyan, 2024). Random Forest and AdaBoost achieve competitive accuracy values and also show significantly higher recall values, reflecting greater ability in terms of churn detection, a feature which would be important, especially in high-cost retention scenarios where undetected attrition is expensive. K-means clustering complemented predictive analysis by identifying three operationally different customer groups based on tenure and monthly billing profiles, each of which was suitable for a differentiated approach to customer retention (Baggyalakshmi et al., 2024). The overwhelming number of monthly-to-monthly contract holders among churned customers strongly suggests that providing incentives for migration of clients to longer term contracts is a high impact retention lever.

ROC and precision-recall curve analysis is consistently in favour of Gradient Boosting in dealing with skewed class distributions, whereas the confusion matrix results of AdaBoost have superior churn identification recall. In aggregate, the results highlight the fact that no single model is dominant in terms of each evaluation dimension, and that deployment context - whether precision or recall are operationally more important - should be taken into consideration when making a final model selection. The sine qua non in building effective, data-informed customer retention architectures is the evidence that ensemble machine learning and segmenting via clusters have played an indispensable role in that effort.

Conclusion

Summary of Findings

This investigation has been able to prove the practical viability of applying machine learning and behavioural clustering techniques to the challenge of predicting and mitigating customer attrition in banking. Gradient Boosting with an accuracy rate of 86.60% proved to be a better performing classifier in the composite predictive accuracy. K-means analysis was able to form three operationally homogenous groups of customers on the basis of which different retention strategies were based (Ogundipe et al., 2024). The strategic importance of the marketing orientation of the promoting trend towards longer term product offers is also reinforced by the fact that the customers of monthly rolling contracts have the best chances of earning the highest department rates of departure. The careful model evaluation, both in the accuracy and precision aspect, and in the recall and AUC aspect are the reasons which ensure that the algorithmic recommendations are backed by empirical information. These outcomes collectively bear witness to the game-changing role that AI-enhanced analytics can play in helping to support the retention objectives of contemporary financial institutions. The set of evidence generated in this study identifies a number of strategic priorities for banking institutions aiming to enhance their customer retention posture. First of all, the introduction of state-of-the-art ensemble classifiers, such as Gradient Boosting and XGBoost, to real-time monitoring systems would allow earlier and more reliable detection of clients at risk, allowing intervention before attrition is irreversible (Modak et al., 2024). Second the application of K-means clustering to segment clients by behavioural signatures should be operationalised as a routine analytical process, with different incentive packages, loyalty rewards and promotional offers depending on the specific characteristics of each cluster (Gurung et al., 2024).

References

Shah, T. (2025). Leadership in digital transformation: Enhancing customer value through AI-driven innovation in financial services marketing. *International Journal of Science and Research Archive, 15*(03), 618-627. https://doi.org/10.30574/ijsra.2025.15.3.1767

Rakesh, N., Mohan, B. A., Kumaran, U., Prakash, G. L., Arul, R., & Thirugnanasambandam, K. (2024). Machine learning-driven strategies for customer retention and financial improvement. *Archives for Technical Sciences, 2(31),* 269–283. https://doi.org/10.70102/afts.2024.1631.269

Babakhanian, M. R., Amin Mousavi, S. A., Soltani, R., & Vakilifar, H. R. (2023). Survey the effect of fintech companies' profitability enhancement on winning customers' loyalty using an artificial intelligence-based optimization algorithm. *International journal of nonlinear analysis and applications*, *14*(1), 2409-2423.

Dhanapal, M., Gopalakrishnan, R., Kabilan, A. S., Nithish, S., & Stephen, A. V. (2023, December). Enrichment of Power Quality Analysis using Machine Learning Algorithm. In *2023 3rd International Conference on Innovative Mechanisms for Industry Applications* (ICIMIA) (pp. 1028-1031). IEEE.

Devan, M., Prakash, S., & Jangoan, S. (2023). Predictive Maintenance in Banking: Leveraging AI for Real-Time Data Analytics. *Journal of Knowledge Learning and Science Technology, 2*(2), 483-490. https://doi.org/10.60087/jklst.vol2.n2.p490

Maseke, B. F. (2024). The Transformative Power of Artificial Intelligence in Banking Client Service. *South Asian Journal of Social Studies and Economics, 21*(3), 93–105.

Madasamy, S., & Aquilanz, L. L. C. (2023). The Evolution of Chatbots: Cloud and AI Synergy in Banking Customer Interactions. *Journal of Emerging Technologies and Innovative Research, 10*(10), 27-37.

Adeniran, I. A., Efunniyi, C. P., Osundare, O. S., & Abhulimen, A. O. (2024). Implementing machine learning techniques for customer retention and churn prediction in telecommunications. *Computer Science and IT Research Journal, 5*(8). https://doi.org/10.51594/csitrj.v5i8.1489

Elias, O., Esebre, S. D., Abijo, I., Timothy, A. M., Babayemi, T. D., Makinde, E. O., Oladepo, O. I., & Fatoki, I. E. (2024). Harnessing artificial intelligence to optimize financial technologies for achieving sustainable development goals. *World Journal of Advanced Research and Reviews, 23*(02), 616–625. https://doi.org/10.30574/wjarr.2024.23.2.2385

Reddy, S. R. B. (2021). Predictive Analytics in Customer Relationship Management: Utilizing Big Data and AI to Drive Personalized Marketing Strategies. *Australian Journal of Machine Learning Research and Applications, 1*(1), 1–12.

Al Faisal, N., Nahar, J., Waliullah, M., & Borna, R. S. (2024). The Role of digital banking features in bank selection an analysis of customer preferences for online and mobile banking. *Frontiers in Applied Engineering and Technology*, *1*(01), 41-58.

Adewumi, A., Ewim, S. E., Sam-Bulya, N. J., & Ajani, O. B. (2024). Advancing business performance through data-driven process automation: A case study of digital transformation in the banking sector. *International Journal of Multidisciplinary Research Updates, 8*(02), 012–022. https://doi.org/10.53430/ijmru.2024.8.2.0049

Adeoye, O. B., Okoye, C. C., Ofodile, O. C., Odeyemi, O., Addy, W. A., & Ajayi-Nifise, A. O. (2024). Integrating Artificial Intelligence in Personalized Insurance Products: A Pathway to Enhanced Customer Engagement. *International Journal of Management and Entrepreneurship Research, 6*(3), 502–511.

Umamaheswari, S., & Valarmathi, A. (2023). Examining artificial intelligence's contributions to the banking sector. *Journal of Survey in Fisheries Sciences, 10*(4S), 2841–2849.

Hassan, M., Aziz, L. A. R., & Andriansyah, Y. (2023). AI's role in contemporary banking: Fraud mitigation, risk governance, and regulatory adherence. *Reviews of Contemporary Business Analytics, 6*(1), 110–132.

Tulcanaza-Prieto, A. B., Cortez-Ordoñez, A., & Lee, C. W. (2023). Customer perception and AI-enabled experience in Ecuadorian banking. *Sustainability, 15*(16), 12441. https://doi.org/10.3390/su151612441

Bhuiyan, M. S. (2024). AI-enhanced personalisation and its influence on customer experience outcomes. *Journal of Computer Science and Technology Studies, 6*(1), 162–169.

Ogundipe, D. O., Odejide, O. A., & Edunjobi, T. E. (2024). Agile methodologies in digital banking: Theoretical underpinnings and implications for customer satisfaction. *Open Access Research Journal of Science and Technology*, *10*(2), 021-030. https://doi.org/10.53022/oarjst.2024.10.2.0045

Modak, C., Ghosh, S. K., Sarkar, M. A. I., Sharif, M. K., Arif, M., Bhuiyan, M., Ahmed, M. P., Pabel, M. A. H., & Devi, S. (2024). Machine Learning Model in Digital Marketing Strategies for Customer Behavior: Harnessing CNNs for Enhanced Customer Satisfaction and Strategic Decision-Making. *Journal of Economics, Finance and Accounting Studies, 6*(3), 178–186. https://doi.org/10.32996/jefas.2024.6.3.14

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. https://doi.org/10.32996/jcsts.2024.6.2.3x

Kolasani, S. (2023). Optimizing Natural Language Processing, Large Language Models (LLMs) for Efficient Customer Service, and hyper-personalization to enable sustainable growth and revenue. *Transactions on Latest Trends in Artificial Intelligence, 4*(4).

Nwosu, N. T., Babatunde, S. O., & Ijomah, T. (2024). Advanced data analytics for customer experience improvement and market penetration in healthcare. *World Journal of Advanced Research and Reviews, 22*(3), 1157–1170.

Baggyalakshmi, N., Harrsini, M. S., & Revathi, R. (2024). Smart Billing Software. *International Academic Journal of Innovative Research*, 11(1), 51–60. https://doi.org/10.9756/IAJIR/V11I1/IAJIR1106

Jacob, J., & Thomson Fredrik, E. J. (2025). Predictive Analytics FOR Customer Churn in Banking: A Machine Learning Approach to Retention. *International Journal of Applied Mathematics, 38*(4, Suppl.), 217. 85-102. https://doi.org/10.12732/ijam.v38i4s.217

Sethi, K., & Kapoor, M. (2024). Data-Driven Marketing in the Age of AI: Reflections from the Periodic Series on Technology and Business Integration. In *Digital Marketing Innovations* (pp. 7-11). Periodic Series in Multidisciplinary Studies.