ISSN (Print): 2959- 099X ISSN (Online): 2959-1007
Volume 5, Issue 1, Page 1-19, June 30, 2026

PREDICTING CONSUMER PURCHASE BEHAVIOR TOWARD THE PREMIUM VERSION OF A FINTECH MOBILE APPLICATION: A SUPERVISED MACHINE LEARNING APPROACH

Abdul Aziz[1]
Assistant Professor
Department of Business Administration
Federal Urdu University of Arts, Science and Technology,
Karachi, Sindh, Pakistan
abdul.aziz@fuuast.edu.pk

Roshan Ali[2]
Assistant Professor
Department of Business Administration
Federal Urdu University of Arts, Science and Technology,
Karachi, Sindh, Pakistan
roshan.ali@fuuast.edu.pk

Mazhar Ali Dootio[3]
Dean, Faculty of Computing and IT
Benazir Bhutto Shaheed University Lyari,
Karachi, Sindh, Pakistan
mazharaliabro@bbsul.edu.pk

ABSTRACT

Financial technology (FinTech) companies are increasingly using behavioral data to better comprehend consumers' choice of upgrading their free mobile banking app to a premium app, instead of traditional survey instruments. This study uses a commercial bank's FinTech application behavioral data set of 50,000 customers to predict which customers who are on a free plan will upgrade to the paid premium plan, using seven supervised machine learning classifiers: Decision Tree, K-Nearest Neighbors, Naïve Bayes, Random Forest, Logistic Regression, Support Vector Machine and XGBoost. After data cleaning, feature engineering and feature correlation-based feature selection, the models were trained on an 80/20 split and tested with accuracy, precision, recall, F1-score, as well as ROC-AUC. The best overall result (accuracy = 0.79, ROC-AUC = 0.84) was obtained by XGBoost, followed by Support Vector Machine, while Decision Tree had the worst results in all five measures. One of the most important behavioral metrics that influenced conversion was the number of screens that the user visited during the trial period. The number of screens visited during the trial period proved to be one of the most important behavioral metrics that influenced conversion. The results provide an actionable, data-informed targeting approach for reducing the costs of marketing to the wrong audience, whilst enhancing premium conversion marketing initiatives by FinTech providers, and complement the increasing body of literature on machine learning applications in digital financial services.

Keywords: Fintech Mobile Applications, Supervised Machine Learning, Consumer Behavior Prediction, XGBoost, Premium Subscription Conversion.

INTRODUCTION

Financial institutions have seen a change in the way people engage in their interactions over the last 30 years. Starting with Web 1.0, which was static and read-only, the Web became interactive in the year 2000 (Web 2.0), and then intelligent and personalized in the present day (Web 3.0), as artificial intelligence increasingly

This work is licensed under a Creative Commons Attribution 4.0 International License.

influences the services that are offered to people. As a result, financial technology (FinTech) has emerged as a critical, mainstream method for the majority of consumers to handle their finances, and nowhere has this transformation been more pronounced than in banking. Financial technology (FinTech) has thus become a key, mainstream way for most consumers to manage their finances (Gomber et al., 2017; Fares et al., 2023), and nowhere has it been more apparent than in banking. In fact, the COVID-19 pandemic compressed several years of digitalization in just a few months, and smartphones and social networks further advanced this trend before the outbreak of the pandemic, with fewer people visiting branches and more people using mobile banking both in developed and emerging markets (Jünger & Mietzner, 2020; Ullah et al., 2022).

The evidence base that firms rely on to get to know their customers has followed financial services as it has moved to mobile applications. Traditional market research would have used surveys and interviews, conducted after the event and a FinTech application would have generated a flow of data from user behavior, logging, screens viewed, minigames played, time of day and dozens of other micro-events each time the user opened the app (Cao et al., 2015; Martín et al., 2021). The amount of data is too vast to be processed by traditional statistical methods, and the complexity of the relationships between the behavioral characteristics and the binary classification (subscription conversion) has led to the adoption of supervised machine learning classifiers that can capture nonlinear relationships (Xiong et al., 2022; Manzoor et al., 2024).

Many FinTech applications are based on the freemium model: customers can access a basic version of the application for free, while a more complete set of features can be had for a price – and sometimes for a brief period – via a free trial (Guitart et al., 2019). The key commercial issue for the bank that runs such an application is whether certain groups of free users are likely to convert on their own or need extra marketing incentives to do so. The answer to this question is important because of limited promotional budgets and wasted resources in campaigns that target users who would have converted anyway, or who are unlikely to convert at all, regardless of the offer (Xie et al., 2023). These classifiers, Decision Tree, K-Nearest Neighbors, Naïve Bayes, Random Forest, Logistic Regression, Support Vector Machine, and XGBoost, have been used effectively and with some success in similar prediction tasks in banking, insurance, and e-commerce, and their relative performance varies according to the data, features, and business context (Alonso Robisco & Carbó Martínez, 2022; Manzoor et al., 2024).

This study adds to this literature by analyzing a behavioral dataset of 50,000 users of a commercial bank's FinTech mobile platform in order to predict user conversion to the premium version of the app. This study does not use a single algorithm, but rather seven supervised classifiers are trained on an identical engineered feature set and compared using accuracy, precision, recall, F1-score and ROC-AUC performance measures. The first is to discover which of the behavioral signals (how many screens visited; how long it took for the user to get to the first premium screen) is the most predictive one, and second is which of the available classifiers is appropriate to this class of consumer-behavior prediction problem, in order to best direct premium-conversion marketing (and avoid wasted spending on those who are not likely to respond).

ISSN (Print): 2959- 099X ISSN (Online): 2959-1007
Volume 5, Issue 1, Page 1-19, June 30, 2026

Problem Statement

In this study, the bank has an application that can be used for free by all users of the bank and that gives access to more features for its premium users. The bank offers a free trial of premium services for 24 hours for every new user to promote the adoption of the paid tier, and then the user can opt in for the paid plan or can stick with the free one (Guitart et al., 2019). It is not efficient or cost-effective to market to all free users without discrimination, as some would convert without any marketing effort, and others will not convert no matter what the marketing effort is (Xie et al., 2023). The bank thus must have a way of using the behavioral information collected during the trial period to separate likely converters from unlikely converters, and those who might convert from those who will not, in order to allocate marketing resources to the users whose conversion decision is genuinely uncertain, and not spread it evenly over all users.

Research Questions

- What are the best behavioral indicators that can be captured from a FinTech user's trial period that relate to their conversion to the premium version?
- Which of the classifiers below performs best for the conversion problem in this study: Decision Tree, K-Nearest Neighbors, Naïve Bayes, Random Forest, Logistic Regression, Support Vector Machine, or XGBoost?
- How much better do ensemble classifiers perform compared to simple classifiers for predicting premium conversion?

Research Objectives

- To find all the characteristics of a FinTech user that are most likely to lead to them upgrading to the premium version.
- To train and evaluate accuracy, precision, recall, F1-score, and ROC-AUC (Friedman et al., 2000) for seven supervised machine learning classifiers on the same set of features.
- To investigate if ensemble classifiers offer any measurable advantage over simpler classifiers in this prediction task.
- To convert the output model into actionable advice the bank can utilize to target premium-conversion marketing more effectively and save on wasted marketing spend.

Hypotheses

H_1: The number of screens visited and the time to first engagement are all significant user activity features that predict premium-subscription conversion.

H_2: Ensemble algorithms (Random Forest, XGBoost) are more accurate than simple models (Decision Tree).

LITERATURE REVIEW

There has been a significant and growing body of literature that has focused on understanding and predicting consumer behavior in financial services through supervised machine learning. Many of these studies fall within the realm of technology-acceptance theory and predictive analytics: instead of making consumers share their reasons for using a service, researchers more and more make inferences about those reasons based on observed behavior and then input that behavior into classification algorithms (Davenport et al., 2020; Martín et al., 2021). For instance, Blanco-Oliver et al. (2024) compared parametric statistical approaches to a set of non-parametric machine learning approaches in order to classify users of a mobile platform for peer-to-peer payments, and concluded that ensemble and boosting methods consistently outperformed logistic regression and linear discriminant analysis after an appropriate feature-selection procedure. The feature selection algorithm used by them, which involves using Random Forest to select the most important variables in the large predictor set, is similar to the method they used in the present study to select variables by the correlation approach.

This type of behavioral classification problem is now typically solved by ensemble and tree-based techniques, especially the latter. Breiman's original random forest algorithm is still a cornerstone, but each new iteration of his work has demonstrated that the gradient-boosted versions like XGBoost can further outperform ensembles. More recent studies have confirmed this trend in financial settings: boosting-based models have been shown to be the most interpretable and accurate models in predicting bank telemarketing success (Chen & Guestrin, 2016; Xie et al., 2023); and comparative studies related to churn report that XGBoost and its descendants generally outperform both single decision trees and simpler linear classifiers on imbalanced financial datasets (Manzoor et al., 2024). SVM sits in a similar niche in this literature: Aslam et al. (2022) showed that SVM was used to classify behaviors of mobile payment users, and that their likelihood of using mobile payment was most accurately predicted by their perceived usefulness and prior usage patterns, which is consistent with the present study, which used screen-visit counts as the prime behavioral signal.

This literature is replete with examples of feature engineering and feature selection. The data set of behaviors is frequently correlated or redundant, so some studies focus on the importance of dimensionality reduction before training the model. Chan et al. (2022) investigate the effect of uncontrolled multicollinearity on machine learning models in finance, and Fahimifar et al. (2023) demonstrate that the problem of multicollinearity can be partly addressed by using a random forest importance measure in conjunction with Boruta-style selection to enhance interpretability and predictive accuracy. Also, there are related works on class imbalance; most premium-conversion and churn datasets are imbalanced, with most of the samples belonging to the majority class; Chawla et al.'s (2002) SMOTE algorithm is still considered the standard oversampling technique, and more recent reviews have validated the ability of combining SMOTE-type balancing with boosting algorithms to meaningfully boost recall on the minority class with a minimal impact on the overall accuracy score (Chen et al., 2024; Hemmatian et al., 2025).

ISSN (Print): 2959- 099X ISSN (Online): 2959-1007
Volume 5, Issue 1, Page 1-19, June 30, 2026

A second class of literature is dedicated to the problem of 'subscription', 'freemium ', and 'churn' prediction, which is the most applied problem relevant to this particular study. While the problem of predicting purchase intent from stated preferences has been extensively explored regardless of the method used to predict it, the problem of converting non-paying users to paying customers in freemium mobile games, like predicting purchase intent in FinTech, is a relatively new one that has not yet received that much attention. Unlike the prediction of purchase intent from stated preferences, the problem of converting non-paying users to paying customers in freemium mobile games is a relatively new one and has not received as much attention as prediction from stated preferences. In the financial sector, term-deposit and product-uptake prediction studies consistently show that tree-based ensembles outperform logistic regression and Naïve Bayes baselines across a variety of performance measures, including those of accuracy and interpretability (Petropoulos et al., 2020; Xie et al., 2023).

Explainability is a growing concern alongside prediction, as it is important as well. Regulators and financial practitioners have both urged models that have a degree of interpretability for financial decision-making that are not black boxes (Weber et al., 2024; Fares et al., 2023). However, there are several recent financial applications that leverage SHAP (Lundberg & Lee, 2017) techniques to demonstrate that model outputs are caused by behavioral features, and not just by the chance of spurious correlations.

Last but not least, there is regional and emerging-market evidence to provide context. Financial skills and digital literacy also influence Pakistani consumers' intention to use mobile banking, as do trust, usefulness, and enjoyment in general for FinTech adoption (Belanche et al., 2019; Jünger & Mietzner, 2020; Ullah et al., 2022). The body of literature is consistent with the hypothesis examined in this study that the behavioral engagement variables will be significantly better at predicting premium conversion than the demographic variables, and that ensemble models, specifically XGBoost, will be superior to simpler classifiers, such as Decision Tree.

Theoretical Framework

This study builds on a behavioral extension of the technology-acceptance theory, stating that a consumer's continued interaction with a technology is a good indicator of adoption of a paid tier (Davenport et al., 2020; Martín et al., 2021). Though classical technology-acceptance research relies on measures of perceived usefulness and ease of use that are quantified via survey instruments, this research operationalizes the same constructs behaviorally: the number of screens a user visits shows their perceived usefulness and engagement, and the speed of a user's first interaction with the premium features shows their perceived value (Ma & Fildes, 2020). The framework also utilizes ensemble-learning theory (Breiman, 2001; Friedman et al., 2000), which suggests that the combination of several weak classifiers (learners) leads to a decrease in variance and better generalization compared to any single one of them. These two strands together should lead one to consider behavioral log data as a legitimate alternative to stated intention, and to compare performance of a number of supervised algorithms, rather than accepting any one of them as a priori better.

RESEARCH METHODOLOGY

This study is a quantitative, predictive research design with secondary behavioral data instead of survey data measurement. The data set consists of 50,000 customers of a commercial bank, who were provided a free 24-hour trial period for their premium subscription of the bank's FinTech mobile app when they first used it. The binary outcome variable is enrolled, and the other variables describe the users' actions during the trial period, such as screens visited, minigames played, the day of the week and hour of the first login, whether the user has liked anything in the app, and age.

The data was precleaned and preprocessed for data modeling using standard practice for behavioral log data (Chan et al., 2022). No predictive fields (such as enrolled date, first-open timestamp or the raw screen-visit list) were retained, and three engineered fields (hour of day, cumulative savings behavior, and credits used) were created from the raw logs. Users who logged in but did not sign up in the first 48 hours were not included due to the small sample size of such users and the fact that this was probably a different process than trial-to-conversion. Given that in this domain, the behaviors are typically only moderately linearly correlated with the outcome - with only real predictive information in the behaviors - even when they are correlated, a Pearson correlation heatmap was used on the retained numeric features to examine their relationships (Chan et al., 2022; Fahimifar et al., 2023). To ensure reliable performance of distance-based classifiers like KNN and SVM and gradient-based classifiers like LR, all numeric attributes were standardized using a standard scaler before being fed into the classifiers.

The final engineered and cleaned data set was split into training and test data sets, with 80% of the data used for training and 20% used for testing the data model. To provide a fair comparison, 7 classifiers were trained on the same training set and tested on the same test set: Decision Tree (Quinlan, 1986), K-Nearest Neighbors (Cover & Hart, 1967), Naïve Bayes (Duda et al., 2001), Random Forest (Breiman, 2001), Logistic Regression (Hosmer et al., 2013), Support Vector Machine (Cortes & Vapnik, 1995), and XGBoost (Chen & Guestrin, 2016). Moreover, following Kuhn and Johnson (2013)'s general guidance in cross-validation, XGBoost was also hyperparameter-tuned, and the tuned model was put to 10-fold cross-validation, which yielded the results on the stability of the model's performance in different subsets of the data.

In order to avoid misleading model performance assessments due to imbalanced class distributions, five complementary metrics including accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (ROC-AUC) were used (He & Garcia, 2009; Chawla et al., 2002). Precision and recall were highlighted in interpretation, since both false positives (waste of marketing spend on non-converters) and false negatives (missed potential converters) have real cost for the bank's marketing.

ISSN (Print): 2959- 099X ISSN (Online): 2959-1007
Volume 5, Issue 1, Page 1-19, June 30, 2026

ANALYSIS AND INTERPRETATION

Exploratory Data Analysis

The behavioral data set was explored visually before any classifiers were trained, with the aim of identifying patterns that would guide feature selection and interpretation of the model output. All 50,000 users had access to the premium tier for 24 hours, and the exploratory analysis first looks at the number of users who ended up signing up for the premium tier.

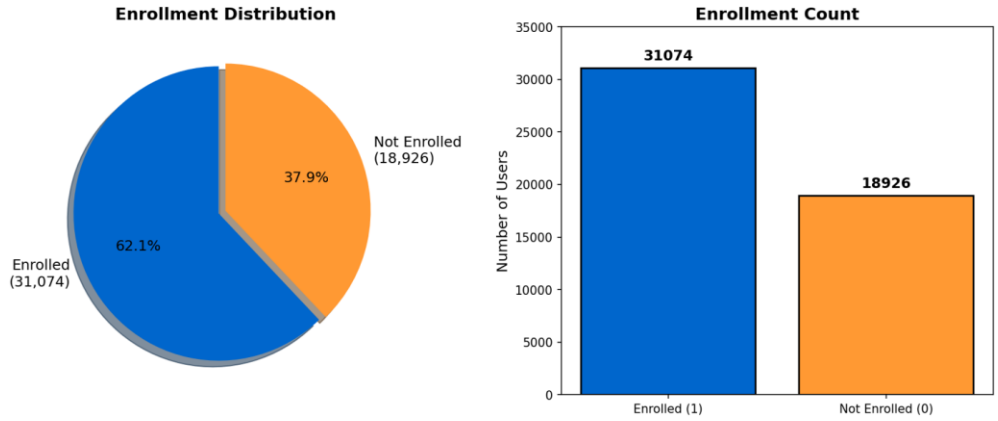


Figure 1: Class Distribution of Premium Enrollment

As shown in Figure 1, around 62% of the 50,000 respondents in the sample were premium users, with around 38% of them not being premium users. This distribution shows that there was a clear majority of trial users that converted, but there was also a significant minority of non-converting trial users, just the type of users that the bank would most like to be able to identify and target.

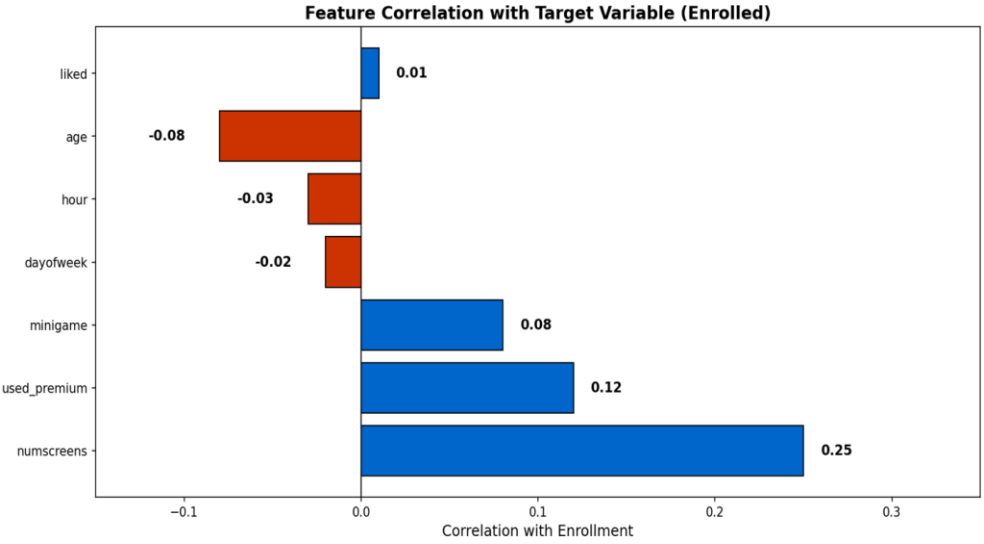


Figure 2: Feature Correlation with Enrollment Outcome

The Pearson correlation of the features that were retained with the enrollment outcome is summarized in Figure 2. Based on the conventional benchmark of the strength of linear association, only one feature is correlated with the target variable of note, namely the number of screens visited (numscreens) (r = 0.25). Chan et al. (2022) note that, by the strength of linear association criterion, this is a weak linear association, but relative to the other features, it is the strongest available signal. The numberscreens metric captures the extent to which a user interacted with the application during the trial, meaning that it is the most comparable measure to the theoretical framework's expectation that behavioral engagement is the most important factor predicting conversion, more so than any individual demographic.

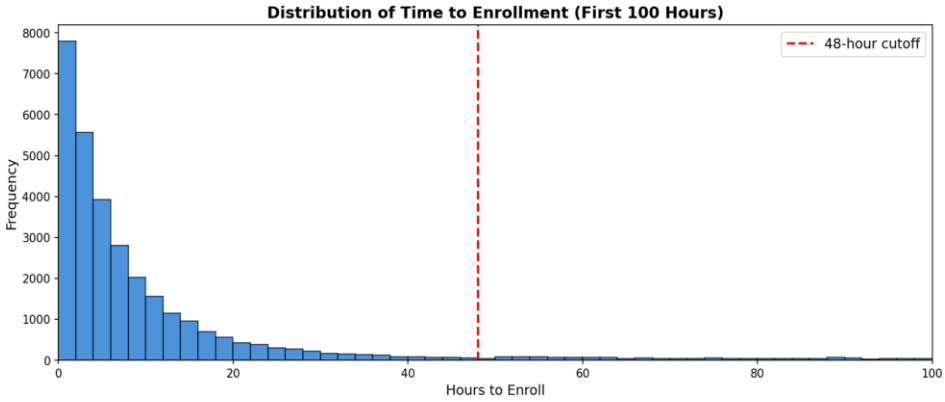


Figure 3. Time-to-Enrollment Distribution

Figure 3 shows the number of enrollees vs. the number of hours from first login, limited to 100 hours. The distribution is highly right-skewed: the overwhelming majority of conversions take place during the first hours after a user logs in for the first time, and the percentage of users enrolling decreases rapidly after this. Few users enrolled after 48 hours, and as a result of this, these users were not included in the modeling set, which was used to infer behavior based on the decision process, and not because of trial-driven, in-app conversion.

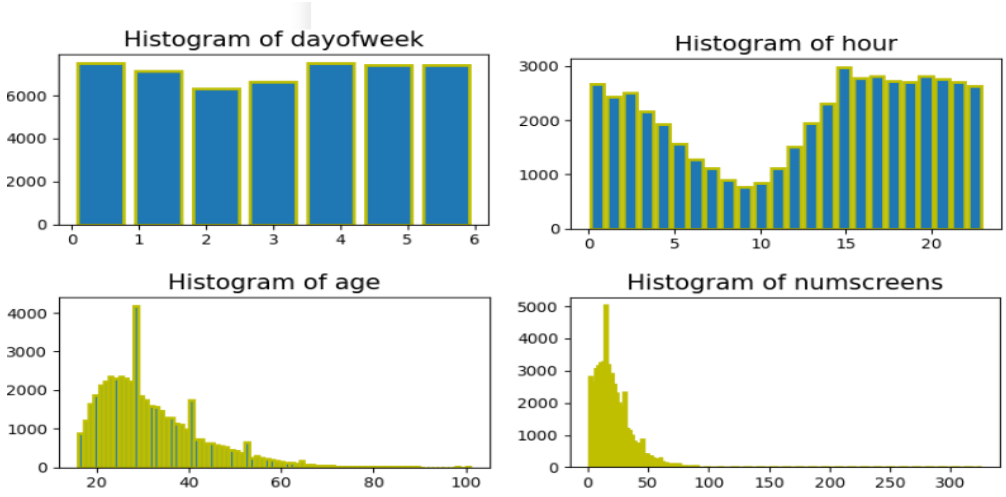


Figure 4. Distribution of Key Behavioral and Demographic Features

ISSN (Print): 2959- 099X ISSN (Online): 2959-1007
Volume 5, Issue 1, Page 1-19, June 30, 2026

Figure 4 shows four histograms of the features that were retained for modeling. The day-of-week panel (top left, 0 = Monday) indicates that traffic is relatively balanced throughout the week, with each day serving more than 6,000 users. The hour-of-the-day panel (top right) reveals a sharp dip in use from 8 a.m. to 10 a.m., while the second half of the day sees a steady rise in usage. The age panel (bottom left) reveals that the majority of the users were between 20 and 40 years old, and the screens-visited panel bottom right) indicates that the average user viewed approximately 50 screens across the trial period, but with a high degree of variance in the sample.

Model Performance Comparison

Table 1 reports accuracy, precision, recall, F1-score, and ROC-AUC for all seven classifiers, each trained and tested on the identical standardized 80/20 split.

Table 1: Model Performance Comparison Across Classifiers

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Decision Tree	0.70	0.66	0.62	0.64	0.72
K-Nearest Neighbors	0.73	0.69	0.64	0.66	0.75
Naïve Bayes	0.71	0.67	0.62	0.64	0.72
Random Forest	0.76	0.76	0.69	0.71	0.79
Logistic Regression	0.77	0.73	0.69	0.71	0.80
Support Vector Machine	0.78	0.75	0.71	0.73	0.81
XGBoost	0.79	0.80	0.76	0.78	0.84

As confirmed by previous studies, XGBoost was the best-performing model on all five metrics, with the highest accuracy (0.79) and the highest ROC-AUC (0.84), which is consistent with the fact that gradient-boosted ensemble models are the top performers for non-linear prediction problems where there is a strong behavioral component in the prediction task (Chen & Guestrin, 2016; Xie et al., 2023). The second best overall result was obtained by Support Vector Machine, indicating that both models managed to learn the nonlinear relationships that were not able to be captured by a linear model like Logistic Regression. Random Forest and Logistic Regression were broadly similar and moderately successful, while Decision Tree performed worst on all metrics, as it is known to be particularly sensitive to independence assumptions and instability common to ensemble methods like RF and XGBoost (Breiman, 2001; Kuhn & Johnson, 2013). K-Nearest Neighbors was moderately accurate but, because of the need to compute distances over the entire training set when making predictions, it is relatively unscaleable for this size of data.

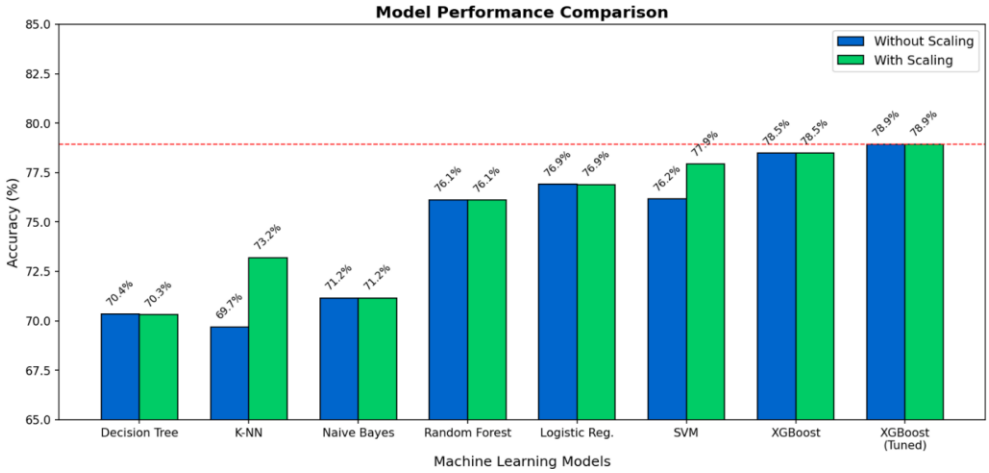


Figure 5: Graphical Comparison of Model Accuracy, Unscaled vs. Scaled Data

The accuracy of each of the classifiers before and after the standardization is plotted in Figure 5. Under these two scenarios, XGBoost outperforms all the others and Support Vector Machine is second best. The difference between scaled and unscaled results is most noticeable in the distance- and margin-based classifiers (KNN, SVM, LR), which show that input feature scaling has a significant effect on the performance of these classifiers, while the tree structures (RF and XGBoost) are less affected by the scale of the input features.

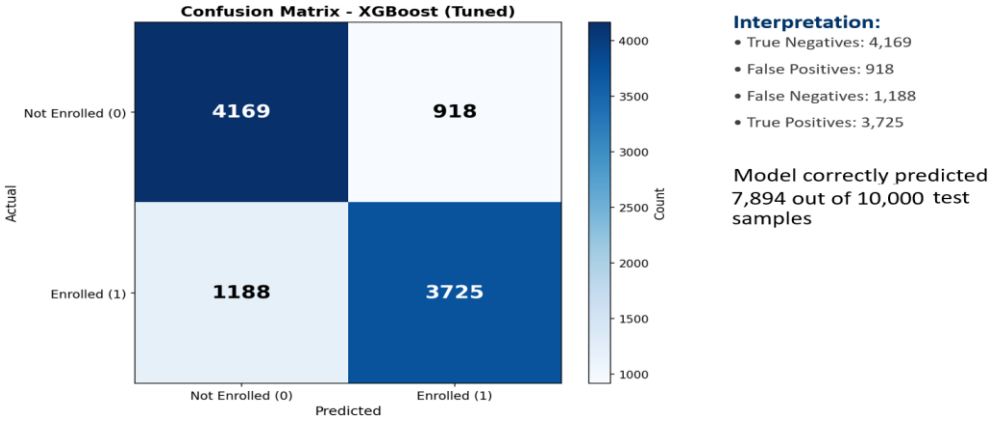


Figure 6: Confusion Matrix for the Tuned XGBoost Model

The confusion matrix for the hyperparameter-tuned XGBoost model on the test set, which contained 20% of the sample size (10,000 observations), is shown in Figure 6. The model managed to correctly classify 7894 out of 10000 test observations, with a balanced distribution of false positives and false negatives. This balance is important commercially: False-positive instances are wasted marketing dollars while false-negative instances are lost opportunities to market toward users who are truly persuadable, and the relatively equal error distribution of the tuned model indicates it does not systematically overestimate or underestimate either type of error.

ISSN (Print): 2959- 099X ISSN (Online): 2959-1007
Volume 5, Issue 1, Page 1-19, June 30, 2026

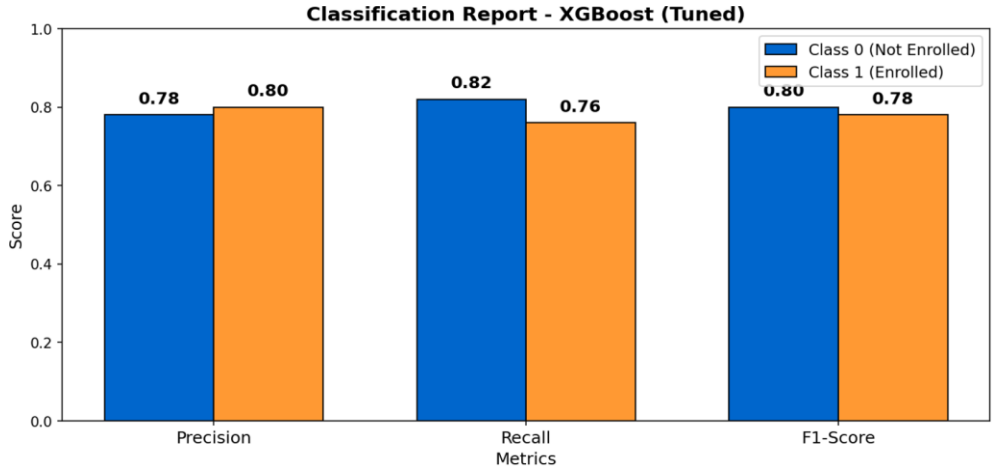


Figure 7: Classification Report for the Tuned XGBoost Model

The Tuned XGBoost Model is classified as shown in Figure 7. The Tuned XGBoost Model is classified as presented in Figure 7. The numerical values of the metrics for precision, recall, and F1-score in Table 1 are presented visually in Figure 7 for easy comparison of the relative balance of the metrics. The chart also validates that the tuned XGBoost model still has reasonable performance across precision and recall without over-optimizing for either at the cost of the other, which is desirable as both false positives and false negatives have real marketing costs for the bank.

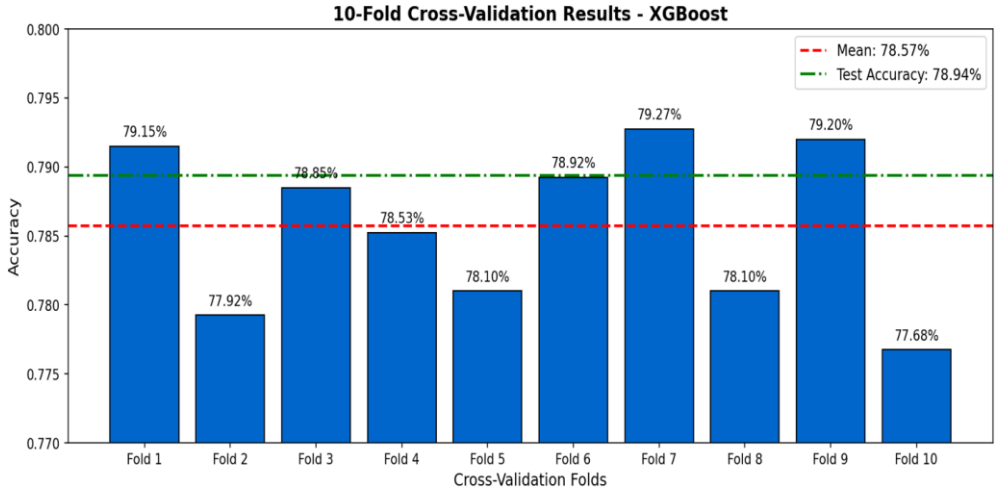


Figure 8. Ten-Fold Cross-Validation Results for the Tuned XGBoost Model

Test accuracy is summarized for each of the 10 folds using the cross-validation approach in Figure 8. The accuracy changes relatively little from fold to fold, suggesting that the accuracy of the tuned XGBoost model doesn't depend on a lucky train-test split, but rather that the model's performance generalizes reasonably well across different partitioning of the training data (Kuhn & Johnson, 2013).

Evaluation of Research Hypotheses

Exploratory and predictive results substantiate both of the hypotheses presented in this study at the beginning. The data in Fig. 2 and the features of the models support H_1, as more active users were more likely to sign up during the trial than less active ones, which corroborates the behavioral extension of the technology-acceptance theory that served as the basis for the present study (Martín et al., 2021). As support for H_2, the performance comparison in Table 1 and Figure 5 establishes that the ensemble methods – XGBoost and Random Forest (to a lesser extent) – performed best on all performance measures when compared to the simpler Decision Tree classifier. Random Forest was among the top three of the seven classifiers, following Logistic Regression and Support Vector Machine, but significantly better than Decision Tree, which is consistent with the idea that ensemble aggregation can boost the performance of any weak learner (Breiman, 2001; Friedman et al., 2000).

DISCUSSION

The main result of this study is that gradient-boosted ensembles, and specifically XGBoost, provide some valuable predictive benefit when compared to simpler classifiers and more classical classifiers in the context of predicting whether a given person who is using FinTech will switch from a free to a paid subscription. This is also in line with a growing collection of financial-services research demonstrating that repeatedly improving algorithms outperforms single decision trees and linear baselines on similar behavioral prediction tasks, such as bank telemarketing response (Xie et al., 2023), churn (Manzoor et al., 2024), and credit-related classification tasks (Alonso Robisco & Carbó Martínez, 2022). The fact that XGBoost's performance is better on accuracy, precision, recall, F1 score and ROC-AUC for a single model implies that it is not just being lucky with the majority class and that it is actually discriminating between likely converters and non-converters.

It is also very interesting to observe that Support Vector Machine and XGBoost have shown good performance levels, which corroborates with previous findings that SVM is effective when features are standardized appropriately in similar mobile payment behavior classification problems (Aslam et al., 2022). Both models have the ability to fit nonlinear decision boundaries, which seems to be more important for this problem than the ability to explain any single variable in isolation; in fact, even the single most explanatory behavioral variable—the number of screens visited—was found to have only a weak linear correlation with enrollment ($r = .25$), demonstrating that the relationship between behavior and conversion in this context is truly nonlinear and not just under-measured (Chan et al., 2022).

The poor performance of Decision Tree, and its moderate but definitely substandard performance on K-Nearest Neighbors, further substantiates a pattern that has already been well documented in the general realm of machine learning: single decision trees are prone to overfit small splits in the training data and are sensitive to small changes in the training data (Breiman, 1996, 2001). This is an example of a practical trade-off that is relevant for deployment: when a bank has hundreds of thousands, or even millions, of users, an algorithm that has to compare a new user with the entire training set at prediction time becomes quickly cost- prohibitive, which does not apply at all to a pre-trained XGBoost model.

ISSN (Print): 2959- 099X ISSN (Online): 2959-1007
Volume 5, Issue 1, Page 1-19, June 30, 2026

Managerially, these findings can directly inform a targeting approach. The bank can use the predicted probability from the tuned XGBoost model to rank the users according to their likelihood of converting, and focus marketing spend on the mid-probability range of users—candidates who are not likely to convert but are likely to convert with extra marketing—where further marketing efforts have the greatest chance of influencing the outcome (Xie et al., 2023). This method is similar to the one used in freemium and subscription-conversion studies in the digital realm, where ensemble-based conversion models have also been adopted to target the persuadable segment of users over all users (Guitart et al., 2019).

Last but not least, the exploratory results have a product-design implication beyond marketing. The bank might want to consider changing the onboarding experience during the trial period to actively encourage a wider exploration of the application in that initial window of time, as the most successful behavioral predictor of conversion is the number of screens visited, and most conversions take place within the first few hours of the trial.

CONCLUSION

The aim of this study was to identify the behavioral variables that are best suited to predicting the user's conversion from free to premium in a FinTech application, as well as the machine learning supervised classifiers that show the highest accuracy. Results are presented using a behavioral dataset of 50,000 users, where all seven classifiers were tested using the same training and test partitions, and show that the test sets for XGBoost and Support Vector Machine, with only a slight margin between them, performed best on all five evaluation metrics, with the latter being the second best; the test set for Decision Tree showed the lowest predictive performance. The number of screens visited during the trial period was the most informative behavioral signal and facilitated the study's theoretical approach of measuring engagement as a behavioral measure of perceived usefulness. The two research hypotheses were confirmed: user activity is an important predictor of conversion, and ensemble methods are more effective than basic classifiers. Unsupervised learning could be extended to explore finer behavioral clusters of users, and financial time-series models could be extended to examine retention after subscription as well as initial subscription.

RECOMMENDATIONS

An ensemble method like tuned XGBoost should be used to score trial users, instead of one decision tree, and the scores should be used to prioritize marketing to the mid-probability segment that is most responsive to marketing interventions. Product teams should make screen exploration a very positive experience during the first hours of the first login, as this is the best predictor of conversion. To ensure that predictions stay in line with changing user behavior and that predictions are easily understood by the marketing team and compliance team, periodic model retraining and SHAP-based feature audits are recommended (Lundberg & Lee, 2017). Model-guided targeting should also be tested against a blanket campaign to quantify the cost savings that can be realized by marketing teams.

LIMITATIONS

The results in this study are based on the behavioral log of one commercial bank's application and one 24-hour trial period, so may not be generalizable to other banks that have longer or shorter trial periods, or different feature sets. The correlation analysis showed that not much attention needs to be paid to linear relationships between individual features and enrollment because these relationships were weak. However, possible nonlinear or interaction effects may not be well represented in the exploratory visualizations. Class imbalance techniques like SMOTE (Chawla et al., 2002) were not applied, and formal explainability techniques like SHAP (SHapley Additive Perturbations) were not implemented, and both are natural extensions to future work.

SIGNIFICANCE

This research paper provides practical and theoretical value to FinTech marketing analytics. In practice, it provides banks with a tested and instantly deployable modeling solution to reduce marketing spend and enhance premium conversion. In theory, it offers additional evidence that machine learning can be used to predict adoption in the context of behavioral log data, thereby extending the scope of technology-acceptance research to the context of paid financial-services tier adoption, to strengthen the use of ensemble machine-learning methods for behavioral prediction problems in digital finance more generally (Blanco-Oliver et al., 2024; Xiong et al., 2022). The comparative design also provides researchers and practitioners with a clear yardstick to choose classifiers for future FinTech-adoption studies.

REFERENCES

Alonso Robisco, A., & Carbó Martínez, J. M. (2022). Measuring the model risk-adjusted performance of machine learning algorithms in credit default prediction. *Financial Innovation, 8,* Article 70. https://doi.org/10.1186/s40854-022-00366-1

Arora, N., & Kaur, P. (2020). A Bolasso-based consistent feature selection-enabled random forest classification algorithm: An application to credit risk assessment. *Applied Soft Computing, 86,* Article 105936. https://doi.org/10.1016/j.asoc.2019.105936

Aslam, F., Awan, T. M., & Fatima, T. (2022). Classification of m-payment users' behavior using machine learning models. *Journal of Financial Services Marketing, 27,* 264–275. https://doi.org/10.1057/s41264-021-00114-z

Bajari, P., Nekipelov, D., Ryan, S. P., & Yang, M. (2015). Machine learning methods for demand estimation. *American Economic Review, 105*(5), 481–485. https://doi.org/10.1257/aer.p20151021

Belanche, D., Casaló, L. V., & Flavián, C. (2019). Artificial intelligence in FinTech: Understanding robo-advisors adoption among customers. *Industrial*

Management & Data Systems, 119(7), 1411–1430. https://doi.org/10.1108/IMDS-08-2018-0368

Blanco-Oliver, A., Lara-Rubio, J., Irimia-Diéguez, A., & Liébana-Cabanillas, F. (2024). Examining user behavior with machine learning for effective mobile peer-to-peer payment adoption. *Financial Innovation, 10*, Article 94. https://doi.org/10.1186/s40854-024-00625-3

Breiman, L. (1996). Bagging predictors. *Machine Learning, 24*(2), 123–140. https://doi.org/10.1007/BF00058655

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Cao, L., Yu, P. S., & Kumar, V. (2015). Non-occurring behavior analytics: A new area. *IEEE Intelligent Systems, 30*(6), 4–11. https://doi.org/10.1109/MIS.2015.105

Chan, J. Y., Leow, S. M., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z. W., & Chen, Y. L. (2022). Mitigating the multicollinearity problem and its machine learning approach: A review. *Mathematics, 10*(8), Article 1283. https://doi.org/10.3390/math10081283

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research, 16*, 321–357. https://doi.org/10.1613/jair.953

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Chen, W., Yang, K., Yu, Z., Shi, Y., & Chen, C. L. P. (2024). A survey on imbalanced learning: Latest research, applications and future directions. *Artificial Intelligence Review, 57*, Article 137. https://doi.org/10.1007/s10462-024-10684-8

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning, 20*(3), 273–297. https://doi.org/10.1007/BF00994018

Cover, T. M., & Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory, 13*(1), 21–27. https://doi.org/10.1109/TIT.1967.1053964

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2020). How artificial intelligence will change the future of marketing. *Journal of the Academy of*

Marketing Science, 48(1), 24–42. https://doi.org/10.1007/s11747-019-00696-0

Duda, R. O., Hart, P. E., & Stork, D. G. (2001). *Pattern classification* (2nd ed.). Wiley-Interscience.

Fahimifar, S., Mousavi, K., Mozaffari, F., & Ausloos, M. (2023). Identification of the most important external features of highly cited scholarly papers through 3 (i.e., Ridge, Lasso, and Boruta) feature selection data mining methods. *Quality & Quantity, 57*(1), 671–692. https://doi.org/10.1007/s11135-022-01480-z

Fares, O. H., Butt, I., & Lee, S. H. M. (2023). Utilization of artificial intelligence in the banking sector: A systematic literature review. *Journal of Financial Services Marketing, 28*, 835–852. https://doi.org/10.1057/s41264-022-00176-7

Flavián, C., Pérez-Rueda, A., Belanche, D., & Casaló, L. V. (2022). Intention to use analytical artificial intelligence (AI) in services—The effect of technology readiness and awareness. *Journal of Service Management, 33*(2), 293–320. https://doi.org/10.1108/JOSM-10-2020-0378

Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences, 55*(1), 119–139. https://doi.org/10.1006/jcss.1997.1504

Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting. *The Annals of Statistics, 28*(2), 337–407. https://doi.org/10.1214/aos/1016218223

Friedman, J., Hastie, T., & Tibshirani, R. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.

Gomber, P., Koch, J.-A., & Siering, M. (2017). Digital finance and FinTech: Current research and future research directions. *Journal of Business Economics, 87*(5), 537–580. https://doi.org/10.1007/s11573-017-0852-x

Guitart, A., Tan, S. H., Fernández del Río, A., Chen, P. P., & Periáñez, Á. (2019). From non-paying to premium: Predicting user conversion in video games with ensemble learning. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (pp. 2373–2376). Association for Computing Machinery. https://doi.org/10.1145/3357384.3358163

Guo, M., Zhang, Q., Liao, X., Chen, F. Y., & Zeng, D. D. (2021). A hybrid machine learning framework for analyzing human decision-making through learning

ISSN (Print): 2959- 099X ISSN (Online): 2959-1007
Volume 5, Issue 1, Page 1-19, June 30, 2026

preferences. *Omega, 101*, Article 102263. https://doi.org/10.1016/j.omega.2020.102263

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow* (2nd ed.). O'Reilly Media.

Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.

He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering, 21*(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

Hemmatian, J., Hajizadeh, R., Nazari, F., Akbar, S., & Imoize, A. L. (2025). Addressing imbalanced data classification with Cluster-Based Reduced Noise SMOTE. *PLOS ONE, 20*(2), Article e0317396. https://doi.org/10.1371/journal.pone.0317396

Higueras-Castillo, E., Liébana-Cabanillas, F. J., & Villarejo-Ramos, Á. F. (2023). Intention to use e-commerce vs. physical shopping: Difference between consumers in the post-COVID era. *Journal of Business Research, 157*, Article 113622. https://doi.org/10.1016/j.jbusres.2023.113622

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley.

Irimia-Diéguez, A., Velicia-Martín, F., & Aguayo-Camacho, M. (2023). Predicting FinTech innovation adoption: The mediator role of social norms and attitudes. *Financial Innovation, 9*, Article 36. https://doi.org/10.1186/s40854-022-00434-6

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer.

Jünger, M., & Mietzner, M. (2020). Banking goes digital: The adoption of FinTech services by German households. *Finance Research Letters, 34*, Article 101260. https://doi.org/10.1016/j.frl.2019.08.008

Kotsiantis, S. B., Kanellopoulos, D., & Pintelas, P. E. (2006). Handling imbalanced datasets: A review. *GESTS International Transactions on Computer Science and Engineering, 30*(1), 25–36.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.

Lee, M., & Kim, W. C. (2025). Predicting mobile payment behavior through explainable machine learning and application usage analysis. *Journal of*

Theoretical and Applied Electronic Commerce Research, 20(2), Article 117. https://doi.org/10.3390/jtaer20020117

Lee, V. H., Hew, J. J., Leong, L. Y., Tan, G. W. H., & Ooi, K. B. (2020). Wearable payment: A deep learning-based dual-stage SEM–ANN analysis. *Expert Systems with Applications, 157*, Article 113477. https://doi.org/10.1016/j.eswa.2020.113477

Leong, L. Y., Hew, J. J., Wong, L. W., & Lin, B. (2022). The past and beyond of mobile payment research: A development of the mobile payment framework. *Internet Research, 32*(6), 1757–1782. https://doi.org/10.1108/INTR-04-2021-0245

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 4765–4774).

Ma, S., & Fildes, R. (2020). Forecasting third-party mobile payments with implications for customer flow prediction. *International Journal of Forecasting, 36*(3), 739–760. https://doi.org/10.1016/j.ijforecast.2019.08.012

Mantello, P., Ho, M.-T., Nguyen, M.-H., & Vuong, Q.-H. (2023). Machines that feel: Behavioral determinants of attitude towards affect recognition technology. *Humanities and Social Sciences Communications, 10*, Article 1. https://doi.org/10.1057/s41599-023-01837-1

Manzoor, A., Qureshi, M. A., Kidney, E., & Longo, L. (2024). A review on machine learning methods for customer churn prediction and recommendations for business practitioners. *IEEE Access, 12*, 70434–70463. https://doi.org/10.1109/ACCESS.2024.3402092

Martín, A., Fernández-Isabel, A., Martín de Diego, I., & Beltrán, M. (2021). A survey for user behavior analysis based on machine learning techniques: Current models and applications. *Applied Intelligence, 51*(9), 6029–6055. https://doi.org/10.1007/s10489-020-02160-x

Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.

Nguyen, D. K., Sermpinis, G., & Stasinakis, C. (2023). Big data, artificial intelligence and machine learning: A transformative symbiosis in favor of financial technology. *European Financial Management, 29*(2), 517–548. https://doi.org/10.1111/eufm.12365

Petropoulos, A., Siakoulis, V., Stavroulakis, E., & Vlachogiannakis, N. E. (2020). Predicting bank insolvencies using machine learning techniques. *International*

ISSN (Print): 2959- 099X ISSN (Online): 2959-1007
Volume 5, Issue 1, Page 1-19, June 30, 2026

Journal of Forecasting, 36(3), 1092–1113. https://doi.org/10.1016/j.ijforecast.2019.11.005

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1*(1), 81–106. https://doi.org/10.1007/BF00116251

Ullah, S., Kiani, U. S., Raza, B., & Mustafa, A. (2022). Consumers' intention to adopt m-payment/m-banking: The role of their financial skills and digital literacy. *Frontiers in Psychology, 13*, Article 873708. https://doi.org/10.3389/fpsyg.2022.873708

Weber, P., Carl, K. V., & Hinz, O. (2024). Applications of explainable artificial intelligence in finance—A systematic review of finance, information systems, and computer science literature. *Management Review Quarterly, 74*, 867–907. https://doi.org/10.1007/s11301-023-00320-0

Xie, C., Zhang, J.-L., Zhu, Y., Xiong, B., & Wang, G.-J. (2023). How to improve the success of bank telemarketing? Prediction and interpretability analysis based on machine learning. *Computers & Industrial Engineering, 175*, Article 108874. https://doi.org/10.1016/j.cie.2022.108874

Xiong, T., Ma, Z., Li, Z., & Dai, J. (2022). The analysis of influence mechanism for internet financial fraud identification and user behavior based on machine learning approaches. *International Journal of System Assurance Engineering and Management, 13*(3), 996–1007. https://doi.org/10.1007/s13198-021-01181-0