IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

Measuring the Profitability and Long-Term Value of BNPL Customers Using Data Analytics

Tanmaykumar Shah
Marketing Analytics, JPMorgan Chase
USA

Abstract—Buy Now, Pay Later (BNPL) applications represent a new type of financial technology that offers convenient payment opportunities to customers and has its own unique issues of fraud prevention. The given paper focuses on the issue of the customer profitability of BNPL services prediction within the Customer Churn dataset provided by Kaggle. There was a lot of exploratory data analysis (EDA) performed in terms of age distribution, churn rates by gender, tenure distribution and the analysis of the frequency of use. The machine learning (ML) models were XGBoost and Random Forest (RF) to predict customer profitability, and their accuracy, precision, recall, F1-score, confusion matrices, and ROC curve were used to measure their performance. The results indicate that XGBoost was much superior, as its accuracy was 99.3% with nearly the perfect classification rate at the second rank. Random Forest took the place accuracy was 97.9%. The effectiveness of ensemble-based models to identify multifaceted occurrences and effective forecasts is presented by evaluating them with benchmark models, such as K-Nearest Neighbours (KNN), Support Vector Machine (SVM), Neural Network (NN), and Isolation Forest. The received findings demonstrate the possibility of XGBoost and Random Forest to predict the availability of profitable BNPL customers accurately, which is valuable in customer retention and specific marketing suggestions.

Keywords: Financial Technology (FinTech), Buy Now Pay Later (BNPL), Credit Card Transactions, Profitability Prediction, Machine Learning, Customer churn Data.

I. Introduction

The speed with which the financial technology (FinTech) sector is developing has transformed the global financial services industry landscape by providing new solutions that address the shortcomings of the existing system. Among these developments is the innovative finance option known as Buy Now, Pay Later (BNPL), which lets customers pay for smaller purchases in instalments without interest [1][2]. It is also gaining some level of popularity as in the UK, the sums of BNPL lending reached over 2.7 billion in 2020, and this is compared to the peak of the payday loans market and the transformation that it brought to consumer financing [3][4]. The expansion of online shopping and the overall merging of these retail marketplaces are closely related to the incorporation of BNPL. Newer buyers, and Gen Z in particular, are resorting to BNPL as an easy way to do the budgeting, to be more flexible, without the negative burden of having to pay out a large loan [5]. For merchants, BNPL reduces payment frictions, increases sales conversions, and expands customer bases, whereas for consumers, it is an affordable and accessible option. Nevertheless, these advantages are also coupled with the issues of repayment risks, possible over-indebtedness, and the absence of steady regulatory control. As a result, the UK, the EU and the US governments have embarked on regulatory review, and measures have already been established by the government of Australia to control the risks of the sector [6][7].

This combination of BNPL presents an interesting question of profitability and customer value in the long run due to the combination of growth opportunities, regulation, and financial issues. The profitability is not just dependent on the volumes of transactions, but also on the behaviour of repayment and the risks of default, and retention of customers [8][9]. In its turn, the long-term value is influenced by the tendencies of repeat use, the cross-selling opportunities, and the implementation of artificial intelligence and ML to enhance personalization and risk mitigation [10]. Machine learning (ML) and other advanced analytic techniques have been emphasized to solve these problems to become critical solutions to BNPL providers [11]. The methods of ML can assist companies to manipulate large volumes of transaction information, assess repayment behaviour, forecast risk, and more precisely assess customer lifetime value (LTV). Besides risk management, ML-based personalization can provide a higher level of customer engagement and retention in addition to a higher level of cross-selling opportunities, which improve short-term profitability and long-term sustainability.

A. *Motivation and Contributions of the Study*

The rapid expansion of BNPL services in the FinTech sector has presented an opportunity and a challenge. BNPL offers easy and interest-free payments to consumers. It also helps e-commerce grow. Nonetheless, it is not well-regulated. This is an unregulated activity that poses threats to financial institutions. These risks include customer defaults, fraud, and misjudging customer profitability. Traditional credit assessment methods are often ineffective. They fail to understand complex modern consumer behaviour. This is especially true in digital payment settings. This study aims to solve this problem. It investigates the use of ML to predict BNPL customer profitability. The focus is on ensemble-based models. The research uses large-scale credit card transaction data. The goal is to give financial institutions better tools. These tools improve risk management and customer targeting. They also enhance decision-making. This helps ensure sustainable growth in the BNPL market. The key contributions offered by the study are discussed below:

- Development of a robust fraud detection framework using the Kaggle Customer Churn dataset.
- Implementation of data pre-processing techniques, including outlier removal, label encoding and min-max scaling for improved model performance.
- A comparison of the fraud categorization capabilities of two potent ML models, Random Forest and XGBoost.
- Comprehensive evaluation of model performance utilizing a variety of criteria, including the ROC curve, F1-score, recall, accuracy, and precision.

B. Significance of the Study

This study holds significant importance as it addresses credit card fraud is becoming more of an issue, endangering both customers and banks. The research helps create more accurate and dependable fraud detection systems that can spot fraudulent activity in real time by utilizing cutting-edge ML models like XGBoost and Random Forest. The evaluation of multiple performance metrics ensures a comprehensive understanding of model effectiveness, highlighting their potential for practical deployment in safeguarding digital financial transactions. Ultimately, the study not only enhances fraud detection accuracy but also strengthens trust and security in modern financial systems.

C. Organization of the Study

This paper is structured as follows: Section II presents a literature review on Measuring the Profitability of BNPL Customers. The methodology is described in Section III, along with the implementation of the ML models. The results are discussed in Section IV, where they are analyzed in detail. Finally, Section V concludes the study and offers guidance for future research.

II. Literature Review

This study is based on a thorough analysis and critical assessment of previous studies on BNPL customer profitability measurement, which helped refine its focus and guide the overall direction of the work. Table I provides a concise summary of current research on BNPL profitability and customer attrition prediction. It highlights the methods used, datasets available, results obtained, and important takeaways from these studies.

Abbas, Usman and Qamar (2022) utilized, including XGboost, KNN, RF, DT, and LR; the effectiveness of classification approaches was examined using Precision, Accuracy, AUC, F1-Score, and Recall. According to this study, the suggested model can assist management in keeping their clientele by predicting customer attrition with a 73% accuracy rate. As the result shows, the XGboost outperformed all other classifiers in every performance assessment metric, making it the most effective classifier for this data set [12].

Kavyarshitha, Sandhya and Deepika (2022) rely on an informative index for a beat model. a collection of Kaggle data. The results are contrasted with what is considered to be the appropriate model. The recollection of the random computation has since emerged. About 87% of it was destroyed. Additionally, the decision tree computation yielded a low definition of 78.3% [13].

Nagaraju and Vijaya (2021) used artificial neural networks (ANN), decision trees with feature selection, and naive Bayes (NB). Following ANN and NB, DT with forward selection yields the greatest results with an accuracy of 91.3111% and a 0.970%. The report provides useful information for further research and CRM tools for improving their automated and vital processes [14].

He et al. (2021) used to assess the impact on F1-score, accuracy, and AUC. The experiment shows that Random Forests and Gradient Boosting Classifier are the most accurate algorithms, with respective accuracy rates of 95.32% and 94.29%. The highest AUC score for both gradient boosting and random forests classifiers was 91%. The Gradient Boosting classifier outperforms the others with the highest F1-score of 97.3% [15].

Johnson, Rodwell and Hendry (2021) Several regulatory gaps are affected by Australia's fee-based BNPL regulation. Consumers may lack the necessary expertise or resources to handle the ever-increasing complexity of many fintech services; as a result, they may need additional protective measures. Studies of regulatory frameworks that take consumer and societal factors into account can facilitate studies of fintech regulation and shift the focus of financial services regulation toward more sustainable and socially beneficial financial services [16].

Fook and McNeill (2020) contributed considering consideration customers' propensity to make impulsive purchases in this setting, a consumer viewpoint on the correlation between BNPL and unsustainable spending behaviours in online shopping is currently lacking in the research. The results, which indicate that BNPL users are more likely than non-BNPL users to participate in online impulse shopping, encourage overconsumption in this environment. Furthermore, it has been demonstrated that sensitivity to sales conversion tools and online impulsive purchasing tendency are directly correlated [17].

Despite the substantial study on BNPL-related consumer behaviour and customer churn prediction, a number of gaps still exist. The majority of research focuses on conventional ML and DL methods for churn prediction, as Table I illustrates, with limited exploration of hybrid or ensemble approaches that can capture complex, nonlinear patterns more effectively. Additionally, many models are tested on relatively small or domain-specific datasets, reducing their generalizability across different industries or geographic regions. While some research examines the regulatory and behavioural aspects of BNPL, there is a lack of integrated studies that combine predictive modelling with consumer behaviour and regulatory insights, limiting the understanding of how financial products influence customer retention and decision-making. Furthermore, few studies comprehensively address feature engineering, managing class imbalance, data preparation, and boosting model robustness are all crucial for making models more applicable in real-world scenarios. Overall, there is a need for research that bridges predictive analytics, consumer behaviour analysis, and regulatory evaluation to provide more actionable insights for businesses and policymakers.

TABLE I. RECENT STUDIES ON MEASURING THE PROFITABILITY OF BNPL CUSTOMERS' CHURN PREDICTION

Author (Year)	Application	Technique / Dataset	Performance	Remarks
Abbas, Usman and Qamar (2022)	Customer churn prediction in retail	Logistic Regression, Random Forest, Decision Tree, KNN, XGBoost; Retail store dataset	Accuracy: 73%	XGBoost outperformed all other classifiers across all evaluation metrics, helping management retain customers
Kavyarshitha, Sandhya and Deepika (2022)	Client retention prediction	SVM, KNN, ANN using Keras & TensorFlow; Beat model dataset from Kaggle	Random Forest Accuracy: 87%; Decision Tree Accuracy: 78.3%	Random Forest achieved the best performance among tested models
Nagaraju and Vijaya (2021)	Customer churn management for insurance	Decision Tree (with forward feature selection), Naive Bayes, ANN; XYZ Insurance dataset (Indonesia)	DT Accuracy: 91.31%, F1-score: 0.970	DT with forward selection provided the highest results; recommended for CRM optimization
He et al. (2021)	Churn prediction in smart intelligent systems	SVM, Random Forest, KNN, Gradient Boosting Classifier	Gradient Boosting Accuracy: 95.32%, F1-score: 97.3%; Random Forest Accuracy: 94.29%, AUC: 91%	Gradient Boosting and Random Forest delivered the highest predictive performance
Johnson, Rodwell and Hendry (2021)	Regulatory assessment of BNPL	Behaviourally informed market failure approach	Highlighted the gaps in consumer understanding and protections regarding complex fintech services	Highlights regulatory failures in fee-based BNPL, emphasizing the need for consumer protection
Fook and McNeill (2020)	BNPL impact on online shopping behaviour	Survey analysis of impulse buying tendencies	BNPL users showed higher impulse buying tendency	BNPL users exhibited higher online impulse buying, promoting overconsumption

III. RESEARCH METHODOLOGY

The proposed methodology utilizes the Kaggle Customer Churn dataset, which has been pre-processed by removing irrelevant columns, handling outliers, and applying label encoding and min-max normalization. The data is then divided into two sets, one for training and one for testing, with a ratio of 80:20. These models are then utilized for classification using Unsupervised Learning with XGBoost. To determine which model is more effective for detecting fraud, compare its accuracy, precision, recall, F1 Score, and ROC curve. Next, the outcomes are examined. In Figure 1, the suggested methodology's flow is illustrated.

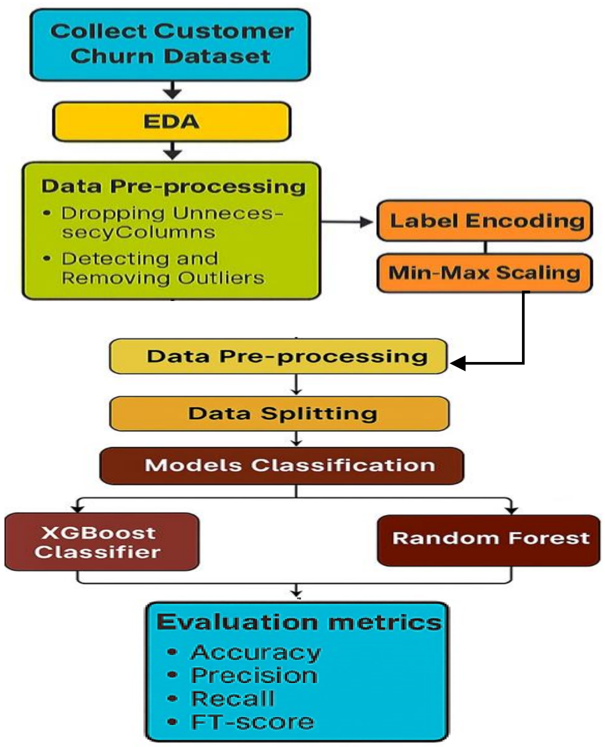

Fig. 1. Proposed Flowchart for Measuring the Profitability of BNPL Customers

The following section details the process depicted in the proposed flowchart for Measuring the Profitability of BNPL Customers.

A. *Data Collection and Visualization*

The Customer Churn dataset [1] From Kaggle, comprising 440,833 records and 12 features, is utilized to analyse customer behaviour and identify churn. In addition to a unique customer ID and the churn goal variable, it contains demographic (age, gender, and region), financial (credit score, salary, and balance), as well as participation indicators (active membership, duration, number of items, and credit card ownership). This dataset supports binary classification for predicting customers at risk. The EDA visualizations are shown below:

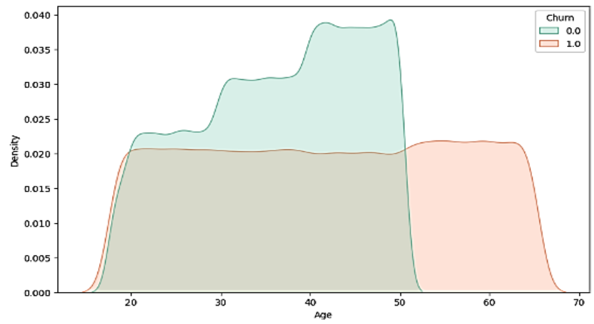

Fig. 2. KDE Plot of Age Distribution by Churn

Figure 2 presents a Kernel Density Estimate (KDE) plot showing the distribution of customer age by churn status. Customers who did not churn (teal) exhibit a bimodal distribution with peaks in the early 20s and mid-40s, tapering off around age 50. Customers who churned (orange) display a more uniform distribution, with the highest density between ages 50 and 65. The overlap between the curves from late teens to around 50 indicates both groups are represented, while the higher density of churned customers over 50 suggests older customers are more likely to leave.

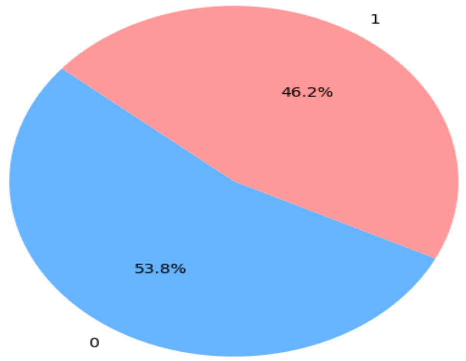

Fig. 3. Pie Chart for Churn Distribution

[1] https://www.kaggle.com/datasets/muhammadshahidazeem/customer-churn-dataset/data

Figure 3 illustrates the churn distribution using a pie chart segmented into two categories: "0" and "1". The blue segment, representing category "0", accounts for 53.8% of the total, indicating the proportion of users who did not churn. In contrast, the red segment labelled "1" comprises 46.2%, reflecting the users who did churn. This visual provides a clear and immediate understanding of the class imbalance in the dataset, which is crucial for designing effective predictive models and evaluating classification performance.

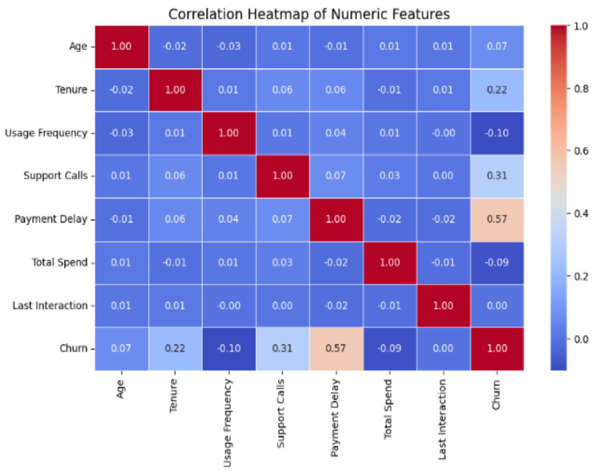

Fig. 4. Correlation Matrix of Numerical Features Of Data

Relationships between important numerical characteristics in the customer dataset are shown in Figure 4. Each cell displays the correlation coefficient, which goes from -1 (very negative) to +1 (strongly positive). Colour gradients show the direction and intensity of these relationships. It is also important to note that the correlation between "Payment Delay" and "Churn" was moderate (0.57), and it can be assumed that delayed payments can be a key predictor of customer attrition. Other attributes, such as Usage Frequency and Spend Total, have lower correlation, whereas the cells in the diagonal indicate that the self-correlation is perfect. The provided matrix assists in identifying strong predictors and the potential multicollinearity, feature selection in churn modelling.

B. *Data Pre-processing*

A major step in data preparation is pre-processing. It involves such processes as dropping redundant columns, outlier detection and control, and label encoding and min-max scaling as follows.

- **Dropping Unnecessary Columns:** A category is assigned a different number in each feature. The categorical features in this study are coded by assigning numeric values to the distinct labels, thereby maintaining the differentiation between categories. The process of dropping such columns makes the dataset easier to work with, removes noise, and improves the efficiency and speed of ML models.
- **Detecting and Removing Outliers:** The transaction amount (amt) can be the sole number that may be applied to outlier identification. In this context, other numerical quantities, such as population, latitude, longitude, etc., have no statistical significance or distribution.

C. *Label Encoding*

The process of turning category data into numerical values is called label encoding, enabling ML models to work with them effectively. A category is assigned a different number in each feature. The label encoding of the categorical features is performed in this study by assigning a numeric code to each unique value, preserving the differences among categories.

D. *Min-Max Scaling*

In this research, the attributes have been brought to a scale of 0 to 1 using the Min-Max normalization. Equation (1) is a mathematical representation of normalization:

$$x' = \frac{x - min\ (x)}{max\ (x) - min\ (x)} \tag{1}$$

where x is the starting point and x' is the adjusted value, $min\ (x)$ represents the data set's minimum, and $max\ (x)$ represents the feature's maximum.

E. *Data Splitting*

Data splitting is the process of dividing a dataset into smaller parts, typically into training and testing sets. The data is divided into two sections: 80% is utilized for training the model, whereas 20% is used to evaluate the model's effectiveness in relation to new data.

F. *Models Classification*

In this section, the proposed ML models, XGBoost and the Random Forest, are applied to discover the profitable BNPL customers.

1) *XGBoost Classifier*

XGBoost is a highly useful and scalable tree boosting ML algorithm that has been applied in different fields to achieve latest findings on specific data problems. Gradient boosting is a structure that improves XGBoost. Rapid and scalable, XGBoost is a version of the Gradient Boosting Machine (GBM) [18]. It is a competitive instrument in the arsenal of artificial intelligence methods as it is user-friendly and its prediction is very accurate. The iterative nature also empowers it with good ability to learn residual and produce a better estimate, particularly with imbalanced data.

$$\mathcal{L}(\Theta) = \sum_{i=1}^{n} \ell(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k) \tag{2}$$

The objective of the XGBoost model can be seen in Equation (2). XGBoost model was trained through hyperparameter tuning and supervised learning, which optimizes parameters such as estimators, max_depth and learning rate by cross-validation. To increase precision and prevent overfitting, especially with unbalanced data, regularization and early stopping were employed.

2) *Random Forest*

A group of decision tree models is used by Random Forest, a supervised ML model, to rank objects and provide predictions [19]. A weak learner is a decision tree because it lacks significant predictive ability. Ensemble learning is the foundation of this approach. To address an issue and enhance the model's accuracy, it utilizes decision tree classifiers extensively. The random forest employs a bagging approach to construct a network of decision trees. To maximize a Random Forest's performance, hyperparameter tuning determines the optimal values for factors like the number of trees and their depth. Using methods like grid search, it tests different combinations to maximize accuracy, resulting in a stronger, more accurate model.

G. *Evaluation Metrics*

Following the classification process, the outcomes can be categorized into four distinct types: TP means that the method correctly found and categorized bad attempts in many samples. TN means that the algorithm correctly finds the proportion of normal samples. Incorrect predictions of FP predictions for negative instances are represented by FN, whereas false negative predictions for positive instances are represented by FP. These outcomes are encapsulated within an error matrix. The formulas for all four parameters are given in Equations (3)-(6) below.

$$Accuracy = \frac{TP+TN}{P+N} \tag{3}$$

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \tag{6}$$

The accuracy of a dataset is the percentage of cases that were accurately labelled. A model's accuracy is defined as the percentage of positive occurrences it can reliably detect from a set of positive examples. The capacity of the model to detect each and every positive case in the dataset is measured by recall, which is also called sensitivity or true positive rate. By combining accuracy and recall into a single number, the F-score achieves a balance between the two. The ROC curve is another tool for measuring how well a binary classification system is doing. True positives and false positives are both displayed on the ROC curve.

IV. RESULTS AND DISCUSSION

These tests were done with a PC based on Windows 10 called a Lenovo Legion Pro Core i9-13900HX with a processing speed of 3.90 GHZ and 32GB RAM. It also uses NVIDIA GeForce RTX 4070, which can process faster. Table II summarizes the results of the performance evaluations of the proposed models using F1-score, recall, accuracy, and precision. The XGB performed best in all measures and was identified as having an accuracy of 99.3%, which proves that it has great potential in identifying profitable customers correctly. The Random Forest (RF) model was also very good, and the accuracy of the model was 97.9% meaning that classification is very reliable, although a little bit lower than XGB. Overall, the table demonstrates the effectiveness of ensemble-based methods, particularly XGB, in accurately assessing the profitability of BNPL customers.

TABLE II. RESULTS OF THE PROPOSED MODELS FOR PROFITABILITY OF BNPL CUSTOMERS

Metric	XGBoost (XGB)	Random Forest (RF)
Accuracy	99.3	97.9
Precision	98.7	96.0
Recall	98.0	96.7
F1-Score	98.3	96.2

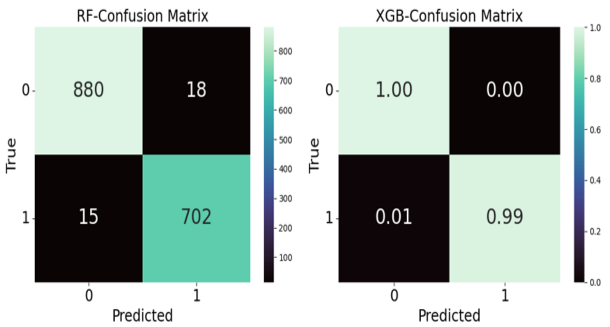

Fig. 5. Confusion Matrix of the Propose RF and XGB Model

The confusion matrices of the suggested RF and XGB models are presented in Figure 5, indicating their ability to classify the data. The RF matrix showed 880 true negatives, 702 true positives, 18 false positives, and 15 false negatives based on the raw counts, indicating that the prediction was quite accurate with a tiny margin of error. Comparatively, the XGB matrix operates on normalized values, which results in almost perfect classification with 1.00 true negatives and 0.99 true positives and low levels of misclassification (0.00 false positives, 0.01 false negatives). This comparison highlights the fact that XGB is more accurate and recalls better, particularly where imbalanced data exists.

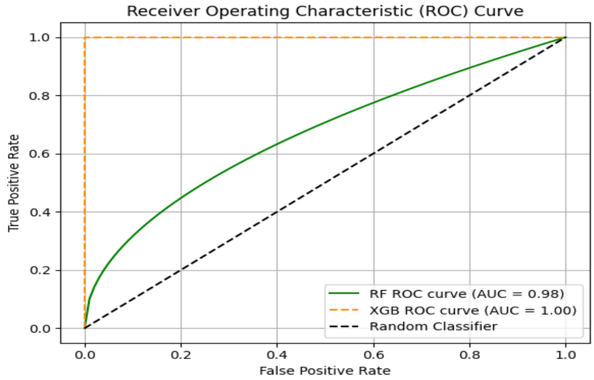

Fig. 6. ROC Curve of XGB and RF Models

The ROC curve of the XGB and the RF model is shown in Figure 6; the performance of these models in terms of categorization is shown graphically. When it comes to Area Under the Curve (AUC), the XGB curve (orange dashed line) is flawless; its value of 1.00 indicates complete discrimination across classes.. The RF curve, depicted as a solid green line, is in close proximity to the top-left boundary, and its predictive accuracy is high, with an AUC of 0.98. The two models perform very well in comparison to the baseline diagonal; that is, they both exhibit the best sensitivity and specificity, with XGB showing slightly better sensitivity and specificity.

A. Comparative Analysis

In order to forecast BNPL clients' profitability, Table III compares and contrasts the performance of several benchmarking models with that of the suggested models. Moderately performing classical models include K-Nearest Neighbours (KNN), Support Vector Machine (SVM), and Neural Network (NN); whereas the Isolation Forest (IF) demonstrated low overall accuracy despite high precision and recall, which implies low reliability. Contrastingly, the suggested ensemble-based models were better than any benchmarks. The better predictive capacity and strength of ensemble algorithms, especially XGBoost, in determining profitable BNPL clients are indicated in this table.

TABLE III. PERFORMANCE COMPARISON FOR THE PROFITABILITY OF BNPL CUSTOMERS BETWEEN BASE AND PROPOSED MODELS

Models	Accuracy	Precision	Recall	F1-Score
KNN [20]	94.0	95.0	95.0	95.0
IF [21]	58.83	94.47	93.9	95.8
SVM [22]	93.49	97.43	89.76	93.36
NN [23]	89.7	92.4	91.2	91.78
XGB	99.3	98.7	98.0	98.3
RF	97.9	96.0	96.7	96.2

XGBoost (XGB) and Random Forest (RF) are both very useful in predicting the profitability of BNPL customers, as they are both ensemble-based. XGB applies gradient boosting to make corrections on errors during the iterative approach, making correct predictions of complex and nonlinear data, whereas RF makes use of bagging and randomness of features to be stable and resistant to noise. Both models work with high-dimensional data, and they capture the intricate interactions among features; they can be trained at scale, interpreted and highly generalised, therefore, they can be used in real-life predictive analytics.

V. CONCLUSION AND FUTURE STUDY

BNPL is the bright new star in the financial technology sector, allowing customers to pay for anything in a series of manageable, interest-free instalments. As illustrated in the paper, under the conditions of massive churn data and ensemble-based models, ML can be used to predict the profitability of BNPL customers with precision. The results indicated that XGBoost had the highest accuracy of 99.3%, followed by random forest at 97.9%, proving that both models are highly predictive and reliable. These models can be used effectively to capture complex relationships in customer demographics, financial, and behavioural characteristics, allowing them to identify profitable and at-risk customers accurately. These benefits enable fintech firms to retain customers more effectively, reduce their financial risk, effectively identify fraud, and personalize financial products, ultimately leading to long-term business growth in the fast-growing and increasingly popular BNPL system. Furthermore, ensemble models possess useful properties such as being strong, interpretable, and scalable, which are relevant to real-life decision systems because they require dynamic consumer behaviour and regulatory variations over time that can be met by dynamic and adaptive, data-driven intelligence to achieve competitiveness and long-term profitability.

The future studies will integrate behavioral analytics and a DL system with explainable artificial intelligence schemes to enhance the interpretability and accuracy of prediction. In addition to that, real-time transactional data and cross-platform behavioral signals could also be used to increase profitability prediction and decision transparency in BNPL ecosystems.

References

[1] M. Barroso and J. Laborda, “Digital transformation and the emergence of the Fintech sector: Systematic literature review,” *Digit. Bus.*, vol. 2, no. 2, 2022, doi: 10.1016/j.digbus.2022.100028.

[2] V. M. L. G. Nerella, “Automated cross-platform database migration and high availability implementation,” *Turkish J. Comput. Math. Educ.*, vol. 9, no. 2, pp. 823–835, 2018.

[3] G. Tepe, U. B. Geyikci, and F. M. Sancak, “FinTech Companies: A Bibliometric Analysis,” *Int. J. Financ. Stud.*, vol. 10, no. 1, Dec. 2021, doi: 10.3390/ijfs10010002.

[4] S. S. S. Neeli, “Ensuring Data Quality: A Critical Aspect of Business Intelligence Success,” *Int. J. Lead. Res. Publ.*, vol. 2, no. 9, p. 7, 2021.

[5] N. Malali, “The Role Of Devsecops In Financial Ai Models: Integrating Security At Every Stage Of Ai/Ml Model Development In Banking And Insurance,” *IJETRM*, vol. 6, no. 11, p. 218, 2022, doi: 10.5281/zenodo.15239176.

[6] H. J. Jung, Y. J. Choi, and K. W. Oh, “Influencing Factors of Chinese Consumers’ Purchase Intention to Sustainable Apparel Products: Exploring Consumer ‘Attitude–Behavioral Intention’ Gap,” *Sustainability*, vol. 12, no. 5, p. 1770, Feb. 2020, doi: 10.3390/su12051770.

[7] H. P. C. Kapadia, “Reducing Cognitive Load in Online Financial Transactions,” *Int. J. Curr. Sci.*, vol. 12, no. 2, pp. 732–797, 2022.

[8] R. C. Cavalcante, R. C. Brasileiro, V. L. F. Souza, J. P. Nobrega, and A. L. I. Oliveira, “Computational Intelligence and Financial Markets: A Survey and Future Directions,” *Expert Syst. Appl.*, 2016, doi: 10.1016/j.eswa.2016.02.006.

[9] M. Ridha and K. Haura Maharani, “Implementation of Artificial Intelligence Chatbot in Optimizing Customer Service in Financial Technology Company PT. FinAccel Finance Indonesia,” 2022. doi: 10.3390/proceedings2022083021.

[10] H. Kali, “Redefining Employee Experience: The Role of Workday HCM in Workplace Digitalization,” *Int. J. Eng. Sci. Math*, vol. 11, no. 09, pp. 1–6, 2022.

[11] S. Pandya, “Predictive Analytics in Smart Grids: Leveraging Machine Learning for Renewable Energy Sources,” *Int. J. Curr. Eng. Technol.*, vol. 11, no. 6, pp. 677–683, 2021.

[12] W. Abbas, M. Usman, and U. Qamar, “Churn Prediction of Customers in a Retail Business using Exploratory Data Analysis,” in *Proceedings - 2022 International Conference on Frontiers of Information Technology, FIT 2022*, 2022. doi: 10.1109/FIT57066.2022.00033.

[13] Y. Kavyarshitha, V. Sandhya, and M. Deepika, “Churn Prediction in Banking using ML with ANN,” in *Proceedings - 2022 6th International Conference on Intelligent Computing and Control Systems, ICICCS 2022*, 2022. doi: 10.1109/ICICCS53718.2022.9788456.

[14] J. Nagaraju and J. Vijaya, “Methodologies used for Customer Churn Detection in Customer Relationship Management,” in *Proceedings of International Conference on Technological Advancements and Innovations, ICTAI 2021*, 2021. doi: 10.1109/ICTAI53825.2021.9673382.

[15] C. He, C. H. Q. Ding, S. Chen, and B. Luo, “Intelligent Machine Learning System for Predicting Customer Churn,” in *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, 2021. doi: 10.1109/ICTAI52525.2021.00085.

[16] D. Johnson, J. Rodwell, and T. Hendry, “Analyzing the Impacts of Financial Services Regulation to Make the Case That Buy-Now-Pay-Later Regulation Is Failing,” *Sustainability*, vol. 13, no. 4, p. 1992, Feb. 2021, doi: 10.3390/su13041992.

[17] L. Ah Fook and L. McNeill, “Click to Buy: The Impact of Retail Credit on Over-Consumption in the Online Environment,” *Sustainability*, vol. 12, no. 18, p. 7322, Sep. 2020, doi: 10.3390/su12187322.

[18] O. Raju, “Credit Card Fraud Detection Using Xgboost,” vol. XIII, no. Iii, pp. 175–187, 2021.

[19] E. Kabir, S. Guikema, and B. Kane, “Statistical modelling of tree failures during storms,” *Reliab. Eng. Syst. Saf.*, 2018, doi: 10.1016/j.ress.2018.04.026.

[20] J. C. Mathew, B. Nithya, C. R. Vishwanatha, P. Shetty, H. Priya, and G. Kavya, “An Analysis on Fraud Detection in Credit Card Transactions using Machine Learning Techniques,” *Proc. 2nd Int. Conf. Artif. Intell. Smart Energy, ICAIS 2022*, pp. 265–272, 2022, doi: 10.1109/ICAIS53314.2022.9742830.

[21] V. N. Dornadula and S. Geetha, “Credit Card Fraud Detection using Machine Learning Algorithms,” *Procedia Comput. Sci.*, vol. 165, pp. 631–641, 2019, doi: 10.1016/j.procs.2020.01.057.

[22] A. RB and S. K. KR, “Credit card fraud detection using artificial neural network,” *Glob. Transitions Proc.*, vol. 2, no. 1, pp. 35–41, Jun. 2021, doi: 10.1016/j.gltp.2021.01.006.

[23] O. F. Seymen, O. Dogan, and A. Hiziroglu, “Customer Churn Prediction Using Deep Learning,” *Adv. Intell. Syst. Comput.*, vol. 1383 AISC, no. April, pp. 520–529, 2021, doi: 10.1007/978-3-030-73689-7_50.