**CHAPTER ONE**

**1.0 Introduction**

The banking industry is one of the most dynamic yet inherently risky sectors of the global economy. A central feature of banking operations is lending, including services like personal loans, mortgages, and credit card loans, which have been instrumental in fueling economic growth over the past few decades. Loans, however, come with inherent risks, most notably the risk of default, where borrowers fail to meet their repayment obligations. The ability to accurately predict and manage this risk is crucial for the sustainability of financial institutions and the broader economy (Steenackers & Goovaerts, 1989; Dornadula & Geetha, 2019).

In recent years, the exponential increase in credit card usage has posed unique challenges for financial institutions. As credit becomes more accessible, the rate of credit card defaults has also risen, placing strain on banking resources and underscoring the need for efficient risk management strategies. Banks have increasingly turned to advanced data analytics and machine learning (ML) techniques to predict credit card defaults and other loan repayment issues. ML models can identify patterns in borrower behavior and predict the likelihood of default, thus allowing financial institutions to take preventive actions to mitigate risks. This paper seeks to provide a comparative analysis of various ML approaches—namely, Logistic Regression, Random Forest, and XGBoost—in predicting credit card defaults, evaluating the performance and robustness of each model in real-world applications (Butaru et al., 2016; Madaan et al., 2021).

**1.1 Background**

Credit risk refers to the potential that a borrower may fail to meet their financial obligations as stipulated in a loan agreement, leading to a "default." Effective credit risk management is essential, as high default rates can threaten the solvency of financial institutions, increase borrowing costs, and destabilize the broader economy. Given the importance of accurately predicting credit defaults, financial institutions have traditionally relied on statistical models such as discriminant analysis and logistic regression. However, recent advancements in ML have led to more accurate and adaptable solutions for credit risk assessment (Lai, 2020; Malekipirbazari & Aksakalli, 2015).

In traditional risk assessment, the applicant’s financial status, loan amount, age, and profession are some of the factors considered to determine loan eligibility. With the advent of ML, predictive models have become highly sophisticated, allowing financial institutions to harness large volumes of data to make more informed lending decisions. The evolution of credit scoring systems from purely statistical approaches to complex ML-based models reflects the industry's need for enhanced accuracy and efficiency in credit risk evaluation (van Liebergen, 2017; Bank of England & Financial Conduct Authority, 2019).

Machine learning has brought transformative changes to credit risk management by enabling models that learn from historical data to predict the likelihood of future defaults. ML models allow for greater accuracy by dynamically incorporating a vast array of variables and interactions that were previously challenging to analyze using traditional methods. For instance, as digital banking grows, financial institutions can now collect extensive transaction data, enabling the application of ML models to recognize patterns in spending behavior, employment history, and debt levels to predict credit risk. These models can adapt to evolving borrower behaviors and economic conditions, making them an invaluable tool for credit risk assessment (Li, 2019; Chou et al., 2018).

In addition to enhancing predictive accuracy, ML models support scalability, enabling financial institutions to process vast amounts of data and evaluate thousands of loan applications in real-time. This capability is especially valuable given the rapid growth of online shopping and mobile payments, which have generated massive amounts of transactional data. ML models such as Random Forests and XGBoost have demonstrated superior performance in handling large datasets and dealing with complex patterns in the data. Random Forest, for example, has outperformed traditional models in some cases due to its ensemble-based approach, which reduces variance and improves accuracy (Malekipirbazari & Aksakalli, 2015). XGBoost, a powerful gradient-boosting algorithm, has also gained traction due to its efficiency and predictive strength, particularly in handling high-dimensional data (Ma et al., 2018).

When choosing an ML model for credit default prediction, it is essential to balance accuracy with interpretability. Logistic regression, a statistical model, is widely used for binary classification problems and offers a relatively interpretable approach to credit risk prediction. However, it may lack the flexibility needed to capture complex patterns in data. On the other hand, Random Forests and XGBoost provide high accuracy but are generally less interpretable, as they involve complex ensembles of decision trees (Ma et al., 2018; Leo et al., 2019).

Random Forest, an ensemble learning method, is particularly effective for high-dimensional data as it aggregates multiple decision trees to improve prediction accuracy and reduce overfitting. Its robustness and ability to handle large datasets make it a preferred choice for credit default prediction, as seen in applications across various financial institutions (Chou, 2015; Madaan et al., 2021). XGBoost, meanwhile, has been designed for efficiency and scalability. This algorithm iteratively improves predictions by minimizing the error of previous models, making it especially suitable for large-scale and complex datasets where accuracy is paramount.

This study’s findings will have significant implications for financial institutions seeking to enhance credit risk assessment using ML techniques. By providing a clear understanding of the strengths and weaknesses of each model, this research aims to help banks and lending institutions make informed choices about their credit scoring systems.

Although much research has been conducted on credit risk prediction using machine learning, there is still a need for comparative studies evaluating model performance in specific contexts, such as credit card defaults. Additionally, while Random Forest and XGBoost offer high predictive accuracy, their complex structures make them less interpretable than simpler models, posing challenges for regulatory compliance and decision-maker trust. Future research could explore methods for improving the interpretability of these advanced models or integrating explainable AI techniques into the credit scoring process (Ribeiro et al., 2016; Vidovic & Yue, 2020).

**1.2 Problem Statement**

The rise in consumer credit has elevated the importance of accurately predicting credit default to maintain financial stability. With the expansion of credit card use, financial institutions face challenges in identifying potential defaulters, resulting in increased credit losses and economic instability (Lai, 2020). Traditional credit scoring systems, while effective to a degree, struggle to accommodate the complexity of modern credit data, prompting institutions to adopt machine learning (ML) approaches for predictive accuracy and efficiency (Madaan et al., 2021). However, the diversity of ML models, such as Logistic Regression, Random Forest, and XGBoost, presents a challenge in determining the most effective method for accurately predicting credit default. Each model offers unique advantages in terms of interpretability, accuracy, and feature handling, but their comparative effectiveness for this purpose remains underexplored (Malekipirbazari & Aksakalli, 2015). Additionally, despite advancements, a significant need remains to identify key factors influencing credit default, which would enable financial institutions to tailor risk management strategies effectively (Li, 2019). This study aims to evaluate and compare the predictive capabilities of various ML models for credit default prediction, seeking to identify the model that offers the highest predictive accuracy while highlighting the most impactful features influencing creditworthiness.

**1.3 Motivation**

The motivation behind this study arises from the growing prevalence of credit defaults, a critical risk factor for financial institutions as credit usage surges globally (Lai, 2020). Traditional methods of assessing credit risk, which often involve manual or statistical approaches, struggle to accurately capture complex, modern data patterns, leading to potential inaccuracies in default predictions (Madaan et al., 2021). Machine learning (ML) methods, such as Logistic Regression, Random Forest, and XGBoost, offer a promising alternative, providing enhanced accuracy by uncovering subtle insights in vast data. However, determining the most effective ML model for credit default prediction remains challenging (Malekipirbazari & Aksakalli, 2015). This study aims to compare these models, offering valuable guidance to improve predictive capabilities and support more robust risk management in lending institutions.

**1.4 Aim and Objectives**

This research aims to provide a comparative analysis of Logistic Regression, Random Forest, and XGBoost for credit default prediction, focusing on credit card consumers. The study will evaluate each model’s accuracy, interpretability, and robustness, examining how well they handle various feature interactions and respond to changes in data patterns. Key objectives include:

1. **Comparing Predictive Accuracy**: To measure and compare the accuracy of Logistic Regression, Random Forest, and XGBoost models in predicting credit defaults.
2. **Feature Impact Analysis**: To analyze how key features—such as financial history, income level, credit score, and spending behavior—impact model performance and prediction accuracy.
3. **Model Interpretability and Practicality**: To assess the interpretability and ease of use of each model in real-world banking scenarios, especially for decision-makers who may need to understand the reasons behind specific predictions.
4. **Recommendation of Best-Fit Model**: To identify and recommend the most effective model for credit default prediction based on the analysis.

**1.5 Research Questions**

1. How do machine learning algorithms such as Logistic Regression, Random Forest, and XGBoost compare in terms of predictive accuracy for credit default prediction?
2. What is the effect of key feature selections on the performance of machine learning models?

**1.6 Outlines for the dissertation**

This dissertation will analyze and compare machine learning approaches for credit default prediction. Chapter 2 will review existing literature on credit risk assessment, highlighting the evolution from traditional statistical methods to contemporary machine learning techniques. Chapter 3 will outline the methodology, detailing data collection, preprocessing, and the chosen algorithms: Logistic Regression, Random Forest, and XGBoost. Chapter 4 will present the results, showcasing model performance metrics. Finally, Chapter 5 will discuss the findings, linking them to prior research, and will conclude with recommendations for financial institutions on utilizing machine learning for credit risk management, along with suggestions for future research directions.

### CHAPTER TWO

### 2.0 Literature Review

### With the financial industry's growing reliance on automated decision-making procedures for lending, credit default prediction has emerged as a critical study topic. By providing strong tools for credit default prediction, machine learning (ML) helps financial institutions improve their risk assessment and management plans.

### 2.1 Definition of Credit Default

### The term "credit default" describes a borrower's inability to fulfill their responsibilities under a loan agreement, particularly their incapacity to make scheduled payments. Both borrowers and financial institutions, especially banks, face serious difficulties as a result of this issue. Default can have serious consequences for borrowers, such as collateral loss, legal penalties, harm to creditworthiness, and limited access to future lending choices (Domeher & Abdulai, 2012). Institutional banks experience significant capital and income losses, more expenses for tracking and collecting past-due loans, and decreased sustainability and profitability as a result of an increase in non-performing assets (Ahlin et al., 2011). Credit default has ramifications that go beyond specific lenders and borrowers; it has the potential to undermine the larger financial system, impacting investor confidence, liquidity, and general economic stability.

### 2.2 Importance of Predicting Credit Default for Financial Institutions

### Predicting loan defaults with accuracy is essential for risk mitigation and efficient financial management, which has a big impact on both specific financial institutions and the overall economy. For banks, anticipating borrower defaults is crucial since it guides their lending policies and risk management plans. According to Louzada et al. (2016), accurate default forecasting is essential for capital allocation optimization, reducing financial losses, and preserving a balanced loan portfolio. Predicting defaults necessitates a thorough examination of numerous risk indicators, which emphasizes the necessity for dynamic credit scoring models that can adjust to shifting borrower patterns and market dynamics.

### The variety of borrower profiles and the increasing complexity of financial markets provide significant challenges for traditional credit scoring methods. Financial organisations have historically found it challenging to manually assess a client's creditworthiness, especially during periods when their clientele is rapidly expanding. The delays in loan processing brought on by this human assessment may impair the effectiveness of banks and other financial organisations. Automating credit assessment procedures with machine learning and artificial intelligence has become essential to increase loan processing speed, accuracy, and economy (Albastaki & RIT Libraries, 2022). By employing AI-driven models, financial institutions may more precisely predict the likelihood of loan defaults, allowing for timely interventions and informed lending decisions. Actions should be taken to stop criminals because credit card fraud is increasing quickly on a global scale. Customers would benefit by setting a limit on those activities as their money would be reclaimed and returned to their accounts, and they wouldn't be billed for goods or services they didn't buy (AlEmad & RIT Libraries, 2022).

**2.3 Brief History of ML Applications in Finance**

The application of machine learning (ML) in the finance sector has a rich history, beginning in the early 1980s. One of the first instances was Apex’s launch of PlanPower in 1982, an AI program designed to provide tax and financial advice to clients with high incomes. The subsequent launch of the Personal Financial Planning System by Chase Lincoln First Bank in 1987 marked another milestone. In 1989, the introduction of the FICO Score revolutionized credit scoring, utilizing algorithms that remain foundational in today's banking practices (Sharma, 2023).

Over the decades, financial institutions have increasingly adopted machine learning techniques, particularly for applications such as fraud detection in credit card transactions. The British fund manager Man Group began employing ML for investment strategies in 2014, and by 2016, Bank of America launched its AI chatbot, Erica, enhancing customer interaction. The momentum continued in 2018 with various financial institutions developing recommendation systems to improve customer engagement and service (Sharma, 2023).

A recent survey conducted by the Bank of England (BoE) and the Financial Conduct Authority (FCA) revealed a growing adoption of machine learning within UK financial services, with respondents anticipating significant growth in ML applications over the coming years. The survey found that the median respondent expects their number of ML applications to more than double within three years, with banks and insurance firms predicting an even larger increase (Machine Learning in UK Financial Services, 2024). This trend indicates a strong recognition of the value that machine learning brings to financial operations, enhancing efficiency and decision-making.

## **2.4 How Machine Learning is Used in Finance**

Machine learning applications in finance are diverse, encompassing various functions that improve operational efficiency and decision-making. Some of the prominent use cases include:

### 2.4.1 Algorithmic Trading

Algorithmic trading leverages machine learning algorithms to enhance trading decisions by analyzing market data in real-time. Traders develop mathematical models that monitor business news and trading activities to identify factors that could influence security prices. These models operate under predetermined sets of parameters—such as timing, price, and quantity—allowing for automated trading without human intervention (Team, 2024). Unlike human traders, algorithmic trading can analyze vast datasets simultaneously, making rapid trading decisions that provide a competitive edge in the market. The emotional neutrality of algorithms also reduces the risk of poor trading decisions driven by psychological biases.

### 2.4.2 Fraud Detection and Prevention

Fraud detection remains a critical concern for banking institutions, leading to billions of dollars in losses annually. Traditional fraud detection systems often relied on rule-based approaches, which could easily be circumvented by sophisticated fraudsters. Modern financial services now utilize machine learning to detect unusual activities and anomalies within large datasets, significantly enhancing fraud prevention efforts (Leyden, 2024). By analyzing transaction data against historical account behaviors, machine learning systems can identify potential frauds in real-time, allowing financial institutions to act swiftly to mitigate losses.

### 2.4.3 Portfolio Management (Robo-Advisors)

Robo-advisors represent another innovative application of machine learning, offering automated financial advice to investors. These online platforms use algorithms to create personalized investment portfolios based on individual goals and risk tolerances (Team, 2024). With lower account minimums and fees compared to traditional portfolio managers, robo-advisors democratize access to investment strategies. By inputting their financial goals, investors can leverage these systems to optimize their portfolios across various asset classes, enhancing long-term financial outcomes.

### 2.4.4 Loan Underwriting

In the banking sector, machine learning streamlines the underwriting process by enabling rapid assessments of loan applications. Algorithms analyze extensive consumer data, such as age, income, occupation, and credit history, to make informed decisions regarding loan approvals (Team, 2024). By training machine learning models to recognize patterns and exceptions in consumer data, financial institutions can significantly reduce the time and resources spent on manual evaluations, leading to more efficient loan processing.

**2.5 Related Works**

The application of machine learning to credit risk assessment has gained traction due to its potential to analyze large datasets more effectively than traditional statistical methods. For instance, Rahman et al. (2018) developed a model that utilizes multiple machine learning techniques to predict loan defaults using a modified version of the German credit dataset. The authors selected 23 features through rigorous feature selection methods and employed algorithms including Logistic Regression, Naïve Bayes, Decision Trees, and Random Forests. Their findings indicated that the Decision Tree model was particularly effective for predicting credit defaults, underscoring the significance of feature selection in enhancing predictive accuracy.

Similarly, Uddin and Rahman (2024) conducted a comparative analysis of various machine learning algorithms, including Support Vector Machine (SVM), K-Nearest Neighbor (K-NN), and Random Forest. Their study demonstrated that Random Forest and Decision Tree models achieved the highest accuracy rates of 92% and 94%, respectively, when trained on a dataset from a private bank in Dhaka, Bangladesh. This reinforces the idea that certain algorithms, especially ensemble methods like Random Forest, can significantly improve predictive performance in credit default scenarios.

A central theme in the literature is the comparative analysis of machine learning algorithms. Bazzana et al. (2024) explored this by evaluating various classifiers on a dataset comprising small Italian companies. Their study found that machine learning models generally outperformed traditional logistic regression in predicting defaults. However, the performance gains were modest, suggesting that while machine learning offers improvements, classical models remain relevant due to their interpretability and ease of implementation.

Similarly, Liu et al. (2022) applied k-nearest neighbor, SVM, and Random Forest algorithms to predict default probabilities in China's online credit market. Their results indicated that machine learning models significantly outperformed logistic regression based on metrics such as accuracy and area under the ROC curve (AUC), thus demonstrating the enhanced predictive capabilities of these algorithms.

Athreyas et al. (2022) also contributed to this comparative discourse by analyzing various machine learning techniques to determine their efficacy in predicting loan eligibility and default. Their findings indicated that different algorithms showed varying levels of success, emphasizing the need for context-specific model selection.

Feature selection is a critical component in building effective machine learning models for credit default prediction. Wu (2022) employed feature engineering techniques such as variance threshold and Variance Inflation Factor (VIF) to filter out irrelevant features before applying Random Forest and XGBoost algorithms. Both models achieved high accuracy rates, demonstrating that effective feature selection can enhance the performance of ML algorithms significantly.

In the context of the Bank of Taiwan, Arora et al. (2022) utilized historical data to identify patterns related to loan defaults. Their analysis revealed that logistic regression provided valuable insights into the relationships between various independent variables and the likelihood of default, emphasizing the importance of interpretability in machine learning applications.

Beyond predictive accuracy, the economic implications of adopting machine learning for credit default prediction are profound. Alonso et al. (2020) investigated the economic benefits of using machine learning models like XGBoost compared to traditional models such as logistic regression. Their simulations indicated substantial savings in regulatory capital, suggesting that implementing advanced ML models could result in significant financial advantages for banks.

Chen (2022) focused on the practical application of linear regression and neural networks to predict loan default behavior. The study highlighted that improved prediction models could effectively reduce default risks, thereby enhancing the overall risk management framework for financial institutions.

Despite the promising results, the literature also highlights several challenges associated with implementing machine learning models for credit default prediction. Neema et al. (2017) emphasized the issue of imbalanced datasets, where non-defaulting customers far outnumber defaulters, complicating the classification task. Their study explored cost-sensitive learning approaches to improve prediction accuracy while maintaining a balance between false positives and false negatives.

The table below summarizes the related works on machine learning approaches for credit default prediction

| **Author(s) with Year** | **Title** | **ML Techniques Used** | **Key Findings** | **Limitations** |
| --- | --- | --- | --- | --- |
| Rahman et al. (2018) | A Machine Learning Approach to Credit Default Prediction and Individual Credit Scoring | Logistic Regression, Naïve Bayes, Decision Trees, Random Forest | Developed a model to output a credit score; Decision Tree was most effective in predicting loan defaults. | Limited generalizability due to dataset modification. |
| Uddin & Rahman (2024) | A Comparative Study of Machine Learning Algorithms for Enhanced Credit Default Prediction | SVM, K-NN, Logistic Regression, Decision Tree, Random Forest | Random Forest and Decision Tree achieved highest accuracy (92% and 94%). | Dataset may not represent all types of credit markets. |
| Arora et al. (2022) | Prediction of credit card defaults through data analysis and machine learning techniques | Logistic Regression, Decision Trees, Random Forest | Found that ML algorithms can accurately predict defaults, with logistic regression providing valuable exploratory insights. | Relied on historical data, which may not reflect current trends. |
| Wu (2022) | Machine Learning Approaches to Predict Loan Default | Random Forest, XGBoost | Random Forest and XGBoost showed similar high accuracy (~90%) in predicting defaults after effective feature selection. | Limited comparison with other advanced ML techniques. |
| Bazzana et al. (2024) | Machine learning techniques for default prediction: an application to small Italian companies | Various ML classifiers, Logistic Regression | ML models slightly outperformed logistic regression, but the performance gain was modest; classical models remain valuable for their interpretability. | Small sample size may limit broader applicability. |
| Liu et al. (2022) | Applying machine learning algorithms to predict default probability in the online credit market | K-NN, SVM, Random Forest | Machine learning models outperformed logistic regression, providing significant benefits to investors based on predictive accuracy. | Focused on a specific market segment, limiting generalization. |
| Neema et al. (2017) | The comparison of machine learning methods to achieve most cost-effective prediction for credit card default | Various ML methods, cost-sensitive approaches | Explored cost-effective predictions by balancing accuracy with costs associated with misclassifications; found Random Forest performed best in cost vs. accuracy. | Imbalance in dataset could skew results. |
| Alonso et al. (2020) | Understanding the performance of machine learning models to predict credit default | Logistic Regression, Lasso, CART, Random Forest, XGBoost, Deep NN | ML models outperformed traditional models in classification and calibration, suggesting significant economic benefits in regulatory capital requirements. | Complexity of models may hinder implementation in smaller institutions. |
| Chen (2022) | Prediction and Analysis of Financial Default Loan Behavior Based on Machine Learning Model | Linear Regression, Neural Networks | Improved logistic regression and neural networks showed advantages in predicting default risk effectively. | May require extensive data preprocessing for optimal performance. |
| Athreyas et al. (2022) | A Comparative Study of Machine Learning Algorithms for Predicting Loan Default and Eligibility | Various ML algorithms | Aimed to determine effective algorithms for predicting loan repayment capacity, highlighting the need for context-specific model selection. | Limited by the availability of relevant training data. |

The application of machine learning in credit default prediction has significantly advanced the field of financial risk assessment. Studies consistently demonstrate that machine learning techniques, particularly ensemble methods, outperform traditional models in accuracy and predictive power. However, challenges such as feature selection, data imbalance, and interpretability remain pertinent. Continued research is essential to navigate these challenges and to harness the full potential of machine learning in credit risk management.

The evolving nature of credit markets necessitates this research to refine machine learning techniques for credit default prediction. There is limited exploration of how different features interact and impact model performance, particularly among credit card consumers. This research will address these gaps by conducting a thorough comparative analysis of these three models (Logistic Regression, Random Forest, and XGBoost), evaluating their predictive accuracy, interpretability, and robustness in handling various feature interactions. By focusing specifically on credit card consumers, this study aims to provide practical insights for decision-makers in banking, enhancing the understanding of model performance in real-world applications.

**CHAPTER THREE**

**3.0 Methodology**

By estimating a customer's probability of missing payments, credit default prediction models help financial organisations manage risk. Machine learning provides strong tools for creating these predictive models as structured data becomes more widely available. This section explores the procedures and methods used to develop a model that uses a variety of machine learning approaches to forecast credit default. Data preparation, feature engineering, model selection, evaluation metrics, and model performance analysis are all covered in this part.

### 3.1 About the Dataset

The UCI Machine Learning Repository is the source of this dataset, which sheds light on the 2006 credit card debt problem that affected Taiwanese banks. Many banks gave credit cards to unqualified applicants in an effort to gain market share, which led to widespread debt buildup and higher cardholder delinquency rates. Customers' trust in personal finance was eroded by this practice, which presented serious problems for banks and consumers alike.

### 3.1.1 Column Information

Repayment status (PAY\_0 to PAY\_6) from April to September 2005 is one of the important features in the dataset that provides information about customer behaviour and financial history. Values in this field reflect whether payments were made on time, late, or significantly past due. Payment amounts (PAY\_AMT1 to PAY\_AMT6) reflect prior payments made each month, whereas bill amounts (BILL\_AMT1 to BILL\_AMT6) reflect the monthly statement balances in NT dollars. These characteristics are crucial for assessing credit risk and comprehending repayment patterns.

### 3.2 Data Preparation

### 3.2.1 Loading and Initial Analysis

### Loading the dataset and performing a preliminary examination to comprehend its structure are the first steps in the data preparation process. The first rows are shown using df.head(), exposing a number of characteristics pertaining to client demographics and credit history. SEX, EDUCATION, MARRIAGE, and default the goal variable that indicates if a consumer has defaulted are important characteristics. This stage helps us find key features for the credit default prediction model by giving us a brief summary of the data. Running df.info() also provides useful information on the counts of non-null values, the data types for each feature, and any possible missing or inconsistent data that should be cleaned up before modelling.

### 3.2.2 Dropping Unnecessary Columns

### It's possible that some of the dataset's columns are irrelevant to the prediction task. In this instance, the ID field has no predictive value and only functions as a unique identifier. Using df.drop('ID', axis=1, inplace=True), it is removed from the dataset because it has no direct correlation to the likelihood of default. By removing superfluous variables from the dataset, this phase lowers the likelihood of overfitting and frees the model to concentrate on more significant characteristics. Additionally, eliminating unnecessary columns improves data processing and storage, which is very advantageous in machine learning workflows.

### 3.2.3 Encoding Categorical Variables

### Since many machine learning algorithms need numerical input, categorical features must be encoded. The categorical columns in this dataset are MARRIAGE, EDUCATION, and SEX. LabelEncoder is used to convert these labels into numerical numbers. For instance, SEX categories like "male" and "female," respectively, are changed to 0 and 1. In a similar manner, categories in MARRIAGE and EDUCATION are represented numerically. Additionally, the target variable default undergoes label-encoding, which transforms it into a binary format appropriate for classification tasks. To guarantee interoperability with machine learning models and to standardise the data format, these variables must be encoded.

### 3.2.4 Exploratory Data Analysis (EDA)

### The purpose of exploratory data analysis is to learn more about the dataset's distribution, relationships, and any problems. Essential metrics, such as the mean, standard deviation, min, and max values for every feature, are provided via basic statistical summaries (df.describe()). These summaries show each variable's range and can indicate the existence of outliers or skewed distributions that could impair model performance.

### For every variable, histograms are plotted in order to better evaluate feature distributions. Skewness, kurtosis, and other anomalies in the data, including extreme values or concentrations around particular spots, can be found with the use of these visualisations. For instance, to enhance model interpretability and performance, adjustments can be necessary if a variable's histogram has a significant skew.

### To investigate the connections between characteristics and spot possible multicollinearity problems, a correlation heatmap is created. We can see which features are significantly associated with the target variable and with each other by computing the correlation matrix and displaying it as a heatmap. While features with high inter-correlation may be redundant, those with a significant correlation to the target variable may be very useful in the model. Therefore, the heatmap acts as a guide for feature selection by highlighting which variables might need to be changed, removed, or combined.

### EDA provides information about feature distributions, correlations, and possible problems with data quality. The subsequent stages of the data preparation process, including feature scaling, managing skewed variables, and selecting feature engineering tactics, are guided by this knowledge. By minimising noise and optimising the information available to the model, these data preparation procedures together produce a refined dataset that is better suited for machine learning, ultimately increasing the predictive ability and resilience of the model.

### 3.3. Feature Selection and Engineering

### 3.3.1 Feature Importance using Random Forest

### A RandomForestClassifier is used to compute feature importance in order to identify which attributes have the most impact on the target variable. The dataset is used to train the classifier, and a bar chart displays the feature importance results. This graphic makes it evident which characteristics are most important in forecasting credit default. Features that have a major influence on the model's predictions are indicated by higher significance ratings. We may improve the model's efficiency and interpretability by concentrating on these important aspects. While keeping crucial predictive information, less significant traits may be eliminated to cut down on noise and the possibility of overfitting.

### 3.3.2 Outlier Detection

### Machine learning models' accuracy and stability may be impacted by outliers. Box plots are created for every feature in order to highlight data points that significantly depart from the normal range. Errors in data entry, measurements, or special circumstances can all lead to outliers. To lessen their influence, techniques like adjustment (such as log scaling) or removing extreme values may be used, depending on the distribution and model performance. When outliers are handled correctly, the model maintains predicted accuracy and generalises well to new data.

### 3.3.3 Target Variable Analysis

It is crucial to examine the distribution of the target variable, especially in order to spot any class imbalance in the dataset. Whether there are an unequal number of default and non-default cases in the dataset is indicated by a count plot for the target variable (default). Model performance can be distorted by class imbalance, frequently favouring the majority class. In order to balance the dataset and increase the model's capacity to forecast minority class outcomes, methods like modifying class weights or using resampling approaches (like SMOTE) may be employed if there is a considerable imbalance.

### 3.4 Data Splitting and Scaling

### 3.4.1 Splitting Data

### A 70-30 split is used to separate the dataset into training and test sets. This guarantees that 30% of the data is set aside for testing and 70% is used to train the models. In order to ensure that any class imbalance in the original data is reflected proportionately in both the training and test data, stratification is used to maintain the target variable's distribution across both sets. As a result, assessment metrics become more trustworthy.

### 3.4.2 Feature Scaling

### StandardScaler is used to scale each feature to a mean of zero and a standard deviation of one in order to standardise feature data. Because it increases convergence speed and model stability, this is crucial for models like logistic regression that are sensitive to feature scales.

**3.5 Modeling and Evaluation**

Three machine learning models—Logistic Regression, Random Forest Classifier, and XGBoost Classifier—are used in this investigation to forecast credit card default. These models were selected because to their diverse strengths, which include robust performance on unbalanced datasets, interpretability, and insights into feature relevance. To counteract any potential class imbalance that can skew models towards the majority class, balanced class weights are given to each model after it has been trained on the dataset. To give a comprehensive evaluation of each model's performance, evaluation measures including accuracy, precision, recall, and F1-score are computed using confusion matrices and classification reports.

**3.5.1 Logistic Regression**

For binary classification applications, logistic regression is frequently used due to its simplicity and interpretability. As a linear model, it assigns probabilities to each instance by assessing the link between the attributes and the chance of credit default. Standardisation approaches are used to train the model using scaled data, which enhances the performance of distance-based models such as Logistic Regression. A confusion matrix that includes the number of true positives, true negatives, false positives, and false negatives is used to assess the predictions on the test set. The categorisation report also displays important parameters including F1-score, recall, accuracy, and precision. These measures offer a thorough analysis of the model's effectiveness in accurately separating defaults from non-defaults, facilitating a deeper comprehension of its predictive power.

**3.5.2 Random Forest Classifier**

In order to increase accuracy and decrease overfitting, Random Forest is an ensemble technique that constructs several decision trees and aggregates their predictions. Because it can handle high-dimensional data and offer insights regarding feature relevance, this model is especially useful for identifying the characteristics most closely linked to default risk. To guarantee that minority classes are treated fairly, the balanced class weight parameter is included. A confusion matrix and a classification report that summarises accuracy, precision, recall, and F1-score are used to assess Random Forest's performance, same like in Logistic Regression. Compared to single-decision-tree models, the model's ensemble nature improves stability and accuracy, and its feature importance rankings provide insightful information about how different qualities affect default risk.

**3.5.3 XGBoost Classifier**

An improved gradient boosting technique called XGBoost is used because of its great efficiency and predicted accuracy, particularly when applied to structured data. XGBoost is a boosting technique that works especially well with imbalanced datasets since it creates trees in a sequential fashion, each one attempting to fix the mistakes of the one before it. Hyperparameters are used to optimise the model in order to avoid overfitting and guarantee that it recognises intricate patterns in the data. A confusion matrix and classification report are used to assess XGBoost's predictions on the test set following training on the scaled dataset. XGBoost is a strong option for credit default prediction because of its high precision and recall scores, which show that it is able to strike a compromise between forecast accuracy and sensitivity to the minority class.

All things considered, the combination of these models offers a thorough assessment of the dataset. XGBoost provides high accuracy and robustness against imbalances, Random Forest provides insights into feature relevance, and Logistic Regression provides interpretability. Based on the precise needs of accuracy, interpretability, and model resilience, the optimum model for credit default prediction can be found by comparing their performances.

### 3.6 Confusion Matrix and Model Comparison

A confusion matrix that shows the numbers of True Positives, True Negatives, False Positives, and False Negatives is created for each of the three models—Logistic Regression, Random Forest Classifier, and XGBoost Classifier—to provide a thorough analysis of performance. This matrix, which shows the accuracy of forecasts for defaulters versus non-defaulters, is a crucial tool for evaluating each model's predictive quality. In particular, the matrix sheds light on each model's capacity to differentiate between true defaulters and non-defaulters, which is critical for assessing model trustworthiness in this situation. We can determine which model is most effective at lowering misclassification—specifically, False Positives and False Negatives—by comparing the confusion matrices. This has a direct bearing on risk management.

**3.7 ROC Curve and AUC Score**

Plotting the True Positive Rate (sensitivity) against the False Positive Rate, the Receiver Operating Characteristic (ROC) curve illustrates the trade-offs between sensitivity and specificity and offers a visual depiction of each model's performance. Because it exhibits a better balance between accurately recognising genuine positives and minimising false positives, a model with a high Area Under the Curve (AUC) score is typically more robust.

**3.8 Conclusion**

In summary, this work uses data from the UCI Machine Learning Repository to show how machine learning models, such as Logistic Regression, Random Forest, and XGBoost, can be used to forecast credit default. These models provide useful insights into credit risk by utilising data preparation strategies like encoding, feature selection, and addressing imbalances. A thorough evaluation is provided by metrics such as accuracy, precision, recall, and AUC, with XGBoost demonstrating the highest performance because of its capacity to manage imbalances and intricate patterns. This emphasises how crucial model tuning and selection are to financial organisations' ability to anticipate loan default effectively.

**CHAPTER FOUR**

**4.0 Result and Analysis**

**4.1 Classification report Analysis**

This section describes the performance evaluation of three machine learning models (Logistic Regression, Random Forest, and XGBoost) on a binary classification problem. The dataset consists of 9000 samples with an unbalanced class distribution: 1991 instances of class 1 (minority class) and 7009 instances of class 0 (majority class). Precision, recall, F1-score, and accuracy were used to evaluate the efficacy of each model.

Table 4.1: Summary of the classification results for the three models

| **Metric** | **Logistic Regression** | **Random Forest** | **XGBoost** |
| --- | --- | --- | --- |
| **Accuracy** | 0.68 | 0.81 | 0.81 |
| **Precision (Class 0)** | 0.87 | 0.83 | 0.84 |
| **Precision (Class 1)** | 0.37 | 0.64 | 0.63 |
| **Recall (Class 0)** | 0.70 | 0.95 | 0.94 |
| **Recall (Class 1)** | 0.63 | 0.34 | 0.36 |
| **F1-Score (Class 0)** | 0.77 | 0.89 | 0.89 |
| **F1-Score (Class 1)** | 0.47 | 0.44 | 0.46 |

The table above provides a clear comparison of the three models across all key performance metrics.

**4.1.1 Logistic Regression**

The overall accuracy of the Logistic Regression model was 68%. With an F1-score of 0.77 for class 0, the model demonstrated high precision (0.87) but comparatively low recall (0.70). The model, on the other hand, had trouble with the minority class (class 1), obtaining a respectable recall of 0.63 but a significantly lower precision of 0.37, which led to an F1-score of 0.47.

With a precision of 0.62, recall of 0.67, and F1-score of 0.62, the macro-average scores—which treat both classes equally regardless of size—indicate modest competence. With an F1-score of 0.71, the weighted averages, which account for class distribution, demonstrate how dominant the majority class is in influencing overall performance.

**4.1.2 Random Forest**

The Random Forest classifier showed an 81% increase in total accuracy. The model produced a high F1-score of 0.89 for the majority class (class 0) by achieving good precision (0.83) and excellent recall (0.95). The model, however, had trouble generalising for the minority class, achieving a recall of just 0.34 and a precision of 0.64, with an F1-score of 0.44.

Precision (0.74), recall (0.64), and F1-score (0.67) are the macro-average scores that show how differently the two classes performed. Despite the model's good majority class predictions, the weighted averages (all around 0.79), show that class imbalance significantly impacts the model's performance on minority occurrences.

**4.1.3 XGBoost**

Similar to Random Forest, XGBoost demonstrated high efficacy for the majority class with an overall accuracy of 81%. Class 0 performed similarly to Random Forest, with precision of 0.84, recall of 0.94, and F1-score of 0.89. On the other hand, XGBoost achieved an F1-score of 0.46 for the minority class with a recall of 0.36 and a precision of 0.63.

The precision (0.73), recall (0.65), and F1-score (0.67) macro-average metrics show somewhat worse performance than Random Forest. The accuracy of the model on the majority class has a significant impact on its performance, as evidenced by the weighted averages (precision and F1-score of 0.79).

**4.1.4 Comparative Analysis**

Although it sacrifices precision for the minority class, logistic regression shows a balanced recall for both classes (0.70 and 0.63, respectively). While Random Forest and XGBoost perform well in majority class predictions, they struggle in minority class management, especially in recall. Despite not substantially improving minority class recognition, both ensemble approaches exceed logistic regression in terms of total predictive potential, with an accuracy of 81%.

The difficulty of class imbalance is highlighted by the low recall for class 1 across all models. Additional strategies like oversampling, cost-sensitive learning, or ensemble modification may improve the ability of ensemble methods like Random Forest and XGBoost to discover minority instances, even though they provide better majority class performance.

**4.2 Confusion Matrix Analysis**

This part uses the confusion matrix findings, which are displayed as heatmaps, to assess the performance of three classification models: Logistic Regression, Random Forest, and XGBoost. A fuller comprehension of the model's advantages and disadvantages is made possible by the confusion matrix elements, which offer insights into true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN).

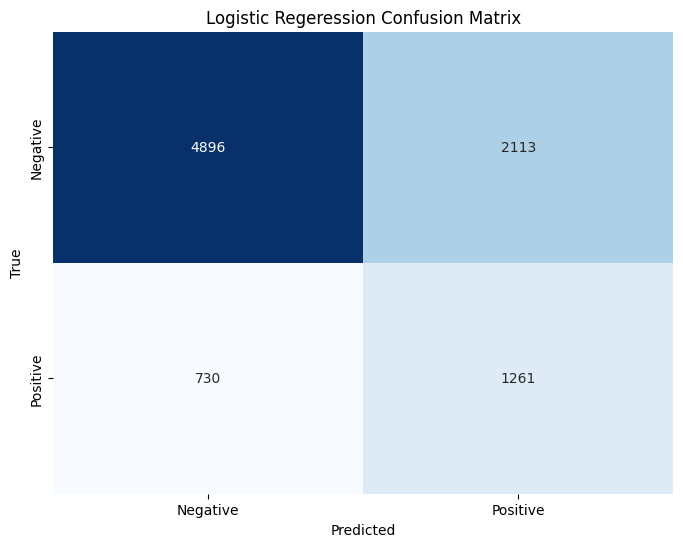


Fig: Logistic Regression Confusion Matrix Heat Map

Although it performed well overall, logistic regression had trouble with the minority class. The model achieved reasonable detection of both classes with 1261 correctly recognised positives (TP) and 4896 correctly identified negatives (TN). While 730 false negatives imply some difficulty in recognising minority class situations, the high false positive rate (2113) suggests a tendency to misclassify negative instances as positive.

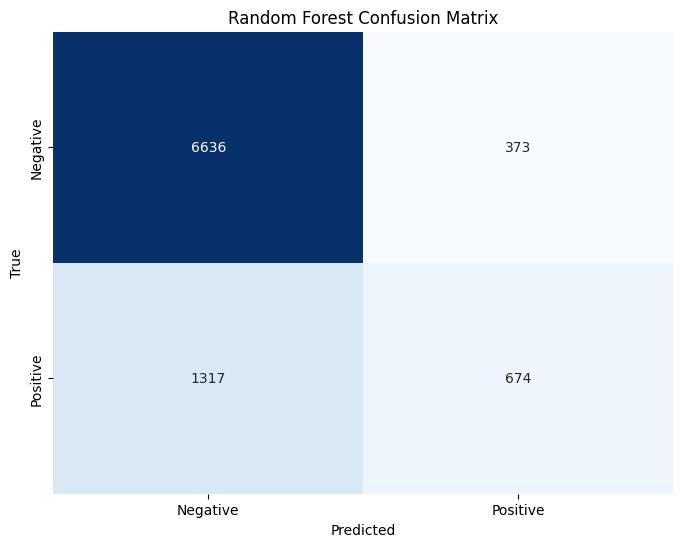


Fig: Random Forest Classifier Confusion Matrix Heat Map

The high TN score (6636) and low FP (373), indicate that the Random Forest model performed exceptionally well in determining the majority class. Nevertheless, its performance on the minority class is less than excellent, with 1317 false negatives and only 674 correctly detected positive occurrences. This demonstrates a notable compromise that favours majority class correctness at the expense of minority class memory.

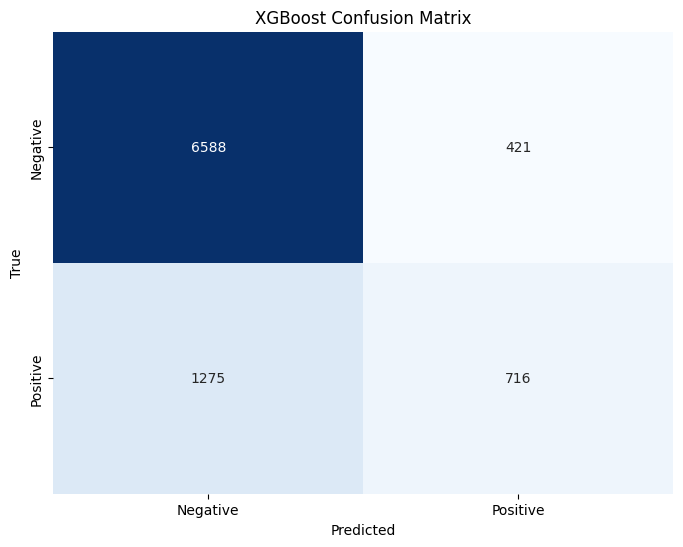


Fig: XGBoost Classifier Confusion Matrix Heat Map

XGBoost achieved strong identification of the majority class (6588 TN), performing comparable to Random Forest. On the minority class, however, it fared somewhat better than Random Forest, as evidenced by the higher TP count (716) and lower FN (1275). In contrast to Random Forest, its FP count (421) is marginally greater, suggesting that more negatives are mistakenly classified as positives.

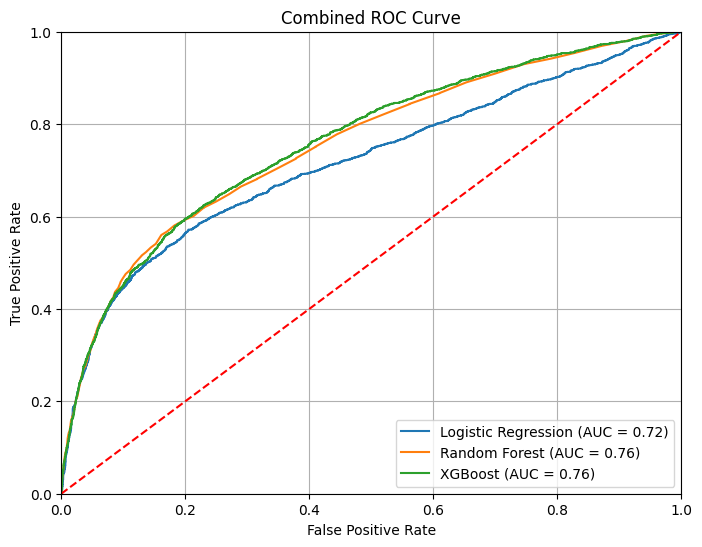
### **Table 4.2: Comparative Insights**

| **Metric** | **Logistic Regression** | **Random Forest** | **XGBoost** |
| --- | --- | --- | --- |
| **True Negatives (TN)** | 4896 | 6636 | 6588 |
| **False Positives (FP)** | 2113 | 373 | 421 |
| **False Negatives (FN)** | 730 | 1317 | 1275 |
| **True Positives (TP)** | 1261 | 674 | 716 |

* **Majority Class (TN & FP):** Random Forest and XGBoost performed significantly better than Logistic Regression, with lower FP values and higher TN values.
* **Minority Class (FN & TP):** Logistic Regression outperformed both ensemble models in recognising minority class instances, achieving the highest TP count (1261) and lowest FN count (730).

Performance across classes is more evenly distributed with logistic regression, especially for the minority class. On the other hand, Random Forest and XGBoost perform poorly when it comes to minority class detection, with XGBoost marginally outperforming Random Forest in minority class detection. Random Forest and XGBoost, on the other hand, are excellent at detecting instances of the majority class. To address the inequalities found, future enhancements such redistributing class weights or using resampling techniques are advised.

**4.3 AUC ROC Curve Analysis**



The combined ROC curve evaluates how well the Random Forest, XGBoost, and Logistic Regression models perform in binary classification. The Area Under the Curve (AUC) measures overall performance, and the plot shows the trade-off between True Positive Rate (TPR) and False Positive Rate (FPR) at different thresholds.

In comparison to the ensemble models, logistic regression has a respectable but lower discriminatory power, as evidenced by its AUC of 0.72. Its curve, which is farther from the top-left corner, illustrates how difficult it is to efficiently discern between classes.

With an AUC of 0.76, Random Forest and XGBoost both perform better than Logistic Regression. Their curves have comparable strengths in balancing TPR and FPR and roughly overlap. By more successfully utilising intricate data linkages, these ensemble approaches exhibit strong categorisation.

All things considered, Random Forest and XGBoost perform better in terms of prediction, but Logistic Regression offers a less complicated but less precise substitute.

**4.4 Conclusion**

The analysis concludes by highlighting significant variations in how well the Random Forest, XGBoost, and Logistic Regression models perform on an unbalanced binary classification task. Although it sacrifices overall accuracy and precision, logistic regression shows balanced recall across classes and successful minority class detection. While Random Forest and XGBoost do better than Logistic Regression in terms of majority class predictions and overall accuracy (81%) but have trouble with minority class recall. In minority class detection, XGBoost performs somewhat better than Random Forest. The AUC values attest to the fact that ensemble approaches make better use of data complexity. Using strategies like oversampling, cost-sensitive learning, or class-weight modifications, future research should concentrate on resolving class imbalance.

### CHAPTER FIVE

### 5.0 Discussion

This study analyses model performance and feature effect to investigate the predictive powers of XGBoost, Random Forest, and Logistic Regression for credit default prediction. All features were included in the modelling process, according to feature importance values generated from ensemble models; the most significant predictors were PAY\_0 (prior payment delay status), AGE, and BILL\_AMT1 (current bill amount). The performance of the models and the contribution of feature importance to answering the research questions are assessed in this discussion.

#### **5.1 Comparison of Model Performance**

Metrics like accuracy, precision, recall, F1-score, and AUC were used to assess the models, giving a fair assessment of their prediction ability.

##### **5.1.1 Logistic Regression**

With an accuracy of 68%, Logistic Regression was the least accurate of the three models. It outperformed Random Forest and XGBoost in detecting true defaults, nevertheless, with a recall of 0.63 for the minority class (credit defaults). Because of this, Logistic Regression is a useful model for reducing false negatives, which is essential for managing credit risk.

With a greater percentage of false positives, its precision for the minority class was 0.37. In situations where the cost of incorrectly classifying a consumer as high-risk is substantial, the low precision is a disadvantage. Reasonable generalisation is indicated by its balanced recall across the majority and minority classes. In contrast to ensemble models, logistic regression's predictive accuracy is limited by its inability to handle non-linear connections, despite its interpretability and simplicity.

##### **5.1.2 Random Forest**

At 81%, Random Forest demonstrated a significant increase in accuracy. With a recall of 0.95 and precision of 0.83, it showed excellent performance in recognising non-defaults (majority class), yielding a high F1-score of 0.89. It is a dependable option for forecasting clients who are unlikely to default because of these metrics.

Nevertheless, Random Forest had the lowest recall of all the models for the minority class, at 0.34. This poor default detection performance indicates that the model puts overall accuracy ahead of recognising clients who are at risk. Although the AUC score of 0.76 suggests a reasonable level of discriminatory strength, the unbalanced recall raises questions regarding its use in situations where robust default identification is required.

##### **5.1.3 XGBoost**

Similar to Random Forest, XGBoost had an 81% accuracy rate. In terms of forecasting non-defaults, it was quite similar to Random Forest, with comparable precision, recall, and F1-score for the majority class. Interestingly, XGBoost scored somewhat better than Random Forest in minority class recall (0.36), suggesting a little higher ability to detect defaults.

Its ability to successfully balance true positive and false positive rates is further demonstrated by its AUC score of 0.76. The effectiveness of boosting algorithms in capturing intricate correlations between features is demonstrated by XGBoost's performance. Nevertheless, the minority class recall continues to be a drawback, illustrating the difficulties in correcting for class imbalance in ensemble models.

##### **5.1.4 Overall Comparison**

##### When compared to Logistic Regression, Random Forest and XGBoost both showed higher accuracy and AUC, demonstrating their capacity to capture intricate patterns in the data. But in minority class recall, Logistic Regression scored better than them, highlighting its usefulness for applications that prioritise default identification. These findings highlight the significance of choosing a model based on particular use-case needs by exposing a trade-off between minority class sensitivity and overall accuracy.

#### **5.2 Impact of Feature Importance**

#### The ensemble models' feature significance values shed light on the variables affecting credit default. The main features—PAY\_0, AGE, and BILL\_AMT1—emphasize how important they are in determining model predictions.

##### **5.2.1 Key Features**

##### **PAY\_0 (Importance: 0.100)**: The status of the most recent payment is the best indicator of default, according to the highest-ranked characteristic, PAY\_0. According to current credit risk theories, customers who have missed payments in the most recent period are more likely to default.

##### **AGE (Importance: 0.067)**: The second most significant factor was age, which may be a reflection of variations in credit usage and financial conduct between generations. Due to their inexperienced credit records or erratic income, younger borrowers may be more vulnerable to default.

##### **BILL\_AMT1 (Importance: 0.061)**: Credit utilisation, a known risk factor in predicting credit default, has a direct correlation with the present bill amount. Financial stress is frequently indicated by larger outstanding balances in comparison to credit limitations.

##### **5.2.2 Inclusion of All Features**

##### All features are included in the modelling process, demonstrating their differing degrees of significance. Model performance was also greatly influenced by payback amounts (PAY\_AMT1, PAY\_AMT2, etc.) and payment history features (PAY\_2, PAY\_3, etc.), with significance scores ranging from 0.05 to 0.04. Together, these factors provide a thorough understanding of credit risk by capturing a borrower's payment patterns across time.

##### Compared to payment history and bill-related parameters, demographic variables like sex, education, and marital status exhibited lower significance scores (0.01 to 0.02), indicating poor predictive potential. Their inclusion, however, gives the models more context, which could enhance their robustness and interpretability.

##### **5.2.3 Feature Importance and Model Performance**

##### The feature importance analysis emphasises how important payment history and bill-related characteristics are in predicting credit default. These characteristics were successfully utilised by Random Forest and XGBoost, which helped explain their great accuracy. Although it also benefited from these factors, the overall performance of logistic regression was limited by its inability to capture intricate interactions.

##### The necessity of feature engineering to improve model performance is also shown by the investigation. For instance, reducing dimensionality and increasing predictive power may be achieved by integrating BILL\_AMT variables into a single feature that represents cumulative debt. Likewise, calculating ratios like credit utilisation (LIMIT\_BAL to BILL\_AMT, for example) may offer more information about borrower behaviour.

#### **5.3 Implications for Credit Default Prediction**

#### The findings have important ramifications for managing credit risk. Because of its higher recall for defaults, logistic regression is appropriate for early warning systems where reducing false negatives is crucial. Its poor accuracy, however, indicates that additional models are required to properly handle non-default predictions.

#### Large-scale credit scoring systems are better suited for Random Forest and XGBoost due to their higher accuracy and discriminatory capacity. Their reliance on bill-related characteristics and payment history is consistent with accepted methods for evaluating credit risk. Their shortcomings in minority class recall, however, point to the need for strategies like cost-sensitive learning or oversampling to overcome class imbalance.

#### Feature importance analysis supports the idea that the main factor influencing credit risk is payment behaviour. The incorporation of demographic factors promotes interpretability and equity in model deployment, despite their low direct predictive power. The findings also emphasise how crucial it is to use domain expertise to improve feature engineering and selection in order to increase model performance and applicability.

**5.4 Limitations**

1. **Class Imbalance**

With most observations falling into the non-default class, the dataset showed a notable class imbalance. Model performance was impacted by this mismatch, especially in the minority class (default), where Random Forest and XGBoost had very low recall scores. The imbalance might have affected the overall accuracy of Logistic Regression even though it did better in this particular sector.

1. **Feature Engineering**

The study didn't use a lot of feature engineering; it just used raw features. The inclusion of derived variables like credit utilisation ratios or cumulative repayment patterns may have increased the prediction accuracy and robustness of the models, even though payment history and bill-related factors were important predictors.

1. **Model Interpretability**

Despite the inherent interpretability of Logistic Regression, the intricacy of Random Forest and XGBoost makes it difficult to comprehend their decision-making processes; while feature importance offers some insight, it falls short in explaining feature interactions and their combined impact on predictions.

1. **No Hyperparameter Optimisation**

Because Random Forest and XGBoost were not extensively hyperparameter tuned in the study, their full potential might have been limited. Better outcomes might have been achieved with optimisation strategies like grid search or Bayesian optimisation.

**5.5 Recommendations**

1. **Addressing Class Imbalance**

Techniques like class-weight modification, undersampling, and oversampling (like SMOTE) should be investigated in order to enhance minority class performance. In line with the objectives of credit risk management, cost-sensitive learning may also give priority to accurately classifying defaults.

1. **Enhanced Feature Engineering**

Richer data could be added to the models by using derived variables like credit utilisation ratios, total bill amounts, or repayment patterns. It is best to use domain expertise to develop features that capture complex borrower behaviour.

1. **Incorporating Interpretability Techniques**

Advanced interpretability tools like as SHAP (Shapley Additive Explanations) or LIME (Local Interpretable Model-agnostic Explanations) ought to be used for Random Forest and XGBoost. These techniques can increase confidence in model outputs by shedding light on how attributes affect certain predictions.

1. **Hyperparameter Optimisation**

To improve the Random Forest and XGBoost parameters, future research should use optimisation methods as grid search, random search, or Bayesian optimisation. Performance metrics can be enhanced by this phase, especially when it comes to minority class prediction.

1. **Explainable AI for Credit Decisions**

It will be crucial to incorporate explainability frameworks into model deployment as regulatory attention shifts towards equity and transparency. This guarantees that credit choices are impartial, comprehensible, and consistent with moral principles.

#### **5.6 Conclusion**

This study provides a comparative analysis of Logistic Regression, Random Forest, and XGBoost for credit default prediction, revealing trade-offs between accuracy, recall, and interpretability. Feature importance analysis highlights the dominance of payment history and bill-related features in shaping model performance. Future research should explore techniques to address class imbalance and further refine feature engineering strategies to enhance the effectiveness of machine learning models in credit risk management.