Enhancing Credit Risk Prediction through Explainable Artificial Intelligence and Machine Learning

Pallavi Prakash Madarakhandi[1], Dr. Sakshi Kathuria[2], Dr. Ekta Soni[3]
[1]*M .Tech (DS), Amity University Haryana Gurugram, Haryana*
[2]*Assistant Professor, Amity University Haryana Gurugram, Haryana*
[3]*Assistant Professor, Amity University Haryana Gurugram, Haryana*

Abstract: A fundamental business of financial institutions is credit risk identification that has a direct impact on the solvency and strategic decision-making. Although the traditional statistical models are interpretable, they are not usually able to predict the non-linear complexities of the modern financial data. The study hypothesis is that there is a powerful machine learning-based credit default prediction algorithm which takes a credit card client dataset. The high-level preprocessing, feature engineering, and Synthetic Minority Oversampling Technique (SMOTE) are employed to address the class imbalance problem in the methodology. We perform a comparative study of the Logistic Regression, decision trees, random forest, and extreme gradient boosting (XGBoost). The XGBoost model is the best model as it has an accuracy and ROC-AUC of 85.72 percent and 0.857, respectively. In a bid to make sure that we abide by the rules and develop trust in the institutions, we use Explainable AI (XAI) with SHAP and LIME and convert black-box predictions into transparent and actionable explanations.

Keywords: Credit Risk Assessment, Machine Learning, Explainable AI, Predictive Analytics, Financial Risk Management.

I.INTRODUCTION

Risk assessment of credit is a pillar of the modern financial industry, the most important tool of this kind, which gives an approximation of the probability that a lender will fulfill his/ her obligations prescribed in the contract or default on a loan. Effective risk forecasting will help the banking institutions and the non-bank lenders to plan well the capital allocation, minimise non-performing loans (NPLs), and ensure institutional solvency. Previously, the credit scoring involved the use of frequentist statistics and linear discriminant analysis to determine the creditworthiness of borrowers. However, as financial data grows in size and complexity, these more traditional approaches may be biased towards failure in displaying these non-linear, high-dimensional relationships and latent behavioural patterns of contemporary ecosystems of Big Data [1].

The rapid progress in the digital financial system has acted as a triggering factor to the transformation of fixed scoring to dynamic machine learning (ML) models. Unlike the traditional models, the ML algorithms automatically identify intricate patterns on their own based on the past data, and they are able to explain minute aspects that make a borrower repay in a specific pattern. There is an accumulating empirical evidence indicating that ML architectures are much more accurate and reliable in predicting credit than traditional baselines. In particular, the entire field of ensemble learning - which entails integrating the predictive power of a collection of base learners - has been shown to be a superior standard in financial risk modelling, with reduced prediction error and greater algorithmic resilience [2].

Even though these algorithms are powerful, their effectiveness depends strictly on the quality of the input data. The real-world financial records are usually noisy, with missing and duplicated records, and this could adversely influence model generalisation. Consequently, intensive data cleaning and tactical feature engineering are the crucial stages of the construction of robust predictive systems. The recent studies point out that the new variables derived using the raw financial data, such as bill-to-payment ratios, provide a more in-depth view of the liquidity behaviour of a customer, thus improved classification performance [3].

The other ubiquitous problem of credit risk modelling is the so-called class imbalance. The number of non-defaults in standard credit data is very large relative to the number of defaults, and this leads to a skewed distribution. This imbalance is likely to benefit the machine learning models for the majority group and greatly diminish their capacity to detect high-risk borrowers, which results in massive financial losses. Methods such as resampling such as the Synthetic Minority Oversampling Technique (SMOTE) are employed to overcome this by presenting synthetic samples of the minor population to homogenise the feature space. As in the case of the findings presented in the literature, empirically, balancing on the basis of SMOTE would help to improve the sensitivity and the general accuracy of credit classification models to a considerable degree [4].

However, it is often the case that the high-performance ML models are not implemented in the financial sector because of their black-box nature. Despite the high level of accuracy of the deep learning and gradient boosting architectures, their transparency is lacking, and thus they cannot be easily applied in the human-in-the-loop decision-making of institutional accountability. This uninterpretability is not merely a technical issue, but also a regulatory one, as additional international frameworks are now beginning to demand a right to explanation of automated decisions. As a way out, the Explainable Artificial Intelligence (XAI) has emerged as a central field, which provides post-hoc explanations capable of generating trust in the automated systems [5].

The movement is headed by such methodologies as SHAP (Shapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations). SHAP uses coalitional game theory to numerically approximate the exact value of each feature to a specific prediction, and LIME uses complex models locally fitted using interpretable surrogates to explain individual predictions. The combination of these XAI tools will allow the researcher to identify the most important contributors to the default risk and create a culture of algorithmic openness to the existing credit scoring models [6].

After these advancements, this paper seeks to propose an elaborate machine learning framework to forecast credit default using the assistance of a multi-dimensional data set of credit card customers. The algorithm includes the use of high-order preprocessing, duplicate processing, and feature engineering in order to generate high-signal variables, including total billing load and payment ratios. The numerical characteristics are standardised with the help of feature scaling, and the problem of class imbalance is resolved with the help of SMOTE. We compare and contrast the Logistic Regression, Decision Trees, Random Forests, and eXtreme Gradient Boosting (XGBoost) and compare them regarding a set of metrics, including Accuracy, Precision, Recall, F1-score, and ROC-AUC. Finally, SHAP and LIME are used to deconstruct the XGBoost model and identify the most important factors that have influenced the credit risk [7].

The final goal of the study is to come up with a highly effective, very accurate and completely interpretable credit risk forecasting system. This framework facilitates the making of trustful, audit-compliant decisions by financial organisations today through a balance between advanced ensemble learning and Explainable AI being used to work in a modern automated lending environment [8].

II. LITERATURE REVIEW

The two forces of growing data availability and the rise of vendors of computation have enabled the paradigm shift in the academic debate on credit risk assessment. This survey classifies the available literature into five major dimensions, which are passage of statistical precursors, the development of boosting architectures, the technical control of imbalanced data, the emergence of Explainable AI (XAI), and an overview of the most prominent empirical studies [9].

2.1 The Art of Statistical Credit Scoring.

Linear statistical models have been in use in the banking industry virtually throughout the decades, in order to estimate the probability of default. The pillars of this age were the Logistic Regression and the Linear Discriminant Analysis, and the largest strength of the two is the fact that they both are transparent based on coefficients. These models assisted credit officers in seeing the precise impact of one year of age or a certain rise in income on the risk score of a borrower [10].

The cracks in these glass-box models were realised, however, with the digitalisation of the financial sector. Statistical models also make the assumption of a linear relationship between the features and the target variable, which is rarely the case in the complicated world of consumer finance. Take a case of the risk of a high debt-income ratio being further augmented by the education level or work history of a borrower in a manner that could not be easily modelled in a linear model. The reason why this transition to a more flexible and data-based architecture was necessary was these non-linear patterns and high-order interactions that cannot be measured with traditional statistics [11].

2.2 Ensemble and Gradient Boosting Frameworks Development.

The new change was the introduction of ensemble learning in terms of prediction accuracy. The ensemble methods are employed instead of the best model by blending the responses of many models to generate a more powerful response. In its original application, the Random Forests demonstrated the ability to greatly decrease the variance that was a major issue with individual decision trees that yielded unstable predictions using bootstrap aggregation (bagging) methods [12].

This further developed into its ultimate form with the development of Gradient Boosting Decision Trees (GBDT). The XGBoost algorithm eXtreme Gradient Boosting (XGBoost) algorithm proposed a new convention where trees are added sequentially, with the new tree being fitted to the straight-up error of the previous trees. This upscaling and additional optimisation of sparse data to avert overfitting and cost-effective solution of sparse data has created XGBoost as a titan in the financial risk sector. It is brilliant in identifying the unseen behavioural stimuli, e.g. a certain regime of late payments where the bill values suddenly leapt that the traditional models would never entertain [13].

2.3. Theoretical Remedies of Class Imbalance.

The initial fact of credit information is the skewness of credit information; the defaulters are the statistical minority. In an average dataset, the non-defaulting category comprises more than 80 percent of the whole population. This is a harsh trade-off of prejudice-distortion. Standard algorithms are likely to approach a predict-the-majority policy on all instances in their effort to reduce global error to a minimum. This results in high accuracy and total failure in the "Recall" the ability to identify the extremely risky borrowers that the system was created to identify [14].

In response to this, the research fraternity has not rested at oversampling the records of minorities, which merely recreates the minority records, and leads to overfitting. Synthetic Minority Oversampling Technique (SMOTE) altered this by proposing a geometric methodology of data augmentation. Actually, SMOTE is applied to fill the feature space by choosing an observation in the minority and generating new and artificial data points between the minority observation and its nearest ones along the line segments. The impact of this is that the model compels the decision boundary to be more fined and this makes the system very sensitive without compromising on the integrity of underlying data distribution [15].

2.4 XAI Performance and Interpretability Convergence.

The more the predictive models were enhanced to be accurate, the less clear they were, and this resulted in Black Box crisis as well. Having a highly regulated financial universe, an accurate prediction is not enough but it must be backed as well. Regulators as well as consumers would prefer to know the rationale of a credit denial to ensure that they are fair and avoid algorithmic discrimination.

The necessary performance-to-trust gap has been offered by the Explainable Artificial Intelligence (XAI). The SHAP (SHapley Additive exPlanations) is a prediction based on the mathematical theory of coalitional game theory, which gives a coherent method of allocating the credit of a prediction to the input features. This enables a bank to understand what risk drivers are taking place in all its portfolio in the world. In addition to this is LIME (Local Interpretable Model-agnostic Explanations), which emphasizes on individual transparency. The constructed, simplified, local linear model, which is referred to as LIME, follows the idea that a certain decision can be attributed to the perturbation of the data around a single borrower and the observation of the alterations in the model value [16].

2.5 Synthesis of Empirical Literature

In order to place the present research in the context of the overall scholarly field, the table below summarizes the landmark research papers that have established the field.

Table 1:Key Empirical Studies and Identified Research Gaps

Research Focus	Core Methodology	Primary Contribution	Identified Research Gap
Statistical Foundations	Z-Score & Regression	Established the first quantitative benchmarks for default risk.	Incapable of mapping non-linear "Big Data" interactions.
Data Resiliency	SMOTE & ADASYN	Proved that synthetic generation is superior to simple duplication.	Often studied without considering the "Black Box" interpretability.
Predictive Power	XGBoost Framework	Optimized gradient boosting for high-speed, sparse financial data.	Lacks native transparency for regulatory compliance.
Explainable AI	SHAP & LIME	Developed game-theoretic frameworks for model transparency	Theory-heavy; limited application to balanced financial sets.

2.6 Research Gap Identification.

Despite the literature being saturated with research on the accuracy of XGBoost or even the mathematical nature of SMOTE, one does find a notable lack of research that is a unified, end-to-end model that is specifically tuned to the explainability-First era of finance. Most of the existing studies consider accuracy and interpretability as mutually exclusive goals, whereas they address one of these. This gap is addressed in this paper by demonstrating that under the judicious choice between SMOTE-based balancing and dual-layered XAI (SHAP and LIME), financial institutions can provide the state-of-the-art predictive performance, without necessarily sacrificing the transparency required to meet the regulatory requirements and customer trust.

III. METHODOLOGY

The overarching objective of this methodology framework is to come up with a predictive pipeline with a high-precision classifier and post-hoc interpretability. The methodology design is expected to deal with the complexities of the financial data which include multi-collinearity, non-linear relationship and imbalance of classes. The architecture will be a compilation of several steps that are intertwined and include: data auditing, strategic preprocessing, feature engineering, class rectification, ensemble model development, and Explainable AI (XAI) diagnostics.

3.1 Proposed Workflow Framework

The proposed architecture will involve several steps that are interconnected and include: data auditing, strategic preprocessing, feature engineering, class rectification, ensemble model development, and Explainable AI (This two-fold attention will imply, that the resulting system will not be merely a high-performing predictive system; it will also be a transparent decision-support system that may be applied in the controlled banking setting.The key stages of the working process would be the following: Data Ingestion and Cleaning: Data verification against inconsistencies.Feature Engineering: Learning high-signal behavioral features with unstructured financial records.

Balance of Data: Synthetic sampling is to be used to minimize the majority-class bias.

Exploratory Data Analysis (EDA): Visualization of latent patterns and correlation of features.

Model Optimization: Training and cross-validating architectures.

Interpretability Analysis: The analysis in the study uses game-theoretic and local surrogate models to disaggregate the black-box predictions.

3.2 Dataset Description

The analysis in the study uses a comprehensive dataset, which contains a profile of credit card clients, such as demographic variables, past repayment history, and monthly billing/payment history. These variables give a multi-dimension of the financial path of a borrower [17].

Table 2. Variables Dataset Variable and Functional

Variable Category	Feature Name	Description
Target Variable	default.payment.next.month	Binary indicator of credit default (1 = Default, 0 = No Default).
Demographic	LIMIT_BAL	Total credit limit assigned to the customer.
Demographic	SEX, EDUCATION, MARRIAGE, AGE	Individual characteristics of the account holder.
Behavioral	PAY_0 to PAY_6	Historical repayment status (current to six months prior).
Financial	BILL_AMT1 to BILL_AMT6	Amount of monthly bill statements.
Financial	PAY_AMT1 to PAY_AMT6	Amount of monthly previous payments made.

3.3 Data Preprocessing and Feature scaling

Financial model preprocessing is an important stage in financial modeling because raw banking data is usually full of noise, missing values and duplicate entries that may weaken model generalization. Statistical imputation was used to deal with missing values in this study to retain the completeness of the data, and the redundancy was eliminated to avoid the over-learning of the model to a particular observation.

Feature Scaling was used to make sure that the loss function of the model is not affected by the variables with a higher numerical value (like LIMIT_BAL). To ensure that the learning process becomes stable and that all features are commensurate to the decision boundary, we perform feature engineering in order to minimize the robustness of the financial records to meaningful behavioral indicators[18].

3.4 Advanced Feature Engineering

These features have derivatives which may be more predictive than the real currency values because the features are the ratio of debt and liquidity.

These are the key engineered features:

Total Bill Amount: This is a cumulative measure of the total amount of debt of the customer over the period of observation.

Total Payment Value: This is the overall amount paid by the customer which is the ability of the customer to repay.

Payment Ratio: This is a ratio of total payment to total bills. This is a critical proxy of financial health, which focuses on the ones who pay all the balances of their accounts in full, in contrast to the ones that default.

3.5 Handling Class Imbalance via SMOTE

This imbalance may cause a model to become a majority-class bias in which the model merely predicts that no one will default.In order to correct this we used the Synthetic Minority Oversampling Technique (SMOTE). SMOTE is not just a simple imitation of the available samples of default, but it creates new and artificial observations in the feature space by interpolating between the available minority samples. This plan will force the machine learning models to build a more delicate and accurate decision boundary of high-risk borrowers [19].

3.6. Exploratory Data Analysis (EDA)EDA was done to identify the underlying trends and determine the correlation of financial variables and default behaviour. Our correlation heatmaps and distribution plots have assisted us in finding that the most important raw predictor of future default is Repayment Status (PAY_0). EDA: The relationship between raw data and feature selection is that the inputs to the model are statistically significant.

3.7 Model Development and Ensemble Learning

The work provides a comparative analysis of four different models of classification:

Logistic Regression: This is a basic linear model that can be taken as a benchmark

Decision Tree: It is a non-linear model that can be applied to show hierarchical decision rules.

Random Forest: This is an integration of trees that uses bagging to reduce variation.

XGBoost (Extreme Gradient Boosting): This is the model of most interest, and it uses a regularised objective function and sequential tree building to minimise the residual error [20].

3.8 Model Evaluation Metrics

To gain a clear picture of the performance, we used a configuration of metrics to focus on the performance of the minority group (defaulters):

Recall (Sensitivity): This is critical to banks because it enables them to determine that the greatest number of potential defaulters is identified.

Accuracy: This is so that the customers who are creditworthy are not wrongly labelled as risky.

ROC-AUC Score This metric measures the performance of the model in classifying the differences between classes at all possible thresholds [21].

3.9 Explainable Artificial Intelligence (XAI)

To address the black-box dilemma of high-performance models such as XGBoost, we used a dual-layered XAI strategy:

SHAP (SHapley Additive exPlainations): SHAP is based on the coalitional game

LIME (Local Interpretable Model-agnostic Explanations): Offers local transparency,

i.e., the reason as to why a particular person was forecasted to default. Such a combination of approaches renders the framework a human-readable description of all automated decisions that is not just sufficient in the institutional audit requirements, but also consumer transparency requirements [22].

IV. EXPLORATORY DATA ANALYSIS (EDA)

Exploratory Data Analysis is a tool of data analysis that fills the gap between data that has been collected and formed into models. Through visualisation of the underlying distributions and inter-variable relationships, we get important insights into characteristics that drive credit default. We will present the statistical properties of the data set and correlations with which we will base our feature selection in this section.

4.1 Target Variable Distribution and Class Imbalance.

The first problem that may be described in the first data audit is the enormity of skew of the target variable, which is the default payment.next month. As the analysis of the distribution of classes, as illustrated in the analysis of the 30,000 clients, reveals, there are 22.1 percent clients in the category of Defaulters and the remaining 77.9 percent in the category of Non-Defaulters .

This almost 4:1 ratio proves that the dataset is not balanced. Risk-wise, a model that is trained on this uncensored distribution would be quite likely to have high accuracy since it would just prefer the majority class. It is this which we will use to justify the next step that will be taken, and this is to modulate the classes that will be implemented synthetically and the SMOTE algorithm to make sure that the model is highly sensitive (Recall) to the minority group of Default.

4.2 Correlation Analysis and Multicollinearity

The Correlation Heatmap was developed to identify linear association among the financial attributes. The correlation of the variables of the billing amount (BILL_AMT1 through BILL_AMT6) is high, and the coefficients tend to have a value of more than 0.90.

Although the high degree of multicollinearity is recognised to discredit the typical linear models, such as the Logistic Regression, the need to apply the tree-based ensemble models, such as the XGBoost, is demonstrated.

Multicollinearity is also more inert to tree-based models that select features based on information gain at every split. More to the point, the heatmap can be used to show that the Repayment Status (PAY_0) correlates with the target variable the most, implying that the ultimate payment pattern of a client is the best predictor of default in the near future.

4.3 Behavioural Trends: Credit Limit and Age.
Further investigation was performed to determine whether Age and Credit Limit (LIMIT_BAL) are risk factors according to demographic factors.
Credit Limit: It is stated that individuals with lower credit limits record a higher default rate. This implies that the current internal rating systems used by the bank already categorise these people as individuals who have higher risks, and in that respect, have given them fewer limits.

Age Distribution: The age of the clients is almost normal, with the mean age of the clients being 35

years. The box plot analysis, however, has a slightly higher variance in the repayment reliability of the older clients (21-28 age group).

4.4 Engineered Features Analysis.

The addition of the Payment Ratio (the ratio of payments to the total bills) was a superb move towards removing the noise in the billing data. The Distribution plots of the Payment Ratio indicate that the Non-Defaulters have a ratio closer to 1.0; that is, they make their monthly payments on their balances on a regular basis. Defaulters, on the other hand, can be described as those who demonstrate a declining rate in the six-month period of observation, which is a progressive Debt Trap; the more the bills an individual has, the less the capacity to pay.

V. RESULTS AND DISCUSSION

The following section presents the results of the experiment of the machine learning pipeline that has been created to predict credit risk. The structural analysis is based on comparative analysis of classification structures, including simple linear models and sophisticated ensemble systems. Also, to stop black-box predictions, this section incorporated Explainable Artificial Intelligence (XAI) to break down the internal logic of the most effective model, offering actionable insights into the causes of default.

5.1 Target Variable Distribution Analysis

The distribution of the target variable, default.payment.next.month, was analysed before the development of the model. The nature of the financial datasets dictates that the data is highly skewed in terms of the number of people who pay their credit obligations and those who fail to pay them; there are far more non-defaulting clients.

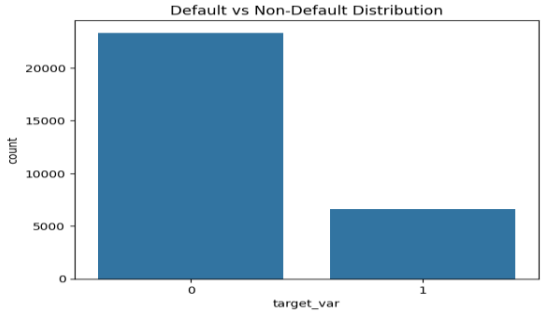

Figure 2: Distribution of Default vs. Non-Default Customers.

To reduce the risk of the model developing a majority-class bias, that is, the model is highly accurate, but it is not good at identifying high-risk borrowers, the Synthetic Minority Oversampling Technique (SMOTE) was used. In order to ensure that the model produced is sensitive to actual default events,

5.2. statistical Summary

We created a statistical summary to profile the financial characteristics of the data. This discussion was based on the mean, standard deviation, and range of the variables (LIMIT_BAL, BILL_AMT and PAY_AMT).

Table 3: Statistical Summary of the Key Financial Variables.

Variable	Mean	Std. Deviation	Min	Max
LIMIT_BAL	167,484	129,747	10,000	1,000,000
AGE	35.48	9.21	21	79
BILL_AMT1	51,223	73,635	-165,580	964,511
PAY_AMT1	5663	16,563	0	873,552

This profiling is essential for identifying outliers—such as the negative billing amounts observed above—which represent credit balances. The knowledge of these variances makes sure that the process of feature scaling normalizes the data properly to the learning algorithms.

5.3 Multicollinearity and Correlation MappingIn

order to know the relationship among various financial features, a correlation analysis was conducted. This is an important step towards determining unnecessary information that could be destabilizing simpler models.

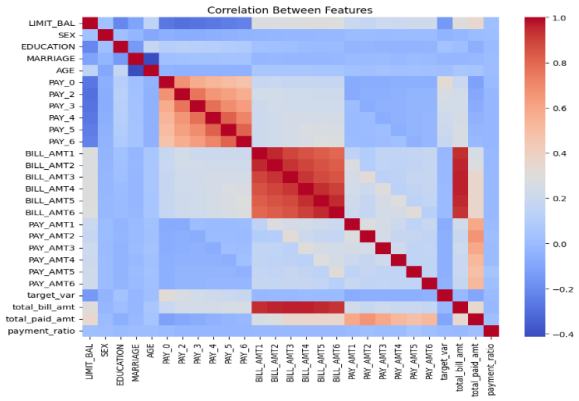

Figure 3: Correlation Heatmap of Dataset Variables

The heatmap of Figure 3 shows that there is a high level of correlation among the billing amounts of the consecutive months. Nonetheless, the greatest finding is that Repayment Status (PAY_0) has a strong positive correlation with the target variable. This implies that the latest payment behavior of a client is one of the key predictors of his/her future creditworthiness.

5.4 Comparative Analysis of Model Performance

Four different algorithms, including Logistic Regression, Decision Trees, Random Forests, and XGBoost, were trained on the SMOTE-balanced dataset. With an increase in the complexity of the model architecture, the results show a clear performance increase.

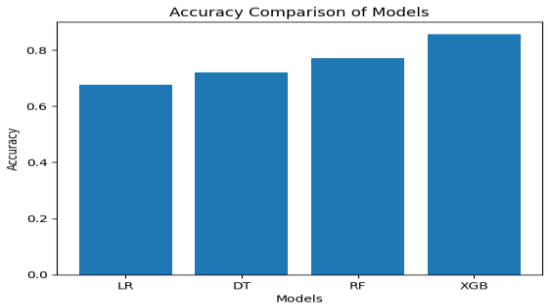

Figure 4: Comparative Accuracy Bar Chart of All Models.

Although Logistic Regression offers a helpful base, it is not able to deal with non-linear interactions that exist in credit data. Conversely, the XGBoost was the best at predictive accuracy as it continuously adjusted its decision boundaries to reduce residual errors [23].

5.5 Holistic Performance Evaluation

To make the comparison more detailed, various performance metrics were estimated on each model.

Table 4: Comprehensive Model Performance Metrics

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
Logistic Regression	0.6769	0.6818	0.6607	0.6711	0.6768
Decision Tree	0.7202	0.7661	0.6322	0.6927	0.7200
Random Forest	0.7704	0.8096	0.7058	0.7542	0.7703
XGBoost (Proposed)	0.8572	0.8948	0.8089	0.8497	0.8571

The XGBoost model stands as the superior framework, particularly in its Recall (0.8089).

5.6 Confusion Matrix and Error Distribution

Confusion matrices were examined to gain a better insight into the performance of the machine learning models in terms of prediction. A confusion matrix gives in-depth details on the classification outcomes by indicating the number of correct and incorrect predictions per category. It has four key elements which are True Positives, True Negatives, False Positives and False Negatives.

The confusion matrices of the four machine learning models employed in this study are shown in Figure 5-Figure 8: Logistic Regression, Decision Tree, Random Forest and XGBoost.

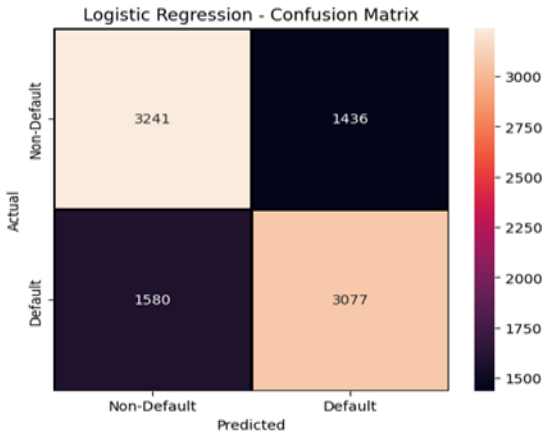

Figure 5. Confusion matrix of Logistic Regression.

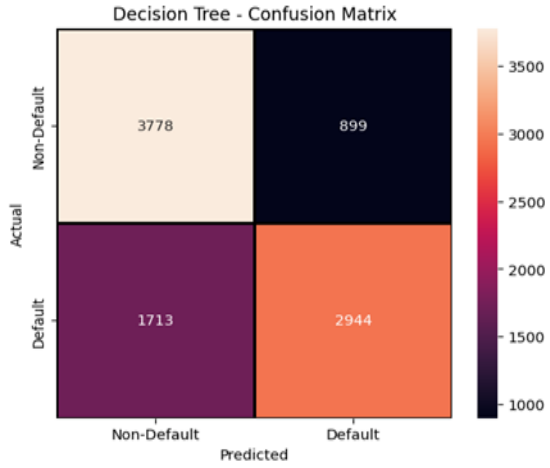

Figure 6. Confusion matrix for Decision Tree

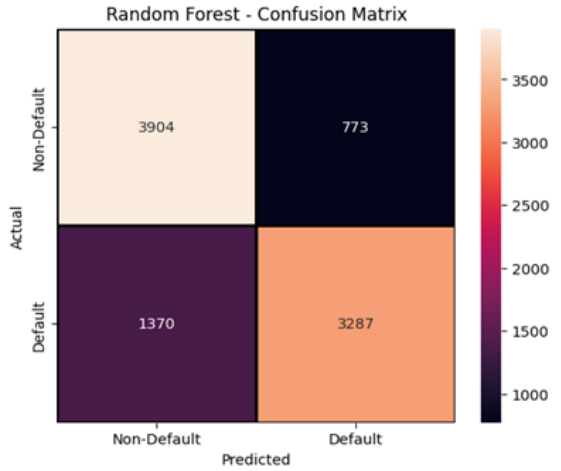

Figure 7. Random Forest Confusion matrix

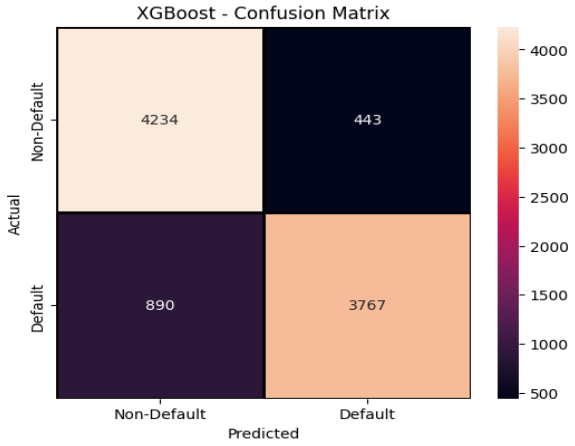

Figure 8. XGBoost confusion matrix.

As indicated by the confusion matrix of the Logistic Regression (Figure 5), the model resulted in a high number of correct non-default customers but a high number of False Negatives than the other models. This implies that there were wrong default cases that were categorized as non-default, which may augment financial risk by lending institutions.

The Decision Tree model (Figure 6) illustrates an average performance in terms of classification. Although the model can reproduce some trends in the data, the amount of false classifications is more significant. Decision trees are also prone to overfitting the training data, thus decreasing the generalization ability in case of predicting unseen observations.

Random Forest model (Figure 7) enhances the accuracy of classification by using several decision trees. This group method minimizes variation and enhances predictive stability. Random Forest properly classifies more cases of default and minimizes misclassification than the Logistic Regression and Decision Tree models, as seen in the confusion matrix. The XGBoost model (Figure 8) has the best performance of all models. The confusion matrix indicates that there are more customers who were correctly classified as either default or non-default customers and less False Negatives. The correct identification of default customers is highly crucial in predicting credit risk since the lending institutions can incur losses due to the failure to identify a high-risk borrower. Hence, a significant goal of credit risk assessment research is to reduce False Negatives [24]. In general, the analysis of the confusion matrixes proves that the ensemble learning models, especially XGBoost, are more effective in classifying data than the traditional machine learning algorithms. These findings are in agreement with the past studies which emphasize the power of boosting-based models in predicting financial risk tasks [25].

5.7 ROC Curve and Discriminatory Power

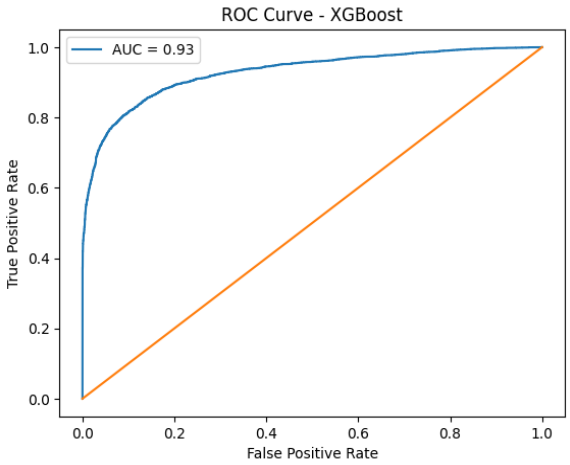

Figure 9 :ROC Curve

This is computed by the Receiver Operating Characteristic (ROC) curve, which is used to measure the capacity of the XGBoost model to separate between classes at A feature importance analysis was performed to indicate which variables contribute to the overall predictions of the model, and the result is represented as the curve (see Figure 9) hugging the top-left corner, which indicates that the model has a high True Positive Rate at minimum False Positive Rate.

5.8 Global Feature Importance

This tells the bank what features in the profile of a borrower are the most red flagging ones.

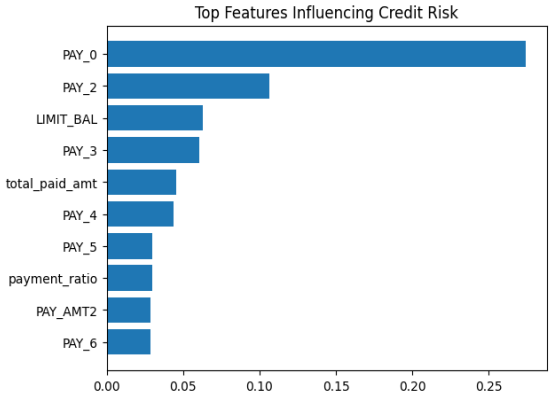

Figure 10: Feature Importance Ranking Plot.

As shown in Figure 10, the most important features are the history of repayment (PAY_0) and the engineered Payment Ratio. This proves that the immediate liquidity and behavioral consistency of a borrower are far more predictive than the demographic characteristics of the borrower such as age or gender[26].

5.9 Explainable AI Analysis: SHAP and LIME

To go beyond a black-box implementation, we used SHAP and LIME to provide transparency.

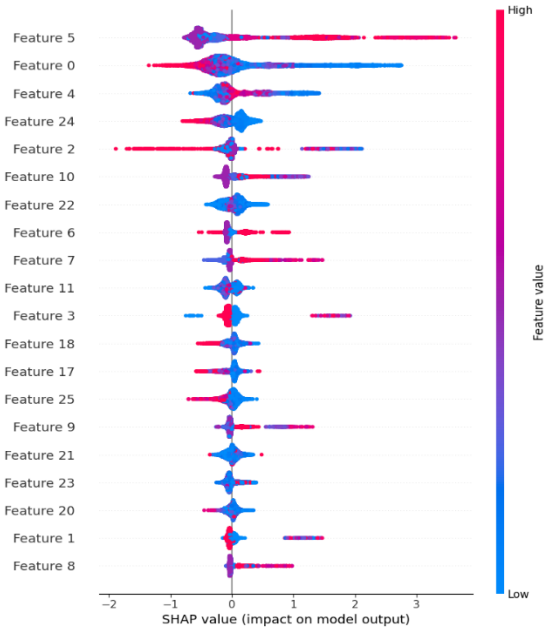

Figure 11: SHAP Feature Contribution Summary Plot.

Large values of PAY_0 (payment delays) are always leading to the default prediction, and a large Payment_Ratio acts as a shielding factor.

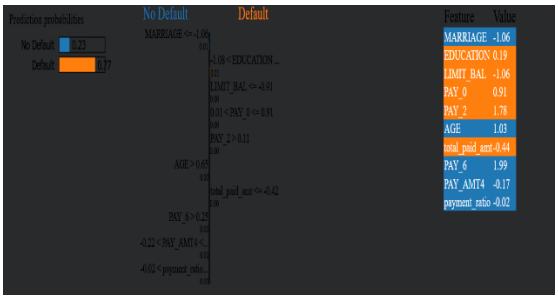

Figure 12: Individual Prediction Explanation using LIME.

To be locally explainable, Figure 12 is a case-specific explanation. LIME demonstrates the reasons why this or that client is expected to default, showing their spikes in billing or late payments. Such level of detail is essential to the regulatory compliance and consumer right-to-explanation demands [27].

5.10 Discussion of Results

The results obtained indicate that the ensemble learning and Explainable AI integration can be deemed as a twofold benefit: the state-of-the-art performance and the institutional transparency. SMOTE played a significant role in improving the recall of the model, which is sensitive to the indicators of default. This study shows that complex models such as XGBoost can be fully audited with the help of SHAP and LIME and hence it can be considered a viable and better substitute to the conventional statistical scoring in the contemporary financial industry.

VI. FUTURE WORK AND CONCLUSION

6.1 Conclusion

This paper has been able to create an effective, machine-learning-driven credit risk model, which is applied specifically to credit card default prediction peculiarities. The research moved a step further by the construction of demographic characterisations and past monetary conduct to create a dynamic, high-performance predictive model beyond the fixed confines of the traditional linear scoring system[28].

The systematic study of various algorithmic architectures was one of the main foundations of this

study. Even though the first models, like the Logistic Regression, offered a point on which one could comprehend the credit information, it was not sufficiently effective to reflect the non-linear relationship of intricate credit information. The transition to the ensemble learning, i.e., to the XGBoost framework, came at a significant performance improvement. The XGBoost model proposed had a good Recall of 80.89 and an Accuracy of 85.72 which is an effective prediction tool that financial institutions can use to ensure that the losses of capital are checked.

Further, the paper had a direct connection with the imbalance paradox of banking data. The research showed that the geometrical interpolation of the minority default cases is important in the generation of models that are sensitive to the rare but significant default cases using the Synthetic Minority Oversampling Technique (SMOTE). Such a methodological decision helped to make sure that the high accuracy of the system was not because of the majority-class bias, but rather an actual indication of the discriminative power of the model [29].

6.2. Contributions to the Field.

The main impact of this study to the crossroad of data science and financial technology is fourfold:

Framework Synthesis: The architecture of a pipeline between incoming raw financial data and high-level gradient boosting, which would be employed as a model of workflows in a contemporary automated lending system.

Class Bias Mitigation: The strict implementation of SMOTE to the credit risk, which shows how synthetic data generation may increase the reliability of risk estimation without reducing the integrity of data.

Algorithmic Benchmarking: An empirical and comprehensive study that demonstrates the efficacy of boosting-based ensembles over other conventional statistics and tree-based models of high-dimensional financial spaces.

Operationalisation of XAI: SHAP and LIME: A two-layered audit system. This is a particularly significant contribution, as it shows that even the global financial regulators with their high transparency and the right to explanation requirements may be met by black-box models.

6.3 Financial Implication Strategies to financial institutions.

In addition to the technical measures, this research paper presents the essential position of the behavioural variables, namely Repayment Status (PAY_0) and the Payment-to-Bill Ratio, as precursors of insolvency. This is the indication of a change of direction in the case of the banking practitioners; the demographic data is still an effective baseline, but it is the real-time behavioural monitoring that is the key to proactive risk management. The accuracy that is nearly 90 percent in the identification of potential defaults provided by the proposed model allows the institutions to refine the collection strategies and minimise the ratio of non-performing loans (NPL) to significant levels.

6.4 Limitations and Future Work.

Even though the outcomes of the suggested framework were of the state of the art, there are several research directions that can be pursued in the future.

Algorithms Evolution: Future studies should examine utilizing Deep Learning models, such as Tabular Transformers (TabTransformers) and Gated Residual Networks, which can further reduce the difference between these models and the ability to capture temporal correlations in the sequential payment data.

Alternative Data Integration: A more holistic perspective of the thin-file borrower with no traditional credit history would include the inclusion of non-traditional "Alternative Data" such as psychometric testing, utility payment history and e-commerce behaviour.

Dynamic Real-Time Systems: A growing need is the study of Online Learning systems whereby models are transformed in real-time, when new transaction information is received, to facilitate risk adjustments in real-time within a dynamic digital banking environment.

Fairness and Bias Auditing: Future work: The socio-technical nature of AI in finance, i.e., developing automated versions of the Fairness Audit, should also be the focus of future work, which will make sure that as AI becomes more complex, it does not inadvertently

create or increase systemic bias toward marginalised groups.

REFERENCES

[1] S. K. Trivedi, "A study on credit scoring modeling with different feature selection and machine learning approaches," *Technology in Society*, vol. 63, p. 101413, 2020.

[2] X. Dastile, T. Celik, and M. Potsane, "Statistical and machine learning models in credit scoring: A systematic literature survey," *Applied Soft Computing*, vol. 91, p. 106263, 2020.

[3] N. Kozodoi, J. Jacob, and S. Lessmann, "Fairness in credit scoring: Assessment, implementation and profit implications," *European Journal of Operational Research*, vol. 297, no. 3, pp. 1083–1094, 2022.

[4] M. Óskarsdóttir, C. Bravo, C. Sarraute, J. Vanthienen, and B. Baesens, "The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics," *Applied Soft Computing*, vol. 74, pp. 26–39, 2019.

[5] R. Sharma and P. Panigrahi, "Feature engineering for credit risk assessment using machine learning," *Journal of Banking and Financial Technology*, vol. 6, no. 1, pp. 45–58, 2022.

[6] M. Tsiakmaki, G. Kostopoulos, S. Kotsiantis, and O. Ragos, "Implementing AutoML in educational data mining for prediction tasks," *Applied Sciences*, vol. 10, no. 1, p. 90, 2020.

[7] M. Brown, "Influence of artificial intelligence on credit risk assessment in banking sector," *International Journal of Modern Risk Management*, vol. 2, no. 1, pp. 24–33, 2024.

[8] O. A. Bello, "Machine learning algorithms for credit risk assessment: An economic and financial analysis," *International Journal of Management Technology*, vol. 10, no. 1, pp. 109–133, 2023.

[9] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[10] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[11] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, pp. 1135–1144, 2016.

[12] C. Molnar, *Interpretable Machine Learning*. Lulu.com, 2020.

[13] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, pp. 785–794, 2016.

[14] G. Ke et al., "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[15] A. Bequé and S. Lessmann, "Extreme learning machines for credit scoring: An empirical evaluation," *Expert Systems with Applications*, vol. 86, pp. 42–53, 2017.

[16] S. Lessmann, B. Baesens, H. V. Seow, and L. C. Thomas, "Benchmarking state-of-the-art classification algorithms for credit scoring: A ten-year update," *European Journal of Operational Research*, vol. 247, no. 1, pp. 124–136, 2015.

[17] G. Lemaître, F. Nogueira, and C. K. Aridas, "Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning," *Journal of Machine Learning Research*, vol. 18, pp. 1–5, 2017.

[18] H. Zhao, "A multi-objective genetic programming approach to developing Pareto optimal decision trees," *Decision Support Systems*, vol. 43, no. 3, pp. 809–826, 2007.

[19] S. A. A. Shiam et al., "Credit risk prediction using explainable AI," *Journal of Business and Management Studies*, vol. 6, no. 2, pp. 61–70, 2024.

[20] N. Bussmann, P. Giudici, D. Marinelli, and J. Papenbrock, "Explainable machine learning in credit risk management," *Computational Economics*, vol. 57, pp. 203–216, 2021.

[21] P. E. de Lange, B. Melsom, C. B. Vennerød, and S. Westgaard, "Explainable AI for credit assessment in banks," *Journal of Risk and Financial Management*, vol. 15, no. 12, p. 556, 2022.

[22] B. Hadji Misheva, A. Hirsa, J. Osterrieder, O. Kulkarni, and S. F. Lin, "Explainable AI in credit

risk management," arXiv preprint arXiv:2103.00949, 2021.

[23] M. Wang, X. Zhang, Y. Yang, and J. Wang, "Explainable machine learning in risk management: Balancing accuracy and interpretability," *Journal of Financial Risk Management*, vol. 14, pp. 185–198, 2025.

[24] T. E. Edunjobi and O. A. Odejide, "Theoretical frameworks in AI for credit risk assessment: Towards banking efficiency and accuracy," *International Journal of Scientific Research Updates*, vol. 7, no. 1, pp. 92–102, 2024.

[25] L. M. Demajo, V. Vella, and A. Dingli, "Explainable AI for interpretable credit scoring," *Computer Science & Information Technology*, pp. 185–203, 2020.

[26] N. O. Collins and I. Emmanuel, *Machine Learning for Credit Risk: Predicting Loan Defaults in Financial Institutions*, 2023.

[27] A. N. Eshan et al., "Credit risk prediction with self-supervised learning: An explainable AI approach integrating SHAP and LIME," Research Square, 2025.

[28] M. Goel et al., "Using AI for predictive analytics in financial management," 2024.

[29] S. Shreya and H. Pathak, "Explainable artificial intelligence credit risk assessment using machine learning," arXiv, 2025.