Explainable AI (XAI) Using SHAP and LIME for Financial Fraud Detection and Credit Scoring

Sophia John Chavakula
Student Member, IEEE
Dept of Computer Engineering
Fr. C Rodrigues Institute of Technology
Vashi, India
sophiachavakula29@gmail.com

Christopher Aseer J Albert
Student Member, IEEE
Dept of Computer Engineering
Fr. C Rodrigues Institute of Technology
Vashi, India
christpheraseer5080@gmail.com

Earnest Ebenezer
Dept of Computer Engineering
Fr. C Rodrigues Institute of Technology
Vashi, India
earnestebenezer777@gmail.com

Mustansir Habil Bhagat
Dept of Computer Engineering
Fr. C Rodrigues Institute of Technology
Vashi, India
mustansirbhagat@gmail.com

Chaitanya Vijaykumar Mahamuni
Dept of Computer Engineering
Fr. C Rodrigues Institute of Technology
Vashi, India
chaitanyamahamuni91@gmail.com

2025 International Conference on Advanced Computing Technologies (ICoACT) | 979-8-3315-0985-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICOACT63339.2025.11005238

Abstract—**Financial fraud detection and credit scoring are two important applications in the financial domain on which high accuracy and interpretability are required. While Random Forests and XGBoost algorithms of the third generation produce good prediction quality, there is no way to explain the prediction results and they do not meet transparency criteria of regulators. The work in this paper focuses on the incorporation of Explainable AI (XAI) methods such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) to combat these issues of concern under trust, compliance and accountability within these models. Decision Trees were employed for the detection of fraud while Random Forest was employed for credit scoring, SHAP was used for global feature importance and LIME for instance explanations. The model for fraud detection had accuracy of 95% and SHAP found the characteristics of transactions such as amount and frequency significant for fraud detection, While, the credit scoring model that had 76% accuracy and with the help of LIME, the debt ratios and payment history of the credit contenders was found important. The integration of SHAP and LIME improves the model interpretability and fairness, and makes the stakeholders and regulating authorities to trust AI solutions that will be used in the future of financial companies.**

Index Terms—**Financial Fraud Detection, Credit Risk Assessment, SHAP Explanation, LIME Interpretability, Machine Learning in Finance, Transparency in Artifical Intelligence, XAI Regulatory Compliance, Decision Trees and Random Forests, Ethical AI Deployment, Feature Importance Analysis**

I. Introduction

Financials plays an important role in the world's economy by relying on accurate and efficient fraud detection and credit scoring systems. These systems maintain stability, trustworthiness, and fairness while avoiding fraudulent activities against financial institutions, as well as offering opportunities for good creditworthiness to the individuals. The fraud detection system detects the anomalies of transactions and saves it from major loss. The credit scoring model measures an individual's creditworthiness for the proper and fair lending decisions. Both applications are vital to maintaining operational integrity and building trust in financial markets. These models, such as Random Forest, XGBoost, and Neural Networks, are strong on predictive accuracy but really struggle because of their complexity and non-linearity. They are ”black-box” models that focus on delivering high predictive power, making them hard to interpret and understand-a critical issue in the financial sector, which requires transparency in regulatory compliance. Decisions such as credit approvals and fraud detection require explanations. Additionally, biases in historical data—such as under-representation of specific demographic groups—and limited data quality can further impact the fairness and accuracy of these models. As fraud tactics evolve over time, regular model updates are necessary to ensure that these systems remain effective in detecting fraud and making fair lending decisions.

A recent survey by PwC shows that about 59% of firms in India fell prey to financial or economic frauds in the last 24 months and this has gone up by 7% compared to 2022 and it is well above the global average of 41%[1]. This is a clear indication that transparency and interpretability is set to rise in the future as AI continues to seep its way into the financial industry. Models like Random Forest, XGBoost, and Neural Networks excel in tasks such as fraud detection and credit scoring, but their complexity and opaque decision-making pose challenges, especially in regulated environments. Financial institutions, regulators, and customers must trust that AI-driven decisions are fair, accurate, and legal. Without transparency, the adoption of these models could be hindered, limiting their potential to enhance financial services. Explaining this complex models is achieved through techniques known as Explainable AI (XAI). XAI not only means that predictions are understandable for stakeholders but also increases trust in the systems that are making critical decisions. This is particularly important in financial services, where loan approvals or fraud

979-8-3315-0985-9/25/$31.00 ©2025 IEEE

detection directly impact individuals and businesses. XAI techniques help meet these requirements by offering transparency into the decision-making process, ensuring that decisions are both explainable and in compliance with regulatory standards. With improved interpretability, XAI builds confidence among users and regulators, making AI systems trustworthy tools for advancing fairness, accountability, and transparency in financial decision-making.

A. Research Objectives and How to Address Them:

1) Use SHAP and LIME in fraud detection based on a decision tree. Applying Random Forest credit scoring models will help increase interpretability as accuracy (all levels should have at least 95% for fraud detection, at least 75% for credit scoring).
2) Assess the influence of SHAP and LIME on trust, compliance and stakeholder. with the help of quantitative interpretability metrics for understanding.

This research advances the use of Explainable AI (XAI) in financial fraud detection and credit scoring by proposing a structured approach to integrating XAI methods into machine learning models. It demonstrates how SHAP and LIME enhance model interpretability, enabling financial institutions to understand decisions, build trust, and meet regulatory requirements. Additionally, the study highlights how explainability positively impacts trust, fairness, and compliance by balancing model interpretability with prediction quality, thereby promoting fair and transparent financial decision-making.

II. Literature Review

The paper titled 'A Comprehensive Study of Data Mining-based Financial Fraud Detection Research' examines how many data mining techniques like Logistic Regression, Artificial Neural Networks (ANN), Decision Trees, and Bayesian Belief Networks can be employed in detecting financial fraud. The strengths of these methods include high accuracy, fast results, and ability to adjust for multiple types of fraud. However, some of the issues discussed include over fitness or over learning and the model dependency on large data sets and with consideration needs high computation power. Data preprocessing and evaluation are presented as crucial steps for successful implementation in the given paper. Future works can be based on deep learning, better time-detection, involvement of researchers and auditors. The paper also points out some research opportunities that are still open and have not been discussed; the effect of price sensitivity on the detection models, need for multi-disciplinary methods, and the scarcity of standard public datasets.[2]

The paper titled 'Implementation of Gradient Boosted Tree, Support Vector Machinery and Random Forest Algorithm to Detecting Financial Fraud in Credit Card Transactions' explores the implementation and comparison of three machine learning algorithms that can be used to identify fraudulent credit card transactions: Random Forest (RF), Support Vector Machines (SVM), and Gradient Boosted Trees (GBT). Based on the data from the 2018 FINHACKS competition, the work is devoted to the problem of class imbalance with the help of the SMOTE and RUS sampling techniques. The perfomance of the above algorithms is assessed by such parameters as AUC Score, Fraud Catching Rate, False Alarm rate, Matthew's Correlation Coefficient (MCC). The results demonstrated that the RF algorithm with RUS technique from the group provided the best all-round performance: AUC = 0.7742, and the fraud-catching rate was 75%. Although the study establishes that RUS is efficient in handling imbalanced data while RF is malleable, it lacks hyperparameter optimization and utilizes datasets and real-world validation using non-synthetic datasets. The future work also aims at a more detailed investigation of different sampling techniques, improvement of algorithms and their interconnection with cloud solutions for further scaling. [3]

The paper 'A Financial Statement Fraud Detection Model Based on Hybrid Data Mining Methods' proposes a hybrid model for classifying financial statement fraud using feature selection and machine learning algorithms. To select what variables are most important, the authors chose a dataset that includes 240 companies (120 are fraudulent, and 120 are nonfraudulent) that are from 2007 to 2016, and they applied Principal Component Analysis (PCA) and feature selection through XGBoost. Five classifiers were used, of which the categories are Random Forest (RF), Support Vector Machine (SVM), Decision Tree (DT), Logistic Regression (LR), and Artificial Neural Network (ANN). The experiments proved that Random Forest has better results than other methods by providing better accuracy, stability and robustness especially when selecting 2 or 5 key parameters. However, this is only applicable to the Chinese firms and does not present the model in its dynamic application to real life fraud detection scenarios. Possible future work areas include expanding data for big nations, moving to real time detection, and incorporating the system to current financial surveillance system to increase versatility and applicability. However, the current study has certain limitations: While the hybrid approach together with the effective feature selection, and the focus on Random Forests' strengths offer insights into the effectiveness of financial fraud detection.[4]

The paper 'AI in credit scoring: A comprehensive review of models and predictive analytics' aims at capturing the historical shift in the credit scoring field where rule-based systems have been substituted with Machine Learning (ML) and Deep Learning (DL) such as XGBoost and Neural Network. It argues that origination data besides transaction histories and data from social media are useful for predicting better credit scoring. Nevertheless, the paper does not contain a comprehensive overview of model performance: real performance indicators and application results are omitted, and certain differences between models are left unexplored. It also lacks coverage of some important preprocessing activities, optimisation methods, and more subtle methods for dealing with the biases, interpretability issues and data privacy of artificial intelligence models. While the paper does include several effective ethical and regulatory questions, it does not contain solutions for them, such as FAIR or PP techniques. Future trends are discussed and highlighted, however, the combined use of these technologies with AI models is not considered. It would be useful to add the specific examples into the paper, especially, describing the ways forward and indicating how, for example, credit scoring may evolve in the future with the help of some of the presented technologies. Areas that still need further investigation include optimization of AI models, adversarial training, and frameworks enhancing secure distributed credit scoring.[5]

The work in the paper 'Explainable AI for Credit Assessment in Banks' contrasts credit scoring models based on LightGBM and Logistic Regression (LR), showing that LightGBM has better predictive performance compared with logistic regression yet remains explainable by SHAP coefficient. The monthly customer application data when combined with daily account data also improves the ability to predict loan defaults and shows the value of combining distinct types of data. Advantages are enhanced efficiency of functioning, improved interpretability with the help of SHAP, and compliance with current regulations using such visualizations as waterfall and dependence. Possible future work directions include comparing other models, such as XGBoost and Neural Networks, use more precise calibration techniques to increase the model reliability, using temporal data to improve the risk prediction, and analyse fairness aspects. More about interpretability and feature engineering may reveal further important insights and the generalisation of results.[6]

This paper 'Explainable Artificial Intelligence (XAI) in Finance: A Systematic Literature Review' analyses 138 papers from 2005-2022 on Explainable AI (XAI) in financial related applications including credit scoring, fraud detection, risk management, and portfolio optimization. In this regard, XAI tools such as SHAP, LIME as well as the rule based approaches elaborated in the paper will help improving model interpretability and compliance with regulatory requirements like GDPR as well as gaining trust from stakeholders. But, the problems are the decrease of generalization and the increase of ininterpretability, high computational requirement of XAI, and dataset biases. Subsequent studies should investigate blended models, establish general set of criteria for comparison, as well as further investigate XAI in novel domains such as in blockchain. The paper also notes deficiencies in the harmonisation of the evaluation criteria, interdisciplinary cooperation, and access to the public databases. Furthermore, solutions to these challenges could inform ways of enhanced and more scalable XAI solutions. These factors are some of the biggest obstacles to the development of the application of the use of AI in the financial field.[7]

This paper 'Explainable AI in Financial Technologies: Balancing Innovation with Regulatory Compliance' XAI is described as an important capability to increase trust and meet the requirements of regulations such as GDPR again thanks to the ability to explain the decision made by AI models. XAI benefits discussed by the author include promoting the accountability and fairness of the AI-driven consumer decisions through implementing SHAP and LIME. However, some limitations include the dynamics in the regulatory environment (for instance, EU's AI Act) and high operating costs mean to putting down XAI frameworks. Moreover, preventing biases in AI systems still present a problem all together. Further research is recommended in the application of XAI and the combination of the technology with blockchain for increased transparency in decentralized networks and the improvement of existing XAI frameworks for compliance with the standards and norms of today's highly regulated industries, including finance and healthcare. Several research gaps are proposed in the paper: the lack of audit tools for ongoing monitoring; the lack of standardized measures for explainability; and least attention to the interdisciplinarity of adopting XAI. To address these gaps will be vital hence propelling the responsible and ethical utilization of AI in financial technology.[8]

The paper 'Explainable AI (XAI):'Core Ideas, Techniques, and Solutions' reveals seven methodologies that are crucial in the domain of Explainable AI (XAI). Integrated Gradients gives the relevance of every input feature based on the gradients of the model, which provides clear understanding in terms of individual predictions. The Contrastive Explanations Method (CEM) provides information on black box models and provides the Pertinent Positives and Pertinent Negatives of a model. Technique such as exploratory data analysis and basic visualization remain relevant in the extraction of meaningful patterns in a dataset. Counterfactual explanations give a perspective into which changes to the features would lead to a different outcome especially useful in binary cases. Model agnostic approaches which include Anchors reveals specific conditions governing the predictions of models in form of if-then rules on different data types. The model-agnostic approach to interpreting models, which we have described in the case of LIME, can be applied to any machine learning model. Finally, distributed learning frameworks are used to spread the training process of models on many computational nodes to increase the scale of efficiency. Combined, these aides help to explain and rationalise the often complex decision making ability of AI.[9]

III. Methodology

A. Problem Domain and Data Collection

To address fraud detection and credit scoring, two primary data sources were utilized: transactional data for fraud detection and personal financial data for credit scoring.

Transactional Data: Includes attributes like transaction amounts, frequency, geolocation, and time of transactions. This data helps in identifying anomalous patterns indicative of fraudulent activities.

Personal and Financial Data: Covers user demographics, financial history, debt ratios, credit history, and payment patterns. These features are essential for determining creditworthiness and risk scoring. The datasets were selected to ensure diversity and relevance, with features such as:

Transaction Amounts: Variations in spending behavior.

User Behavior: Patterns in transaction frequency, geolocation deviations.

Financial History: Historical data on loans, defaults, and repayment schedules.

Debt Ratios: Proportions of debt to income and overall credit utilization.

B. Data Preprocessing

Preprocessing steps were meticulously designed to enhance model performance and ensure data integrity:

Handling Missing Values: Imputation strategies such as mean or median replacement for numerical features. Mode imputation for categorical variables.

Feature Scaling: Standardization of numerical features to ensure uniformity across dimensions. Normalization for features with skewed distributions, such as transaction amounts. Label encoding for ordinal categories where a natural order exists.

Feature Engineering: Creation of derived features like transaction frequency, average transaction value, and credit utilization ratios. Aggregated features such as rolling averages over different time windows for enhanced temporal insights SMOTE synthesizes new samples for minority on the basis of border points in order to enhance the performance of a model, which is often applied in fraud and diagnosis. It may develop unrepresentative samples and is laborious when working with a huge data set.

C. Machine Learning Models

The models used for fraud detection and credit scoring were chosen for their interpretability, robustness, and performance on structured data.

- Decision Trees: Useful for capturing nonlinear relationships and producing interpretable decision rules.
- Random Forest: An ensemble approach offering improved accuracy and resilience to overfitting by aggregating multiple decision trees.
- XGBoost: A gradient-boosted decision tree algorithm, highly efficient in handling large datasets and producing state-of-the-art performance.

Justification:

- Interpretability: Essential in financial domains for regulatory compliance.
- Performance: High accuracy and recall metrics, crucial for minimizing false positives and negatives.
- Suitability: These models are adept at handling numerical and categorical data, making them well-suited for structured financial datasets.

D. Explainable AI Integration

Transparency in AI models was prioritized by integrating explainability techniques: SHAP (SHapley Additive exPlanations): Used for global explainability, providing a comprehensive view of feature importance across the dataset. This helps stakeholders understand the primary drivers behind predictions for fraud detection and credit scoring. LIME (Local Interpretable Model-agnostic Explanations): Applied for local explainability to explain individual predictions. This is particularly valuable in justifying adverse decisions, such as loan denials, to end-users.

E. Model Training and Evaluation

The models were trained and evaluated following rigorous procedures:

Training Procedures:Data was split into training, validation, and test sets. Cross-validation was employed to fine-tune hyperparameters and prevent overfitting.

Evaluation Metrics for Fraud Detection and Credit Scoring:

Accuracy: Used to estimate the level of accuracy in fraud detection as well as credit scoring.

Precision: Checks the model's performance in accurately sorting out true positives, defrauding transactions or low risk personnel.

Recall: Describes the ability of the model in identifying all or most relevant instances, such as fraudulent transactions or credit worthy clients.

F1-Score: Used in cases where precision and recall have been estimated giving an average of the two in a bid to produce a more accurate result.

ROC-AUC: Compares the true positive rate with the false positive rate, showing discrimination capability of the model both in true positive and false positive conditions.

IV. Proposed System

Today's challenges of the financial sphere require accuracy and transparency of decisions: fraud prevention and fair credit scoring. The idea of the system resolves issues related to financial risk since it offers clear and economical fraud detection and credit management. It uses personal and transactional information to label the transactions as 'non fraudulent' or 'fraudulent'-represented as 0 and 1 respectively -as well as to rate the creditworthiness of a client. By employing Decision Tree model for checking for frauds and Random Forest for credit risk assessment, the system analyzes several functions such as amount transferred, credit profile, income, and payment behaviour. Currently, SHAP adds global feature importance, while LIME considers local instance-based explanations of the given predictions. Outcomes are presented in an automatically generated dynamic table with SHAP plot summaries and LIME interpretation for full disclosure of the analysis. With compliance considerations in mind, the system is accurate, easy to interpret and built for scaling to incorporate real-time fraud detection and credit scoring.

A. System Architecture

The architecture proposed in this paper has been designed with focus on high performance, explainability and ability to scale for fraud detection as for credit scoring. The data is obtained from personal, financial, and transactional records and is streamed through the preprocessing layer where data is cleaned, normalized and encoded to convert categorical data. The feature extraction is also aimed at accuracy of models in the process of extracting them. Decision trees are used in the fraud detection while Random Forest models are used in credit scoring with cross validation to check on the models' accuracy.

Figure 1 shows the architecture of the proposed system deployed with explainability into account with the utilization of SHAP for global feature importance and LIME for feature specific analysis by instance importance. The model is evaluated for aspects like Accuracy, Precision, Recall, F1-Score and so on which is presented in a Dashboard view along with regulatory check for compliance. These outputs are presented in a more human-understandable form to maintain transparency for financial institutions along with compliance reports for the authorities so that a fair and trusting environment can be created.

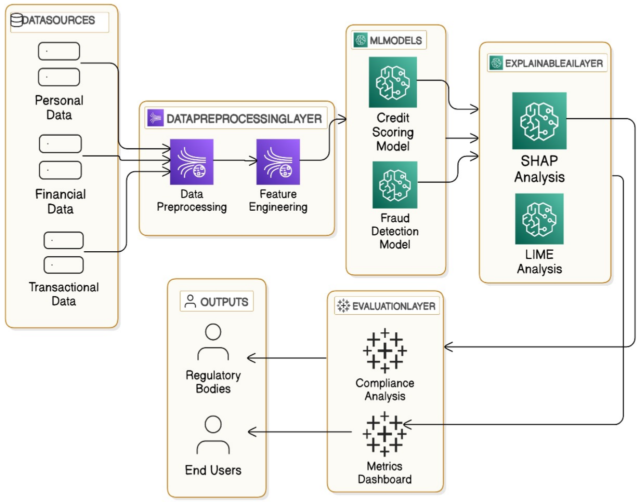

Fig. 1. System Architecture

V. Implementation

The work of the project involves training a machine learning model each for both financial fraud detection and credit scoring using Python libraries and especially with the XAI techniques implemented. **Pandas**: is used to handle datasets and Common data preprocessing techniques on the dataset such as missing values imputation, encoding of categorical features and feature extraction. **Scikit-learn** is used to fit **Decision Tree** for handling the fraud detection, which was selected for its interpretability, and capability in regards to handling non-linearity in financial variables. To achieve credit scoring, **Random Forest** is used and this approach is effective when it comes to ensemble learning. The case of **SHAP** guarantees an accurate interpretation of how the model works through the display of features. On the other hand, **LIME** allows feature interpretations at a local level as it approximates the features under a larger local model framework. Besides, SHAP's visualization of the summary and dependence plots and the instance-level explanation of the model employed, LIME guarantees full explainability. This is especially the case in fraud detection and credit scoring, which require model explanations to provide compliance, accountability, and mitigation of harms that might affect customers with low credit scores.

A. Model Training and Evaluation

The proposed model was trained utilizing **Decision Tree (DT)** and **Random Forest (RF)** techniques, As both the techniques are quite efficient in different way. The Decision Tree algorithm was selected to achieve a greater interpretability if compared to other algorithms and as it can handle non-linear transaction data that can occur in decision making processes. However, the Random Forest algorithm was used due to its efficiency for dealing with the interactions of different features, as well as for minimizing overfitting through bagging decision trees.

Cross-validation training, regularization was performed where the efficacy of both models on different partitions of data involved in re-architecting through cross-validation

TABLE I
Classification Report for Fraud Detection

	Precision	**Recall**	**F1-score**	**Support**
0	0.93	0.96	0.95	56750
1	0.96	0.93	0.95	56976
Accuracy			0.95	113726
Macro avg	0.95	0.95	0.95	113726
Weighted avg	0.95	0.95	0.95	113726

*1) **Fraud Detection Model:*** The model was trained on a balanced dataset consisting of 284,315 samples in each class: As binary classes: Class 0 (non-fraudulent transactions) and Class 1 (fraudulent transactions). The DT algorithm was used to separate the data using features where splits were based on Gini impurity or entropy and aimed at minimizing classification error. Other hyperparameters that include maximum depth and minimum samples per leaf were adjusted to get the best split measurements to avoid overfitting.

Thus, the assessment of the trained model as shown in Table I proves the high efficiency of the proposed model for fraud detection. It is noteworthy to mention that the classification report demonstrates high precision (0.93 for class 0 and 0.97 for class 1) and recall (0.96 for class 0 and 0.93 for class 1), which amounts to an accuracy of 95%. These measures express the model's high predictive capability in correctly identifying credit card transactions belonging to the fraud and non-fraud classes. High precision and recall indicate the model's effectiveness in identifying fraudulent transactions.

TABLE II
Confusion Matrix for Decision Tree

Predicted	0	1
Actual 0	54576	2174
Actual 1	3922	53054

Table II confusion matrix for Decision Tree model highlights 54,576 true positives (fraudulent transactions correctly classified) and 53,054 true negatives (non-fraudulent transactions correctly classified), with relatively low misclassification rates:

Measuring the performance of this approach, we have identify 3,922 instances of false positives (non-fraudulent transaction that is classified as a fraud) and 2,174 instances of false negative (fraudulent transaction that is categorized as non-fraudulent). Declared these outcomes substantiated the validity plus interpretability of the model in differentiating fraudulent deals suitably.

TABLE III
CLASSIFICATION REPORT FOR CREDIT SCORING

Class	Precision	Recall	F1-Score	Support
0	0.62	0.78	0.69	1736
1	0.76	0.80	0.78	3351
2	0.83	0.73	0.77	5523
Accuracy			0.76	10610
Macro avg	0.73	0.77	0.75	10610
Weighted avg	0.77	0.76	0.76	10610

*2) **Credit Scoring Model:*** The model was trained on a dataset consisting of 10,610 samples across three classes: Class 0 for low credit score, Class 1 for medium credit score, and Class 2 for high credit score. The classification was performed using the Random Forest (RF) algorithm; it works together with the collection of decision trees to raise the rate of classification. These splits were done based on features and the Gini impurity was used in order to reduce the classification error on the available data which was used to train the model. To overcome this problem, hyperparameters including the number of trees in the forest, maximum depth etc., have been tuned to ensure that the model should not overfit to the training data set and the new data set which will be used in future.

The analysis of the trained model as shown in Table III proves that the proposed system provides high accuracy rate in the credit scoring classification task. The results of the classification report include precision of 0.62 to 0.83, recall of 0.73 to 0.80, and F1-score of 0.75 macro average. The total accuracy of the model analyzed is at a satisfactory record of 76%, that is, an impressive metric in terms of totality for the right credit scores from the given data set. These measures reveal the degree of accuracy with which the specified credit score categories have been classified.

TABLE IV
CONFUSION MATRIX FOR CREDIT SCORING

Predicted	0	1	2
Actual 0	1359	20	357
Actual 1	166	2693	492
Actual 2	671	835	4017

Table IV confusion matrix for Random Forest model reveals the following results: The results are 1359 true positives, 837 false positives and 377 false negatives for Class 0, 855 false positives, 658 false negatives and 2693 true positives for Class 1, and 849 false positives, 1506 false positives and 4017 true positives for Class 2. From this it can be seen that the proposed model is effective in discrimating most of the credit score classifications, although there is higher misclassification in Class 2, in terms of both false positives and false negatives. Nevertheless, such a degree of misclassification suggests the model's interpretability and efficiency for credit scoring as a whole.

B. Applications of SHAP and LIME

For increased interpretability of the proposed system, the XAI methods, namely SHAP and LIME, are used.These methods enable model interpretability to make the decision reasonable and justify it to the stakeholders concerned.

SHAP (SHapley Additive exPlanations) is used to analyze the global behavior of the model by calculating the impact of each feature on the predictions. SHAP values are derived from the Shapley value concept in cooperative game theory, where the SHAP value ϕ_i for feature i is defined as:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!} \Big[f(S \cup \{i\}) - f(S) \Big]$$

Where:

- N is the set of all features.
- S is a subset of features excluding feature i.
- $f(S)$ is the model output using only the features in S.
- $f(S \cup \{i\})$ is the output of the model when the feature i is added to subset S.
- $|N|!$ is the factorial of the total number of features, acting as a normalization factor.

This allows the system to determine the relative importance of features like transaction amount or credit history in the predictions.

LIME (Local Interpretable Model-agnostic Explanations), on the other hand, provides local explanations for individual predictions by approximating the complex model $f(x)$ with a simpler, interpretable surrogate model $g(z')$. The objective is to minimize the following loss:

$$\min_{g \in G} \sum_{z' \in Z} \pi_x(z) \left(f(z) - g(z') \right)^2 + \Omega(g)$$

Where:

- $g(z')$ is the interpretable model (e.g., linear regression) trained on perturbed samples.
- $\pi_x(z)$ is a proximity kernel that weighs the samples based on their similarity to the instance x.
- $\Omega(g)$ is a penalty for model complexity to ensure interpretability.

This allows the system to explain why a specific transaction was flagged as fraudulent or why a credit application was denied.

To address the computational challenges of SHAP and LIME, the following optimization strategies were employed: Regarding SHAP , the high computational complexity when tested on large datasets was handled by considering the top 10 features only starting from which importance decreases significantly. This reduction in the number of features made the calculations manageable while at the same time making the results easily interpretable. This reduced SHAP's cubic time complexity in large data sets, and made it more efficient while not sacrificing meaningful global explanations.Similarly, for LIME, the instance selection task was performed with the same goal to choose instances of data sample. A departure from approximating the entire dataset with a simpler model was done using LIME on a subset of instances that best represented the underlying data patterns. This approach proved useful in eliminating a large part of the complexity of the model, as well as the computational cost associated with the derivation of the said explanations, which held parsimony and local interpretability responsible for providing relevant and valuable explanations for the specific predictions made by the model.

These were necessary to augment scalability of SHAP and LIME while at the same time keeping the model both explainable and efficient.

VI. Results and Discussions

In financial fraud detection and credit scoring, performance is critical to ensure accurate, reliable, and fair decisions. Fraud detection models must strike a balance between identifying fraudulent activities with high sensitivity and minimizing false positives to avoid unnecessary disruptions to legitimate transactions. Similarly, credit scoring systems need to be precise in assessing creditworthiness to reduce the risk of defaults while ensuring fair access for applicants. These decisions carry significant consequences, not only financially but also in terms of customer trust and regulatory compliance. Thus, robust evaluation metrics and explainable AI (XAI) are essential components of these systems.

In this section, we evaluate the fraud detection and credit scoring models, providing insights through explainability techniques. We use SHAP (SHapley Additive exPlanations) for the Decision Tree (DT) model in fraud detection and LIME (Local Interpretable Model-Agnostic Explanations) for the Random Forest (RF) model in credit scoring. These methods demonstrate the models' effectiveness and their ability to provide understandable explanations for predictions, which can foster trust in financial decisions.

A. Explainability Analysis for Fraud Detection Model

The Decision Tree model achieved an impressive accuracy of 95%. Using SHAP, we can visualize how features contribute to classifying data into two categories: Class 0 and Class 1, as illustrated in Figure 2. The plot shows the degree of feature importance and their influence on predictions.

For example, features generating positive contributions toward "Class 1" are displayed on the right, while those with negative contributions are on the left. The length of each bar indicates the magnitude of a feature's impact.

In this case, the feature "V14" has a significant negative impact (-0.87 or less), favoring "Class 1," while "V8" has a positive contribution (above 0.05) toward "Class 0." This visualization highlights how individual features drive the model's decisions, providing critical insights for interpretability.

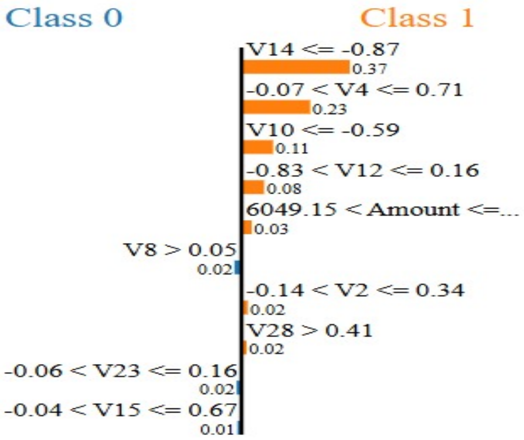

Fig. 2. SHAP Visualization for Fraud Detection

B. Explainability Analysis for Credit Scoring Model

The LIME explanation for the Random Forest model provides an intuitive breakdown of how specific features influence predictions for credit scoring (Figure 3).

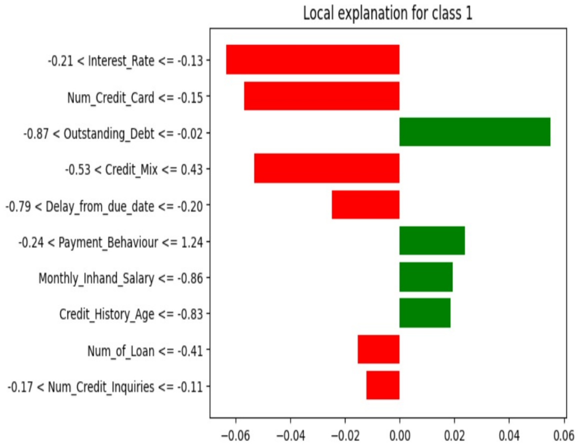

Fig. 3. Local Explanation for LIME Classes

In this visualization, green bars on the right represent features that increase the probability of the instance being classified as Class 1, while red bars on the left show features that decrease this probability. The length of each bar reflects the strength of the feature's influence.

For instance, features like a strong repayment history or low credit utilization appear as longer green bars, signaling favorable credit traits. Conversely, missed payments or high debt levels show up as longer red bars, highlighting risk factors. This level of interpretability ensures transparency in credit scoring, allowing financial institutions to identify and address areas where models may require improvements to align with ethical and regulatory standards.

Additionally, Figure 4 reveals how specific features contribute to the model's decision-making process across multiple classes. For example, Class 2 shows the highest predicted probability (86%), followed by Class 1 (10%) and Class

0 (4%). Positive contributors (e.g., ”Outstanding Debt” and ”Payment Behavior”) drive the model toward Class 2, while negative contributors (e.g., ”Interest Rate” and ”Credit Mix”) detract from this classification. This breakdown highlights how the model integrates various factors to arrive at its final prediction.

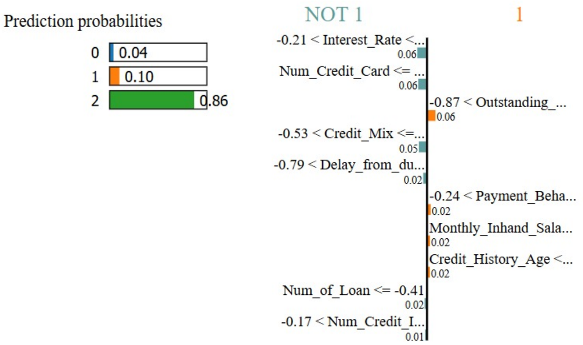

Fig. 4. LIME Visualization for Credit Scoring

C. Trust, Regulatory Compliance, and Fairness

Incorporating XAI techniques like SHAP and LIME into financial systems ensures that the model predictions are transparent and easy to understand. This transparency can improve trust among stakeholders by explaining decisions, such as why a credit application was approved or flagged for fraud detection.

XAI also plays a pivotal role in regulatory compliance, such as adhering to the GDPR and Fair Lending Laws. By making model decisions interpretable, it reduces biases and ensures fairness. For example, SHAP and LIME can uncover hidden prejudices in models, such as those linked to demographic characteristics.

That's why as XAI techniques develop, its application in financial systems will be even more significant. Real-time decision explaining could bring added value by giving customers instant and simple explanation of the decision made in the financial sphere.To achieve this, organizations involved in delivering of financial solutions need to integrate XAI within their operations so as to develop legally compliant, reliable, and responsible models that will create foundation to improved AI solutions in the financial sector.

Conclusion

This paper shows how these two techniques, SHAP and LIME, could be used to improve the quality and interpretability of machine-learning models for fraud detection and credit scoring. From the models developed by decision trees for fraud detection and random forests for credit scoring, the following performance metrics were realized; Accuracy, precision, recall, and f1 score. The incorporation of SHAP and LIME made it easier to interpret the results and revealed the key features that influenced the decision making of the model. These techniques are especially relevant in the financial industry where accuracy and honesty with customers and shareholders are needed, to conform to GDPR and Fair lending laws, and to avoid biases when using AI. Subsequent studies should focus on examining the application of these methods to actual-time systems in relation to dynamism in fraud and credit patterns. Furthermore, research on the combined methods of XAI and the integration of new methodologies such as attention mechanisms could increase scalability and model performance. Possible future additions include using blockchain to improve the accuracy of records and increasing the robustness of the model against biases at the time of model design. The following limitations: aspect of the dataset, computational aspect, and comparing the simpler DT and RF to more complex deep learning models will take the field even further. Such endeavors can therefore help in creating scalable, transparent and trustworthy frameworks and solutions to the application of AI in finance.

Acknowledgment

The work is partially funded by Agnel Charities Fr C Rodrigues Institute of Technology, Vashi. We also thank our mentor, Mr. Chaitanya Mahamuni Sir, for his valuable guidance.

References

[1] Economic Times, ”59% of Indian companies suffered financial fraud in 24 months,” The Economic Times, [Online] Available: https://economictimes.indiatimes.com/news/company/corporate-trends/59-of-indian-companies-suffered-financial-fraud-in-24-months/articleshow/116435203.cms?from=mdr

[2] A. Jain and S. Shinde, ”A Comprehensive Study of Data Mining-based Financial Fraud Detection Research,” 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), Bombay, India, 2019, pp. 1-4, doi: 10.1109/I2CT45611.2019.9033767.

[3] F. Salomo Leuwol, A. Ady Bakri, M. N. Bailusy, H. Setia Putra, and N. K. Sukanti, “Implementation of Gradient Boosted Tree, Support Vector Machinery and Random Forest Algorithm to Detecting Financial Fraud in Credit Card Transactions”, jidt, vol. 5, no. 3, pp. 26-30, Aug. 2023.

[4] J. Yao, J. Zhang and L. Wang, ”A financial statement fraud detection model based on hybrid data mining methods,” 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 2018, pp. 57-61, doi: 10.1109/ICAIBD.2018.8396167.

[5] W. Addy, A. Ajayi-Nifise, B. Bello, S. T. Tula, O. Odeyemi, and T. Falaiye, ”AI in credit scoring: A comprehensive review of models and predictive analytics,” Global Journal of Emerging Technologies and Applications, vol. 18, no. 2, pp. 118-129, 2024, doi: 10.30574/gjeta.2024.18.2.0029.

[6] 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, doi: 10.3390/jrfm15120556.

[7] J. Černevičienė and A. Kabašinskas, ”Explainable artificial intelligence (XAI) in finance: A systematic literature review,” Artificial Intelligence Review, vol. 57, p. 216, 2024, doi: 10.1007/s10462-024-10854-8.

[8] A. N. Anang, O. Ajewumi, T. Sonubi, K. Nwafor, and I. Akinbi, ”Explainable AI in financial technologies: Balancing innovation with regulatory compliance,” International Journal of Science and Research Archive, vol. 13, pp. 1793-1806, 2024, doi: 10.30574/ijsra.2024.13.1.1870.

[9] Dwivedi, Rudresh, Dave, Devam, Naik, Het, Singhal, Smiti, Rana, Omer , Patel, Pankesh, Qian, Bin, Wen, Zhenyu, Shah, Tejal, Morgan, Graham and Ranjan, Rajiv 2023. Explainable AI (XAI): core ideas, techniques and solutions. ACM Computing Surveys 55 (9) , 835. 10.1145/3561048

[10] Dataset: N. Yewithana, ”Credit Card Fraud Detection Dataset 2023,” Kaggle. [Online]. Available: https://www.kaggle.com/datasets/nelgiriyewithana/credit-card-fraud-detection-dataset-2023

[11] Dataset: R. Paris, ”Credit Score Classification,” Kaggle. [Online]. Available: https://www.kaggle.com/datasets/parisrohan/credit-score-classification