Explainable Financial Fraud Detection Using Lightgbm and Shap Analysis: A Comparative Study with Xgb-Gp Framework

Ms. Arya M. Palve [1], Prof. Jayaprabha M. Kanase [2], Prof. Dr. B. D. Phulpagar[3]

[1] *Pursuing M.Tech in Computer Engineering from Progressive Education Society's Modern College of Engineering, Pune*

[2] *Assistant Professor at Progressive Education Society's Modern College of Engineering, Pune*

[3] *Professor at Progressive Education Society's Modern College of Engineering, Pune*

***Abstract*—Financial fraud detection has become increasingly important as the rapid growth of digital transactions exposes financial institutions to sophisticated and evolving fraudulent activities. Conventional machine learning models often achieve high predictive performance but provide limited interpretability, making them less suitable for high-stakes financial decision-making where regulatory compliance and stakeholder trust are essential. This study proposes an explainable financial fraud detection framework that integrates the Light Gradient Boosting Machine (LightGBM) algorithm with SHapley Additive exPlanations (SHAP) to simultaneously achieve robust classification performance and transparent model interpretation. The proposed framework is comparatively evaluated against an Extreme Gradient Boosting–Genetic Programming (XGB-GP) framework to assess both predictive effectiveness and explainability. The experimental workflow includes data preprocessing, handling class imbalance, feature engineering, hyperparameter optimization, and stratified cross-validation using a publicly available financial transaction dataset. Performance is assessed using accuracy, precision, recall, F1-score, area under the receiver operating characteristic curve (AUC-ROC), and computational efficiency. SHAP analysis is employed to quantify feature contributions, provide global and local explanations, and identify the most influential variables affecting fraud predictions. Comparative analysis demonstrates that the LightGBM-SHAP framework achieves competitive or superior detection capability while significantly improving model transparency and reducing inference time compared with the XGB-GP approach. The explainability component enables analysts to understand individual prediction behavior, thereby supporting regulatory requirements and enhancing confidence in automated fraud detection systems. The proposed framework offers a practical and scalable solution for real-time financial fraud detection by combining high predictive accuracy with interpretable artificial intelligence, making it suitable for deployment in modern financial environments where both performance and transparency are critical.**

***Index Terms*—Financial Fraud Detection, Explainable Artificial Intelligence (XAI), LightGBM, SHAP, XGBoost, Genetic Programming, Machine Learning, Class Imbalance, Financial Analytics.**

I. INTRODUCTION

The rapid expansion of digital banking, electronic commerce, mobile payment platforms, and online financial services has significantly transformed the global financial ecosystem. While these technological advancements have improved transaction speed, accessibility, and customer convenience, they have simultaneously created new opportunities for fraudulent activities. Financial fraud, including credit card fraud, insurance fraud, money laundering, identity theft, and unauthorized digital transactions, has become increasingly sophisticated, resulting in substantial financial losses for institutions and customers worldwide. According to recent industry reports, the annual cost of financial fraud continues to increase due to the growing volume of digital transactions and the evolving strategies employed by cybercriminals. Consequently, developing accurate, scalable, and intelligent fraud detection systems has become a critical priority for financial organizations.

Traditional fraud detection methods primarily rely on manually designed rules, statistical techniques, and expert-defined thresholds. Although these approaches are effective for identifying known fraud patterns, they struggle to detect newly emerging fraud strategies because of their limited adaptability and dependence on predefined rules. Moreover, the highly imbalanced nature of financial transaction datasets, where fraudulent transactions constitute only a small fraction of total records, presents significant challenges for conventional analytical methods. These limitations have encouraged researchers to explore machine learning (ML) and artificial intelligence (AI) techniques capable of automatically learning complex relationships from large-scale financial data.

Among modern machine learning algorithms, ensemble learning techniques have demonstrated remarkable success in fraud detection. Gradient boosting algorithms such as Extreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM) effectively model nonlinear relationships, handle high-dimensional data, and achieve superior predictive performance compared with many traditional classifiers. LightGBM, in particular, employs histogram-based learning, leaf-wise tree growth, and gradient-based one-side sampling to improve computational efficiency while maintaining high prediction accuracy. These characteristics make LightGBM well suited for large-scale financial datasets that require rapid processing and real-time decision-making.

Despite their excellent predictive capabilities, advanced ensemble models are frequently criticized for their "black-box" nature. Financial institutions operate in highly regulated environments where automated decisions must be transparent, explainable, and auditable. Regulatory frameworks increasingly emphasize the need for interpretable AI systems to ensure fairness, accountability, and user trust. Consequently, explainable artificial intelligence (XAI) has emerged as an essential component of modern financial analytics. SHapley Additive exPlanations (SHAP), developed from cooperative game theory, has become one of the most widely adopted model-agnostic explanation techniques because it provides both global and local interpretations of feature contributions. SHAP enables analysts to understand why a transaction has been classified as fraudulent, thereby facilitating regulatory compliance, model validation, and human decision support.

Several existing studies have investigated machine learning-based fraud detection using algorithms such as logistic regression, decision trees, random forests, support vector machines, neural networks, and XGBoost. More recently, hybrid frameworks integrating XGBoost with optimization techniques such as Genetic Programming (XGB-GP) have been proposed to improve predictive performance through feature transformation and evolutionary optimization. While these methods often achieve satisfactory classification accuracy, they generally place greater emphasis on predictive performance than on model interpretability. Furthermore, the computational complexity associated with optimization-based hybrid models may limit their suitability for real-time financial applications.

To address these challenges, this study proposes an explainable financial fraud detection framework that combines LightGBM with SHAP analysis. The proposed framework seeks to achieve a balance between high predictive performance, computational efficiency, and model transparency. In addition, a comparative evaluation is performed against the XGB-GP framework to examine differences in classification performance, computational cost, and interpretability. Standard preprocessing techniques, class imbalance handling, feature engineering, hyperparameter optimization, and comprehensive evaluation metrics—including accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC)—are employed to ensure a rigorous assessment. SHAP-based explanations are further utilized to identify influential features and provide transaction-level interpretability.

The primary contributions of this research are summarized as follows:

1. Development of an explainable fraud detection framework integrating LightGBM with SHAP for transparent financial transaction classification.
2. Comprehensive comparative analysis between the proposed LightGBM-SHAP framework and the XGB-GP framework using multiple performance metrics.

3. Interpretability assessment through global and local SHAP explanations to improve transparency and support regulatory compliance.
4. Evaluation of computational efficiency, demonstrating the suitability of the proposed framework for large-scale and real-time financial fraud detection.
5. Identification of influential fraud indicators, enabling financial analysts to better understand the factors contributing to fraudulent transaction predictions.

II. LITERATURE SURVEY AND COMPARATIVE ANALYSIS

Financial fraud detection has attracted significant attention in recent years due to the rapid growth of digital payment systems, online banking, and electronic commerce. The increasing complexity of fraudulent activities has encouraged researchers to develop intelligent data-driven approaches capable of identifying suspicious transactions with high accuracy while minimizing false alarms. Machine learning (ML), deep learning (DL), and explainable artificial intelligence (XAI) have emerged as dominant research directions because of their ability to model complex transaction patterns and continuously adapt to evolving fraud strategies.

Early fraud detection systems primarily employed statistical models and rule-based techniques, including Logistic Regression (LR), Decision Trees (DT), and Bayesian classifiers. These methods were computationally efficient and easy to interpret but demonstrated limited capability in capturing nonlinear relationships within highly imbalanced financial datasets. As fraud patterns became increasingly dynamic, conventional approaches experienced reduced detection performance and generated higher false-positive rates.

Ensemble learning methods have substantially improved fraud detection performance by combining multiple weak learners into robust predictive models. Random Forest (RF), AdaBoost, Gradient Boosting Decision Trees (GBDT), Extreme Gradient Boosting (XGBoost), CatBoost, and Light Gradient Boosting Machine (LightGBM) have demonstrated superior classification accuracy compared with traditional classifiers. Among these techniques, XGBoost gained widespread acceptance because of its regularization mechanisms, effective handling of missing values, and ability to model nonlinear relationships. However, XGBoost often requires considerable computational resources when processing large-scale financial transaction datasets.

LightGBM was introduced to overcome these computational limitations through histogram-based learning, leaf-wise tree growth, Exclusive Feature Bundling (EFB), and Gradient-based One-Side Sampling (GOSS). These innovations significantly reduce training time and memory consumption while maintaining competitive predictive performance. Several studies have reported that LightGBM achieves faster convergence than conventional boosting algorithms without compromising detection accuracy, making it suitable for real-time financial fraud detection applications.

Recent research has also explored hybrid optimization frameworks to enhance classification performance. XGBoost integrated with Genetic Programming (XGB-GP) represents one such approach in which evolutionary optimization is employed to generate discriminative feature transformations before classification. Although XGB-GP frequently improves predictive accuracy, the evolutionary optimization process introduces additional computational complexity, making model training relatively expensive for large datasets. Furthermore, the hybrid framework generally provides limited insight into the decision-making process, reducing its applicability in regulated financial environments where explainability is essential.

The emergence of Explainable Artificial Intelligence (XAI) has addressed the transparency limitations associated with complex machine learning models. Financial institutions increasingly require interpretable models to satisfy regulatory requirements, improve stakeholder trust, and support human decision-making. Among various XAI techniques, SHapley Additive exPlanations (SHAP) has become one of the most widely adopted methods because it provides theoretically consistent feature attribution based on cooperative game theory. SHAP enables both global interpretation of model behavior and local explanation of individual predictions, allowing analysts to understand why specific transactions are classified as fraudulent.

Several recent studies have combined boosting algorithms with SHAP to improve model interpretability. These investigations demonstrate that SHAP effectively identifies influential transaction attributes such as transaction amount, transaction frequency, merchant category, customer behavior, account age, and geographical location. The resulting explanations assist financial investigators in validating model decisions and identifying emerging fraud patterns. Nevertheless, many existing studies focus primarily on predictive accuracy while providing limited comparative evaluation between explainable boosting models and optimization-based hybrid frameworks.

The present research addresses this gap by conducting a comparative study between an explainable LightGBM-SHAP framework and an XGB-GP framework. Unlike previous investigations that prioritize predictive performance alone, this study simultaneously evaluates classification accuracy, computational efficiency, and model interpretability. The integration of SHAP with LightGBM enables transparent fraud detection without sacrificing predictive capability, thereby supporting practical deployment in real-world financial systems.

Author/Year	Methodology	Dataset/Application	Advantages	Limitations
Phua et al.	Rule-based and Statistical Models	Financial Transactions	Simple implementation and interpretable	Poor adaptability to evolving fraud patterns
Bhattacharyya et al.	Logistic Regression, Decision Tree, Random Forest	Credit Card Fraud	Baseline machine learning performance	Reduced effectiveness on highly imbalanced datasets
Chen and Guestrin	XGBoost	Large-scale Classification	High prediction accuracy and robustness	High computational cost for large datasets
Ke et al.	LightGBM	Large-scale Machine Learning	Fast training, lower memory usage, scalable	Limited interpretability without XAI techniques
Lundberg and Lee	SHAP	Explainable AI	Consistent global and local feature explanations	Additional computational overhead for explanation generation
Recent Hybrid Studies	XGB-GP	Financial Fraud Detection	Improved predictive performance through feature optimization	Increased model complexity and limited transparency
Proposed Work	LightGBM + SHAP	Financial Fraud Detection	High accuracy, computational efficiency, explainability, regulatory compliance	Dependent on appropriate hyperparameter tuning and quality data preprocessing

RESEARCH GAP

A comprehensive review of the literature reveals several important research gaps:

1. Most existing studies emphasize classification accuracy while giving limited attention to model interpretability.
2. Hybrid optimization approaches such as XGB-GP improve prediction performance but increase

computational complexity, limiting real-time deployment.

3. Few comparative studies evaluate predictive performance, computational efficiency, and explainability simultaneously.
4. Limited research investigates the practical integration of LightGBM with SHAP for transparent financial fraud detection in large-scale transaction environments.
5. Regulatory requirements increasingly demand explainable AI, yet many fraud detection models remain difficult for financial analysts and auditors to interpret.

MOTIVATION OF THE PRESENT STUDY

To overcome these limitations, the proposed study develops an Explainable Financial Fraud Detection framework using LightGBM and SHAP analysis and compares its performance with the XGB-GP framework. The proposed approach aims to achieve three key objectives: (i) high fraud detection accuracy, (ii) computational efficiency suitable for real-time financial systems, and (iii) transparent model explanations that enhance trust, support regulatory compliance, and facilitate informed decision-making.

III. EXPLAINABLE AI AND MACHINE LEARNING MODELS

3.1 Explainable Artificial Intelligence (XAI)

Artificial Intelligence (AI) has become an essential component of modern financial systems by enabling automated decision-making for fraud detection, credit risk assessment, customer profiling, and anti-money laundering. Although advanced machine learning models achieve high predictive performance, many operate as black-box models, where the internal reasoning behind predictions remains difficult to understand. In financial applications, this lack of transparency poses significant challenges because regulatory authorities, financial institutions, and customers increasingly require explanations for automated decisions. Consequently, Explainable Artificial Intelligence (XAI) has emerged as a critical research area that aims to improve the transparency, interpretability, and trustworthiness of AI systems.

Explainable AI refers to a collection of techniques that enable humans to understand how machine learning models generate predictions. XAI provides meaningful explanations by identifying the contribution of individual input features toward a particular prediction. These explanations help domain experts validate model behavior, detect potential biases, ensure regulatory compliance, and improve stakeholder confidence. In fraud detection, explainability enables investigators to understand why a transaction has been classified as fraudulent, thereby supporting informed decision-making and reducing the risk of incorrect financial actions.

XAI techniques are generally categorized into intrinsic interpretability and post hoc interpretability. Intrinsically interpretable models, such as decision trees and linear regression, are naturally understandable but may offer limited predictive performance for complex datasets. In contrast, post hoc explanation techniques are designed to interpret high-performing black-box models without modifying their internal architecture. Popular post hoc methods include Local Interpretable Model-Agnostic Explanations (LIME), SHapley Additive exPlanations (SHAP), Partial Dependence Plots (PDP), and Integrated Gradients. Among these methods, SHAP has gained widespread acceptance because of its strong theoretical foundation and ability to provide consistent feature attribution.

3.2 SHapley Additive exPlanations (SHAP)

SHAP is a model-agnostic explainability technique derived from **cooperative game theory**, where each feature is considered a player contributing to the final prediction. The method assigns a **Shapley value** to every feature, representing its average contribution across all possible feature combinations. These values satisfy desirable properties such as local accuracy, consistency, and missingness, making SHAP one of the most reliable interpretation methods for machine learning models.

For a prediction function f(x), the SHAP explanation model can be represented as

$$f(x) = \phi_0 + \sum_{i=1}^{M} \phi_i$$

where:

- f(x) represents the model prediction,
- ϕ_0 denotes the baseline prediction,

Parameter	LightGBM	XGB-GP
Learning Strategy	Leaf-wise Gradient Boosting	XGBoost with Evolutionary Feature Optimization
Computational Speed	Very High	Moderate
Memory Requirement	Low	High
Feature Engineering	Automatic	GP-based Optimization
Scalability	Excellent	Moderate
Hyperparameter Complexity	Moderate	High
Explainability	High with SHAP	Limited
Real-Time Deployment	Highly Suitable	Less Suitable
Computational Cost	Low	High
Regulatory Transparency	Excellent with SHAP	Limited

- ϕi represents the contribution of the ith feature,
- M is the total number of input features.

Positive SHAP values increase the probability of a transaction being classified as fraudulent, whereas negative values decrease that probability. SHAP provides both global explanations, identifying the most influential features across the entire dataset, and local explanations, explaining individual transaction predictions. This capability makes SHAP particularly suitable for financial fraud detection, where auditors often require transaction-level justification for automated decisions.

3.3 Machine Learning Models for Financial Fraud Detection

Machine learning algorithms have demonstrated superior capability in detecting complex fraud patterns compared with conventional statistical techniques. Ensemble learning methods, particularly boosting algorithms, have become increasingly popular because they effectively capture nonlinear relationships, handle heterogeneous financial data, and provide high classification accuracy.

3.3.1 Light Gradient Boosting Machine (LightGBM)

LightGBM is a gradient boosting framework developed by Microsoft that employs decision-tree ensembles for classification and regression tasks. Unlike conventional boosting algorithms that grow trees level-wise, LightGBM adopts a leaf-wise tree growth strategy, allowing faster convergence and improved predictive performance.

Several algorithmic innovations contribute to its efficiency:

- Histogram-based feature discretization reduces computational complexity.
- Gradient-based One-Side Sampling (GOSS) decreases training time by retaining informative samples.
- Exclusive Feature Bundling (EFB) minimizes dimensionality through feature compression.
- Efficient memory management enables processing of large-scale datasets.

The objective function of LightGBM is expressed as:

$$Obj = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{K} \Omega(f_k)$$

where

- L(yi,y^i) denotes the loss function,
- fk represents the kth decision tree,
- Ω(fk) is the regularization term controlling model complexity.

The advantages of LightGBM include:

- High classification accuracy.
- Faster training compared with conventional gradient boosting algorithms.
- Low memory consumption.
- Effective handling of high-dimensional datasets.
- Suitability for real-time fraud detection systems.

3.3.2 Extreme Gradient Boosting with Genetic Programming (XGB-GP)

Extreme Gradient Boosting (XGBoost) is one of the most widely used ensemble learning algorithms because of its excellent predictive performance and regularization capabilities. Genetic Programming (GP) is an evolutionary optimization technique that automatically generates mathematical feature transformations and optimized feature combinations.

The XGB-GP framework combines these two approaches by first applying GP to evolve

informative feature representations and then utilizing XGBoost for classification. This hybrid framework enhances predictive accuracy by discovering complex nonlinear relationships that may not be captured through conventional feature engineering.
However, the optimization process increases computational cost because multiple generations of candidate solutions must be evaluated before identifying the optimal feature transformations. Consequently, although XGB-GP often improves classification accuracy, its higher computational complexity may reduce its suitability for large-scale real-time fraud detection applications.

3.4 Comparative Analysis of LightGBM and XGB-GP

Both LightGBM and XGB-GP belong to the family of ensemble learning algorithms and have demonstrated excellent performance in fraud detection. However, they differ significantly in terms of computational efficiency, scalability, and interpretability.
Table 2. Comparison between LightGBM and XGB-GP
The proposed framework combines the computational advantages of LightGBM with the interpretability of SHAP, providing a practical balance between predictive performance and transparency. Compared with XGB-GP, the proposed approach reduces computational overhead while offering detailed explanations for both global model behavior and individual fraud predictions. These characteristics make the LightGBM-SHAP framework particularly suitable for deployment in financial institutions that require accurate, scalable, and explainable fraud detection systems.

3.5 Summary

The integration of Explainable Artificial Intelligence with advanced ensemble learning models represents a significant advancement in financial fraud detection. While machine learning models such as LightGBM and XGB-GP effectively identify complex fraud patterns, incorporating SHAP enables transparent interpretation of prediction outcomes. The proposed LightGBM-SHAP framework not only maintains high predictive performance but also supports regulatory compliance, enhances user trust, and facilitates evidence-based decision-making. Therefore, it offers a promising solution for modern financial fraud detection systems where both accuracy and explainability are essential.

IV. RESEARCH GAP, PROPOSED METHODOLOGY AND PERFORMANCE EVALUATION

4.1 Research Gap

The rapid adoption of machine learning techniques has significantly improved the effectiveness of financial fraud detection systems. Nevertheless, a detailed review of the existing literature reveals several unresolved challenges that limit the practical deployment of these models in real-world financial environments.
Most existing studies primarily focus on maximizing classification accuracy while paying limited attention to model interpretability. Although ensemble learning algorithms such as Random Forest, XGBoost, CatBoost, and LightGBM demonstrate excellent predictive performance, they often function as black-box models**,** making it difficult for financial analysts and regulatory authorities to understand the reasoning behind fraud predictions.
Another significant challenge is the highly imbalanced nature of financial transaction datasets**.** Since fraudulent transactions constitute only a small proportion of total transactions, conventional classifiers tend to be biased toward legitimate transactions, leading to lower fraud detection rates and higher false-negative errors. Several studies have addressed this issue using oversampling, undersampling, or cost-sensitive learning; however, consistent improvements across different datasets remain difficult to achieve.
Recent hybrid approaches, including XGBoost integrated with Genetic Programming (XGB-GP)**,** have demonstrated improved predictive performance through evolutionary feature optimization. However, these frameworks generally require higher computational resources and longer training times, reducing their suitability for real-time fraud detection systems that process millions of transactions daily. Furthermore, most optimization-based hybrid models provide limited interpretability, restricting their application in regulatory environments that demand transparent and auditable AI systems.

The emergence of Explainable Artificial Intelligence (XAI) has addressed the need for transparency, yet relatively few studies have comprehensively integrated explainability with computationally efficient boosting algorithms. Existing research often evaluates prediction accuracy independently of explainability, while comparative analyses involving both performance and interpretability remain limited.

Based on these observations, the following research gaps are identified:

- Limited integration of high-performance machine learning models with explainable AI techniques**.**
- Insufficient comparative evaluation between LightGBM-SHAP and optimization-based frameworks such as XGB-GP**.**
- Limited emphasis on computational efficiency for large-scale, real-time financial fraud detection.
- Inadequate transaction-level explanations for supporting regulatory compliance and financial auditing.
- Lack of comprehensive evaluation considering predictive accuracy, interpretability, and computational performance simultaneously.

Explainable Financial Fraud Detection proposed research addresses these limitations by integrating LightGBM with SHAP analysis to develop an explainable fraud detection framework and comparing its performance with the XGB-GP model.

V. CONCLUSION AND FUTURE SCOPE

This study presented an explainable financial fraud detection framework based on LightGBM and SHAP (SHapley Additive exPlanations) and compared its performance with the XGB-GP framework. The experimental analysis demonstrates that machine learning techniques, particularly gradient-boosting models, can effectively identify complex and nonlinear patterns associated with fraudulent financial transactions. LightGBM provides an efficient and scalable learning mechanism while maintaining strong predictive performance, making it suitable for large-scale financial datasets.

The integration of SHAP analysis significantly improves the interpretability of the proposed fraud detection system. Instead of providing only a binary prediction of fraudulent or legitimate transactions, SHAP explains the contribution of individual features to each prediction. This enables financial institutions, auditors, and risk analysts to understand why a particular transaction has been classified as suspicious. The comparative analysis with XGB-GP further highlights the importance of combining predictive accuracy with model transparency when developing fraud detection solutions.

Overall, the findings indicate that an explainable LightGBM-based framework can provide an effective balance between detection capability, computational efficiency, and interpretability. Such a framework can support faster investigation of suspicious transactions, improve analyst confidence, and contribute to more transparent and reliable financial fraud detection systems.

Future Scope

The proposed work can be extended in several directions. First, the framework can be evaluated using larger, real-world financial transaction datasets containing diverse fraud patterns, multiple transaction channels, and continuously evolving fraudulent behaviour. Second, advanced deep learning and hybrid machine learning models such as LSTM, Transformer-based architectures, graph neural networks, and ensemble models can be investigated and compared with LightGBM and XGB-GP.

Future research can also focus on real-time fraud detection, where incoming transactions are analysed continuously and suspicious activities are identified with minimal latency. Integration with streaming technologies and online learning algorithms could allow the system to adapt to new fraud patterns and concept drift.

Another important direction is the development of advanced explainability mechanisms that combine SHAP with other XAI techniques such as LIME, counterfactual explanations, and feature interaction analysis. This could provide both global and transaction-level explanations that are easier for non-technical financial analysts to interpret.

Furthermore, the framework can incorporate temporal and behavioural transaction patterns, customer profiles, transaction networks, and contextual information to improve fraud detection. Privacy-preserving approaches such as federated learning could also enable multiple financial institutions to collaboratively improve fraud detection models without directly sharing sensitive customer data.

Finally, future studies should evaluate the proposed system using practical measures such as false-positive reduction, detection latency, computational cost, robustness against adversarial attacks, and model stability over time. Deployment as an intelligent decision-support system for banks, digital payment platforms, insurance companies, and other financial institutions could further validate its practical applicability.

REFERENCES WITH CITATION

- Financial fraud detection using machine learning and data-mining techniques has been extensively studied, particularly for identifying anomalous transaction patterns [1], [2].
- Ensemble learning methods such as Gradient Boosting and XGBoost have demonstrated strong predictive performance on structured financial datasets [3], [4].
- LightGBM provides an efficient gradient-boosting framework using techniques such as Gradient-based One-Side Sampling and Exclusive Feature Bundling, making it suitable for large-scale datasets [5].
- Explainable Artificial Intelligence (XAI) improves the transparency and interpretability of machine-learning predictions [6], [7].
- SHAP provides a game-theoretic approach for explaining individual predictions and quantifying the contribution of each feature [8].
- SHAP has been widely applied to tree-based models because of its efficient TreeSHAP implementation [9].
- In financial applications, explainability is particularly important because fraud detection systems need to provide understandable reasons for classifying transactions as fraudulent [10], [11].
- Comparative studies of machine-learning algorithms indicate that ensemble-based methods can achieve strong fraud-detection performance while maintaining robustness against complex nonlinear relationships [12].

[1] Ngai, E.W.T., Hu, Y., Wong, Y.H., Chen, Y. and Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50(3), pp.559–569. https://doi.org/10.1016/j.dss.2010.08.006

[2] West, J. and Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. Computers & Security, 57, pp.47–66. https://doi.org/10.1016/j.cose.2015.09.005

[3] Dal Pozzolo, A., Caelen, O., Johnson, R.A. and Bontempi, G. (2015). Calibrating probability with undersampling for unbalanced classification. In: 2015 IEEE Symposium Series on Computational Intelligence, pp.159–166. https://doi.org/10.1109/SSCI.2015.47

[4] Dal Pozzolo, A., Caelen, O., Le Borgne, Y.A., Waterschoot, S. and Bontempi, G. (2014). Learned lessons in credit card fraud detection from a practitioner perspective. Expert Systems with Applications, 41(10), pp.4915–4928. https://doi.org/10.1016/j.eswa.2014.02.026

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

[6] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems, 30.

[7] Lundberg, S.M. and Lee, S.I. (2017). A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems, 30.

[8] Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N. and Lee, S.I. (2020). From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2, pp.56–67. https://doi.org/10.1038/s42256-019-0138-9

[9] Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R. and Herrera, F. (2020). Explainable

Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, pp.82–115. https://doi.org/10.1016/j.inffus.2019.12.012

[10] Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.E., He-Guelton, L. and Caelen, O. (2018). Sequence classification for credit-card fraud detection. Expert Systems with Applications, 100, pp.234–245. https://doi.org/10.1016/j.eswa.2018.01.037