Demystifying Interpretable AI in Finance: A Review of SHAP and LIME

Sarthak Durgesh Marathe
Student – SPPU

***Abstract*— SHAP and LIME have become essential tools for interpreting complex machine learning models, particularly in finance, where predictive systems influence critical decisions and economic stability. These methods provide detailed insights into how algorithms make predictions across a wide range of financial tasks, including credit scoring, fraud detection, and environmental, social, and governance evaluation. This review compiles findings from recent studies that apply SHAP and LIME in financial contexts and compares their theoretical foundations, practical effectiveness, and current limitations. It also considers the direction of ongoing improvements aimed at achieving scalability, reliability, and domain adaptation. Explainable artificial intelligence is shown to be a key component of transparency and accountability in financial technology, though much progress is still needed before interpretability becomes standard practice across the Finance sector.**

I. INTRODUCTION

Artificial intelligence is dramatically reshaping financial services. Increasingly complex machine learning models are directly influencing decisions related to lending, investing and risk management. However, the very complexity that gives these models power also renders their decision processes opaque, creating a need for interpretability frameworks. SHAP and LIME are two such techniques that have risen to the challenge. They are used to generate explanations both at the local level (individual predictions) and the global level (for the model's overall behaviour). This review synthesizes the latest research focusing on the application of these explainability techniques in finance, highlighting their role in improving transparency and regulatory compliance and helping stakeholders, from data scientists to regulators, understand AI-driven financial decisions without needing extensive knowledge in machine learning.

"A doctor would never operate on a patient because 'the model said so'" (Nieto Juscafresa, An Introduction to Explainable Artificial Intelligence with LIME and SHAP).

BACKGROUND

SHAP derives from Shapley values in cooperative game theory, where each input feature's contribution to the model's decision is fairly assigned using Shapley values. This approach comes with strong theoretical guarantees, such as consistency and local accuracy, making SHAP explanations trustworthy and rigorous. On the flip side, the computation of exact Shapley values requires evaluating many coalitions of features which grows exponentially with feature count and can turn into a bottleneck for real-time or large-scale applications. Fortunately, various approximations and model specific optimizations like TreeSHAP exist to reduce this cost.

"As Gramegna and Giudici (2021) define it, 'SHAP values are an explanatory model that locally approximate the original model, for a given variable value x (local accuracy); with the property that, whenever a variable is equal to zero, so is the Shapley value (missingness); and that if in a different model the contribution of a variable is higher, so will be the corresponding Shapley value (consistency)'" (Gramegna & Giudici, SHAP and LIME: An Evaluation of Discriminative Power in Credit Risk).

LIME takes a different but complementary route by creating simple and easy to understand surrogate models locally around a specific prediction. It slightly changes the input data and fits a lightweight model, such as linear regression or decision trees on this neighbourhood to explain the complex model's behaviour nearby. LIME's flexibility and speed allow for rapid explanations, particularly valuable in operational settings where fast turnaround is needed. But since it focuses only on the local approximation, results can sometimes be unstable and sensitive to

parameters like kernel width or sample size. Both methods satisfy important roles in the finance ecosystem where transparency for individual cases and model wide explanations is both essential.

"According to Knab et al. (2023), 'LIME explains the decisions of a neural network f in a model-agnostic and instance-specific (local) manner, applicable to images, text, and tabular data' by training 'a local, interpretable surrogate model g to approximate f around an instance x'" (Knab et al., Which LIME should I trust? Concepts, Challenges, and Solutions).

APPLICATIONS IN FINANCE

"The emergence of explainable AI in finance addresses a critical challenge where 'the incredible steps forward made in IT gave a real shake to the way [credit evaluation] was performed by the industry,' but 'the increase in prediction power of new algorithms takes a toll on explainability, since the models are now so complex that it is close to impossible to establish clear links between the inner workings of the model and the given output'" (Gramegna & Giudici, SHAP and LIME: An Evaluation of Discriminative Power in Credit Risk).

CREDIT SCORING AND DEFAULT PREDICTION

A significant number of studies focus on credit risk assessment as an area where explainability is urgently needed. The responsibility is huge because decisions influence whether individuals can secure loans or mortgages, which affects their financial futures. SHAP has been widely used to uncover and clearly communicate the contributions of features such as income, past delinquencies, debt-to-income ratios, and employment status to the probability of default. These detailed attributions enable lenders to justify decisions and detect biases or inequities in their models. In parallel, LIME offers a more agile solution, allowing lenders and loan officers to generate quick and intuitive explanations for single loan cases making it easier to communicate effectively with customers when denying or approving financing. This tailored use of SHAP and LIME ensures that automated credit assessments do not become inscrutable black boxes that lack interpretability but are tools that can be examined and trusted.

Financial institutions are increasingly combining SHAP with model governance platforms to produce explanation reports automatically. These reports are often shared with the respective internal teams to ensure fairness and compliance with regulations. SHAP's consistency makes it particularly useful in defending decisions when challenged by regulators or customers. In contrast, LIME's flexibility is valuable for customer-facing applications, where loan officers can quickly show why a certain decision was made. The growing use of these methods has encouraged the adoption of "explainability dashboards," which visualize the top factors influencing each applicant's credit outcome. This not only improves internal transparency but also promotes ethical lending practices.

FRAUD DETECTION

Fraud detection is inherently challenging due to its adversarial nature. SHAP's ability to group correlated features and provide aggregated importance scores offers detailed insight into anomalous patterns of suspicious behaviours that are buried deep within complex models. This enables investigators to identify suspicious activity more effectively. Explaining these subtle signals helps investigators prioritize alerts more effectively and strengthens the audit trail required for compliance. Meanwhile, LIME's ability to rapidly generate visual case specific explanations aids analysts in validating fraud alerts in real time reducing false positives and increasing operational efficiencies. Both tools play complementary roles in fighting financial crime by making the "why" behind suspicious activities more explicit.

In most real-world setups, SHAP is used during periodic audits to assess model performance, while LIME serves in real-time fraud detection dashboards. This division ensures that while overall trends are well understood, immediate cases still receive human verification before any financial block or report is issued.

BOND AND MARKET RISK

Volatility and risk prediction in bond markets are notoriously complex problems with many intertwined factors. SHAP helps explain how different market factors like interest rates, credit spreads, and economic

indicators such as GDP growth, inflation, and unemployment affect the model's predictions in detail. This transparency is highly valuable for traders and risk managers who must understand the drivers and causes of these risks under different scenarios. While LIME enhances this by enabling "what-if" scenario analyses, allowing stakeholders to explore how small changes in factors could alter predictions. This supports more effective scenario planning and stress testing.

Together these explanations create a richer and more intuitive understanding of risk, moving financial decision-making beyond opaque black box models and guesswork.

One of the major uses of SHAP in this area is portfolio sensitivity analysis. SHAP values are often used to quantify which features cause changes in predicted bond prices or risk scores. This helps detect hidden dependencies between macroeconomic indicators that would otherwise go unnoticed. LIME contributes to these insights by simulating local modifications, helping to understand how minor interest rate shifts might impact overall volatility. These explanations are also becoming popular in risk reporting, where regulators expect high transparency on how predictive models behave under stress test conditions.

Another emerging application for this would be in automated trading systems. Integrating SHAP into trading models allows institutions to verify whether the system is relying on legitimate market indicators or spurious correlations. The author thinks this level of explainability is crucial for preventing overreliance on algorithmic trading decisions that could otherwise introduce systemic risks in the market.

Environmental Social and Governance Ratings

Environmental, Social, and Governance (ESG) factors are used to measure how responsibly a company operates beyond just financial performance. These include how it manages its environmental impact, how it treats its employees and communities, and how it upholds ethical and transparent corporate governance. ESG scores have become a key part of modern investing, helping stakeholders assess the long-term sustainability and ethical standing of businesses.

With the recent surge in sustainable investing, the importance of transparent ESG ratings has been highlighted. SHAP, when combined with models like XGBoost helps reveal which aspects of a company's ESG profile influence its rating the most. This transparency encourages responsible investing, allowing asset managers to align their portfolios with their values while maintaining clear and rigorous audit trails. Making ESG ratings understandable also empowers regulators and stakeholders eager for transparency to trust the methodologies behind these increasingly important metrics.

CUSTOMER SEGMENTATION AND MARKET ANALYTICS

Going beyond predictions, some studies apply SHAP and LIME to unsupervised learning tasks such as customer segmentation and market analysis. By using feature attributions to explain cluster assignments, financial institutions can better understand customer behaviour and tailor products accordingly. This adds an important layer of interpretability to models often deemed opaque (black box models).

In most cases, SHAP helps identify which attributes make customers fall into high-value or low-risk segments, while LIME provides quick explanations for individual data points. Though simpler compared to other use cases, these explanations play a key role in improving personalization and customer experience.

COMPARISON AND DISCUSSION

When it comes to accuracy and consistency in explanations SHAP's game theoretic basis really shines. Its results are stable across datasets and models, making it particularly suitable for contexts demanding high accountability such as regulatory audits and risk reporting. LIME, while generally faster and more adaptable across diverse models may produce less stable outputs depending on surrogate model parameters and sampling methods. Both approaches integrate well with the tree-based ensembles commonly used in finance such as Random Forests and XGBoost but LIME's performance may degrade when dealing with highly nonlinear and high dimensional models.

Practically, SHAP tends to be favoured for comprehensive risk optimization model validation and transparency reports. LIME on the other hand supports

interactive exploration and communication efforts breaking down barriers between technical teams and business stakeholders. Challenges remain for both particularly in managing feature interactions and addressing correlated features which can confound explanations and lead to misleading interpretations without nuanced treatment

FUTURE DIRECTIONS

The path forward involves overcoming key challenges. Enhanced methods that account more precisely for feature correlations and nonlinear dependencies such as grouped Shapley values and advanced LIME variants that model local nonlinear behaviour promise improved interpretative fidelity. Scalability is another pressing issue particularly for SHAP which demands efficient algorithms to keep pace with the vast and rapid flow of financial data. Robustness against adversarial examples and noisy inputs must also improve to maintain trust in high stakes environments like fraud detection. Further customizing interpretability tools for domain specific tasks ranging from derivatives pricing to insurance claim analysis will help expand their usability and impact.

Additionally developing explanation systems that are tailored to user expertise and cognitive styles will help bridge the gap between complex model outputs and actionable insights. This user centric focus aims to democratize AI understanding making high quality explanations accessible both to regulators and to analysts and consumers alike.

"As noted by Knab et al. (2023), 'LIME faces several challenges, including instability, computational inefficiency, and limitations in the handling of certain types of data,' highlighting the need for 'numerous studies that have proposed enhancements to address these issues'" (Knab et al., Which LIME should I trust? Concepts, Challenges, and Solutions).

Conclusion

"Both LIME and SHAP are powerful tools for model interpretability, each with its own strengths and weaknesses" (Nieto Juscafresa, An Introduction to Explainable Artificial Intelligence with LIME and SHAP).

The widespread adoption of SHAP and LIME across these diverse financial applications reveals an interesting paradox. While institutions embrace these tools for regulatory compliance and stakeholder trust, the practical implementations often favour convenience over theory. Credit scoring teams lean heavily on SHAP's mathematical foundation because auditors demand it, yet fraud detection systems frequently choose LIME for its speed despite explanation inconsistencies.

CONCLUSION

What's particularly fascinating is how these tools are reshaping the skill requirements in finance. The emergence of "model explainers" as a distinct role shows that technical interpretability has become a business necessity rather than just an academic exercise. The ESG application especially highlights this trend, where explanation quality directly impacts investment decisions and regulatory approval.

From the author’s perspective, the real value lies not in the individual explanations these tools provide, but in how they're forcing financial institutions to think more systematically about algorithmic accountability. The technology is mature enough for practical deployment, but the organizational processes around explanation validation and stakeholder communication are still evolving rapidly.

REFERENCE

[1] ACM Digital Library. (2022). Model interpretability of financial fraud detection by group SHAP.
[2] arXiv. (2024). Provably stable rankings with SHAP and LIME.
[3] arXiv. (2025). A method for evaluating the interpretability of machine learning models in predicting bond default risk based on LIME and SHAP.
[4] arXiv. (2025). A systematic review of explainable AI in finance.
[5] arXiv. (2025). Interpretable credit default prediction with ensemble learning and SHAP.
[6] arXiv. (2025). Which LIME should I trust? Concepts, challenges, and survey.
[7] C3 AI. (2023). LIME: Local interpretable model agnostic explanations.

[8] European Data Protection Supervisor. (2023). Explainable artificial intelligence technical dispatch.
[9] Frontiers in Artificial Intelligence. (2021). SHAP and LIME: An evaluation of discriminative power in credit risk.
[10] IOSR Journal of Economics and Finance. (2025). Finance modeling approach using machine learning.
[11] International Journal of Computer Applications. (2023). Review on explainable AI by using LIME and SHAP.
[12] MarkovML. (2024). LIME vs SHAP: A comparative analysis of interpretability tools.
[13] PMC. (2021). An evaluation of discriminative power in credit risk.
[14] ScienceDirect. (2023). Interpretable machine learning for imbalanced credit scoring.
[15] ScienceDirect. (2024). BMB-LIME: LIME with modeling local nonlinearity and bias.
[16] ScienceDirect. (2024). Machine learning model interpretability using SHAP values.
[17] SSRN. (2024). Corporate ESG rating prediction based on XGBoost-SHAP.
[18] Svitla Systems. (2024). Interpreting machine learning models using LIME and SHAP.
[19] University of Barcelona. (2023). An introduction to explainable artificial intelligence with LIME and SHAP.
[20] Wiley Online Library. (2024). A perspective on explainable artificial intelligence methods.