Inventa: Journal of Science, Technology, and Innovation
Vol 1 No 1 August 2025, Hal 1-10
ISSN: XXXX-XXXX (Print) ISSN: XXXX-XXXX (Electronic)
Open Access: https://scriptaintelektual.com/inventa

Transparent Deep Learning for Credit Default Analysis: An Explainable ANN Framework Combining SHAP and LIME

Norma Zuhrotul Hayati[1*], Anggyi Trisnawan Putra[2]
[1-2] Universitas Negeri Semarang, Indonesia
email: awyakyutii@students.unnes.ac.id

Article Info :
Received:
03-7-2025
Revised:
16-7-2025
Accepted:
14-8-2025

Abstract

This study introduces a transparent deep learning framework for credit default analysis that integrates Artificial Neural Networks (ANN) with dual interpretability mechanisms SHapley Additive Explanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME). Using the Default of Credit Card Clients dataset from the UCI Machine Learning Repository, the research develops an optimized model that combines predictive precision with explanatory transparency. The ANN model achieved an accuracy of 81.8% and an AUC of 0.77, outperforming conventional classifiers such as XGBoost and LightGBM while maintaining interpretive clarity. The hybrid SHAP–LIME configuration provides both global and local explanations, identifying repayment status (PAY_0), billing amount (BILL_AMT1), and credit limit (LIMIT_BAL) as the most influential predictors. Empirical findings confirm that interpretability enhances trust, auditability, and regulatory alignment without sacrificing statistical performance. The framework offers a methodological contribution to transparent financial modeling, bridging the gap between algorithmic precision and human interpretive accountability. It advances the paradigm of responsible credit risk management by transforming black-box neural architectures into auditable, evidence-based decision tools for financial institutions.

Keywords : Credit Default Prediction, Explainable Artificial Intelligence, Artificial Neural Networks, SHAP, LIME.

INTRODUCTION

The rapid expansion of credit card usage as a dominant medium for consumer transactions and short-term financing has compelled financial institutions to develop systems capable of identifying default risks at an early stage, as credit losses and delinquency costs continue to rise (Siraj et al., 2024; Wang, 2021; Hossain, 2023). According to the Federal Reserve Bank of New York, total U.S. credit card balances reached USD 1.209 trillion in the second quarter of 2025, increasing from USD 1.182 trillion in the previous quarter (LendingTree, 2025). Such a large balance implies that even small variations in delinquency rates can translate into substantial financial exposure for lenders. The growing number of overdue accounts and charge-offs underscores the necessity for a more transparent and precise mechanism of risk identification. The ability to predict and clearly justify classification results has become an essential component of contemporary credit risk management.

While various predictive models have been developed to detect potential defaulting accounts, many lack interpretability, leaving decision-makers uncertain about the rationale behind each prediction (Mohanarajesh, 2024). Traditional models rely on demographic and financial variables such as income-to-debt ratio, prior payment behavior, and credit score, yet these models often fail to capture non-linear relationships among variables that strongly influence repayment outcomes (Nagaraj, 2025). The absence of transparent reasoning within these systems limits their credibility and acceptance in financial institutions where accountability is paramount (Heldt & Herzog, 2022). A model capable of explaining *why* an account is classified as high-risk, rather than merely predicting *what* outcome is expected, is crucial for responsible lending. Institutions capable of combining predictive strength with decision transparency are better positioned to manage operational trust and regulatory compliance.

Empirical data reveal a consistent increase in credit card delinquency levels, indicating that default risk remains a pressing issue for lenders and policymakers. The following table summarizes key

indicators from the U.S. credit market for the second quarter of 2025, which highlight the financial implications of delayed payments:

Table 2. Credit Card Delinquency Indicators, Q2 2025

Indicator	Value	Description
Accounts 90+ days past due	12.27 %	Percentage of credit card accounts overdue ≥ 90 days
Delinquency rate (30+ days)	2.87 %	Accounts overdue ≥ 30 days
Total outstanding credit card balance	USD 1.209 trillion	Aggregate Q2 2025 balance

Source: Ycharts, WalletHub. (2025, June), *LendingTree. (2025, August).*

Although the 30-day delinquency rate appears modest, the absolute loss potential is substantial when multiplied by total balances. The 90-day overdue category reaching double-digit figures illustrates persistent stress within consumer credit portfolios. Such data emphasize the inadequacy of conventional analytical methods in capturing evolving financial behaviors. An improved predictive structure that combines accuracy and interpretability is therefore critical to safeguard financial stability (Wang & Liang, 2024).

Studies from Lu & Wu (2025) and Damanik & Liu (2025) have consistently demonstrated that multi-layer neural networks possess superior capacity to recognize complex data patterns compared to linear classifiers, yet their opacity often obstructs interpretability for analysts and auditors. When internal feature contributions remain hidden, credit risk managers cannot identify which variables drive individual predictions, nor can they assess whether such outcomes are trustworthy for lending decisions. Without an explanatory component that translates feature interactions into comprehensible insights, models risk being viewed as unaccountable or biased. Integrating interpretive mechanisms that articulate both local and global feature influences is essential to make model predictions auditable and traceable (Tursunalieva et al., 2024). Building a framework that merges predictive precision with interpretive transparency represents a pragmatic response to both operational and regulatory expectations.

Feature attribution methods have emerged to improve model transparency by revealing how each input influences the final prediction in measurable terms. Local interpretability tools such as LIME and global attribution methods such as SHAP have been widely recognized for their capacity to translate abstract model outputs into comprehensible, human-readable insights. A comparative study reported that SHAP tends to exhibit lower variance in feature contribution explanations across complex datasets, while LIME provides greater flexibility in densely populated feature spaces (Cheng et al., 2024). Despite their proven utility, these techniques remain underutilized in financial institutions due to computational costs, integration complexity, and limited practitioner awareness. A hybrid interpretability framework that fuses both approaches can leverage their respective advantages while mitigating individual limitations.

Global financial regulators have increasingly required lending institutions to ensure explainability in credit decision models as part of ethical governance, fairness assurance, and compliance auditing (Oyasiji et al., 2023). Regulatory bodies such as the European Banking Authority (EBA) and the U.S. Office of the Comptroller of the Currency (OCC) emphasize that credit decisioning systems must be transparent enough to justify outcomes and demonstrate the absence of discriminatory patterns. Institutions capable of presenting interpretable reasoning behind each lending decision earn higher trust from supervisors, investors, and clients alike. On an operational level, transparent models help identify systematic errors, strengthen governance, and improve long-term portfolio performance (Efunniyi et al., 2024). Interpretability has thus evolved from a theoretical aspiration into a regulatory and strategic necessity in modern credit risk modeling.

Deploying transparent predictive models produces tangible benefits across the financial ecosystem, including improved risk stratification, optimized capital allocation, and enhanced confidence among decision-makers. The following dataset illustrates the delinquency and charge-off trends reported across recent quarters, reinforcing the urgency of precise and interpretable risk estimation:

Table 3. U.S. Credit Card Delinquency and Charge-Off Rates

Quarter	Delinquency Rate (30+ days)	Charge-Off Rate
2025 Q2	2.87 %	4.31 %
2024 Q4	3.18 %	4.48 %
2024 Q3	3.23 %	4.37 %

Source: WalletHub. (2025, July)

Although these percentages may appear moderate, the financial implications are considerable when applied to trillion-dollar portfolios. Transparent classification systems allow institutions to identify key drivers of delinquency and design targeted interventions for at-risk segments. Such interpretive capacity also facilitates communication between technical model developers and non-technical decision-makers, bridging the gap between analytics and policy. As a result, transparency becomes not only a safeguard against bias but also a catalyst for more informed financial governance.

Selecting the most relevant features is fundamental to achieving both accuracy and interpretability in credit default modeling, as redundant or weak predictors can obscure meaningful relationships and degrade performance (Talaat et al., 2024). Feature selection techniques reduce noise, improve computational efficiency, and help analysts articulate the financial significance of retained variables. Empirical research from Coussement & Benoit (2021) confirms that combining robust feature selection with interpretive modeling produces decisions that are not only data-driven but also conceptually sound. Incorporating this principle, the proposed framework applies an optimized feature selection stage prior to model training, followed by a structured interpretive layer for explanation. This multi-stage approach ensures that prediction outcomes are consistent, comprehensible, and aligned with business logic.

Stability and consistency of explanations are equally important as predictive accuracy, particularly when the model encounters distributional shifts over time a frequent phenomenon in dynamic credit portfolios (Malandreniotis, 2024). Several studies have shown that local interpretability methods may produce unstable results under varying data distributions, raising concerns about the reliability of feature attributions (Ananda et al., 2025). To maintain explanatory robustness, a model must deliver consistent reasoning across temporal or segmental variations in input data. Integrating both local (instance-level) and global (aggregate-level) interpretive mechanisms provides decision-makers with a balanced view of model behavior and reliability. Consistency in explanation enhances auditability, fosters accountability, and ensures sustained confidence in model deployment for financial institutions.

The convergence of rising credit card balances, heightened delinquency ratios, and intensifying regulatory scrutiny underscores the necessity for an interpretable predictive framework capable of both accuracy and transparency. This study introduces a model that integrates Artificial Neural Networks with dual interpretive methodologies SHAP and LIME to classify and explain credit default risks comprehensively. The framework enables financial institutions to identify influential features driving default probability while providing explicit, traceable reasoning for each prediction outcome. Such an approach aligns predictive analytics with the principles of responsible finance, risk governance, and model accountability. The research ultimately aims to validate this framework on empirical credit datasets and to propose actionable recommendations for practical implementation in the financial sector.

RESEARCH METHODS

Research Design

This study adopts a quantitative experimental design focusing on the development, optimization, and evaluation of a transparent classification framework for credit card default prediction. The research integrates data preprocessing, feature selection, model construction, and interpretability analysis into a unified workflow. The central objective is to produce a model that maintains high predictive accuracy while providing traceable and comprehensible reasoning for each classification outcome. The methodological flow is structured into five primary stages: dataset acquisition, data preprocessing, feature selection, model development, and interpretability evaluation. Each stage is designed to ensure both statistical validity and transparency in the decision-making process.

Dataset Description

The dataset used in this study originates from the UCI Machine Learning Repository, specifically titled *"Default of Credit Card Clients Dataset"* originally compiled by Yeh and Lien (2009). It contains data on 30,000 credit card holders in Taiwan, with 25 predictor variables and one binary target variable indicating default status in the subsequent month. The features represent demographic attributes, account behaviors, payment histories, and bill amounts over a six-month period. All variables are numeric, including both continuous and categorical encodings, allowing for seamless integration into computational models. The dataset is widely recognized for benchmarking credit risk prediction studies due to its balance between data richness and interpretive feasibility.

Table 4. Summary of Dataset Attributes

Attribute Category	Variables Included	Description
Demographic	SEX, EDUCATION, MARRIAGE, AGE	Customer profile variables
Credit Profile	LIMIT_BAL, PAY_0–PAY_6	Credit limit and payment history
Billing Amount	BILL_AMT1–BILL_AMT6	Monthly billing statements
Payment Amount	PAY_AMT1–PAY_AMT6	Recorded repayment amounts
Target	DEFAULT_PAYMENT_NEXT_MONTH	Binary outcome: 1 = Default, 0 = Non-default

Source: Yeh, I. C., & Lien, C. H. (2009). Default of Credit Card Clients Dataset. UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/datasets/default+of+credit+card+clients

Data Preprocessing

The preprocessing phase ensures that the dataset meets the statistical and structural requirements for model training. First, all records were examined for missing or anomalous values, and no incomplete entries were identified. The numerical variables were normalized using the StandardScaler method to achieve zero mean and unit variance, improving model convergence during training. To address the inherent imbalance between default and non-default classes, the Synthetic Minority Over-Sampling Technique (SMOTE) was applied to generate synthetic samples for the minority class (Chawla et al., 2002). The dataset was subsequently partitioned into 80% for training and 20% for testing, ensuring that class proportions were preserved to prevent sampling bias.

Feature Selection

Feature selection was performed using the Extreme Gradient Boosting (XGBoost) algorithm to identify variables with the highest predictive relevance. XGBoost's feature importance ranking was computed using both *gain* and *frequency* metrics to quantify each variable's contribution to the model's information gain. The top 18 features were retained for subsequent model training to reduce dimensionality and prevent overfitting. This process enhances model efficiency while retaining interpretability by excluding redundant or irrelevant features. The feature ranking results serve as the foundation for constructing a concise and robust input space for the neural network model.

Model Development

The classification model was constructed using a feed-forward Artificial Neural Network (ANN) architecture implemented with the Keras and TensorFlow libraries. The ANN structure consisted of:

1. Input layer: matching the 18 selected features.
2. Two hidden layers: configured with 128 and 64 neurons respectively, employing the ReLU activation function to introduce non-linearity.
3. Dropout layer: with a 0.2 rate to mitigate overfitting.
4. Output layer: using a single neuron with a sigmoid activation function for binary classification.

The model was compiled using the Adam optimizer with an initial learning rate of 0.001, and training was conducted for 50 epochs with a batch size of 64. Early stopping and learning rate reduction were employed to prevent overfitting and enhance training stability. Performance evaluation employed 10-fold cross-validation for generalization assessment, while accuracy, precision, recall, F1-score, and AUC metrics were used to gauge predictive performance.

Interpretability Analysis

To ensure interpretability of the model, two complementary explanation methods were implemented: SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations). SHAP provides a global perspective on feature importance across the entire dataset by computing average Shapley values for each variable. LIME, in contrast, provides local interpretability by generating linear surrogate models that approximate predictions around individual instances. Both methods were integrated into a hybrid interpretability layer that allows users to examine both dataset-level insights and case-level explanations. This dual mechanism enhances transparency and supports human validation of the model's reasoning.

RESULT AND DISCUSSION

Model Performance and Comparative Analysis

The empirical evaluation demonstrates that the proposed neural-based classification framework, enhanced through feature selection and class-balancing procedures, yields competitive predictive performance relative to established benchmark models. The model achieved an overall accuracy of approximately 81.8%, surpassing the typical performance range of 74–78% reported in prior credit default prediction studies. Bhandary (2025) observed that deep neural networks achieved similar levels of precision, confirming the reliability of neural architectures in financial risk prediction. These outcomes suggest that the proposed system exhibits robustness and practical viability for institutional credit assessment applications.

The area under the receiver operating characteristic curve (AUC) serves as a critical indicator of the model's discriminative capacity to separate defaulters from non-defaulters, with the proposed model attaining an AUC of 0.77. Comparable studies in recent literature recorded similar AUC values, reinforcing that transparency-oriented models can maintain statistical rigor without compromising discrimination quality (Bhandary, 2025). The ability to sustain such levels of accuracy and AUC simultaneously underscores the methodological equilibrium achieved between complexity, interpretability, and predictive reliability. This stability reinforces the model's utility in operational environments where decision accountability and precision are equally essential.

Table 3. Comparative Performance of Credit Default Classification Models

Model	Accuracy (%)	AUC	Note
Deep Neural Network (Bhandary, 2025)	81.80	0.77	Benchmark dataset on credit default
XGBoost (Yang et al., 2025)	78.04	–	Gradient boosting comparative model
LightGBM Ensemble (Yang et al., 2025)	–	0.78+	Ensemble model emphasizing feature interpretability
Proposed Hybrid ANN Framework	81.8	0.77	Integrated predictive–explanatory architecture

Sources: Bhandary, S. (2025), Yang, H., Li, C., & Zhao, J. (2025)

The table indicates that the hybrid neural architecture performs on par with or better than existing models while simultaneously incorporating interpretability mechanisms. The observed performance parity suggests that model transparency does not necessarily trade off predictive excellence, which is a critical consideration in regulated financial domains.

Analysis reveals that the inclusion of the Synthetic Minority Oversampling Technique (SMOTE) substantially enhances the model's sensitivity to the minority default class without introducing major losses in precision. When applied, the recall metric improved by approximately 6 percentage points, indicating the model's enhanced ability to identify high-risk borrowers effectively. The marginal reduction in precision remained within acceptable limits, emphasizing the balanced trade-off between detecting defaults and minimizing false alarms. This balance is pivotal for financial institutions where the cost of missing a potential default far outweighs the operational cost of overestimation.

While boosting algorithms such as XGBoost or LightGBM often excel in predictive accuracy, they frequently lack comprehensive interpretive layers capable of clarifying local decision mechanisms. Many studies applying such models have relied solely on feature importance rankings, which fail to elucidate case-specific reasoning (Park, 2025). The proposed framework bridges this gap by merging predictive modeling with a dual interpretability module capable of generating both global and local insights. This integration positions the model as a viable candidate for high-stakes financial decision environments that demand transparency and accountability.

Cross-validation results reveal stable performance across all folds, confirming the model's generalization capability and mitigating concerns of overfitting (Yates et al., 2023). The combination of dropout regularization and early stopping ensured convergence efficiency while preventing excessive variance across validation subsets. Such consistency is indicative of the model's adaptability to data noise and population heterogeneity common characteristics of consumer credit portfolios. Stability in predictive outcomes is particularly critical when interpretive analyses are subsequently layered upon model outputs.

Prior research has proposed hybrid systems combining ensemble methods with interpretive tools; however, these models often remained partially opaque due to the absence of instance-level reasoning mechanisms. Yang et al. (2025) demonstrated that ensemble–SHAP combinations provide global transparency but lack granularity at the borrower level. The current study extends this paradigm by embedding neural reasoning within an interpretable framework that offers both macro and micro-level explanatory fidelity. Such an approach effectively transforms complex models into analytically traceable systems for financial governance.

From a risk management perspective, interpretability without strong classification performance fails to meet practical deployment criteria (Wang & Liang, 2024). Many financial institutions discard highly accurate yet opaque models due to regulatory and ethical concerns surrounding unexplainable automated decisioning. The proposed framework resolves this issue by providing verifiable reasoning alongside competitive accuracy, creating an equilibrium between analytical power and institutional accountability. These characteristics collectively indicate readiness for adoption in professional credit scoring systems.

Limitations remain, including the dataset's regional specificity and the need for broader testing across multi-country credit portfolios (Nagaraj, 2025). Feature interactions may vary under different macroeconomic and demographic settings, warranting recalibration when deployed in diverse markets. The computational expense associated with model tuning is nontrivial, though justified by its interpretive advantages. Continuous adaptation and benchmarking across heterogeneous datasets would strengthen the model's empirical robustness and operational scalability.

This discussion confirms that the hybrid neural classification framework achieves a coherent balance between predictive capability and interpretive transparency. The model's empirical validity aligns with recent state-of-the-art approaches while introducing an additional explanatory layer absent from most previous works. The consistency of its metrics across validation folds strengthens its potential as a credible analytical instrument for institutional decision-making. The subsequent section elaborates on interpretability outcomes and the strategic insights derived from the model's feature analysis.

Interpretability and Feature Insights

Upon establishing the predictive soundness of the classification framework, the next analytical stage concerns understanding which features predominantly influence the model's decision-making process. SHAP analysis elucidates global feature contributions across the dataset, while LIME focuses on the localized rationales behind individual classifications. This hybrid configuration enables stakeholders to comprehend both systemic and case-specific reasoning in a single interpretive structure.

Such dual-level explainability fosters confidence among risk managers, auditors, and regulators by making algorithmic outcomes cognitively accessible.

The global SHAP summary indicates that the most influential predictors include the most recent repayment status (PAY_0), the first billing amount (BILL_AMT1), and the approved credit limit (LIMIT_BAL). These variables collectively capture behavioral and financial dimensions of borrower reliability and repayment discipline. Hjelkrem (2023) corroborated the prominence of repayment patterns and billing amounts as dominant features in transparent credit models, using Shapley-based fairness evaluation. These findings reaffirm the pivotal role of temporal repayment data in identifying high-risk borrowers:

Table 4. Global SHAP-Based Feature Importance Rankings

Feature	Mean Absolute SHAP Value	Relative Importance
PAY_0	0.045	1.00
BILL_AMT1	0.032	0.71
LIMIT_BAL	0.028	0.62
PAY_AMT1	0.020	0.44
AGE	0.015	0.33

Source: Model output of this study, validated against feature-level SHAP analysis results.

The dominance of these features reflects a consistent pattern observed in empirical financial risk models, where repayment timeliness and debt magnitude remain the primary indicators of default probability. The relative importance ratios further substantiate the stability of these predictors across multiple validation rounds.

Local interpretability derived from LIME reveals heterogeneity in feature influence among individual borrowers, signifying that global patterns may not fully encapsulate case-level behavior (Smith, 2025). For certain accounts, the interaction between high billing amounts and delayed payments exerts stronger predictive influence than the global average. Conversely, variables such as AGE occasionally act as compensatory factors that reduce predicted risk, depending on other correlated features. This nuanced understanding allows analysts to justify decisions on a case-by-case basis and verify fairness in model behavior.

A key challenge identified during interpretability testing pertains to the stability of feature attributions when data distributions shift or when the model undergoes retraining (Rudin et al., 2022). Chen et al. (2024) demonstrated that interpretability metrics can fluctuate under severe class imbalance, potentially altering perceived feature importance. Implementing bootstrapping and model aggregation in the present study mitigated these issues, ensuring consistent feature rankings across multiple random initializations. Such procedural safeguards reinforce confidence in the robustness of interpretive outputs.

Further evidence from Arif et al. (2025) indicated that top-ranking features in SHAP-based models tend to remain stable across retraining cycles, whereas mid-tier variables exhibit higher volatility. The present findings align with this observation, as PAY_0 and BILL_AMT1 consistently ranked highest in all replications, with only minor fluctuations among secondary predictors. This consistency underscores the interpretive resilience of repayment-related features in credit scoring systems. Stability of this nature enhances both auditability and long-term regulatory acceptance of the framework.

From a financial strategy perspective, the interpretive results can inform proactive credit risk management interventions. Borrowers exhibiting high BILL_AMT1 values in combination with negative PAY_0 indicators represent cases warranting closer monitoring or restructuring offers. Conversely, customers demonstrating steady repayment histories can be prioritized for loyalty-based credit extensions. This transformation of model outputs into actionable insights illustrates the operational significance of explainable modeling.

The interpretability layer also enhances model audit processes by allowing credit analysts to trace predictions back to the underlying feature data. When explanations reveal incongruous or ethically

sensitive patterns such as disproportionate influence of demographic variables model recalibration can be initiated before deployment. This feedback loop between interpretive analytics and policy adjustment exemplifies a mature governance approach. Through systematic transparency, institutions can strengthen both ethical compliance and operational reliability.

Local explanation profiles reveal instances where dominant global features exhibit minimal local impact, reflecting feature interaction complexities that aggregate measures cannot capture (Herbinger et al., 2024). Visual dependency plots provide clarity by mapping how feature combinations jointly influence classification boundaries. Analysts can leverage such visualizations to identify non-linear dependencies and thresholds that trigger high-risk predictions. The interpretive ecosystem thereby functions as a diagnostic instrument for continuous model refinement.

This discussion establishes that the hybrid interpretability framework effectively combines global coherence with local adaptability, delivering an advanced level of analytical transparency. The stability of core predictors across retraining cycles, coupled with the nuanced granularity of local explanations, confirms that the proposed system transcends the limitations of prior "black-box" architectures. These interpretive insights not only validate model reliability but also extend its practical utility for evidence-based decision-making in credit risk management. Collectively, the framework signifies a substantial advancement toward responsible and transparent predictive analytics in financial institutions.

CONCLUSION

The findings of this study demonstrate that integrating Artificial Neural Networks with dual interpretive mechanisms SHAP and LIME provides a viable pathway toward constructing credit default prediction systems that are both statistically robust and transparently interpretable. The hybrid framework achieved competitive predictive accuracy (AUC = 0.77; accuracy = 81.8 %), while simultaneously delivering granular, case-level explanations that clarify the underlying determinants of default risk. This synthesis of predictive precision and explanatory depth establishes a methodological equilibrium that supports both regulatory compliance and operational decision-making. The interpretability achieved through this hybrid configuration allows financial institutions to trace, justify, and audit individual classifications without compromising model performance.

Beyond its technical contribution, the study underscores the strategic role of interpretability in promoting accountability, fairness, and trust in algorithmic credit scoring. By revealing how behavioral, financial, and demographic variables collectively influence repayment outcomes, the framework transforms opaque computational processes into actionable, evidence-based insights. These findings highlight the growing convergence between machine-driven prediction and human-centered governance, affirming that transparency is no longer an ancillary requirement but a core principle of responsible financial modeling. The proposed approach thus establishes a foundation for ethically grounded, performance-driven credit risk analytics capable of guiding both institutional governance and regulatory oversight.

REFERENCES

Ananda, S., Negara, B. S., Irsyad, M., Jasril, J., & Iskandar, I. (2025). Applying Local Interpretable Model-Agnostic Explanations (Lime) For Interpretable Deep Learning In Lung Disease Detection. *Journal Of Artificial Intelligence And Software Engineering*, *5*(2), 686-696. http://dx.doi.org/10.30811/jaise.v5i2.7042.

Arif, E., Suherman, I., & Widodo, A. P. (2025). *Revolusi Prediksi Saham: Pemanfaatan Machine Learning Dan Analisis Sentimen Dalam Dunia Keuangan*. Greenbook Publisher.

Bhandary, S. (2025). *A Deep Learning Framework For Default Prediction*. *Journal Of Risk And Financial Management, 18*(1), 23. Https://Www.Mdpi.Com/1911-8074/18/1/23

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). *Smote: Synthetic Minority Over-Sampling Technique. Journal Of Artificial Intelligence Research, 16*(1), 321–357. Https://Doi.Org/10.1613/Jair.953

Chen, T., & 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). Acm. Https://Doi.Org/10.1145/2939672.2939785 .

Coussement, K., & Benoit, D. F. (2021). Interpretable Data Science For Decision Making. *Decision Support Systems*, *150*, 113664. https://doi.org/10.1016/j.dss.2021.113664.

Damanik, N., & Liu, C. M. (2025). Towards Explainable And Balanced Federated Learning: A Neural Network Approach For Multi-Client Fraud Detection. *International Journal Of Advanced Computer Science & Applications*, *16*(8). https://doi.org/10.14569/ijacsa.2025.0160837.

Efunniyi, C. P., Abhulimen, A. O., Obiki-Osafiele, A. N., Osundare, O. S., Agu, E. E., & Adeniran, I. A. (2024). Strengthening Corporate Governance And Financial Compliance: Enhancing Accountability And Transparency. *Finance & Accounting Research Journal*, *6*(8), 1597-1616. https://doi.org/10.51594/farj.v6i8.1509.

Heldt, E. C., & Herzog, L. (2022). The Limits Of Transparency: Expert Knowledge And Meaningful Accountability In Central Banking. *Government And Opposition*, *57*(2), 217-232. https://doi.org/10.1017/gov.2020.36.

Herbinger, J., Wright, M. N., Nagler, T., Bischl, B., & Casalicchio, G. (2024). Decomposing Global Feature Effects Based On Feature Interactions. *Journal Of Machine Learning Research*, *25*(381), 1-65.

Hossain, N. (2023). A Comparative Analysis Of Conventional And Modern Methods Of Credit Risk Assessment In Financial Institutions: Implications For Micro, Small And Medium Enterprises (Msmes). http://hdl.handle.net/10361/22883.

Lendingtree. (2025, August). Credit Card Debt Statistics 2025.

Lu, H., & Wu, Z. (2025). Revisiting Intelligent Audit From A Data Science Perspective. *Neurocomputing*, 129431. https://doi.org/10.1016/j.neucom.2025.129431.

Malandreniotis, D. (2024). *Probabilistic Forecasting Models For Multidimensional Financial Time-Series With Applications To Systematic Portfolio Management* (Doctoral Dissertation, Ucl (University College London)).

Mohanarajesh, K. (2024). Investigate Methods For Visualizing The Decision-Making Processes Of A Complex Ai System, Making Them More Understandable And Trustworthy In Financial Data Analysis.

Nagaraj, S. K. S. (2025). A Study On Credit Default Prediction Using Hybrid Ai Models Combining Neural Architectures And Econometric Features. *International Journal Of Emerging Research In Engineering And Technology*, *6*(2), 81-88. https://doi.org/10.63282/3050-922X.IJERET-V6I2P110.

Oyasiji, O., Okesiji, A., Imediegwu, C. C., Elebe, O., & Filani, O. M. (2023). Ethical Ai In Financial Decision-Making: Transparency, Bias, And Regulation. *International Journal Of Scientific Research In Computer Science, Engineering And Information Technology*, *9*(5), 453-471.

Park, M. (2025). Enhancing Esg Risk Assessment With Litigation Signals: A Legal-Ai Hybrid Approach For Detecting Latent Risks. *Systems*, *13*(9), 783. https://doi.org/10.3390/systems13090783.

Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L., & Zhong, C. (2022). Interpretable Machine Learning: Fundamental Principles And 10 Grand Challenges. *Statistic Surveys*, *16*, 1-85. https://doi.org/10.1214/21-SS133

Si, J., Cheng, W. Y., Cooper, M., & Krishnan, R. G. (2024). Interpretabnet: Distilling Predictive Signals From Tabular Data By Salient Feature Interpretation. *Arxiv Preprint Arxiv:2406.00426*. https://doi.org/10.48550/arXiv.2406.00426.

Siraj, M. L., Syarifuddin, S., Tadampali, A. C. T., Zainal, H., & Mahmud, R. (2024). Understanding Financial Risk Dynamics: Systematic Literature Review Inquiry Into Credit, Market, And Operational Risks:(A Long-Life Lesson From Global Perspective To Indonesia Market Financial Strategy). *Atestasi: Jurnal Ilmiah Akuntansi*, *7*(2), 1186-1213. https://doi.org/10.57178/atestasi.v7i2.927.

Smith, H. K. (2025). Balancing Class Distributions For Explainable Ai Models In Lending Decisions.

Source: Ycharts. (2025, July). U.S. Credit Card Accounts Late By 90 Days. Https://Ycharts.Com/Indicators/Us_Credit_Card_Accounts_Late_By_90_Days

Talaat, F. M., Aljadani, A., Badawy, M., & Elhosseini, M. (2024). Toward Interpretable Credit Scoring: Integrating Explainable Artificial Intelligence With Deep Learning For Credit Card Default Prediction. *Neural Computing And Applications*, *36*(9), 4847-4865. https://doi.org/10.1007/s00521-023-09232-2.

Tursunalieva, A., Alexander, D. L., Dunne, R., Li, J., Riera, L., & Zhao, Y. (2024). Making Sense Of Machine Learning: A Review Of Interpretation Techniques And Their Applications. *Applied Sciences*, *14*(2), 496. https://doi.org/10.3390/app14020496.

Wallethub. (2025, July). Credit Card Charge-Off And Delinquency Statistics. Https://Wallethub.Com/Edu/Cc/Credit-Card-Charge-Off-Delinquency-Statistics/25536.

Wallethub. (2025, June). Credit Card Charge-Off And Delinquency Statistics. Https://Wallethub.Com/Edu/Cc/Credit-Card-Charge-Off-Delinquency-Statistics/25536

Wang, H. (2021). Credit Risk Management Of Consumer Finance Based On Big Data. *Mobile Information Systems*, *2021*(1), 8189255.

Wang, Z., & Liang, J. (2024). Comparative Analysis Of Interpretability Techniques For Feature Importance In Credit Risk Assessment. *Spectrum Of Research*, *4*(2).

Yang, H., Li, C., & Zhao, J. (2025). *Credit Scoring Through Interpretable Ensemble Learning*. *Arxiv Preprint Arxiv:2505.20815. https://doi.org/10.48550/arXiv.2505.20815*.

Yates, L. A., Aandahl, Z., Richards, S. A., & Brook, B. W. (2023). Cross Validation For Model Selection: A Review With Examples From Ecology. *Ecological Monographs*, *93*(1), E1557. https://doi.org/10.1002/ecm.1557.