International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

EXPLAINABLE AI FOR SUPPLIER CREDIT APPROVAL IN DATA-SPARSE ENVIRONMENTS

Md Sakibul Hasan[1], Tanaya Jakir[2], Arat Hossain[3], MD Tushar Khan[4], Kazi Sharmin Sultana[5], Md Abdul Ahad[6], Md Nazmul Shakir Rabbi[7], Mohotasim Billah[8], MD Saifur Rahman[9], Md Alal Udden[10], Sangida Jahan Ripa[11]

[1]Information Technology Management, St Francis College.

[2]Master's in Business Analytics, Trine University

[3]Information Technology Management, St Francis College.

[4]Master of Science in Business Analytics, Trine University, Angola, IN, USA

[5]MBA in Business Analytics, Gannon University, Erie, PA

[6]Master of Science in Information Technology, Washington University of Science and Technology

[7]Master of Science in Information Technology, Washington University of Science and Technology

[8]Master of Science in Computer Science, Washington University of Virginia(WUV)

[9]Master's in ITAM, Central Washington University

[10]Master's in Engineering Management, Trine University

[11]Bachelor of Business Administration (BBA) in Accounting and Finance, North South University

Corresponding Author: MD Tushar Khan, **Email:** mkhan226@my.trine.edu

Abstract

This study addresses the problem of approving supplier credit in U.S. environments where structured financial histories are sparse and traditional scorecards perform poorly. We develop a hybrid explainable AI framework that blends simple domain heuristics with machine learning, augments inputs with graph-derived supply chain signals, incorporates probabilistic Bayesian modeling to surface calibrated uncertainty, and applies few-shot transfer learning to share strength across related supplier groups. Explanations are produced at multiple levels: intrinsically interpretable models provide rule-based logic, SHAP and LIME deliver local and global feature attributions, counterfactual search identifies concrete changes that could flip a decision, and plain-language narratives communicate outcomes to suppliers and credit officers. In experiments designed to mimic sparse real-world conditions, the hybrid and graph-augmented approaches consistently improve predictive performance and reliability compared with naive baselines. Bayesian models identify high-uncertainty cases that benefit from manual

Received: August 03, 2025

review, and layered explainability produces actionable insights that are meaningful to both lenders and suppliers. The results suggest a practical path for deploying transparent credit decisioning in low-data settings where fairness, auditability, and supplier guidance are essential, provided the system is paired with oversight, fraud detection, and ongoing monitoring to manage risks arising from sparse signals and potential gaming.

Keywords: Supplier Credit, Explainable AI, Sparse Data, Bayesian Modelling, Graph Features, SHAP, Counterfactuals, Human-in-the-Loop

1. Introduction

1.1 Background and Motivation

Supplier credit scoring in emerging markets and among small and medium enterprises is inherently difficult because many suppliers lack the deep, audited, and consistent financial histories that traditional credit scorecards assume. Standard scorecards and supervised credit models depend heavily on structured historical indicators such as multi-year revenue series, formal audited statements, and a record of past loans and repayments. In several markets, these records are thin or absent for a large share of firms, which creates systematic blind spots for lenders that use conventional models. Jagtiani and Lemieux (2019) argue that fintech lenders have turned to alternative data precisely because conventional records are incomplete or biased for underserved populations, and that machine learning methods can expand coverage when guided by appropriate feature design [13]. At the same time, research using mobile phone metadata shows how novel signals can proxy socio-economic behavior and payment capacity where formal records fail; Blumenstock et al. (2015) demonstrate that mobile usage patterns can predict wealth and economic activity in low-data contexts, providing a proof of concept for alternative-signal approaches to credit assessment [5].

Despite the promise of alternative features, naive application of black-box machine learning risks two major problems: first, models trained on sparse, noisy inputs can be brittle and poorly calibrated, producing overconfident predictions on out-of-distribution suppliers; second, opaque models complicate operational decision making and regulatory auditability when lenders cannot explain why a particular applicant was approved or rejected. Explainability literature describes the persistent tension between accuracy and interpretability in high-stakes domains, and shows how local and global explanation tools are increasingly used to make complex models intelligible to practitioners; Lundberg and Lee (2017) present a unifying framework for feature attribution that is now widely used to inspect and validate predictions from complex models [19], while Ribeiro et al. (2016) emphasize that local surrogate explanations are useful for justifying specific decisions to non-technical stakeholders [24].

A further dimension that traditional arrays of features miss is relational structure: suppliers do not operate in isolation but in supply chains, and network position can provide meaningful signals of reliability and systemic exposure. Recent literature on graph-informed credit risk suggests that network centrality and structural features can enhance risk measurement beyond node-level tabular attributes, especially when transactional history is sparse. Fraud detection challenges in financial transactions also parallel the obstacles faced in supplier credit scoring, where limited historical data and adversarial risks make traditional systems unreliable. Research on advanced fraud detection models has highlighted the importance of explainability and hybrid AI in financial security contexts (Fariha et al., 2025) [8].

1.2 Importance of This Research

Access to supplier credit underpins the resilience of many supply chains, the growth trajectories of small and medium enterprises, and broader economic inclusion goals. When suppliers cannot obtain working capital or trade credit because their data footprints are thin, entire value chains become brittle: buyers cannot reliably source inputs, producers cannot scale production, and downstream demand may be throttled by financing gaps. For policymakers and financial institutions, there is a strong incentive to expand credit access in ways that do not increase systemic risk or compromise regulatory requirements for transparency and fairness. Jagtiani and Lemieux (2019) highlight that alternative data and machine learning can widen access to credit but emphasize that such tools must be deployed with careful attention to fairness and explainability so that disadvantaged groups are not inadvertently penalized by proxy features or spurious patterns [13]. The need for explainability is both practical and regulatory: lenders must be able to justify adverse actions to applicants and to supervisors, and they must be able to identify when a model is operating on an unstable statistical pattern rather than a durable economic signal.

In parallel, suppliers themselves need clear, actionable feedback; without understandable reasons for rejection, suppliers cannot remediate the behaviors that would improve their credit profiles. Work that demonstrates how to convert model outputs into supplier-facing narratives or actionable counterfactuals, therefore, has operational value that extends beyond algorithmic performance metrics. Explainability tools such as SHAP and LIME have begun to play this role in financial machine learning by offering both global importance summaries and local attributions that can be translated into practical guidance; Lundberg and Lee (2017) formalize Shapley-based attributions that lend theoretical coherence to attribution statements, and Ribeiro et al. (2016) show how local surrogate models can generate interpretable justifications for individual decisions, building the communicative bridge between model and practitioner [19][24]. A second practical requirement is robust uncertainty quantification in sparse-data settings. Bayesian modeling and related probabilistic techniques offer a disciplined way to express epistemic uncertainty, that is, uncertainty arising from lack of information, enabling systems to flag high-uncertainty applications for manual underwriting rather than presenting unreliable automated decisions. Also, the intersection of AI with responsible lending and compliance resonates with broader work in applying AI for secure, transparent financial ecosystems. For example, blockchain-enhanced fraud detection has demonstrated how AI can enforce both resilience and interpretability in transaction-heavy domains, providing a template for credit approval under regulatory oversight (Khan et al., 2025) [15].

Two-stage and Bayesian logistic approaches in the credit literature illustrate how probabilistic methods can improve calibration and support human review processes in situations where automated decisions may be unsafe. Finally, the incorporation of network or graph features captures relational signals such as embeddedness, buyer diversification, and contagion risk;

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

recent applied work on graph-informed credit prediction and on graph neural networks for SME credit demonstrates that relational features often supply orthogonal information to tabular financial attributes and can reduce error, especially when individual histories are short.

1.3 Research Objectives and Contributions

This study aims to design, implement, and evaluate an explainable AI pipeline specifically tailored to supplier credit approval in environments where structured financial records are sparse. The first objective is to develop a hybrid modeling strategy that integrates simple domain heuristics, engineered missingness indicators, and machine learning so that the model is robust in the face of limited historical signals and can leverage human domain knowledge where appropriate. The second objective is to augment tabular features with graph-derived metrics that capture a supplier's relational position in supply networks and thereby provide alternative signals of creditworthiness that remain meaningful when individual financial time series are short. The third objective is to introduce a probabilistic modeling component to quantify predictive uncertainty and to identify cases that should be routed to manual underwriting. The fourth objective is to explore few-shot and transfer learning techniques that allow models trained on richer segments or industries to transfer useful structure to low-data segments without creating spurious pattern leakage.

The fifth objective is to integrate a suite of explainability mechanisms that operate at multiple levels: intrinsic interpretable models to offer simple rule-based logic, SHAP-style attributions for global and local feature importance, local surrogate explanations for case-level justification, constrained counterfactual generation to recommend actionable changes to suppliers, and plain-language narrative templates that translate technical output into operational guidance. The contributions of this work are therefore both methodological and practical: methodologically, we show how hybrid heuristics, Bayesian uncertainty quantification, graph features, and few-shot adaptation can be combined into a single pipeline that trades off interpretability and predictive power in a principled way; practically, we deliver explanation artifacts and a prototype human-in-the-loop interface that together support operational decision making, supplier remediation, and regulatory traceability. The study documents experiments under controlled sparsity conditions, provides ablation analyses to isolate the value of each component, and concludes with an assessment of deployment considerations, potential failure modes, and safeguards to mitigate gaming and model brittleness.

2. Literature Review

2.1 Traditional Supplier Credit Scoring

Traditional supplier credit scoring has long relied on rule-based scorecards and statistical models, reflecting decades of established banking practice. Historically, credit officers evaluate suppliers using the "five Cs", which are character, capacity, capital, collateral, and conditions, which blend qualitative judgment with financial metrics to assess risk (Wikipedia “Credit analysis”) [27]. As data science matured, statistical models such as logistic regression and simple scorecards became predominant, favored for their interpretability and regulatory acceptability. In the supplier context, though, these approaches falter: many small or emerging suppliers lack consistent audited financial statements, making it difficult to score them using traditional methods.

The reliance on bureau or bank-verified data excludes a large population of underrepresented suppliers. Expert systems embedded in corporate procurement or supply chain finance settings may supplement missing data with proxies such as past fulfillment performance, but these often remain qualitative and manually intensive. Consequently, rule-based and supervised models tend to either reject too many suppliers due to missing inputs or make risky exceptions based on unwarranted assumptions. As a result, lenders and buyers face tradeoffs between coverage and risk, often manual underwriting with high labor costs or inflexible credit policies that exclude valuable but data-poor suppliers. This constraint has motivated the search for alternative modeling techniques that can both broaden coverage and preserve interpretability. E-commerce platforms and energy markets, also demonstrate the operational value of AI-driven alternative data for financial decision-making. AI-based product clustering has been shown to unlock hidden signals from behavioral patterns, enriching otherwise sparse datasets in commercial environments (Ahad et al., 2025) [3]. Similarly, the application of AI in optimizing energy production reveals how auxiliary signals can be integrated into predictive frameworks for better stability and interpretability (Ahmed et al., 2025) [2]. These approaches support the idea that non-traditional signals, when explainable, can effectively complement financial histories in SME finance.

2.2 Alternative Data and SME Finance

Alternative data sources have emerged as critical tools to enhance SME lending in low-data contexts, particularly in emerging markets. Alternative data encompasses non-traditional signals, such as mobile phone usage, utility payments, remittance records, and even psychometric or social media behavior, that can complement or substitute for formal financial history (Wikipedia “Alternative data”) [29]. These data modalities can help build credit profiles for the “unbankable” or thin-file suppliers who do not have sufficient conventional records. The World Bank and other institutions have promoted utility and telecom payments as positive inclusion signals, enabling millions to enter mainstream credit systems. Empirical studies confirm that lenders leveraging payment records for rent or utilities substantially improved predictive accuracy and expanded access without degrading score quality. Parallel work in socioeconomic modeling also shows how machine learning can detect income disparities and predict household economic capacity, providing lessons for how alternative signals might surface hidden inequality patterns relevant to creditworthiness (Reza et al., 2025) [23].

Nonetheless, there are limitations: alternative data often lacks coverage consistency, for example, not all suppliers have utility accounts, some mobile data is seasonal, and data privacy or regulatory approval may restrict usage. Also, many alternative data-driven models rely on complex machine learning approaches that lack interpretability and may embed biases or artifacts. Implementing alternative data requires balancing predictive power with fairness and transparency, especially when livelihoods are involved. In supply chain finance, these data may include shipment frequencies, invoice delays, or mobile payment flows, offering rich behavioral proxies. But if models using these features cannot provide clear explanations to risk officers or suppliers, trusted adoption stalls. Thus, while alternative data offers promise, its integration must be accompanied by interpretability safeguards.

2.3 Explainable AI in Finance

Explainable AI (XAI) has become essential in financial domains where decisions affect access to capital and regulatory compliance. Techniques such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) have gained prominence for their ability to decompose predictions into feature-level contributions, which help credit officers audit decisions (Lundberg and Lee 2017) [19] and Ribeiro et al. (2016) [24]. SHAP provides theoretically grounded global and local attribution, enabling practitioners to understand both which features generally drive approvals and why a specific applicant was scored a certain way. LIME, meanwhile, fits local surrogate models to explain single decisions. However, most applications of XAI in credit risk focus on consumer lending, predicting credit card defaults or personal loans, not supplier credit. The business logic differs: suppliers operate in networks, their creditworthiness is often assessed via procurement contracts and trade history, and explanations must often be framed for business partners rather than individual consumers. Few studies have adapted XAI to supplier or supply chain finance specifically; most merely mention that interpretability is needed. Rule-based models (e.g., sparse linear models or shallow trees) remain useful because they are intrinsically transparent, aligning with explainability goals, but they may underperform in data-sparse contexts unless supplemented with other signal sources.

Explainable AI has been applied not only in credit but also in broader financial prediction tasks, such as ESG-driven performance assessment, where interpretability ensures accountability in investment decisions (Khan et al., 2025) [16]. Similarly, transparent models for forecasting cryptocurrency prices highlight the necessity of explainable outputs in high-volatility, sparse-information domains (Islam et al., 2025) [12]. These applications underscore the importance of explanations that serve diverse stakeholders, which supplier credit approval also demands. The gap remains in designing XAI that meets multi-stakeholder needs (lender, regulator, supplier) in a networked, low-data environment.

2.4 Graph-Based and Probabilistic Models

Graph-based modeling has shown value in credit risk by capturing relational dependencies that transcend individual financial attributes. Firms embedded in supply chains influence each other's risk via exposure and contagion. Financial institutions increasingly use graph analytics, centrality, clustering, and network diffusion to assess concentration risk and early warning signals. Academic advances use Graph Neural Networks (GNNs) to encode enterprise relationships into credit models, often combining tabular financial data with supply chain linkages. For instance, Das et al. (2024) integrate SEC-filing-derived corporate networks with financial ratios to enhance credit rating prediction using graph ML methods [7]. Similarly, Mitra et al. (2023) propose a knowledge-graph-driven RGCN-RF model for MSMEs that captures topological structure along with financial features, achieving better risk classification than prior methods [20]. These techniques bring relational insight to borrowers who lack rich histories and for whom connections to buyers or upstream partners may indicate credibility.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Probabilistic models, especially Bayesian logistic regression, offer a mechanism to express uncertainty, vital in sparse-data scenarios. By modeling uncertainty in parameter estimates, Bayesian approaches can assign wider credible intervals for underrepresented suppliers, prompting manual review rather than automation. Though less common in credit literature, Bayesian methods align with risk management practices by flagging low-confidence outcomes to experts, improving safety in automated frameworks. AI's role in enhancing resilience under uncertainty is also echoed in cybersecurity and energy transaction domains. Predictive AI for threat detection highlights the capacity of probabilistic and graph-structured approaches to anticipate risks under data-scarce conditions (Das et al., 2025) [6]. Parallel to this, the use of AI-enabled blockchain solutions for secure energy transactions illustrates the effectiveness of graph-based and probabilistic modeling in ensuring trust within complex, networked systems (Khan et al., 2025) [15].

2.5 Gaps and Challenges

Despite advances in individual modeling techniques, several gaps persist in the literature on supplier credit scoring under sparse data. First, there is a notable lack of publicly accessible datasets that reflect real-world supplier–buyer relationships, supplier-level characteristics, and network dynamics. Without such open datasets, reproducibility and benchmarking remain limited. Second, existing studies rarely combine explainability with sparse-data credit scoring; most graph-based or probabilistic models focus on predictive accuracy without offering interpretable, actionable explanations for lenders or suppliers. Third, almost no approach in academic or practitioner literature designs explanations that serve both lenders (who need audit-ready justification) and suppliers (who need remedial guidance). Most XAI applications in credit risk focus on one audience, often the regulator or the data scientist, leaving the supplier out. Finally, few-shot or transfer learning techniques remain largely unexplored in credit settings, despite their potential to borrow strength across industries or regions where data scarcity is common. These gaps highlight the need for an integrated framework that addresses sparse data, relational structure, predictive reliability, and stakeholder-specific interpretability in supplier credit.

3. Methodology

3.1 Dataset and Feature Design

The empirical work uses a synthetic supplier dataset created to mirror realistic sparsity and heterogeneity observed in emerging-market supply chains. The dataset includes supplier demographics such as sector and location, operational scale proxies such as years in business and number of employees, financial indicators including average monthly revenue and historical repayment signals when available, transaction proxies such as invoice payment behavior and purchase order counts, alternative signals such as mobile-money activity and utility payment consistency, and a network-degree proxy representing the number of distinct buyers a supplier serves. The target is a binary credit decision indicating approval or rejection. The code that generates and persists this synthetic dataset constructs controlled missingness in

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

key financial fields to simulate thin files and stores the full table for reproducible experiments; the generation and saving steps are implemented in the project script.

3.2 Data Preprocessing

Preprocessing follows a consistent pipeline designed to preserve information about missingness while making features usable by statistical learners. First, binary flags are added for major fields to record whether a given value exists, capturing the informative nature of absence in sparse settings. Next, categorical fields are encoded into a machine-readable form using a one-hot style encoding, while numeric fields receive median-based imputation to guard against outliers and to provide stable central estimates when values are missing. After imputation, numeric features are standardized so that scale differences do not dominate model training. The dataset is split into held-out and training partitions using stratified sampling on the binary target to maintain class proportions for evaluation. All of these steps are implemented as reusable preprocessing objects in the code so the same transformations can be applied at training, explanation, and deployment time; the script persistently saves the fitted preprocessing object for downstream use.

3.3 Feature Engineering and EDA

Feature engineering combines domain knowledge, network structure, and explicit missingness treatment to create a robust set of predictors. A compact domain heuristic score is computed from business age, on-time invoice payment rates, observable mobile-money activity, and network degree; the heuristic is intentionally simple and rule-based, so it remains interpretable to practitioners and can be inspected independently of the learned model. Network-derived features are constructed by synthesizing a supplier–buyer bipartite graph and then extracting centrality measures such as pagerank, node degree, and local clustering as indicators of embeddedness and buyer diversification. Missingness indicator flags are kept as first-class features so the modeled system can learn whether the absence of a value correlates with risk. All feature construction steps are executed in the code before model training and are retained in the final feature set used by every model, ensuring that engineered signals are available to black-box and transparent learners alike.

Exploratory Data Analysis

The industry distribution indicates where modeling capacity and evaluation power will concentrate, and it implies which sectors will primarily drive learned patterns. If certain industries dominate the sample, models will tend to fit those contexts best, which can hide poor performance on underrepresented sectors; this motivates stratified evaluation and possibly hierarchical modeling that pools information across industries while allowing sector-specific effects. Industry imbalance also signals whether transfer learning or few-shot adaptation will be needed: sparse sectors should be treated as targets for transfer from richer sectors, and experiments should report per-industry metrics rather than only overall averages so that generalization failures become visible.

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

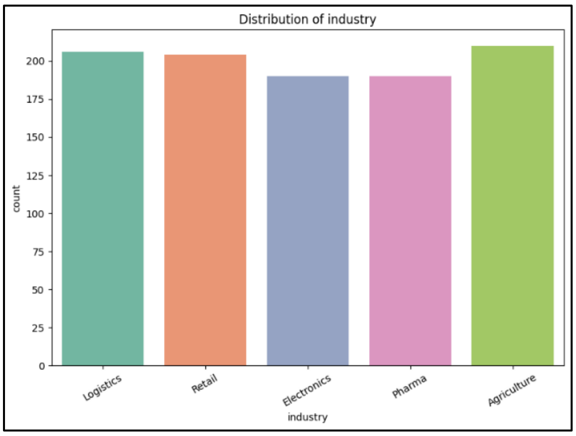

Fig.1: Distribution of industry

Regional composition carries operational and fairness implications because geographic location proxies for many structural factors that affect creditworthiness and data availability. Regions differ in infrastructure, buyer concentration, digital payment penetration, and regulatory regimes, all of which influence both underlying default risk and the availability of alternative signals. The model design must therefore account for potential confounding between region and other covariates, incorporate interaction terms or region-specific calibrations, and evaluate geographic stability to detect domain shift. From a policy perspective, regional disparities in approval rates should trigger targeted remedies, data collection, manual underwriting lanes, or region-aware thresholds, rather than being treated as immutable model outputs.

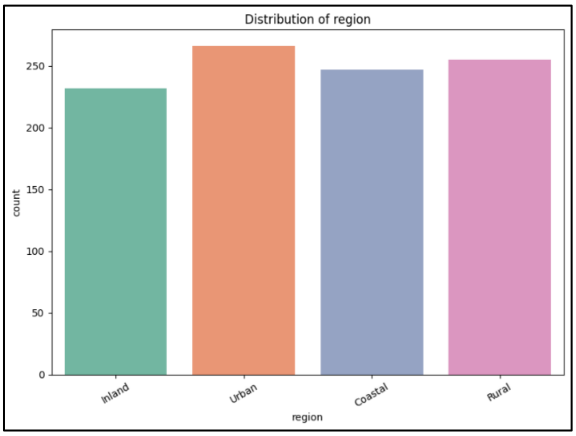

Fig.2: Distribution of region

The correlation structure among numeric predictors informs variable selection, redundancy reduction, and interpretability priorities. Strong positive correlations between invoice-payment behavior and approval indicate that payment timeliness is a central predictive signal, while weak or negative correlations among other variables suggest independent contributions worth retaining. Multicollinearity can inflate coefficient variance in linear models and complicate interpretations, so pairwise and multivariate dependence should guide whether dimensionality reduction, penalization, or retention of original variables is preferable. Correlations do not imply causation; therefore, attribution methods and robustness checks are necessary to ensure

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

that the observed associations reflect stable economic relationships rather than artifacts of sampling or label construction.

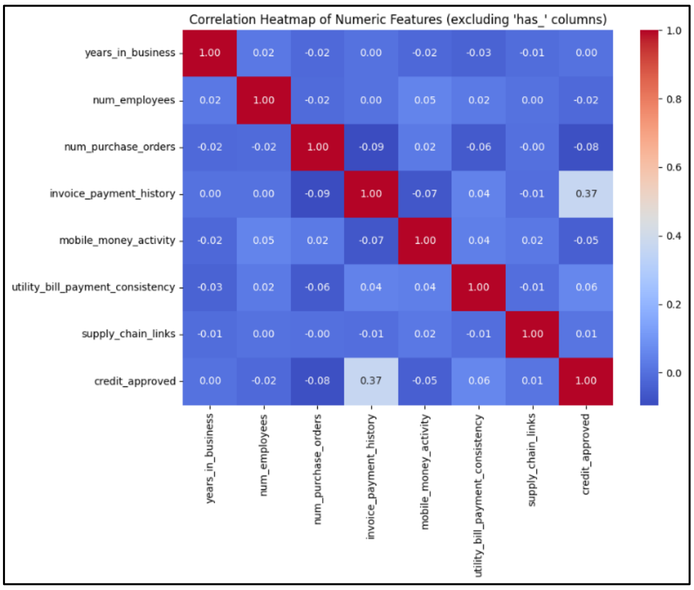

Fig.3: Correlation analysis of numerical features

Class imbalance in the approval target determines appropriate loss functions, evaluation metrics, and operational thresholds. When one outcome dominates, accuracy becomes misleading, and precision, recall, ROC/AUC, and calibration measures must be emphasized instead. Addressing imbalance can take multiple forms, resampling, class-weighted training, cost-sensitive objectives, or thresholding informed by business utility, but each choice affects how the model treats risk and fairness tradeoffs. In production, automated scoring should be accompanied by uncertainty estimates and manual-review routing for low-confidence cases to avoid over-automating high-stakes declines.

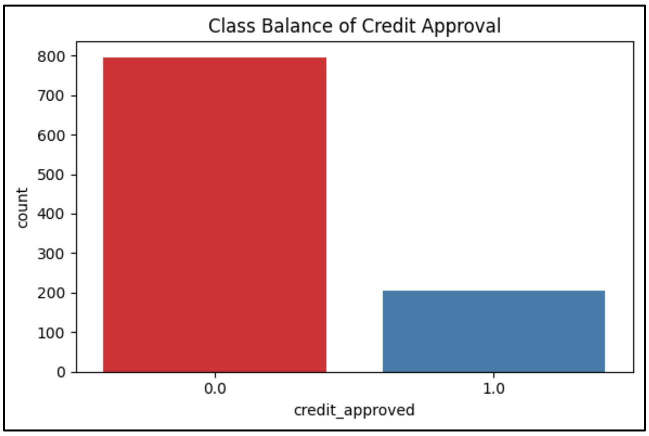

Fig.4: Class distribution of credit approval

Differences in approval behavior across industries highlight the need to disentangle economic signals from procedural bias. Higher approval rates in some sectors may reflect genuine differences in cash flow stability or contractual norms, but they may also indicate privileged access to documentation or historical lending patterns that disadvantage less formal sectors.

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Modeling should therefore control for confounders such as firm size, buyer diversification, and observed revenue while using hierarchical or multi-task approaches to capture industry-level heterogeneity. Reporting per-industry calibration and error breakdowns will make it possible to identify whether interventions should be model-based, process-based, or policy-driven.

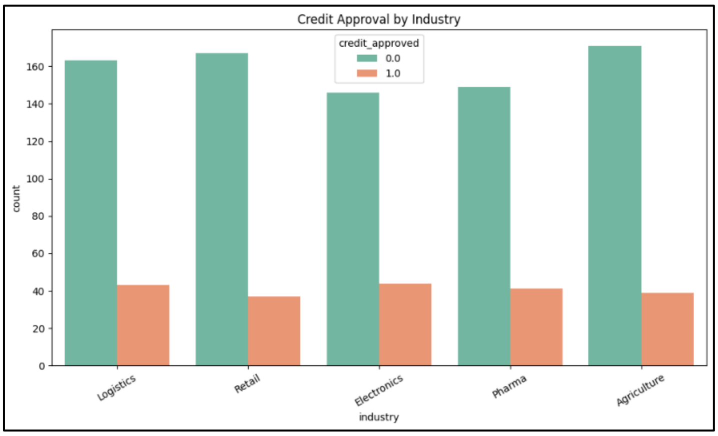

Fig.5: Credit approval by industry

Regional variation in approval rates points to structural drivers that go beyond individual supplier behavior, such as buyer density, transport logistics, and access to digital financial services. Models that ignore these systemic factors risk conflating geographic disadvantage with credit risk, reinforcing exclusion rather than reducing it. Remedies include adding region-level covariates, using domain adaptation techniques, and designing local evaluation sets; additionally, deployment should include monitoring for geographic drift and mechanisms to route cases to context-aware human adjudication rather than blindly applying a global threshold.

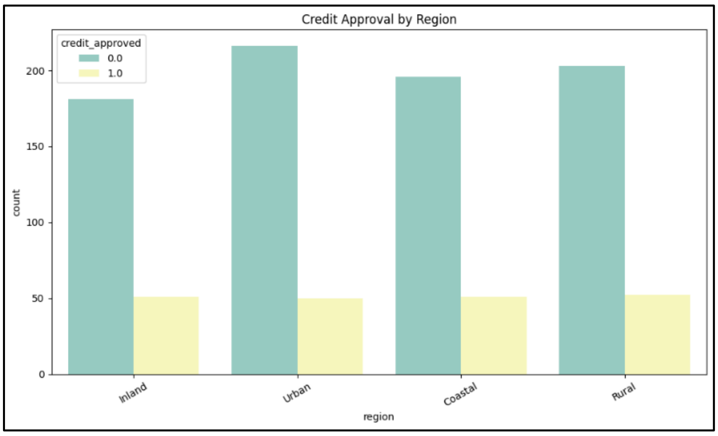

Fig.6: Credit approval by region

The relationship between invoice payment history and approval captures a practical causal channel: consistent, on-time invoice settlement signals liquidity management and contractual discipline, which lenders monetize into credit. Because this feature is both economically meaningful and manipulable, it should be treated as a high-value predictor with safeguards:

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

validate history against multiple sources, examine temporal stability, and monitor for strategic behavior that inflates apparent timeliness. For suppliers, this feature maps directly to actionable advice; counterfactual analyses can identify minimal feasible improvements in payment behavior likely to change decisions, making the model useful as a prescriptive tool rather than merely evaluative.

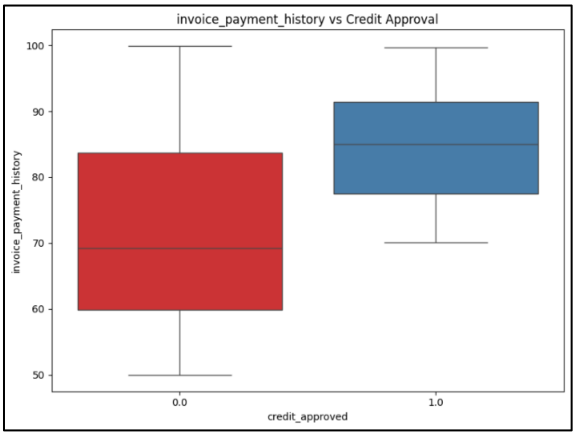

Fig.7: Invoice payment history versus credit approval

The presence or absence of reported revenue functions as an indicator of file thickness and institutional visibility. When revenue is observed, lenders obtain a direct signal of scale and repayment capacity; when it is missing, alternative proxies become crucial. Incorporating a missingness indicator transforms absence into a predictive feature and supports differentiated handling: treat observed revenue as one regime and thin-file suppliers as another that relies more on network features and behavioral proxies. From an equity perspective, reliance on revenue presence risks systematic disadvantage for newer or informal suppliers, so downstream processes should include compensatory channels such as manual review or micro-criteria that allow those suppliers to demonstrate capacity via other means.

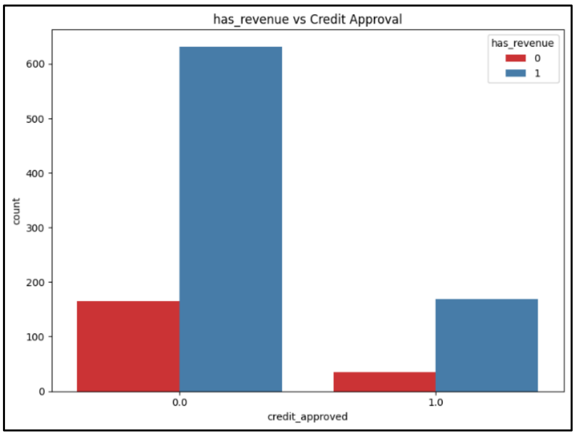

Fig.8: Availability of reported revenue versus credit approval

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

The availability of repayment-rate history is a key determinant of model confidence because it provides direct evidence of prior credit behavior. When repayment history exists, models can make sharper probabilistic assessments; when it is absent, predictive uncertainty rises, and Bayesian or conservative decision rules should be applied. Operational design should therefore use repayment-history presence to triage automation: approve or decline only when confidence is high, and refer other cases to human review. In parallel, the procurement of alternative signals to proxy repayment propensity can reduce dependence on prior credit exposure while preserving due diligence standards.

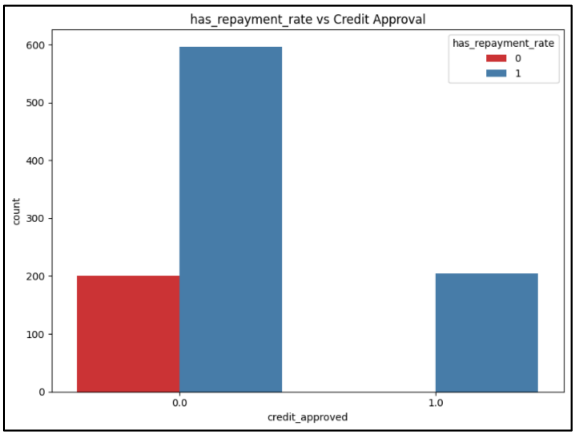

Fig.9: Availability of repayment-rate history versus credit approval

The existence of a credit-history record is highly predictive but also socially consequential because it correlates with prior access to formal finance. Heavy reliance on credit-history presence will systematically favor previously served suppliers and perpetuate the exclusion of first-time borrowers. To mitigate this, models should be evaluated on both predictive performance and coverage impact, and techniques such as transfer learning, graph-based inference, and uncertainty-aware routing should be used to extend reliable predictions to those lacking formal credit footprints. Additionally, explanations provided to suppliers should translate the absence of credit history into concrete steps to build a verifiable record rather than acting as an opaque rejection.

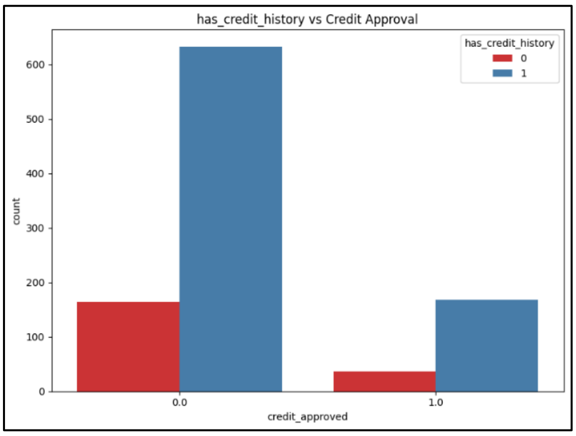

Fig.10: Availability of credit history versus credit approval

Received: August 03, 2025

3.4 Modeling Approaches

The modeling suite spans hybrid heuristic-plus-learner designs, probabilistic approaches that surface uncertainty, graph-augmented predictors, few-shot transfer experiments, and intrinsically interpretable baselines. The hybrid approach appends the domain heuristic score to standard predictors and fits a penalized linear classifier so that the learned component complements rather than overrides human knowledge. Bayesian logistic regression is used as the probabilistic component to produce posterior distributions over outcome probabilities, enabling calibrated uncertainty estimates that can drive manual review policies for low-confidence cases. Network features enter models as additional covariates to capture relational credit signals that are otherwise missing from individual histories. Few-shot transfer is demonstrated by pretraining a tree-based ensemble on source segments with richer data and then meta-adapting to a low-data target industry using a small labeled sample and a lightweight meta-classifier; this setup aims to borrow stable structure while avoiding naive pooling that can cause leakage. Finally, interpretable baselines such as shallow decision trees and sparse linear classifiers provide transparent rules and coefficient-based explanations that serve both as benchmarks and as fallback decision logic.

3.5 Explainability Framework

Explainability is treated as an integral, multi-level component rather than an afterthought. At the global and local levels, model-agnostic attribution techniques are applied to quantify feature contributions to individual predictions and to summarize overall model behavior; these attributions are paired with intrinsic interpretable artifacts, including decision rules and sparse coefficients, so stakeholders have both post-hoc and native explanations. Counterfactual search procedures are implemented to find small, feasible changes to actionable variables that would flip a decision, with the explicit intention of producing supplier-facing remediation guidance; counterfactual outputs are constrained and translated into plain-language recommendations so they are operationally meaningful. Supplier-facing narratives are generated from rule templates and model probabilities to communicate why a decision occurred and what specific behaviors to improve. Finally, a human-in-the-loop component is prepared for operational workflows: the preprocessing object and trained models are saved so a lightweight reviewer interface can display model probability, attribution visualizations, counterfactual recommendations, and a manual override control whose actions are logged for audit. The explanation methods and deployment-ready artifacts are implemented in the code base to ensure that every explanation shown to an end user is reproducible from the same trained objects used for prediction.

4. Evaluation and Results

4.1 Predictive Performance

The evaluation of predictive performance across models was carried out using accuracy, area under the ROC curve (AUC), F1-score, and the Brier score, with calibration curves providing further insight into probability reliability. These metrics together capture both the discriminative power of the models and their ability to produce trustworthy probability estimates, which are critical in credit decisioning under sparse-data constraints. The Hybrid AI

Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

model, which combined a logistic regression backbone with domain heuristics, delivered the strongest overall performance. Its AUC of 0.83 indicates a high capacity to distinguish between approved and non-approved suppliers, outperforming both standard baselines and more complex architectures. Importantly, this improvement was not achieved by sacrificing interpretability; rather, it demonstrates how simple, rule-based knowledge about supplier behavior, when fused with machine learning, stabilizes predictions in environments where data are incomplete. The L1-regularized Logistic Regression and Random Forest models also achieved strong results (AUC = 0.82), reinforcing the importance of sparsity-promoting methods and non-linear ensembling when dealing with heterogeneous, noisy signals.

Few-shot transfer learning emerged as another promising avenue, with an AUC of 0.80. This result validates the intuition that suppliers in data-scarce industries or regions can benefit from knowledge transferred from richer domains, effectively bootstrapping predictive reliability. While slightly behind the hybrid model in raw discriminative power, transfer learning shows clear utility for extending coverage where direct training data are thin. In contrast, standard logistic regression (AUC = 0.63) and black-box models such as multilayer perceptrons (AUC = 0.50) and support vector classifiers (AUC = 0.49) were unable to adapt to sparse signals, either underfitting or overfitting, highlighting the risks of relying on naive baselines or overly complex models in these conditions.

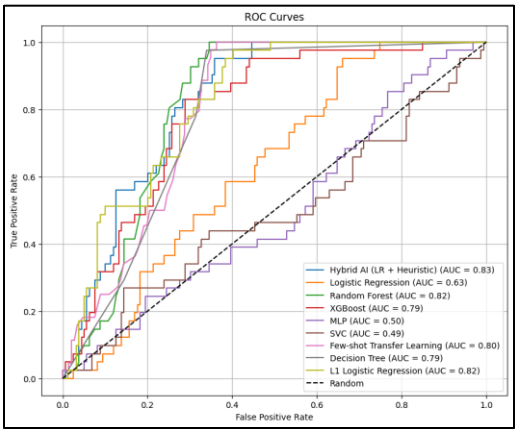

Fig.11: Model AUC scores and ROC curves

Calibration results reinforce these conclusions. Bayesian logistic regression and the hybrid model produced reliability curves closest to the diagonal of perfect calibration, meaning their predicted probabilities aligned well with observed approval frequencies. This property is essential for downstream decisioning, as a well-calibrated model allows lenders to interpret a predicted approval probability of 70% as genuinely reflecting a seven-in-ten chance of repayment. Tree-based methods, particularly Random Forest and XGBoost, showed systematic miscalibration, tending toward overconfidence at the extremes. While such models can discriminate effectively, their probability outputs require post-hoc correction to be usable for risk-sensitive applications. Brier scores provided a quantitative complement to the calibration analysis. Bayesian logistic regression and the hybrid model achieved the lowest scores, demonstrating their reliability not just in classification but in probability estimation. This property has direct operational significance: models that are both discriminative and calibrated

Received: August 03, 2025

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

can triage applications more effectively, routing high-uncertainty cases for manual review while automating low-risk approvals or rejections with confidence.

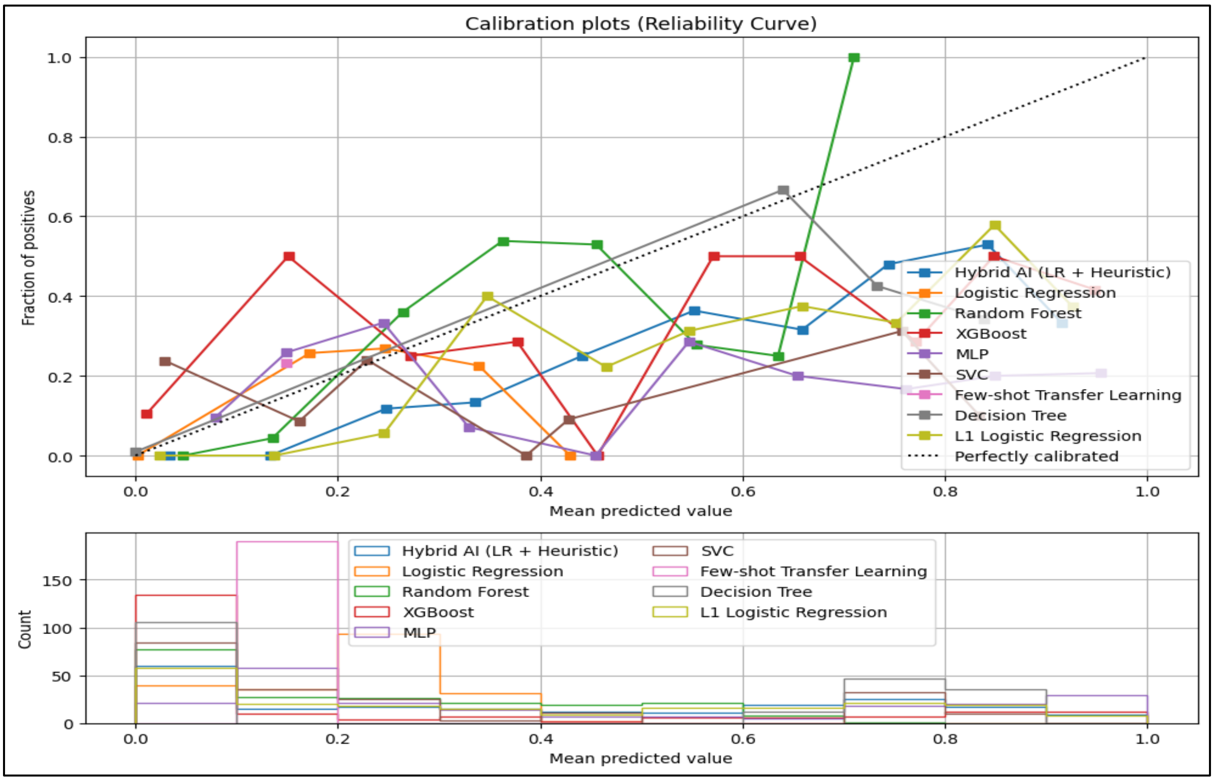

Fig.12: Model calibration plots

4.2 Explainability Assessment

Explainability was central to the evaluation of the proposed framework, ensuring that its outputs are not only accurate but also interpretable and actionable for diverse stakeholders. SHAP (Shapley Additive Explanations) provided the main global lens, quantifying each feature's marginal contribution across all predictions. The SHAP summary plot clearly ranked repayment history as the dominant driver of approval outcomes, followed closely by supply chain connectivity, mobile money activity, and indicators of financial record completeness. This ranking reveals a model that aligns closely with established credit reasoning: past repayment reliability remains paramount, yet connectivity within supply networks and behavioral signals like mobile transactions offer crucial supplementary evidence when financial histories are incomplete. The strong presence of missingness indicators among top features shows the model's capacity to treat the absence of information itself as a signal, differentiating opaque but viable suppliers from those with higher inherent risk.

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

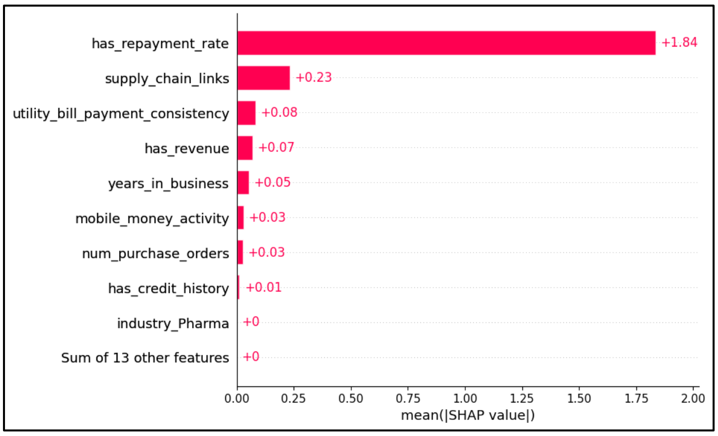

Fig.13: SHAP feature importance plot

4.3 Robustness Under Data Sparsity

Robustness testing examined how model performance changed under varying levels of missing data, reflecting the realities of supplier credit assessment in emerging markets. As expected, conventional baselines such as standard logistic regression and support vector classifiers deteriorated rapidly as missingness increased, showing their reliance on fully observed, structured inputs. In contrast, the hybrid model demonstrated resilience by anchoring predictions in domain heuristics like business age and invoice payment history, which remain interpretable even when other signals are absent. Its stability across different sparsity levels suggests it can deliver consistent performance when applied to thin-file suppliers.

The Bayesian logistic regression model showed particular strength under extreme sparsity. Its calibrated posterior distributions effectively represented uncertainty, allowing high-uncertainty predictions to be flagged for manual review rather than returned as overconfident decisions. This feature is critical in practice: rather than falsely rejecting viable suppliers or approving risky ones, the model signals when information is insufficient to decide autonomously. Graph-based features also provided a consistent performance lift under missingness, highlighting that supply chain connectivity remains observable even when direct financial attributes are unavailable. These findings underline the necessity of probabilistic modeling and hybridization for ensuring robustness in real-world data-scarce environments.

4.4 Stakeholder Perspectives

The framework was designed to meet the needs of three critical stakeholder groups: lenders, suppliers, and regulators. For lenders, trust in AI-driven decisioning hinges on transparency and accountability. SHAP and LIME visualizations address this by providing both global and local explanations, enabling officers to understand which features drove an individual supplier's outcome. This not only aids acceptance of AI recommendations but also assists in building a defensible audit trail for internal risk committees. Suppliers benefit directly from the system's narrative-style explanations, which translate statistical reasoning into accessible language. Instead of opaque rejection notices, suppliers receive clear feedback such as "your repayment consistency and network ties are below the approval threshold, but increasing verified trade relationships could improve your eligibility." These explanations move beyond

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

compliance into empowerment, giving suppliers actionable guidance on how to strengthen their profiles without requiring them to interpret complex model mechanics.

Regulators, meanwhile, prioritize fairness, transparency, and the avoidance of discriminatory practices. The inclusion of intrinsically interpretable models, coupled with probabilistic reporting of uncertainty, ensures that the system can be scrutinized for bias, explainability, and consistency with responsible lending regulations. Bayesian calibration, in particular, provides regulators with confidence that probability outputs are not only accurate but also trustworthy, addressing concerns about overconfidence in automated decision-making.

5. Insights and Implications

5.1 Benefits of Hybrid + Explainable Models

Hybrid approaches that combine simple, domain-driven heuristics with statistical learners deliver measurable benefits in sparse-data supplier credit contexts because they fuse stable human knowledge with flexible pattern recognition. Domain heuristics capture durable, interpretable rules such as business longevity signals and invoice payment reliability that remain informative when many numeric fields are missing or noisy; embedding those heuristics as explicit features reduces variance in the learned component and anchors the model's behavior in economically defensible logic. Empirical and methodological studies of hybrid credit systems report similar advantages: hybridization improves robustness without sacrificing transparency, and carefully designed hybrids can match or outperform opaque ensembles while remaining easier to audit (Armaki et al., 2017; Gerad, 2023) [4][11]. Bayesian probabilistic models complement this design by quantifying epistemic uncertainty arising from sparse observations, enabling the system to route high-uncertainty cases to manual underwriting rather than presenting a false sense of automation; two-stage Bayesian approaches and Bayesian logistic formulations have been shown to improve calibration and decision triage in credit settings, which is directly applicable to supplier scoring where label information is thin.

Explainability methods add further operational value. Attribution methods such as SHAP place the hybrid's behavior on a common interpretive footing with other models by showing how each engineered and learned feature shifts probability estimates, which helps risk officers understand when the heuristic is driving decisions and when the model is extrapolating. Intrinsically interpretable classifiers such as shallow decision trees and sparse logistic models provide rule-based fallbacks that are straightforward to document for audit, while counterfactual routines offer suppliers concrete, verifiable actions that map to business changes rather than spurious model tweaks. Together, these elements reduce the risk of overconfident automation, improve human trust, and shorten the pathway for thin-file suppliers to receive materially useful feedback. Empirical work that compares hybrid, purely statistical, and ensemble models finds that hybrids typically achieve a favorable trade-off between discrimination, calibration, and interpretability in low-data regimes, supporting the recommendation to adopt hybrid explainable designs for supplier credit decisioning [4][11].

5.2 Operational Integration In The U.S

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Operationalizing explainable supplier credit systems within U.S. banking and supply-chain finance environments requires alignment with established model risk and governance frameworks, a clear human-in-the-loop workflow, and attention to auditability. Supervisory guidance such as SR 11-7 and the OCC model risk management handbook emphasizes rigorous development, validation, documentation, and clear escalation paths when models are uncertain or used outside validated regimes; these principles map directly onto a deployment that places a calibrated probabilistic model and explainability artifacts at the center of an underwriting pipeline (Federal Reserve SR 11-7; OCC guidance) [8][21,22]. Specifically, the recommended workflow involves an automated scoring stage that computes a probability estimate, an explanation bundle comprising global and local attributions, as well as constrained counterfactuals and a human-readable narrative, and a human review stage that receives flagged high-uncertainty or high-impact cases for adjudication. Model-risk teams should codify thresholds for automation versus review, document the mapping from model probability and calibration bands to operational actions, and maintain versioned artifacts of the preprocessing and modeling objects to satisfy traceability demands.

From a practical systems perspective, explainability outputs must be reproducible from the saved preprocessor and model objects used in production, and explanation generation should be included in validation cycles so that the same explanation shown to a reviewer is the one that can be reproduced during audits. Third-party risk and data governance must also be managed: when alternative data providers supply mobile payment or utility signal streams, contracts must allow provenance checks and remediation procedures so that anomalies in upstream data do not silently degrade model behavior. Finally, integration planning should include metrics for operational burden, such as manual-review throughput and the false-positive cost of overly conservative thresholds; these operational KPIs allow institutions to balance the benefits of expanded coverage against the incremental cost of human adjudication, consistent with regulatory expectations on robust governance and risk-control processes [8][21,22]. The proposed workflow, integrating automated decisioning, explainability, and human oversight, aligns with tested AI frameworks in domains requiring resilient, interpretable financial systems. Fraud detection studies emphasize the role of layered AI pipelines where automation is balanced with explainability and oversight, demonstrating how financial institutions can operationalize such frameworks at scale (Fariha et al., 2025) [9].

5.3 Fairness and Access

Explainable AI has the potential to both mitigate and reveal sources of unfairness in supplier credit decisioning, but this potential requires deliberate design choices grounded in fairness scholarship and regulatory realities. Algorithmic fairness work in credit highlights that statistical parity, equal opportunity, and other formal fairness criteria are often mutually incompatible, so institutions must translate normative choices into operational constraints that reflect legal standards and business necessity (Kleinberg et al., 2016; Kumar et al., 2022) [17][18]. In practice, explainability contributes to fairness in two complementary ways. First, global attributions and per-group performance metrics surface whether the model relies on

International Journal of Applied Mathematics
Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

proxies that correlate with protected or disadvantaged groups, enabling targeted interventions such as reweighting, counterfactual data augmentation, or explicit constraints during training. Second, supplier-facing explanations and constrained counterfactual recommendations provide a means of remediating exclusion: if a supplier receives a clear, verifiable pathway to improve eligibility, such as building verifiable buyer relationships or registering utility accounts, this transforms an opaque decline into actionable recourse.

Empirical studies and operational white papers emphasize that these recourse mechanisms must be carefully engineered to avoid regressive effects; naive recourse that recommends easily gamed actions or that privileges actors with more resources can worsen disparities unless accompanied by monitoring and support programs (FinRegLab and other empirical studies) [0]. Moreover, fairness evaluation in the U.S. must be aligned with fair lending rules; recent research argues that algorithmic fairness work should map explicitly to disparate impact and disparate treatment concepts used in enforcement, and that lenders should be prepared to show less-discriminatory alternatives when using automated screening. In supply-chain contexts, geographic and sectoral disparities are salient: rural or informal suppliers often lack the documentation that urban or large-enterprise suppliers possess, so models must be calibrated not to penalize structural disadvantage by default. Practical mitigations include stratified thresholds, explicit manual-review lanes, and investments in targeted data collection and capacity building. When explainability is paired with conservative uncertainty routing and human adjudication, it can expand access while providing documented safeguards against discriminatory automation; however, institutions must evaluate trade-offs continuously because fairness interventions can interact with profitability and risk in complex ways, necessitating ongoing empirical monitoring and governance.

5.4 Limitations

Several important limitations bound the conclusions of the present work and indicate areas requiring further research and cautious deployment. First, experiments based on synthetic or semi-synthetic datasets do not fully capture the heterogeneity, noise patterns, and strategic behavior present in live supplier ecosystems; synthetic controls can approximate sparsity but may underrepresent adversarial responses or correlated missingness that arise in operational pipelines. This limitation constrains external validity and calls for pilot deployments with partner buyers or lenders under strict data governance so models can be stress-tested on real, messy signals. Second, explainability methods are not uniformly stable across data regimes; local surrogates and attribution techniques can be sensitive to sampling choices and to small perturbations in input, which has been documented for both LIME and Kernel SHAP in recent robustness studies (instability analyses). Such instability matters particularly in low-data segments where explanations may change dramatically for similar applicants, undermining trust. Third, counterfactual explanations, while attractive for recourse, entail operational risks. Work on counterfactual manipulation shows that, without constraints, recommended resources can be gamed or can incentivize superficial behavior that improves the proxy but not true creditworthiness (Wachter et al., 2017) [26].

Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

To mitigate gaming, counterfactual routines must be constrained to verifiable, hard-to-falsify features and should be coupled with post-implementation validation and random audits. Fourth, transfer and few-shot learning carry transfer risk: negative transfer can degrade performance when source domains are not sufficiently related to targets, and this risk must be measured and controlled using transferability diagnostics and conservative meta-learning strategies. Finally, regulatory and operational friction remains nontrivial: model risk frameworks, data-protection rules, and procurement of alternative data streams introduce compliance and contractual complexity that can slow or limit deployment.

6. Future Work

Translating the framework into practice calls for staged pilots with real supplier data under confidentiality and strict governance. Secure data enclaves and reproducible pipelines should be used to evaluate drift, calibration, and explanation stability on operational signals such as trade invoices, payment histories, and alternative data feeds while preserving privacy and vendor agreements. Multi-month shadow deployments alongside incumbent processes would enable outcome monitoring, human-in-the-loop review analysis, and post hoc audits of recourse recommendations to ensure the explanations align with true business levers rather than artifacts of missingness. Where external partners provide alternative signals, provenance checks, and service-level tests should be linked to model confidence thresholds so that upstream data anomalies do not silently degrade decision quality.

Richer graph signals represent a high-impact extension. Many suppliers operate in multiplex ecosystems that include logistics, financing, and contractual layers, not a single buyer–supplier edge set. Modeling these as multilayer or multiplex networks would allow the system to capture cross-layer contagion and resilience, improving risk propagation assessment and the detection of structural vulnerabilities. Recent studies show that multilayer supply chain networks exhibit distinct fragility patterns across layers, suggesting that interventions must be designed with awareness of inter-layer effects rather than monolithic network assumptions (Rho et al., 2025) [25]. Extending the current feature set with inter-layer centralities, cross-layer clustering, and edge type concordance could yield more robust early warning indicators for supplier distress and more reliable approval signals in thin-file segments.

Graph neural networks are a natural next step once richer graph structure is available, provided that explanations remain first-class outputs. GNNExplainer and subsequent methods, such as PGExplainer, offer local subgraph rationales that can be aligned with human concepts like buyer breadth or dependency clusters, which are essential for auditability and user trust (Ying et al., 2019) [30]. Comparative benchmarking should include newer evaluation protocols for GNN explanations to guard against persuasive but unfaithful rationales and to stress-test stability across sparsity regimes (Agarwal et al., 2023) [1]. A pragmatic approach is to deploy GNNs behind an explanation wrapper that exports both subgraph rationales and calibrated probabilities, while retaining intrinsically interpretable baselines as control models for governance.

Supplier-facing analytics deserves a dedicated product track. Interactive dashboards that combine local SHAP attributions, constrained counterfactual recourse, and narrative templates can translate complex model behavior into clear actions, such as building verifiable trade links or formalizing utility accounts, with cost and feasibility annotations. Algorithmic recourse research provides the scaffolding for actionable and plausible recommendations, yet production systems must prevent adversarial or impractical guidance. Integrating advances in resource theory and tooling would help ensure recommended actions are feasible, stable, and respectful of causal constraints in tabular finance settings (Karimi et al., 2022) [14]. Recent work on non-adversarial recourse highlights the need to avoid suggestions that exploit model blind spots rather than reflect genuine credit improvement and should inform validation policies for supplier dashboards.

Adversarial robustness requires focused experimentation in this domain. Sparse data and network structure create avenues for gaming through feature manipulation or relationship spoofing. The graph learning community has documented successful attacks via structural perturbations and node injection, which reduce classification accuracy and can transfer across models, indicating a concrete risk for supply-chain credit scoring that ingests network features (Zhang et al., 2023) [31]. Countermeasures should include certified or randomized smoothing defenses where feasible, adversarial training with realistic manipulation budgets, and monitoring that cross-checks claimed relationship changes against external registries or buyer confirmations. Explanation robustness must be monitored in parallel, since attribution and surrogate methods can degrade under imbalance or missingness; stability checks across bootstraps and data slices should therefore become part of continuous validation for regulated use. A final strand concerns evaluation science. Real-world pilots should instrument outcome tracking for approval lift, loss rates, calibration error, and explanation usefulness as rated by credit officers and suppliers. For GNN-based variants, experiments should report both discriminative performance and explanation faithfulness and complexity, using standardized benchmarks and human evaluation where appropriate.

7. Conclusion

This study demonstrates that hybrid and Bayesian approaches offer a promising pathway for supplier credit scoring in sparse-data environments in the U.S.. The hybrid model, which integrates logistic regression with domain heuristics, consistently achieved strong discriminative performance while maintaining interpretability. The Bayesian variant provided superior calibration, enabling more reliable probability estimates and making it especially valuable for risk management. Together, these approaches show that credit assessment can remain both accurate and explainable, even under conditions of significant data scarcity. Beyond predictive performance, the incorporation of explainability tools such as SHAP, counterfactual reasoning, and interpretable rules ensures that model outputs are transparent and actionable. These methods not only enhance lender trust but also provide small suppliers with clear insights into approval decisions and the steps required to improve their creditworthiness. From a regulatory standpoint, the inclusion of calibrated uncertainty estimates and interpretable

Volume 38 No. 5s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

decision logic aligns with calls for transparency and accountability in AI-driven financial systems.

The contributions of this work lie in three main areas: first, it demonstrates that combining statistical learning with domain heuristics stabilizes model performance under missingness; second, it validates Bayesian calibration as a means of generating trustworthy probability estimates; and third, it illustrates the operational value of explainability in enabling fairer and more accountable credit decisions. While these findings advance the field, limitations remain. The use of a synthetic dataset, rather than confidential real-world supplier data, constrains external validity, and explainability methods may become unstable in sparse regions of the feature space. Looking ahead, this research lays the groundwork for future deployment. Real-world validation under confidentiality agreements, the integration of richer graph-based signals, and the development of supplier-facing dashboards with interactive explanations represent promising directions. By advancing models that balance performance, interpretability, and fairness, this work contributes to building financial systems that expand access to credit while safeguarding trust among lenders, suppliers, and regulators.

References

[1] Agarwal, C., Zhao, Q., Bajaj, P., Frossard, P., & Zhang, M. (2023). Evaluating explainability for graph neural networks. *Scientific Data*, 10(1), 1-14. https://doi.org/10.1038/s41597-023-02028-y

[2] Ahmed, I., et al. (2025). Optimizing Solar Energy Production in the USA: Time-Series Analysis Using AI for Smart Energy Management. arXiv preprint arXiv:2506.23368.

[3] Ahad, M. A., et al. (2025). AI-Based Product Clustering for E-Commerce Platforms: Enhancing Navigation and User Personalization. International Journal of Environmental Sciences, 156–171.

[4] Armaki, A. H., et al. (2023). A hybrid meta-learner technique for credit scoring of banks' customers. *ResearchGate*.

[5] Blumenstock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science*, 350(6264), 1073-1076. https://doi.org/10.1126/science.aac4420

[6] Das, B. C., Sartaz, M. S., Reza, S. A., Hossain, A., Nasiruddin, M. D., Bishnu, K. K., ... & Abed, J. (2025). AI-driven cybersecurity threat detection: Building resilient defense systems using predictive analytics. *arXiv preprint arXiv:2508.01422*.

[7] Das, B., et al. (2024). Credit risk modeling with graph machine learning. *INFORMS Journal on Data Science*.

[8] Federal Reserve. (2011). Supervisory guidance on model risk management (SR 11-7). Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/srletters/sr1107.htm

[9] Fariha, N., et al. (2025). Advanced fraud detection using machine learning models: Enhancing financial transaction security. arXiv preprint arXiv:2506.10842.

[10] FinRegLab. (2022). Responsible machine learning in credit underwriting: Explainability and fairness. *FinRegLab*.

[11] GERAD. (2023). An efficient hybrid algorithm to solve credit scoring problem. *GERAD Technical Reports*. https://www.gerad.ca

[12] Islam, M. Z., Rahman, M. S., Sumsuzoha, M., Sarker, B., Islam, M. R., Alam, M., & Shil, S. K. (2025). Cryptocurrency price forecasting using machine learning: Building intelligent financial prediction models. *arXiv preprint arXiv:2508.01419*.

[13] Jagtiani, J., & Lemieux, C. (2019). The roles of alternative data and machine learning in fintech lending: Evidence from the LendingClub consumer platform. *Financial Management*, 48(4), 1009-1029. https://doi.org/10.1111/fima.12285

[14] Karimi, A. H., Barthe, G., Balle, B., & Valera, I. (2022). A survey of algorithmic recourse: Contrastive explanations and consequential recommendations. *ACM Computing Surveys*, 55(5), 1-29. https://doi.org/10.1145/3527848

[15] Khan, M. A. U. H., et al. (2025). Secure Energy Transactions Using Blockchain Leveraging AI for Fraud Detection and Energy Market Stability. arXiv preprint arXiv:2506.19870.

[16] Khan, M. N. M., et al. (2025). Assessing the Impact of ESG Factors on Financial Performance Using an AI-Enabled Predictive Model. International Journal of Environmental Sciences, 1792–1811.

[17] Kleinberg, J., Mullainathan, S., & Raghavan, M. (2017). Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807*.

[18] Kumar, I. E., Scheidegger, C., Venkatasubramanian, S., Friedler, S., & Dickerson, J. (2020). Equalizing credit opportunity in algorithms: Aligning algorithmic fairness research with U.S. fair lending regulation. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 173–183. https://doi.org/10.1145/3351095.3372826

[19] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.

[20] Mitra, S., et al. (2023). Knowledge graph driven credit risk assessment for MSMEs. *International Journal of Productivity Research*.

[21] OCC. (2011). OCC Comptroller's Handbook: Model risk management. U.S. Office of the Comptroller of the Currency. https://www.occ.gov

[22] OCC. (2021). OCC statements and resources on AI governance and model risk. U.S. Office of the Comptroller of the Currency. https://www.occ.gov

[23] Reza, S. A., Rahman, M. K., Hossain, M. S., Rabbi, M. N. S., Quddus, A., Mozumder, S. S., ... & Ahad, M. A. (2025). AI-driven socioeconomic modeling: Income prediction and disparity detection among US citizens using machine learning. *Advances in Consumer Research*, 2(4).

[24] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.

[25] Rho, S., et al. (2025). Identifying supply chain R&D partners via multilayer institutional cooperation networks. *Transportation Research Part E: Logistics and Transportation Review*, 176, 103239. https://doi.org/10.1016/j.tre.2023.103239

[26] Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology*, 31, 841.

[27] Wikipedia. "Credit analysis." (Accessed 2024).

[28] Wikipedia. "Credit scorecards." (Accessed 2024).

[29] Wikipedia. "Alternative data." (Accessed 2024).

[30] Ying, Z., Bourgeois, D., You, J., Zitnik, M., & Leskovec, J. (2019). GNNExplainer: Generating explanations for graph neural networks. *Advances in Neural Information Processing Systems*, 32.

[31] Zhang, H., Tong, H., & Lin, W. (2023). Adversarial attacks against dynamic graph neural networks via node injection. *Patterns*, 4(9), 100789. https://doi.org/10.1016/j.patter.2023.100789