

Interpretable MSME Credit Risk Prediction Through XGBoost Feature Selection and LSSVM Classification

Sharmila Jeyarani S.¹, Sanjay Raj J.², Rajamannar N.G.³, Sabari Anandh M.⁴

¹Assistant Professor, ^{2,3,4} Student, Department of Computer Science and Engineering, NPR College of Engineering and Technology, Natham, Dindigul, India.

E-mail: ¹sharmilajeyaranis@nprcolleges.org, ²sanjayraj012@gmail.com, ³rajamannar98@gmail.com,

⁴sbranandh@gmail.com

Abstract

This research work proposes an interpretable hybrid ML approach to credit risk evaluation of MSMEs in supply chain financing, combining feature selection using XGBoost with classification using RBF-LSSVM. The credit information of the multi-sourced variety is standardized into a common set of financial and operational variables followed by stratification, training normalization, and class balance through SMOTE technique to minimize the impact of minority class bias while maintaining test set independence. Using the XGBoost classifier with TreeSHAP analysis of features selects the relevant predictors that are further classified using the RBF-LSSVM classifier with calibrated probability estimation and optimized decision threshold for risk-aware classification. This proposed approach achieves 97.96% accuracy, 94.52% precision, 97.43% recall, and 95.95% F1-score on the held-out evaluation set. SHAP and counterfactual explanations offer feature level interpretations of the risk assessments.

Keywords: Supply Chain Finance, MSME Credit Risk, XGBoost Feature Selection, RBF-LSSVM, SHAP Explainability, Counterfactual Analysis, SMOTE Balancing.

1. Introduction

Supplies Chain Finance (SCF) plays a crucial role in financing the working capital of firms operating under the connected buyer-supplier chains. The Micro, Small, and Medium Enterprises

(MSMEs) make a great contribution to the Indian economy; however, they still struggle to secure necessary financing because of deficiencies in traditional evaluation techniques used to evaluate their credit risk. They might rely on static financial data that does not consider nonlinear associations between various financial variables. Moreover, financial distress experienced by one MSME might be reflected in other firms in the supply chain.

ML algorithms like Random Forest, XGBoost, and Support Vector Machines have shown promise in the modeling of complicated credit-risk patterns. Nevertheless, including diverse financial attributes without feature selection may result in redundancy and noise, while highly predictive models may not help much in understanding the rationale of individual decision-making. In addition, existing researches do not provide enough actionable information to rejected customers on how to improve their risk status.

To address these limitations, the research proposes an explainable hybrid XGBoost-LSSVM approach to perform credit-risk analysis of MSMEs in SCF applications. In the proposed approach, XGBoost is used to perform feature selection based on SHAP importance, whereas LSSVM with RBF kernel is used for performing non-linear risk classification on the selected features. Furthermore, SMOTE is applied only to the training data set due to class imbalance.

This framework is tested on 136,960 instances which have been characterized by 19 financial features and benchmarked against the baseline models. The whole process has been further extended to build a web-based decision-support system capable of making predictions in batches and real-time, as well as report generation and auditing. This combination of hybrid classification, interpretability of features, and counterfactual analysis is an effective solution to assessing credit risk for MSMEs in SCFs.

2. Related Works

2.1 Traditional Credit Risk Assessment

Traditional methods of credit risk analysis have made use of statistics and financial indicators to measure the solvency and default possibility of borrowers. While this type of method allows a fair amount of transparency in the decision process, its capacity to model the nonlinear interrelationship between the financial indicators is inadequate. As such, researches

have explored machine learning-based support vector methods to be used in supply chain financing, where the optimization of the kernel parameters would result in better classification performance than the traditional SVM design [3]. Feature selection methods have also been considered to eliminate unnecessary features and improve classification [7].

2.2 Machine Learning and Hybrid Credit-risk Models

Machine learning techniques provide greater flexibility for modelling complex interactions within structured financial data. Hybrid ensemble approaches have demonstrated improved credit-risk prediction by combining complementary learning mechanisms, particularly for SME-oriented supply chain finance applications [2]. XGBoost provides an efficient gradient-boosting framework capable of learning nonlinear relationships while incorporating regularization to control model complexity [5]. Machine learning has also been evaluated across different banking environments, demonstrating its broader applicability to automated credit-risk assessment [15]. These studies indicate that combining robust predictive learning with appropriate feature representation can improve the reliability of automated financial-risk decisions.

Machine learning has also been applied to MSME-oriented analytical problems beyond conventional credit assessment. A data-driven framework employing supervised ensemble learning, clustering, and dimensionality-reduction techniques was developed to examine MSME incubation dynamics and identify relationships among institutional affiliation, geographical location, and innovation-related characteristics [16]. This research demonstrates the applicability of machine learning for extracting structured patterns from heterogeneous MSME data and provides a broader analytical basis for data-driven enterprise-level decision systems.

2.3 Deep Learning for Credit-risk Prediction

Deep Learning (DL) models have also been considered in credit risk assessments, as well as in the ensemble of DL models that attempt to increase the model performance in cases of class imbalance. In addition to this, sampling strategies like SMOTE have been used in the deep learning models to handle the issues regarding minority classes [10]. Nevertheless, the use of advanced neural networks in structured finance data comes with many complexities. In the context of financial risks in supply chains, there have also been efforts to use CNNs for early

warning purposes [6].

2.4 Feature Selection and Model Interpretability

Feature selection remains as an integral part of credit risk modelling as unnecessary or poor predictors may make the model more complex and limit the model generalization. Cost-sensitive feature selection methods have shown that selecting small and informative sets of features can be beneficial for solving credit scoring problems [7]. In case of decision trees, SHAP offers a novel approach for estimating the importance of individual features and enhancing the interpretability of complex decisions [8]. Tree-based SHAP analysis allows to analyze global and local impact of features on model decisions [9].

2.5 Explainable and Actionable Credit Decisions

Explaining the decisions made by automated systems is especially crucial for applications in finance since such decisions need to be able to explain the result they reached. SHAP-based methods can discover how much individual features in finance influence the certain decision made, thus adding more transparency to the decisions made [8], [12]. XGBoost in combination with SHAP method has been used to build predictive models that can be explained, showing the applicability of feature attributions in tree-based learning systems [13]. In addition to feature attributions, counterfactual explanations represent another type of explanations which may influence an automated decision [11].

2.6 ESG Factors in Financial Risk

The assessment of financial risks has also started being shaped by unconventional measures apart from traditional financial metrics. The use of ESG-oriented features in the context of financial risk has also been explored in the research literature, demonstrating that sustainability performance as well as other organizational attributes may serve as another source of knowledge regarding risk evaluation [14]. Hence, the use of these indicators in credit risk modeling will allow expanding the coverage of enterprise risk, especially within the supply chain, where financial and operational conditions are connected. The growing prevalence of machine learning in various banking settings also justifies such an approach [15].

2.7 Research Gap

The existing literature shows considerable development in the fields of credit-risk prediction in the supply chain [1], hybrid machine learning [2], optimised SVM modelling [3], risk propagation in the supply chain [4], gradient boosting techniques [5], deep learning [6], feature selection [7], SHAP interpretability [8], [9], imbalanced learning [10], counterfactual explanation [11], SHAP application [12], [13], ESG-based financial risk analysis [14] and banking credit assessment using machine learning [15]. These issues are generally discussed independently from each other as separate models or interpretability techniques. The current research, on the other hand, combines all these issues into one pipeline for MSME supply chain finance, which includes feature selection using XGBoost, nonlinear classification using LSSVM, SHAP interpretability and counterfactual explanation.

3. Proposed Methodology and System Implementation

3.1 Multi-Source Dataset Integration and Feature Representation

The proposed framework utilizes several credit risk datasets, which include various financial and operational aspects that are significant for MSME credit risk assessment within supply chain finance. The Credit Risk Dataset [17], Leading Indian Bank & CIBIL Real-world Dataset [18], and an additional credit risk assessment dataset [19] complement each other to formulate the credit risk representation. Because each of these datasets varies in terms of features, dimensions, and attribute descriptions, a feature harmonization method is used to form a common representation prior to model construction.

Each business instance corresponds to a predictor vector in 19 dimensions,

$$x_i = x_{i1}, x_{i2}, \dots, x_{i19} \quad (1)$$

where x_{ij} denotes the j -th financial indicator of the i -th enterprise. The corresponding target variable denotes the risk level, where the goal is to classify instances into the categories *Low Risk* and *High-Risk* enterprises. The common representation of features is intended to capture financially meaningful attributes while being consistent with the original feature lexicon of the source datasets.

Table 1 summarizes the source datasets, their record counts, and their original feature dimensions. The combined data contain 136,960 records, which are subsequently represented using the unified 19-indicator schema.

Table 1. Dataset composition and feature summary

No.	Dataset	Records	Features
1	MSME SCF Credit Data [19]	1,707	89
2	Credit Risk Dataset [17]	32,581	12
3	Internal Bank Dataset [18]	51,336	26
4	External CIBIL Dataset [18]	51,336	62
Total	Combined Dataset	136,960	19 unified indicators

3.2 Data Preprocessing and Training-set Balancing

The integrated dataset goes through a series of preprocessing steps to create a leakage-free, uniform training and evaluation environment. After feature harmonization, the dataset becomes ready to undergo supervised learning through label processing, stratified splitting, feature scaling, and class balancing. The purpose of the preprocessing is to ensure that the natural distribution of the independent test dataset is preserved while the training process handles the class imbalance.

Initially, the datasets with integrated records are combined into one dataset. In the basic algorithm, 0.2% label jittering procedure is used where some percentage of class labels is changed to add slight noise to annotations. Further, this dataset is split into train and test sets with a stratified ratio of 80:20. In the stratified split, the proportion of both Low Risk and High-Risk classes is maintained in both datasets.

In partitioning, Z-score normalization is calculated only using training data and is then applied to training and testing features respectively. The standardized value is represented by

$$z_{ij} = \frac{x_{ij} - \mu_j^{\text{train}}}{\sigma_j^{\text{train}}} \quad (2)$$

where x_{ij} represents the original value of feature j for sample i , while μ_j^{train} and σ_j^{train} denote the training-set mean and standard deviation, respectively. The use of training-based statistics ensures that the test data is not used to preprocess the data.

Due to the prevalence of imbalanced data sets for credit risk classification problems, SMOTE technique is later employed only on the training data set. The minority class is over-sampled until balance between both classes is achieved on the training set at 50:50 ratio, while on the testing set there still exists an original imbalance of 75:25 between low risk and high-risk classes respectively.

Figure 1 shows the entire preprocessing pipeline, including integration and mapping of multi-source data, label processing, stratified splitting, training-specific normalization, and balancing using SMOTE technique. Hence, Figure 1 gives a pictorial depiction of the order of steps followed prior to the feature selection process that is XGBoost-based.

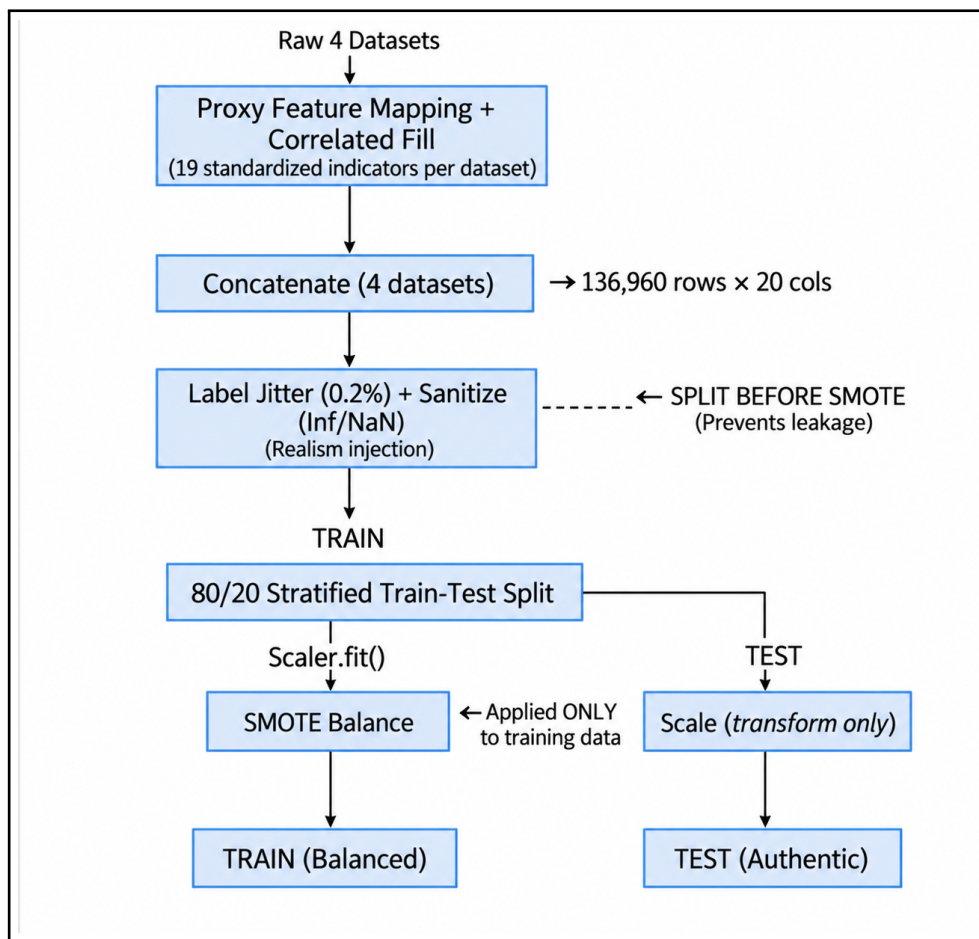


Figure 1. Data preprocessing and class-balancing workflow

3.3 SHAP-guided XGBoost Feature Selection

After performing the preprocessing and balancing of classes, XGBoost is used as the initial step of the proposed hybrid algorithm to select informative features from the integrated

financial feature space. XGBoost is chosen since the gradient-boosting mechanism of the decision trees can discover nonlinear interactions between the financial indicators in an efficient way of feature importance analysis. XGBoost is configured to have 300 boosting iterations with the maximum tree depth equal to 6 and learning rate of 0.08, according to the base configuration.

XGBoost prediction for the training sample is represented by an additive combination of the regression trees:

$$\hat{y}_i = \sum_{k=1}^K f_k x_i, \quad f_k \in \mathcal{F} \quad (3)$$

where K denotes the number of boosting rounds, x_i is the input feature vector, and f_k represents the k -th decision tree.

To enhance the interpretability of feature selection, the TreeSHAP method is used for the trained XGBoost model. The contribution of each variable is quantified using SHAP as the attribution score. The global importance of feature j is determined from the mean absolute SHAP contribution:

$$I_j = \frac{1}{N} \sum_{i=1}^N |\phi_{ij}| \quad (4)$$

where N is the number of evaluated samples and ϕ_{ij} is the SHAP value of feature j for sample i . Features with larger I_j values have greater overall influence on the model's predictions. This provides more informative insights than tree-based importance alone, since SHAP also reveals the direction and magnitude of the influence.

The final set of features will be selected for integrating the feature importance parameters obtained from the model along with indicators that have financial relevance. Out of the 15 indicators, 13 are predefined and include, for example, debt ratio, debt service coverage ratio, current ratio, operating margin, and ESG rating. The rest of the two spots will be filled by the two most important features based on global SHAP importance.

This produces a 15-dimensional feature representation, which is subsequently fed to the RBF-LSSVM classifier, which serves as the second phase of the proposed hybrid learning algorithm. This sequence ensures that feature extraction is independent of non-linear

classification, while reducing the dimensionality of input to the final classifier.

3.4 RBF-LSSVM Credit-risk Classification and Decision Thresholding

The 15 selected financial factors are given to a Least Squares Support Vector Machine (LSSVM) algorithm for credit risk classification. In contrast to traditional SVM, the LSSVM uses an equality constraint to formulate the learning task and hence leads to a system of linear equations which might train efficiently but still have the ability to model nonlinearity. This is because an RBF kernel function is used to model nonlinearity between selected financial factors and credit risk classes. The regularization parameter C is set to 10.0 in the base implementation.

The RBF kernel employed by the classifier is expressed as

$$K_{x_i, x_j} = \exp \left(-\gamma \|x_i - x_j\|^2 \right) \quad (5)$$

where x_i and x_j denote two selected-feature vectors, and γ controls the width of the radial basis function. The kernel allows the classifier to form a nonlinear boundary in the higher-dimensional feature space.

The LSSVM results are then converted to a calibrated risk probability through Platt scaling. Rather than use the standard probability of 0.5 as the threshold value, the program calculates thresholds for decisions since an incorrect approval of a high-risk venture has higher financial implications than the denial of a low-risk venture. The calculated optimum threshold is 0.364 out of 551 threshold candidates.

Accordingly, the risk decision is represented by

$$\hat{y} = \begin{cases} \text{High Risk,} & P_{\text{High Risk}} | x \geq 0.364, \\ \text{Low Risk,} & P_{\text{High Risk}} | x < 0.364. \end{cases} \quad (6)$$

This thresholding process forms the final classification decision which will be used in the next explainability layer.

3.5 Explainability and Counterfactual Risk Analysis

SHAP local explanations have been used within the framework in order to find the financial characteristics that play a role in the prediction of credit risk. It is possible to analyze the direction and importance of the factors that have an impact on the decisions made.

In high-risk cases, counterfactual analysis helps identify feasible adjustments in certain financial variables to minimize the expected risk. Both the SHAP and counterfactual values are subsequently used by Gemma-2-2B-IT through NVIDIA NIM to produce a succinct explanation for the decision.

3.6 Web-based System Architecture and Implementation

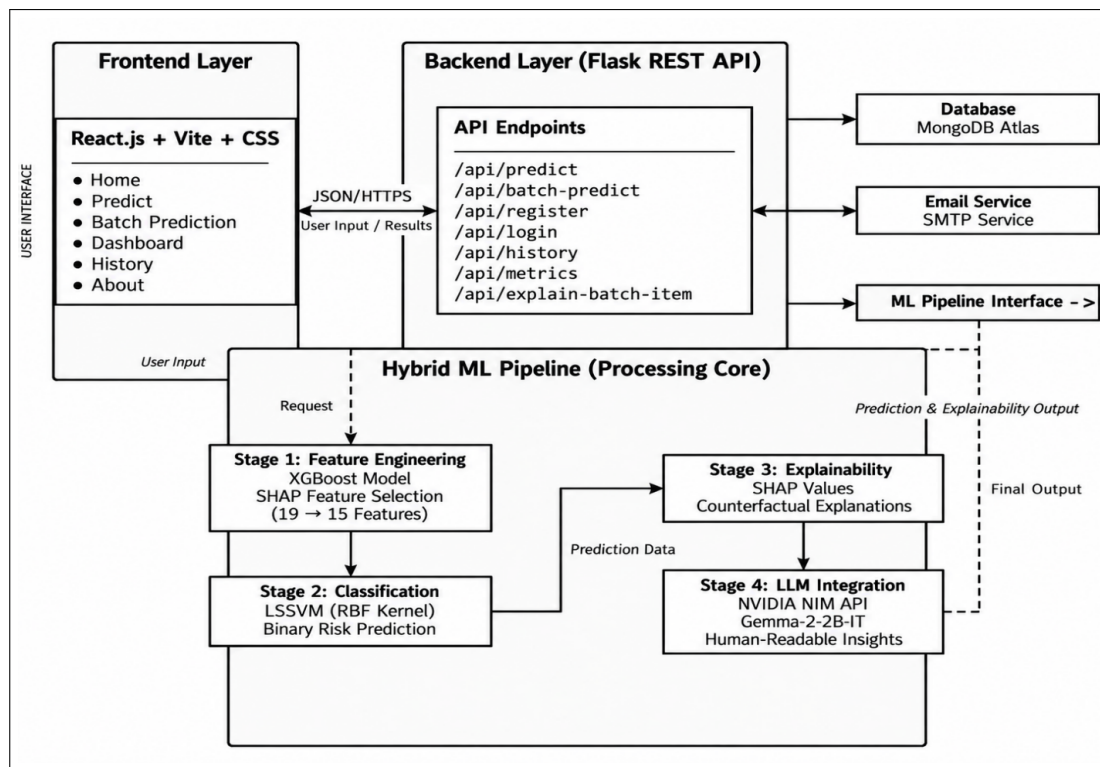


Figure 2. Web-based architecture of the proposed credit-risk system.

The proposed system is realized as an online credit-risk decision aid. The front end is implemented using React.js 18 with Vite, while Flask 2.0 serves as the back-end API to link the user interface to the XGBoost–LSSVM model prediction pipeline. The system is deployed with the trained XGBoost selector, LSSVM classifier, scaler, SHAP explainer, and decision threshold. SHAP and counterfactual explanations are provided to allow interpretability of credit risk decisions, and NVIDIA NIM (Gemma-2-2B-IT) is used to produce textual explanations.

MongoDB Atlas is used to store prediction results and logs, while jsPDF and Gmail SMTP with TLS are used for reporting and notification via e-mail. The architecture of the system is given in Figure 2. The technologies used to implement the system are summarized in Table 2.

Table 2. System components and technologies

Component	Technology
Frontend	React.js 18 + Vite
Backend API	Flask 2.0
ML Pipeline	XGBoost + LSSVM
Explainability	SHAP 0.42+
LLM Engine	NVIDIA NIM (Gemma-2-2B-IT)
Database	MongoDB Atlas
Report Generation	jsPDF
Email Alerts	Gmail SMTP + TLS
Data Balancing	SMOTE

4. Experimental Results and Discussion

4.1 Performance of the Proposed Hybrid Model

The proposed XGBoost-LSSVM system was validated using 27,392 test records held out entirely, retaining the actual 75:25 Low Risk to High Risk ratio. The performance of the proposed method is presented in Table 3 in terms of accuracy, precision, recall, and F1-score. The proposed model achieves 97.96% accuracy, 94.52% precision, 97.43% recall, and 95.95% F1-score, demonstrating strong discrimination between the two risk classes.

The results demonstrate the advantage of using the sequential hybrid approach. XGBoost alone has an accuracy of 95.75% and F1 score of 95.29%, while LSSVM gives an accuracy of 90.52% and F1 score of 87.66%. Combining XGBoost feature selection with LSSVM boosts the accuracy by 2.21% compared to XGBoost and 7.44% compared to LSSVM. The proposed approach also gives the highest recall of 97.43%, which shows that the high-risk patients can be

identified effectively.

Table 3. Performance comparison of different models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN [1]	82.60	85.30	87.70	86.50
LSSVM [6]	90.52	84.78	90.90	87.66
Random Forest [9]	95.36	95.05	94.66	94.91
XGBoost [5]	95.75	95.20	95.47	95.29
Proposed XGBoost–LSSVM	97.96	94.52	97.43	95.95

The confusion matrix in Figure 3 serves as an alternative class-based evaluation by displaying the number of true and false classifications made to both Low-Risk and High-Risk classes. Combined with the summary metrics in Table 3, this evaluation proves the ability of the proposed hybrid framework to predict new, unseen records.

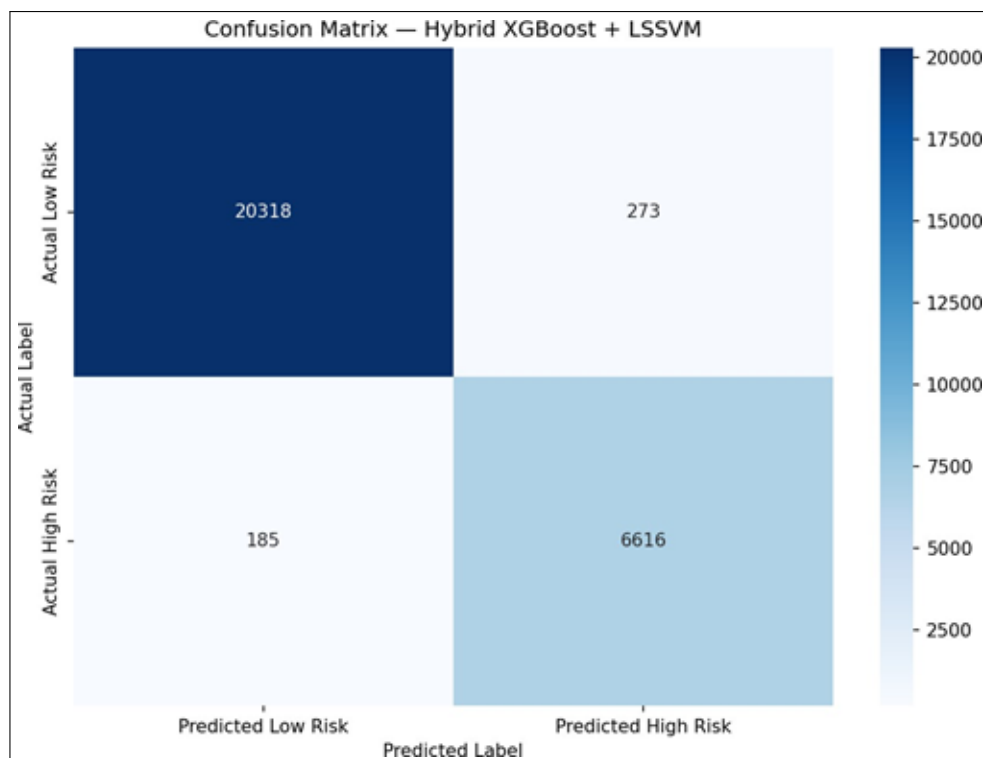


Figure 3. Confusion matrix of the proposed hybrid XGBoost–LSSVM model on the held-out test set.

4.2 Cross-validation and Prediction Stability

Stratified fivefold cross-validation technique was employed for the evaluation of the consistency of the proposed hybrid model on varying sets of data. The model showed consistent accuracy in the 97-98% interval, with standard deviation of less than 1%. It means that the model results are relatively independent from the chosen training/validation split.

This consistent cross-validation result also supports the efficiency of the preprocessing and feature selection technique used. Specifically, the use of SMOTE on the training set but keeping the true proportion in the test set is a more controlled experiment for measuring generalization ability. Hence, the use of stratification, balance in the training set, and dimensionality reduction makes the model behave consistently with varying data splits.

4.3 SHAP-based Feature Importance and Risk Interpretation

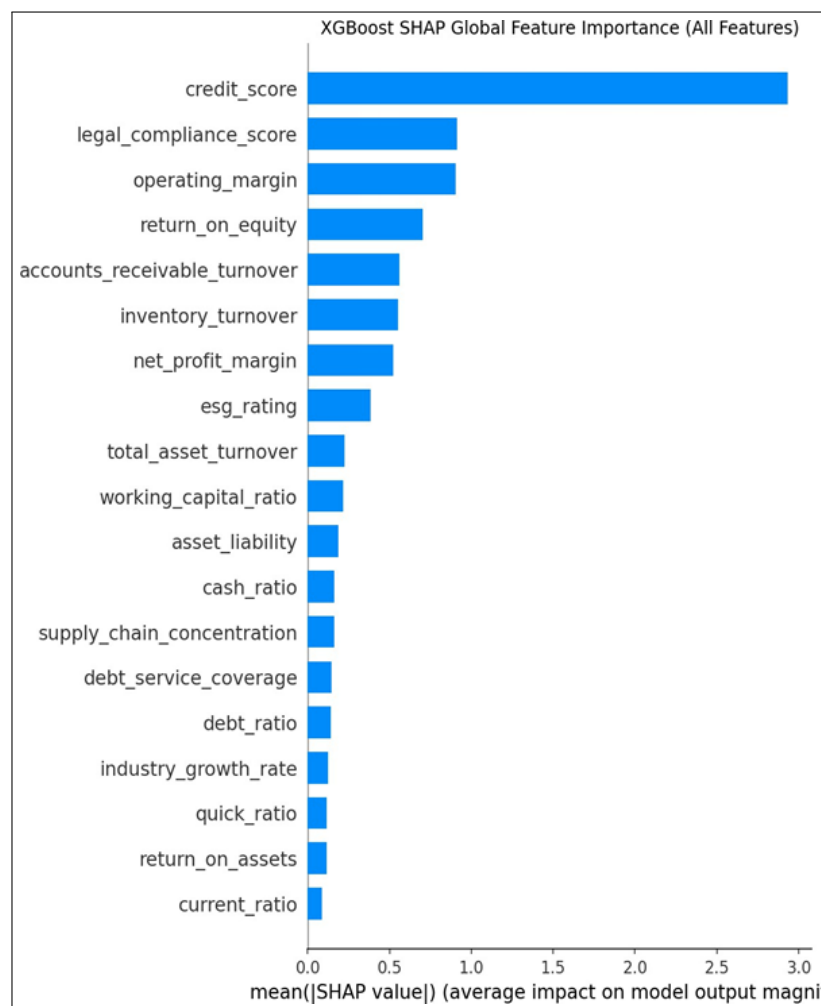


Figure 4. SHAP global feature-importance ranking for the XGBoost stage.

Figure 4 shows the SHAP-based ranking for the importance of features from the XGBoost phase globally. Credit score is the most important feature, followed by legal compliance score, operating margin, return on equity, and accounts receivable turnover. Several other financial, operational, ESG, and supply chain measures also play a role in constructing the risk profile.

The SHAP analysis further indicates that an increase in debt ratio beyond approximately 0.40 is associated with increased predicted risk, while favourable ESG-related characteristics contribute toward lower risk. Thus, Figure 4 demonstrates how SHAP converts the feature-selection output into an interpretable representation of the factors influencing credit-risk predictions.

4.4 Counterfactual Risk Analysis

The counterfactual approach adds to SHAP in terms of determining how changing certain financial variables could affect the predicted outcome. In one of the example cases where there is a high probability of defaulting, the reduction of the debt ratio from 0.72 to 0.40 is found to lower the predicted risk of defaulting by about 23% points. Further improvement in operating margin could also be helpful in reducing risks. This analysis makes the above model action-oriented by correlating the output prediction to possible changes in the financial indicators under control.

The combination of feature attribution and counterfactual reasoning, the approach moves beyond just labeling an MSME as either a Low Risk or High-Risk business. The approach not only explains the reasons for a certain risk classification but also demonstrates how certain financial situations might influence that risk classification.

4.5 Discussion of the Hybrid Pipeline and System Implementation

The proposed architecture, result is using for the merits of XGBoost and LSSVM in a sequential way. Initially, the XGBoost model is used to analyze and select the most important features among the input variables, decreasing the number of features from 19 to 15 variables. Then, the selected features will be fed to the LSSVM model for doing the classification of risk.

The entire process is integrated within a decision support system based on the web. The user will feed the credit data via the web, and the backend will take the record into account and generate the risk assessment. SHAP will provide the explanation at the level of features, and counterfactual will investigate possible alterations of financial metrics.

5. Conclusion and Future Work

The proposed framework integrated feature selection through XGBoost, RBF-LSSVM classification, SHAP, and counterfactual methods in an interpretable credit risk decision-making process for MSMEs in supply chain finance. Through the use of the sequential learning approach, this framework eliminated irrelevant feature information while still incorporating important risk signals with financial significance, and the explainability component makes connections between the predictions of the model and its determinants along with possible risk mitigation actions. Integration with a Flask-based web application has made the whole process of credit evaluation, credit reporting, and decision making more efficient for auditing purposes. Future research may involve federated learning to create models collaboratively among financial institutions without the need to reveal confidential data about borrowers. The risk propagation through graphs may include interrelationships among all the supply chain members. The framework may be extended from the binary classification problem to the multi-class strategies.

References

- [1] Zhou, Guanglan, and Shiru Wang. "Enhancing Credit Risk Decision-Making in Supply Chain Finance with Interpretable Machine Learning Model." *IEEE Access*, 2025, vol. 13: 14239-14251.
- [2] Zhu, You, Li Zhou, Chi Xie, Gang-Jin Wang, and Truong V. Nguyen. "Forecasting Smes' Credit Risk in Supply Chain Finance with an Enhanced Hybrid Ensemble Machine Learning Approach." *International Journal of Production Economics*, 2019, vol. 211: 22-33.
- [3] Zhang, Hao, Yuxin Shi, Xueran Yang, and Ruiling Zhou. "A Firefly Algorithm Modified Support Vector Machine for the Credit Risk Assessment of Supply Chain Finance." *Research in International Business and Finance*, 2021, vol. 58: 101482.
- [4] Ojha, Ritesh, Abhijeet Ghadge, Manoj Kumar Tiwari, and Umit S. Bititci. "Bayesian Network Modelling for Supply Chain Risk Propagation." *International Journal of Production Research*, 2018, vol. 56, no. 17: 5795-5819.
- [5] Chen, Tianqi, and Carlos Guestrin. "XGBoost: A Scalable Tree Boosting System." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, 785-794.
- [6] Tang, Qian, Yan Lu, Bin Wang, and Zhen Li. "A Deep Convolutional Neural Network Based Risk Identification Method for E-Commerce Supply Chain Finance." *Scientific Programming*,

- 2022, vol. 2022, no. 1: 6298248.
- [7] Maldonado, Sebastián, Juan Pérez, and Cristián Bravo. "Cost-Based Feature Selection for Support Vector Machines: An Application in Credit Scoring." *European Journal of Operational Research*, 2017, vol. 261, no. 2: 656-665.
 - [8] Lundberg, Scott M., and Su-In Lee. "A Unified Approach to Interpreting Model Predictions." *Advances in Neural Information Processing Systems*, 2017, vol. 30: 1-10.
 - [9] Lundberg, Scott M., Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M. Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. "From Local Explanations to Global Understanding with Explainable AI For Trees." *Nature Machine Intelligence*, 2020, vol. 2, no. 1: 56-67.
 - [10] Shen, Feng, Xingchao Zhao, Gang Kou, and Fawaz E. Alsaadi. "A New Deep Learning Ensemble Credit Risk Evaluation Model with an Improved Synthetic Minority Oversampling Technique." *Applied Soft Computing*, 2021, vol. 98: 106852.
 - [11] Wachter, Sandra, Brent Mittelstadt, and Chris Russell. "Counterfactual Explanations Without Opening the Black Box: Automated Decisions and The GDPR." *Harvard Journal and Technology*, 2017, vol. 31: 841.
 - [12] Ekanayake, I. U., D. P. P. Meddage, and Upaka Rathnayake. "A Novel Approach to Explain the Black-Box Nature of Machine Learning in Compressive Strength Predictions of Concrete using Shapley Additive Explanations (SHAP)." *Case Studies in Construction Materials*, 2022, vol. 16: e01059.
 - [13] Mo, Tiexiang, Shanshan Li, and Guodong Li. "An Interpretable Machine Learning Model For Predicting Cavity Water Depth and Cavity Length Based on Xgboost–SHAP." *Journal of Hydroinformatics*, 2023, vol. 25, no. 4: 1488-1500.
 - [14] Shakil, Mohammad Hassan. "Environmental, Social and Governance Performance and Financial Risk: Moderating Role of ESG Controversies and Board Gender Diversity." *Resources Policy*, 2021, vol. 72: 102144.
 - [15] Gafsi, Nesrine. "Machine Learning Approaches to Credit Risk: Comparative Evidence from Participation and Conventional Banks in The UK." *Journal of Risk and Financial Management*, 2025, vol. 18, no. 7: 345.
 - [16] Raj, Taflin S, and Radhika R. 2026. "A Machine Learning Framework for Evaluating MSME Incubation Dynamics". *Journal of Trends in Computer Science and Smart Technology* 8 (1): 114-29. <https://doi.org/10.36548/jtcsst.2026.1.006>.
 - [17] Credit Risk Dataset - <https://www.kaggle.com/datasets/laotse/credit-risk-dataset>

- [18] Leading Indian Bank & CIBIL Real-World Dataset – <https://www.kaggle.com/datasets/saurabhbhadole/leading-indian-bank-and-cibil-real-world-dataset>
- [19] Ensemble Learning with Feature Optimization for Credit Risk Assessment –https://www.researchgate.net/publication/382688309_Ensemble_Learning_with_Feature_Optimization_for_Credit_Risk_Assessment