Expert Systems With Applications 208 (2022) 118026

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Financial supply chain analysis with borrower identification in smart lending platform

Rony Mitra [a], Adrijit Goswami [a], Manoj Kumar Tiwari [b,c,*]

[a] Department of Mathematics, Indian Institute of Technology Kharagpur, Kharagpur 721302, West Bengal, India
[b] National Institute of Industrial Engineering, Mumbai, India
[c] Department of Industrial and Systems Engineering, Indian Institute of Technology Kharagpur, Kharagpur 721302, West Bengal, India

ARTICLE INFO

Keywords:
Supply chain finance
Peer-to-peer lending platform
Machine learning
Random forest
Collaborative filtering
Similarity estimation

ABSTRACT

The popularity of the online peer-to-peer (P2P) lending platform in the financial supply chain (FSC) has grown tremendously in the past few years. However, it is pretty challenging to develop an efficient financial supply chain for Micro, Small, and Medium-Sized Enterprises (MSMEs) by providing credits and services. In this research, an FSC model analyzes by emphasizing the retailer and supplier relationship. The retailer predicts the market demand to maximize profit, and accordingly, orders to the suppliers and suppliers utilize the production capabilities in limited monetary value. In such a case, the supplier also needs to apply for loans to meet the production commitment. With an intelligent lending platform, applicants can get loans quickly and more conveniently from the lenders. Still, identifying defaulter borrowers is a difficult task on the peer-to-peer lending platform. The main focus of the lenders is to maximize profit and minimize risk by giving the loan to non-defaulter suppliers. The present study describes the importance of the financial parameters (λ) of the retailers and suppliers (MSMEs) and their impact on the related supply chain. We provide the proof of dependency of the financial parameter on the cost of debt and the ratio of the debt-to-equity by developing two propositions. We propose an innovative k-Random Boosting Classifiers (k-RBC) algorithm for identifying potential good and bad borrowers to capture this case. The time complexity of the k-RBC algorithm is $\mathcal{O}(n^2 log(n))$. The results obtained from the study show a significant improvement in comparison to the outputs from existing approaches on the same datasets. Our algorithm gives 90% accuracy to identify potential good borrowers, whereas existing algorithms achieve up to 87% accuracy. Furthermore, the importance of the borrowers' features and its impact are analyzed of the lending platform in FSC.

1. Introduction

Recently, an increasing interest in the financial flow has been noticed in both industry practitioners and academics. Supply chain (SC) obligations with consignments of purchasing have fastened a place in the research of operations management. Bassok and Anupindi (1997) were amongst the earlier to interpret the fact that a retailer promises to buy the smallest amount defined from a supplier over a given outlining range can reduce the vulnerability of SC. Researchers (Durango-Cohen & Yano, 2006; Li & Kouvelis, 1999) have consequently dispensed that supply agreements can help the organization to generate value by facilitating better planning of operations and reducing uncertainties of surplus or shortage. Supply chain finance (SCF) or Financial supply chain has become one of the trending topics among SC managers and researchers in countries that are facing dramatic changes in their economic systems. SCF aims at collaborative and innovative business models for providing services and credit to Micro, Small & Medium Enterprises (MSMEs) by changing convertible assets (i.e., inventory, raw materials, account receivable, etc.) into cash. There are plentiful possibilities for improving profits through the enhancement of economic flows. As companies or organizations have made significant advancements in the integration and coordination of information flows and material, they can identify various opportunities for delivering better benefits through the coordination of financial flows as well as material and information transparency. According to Financing MSMEs and Entrepreneurs 2019: the OECD Scoreboard reported that more than 98% of all companies in China are MSMEs account. Moreover, in 2018 they shared 50% of tax revenue and 58% of GDP. Moreover, they shared 50% of tax revenue and 58% of GDP in 2018. In many supply chain, MSMEs are compressed by their dominant trading companions. They encounter complexities in maintaining SC activities due to the lack of funds. Furthermore, traditional banks are not willing to provide

* Correspondence to: National Institute of Industrial Engineering (NITIE), Vihar Lake Road, Mumbai 400087, Maharashtra, India.
E-mail addresses: mitrarony92@gmail.com (R. Mitra), goswami@maths.iitkgp.ac.in (A. Goswami), mkt09@hotmail.com (M.K. Tiwari).

https://doi.org/10.1016/j.eswa.2022.118026
Received 24 February 2022; Received in revised form 26 June 2022; Accepted 29 June 2022
Available online 6 July 2022
0957-4174/

the credit needed by these MSMEs due to a lack of collateral and information asymmetry.

Due to the asymmetry of data and shortage of collateral, small companies are faced with troubles in funding from lenders or bankers. In conclusion, these tiny suppliers in manufacturing are budget-constrained. Ma et al. (2020) have investigated the significance of the supply chain collaborative factors in SCF. They specified eight collaborative factors, such as decision synchronization, IT infrastructure, integrated supply chain process, information sharing, the Collaborative performance system, top management support, trust, and incentive alignment. Supplier need some money in advance to meet the optimal production decision. But for lenders, it is difficult to identify good borrowers and defaulters. So the vulnerability of a financial system had become a big difficulty in financial engineering because of the financial disaster in 2007–2008 when the large market collision led to the bankruptcy of FI, such as Lehman Brothers. In today's world, the credit score is one of the key features to get a loan in an online P2P lending platform, and assessing financial credit risk is of immense importance in the accounting and finance areas. FI needs to keep the credit default risk to an allowable level so that higher profitability can be secure. As the number of applicants increases significantly, the value of selecting a suitable model for predicting credit risk is crucial. For that, an increasing interest in the financial flow has been noticed in both industry practitioners and academics.

We study an SC where a supplier gets orders from a retailer, meets economic limitations. Informational uncertainties about demand and production capacities restrict the supplier to access resources. By perpetrating at least the purchase amount, the retailer can decrease the informational uncertainties and increase the supplier's achievable production quantity. To meet the optimum production decision, suppliers need some money in advance. For that, they apply for the loan in the peer-to-peer (P2P) lending platform. So, we proposed a method for the borrower identification on the P2P lending platform, which combines collaborative filtering technology and machine learning (ML) algorithms to distinguish unidentified borrowers to reduce risk.

The remainder of the paper is organized as follows. After a relevant literature review in the next section, we described the problems in purchase order financing in Section 3 and analyzed the impact of the financial parameter in SCF. In Sections 4 and 5 we developed an algorithm to identify borrower's characteristics and analyzed its accuracy with existing algorithms. Finally, Section 6 concludes this paper and offers some future research directions.

2. Literature survey

An intercommunication between operations and finance can only occur when financial markets are flawed and therefore fail to complete the necessities of the Modigliani-Miller theorem (MMT). The MMT states that the market value of an enterprise is estimated using its earning potential and the uncertainty of its underlying assets and is autonomous of the way it distributes dividends or finances investments. Financing is one of the key factors in SCM. There are three approaches an enterprise can decide to finance: borrowing, straight issuance of shares, and spending profits (versus returning them out to stockholders in the form of shares).

We explore the general framework of the SCF for the MSMEs including both public and private financial support. The supplier can apply for loan direct from banks independently (Fig. 1), where he has to provide credit guarantee. In this case supplier has to repay the loan and compensate the bank for the loss. A distinction to the orientation of recent studies is presented by Reindorp et al. (2018), where the financial parameters either affect the retailer (Caldentey & Chen, 2009) or both suppliers and retailer (Kouvelis & Zhao, 2012; Lai et al., 2009).

Reindorp et al. (2018) show that purchase assurances suggest a comparable risk-sharing effect, but also differentiate this from a contemporary financing influence. While a prosperous retailer's total gain from a purchase assurance is definite, its benefit from the risk-shifting effect individually is negative. Abbasi et al. (2018) has examined the critical financing challenges in the supply chain faced by small and medium-sized enterprises, such as Lack of Coordination & Cooperation Between Parties, Unrealistic Approaches, Limited Accessible Technology, Potential Risk Negligence, and Locked Working Capital. They have discussed the benefits (quantitative and qualitative) of SCF to overcome those challenges. Also, cash flow management influences the financial performance of MSMEs and is also essential for their growth and survival. Chen et al. (2019) introduced a supply chain financing model to capture the cash flow dynamics for all the parties in the supply chain. The role of intermediary parties is crucial for the cash flow in the supply chain. They also identified the effective grace period for the payment delay and transportation time using the Nash bargaining solution to occur more buyers. Wu (2022) examined SMEs' optimal purchase order financing concerning uncertainty and operational flexibility. They have proposed the Stackelberg model for two suppliers and one buyer to analyze the strategies for a buyer. They have not considered the financial factor of the buyer and suppliers. Information asymmetry of MSMEs consistently directs difficulty to determine their credit risk for lenders or banks.

If supplier is a MSMEs and may not get financing from banks independently due to asymmetric information, little collateral, and high default risk. The rate of refusal of bank loan applications for SMEs' is almost thrice that of large companies. Even though on average only 53.1% amounts of funding requests of the bank loan amount are ultimately granted for the accepted loan of MSMEs. Somjai et al. (2019) analyzed negotiation plays an essential role among the MSMEs and financial service providers to improve the entire supply chain. Also, it can help to improve the performance of MSMEs. There are many available financial schemes for the MSMEs offered by the public or private sectors, such as Pradhan Mantri Mudra Yojana, Credit Guarantee Trust Fund for Micro & Small Enterprises, Credit Linked Capital Subsidy Scheme, Equity Infusion for MSMEs through Fund of Funds, Credit Guarantee Scheme for Subordinate Debt, and SIDBI Make in India Loan for Enterprises. Park et al. (2020) studied the role of government support services and loan financing in the survival and growth of SMEs. They explored that governmental economic support can only help SMEs to survive, but do not ensure sales growth and achieve higher annual assets. If governmental economic services are integrated with financial institutions, that is more effective in improving the SMEs' sales growth and annual assets. Though various schemes are offered by governments for MSMEs survival including financial assistance & incentives, tax allowances, reducing administration costs, technical & management skills, technical support, and assembling connections with external business, in arising economics majority of MSMEs are unable to receive government support for their survival and growth. But, the survival of MSME is dependent on their performance which can be ensured through good supply chain practices.

Our research benefits all the parties, such as MSMEs, banks, e-commerce platforms in the supply chain finance so that the SCF model is sustainable. To improve the financing and funding of MSMEs', numerous innovations have developed (Asai, 2019). Yoshino and Taghizadeh-Hesary (2019) studied how to implement financing for SMEs intelligently. Therefore the popularity of individual micro-loan online service increases. So, more and more P2P lending platforms (Upstart, Funding Circle, Prosper, Peerform etc.) established in last ten years. Peer group lending has been emerging in local communities, and then Conlin (1999) proposed a model to explain the micro-lending businesses in America, which shows the peer groups' marketing mechanism. Berger and Gleisner (2009) examines the intermediary role by analyzing 14 000 businesses of the loan in an electronic market. However, the big difficulty in individual lending markets is the information asymmetry (Emekter et al., 2015). The Dominance-based Rough Set Balanced Rule Ensemble is proposed by Błaszczyński et al. (2021) for predicting financial fraud.

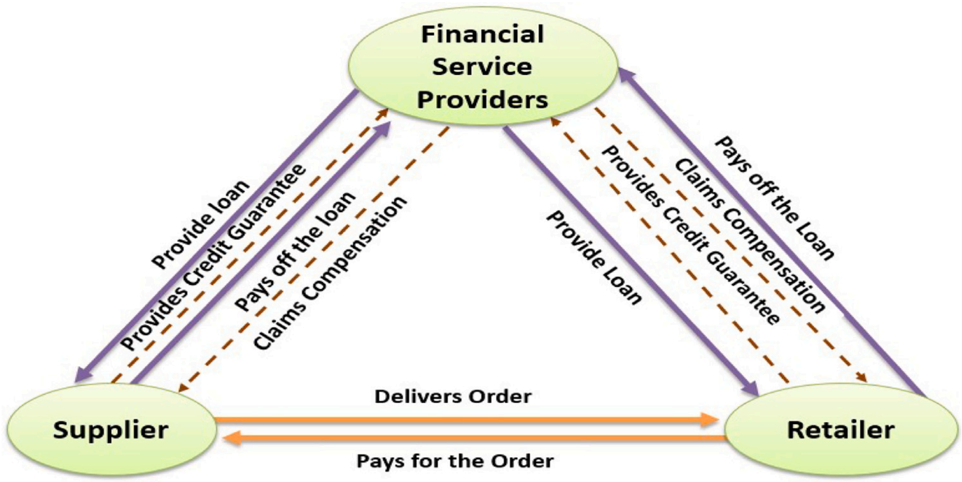


Fig. 1. Supply Chain Finance (SCF) framework.

Borrower identification belongs to credit risk evaluation, which is useful to control the probability of default. Giudici and Bilotta (2004) analyzed that Bayesian networks can constitute a valid model to measure the operational risks of banks. They have calculated the operational risk VaR for the banks. Numbers of approaches have been proposed, including statistical methods (e.g., decision trees (Fang et al., 2004), naive Bayes (Gao et al., 2017), logistic regression (Skordilis & Moghaddass, 2017)), and data mining methods (e.g., support vector machine (Huang et al., 2004) and neural network (Baesens et al., 2003)). Hand and Henley (1997) have classified loan applicants based on their credit scores by using Statistical Classification Methods. They also analyzed k-nearest neighbor, Logistic Regression, Linear Regression, and Decision Tree considering bad risk rate. Wei et al. (2016) analysed the impact of the social network on credit risk score accuracy and develop a series of models for comparing without the social network data. Chen et al. (2016) developed an equilibrium constrained optimization approach to model the systemic risk in a banking system. Agarwal et al. (2016) suggest by the text mining result that the credit rating performance records include new default-related data. Le et al. (2021) applied sequence-based clustering for default behavior prediction.

Credit access decreases MSMEs' sales volatility. In particular, the effect is stronger for MSMEs with more irregular sources of financing. Roa et al. (2021) proposed an application (super-app) based on behavioral patterns to develop a credit risk model. Chen et al. (2022) analyze that MSMEs with access to FinTech credit are less probable to go bankrupt or leave the industry in near future. By using Additional channel tests, they demonstrate that, specifically during business downturns, the MSMEs with strong credit invest more in the sector or product diversification and promotion, which operates as a valuable practice to decrease MSMEs volatility by credit access. To analyze the credit risk, Deo and Juneja (2021) proposed closed-form approximations to the maximum likelihood estimator based on discrete default intensity and logit-type reduced-form for conditional default probabilities. They considered that the conditional default probabilities follow a stationary Gaussian process for corporate loans. Also, to determine the performance of the estimator, they have introduced a rarity parameter based on the sequence of specified statistical models for the underlying defaults. There are three types of risk associated with financial institutions like operational risk, market risk, and credit risk. For the high dimensional complex dataset, Giudici (2001) proposed a credit scoring model by using Bayesian methods and Markov chain Monte Carlo computational techniques. They employed graphical models and capital asset pricing model along with arbitrage pricing theory to localize model specification & inferences and determine the extra return respectively. Xu et al. (2020) developed a framework for operational risk management of financial firms. They have represented optimal control policies by incorporating preventive control, corrective control, and joint control. Moscatelli et al. (2020) compared statistical models with ML models in credit risk modeling. Yıldırım et al. (2021) introduced default prediction models using a graph-theoretical approach based on ML algorithms (such as logistic regression, decision tree, random forest, gradient boosting). Liu and Wu (2021) studied the P2P platform and applied RF to predict the default status of the P2P customers. Not only P2P, but we have also analyzed the impact of financial parameters in supply chain financing. Andreeva et al. (2016) have compared the predictors of SMEs failure between United Kingdom and Italy based on Generalized Extreme Value regression. Also, they have used non-parametric additive model to relax the assumption of non-linearity. For a better understanding of the components of credit risk or financial risk in P2P lending platforms, Bussmann et al. (2021) have introduced an Artificial Intelligence model based on XGBoost and TreeSHAP. They analyzed their model by using 15,045 SMEs' credit risks. They compared XGBoost & LR algorithms based on the AUROC curve and employed the Minimum Spanning Tree algorithm for better representation in the form of a cluster dendrogram. Vanara et al. (2021) developed a credit scoring model to find valid customers by using RF with 11 features. We have developed a classification model by using novel k-Random Boosting Classifiers with 23 features.

As there is no study that combines borrowers' similarities with ML algorithms for borrower identification in P2P lending platform, we fill this void in some extent Taiwan dataset. We analyzed the impact of the financial parameter in purchase order financing and developed an algorithm to predict bad borrower to reduce the risk and prevent bankruptcy of the financial institutions.

3. Purchase order financing in supply chain consequences

Multiple expenses arise when retailers order goods to a supplier. Buying materials, as well as manufacturing, packaging and shipping the goods can put a huge strain on supplier's cash flow. Without adequate cash flow on hand, it can be impossible to fill all order requests. Purchase order financing provides a cash in advance for the goods that retailers have ordered. This way, supplier can meet production and delivery expenses while waiting for payment.

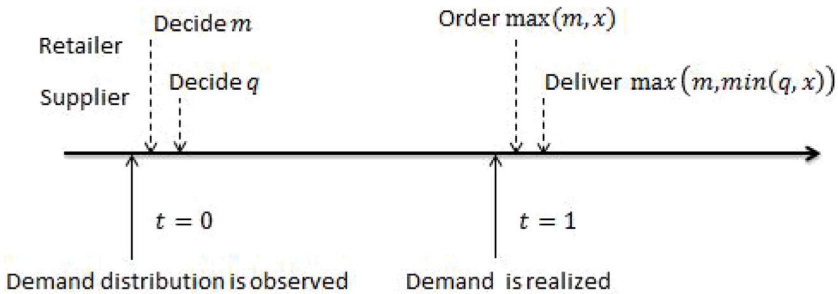


Fig. 2. Timing of events.

3.1. Primary notations

Symbol	Definition
X	demand (random variable)
$f(x)$	probability density function (pdf) of X
$F(x)$	cumulative distribution function (cdf) of X
t	time index
p	supplier's unit cost of production
r_p	retail price
g_r, g_s	retailer (r) and supplier (s) gross margin
q_j^n	retailer (r) and supplier (s) newsvendor quantity
κ_s	supplier's independent debt capacity
κ_r	retailer's independent debt capacity
λ_s	supplier's informational transparency
λ_r	retailer's informational transparency
m	retailer's purchase order commitment
m^*	retailer's optimal purchase order commitment
q	supplier's production decision
q^*	supplier's optimal production decision
Z_r, Z_s	retailer (r) and supplier (s) profit

3.2. Problem description

In a double-stage SC, a retailer sells goods that sources from one particular supplier and meets a demand of X units which is random. The corresponding pdf $f(x)$ is apprehended to both farms. The unit production cost of the supplier is $p > 0$, and the retailer sells at the price of $r_p > c$ to the final market. p and r_p are both exogenous. The retail price for the retailer is $p(1 + g_s)$ per unit, where the supplier's gross margin is g_s per unit on sales. Also, the retailer's gross margin is g_r per unit for the wholesale market. Also, there is no salvage value for the unsold inventory in each firm. The supplier can ask for funds to a maximum of $\kappa_s \geq 0$, and at the end of the production or loan period, he must repay the principal with interest. Let, the rate of interest be i%. Besides the risk-free (r_f) rate, the rate of interest (i) may cause transaction costs in the money market charges when assigning the loan. Borrowing to be economically achievable for the customer if $i < g_s$ but setting $i > 0$ causes only a consistent change in the decision, therefore estimated $i = 0$ without loss of generality. For $j \in \{r, s\}$, let q_j^n denote the news vendor optimum of each firm: $q_j^n \equiv F^{-1}(\alpha_j)$, where $\alpha_j = \frac{g_j}{1+g_j}$ is the relevant analytical factor. The supplier would carry off pq_s^n and produce q_s^n units. $0 \leqslant \kappa_s < pq_s^n$ is needed to assure that purchase order financing is consistent when g_s is exogenous. The retailer can increase profit with the commitment to purchase $m > 0$ units. The commitment slackens of the predefined amount is the financial constraint on the supplier. It reaches his credit limit, but only to a fraction of its total value called the financial-parameter, denoted by $\lambda_s \in [0, 1]$, that is, a consignment m allows extra borrowing of $p\lambda_s m(1 + g_s)$. If $\lambda_s = 1$, the supplier is sufficiently informationally transparent, then the assurance of the retailer can be met. If $\lambda_s = 0$, the supplier is informationally opaque, and there is no indication that the supplier will comply with the contract.

Fig. 2 shows at $t = 0$, supplier and retailer observe the cdf of X. Retailer ensures to purchase $m \geq 0$ units and supplier chooses to produce $q(m)$ quantity. At $t = 1$, the actual market demand x is realized. If $x \leq m$, retailer purchase the dedicated volume. Otherwise, the retailer orders x unit and the minimum of (q, x) unit is delivered by the supplier.

The Supplier's Problem:

$$maximize\ Z_s(q) = E_X[p(1 + g_s)\max(m, \min(q, X))] - pq \tag{1}$$

$$subject\ to, \quad pq \leq \kappa_s + mp\lambda_s(1 + g_s) \tag{2}$$

The profit of the supplier is quantified by objective function (1). The inequality (2) represents the financial constraint of the supplier which implies that the production value cannot pass the investment available from debt. and the loan given by the financial institution.

Supplier's Optimum Manufacturing Decision: According to the retailer's predefined commitment m, the supplier's optimum manufacturing volume decision $q^*(m)$ is as follows:

$$q^*(m) = \begin{cases} q_s^k + m\lambda_s(1 + g_s) & if \quad 0 \leq m^n \\ q_s^n & if \quad m^n \leq m < q_s^n \\ m & if \quad q_s^n \leq m \end{cases} \tag{3}$$

The detailed proof of the supplier's optimum manufacturing decision is described in Appendix A.

The Retailer's Problem:

$$maximize\ Z_r(\omega) = E_X[r_p \min(q^*(m), X) - p(1 + g_s)\max(m, \min(q^*(m), X))] \tag{4}$$

$$subject\ to, \quad mp(1 + g_s) \leq \kappa_r + p\lambda_r q^*(m)(1 + g_s) \tag{5}$$

The objective function (4) quantifies the retailer's profit. The retailer's earnings depend on market demand and the supplier's production decision. The inequality (5) represents the financial constraint of the retailer which implies that the cost of the committed amount cannot exceed the financing possible from debt and the loan given by the financial institution.

Retailer's Optimum Commitment: In equilibrium, the retailer always commits. The stability commitment amount m^* is:

$$q^*(m) = \begin{cases} \max(\tilde{m}, m_r^s) & if\ \ 0 < \frac{\lambda_r}{\hat{\lambda}_r} < \frac{\hat{\lambda}_s}{\lambda_s} \\ \max(m_r^n, m_r^k) & if\ \ m^n \leq \frac{\hat{\lambda}_r}{\lambda_s} < 1 \\ m^n & otherwise \end{cases} \tag{6}$$

The detailed proof of the retailer's stability commitment amount is described in Appendix B.

The Lender's Problem: The lender's problem is to set the financial parameter (λ_r, λ_s) for the borrower and forecasting bankruptcy. The information asymmetry of borrowers brings the risk for lenders. To identify whether an applicant has default risk (i.e., a potential bad borrower) or not (i.e. a potential good borrower) becomes significant. Therefore we proposed a method for classifying unidentified applicants into good borrowers or non-defaulter and bad borrowers or defaulter, which combines the advantages of machine learning methods and recommender systems.

Theorem 3.1. *Financial parameter $\lambda(\lambda_r, \lambda_s)$ depends on the cost of debt and the ratio of the debt-to-equity.*

The detailed proof of Theorem 3.1 is described in Appendix C. From the three cases of the proof, it can be conclude that the financial parameter depends on the cost of debt and the ratio of the debt-to-equity.

4. Methodology

We formulate the purchase order financing model concerning demand uncertainty and the financial capabilities of MSMEs (suppliers [(1)] and retailers [(4)]). Our experiment came up with the retailers' stability commitment amount [(3)] and suppliers' optimum manufacturing decision [(6)]. The novel finding is how a retailer's optimum commitment and supplier's optimum manufacturing decision depend on the financial parameter. Since optimum commitment depends on the financial capabilities of MSMEs, financial institution (public sector bank and private sector bank) plays an important role in supply chain finance. Banks provide loans to MSMEs to meet the demand. Therefore, banks have to identify potential good borrowers and defaulters to minimize their risk and maximize profit. Researchers have applied various machine learning algorithms such as Neural Network and regression analysis for prediction. Ensemble ML techniques have drawn the attention of researchers in the domain of borrower identification and credit risk predictions because of their high performance and strategic efficiency. We proposed a novel k-Random Boosting Classification (k-RBC) algorithm (Fig. 3) to solve the lender's problem. Wang and Ma (2012) proposed the RSB-SVM method based on bagging and random subspace and Support Vector Machine for the credit risk assessment. Pan et al. (2021) have proposed a hybrid model (CF-NN) by combining collaborative filtering and neural networks. to identify risky borrowers and minimize risk for the lenders. Our model introduces a hybrid k-RBC algorithm to classify potential borrowers and defaulters. We have selected a top-k similar borrower for a new applicant to improve the accuracy and random forest regression for better prediction. Let, $B = \{b_i | i \in \{1, 2, \dots, n\}\}$ and $A = \{a_j | j \in \{1, 2, \dots, m\}\}$ represent the set of n borrowers and m attributes of each borrowers respectively. $C = \{c_{ij} | i \in \{1, 2, \dots, n\}, j \in \{1, 2, \dots, m\}\}$ denotes attribute matrix containing $n \times m$ elements of the borrower and $D = \{d_i | i \in \{1, 2, \dots, n\}\}$ is the set of classified results. We consider d_i as a binary variable, which represents that b_i is a non-default or good borrower if $d_i = 0$ and default or bad borrower if $d_i = 1$.

4.1. Data cleaning

Since the attributes of the data are in different dimension, we normalized the data into $[0, 1]$ by the following formula:

$$\alpha_{ij} = \frac{c_{ij} - \min_j c_{ij}}{\max_j c_{ij} - \min_j c_{ij}} \tag{7}$$

where $\mathfrak{C} = \{\alpha_{ij} | i \in \{1, 2, \dots, n\}, j \in \{1, 2, \dots, m\}\}$ is the normalized result of C and $\alpha_{ij} \in [0, 1]$ for all $i \in \{1, 2, \dots, n\}$ and $j \in \{1, 2, \dots, m\}$.

4.2. Top-k selection

To select top k similar borrower, we use a non parametric method. Therefore, for a target applicant or unidentified borrower b_p, we find k most similar borrowers from the existing database by formula (8), where $dist(b_p, b_q)$ is the distance between unidentified borrower b_p and identified borrower $b_q (q = \{1, 2, \dots, n\})$ denoted by the formula (9). K_{b_p} is the set of existing borrower whose distance with the unidentified borrower is less than δ, which is user defined parameter.

$$K_{b_p} = \{b_q | b_q \in B, dist(b_p.b_q) < \delta\} \tag{8}$$

$$dist(b_p, b_q) = \sqrt{\sum_{j=1}^{m} (\alpha_{pj} - \alpha_{qj})^2} \tag{9}$$

After that we add the distance and the index of the records to an ordered collection. Thereafter sort the distances in increasing order. Also, pick the first k entries from the sorted collection and return top k most similar borrowers from the records. Also, Pearson correlation coefficient (PPC) is used to similarity measurement, which shows the numerical distance between two borrowers. The similarity of b_p and b_q can be written as:

$$sim(b_p, b_q) = \frac{\sum_{j=1}^{m} (\alpha_{pj} - \bar{\alpha_p})(\alpha_{qj} - \bar{\alpha_q})}{\sqrt{\sum_{j=1}^{m} (\alpha_{pj} - \bar{\alpha_p})^2} \sqrt{\sum_{j=1}^{m} (\alpha_{qj} - \bar{\alpha_q})^2}} \tag{10}$$

where $\bar{\alpha_p}$ and $\bar{\alpha_q}$ are the attribute average values of unidentified borrower b_p and identified borrower b_q. The value range of sim is -1 to 1. The negative values denote dissimilarity between two borrowers. Based on the similarities calculated by the formula (10), for an unidentified borrower b_p, k identified borrowers are selected as b_p's neighbors based on similarities in descending order.

4.3. Borrower identification

After getting K_{b_p} we calculate ζ and ξ by the formula (11) and (12) respectively. Also, set ϵ a strong threshold.

$$\zeta = \frac{1}{k} \sum_{q=1}^{k} d_q \tag{11}$$

where $d_q \in \{0, 1\}$ is the classification result of the existing borrower $b_q \in K_{b_p}$.

$$\xi = 1 - \zeta \tag{12}$$

If $\zeta \leq \epsilon$ for a unidentified borrower b_p and the result for that borrower from RFC is 0, then we classify that borrower b_p as a potential good borrower. If lender wants to minimize his risk, then he will provide loan to those potential good borrower. To identify potential bad borrower or defaulter he can choose the borrower whose output from RFC is 1 and corresponding $\zeta > \epsilon$.

4.4. Performance metrics

To evaluate the performance of the classification (binary) model, we illustrates the confusion matrix, which represents two numbers accurately predicted good borrowers or non-defaulter and bad borrowers or defaulter as true positive (TP) and true negative (TN) respectively.

Precision is defined as the ratio of relevant instances among the retrieved instances. The defaulter precision (BP) and non-defaulter precision (GP) can be stated as follows:

$$BP = \frac{TN}{TN + FN} \tag{13}$$

$$GP = \frac{TP}{TP + FP} \tag{14}$$

The defect detection rate or recall is defined by the ratio of the total amount of relevant instances that were actually retrieved. The recall for defaulter (BR) and non-defaulter (GR) can also be described as the formulas (15) and (16) respectively.

$$BR = \frac{TN}{TN + FP} \tag{15}$$

$$GR = \frac{TP}{TP + FN} \tag{16}$$

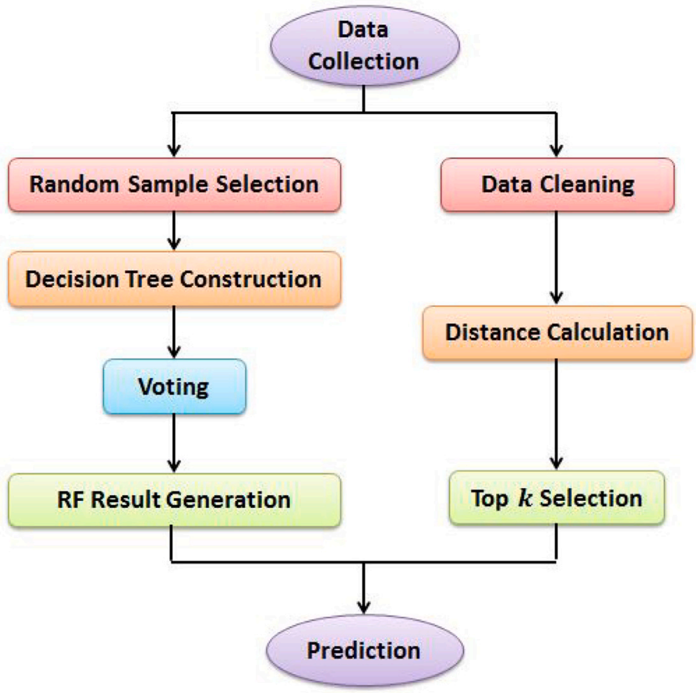


Fig. 3. The structure of the model.

Table 1
Performance matrix.

Performance Matrix		Predicted Classification	
		Defaulter	Non-defaulter
Actual Classification	Defaulter	TN	FP
	Non-defaulter	FN	TP

There is a trade off between recall and precision. Therefore F-measure is used for balancing precision and recall in classification evaluation. The F-measure for bad borrower (BF) and good borrower (GF) can be defined by the formulas (17) and (18) respectively.

$$BF = \frac{2 \times BR \times BP}{BR + BP} \tag{17}$$

$$GF = \frac{2 \times GR \times GP}{GR + GP} \tag{18}$$

The accuracy (correct classification rate) for evaluating the overall performance of the method can be calculated by the formula (19).

$$AC = \frac{TN + TP}{TN + TP + FN + FP} \tag{19}$$

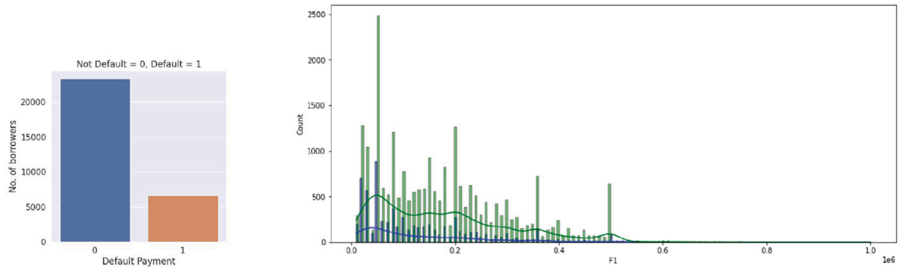
Theorem 4.1. *The time complexity of the k-RBC algorithm is $\mathcal{O}(n^2 log(n))$.*

Proof. Let the dataset contains m attributes and n number of borrowers. Therefore the depth of the tree is $\mathcal{O}(log(n))$. Since our dataset contains m attributes at each node the complexity to build one tree is $\mathcal{O}(mlog(n))$. Now for building RF with n number of borrowers, the complexity would be $\mathcal{O}(mnlog(n))$. The complexity for distance calculation is $\mathcal{O}(n(n-1))$ i,e; $\mathcal{O}(n^2)$. Therefore time complexity of k-RBC algorithm is $\max\{\mathcal{O}(mnlog(n)), \mathcal{O}(n^2)\}$ i.e; for n attributes and n number of borrowers time complexity of k-RBC algorithm is $\mathcal{O}(n^2 log(n))$. □

4.5. Dataset description

In this work, we have used the "Taiwan" dataset of Taiwan's clients' default cases which has 23 features ($F1, F2, .., F23$) and 30,000 instances (data source: 'https://archive.ics.uci.edu/ml/machine-learning-databases/00350/"), out of which 6636 (22.1%) are default cases. Therefore, this is a unbalanced dataset (Fig. 4(a)). The same dataset has also been used in some other research works. Some of the features of this dataset are credit limit ($F1$), gender ($F2$: 1 = male; 2 = female), educational qualification ($F3$: 1 = graduate school; 2 = university; 3 = high school; 4 = others), marital status ($F4$: 1 = married; 2 = single; 3 = others), age ($F5$), last 6 months bills payment status ($F6, \ldots, F11$), last 6 months bill amounts ($F12, \ldots, F17$), and last 6 months re-payment status ($F18, \ldots, F23$). Records are labeled (Y) as either 0 (non-default) or 1 (default). In our experiment, we consider $F2, F3, F4, F6, F7, F8, F9, F10, F11$ as categorical features and the rest of the features as nominal features. Statistical analysis is listed (Table 2) in terms of minimum, maximum, mean, standard deviation, and distribution for the nominal features. In Fig. 4(b), the x-axis represents the credit limit of the borrower and the y-axis represents the number of borrowers. It signifies the default amount with respect to the credit limit. The box plot (Fig. 5) indicates the lower extreme, lower quartile ($Q1$), median ($Q2$), upper quartile ($Q3$), the upper extreme of the borrowers in terms of non-defaulter and defaulter concerning educational qualification and age. Such as, the $Q2$ and $Q3$ are almost similar, but the $Q1$ of defaulters is larger than the non-defaulter for educational qualification 6. Box plot of the borrowers concerning the educational qualification and credit limit (Fig. 6) suggests that Q1, Q2, and Q3 are higher for non-defaulters, and also population density of defaulters are minimum. Fig. 7 indicates that $Q1, Q2$, and $Q3$ are increasing for both non-defaulter and defaulter with age until around 32 years, and they are oscillating and get to the maximum of the upper extreme for non-defaulter at age 46. Also, it signifies that $Q1, Q2$, and $Q3$ are higher for no-defaulter to the defaulter. The correlation graph (Fig. 8) intimates that correlations are declining with the distance between months. The lowest correlations are between April and September (see Table 1).

To explain the black box AI models, Giudici and Raffinetti (2021) have proposed a model based on the Shapley-Lorenz decomposition. Also, they have used the Marginal Gini Contribution and the partial Gini Contribution based on Lorenz Zonoids to measure a single covariate's absolute explanatory power and a new covariate's addi-



(a) Unbalance Data (b) Default amount concerning the Credit Limit

Fig. 4. Data visualization.

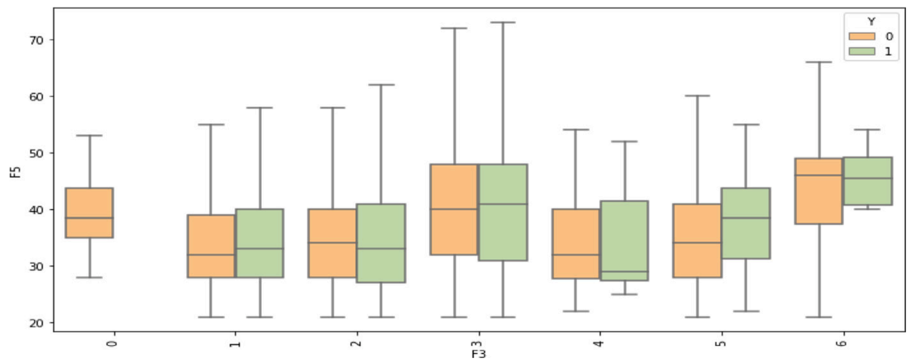


Fig. 5. Box plot of the borrowers concerning the educational qualification and age.

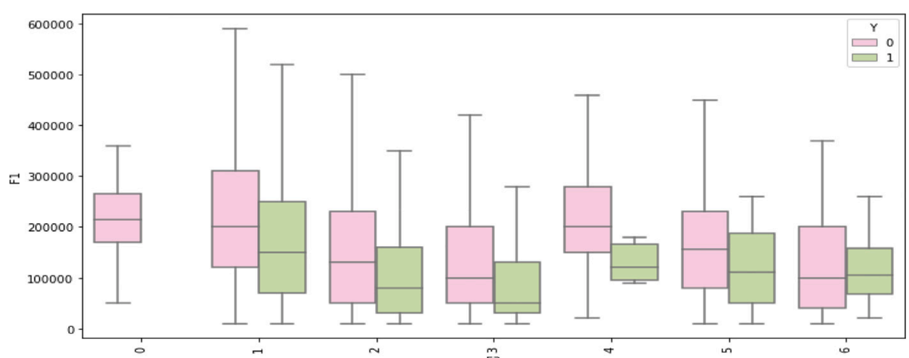


Fig. 6. Box plot of the borrowers concerning the educational qualification and credit limit.

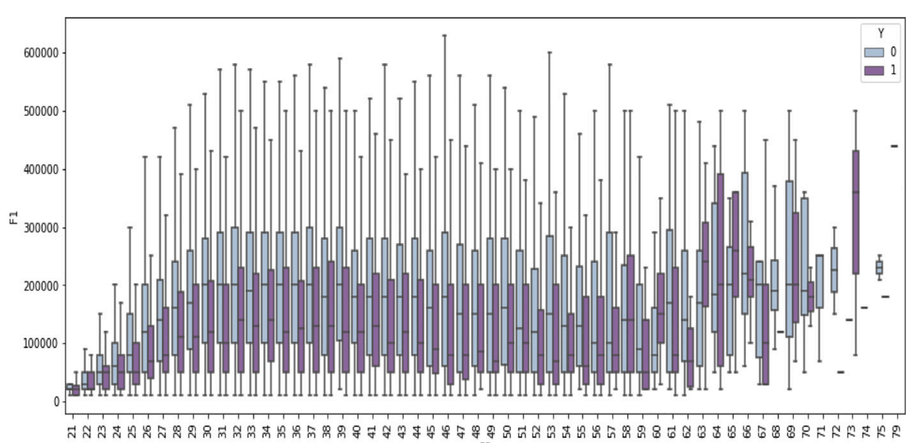
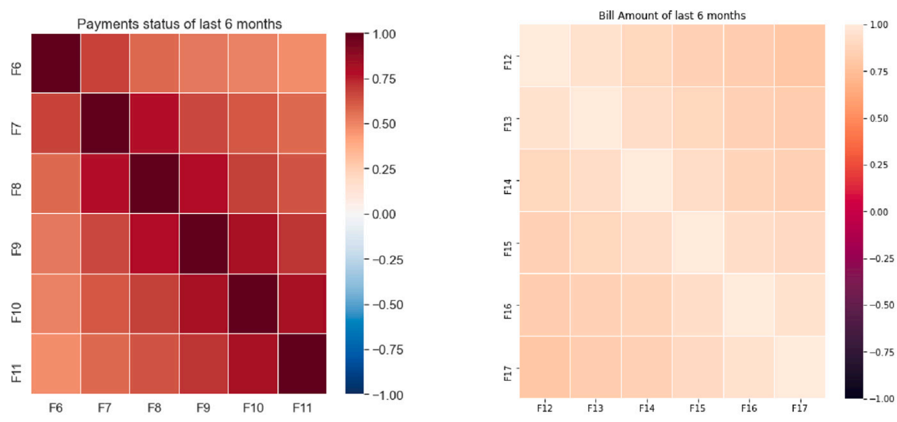


Fig. 7. Box plot of the borrowers concerning the age and credit limit.

Table 2
Statistical analysis of nominal features.

	mean	σ	min	25%	50%	75%	max
F1	167484.3227	129747.6616	10000	50000	140000	240000	1000000
F12	51223.3309	73635.86058	−165580	3558.75	22381.5	67091	964511
F13	49179.07517	71173.76878	−69777	2984.75	21200	64006.25	983931
F14	4.70E+04	6.93E+04	−1.57E+05	2.67E+03	2.01E+04	6.02E+04	1.66E+06
F15	43262.94897	64332.85613	−170000	2326.75	19052	54506	891586
F16	40311.40097	60797.15577	−81334	1763	18104.5	50190.5	927171
F17	38871.7604	59554.10754	−339603	1256	17071	49198.25	961664
F18	5663.5805	16563.28035	0	1000	2100	5006	873552
F19	5.92E+03	2.30E+04	0.00E+00	8.33E+02	2.01E+03	5.00E+03	1.68E+06
F20	5225.6815	17606.96147	0	390	1800	4505	896040
F21	4826.076867	15666.1597	0	296	1500	4013.25	621000
F22	4799.387633	15278.30568	0	252.5	1500	4031.5	426529
F23	5215.502567	17777.46578	0	117.75	1500	4000	528666



(a) Correlation graph of last 6 months payment status

(b) Correlation graph of last 6 months bill amounts

Fig. 8. Correlation graph.

Fig. 9. Features contribution to the prediction using LIME.

tional contribution to an existing model respectively. In the same line, Fig. 9 illustrates the explanations produced with Local Interpretable Model-agnostic Explanations (LIME) for the test data. Class 0 and class 1 represent the good borrower and defaulter respectively. The blue-colored features support class 0, and the orange-colored features support class 1 (defaulter). The real numbers on the horizontal bars depict the significance of the important features. Also, we have explained Shapley values to capture the consistency and local accuracy of our model. Fig. 10 represents the local explanations created with SHAP for the test data. The magnitude and direction across the features denote the impacts on the model output. We have used LIME and SHAP python libraries for these explanations.

5. Results

Fig. 11 represents the Receiver Operator Characteristic (ROC) curve, where a higher X-axis value implies a higher number of False positives than True negatives, while a higher Y-axis value implies a higher number of True positives than False negatives. The area under the ROC (AUROC) curve is 0.8579, which indicates that the performance of our model is excellent. For evaluating the performances of proposed method, we first compare them with classical binary classification methods, including Naive Bayes (NB), Logistic Regression (LR), Decision Trees (DT), Random Forests (RF), K-Nearest Neighbor (KNN), Support Vector Machines (SVM), Neural Network (NN), Collaborative Filtering (CF), hybrid random subspace & Bagging SVM (RSB-SVM) and hybrid Collaborative Filtering & Neural Network (CF-NN). All

Table 3
Performance of the machine learning algorithms.

Methods	Defaulter identification			Non-defaulter identification			AC
	BP	BR	BF	GP	GR	GF	
k-RBC	0.70	0.62	0.50	0.90	0.95	0.92	0.85
NN	0.62	0.31	0.42	0.82	0.94	0.88	0.80
CF	0.20	0.26	0.22	0.87	0.92	0.90	0.82
SVM	0.67	0.16	0.27	0.80	0.97	0.88	0.79
NB	0.60	0.32	0.41	0.82	0.93	0.88	0.80
LR	0.67	0.14	0.23	0.79	0.98	0.88	0.80
DT	0.35	0.37	0.36	0.82	0.93	0.87	0.71
RF	0.59	0.31	0.41	0.81	0.93	0.87	0.80
KNN	0.51	0.30	0.37	0.82	0.91	0.86	0.78
GB	0.68	0.12	0.21	0.79	0.89	0.88	0.79
RSB-SVM	0.69	0.13	0.27	0.79	0.97	0.89	0.79
CF-NN	0.43	0.45	0.48	0.9	0.89	0.92	0.84

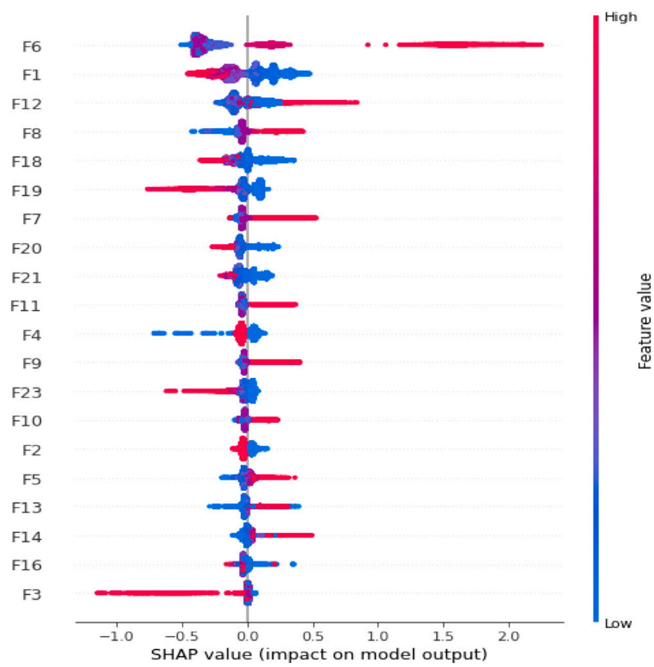


Fig. 10. Features contribution to the prediction using SHAP.

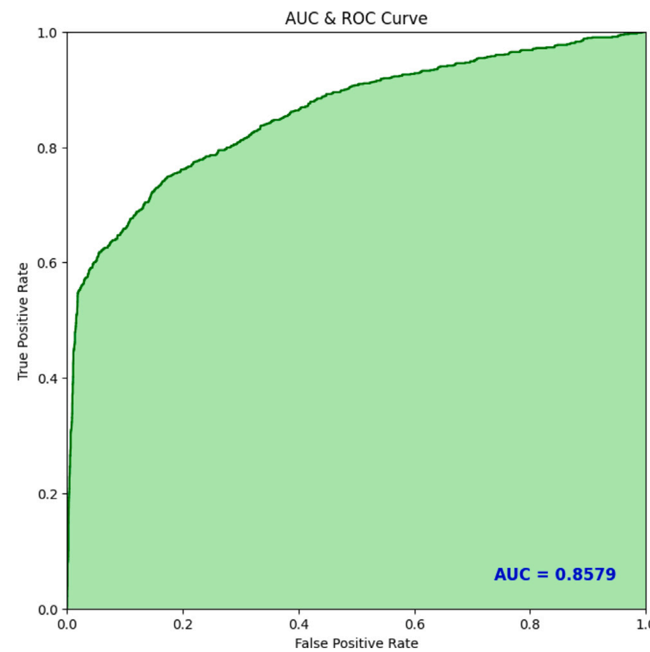


Fig. 11. ROC & AUC curve.

experiments were carried out on Python3.0 with the library Scikit-learn (sklearn) for machine learning algorithms on a HP workstation with Intel Core i5-4210U processor, which has 2 cores and 4 GB RAM. The parameters used in this model are $n\_estimators$, $max\_features$, max_depth, $min\_samples\_split$, bootstrap, $random_stat$e, $learning\_rate$. The parameters $n\_estimators$ and $max\_features$ create the number of trees in RBC. The larger tree is better for accuracy but it will take longer to compute. And also, beyond a certain number of trees, the outcomes will stop obtaining better solutions. Therefore, we choose $n\_estimators = 100$. To consider all the features we set $max\_features = None$. For better prediction and completely devising the trees, we used the combination of $max\_depth = 1$ and $min\_samples\_split = 2$. The settings of the parameters are $bootstrap = True$, $random\_state = 0$, and $learning\_rate = 1.0$ for easier training. In this experiment 75% data (i.e. 22 500 samples) is randomly selected as a training set that is used to train the model, and 25% data (i.e. 7500 samples) is used for testing to evaluating model performance. All performance metrics are used to validate the k-RBC model.

Our data set contains 77.9% of samples of good borrowers. So accuracy of good borrower identification is higher than bad borrower identification. The accuracy of k-RBC algorithm for bad borrower precision is 70%, recall of bad borrower is 62%, F-measure for bad borrower is 50%, good borrower precision is 90%, recall of good borrower is 95% and F-measure for good borrower 92%. Overall accuracy of our algorithm is 85%, which is better than any other existing algorithm (Table 3). While comparing with the existing algorithms, we have used python libraries and its default settings. The parameters for SVM are The parameters we have used for SVM are $Cfloat = 1.0$, $kernel =' rbf'$, $degreeint = 3$, and $decision\_function\_shape =' ovr'$; for collaborative filtering algorithm are $metric =' cosine'$, $n\_neighbors = 3$, $n\_jobs = -1$; for decision tree algorithm are $criterion =' gini'$, $splitter =' best'$, $min\_samples\_split = 2$, and $min\_samples\_leaf = 1$; for random forest classifier are $n\_estimator = 100$, $criterion =' gini'$, $min\_samples\_split = 2$, and $min\_samples\_leaf = 1$; for KNN classifier are $weights =' uniform', algorithm =' auto', leaf\_size = 30$, and $metric =' minkowski'$; for gradient boosting algorithm are $loss =' deviance', learning\_rate = 0.1$, and $n\_estimator = 100$; for CF-NN are $num\_neg = 4, epochs = 20, batch\_size = 256$, and $learning\_rate = 0.001$. From the performance graph (Fig. 12) we can see that our proposed methodology (k-RBC) improved the accuracy of selecting potential good borrower and potential bad borrower.

6. Conclusion

From the purchase order financing, it can be concluded that the economic features of the supplier control the operational decisions and

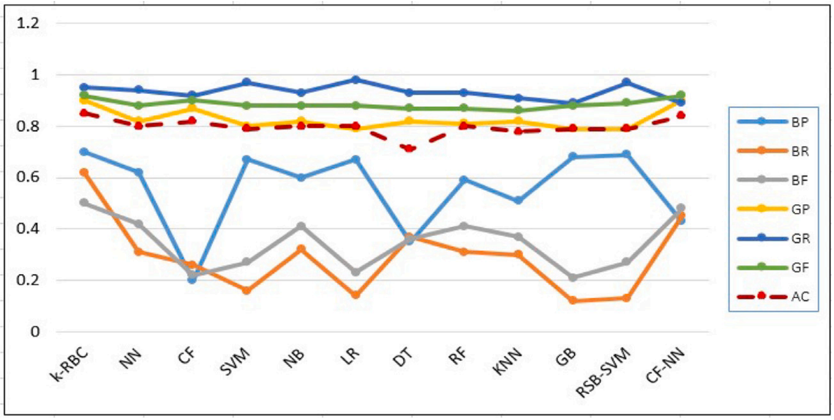


Fig. 12. Performance graph.

profits of both retailer and supplier. It can be shown that the purchase quantity funding significantly alleviates the impact of the capital market friction. In an SCF network, the financial parameter depends on the cost of debt and the ratio of the debt-to-equity. Also, the impact of the financial parameter in the financial supply chain is analyzed and both the total asset and the asset inequality may affect the stability of the financial network. Here this study contributes to the literature on purchase order financing concerning the financial conditions of MSMEs (retailers and suppliers). Our novel finding is how a retailer's optimum commitment and supplier's optimum manufacturing decision depend on the financial parameter. Also, we introduced the k-RBC algorithm to identify potential good or potential bad borrowers for lenders to avoid defaulters. Ghildiyal et al. (2022) applied ML algorithms such as RF, SVM to automate the loan approval for banks with 82% precision. Singh et al. (2021) used the RF classifier, support vector machine classifier to predict whether the Client will pay back the loan or not with an accuracy of 77%. Whereas, our k-RBC algorithm gives 90% accuracy for good borrower prediction. The k-RBC algorithm performs better than existing algorithms and time complexity is $\mathcal{O}(n^2 log(n))$ same as RF but requires less space than RF. In the future, a framework will be developed to identify the structure of the liability matrix such that the resulting system is the most stable one and also originate some policies which can be implemented in advance to prevent the catastrophic disaster. Also, suggest that the financing model based on multi-source information can be developed and explore the details regarding the industrial network.

CRediT authorship contribution statement

Rony Mitra: Methodology, Writing – original draft, Writing – review & editing, Software, Validation. **Adrijit Goswami:** Investigation, Conceptualization, Supervision. **Manoj Kumar Tiwari:** Resources, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

The authors acknowledge the support provided by Indian Institute of Technology Kharagpur.

Appendix A. Supplier's optimal production decision

Proof. Use Lagrange multiplier μ_s and Lagrange function $L(q,\mu_s)$ to solve the supplier's problem.

$$L(q,\mu_s) = Z_s(q) - \mu_s[pq - \kappa_s - mp\lambda_s(1+g_s)] \tag{20}$$

Now use Karush–Kuhn–Tucker (KKT) conditions to use optimal solution:

$$\frac{\partial L}{\partial q} = 0 \;\Rightarrow\; \frac{\partial}{\partial q}\left\{Z_s(q)\right\} - \mu_s p = 0 \tag{21}$$

$$\begin{aligned}\because Z_s(q) &= E_x[p(1+g_s)\max(m,\min(q,X))] - pq\\ &= p(1+g_s)\left[\int_0^m mf(x)dx + \int_m^q xf(x)dx + \int_q^\infty qf(x)dx\right] - pq\end{aligned} \tag{22}$$

$$\therefore p(1+g_s) + \int_q^\infty f(x)dx - p - \mu_s p = 0$$
$$\Rightarrow p(1+g_s) + (1-F(q)) - p - \mu_s p = 0 \quad [From\ (21)]$$

Case-1: If $\mu_s = 0$ then $p(1+g_s)(1-F(q)) - p = 0 \Rightarrow F(q) = \frac{g_s}{1+g_s}$ Therefore, $q^*(m) = F^{-1}\left(\frac{g_s}{1+g_s}\right) \equiv q_s^n$. Now from the constraint (2),

$$\begin{aligned} pq^*(m) - \kappa_s - mp\lambda_s(1+g_s) \le 0 &\Rightarrow q^*(m) \le \frac{\kappa_s}{p} + m\lambda_s(1+g_s)\\ &\Rightarrow q_s^n \le q_s^k + m\lambda_s(1+g_s)\end{aligned} \tag{23}$$

Where q_s^k ≡The production that the supplier can independently finance= $\frac{\kappa_s}{p}$. Let the loan amount, $q_s^\Delta = q_s^n - q_s^k$ and $m^n \equiv \frac{q_s^\Delta}{\lambda_s(1+g_s)}$. Then from (23),

$$q_s^\Delta \le m\lambda_s(1+g_s) \Rightarrow \frac{q_s^\Delta}{\lambda_s(1+g_s)} \le m \Rightarrow m^n \le m \tag{24}$$

Let $\hat{\lambda}_s$ be the minimum allowable value of λ_s. Therefore $\hat{\lambda}_s \equiv \frac{1}{1+g_s} \le \lambda_s \Rightarrow \frac{1}{\lambda_s(1+g_s)} \le 1$ which implies $m^n \le q_s^\Delta \le q_s^n$. Also it is clear that $q_s^n \le m$ is sub-optimal. Since the retailer has committed to buy m quantity, therefore $q^*(m) = m$ is optimal when $m \ge q_s^n$.

Case-2: If $\mu_s > 0$ then,

$$\mu_s = (1+g_s)(1-F(q)) - 1 > 0 \tag{25}$$

$$\therefore \frac{\partial L}{\partial \mu_s} = 0 \Rightarrow pq - \kappa_s - mp\lambda_s(1+g_s) = 0 \Rightarrow q^*(m) = q_s^k + m\lambda_s(1+g_s) \tag{26}$$

The inequality (42) is satisfied for $0 \le m < m^n$. Also $\lambda_s \ge \hat{\lambda}_s$ and $\kappa_s \ge 0$ guarantee the feasibility of the solution. Therefore the optimal solution for the supplier is,

$$q^*(m)=\begin{cases} q_s^k+m\lambda_s(1+g_s) & if \quad 0\le m^n\\ q_s^n & if \quad m^n\le m<q_s^n\\ m & if \quad q_s^n\le m\end{cases} \quad \square \tag{27}$$

Appendix B. Retailer's optimal commitment

Proof. Use Lagrange multiplier $\mu_r \ge 0$ and Lagrange function $L(m,\mu_r)$ to solve the retailer's problem.

$$L(m,\mu_r)=Z_r(m)-\mu_r[mp(1+g_s)-\kappa_r-r_p\lambda_r q^*(m)] \tag{28}$$

Now use KKT conditions to find the optimal solutions:

$$\therefore \frac{\partial L}{\partial m}=0 \;\Rightarrow\; \frac{\partial}{\partial m}Z_r(m)-\mu_r\left[p(1+g_s)-r_p\lambda_r\frac{\partial}{\partial m}\{q^*(m)\}\right]=0 \tag{29}$$

$$\begin{aligned}\because Z_r(m)&=r_pE_x\left[\min(q^*(m)),X\right]-p(1+g_s)E_x\left[\max(m,\min(q^*(m),X))\right]\\ &=r_p\left[\int_0^{q^*(m)}xf(x)dx+\int_{q^*(m)}^{\infty}q^*(m)f(x)dx\right]\\ &\quad -p(1+g_s)\left[\int_0^m mf(x)dx+\int_m^{q^*(m)}xf(x)dx\right.\\ &\quad \left.+\int_{q^*(m)}^{\infty}q^*(m)f(x)dx\right]\end{aligned} \tag{30}$$

$$\begin{aligned}\therefore\frac{\partial}{\partial m}\{Z_r(m)\}&=r_p\frac{\partial}{\partial m}\{q^*(m)\}\int_{q^*(m)}^{\infty}f(x)dx-p(1+g_s)\\ &\quad\times\left[\int_0^m f(x)dx+\frac{\partial}{\partial m}\{q^*(m)\}\int_{q^*(m)}^{\infty}f(x)dx\right]\\ &=[r_p-p(1+g_s)]\frac{\partial}{\partial m}\{q^*(m)\}\\ &\quad\times\int_{q^*(m)}^{\infty}f(x)dx-p(1+g_s)\int_0^m f(x)dx\end{aligned} \tag{31}$$

Constrained Solution: Here in this case $0\le m<m^n$. So $q^*(m)=q_s^k+m\lambda_s(1+g_s)$. Since $r_p=p(1+g_s)(1+g_r)$, therefore $r_p-p(1+g_s)=pg_r(1+g_s)$

$$\therefore\frac{\partial}{\partial m}Z_r(m)=p\lambda_sg_r(1+g_s)^2\left[1-F(q^*(m))\right]-p(1+g_s)F(m) \tag{32}$$

$$\begin{aligned}\therefore\, & p\lambda_sg_r(1+g_s)^2\left[1-F(q^*(m))\right]-p(1+g_s)F(m)\\ &-\mu_r\left[p(1+g_s)-r_p\lambda_r\frac{\partial}{\partial m}\{q^*(m)\}\right]\quad [From\ (29)]\end{aligned}$$

If $\mu_r=0$ then,

$$\begin{aligned}&p\lambda_sg_r(1+g_s)^2\left[1-F(q^*(m))\right]-p(1+g_s)F(m)=0\\ &\Rightarrow F(m)=\lambda_sg_r(1+g_s)\overline{F}(q^*(m))\end{aligned} \tag{33}$$

Let the point $m=\tilde{m}$ satisfies Eq. (47). Now from the constraint (45),

$$\begin{aligned}&mp(1+g_s)\le\kappa_r+r_p\lambda_r[q_s^k+m\lambda_s(1+g_s)]\\ &\Rightarrow m(1+g_s)\left[p-r_p\lambda_r\lambda_s\right]\le\kappa_r+r_p\lambda_rq_s^k\\ &\Rightarrow m\le\frac{\kappa_r+r_p\lambda_rq_s^k}{(1+g_s)\left[p-r_p\lambda_r\lambda_s\right]}\equiv m_r^s\end{aligned} \tag{34}$$

Since, $m>0$, therefore $p-r_p\lambda_r\lambda_s>0\Rightarrow\frac{\lambda_r}{\hat{\lambda}_r}<\frac{\hat{\lambda}_s}{\lambda_s}$ where $\hat{\lambda}_r$ be the minimum allowable value of λ_r. Therefore $\hat{\lambda}_r\equiv\frac{1}{1+g_r}\le\lambda_r$. If $\mu_r>0$ then,

$$\begin{aligned}\frac{\partial L}{\partial\mu_r}=0&\Rightarrow mp(1+g_s)-\kappa_r-r_p\lambda_rq^*(m)=0\\ &\Rightarrow mp(1+g_s)-\kappa_r-r_p\lambda_r[q_s^k+m\lambda_s(1+g_s)]=0\\ &\Rightarrow m(1+g_s)\left[p-r_p\lambda_r\lambda_s\right]=\kappa_r+r_p\lambda_rq_s^k\\ &\Rightarrow m=\frac{\kappa_r+r_p\lambda_rq_s^k}{(1+g_s)\left[p-r_p\lambda_r\lambda_s\right]}\equiv m_r^s\end{aligned}$$

Unconstrained Solution: Here in this case $m^n\le m<q_s^n$ and $q^*(m)=q_s^n$. Therefore $\frac{\partial}{\partial m}\{q^*(m)\}=0$

$$\therefore\frac{\partial L}{\partial m}=0\Rightarrow p(1+g_s)F(m)-\mu_r[p(1+g_s)]=0\Rightarrow F(m)=-\mu_r \tag{35}$$

Since $\mu_r\ge0$, therefore $F(m)\le0\Rightarrow F(m)=0$. Therefore the only possibility is $q^*(m)=q_s^n$ and $m=m^n$. Since at (m^n,q_s^n), $\frac{d}{dm}\{Z_r(m)\}=-p(1+g_s)F(m)<0$ and $\frac{d^2}{dm^2}\{Z_r(m)\}=-p(1+g_s)f(m)<0$, therefore $Z_r(m)$ is concave decreasing in $m^n\le m\le q_s^n$ and $Z_r(m)$ has local maximum at (m^n,q_s^n).

Fulfilling Solution: Here in this case $q_s^n\le m$ and $q^*(m)=m$.

$$\begin{aligned}\therefore\frac{\partial L}{\partial m}=0&\Rightarrow[r_p-p(1+g_s)]\int_m^{\infty}f(x)dx-p(1+g_s)\\ &\quad\times\int_0^m f(x)dx-\mu_r[p(1+g_s)-r_p\lambda_r]=0\\ &\Rightarrow r_p\int_m^{\infty}f(x)dx-p(1+g_s)\\ &\quad\times\int_0^{\infty}f(x)dx-\mu_r[p(1+g_s)-r_p\lambda_r]=0\\ &\Rightarrow r_p(1-F(m))-p(1+g_s)-\mu_r[p(1+g_s)-r_p\lambda_r]=0\end{aligned} \tag{36}$$

If $\mu_r=0$ then,

$$\begin{aligned}r_p(1-F(m))-p(1+g_s)=0\Rightarrow F(m)&=\frac{1}{r_p}[r_p-p(1+g_s)]\\ &=\frac{p(1+g_s)(1+g_r-1)}{p(1+g_s)(1+g_r)}=\frac{g_r}{1+g_r}\end{aligned}$$

$$\Rightarrow m=F^{-1}\left(\frac{g_r}{1+g_r}\right)\equiv m_r^n \tag{37}$$

Now from the constraint (45),

$$\begin{aligned}&mp(1+g_s)\le\kappa_r+r_p\lambda_rm\Rightarrow m[p(1+g_s)-r_p\lambda_r]\le\kappa_r\\ &\qquad\Rightarrow m\le\frac{\kappa_r}{p(1+g_s)-r_p\lambda_r}\equiv m_r^k\end{aligned} \tag{38}$$

Since $m>0$ and $\kappa_r\ge0$, therefore

$$\begin{aligned}&p(1+g_s)-r_p\lambda_r>0\Rightarrow p(1+g_s)[1-\lambda_r(1+g_r)]>0\\ &\qquad\Rightarrow\lambda_r(1+g_r)<1\Rightarrow\frac{\lambda_r}{\hat{\lambda}_r}<1\end{aligned} \tag{39}$$

If $\mu_r>0$ then,

$$\frac{\partial L}{\partial m}=0\Rightarrow mp(1+g_s)=\kappa_r+r_p\lambda_rm\Rightarrow m[p(1+g_s)-r_p\lambda_r]=\kappa_r\Rightarrow m=m_r^k \tag{40}$$

Therefore the optimal commitment for the retailer is,

$$m^*=\begin{cases}\min(\tilde{m},m_r^s) & if\ \ 0<\frac{\lambda_r}{\hat{\lambda}_r}<\frac{\hat{\lambda}_s}{\lambda_s}\\ \min(m_r^n,m_r^k) & if\ \ \frac{\hat{\lambda}_s}{\lambda_s}\le\frac{\lambda_r}{\hat{\lambda}_r}<1\\ m^n & otherwise\end{cases}\quad\square \tag{41}$$

Appendix C. Proof of Theorem 3.1

Proof. Let, cost of debt be b, return on asset (ROA) be $A(A\in[-1;1])$, realized return of debt (ROD) be $D(D\in[-1;b])$ and return on equity (ROE) be $E(E=A-D)$. To find x_1 at the point P (Fig. 13),

$$A(1+x_1)=D(1+b)\Rightarrow1+x_1=\frac{D}{D+E}(1+b)\Rightarrow x_1=\frac{\frac{D}{E}}{\frac{D}{E}+1}(1+i)-1 \tag{42}$$

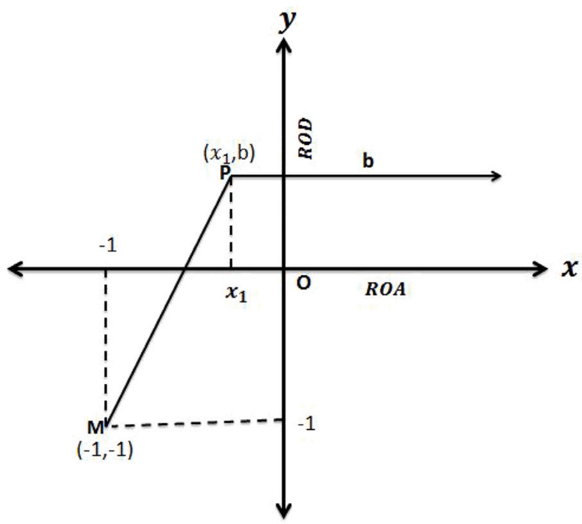


Fig. 13. ROA Vs ROD.

∴

The equation of the line MP,

$$\frac{y+1}{x+1} = \frac{b+1}{x_1+1} \Rightarrow y = \left(\frac{b+1}{x_1+1}\right) + \left(\frac{b-x_1}{x_1+1}\right) \tag{43}$$

Let, $y = mx + c$ where $x \in [-1, \infty], y \in [-1, 1], m = \frac{b+1}{x_1+1}$ and $c = \frac{b-x_1}{x_1+1}$. Therefore,

$$y = \begin{cases} b & \text{for } x \geq x_1 \\ mx + c & \text{for } x < x_1 \end{cases} \tag{44}$$

Since ROA is random. Let x is random variable, f_x and F_x are pdf and cdf of x respectively. Since y is a function of x then y is also a random variable. Now let μ_x and σ_x are the mean and standard deviation of x respectively. $E(y)$ and $V(y)$ are the mean and variance of y respectively. Where $E(y)$ = Expected return on debt and $V(y)$ = Risk on debt (Credit Risk). To prove the proposition we are going to show following three cases.

Case-1: Expected return $(E(y))$ is increasing as debt (b) increase. To prove this let $y = h(x)$

$$\begin{aligned}\therefore E(y) = E(h(x)) &= \int_{-1}^{\infty} h(x) f_x(x)dx \\ &= \int_{-1}^{x_1} (mx+c) f_x(x)dx + \int_{x_1}^{\infty} b f_x(x)dx\end{aligned}$$

$$\therefore E(y) = m\int_{-1}^{x_1} x f_x(x)dx + cF(x_1) + b(1 - F_x(x_1)) \tag{45}$$

Let $k = \frac{1}{\frac{D}{E}}$, therefore $x_1 = \frac{b-k}{1+k}, m = 1 + k$ and $c = k$. Therefore $\frac{\partial x_1}{\partial b} = \frac{1}{1+k}, \frac{\partial m}{\partial b} = 0$ and $\frac{\partial c}{\partial b} = 0$.

$$\begin{aligned}\therefore \frac{\partial E}{\partial b} &= m\frac{\partial}{\partial b}\int_{-1}^{x_1} x f_x(x)dx + (c-b)\frac{\partial}{\partial b}F_x(x_1) + 1 \\ &= mx_1 f_x(x_1)\frac{\partial x_1}{\partial b} + (c-b) f_x(x_1)\frac{\partial x_1}{\partial b} - F_x(x_1) + 1 \quad [Leibniz's\ rule] \\ &= mx_1 f_x(x_1)\frac{1}{1+k} + (c-b) f_x(x_1)\frac{1}{1+k} - F_x(x_1) + 1 \\ &= 1 - F_x(x_1) \geq 0 \quad [\because F_x(x_1) \leq 1]\end{aligned}$$

Therefore $\frac{\partial E}{\partial b} \geq 0$, which implies that as debt increase return is also increase.

Case-2: Expected return $(E(y))$ is increasing as the ratio of debt-to-equity $\left(\frac{D}{E}\right)$ decrease. Since k is the inverse of $\frac{D}{E}$, to prove this we will show that expected return $(E(y))$ is increasing as k increase. Also, $\frac{\partial x_1}{\partial k} = -\frac{1+b}{(1+k)^2}, \frac{\partial m}{\partial k} = 1$ and $\frac{\partial c}{\partial k} = 1$.

$$\begin{aligned}\therefore \frac{\partial E}{\partial k} &= m\frac{\partial}{\partial k}\int_{-1}^{x_1} x f_x(x)dx \\ &\quad + \frac{\partial m}{\partial k}\int_{-1}^{x_1} x f_x(x)dx + (c-b)\frac{\partial}{\partial k}F_x(x_1) + F_x(x_1)\frac{\partial c}{\partial k} \\ &= mx_1 f_x(x_1)\frac{\partial x_1}{\partial k} \\ &\quad + \int_{-1}^{x_1} x f_x(x)dx + (c-b) f_x(x_1)\frac{\partial x_1}{\partial k} + F_x(x_1) \quad [Leibniz's\ rule] \\ &= \int_{-1}^{x_1} x f_x(x)dx + F_x(x_1) \quad [\because mx_1 + c - b = b - k + k - b = 0]\end{aligned}$$

$$\begin{aligned}&\because (-1)\int_{-1}^{x_1} f_x(x)dx < \int_{-1}^{x_1} x f_x(x)dx < x_1\int_{-1}^{x_1} f_x(x)dx \\ &\quad \Rightarrow 0 < \frac{\partial E(y)}{\partial k} < (1 + x_1)F_x(x_1)\end{aligned} \tag{46}$$

Therefore return increase as the ratio of debt-to-ratio decrease.

Case-3: Risk on debt or credit risk $(V(y))$ increase as debt increase. We know that risk on debt $= V(y) = E(y^2) - \{E(y)\}^2$

$$\begin{aligned}\therefore V(y) &= \int_{-1}^{x_1} (mx+c)^2 f_x(x)dx + \int_{x_1}^{\infty} b^2 f_x(x)dx - E(y)^2 \\ &= m^2\int_{-1}^{x_1} x^2 f_x(x)dx + c^2\int_{-1}^{x_1} f_x(x)dx + 2mc \\ &\quad \times \int_{-1}^{x_1} x f_x(x)dx + b^2\int_{x_1}^{\infty} f_x(x)dx - E(y)^2\end{aligned}$$

$$\begin{aligned}\therefore \frac{\partial V}{\partial b} &= \left((1+k)^2 x_1^2 + k^2 + 2k(1+k) - b^2\right) f_x(x_1)\frac{\partial x_1}{\partial b} \\ &\quad + 2b(1 - F_x(x_1)) - 2E(y)\frac{\partial}{\partial b}E(y) \\ &= 2(1+k)(1 - F_x(x_1))\left\{x_1 F_x(x_1) - \int_{-1}^{x_1} x f_x(x)dx\right\}\end{aligned} \tag{47}$$

$$\begin{aligned}&\because (-1)\int_{-1}^{x_1} f_x(x)dx < \int_{-1}^{x_1} x f_x(x)dx < x_1\int_{-1}^{x_1} f_x(x)dx \\ &\Rightarrow F_x(x_1) > -\int_{-1}^{x_1} x f_x(x)dx > -x_1 F_x(x_1)\end{aligned}$$

$$\therefore (x_1 + 1)F_x(x_1) > F_x(x_1) - \int_{-1}^{x_1} x f_x(x)dx > 0 \tag{48}$$

Since $2(1+k)(1-F_x(x_1)) > 0$ and $x_1 F_x(x_1) - \int_{-1}^{x_1} x f_x(x)dx > 0$. Therefore $\frac{\partial V}{\partial b} > 0$ [by (47) and (48)], which proves that risk on debt or credit risk increase as debt increase. □

References

Abbasi, W. A., Wang, Z., & Alsakarneh, A. (2018). Overcoming SMEs financing and supply chain obstacles by introducing supply chain finance. *HOLISTICA–Journal of Business and Public Administration*, *9*(1), 7–22.

Agarwal, S., Chen, V. Y. S., & Zhang, W. (2016). The information value of credit rating action reports: A textual analysis. *Management Science*, *62*(8), 2218–2240.

Andreeva, G., Calabrese, R., & Osmetti, S. A. (2016). A comparative analysis of the UK and Italian small businesses using generalised extreme value models. *European Journal of Operational Research*, *249*(2), 506–516.

Asai, Y. (2019). Why do small and medium enterprises (SMEs) demand property liability insurance? *Journal of Banking & Finance*, *106*, 298–304.

Baesens, B., Setiono, R., Mues, C., & Vanthienen, J. (2003). Using neural network rule extraction and decision tables for credit-risk evaluation. *Management Science*, *49*(3), 312–329.

Bassok, Y., & Anupindi, R. (1997). Analysis of supply contracts with total minimum commitment. *IIE Transactions*, *29*(5), 373–381.

Berger, S. C., & Gleisner, F. (2009). Emergence of financial intermediaries in electronic markets: The case of online P2P lending. *BuR Business Research Journal*, *2*(1).

Błaszczyński, J., de Almeida Filho, A. T., Matuszyk, A., Szelag, M., & Słowiński, R. (2021). Auto loan fraud detection using dominance-based rough set approach versus machine learning methods. *Expert Systems with Applications*, *163*, Article 113740.

Bussmann, N., Giudici, P., Marinelli, D., & Papenbrock, J. (2021). Explainable machine learning in credit risk management. *Computational Economics*, *57*(1), 203–216.

Caldentey, R., & Chen, X. (2009). *The role of financial services in procurement contracts*. New York: New York University.

Chen, X., Cai, G., & Song, J.-S. (2019). The cash flow advantages of 3pls as supply chain orchestrators. *Manufacturing & Service Operations Management*, *21*(2), 435–451.

Chen, T., Huang, Y., Lin, C., & Sheng, Z. (2022). Finance and firm volatility: Evidence from small business lending in China. *Management Science*, *68*(3), 2226–2249.

Chen, N., Liu, X., & Yao, D. D. (2016). An optimization view of financial systemic risk modeling: Network effect and market liquidity effect. *Operations Research*, *64*(5), 1089–1108.

Conlin, M. (1999). Peer group micro-lending programs in Canada and the United States. *Journal of Development Economics*, *60*(1), 249–269.

Deo, A., & Juneja, S. (2021). Credit risk: Simple closed-form approximate maximum likelihood estimator. *Operations Research*, *69*(2), 361–379.

Durango-Cohen, E. J., & Yano, C. A. (2006). Supplier commitment and production decisions under a forecast-commitment contract. *Management Science*, *52*(1), 54–67.

Emekter, R., Tu, Y., Jirasakuldech, B., & Lu, M. (2015). Evaluating credit risk and loan performance in online peer-to-peer (P2P) lending. *Applied Economics*, *47*(1), 54–70.

Fang, G., Gao, W., & Zhao, D. (2004). Large vocabulary sign language recognition based on fuzzy decision trees. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *34*(3), 305–314.

Gao, Y., Xu, A., Hu, P. J.-H., & Cheng, T.-H. (2017). Incorporating association rule networks in feature category-weighted naive Bayes model to support weaning decision making. *Decision Support Systems*, *96*, 27–38.

Ghildiyal, B., Garg, S., & Raturi, V. (2022). Analyze of different algorithms of machine learning for loan approval. In *Smart Trends in Computing and Communications* (pp. 719–727).

Giudici, P. (2001). Bayesian Data mining, with application to benchmarking and credit scoring. *Applied Stochastic Models in Business and Industry*, *17*(1), 69–81.

Giudici, P., & Bilotta, A. (2004). Modelling operational losses: a Bayesian approach. *Quality and Reliability Engineering International*, *20*(5), 407–417.

Giudici, P., & Raffinetti, E. (2021). Shapley-Lorenz Explainable artificial intelligence. *Expert Systems with Applications*, *167*, Article 114104.

Hand, D. J., & Henley, W. E. (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *160*(3), 523–541.

Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H., & Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems*, *37*(4), 543–558.

Kouvelis, P., & Zhao, W. (2012). Financing the newsvendor: supplier vs. bank, and the structure of optimal trade credit contracts. *Operations Research*, *60*(3), 566–580.

Lai, G., Debo, L. G., & Sycara, K. (2009). Sharing inventory risk in supply chain: The implication of financial constraint. *Omega*, *37*(4), 811–825.

Le, R., Ku, H., & Jun, D. (2021). Sequence-based clustering applied to long-term credit risk assessment. *Expert Systems with Applications*, *165*, Article 113940.

Li, C.-L., & Kouvelis, P. (1999). Flexible and risk-sharing supply contracts under price uncertainty. *Management Science*, *45*(10), 1378–1398.

Liu, S., & Wu, S. (2021). Research on the identification and prediction of default risk of online lending platform customers. In *LISS 2020* (pp. 823–835).

Ma, H.-L., Wang, Z., & Chan, F. T. (2020). How important are supply chain collaborative factors in supply chain finance? A view of financial service providers in China. *International Journal of Production Economics*, *219*, 341–346.

Moscatelli, M., Parlapiano, F., Narizzano, S., & Viggiano, G. (2020). Corporate default forecasting with machine learning. *Expert Systems with Applications*, *161*, Article 113567.

Pan, Y., Chen, S., Wu, D., & Dolgui, A. (2021). CF-NN: A novel decision support model for borrower identification on the peer-to-peer lending platform. *International Journal of Productions Research*, *59*(22), 6963–6974.

Park, S., Lee, I. H., & Kim, J. E. (2020). Government support and small-and medium-sized enterprise (SME) performance: The moderating effects of diagnostic and support services. *Asian Business & Management*, *19*(2), 213–238.

Reindorp, M., Tanrisever, F., & Lange, A. (2018). Purchase order financing: Credit, commitment, and supply chain consequences. *Operations Research*, *66*(5), 1287–1303.

Roa, L., Correa-Bahnsen, A., Suarez, G., Cortés-Tejada, F., Luque, M. A., & Bravo, C. (2021). Super-app behavioral patterns in credit risk models: Financial, statistical and regulatory implications. *Expert Systems with Applications*, *169*, Article 114486.

Singh, V., Yadav, A., Awasthi, R., & Partheeban, G. N. (2021). Prediction of modernized loan approval system based on machine learning approach. In *2021 International Conference on Intelligent Technologies (CONIT)* (pp. 1–4). IEEE.

Skordilis, E., & Moghaddass, R. (2017). A condition monitoring approach for real-time monitoring of degrading systems using Kalman filter and logistic regression. *International Journal of Productions Research*, *55*(19), 5579–5596.

Somjai, S., Vasuvanich, S., Laosillapacharoen, K., & Suteerachai, B. (2019). Governing role of trade digitalization in global supply chain finance, negotiation and SMEs performance. *International Journal of Supply Chain Management*, *8*, 660–672.

Vanara, R., Wani, P., Pawar, S., More, P., & Patil, P. (2021). Predication approval for bank loan using random forest algorithm. *International Journal of Progressive Research in Science and Engineering*, *2*(7), 137–142.

Wang, G., & Ma, J. (2012). A hybrid ensemble approach for enterprise credit risk assessment based on support vector machine. *Expert Systems with Applications*, *39*(5), 5325–5331.

Wei, Y., Yildirim, P., Van den Bulte, C., & Dellarocas, C. (2016). Credit scoring with social network data. *Marketing Science*, *35*(2), 234–258.

Wu, A. (2022). Purchasing from multiple capital-constrained suppliers using order financing. *International Journal of Productions Research*, 1–16.

Xu, Y., Zhu, L., & Pinedo, M. (2020). Operational risk management: A stochastic control framework with preventive and corrective controls. *Operations Research*, *68*(6), 1804–1825.

Yıldırım, M., Okay, F. Y., & Özdemir, S. (2021). Big data analytics for default prediction using graph theory. *Expert Systems with Applications*, *176*, Article 114840.

Yoshino, N., & Taghizadeh-Hesary, F. (2019). Role of SMEs in Asia and the financing challenges they face. In *Unlocking SME Finance in Asia: Roles of Credit Rating and Credit Guarantee Schemes* (p. 1). Routledge.