Borsa Istanbul Review 24 (2024) 352–362

Contents lists available at ScienceDirect

Borsa Istanbul Review

journal homepage: www.elsevier.com/journals/borsa-istanbul-review/2214-8450

Exploring the impact of financial literacy on predicting credit default among farmers: An analysis using a hybrid machine learning model

Zhiqiang Lu [a], Hongyu Li [a], Junjie Wu [b,*]

[a] *Business School, Ningbo University, Ningbo, China*
[b] *Leeds Business School, Leeds Beckett University, Leeds, UK*

A R T I C L E I N F O

JEL classification:
G21
G32
G51
G53

Keywords:
Credit default
Credit risk prediction
Farmers
Financial literacy
Machine learning

A B S T R A C T

This study explores whether financial literacy can enhance the ability to predict credit default by farmers using machine-learning models. It introduces a hybrid model combining k-means clustering and Adaboost to predict loan default using data on 10,396 farmers who obtained credit from Chinese rural commercial banks, including demographics, household finance, credit history, and financial literacy. We systemically compare the results of models with and without financial literacy variables, which indicate significant improvement in the predictive accuracy about credit risk when financial literacy factors are included. Our findings confirm that financial literacy is a crucial indicator of farmers' ability to make informed financial decisions, reducing their likelihood of loan default and suggesting its utility as a screening tool or supplementary credit risk assessment variable. This research has profound implications for financial inclusion and credit risk management, indicating that financial institutions can leverage financial literacy data to evaluate farmers' creditworthiness and design effective financial education programs. This study enriches the literature on credit risk prediction by introducing financial literacy as a predictor of credit default.

1. Introduction

Credit default, the failure to fulfill the agreed terms for loan repayment, poses significant challenges for borrowers and banks, with negative consequences. Borrowers may experience the loss of collateral, legal repercussions, damage to creditworthiness, and lower future access to credit (Domeher & Abdulai, 2012). At the same time, banks encounter income and capital losses, higher costs for monitoring and recovery, and lower profitability and sustainability because of overdue loans (Ahlin et al., 2011).

Agriculture consists primarily of small-scale and disadvantaged farmers, who face significant challenges in accessing formal finance, making them susceptible to credit default. These farmers lack access to modern farming techniques, essential equipment, and resilient crops, hindering the adoption of more advanced practices. Additionally, factors such as high seed costs, unpredictable weather, insufficient government support, and inadequate rural infrastructure further impede their progress (Isakson, 2015; Poon & Weersink, 2011). Farmers' capacity to take risk and diversify their sources of income are hampered by their limited access to credit markets, due to their low ownership of assets and lack of collateral (Awunyo-Vitor, 2018; Raza et al., 2023).

China has about 500 million farmers, comprising 36.11 percent of its population in 2020 (Xu et al., 2023). In China, farmers have severe credit default challenges for several reasons. First, farmers often have unexpected setbacks, making prompt repayment difficult (Skees et al., 2007). Moreover, Chinese farmers lack sufficient collateral, as the state owns rural land through local collectives, hence, farmers only have the right to use it (Zhang et al., 2020). After natural disasters, the incidence of credit default by farmers increases significantly, and many of them experience a decline in their family's cash flow (Mi et al., 2022).

Second, farmers face challenges in securing finance from major state-owned commercial banks, relying instead on smaller financial institutions, such as rural commercial banks, rural cooperative banks, rural credit cooperatives, and village and township banks. These smaller banks and credit cooperatives have lower credit capacity and higher risk than their state-owned counterparts. According to the most recent *China Financial Statistics Yearbook* (2010–2018), the average nonperforming loan (NPL) ratio for Chinese financial institutions was 1.8 percent, whereas for rural commercial banks and agricultural loans it was 2.3 percent and 3.3 percent, respectively. This indicates higher credit

Peer review under responsibility of Borsa İstanbul Anonim Şirketi.

* Corresponding author. Leeds Business School, Leeds Beckett University, Rose Bowl, City Campus, Leeds, LS1 3HB, UK.
E-mail addresses: J.Wu@leedsbeckett.ac, j.wu@leedsbeckett.ac.uk (J. Wu).

https://doi.org/10.1016/j.bir.2024.01.006
Received 31 August 2023; Received in revised form 7 January 2024; Accepted 7 January 2024
Available online 18 January 2024

default risk for small and agriculture-related banks.[1] Rural credit institutions also lack effective methods for assessing potential borrowers' creditworthiness, leading to adverse selection issues and credit default (Crawford et al., 2018; Mi et al., 2022).

Third, rural China lacks tools for risk-sharing. Many Chinese farmers, because they are economically disadvantaged, lack an established credit history, and the rural credit market faces information asymmetry due to the high cost of information screening and identification (Petrick, 2005). The absence of a credit system in rural areas means that farmers generally do not value their credit, leading to higher credit defaults. Mi et al. (2022) analyze data from the China Household Finance Survey (CHFS) in 2015 and 2017, which reveals that farmer lenders have a higher probability of credit default than their urban counterparts. For these reasons, rural financial institutions hesitate to extend credit to farmers in gernal to minimize credit default, hindering their access to essential inputs and significantly curtailing their investment in agricultural technology.

Financial literacy comprises the knowledge and skills that are essential for informed money management, encompassing responsible debt handling (Kumar et al., 2023; Singh, 2014). Prior studies indicate that people who possess higher financial literacy are more likely to adopt good financial practices, such as maintaining records, making deposits at banks, and accessing affordable credit, which reduces their incidence of credit default (Kidwell & Turrisi, 2004; Nkundabanyanga et al., 2014). Financial literacy is considered a critical factor for farmers in overcoming funding barriers, with the potential to significantly enhance their access to credit and savings despite having low income (Klapper et al., 2013; Raza et al., 2023). Research in China indicates a positive correlation between financial literacy and household financial status and behavior (Xu et al., 2023). Xu et al. (2022), using data from the CHFS in 2015 and 2017, show that financial literacy has a positive impact on having a formal bank account, commercial insurance, a pension plans, and a credit card as well as participating in the financial market.

Despite its potential benefits of financial literacy in helping agricultural business, financial literacy among Chinese farmers remains low. A survey of 2278 farmer households in Jiangsu Province in 2021 reveals that only 11.2 percent of them initiated businesses. Moreover, despite financial literacy's effectiveness in alleviating traditional credit constraints and promoting entrepreneurship, Liu et al. (2023) stress the unequal distribution of credit resources, as China's credit policy favors cities, imposing stricter conditions on rural farmers who are seeking financial support. In this context, higher financial literacy enables people to better understand the borrowing process, interest rates, inflation, and the time value of money.

Although researchers have explored the connection between financial literacy and financial behavior by farmers, they have not explained how financial literacy contributes to prediction of credit default by farmers. In this paper, we fill this research gap in order to enable financial institutions and policy makers to evaluate and address risks associated with agricultural lending. If financial literacy is a significant factor in predicting credit default, banks and lending institutions can adjust their criteria and allocate resources efficiently to farmer-targeted financial education programs. Governments can also use our findings to design initiatives for promoting financial literacy among farmers, thereby fostering financial stability in the agricultural sector and contributing to overall economic development and well-being.

In China, where agriculture plays a central role, and many farmers lack access to formal financial education, understanding the impact of financial literacy on credit default is crucial. The government's ambitious shift from traditional to modern agriculture involves changes in farming practices, technology adoption, and market dynamics. Evaluating the influence of financial literacy on credit default can assist financial institutions in tailoring financial services to meet evolving farmer needs. China's financial inclusion initiatives also aim to integrate more individuals into the formal financial system. Recognizing the role of financial literacy in the prediction of credit default is essential for refining these initiatives to effectively address challenges unique to rural and agricultural communities. Moreover, financial institutions encounter significant risks in agricultural lending, and the knowledge that financial literacy is a valuable predictor of credit default can guide Chinese financial institutions in refining risk assessment models.

This study investigates the predictive role of financial literacy in credit default by farmers using innovative machine-learning techniques. Although machine-learning applications have expanded beyond the traditional accuracy assessments of classifier algorithms (Lessmann et al., 2015; Rishehchi et al., 2021), which explore "alternative" factors,[2] machine-learning has not been used to predict the probability of credit default. We employ a hybrid model integrating k-means clustering with the Adaboost algorithm to assess the impact of financial literacy on predicting credit default by farmers. Our performance comparison involves two sets of borrower characteristics: the first is exclusively traditional, excluding financial literacy, and the second combines traditional characteristics with variables for financial literacy. These variables come from a preloan survey by a rural commercial bank in China, covering attentiveness to financial information, familiarity with financial products, investment experience, risk preferences, and years of investment. Our dataset comprises 10,396 farmers who received loans from this bank, in order to demonstrate that integrating financial literacy factors enhances credit risk models, particularly in agricultural lending.

This paper contributes to the literature in two ways. First, it shows that incorporating financial literacy variables into a credit risk prediction model makes it more accurate than models without them. Financial literacy variables capture critical aspects of farmers' financial behavior and preferences that are not encompassed in traditional characteristics. For instance, our findings show that farmers who are more interested in financial news or more familiar with stocks have a lower likelihood of default. Second, we demonstrate that financial literacy alone has a remarkable capacity for predicting credit risk, emphasizing its substantial power and importance. This suggests that financial literacy can serve as a powerful screening tool or supplementary variable in credit risk assessment, emphasizing its valuable role in enhancing credit risk management.

The paper is organized as follows. Section 2 constructs a theoretical framework for clarifying the relationship between financial literacy and credit default risk. Section 3 outlines our research methodology. Section 4 presents the research findings. Section 5 concludes the paper, emphasizing our key contributions, implications, and limitations.

2. Theoretical framework

In this section, we critically review relevant literature to establish a conceptual framework that connects financial literacy, its dimensions, associated mechanisms, and credit default risk, which helps to clarify the intricate relationships among these variables, and this framework is the analytical foundation of our paper.

Financial literacy encompasses farmers' ability to understand and use various financial skills, including managing money, budgeting, and

[1] For example, the People's Bank of China's (PBoC) risk rating for the fourth quarter of 2022 indicates risk levels of 1–5 for large state-owned banks and 2–10 for agriculture-related smaller banks; http://camlmac.pbc.gov.cn/jinrongwendingju/146766/146772/4979810/index.html.

[2] Such as textual characteristics (Gao et al., 2023; Netzer et al., 2019), psychometrics (Dlugosch et al., 2018; Fine, 2023; Liberati & Camillo, 2018; Owusu et al., 2023), social networks (De Cnudde et al., 2015; Ge et al., 2017; Lu et al., 2023), mobile phone usage (Björkegren & Grissen, 2020; Pedro et al., 2015), and web browsing behavior (Lu et al., 2023; Rozo et al., 2023).

making informed financial decisions. Financial literacy is essential for active participation in financial markets, efficient financial resource management, long-term goal setting, and resilience in the face of economic uncertainty (OECD/INFE, 2011). Financial literacy not only can improve farmers' wealth but also reduce their risk of default. Therefore, the theoretical discourse is a foundational step in understanding the significance of farmers and their financial behavior. We discuss financial literacy in terms of the following four dimensions.

2.1. Attention to financial information

A farmer's attention to financial information involves staying informed about market trends, commodity prices, and economic indicators, which is crucial for making informed decisions about agricultural activities and financial management, ultimately reducing the risk of credit default (Chang et al., 2022). Disney and Gathergood (2013) and others support the idea that higher financial literacy makes people more willing to gather financial information, thereby enhancing their ability to find appropriate credit products. In essence, financial literacy enables farmers to evaluate their financial needs more accurately, preventing excessive debt and reducing the chance of a debt crisis (Gaurav & Singh, 2012; Kurowski, 2021).

This dimension of financial literacy plays a pivotal role in predicting credit default by reflecting a farmer's capacity to respond to changing economic conditions. Farmers who pay high attention to financial information are likely to make more strategic financial decisions, thereby reducing the likelihood of defaulting on credit obligations. Conversely, those who pay less attention may face challenges in adapting to market fluctuations, potentially increasing their credit default risk.

2.2. Understanding financial products

The second dimension of financial literacy centers on farmers' understanding of the diverse financial products and services available to them. Thorough comprehension of financial products is paramount for making judicious investment choices, managing debt responsibly, and optimizing financial resources in agriculture (Hastings & Tejeda-Ashton, 2008). Prior research indicates that financial literacy can enhance farmers' knowledge of financial products (Xia et al., 2014) and assist them in selecting the most suitable options (Sarfo et al., 2023). By comparing interest rates and terms across banks, financially literate farmers can reinforce their understanding of credit risks and minimize the default risk from borrowing (Disney & Gathergood, 2013; Nkundabanyanga et al., 2014).

Farmers who understand financial products can skillfully navigate diverse credit options, insurance plans, and investment opportunities. This comprehension empowers them to align their financial decisions with the specific needs and risks associated with agricultural activities, thereby reducing their likelihood of default (Cao-Alvira et al., 2021; Mandell, 2006; Sivaramakrishnan et al., 2017). However, farmers who lack a comprehensive understanding of financial products may struggle to select the most suitable financial instruments, exposing themselves to unnecessary risks. The nexus between this dimension of financial literacy and prediction of credit default is the farmer's capacity to make well-informed choices about the financial tools used in their farming operations (Lusardi, 2019; Wang et al., 2021; Yuesti et al., 2020).

2.3. Investment experience and duration

Financial literacy transcends theoretical knowledge to extend to practical application, and a crucial facet of it is farmers' investment experience and the duration of their financial involvement. Krische (2019) indicates that people with higher financial literacy have more investment experience and longer investment horizons. Farmers with substantial investment experience demonstrate a practical understanding on the risks and returns associated with various agricultural ventures, contributing to a more accurate prediction of credit default (Van Rooij et al., 2011).

Other studies provide evidence that financial literacy enables households to attain better comprehension of financial terms, investment returns, and the risks associated with financial products (Mandell, 2006; Sivaramakrishnan et al., 2017). It improves their information screening and financial calculations (Albaity & Rahman, 2019; Hastings & Tejeda-Ashton, 2008), mitigates risk aversion among investors, increases willingness to participate in financial investment (Dohmen et al., 2010), encourages the purchase of risky financial assets (Li et al., 2020), and diversifies investment portfolios (Abreu & Mendes, 2010; Krische, 2019). These findings support the notion that enhanced financial literacy increases participation in financial investment and raises the effectiveness of financial investment decisions. However, deficiencies in financial literacy can lead to flawed financial decision-making (Calvet et al., 2007). In the Chinese context, Li et al. (2020) indicate that households with varying levels of financial literacy and characteristics can make better financial decisions, leading to better investment outcomes. As in the results observed in other countries, financially knowledgeable investors in China tend to invest in the stock market (Guo & Liang, 2014; Li, 2006).

The duration of investment activities is equally significant, reflecting a farmer's resilience and adaptability to market dynamics over time. Long-term investment experience suggests a capacity to withstand economic fluctuations, whereas short-term experience may indicate vulnerability to immediate challenges. Farmers with extensive investment experience and sustained financial activity are more likely to exhibit financial stability and have a lower likelihood of credit default. However, those with limited exposure may face higher default risk due to a lack of practical knowledge in navigating the complexities of agricultural finance.

2.4. Risk preference

Risk preference means an individual's readiness to take on financial risks in order to obtain potential rewards (Demirguc-Kunt et al., 2017). Recognizing farmers' risk preferences is vital for predicting credit default, offering insight into their tolerance for financial uncertainty (Van Rooij et al., 2011). Farmers with a higher preference for risk might tend toward innovative agricultural practices or higher-return, higher-risk ventures. But those with a more conservative risk preference may opt for stable, low-risk investment, reducing the likelihood of default but also reducing the potential for greater profit (Krische, 2019; Li et al., 2020).

Financial literacy can enhance farmers' risk management, reducing the chance of loan default. Although this may boost profitability, it also heightens the risk of default, especially in challenging economic times (Abreu & Mendes, 2010; Calvet et al., 2007; Dohmen, 2010). Financially literate farmers are more likely to use financial markets and instruments to mitigate risk, such as opening and managing bank accounts (Grohmann et al., 2018), saving in case of unexpected events (Jappelli & Padula, 2013), joining pension plans (Xu et al., 2023), investing in the stock market (Van Rooij et al., 2011), and purchasing insurance (Pitthan & De Witte, 2021). Consequently, more financially literate farmers can adapt more effectively to uncertain shocks, decreasing their risk of loan default.

Furthermore, financial literacy can deepen farmers' understanding of loan credit, enabling them to be more vigilant about credit records, ratings, and the repercussions of loan default (Santos & Gallucci, 2020). Meoli et al. (2022) reveal that financial literacy shapes farmers' credit attitudes and risk tolerance, reducing moral hazard and adverse selection. Thus more financially literate farmers may actively adhere to loan terms, assess project risks, and maintain good credit, and are less inclined to default.

Combining risk preference with a financial literacy analysis enhances the predictive power of credit default models by considering the farmer's individual attitudes toward financial risk. This dimension adds a nuanced layer to an evaluation of financial literacy, providing a more comprehensive understanding of factors that influence creditworthiness in the agricultural sector.

Analyzing financial literacy in terms of these four dimensions offers a comprehensive framework for predicting farmers' credit default. To explain how these dimensions mitigate credit default risk, we explore three mechanisms: identifying the need for financial products, improving the ability to manage risks, and understanding loan credit. This reveals the complex ties between financial literacy and farmers' decision-making, which affects their creditworthiness (Disney & Gathergood, 2013).

The focus on financial information and grasp of financial products is closely linked to identifying the need for financial products—a pivotal strategy for reducing credit default by farmers. Farmers who actively monitor market trends and comprehend financial products can effectively pinpoint their specific financial requirements. For instance, farmers attuned to financial information can recognize the need to diversify financial tools for risk mitigation, and those with a deep understanding of financial products can select instruments tailored for their needs. Staying informed about various financial instruments allows farmers to align products with their agricultural operations (Xu et al., 2022, 2023).

Investment experience and risk preference significantly contribute to improving risk management abilities—a crucial factor in reducing credit default in agriculture. Farmers with extensive investment experience have likely navigated various risk scenarios, enhancing their practical understanding of risk management strategies (Li et al., 2020). Additionally, farmers' risk preferences directly influence their approach to risk management. Those with a higher preference for risk adopt risk mitigation strategies, such as investment in financial derivatives or insurance products, whereas those with a more conservative risk preference can prioritize low-risk investment, inherently reducing exposure to financial setbacks (Guo & Liang, 2014).

Attention to financial information, understanding of financial products, and risk preference collectively contribute to improving understanding of loan credit by farmers. Farmers who actively pay attention to financial information are more likely to stay informed about loan credit terms, conditions, and nuances, reducing the likelihood of misunderstanding that leads to default.

Moreover, a deep understanding of financial products, especially loan credit, empowers farmers to select credit options that are aligned with their financial needs and risk tolerance. These farmers can evaluate interest rates, repayment terms, and associated costs, supporting informed decision-making and successful loan management (Xia et al., 2014). Additionally, farmers with a conservative risk preference may choose loans with lower interest rates and stable repayment structures, minimizing default risk. However, those with a higher risk preference may strategically leverage credit for potentially higher returns while implementing robust risk management practices (Gaurav & Singh, 2012; Kurowski, 2021; Sarfo et al., 2023).

Fig. 1 illustrates the structure of our theoretical framework.

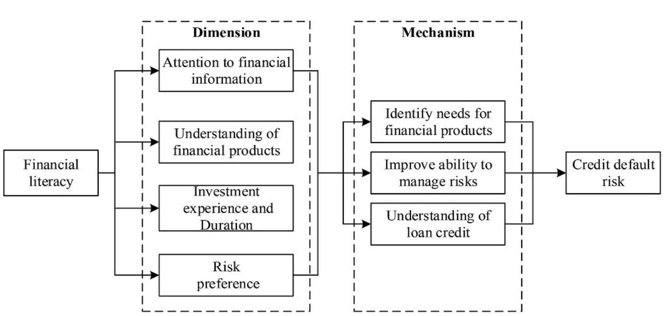

Fig. 1. The theoretical framework.

3. Methodology

3.1. The analytical process

The analytical process of this paper is illustrated in Fig. 2. It begins with gathering information that can influence farmers' credit risk, encompassing data collection and preprocessing. This dataset incorporates both indicators of conventional credit risk prediction and financial literacy metrics with data from a rural commercial bank in China. Subsequently, we engaged in data preprocessing, consistent with best practices, to refine the dataset for the creation of more robust and precise models, as advocated by Maharana et al. (2022).

Next, because hybrid machine-learning models can mitigate potential weaknesses found in individual models (Machado & Karray, 2022), we employ these hybrid techniques to formulate prediction models for farmers' credit default. Specifically, we begin the process by implementing the k-means clustering algorithm to group the data effectively. Following this clustering step, we embarked on the application of various supervised learning techniques. These encompass logistic regression (LR), decision tree (DT), support vector machine (SVM), random forest (RF), bagging, gradient boosting tree (GBT), and Adaboost, each applied to the clusters obtained from the previous step to predict credit default. Through a systematic evaluation process, we identify the best-performing model by comparing the performance of these diverse algorithms. This rigorous comparison allows us to select the model that demonstrates the highest predictive accuracy and relevance for the prediction of credit default by farmers.

In the third step, using the best-performing model, we examine the value added with financial literacy variables. This analysis involves a direct comparison of model performance between those with and without financial literacy variables. To rigorously evaluate the significance of this difference, we conducted a DeLong test. This statistical test is a critical tool for determining whether a statistically significant difference in the prediction of farmers' credit defaults exists between the two sets of models—those with financial literacy variables and those without them. This process allows us to ascertain whether including financial literacy variables contributes meaningfully to the accuracy and

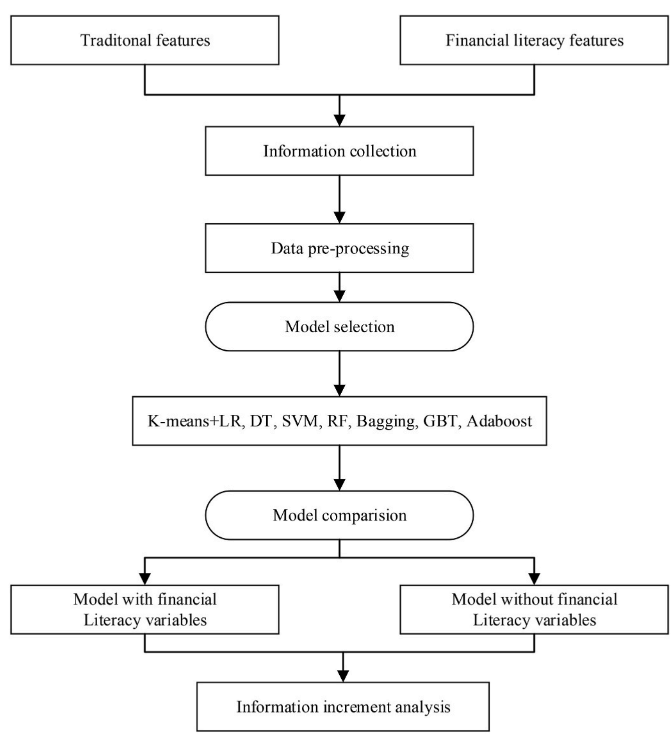

Fig. 2. The analytical process.

effectiveness of predictions of credit default by farmers.

3.2. Machine-learning methods

3.2.1. K-means

K-means is a straightforward and highly effective unsupervised learning technique employed in the realm of clustering (Yu et al., 2018). Its core concept relates to the identification of k centroids, which are representative markers. These centroids are strategically positioned to minimize the collective sum of the squared distance that separates every data point from its closest centroid (Singh et al., 2011). The process of refining these centroids unfolds iteratively by assigning each data point to its nearest centroid and then reconfiguring the centroids as the mean of the data points assigned to them.

The primary parameter governing the k-means algorithm is the number of clusters, denoted as k. In this study, we set k as equal to 7.[3] We then use traditional credit risk variables to divide borrowers into seven distinct clusters.

3.2.2. LR

Logistic regression (LR) is a prominent ML technique extensively used for binary classification tasks (Schein & Ungar, 2007), such as credit risk prediction. It hinges on the logistic function, which transforms the predicted output into a probability value, from 0 to 1. The logistic function can be mathematically represented as:

$$y = \frac{e^{(b0+b1x)}}{\left(1 + e^{(b0+b1x)}\right)}$$

where y is the predicted output, $b0$ is the intercept term, $b1$ is the coefficient for the input variable x, and e is the base of the natural logarithm. These coefficients are determined through maximum likelihood (ML) estimation, which is employed during the training phase to optimize the model parameters. The primary objective of ML estimation is to maximize the likelihood of observing the data given these model parameters, effectively fine-tuning the model to fit the data as closely as possible.

The LR model by Lee and Jun (2018) has the virtues of simplicity, interpretability, and computational efficiency. It lends itself effectively to the analysis of a borrower's likelihood of default, enabling the classification of the borrower based on whether the estimated probability exceeds a predefined threshold.

3.2.3. DT

Decision trees (DT) offer another machine-learning technique for tackling binary classification problems (Charbuty & Abdulazeez, 2021). They are rooted in a tree-like structure, in which each branch symbolizes an outcome or class label. This tree is constructed through recursive data partitioning, guided by the input variables' values that best distinguish the classes. The partitioning process ends when a node becomes pure (containing only one class) or after it reaches a predefined stopping criterion, such as maximum tree depth or minimum node sample size. Decision trees are known for their intuitive, transparent nature and ease of comprehension. They exhibit versatility in handling numerical and categorical variables as well as the capacity to capture nonlinear relationships and interactions (Gulati et al., 2016).

3.2.4. SVM

The support vector machine (SVM; George & Vidyapeetham, 2012) is another supervised learning approach used for binary classification tasks. It works by identifying a hyperplane that effectively divides the data into two classes while maximizing the margin between them. SVM has versatility, as it can address both linear and nonlinear problems by using diverse kernel functions, which facilitate the transformation of the input space into a higher-dimensional feature space (Pietrzak, 2022; Zhou et al., 2010).

3.2.5. RF

The random forest (RF) is an additive model that generates predictions by amalgamating the outcomes of DT models (Luo, 2020). Its foundation rests on amalgamating several DTs, each trained on a random subset of the data and input variables. The ultimate prediction is derived by averaging or voting on the forecasts of these individual trees. RF can handle numerical and categorical variables, effectively capturing nonlinear relationships and interactions (Boulesteix et al., 2012).

3.2.6. Bagging

Bagging, an abbreviation for "bootstrap aggregating," is a machine-learning technique suitable for binary classification tasks, as described by Hsiao et al. (2020). It revolves around the construction of multiple models, each trained on a bootstrap sample of the data, and then amalgamating their predictions through averaging or voting. A bootstrap sample is randomly drawn from the data with replacement, implying that some data points may appear multiple times or not within the sample.

The utility of bagging is its ability to diminish variance and enhance the accuracy of a single model, mainly when that model has instability or a tendency to overfit, as discussed by Kim and Kang (2010). Moreover, bagging has the versatility to handle both numerical and categorical variables while also being proficient at capturing nonlinear relationships and interactions (Drechsler & Reiter, 2011).

3.2.7. Gradient boosting tree

Gradient boosting trees (GBT), introduced by Krauss et al. (2017), are another valuable machine-learning approach that applies to binary classification problems. GBT is grounded in constructing an ensemble of weak learners, typically, decision trees, and then enhancing their performance through an iterative process of introducing new trees that correct the errors of their predecessors. These new trees are tailored to fit the negative gradient of the loss function, which quantifies the disparity between the actual and predicted outcomes. GBT can handle both numerical and categorical variables while demonstrating proficiency in capturing nonlinear relationships and interactions (Ding et al., 2018).

3.2.8. Adaboost

Adaboost, short for "adaptive boosting," is a supervised learning technique suitable for binary classification problems (Walker & Jiang, 2019). It revolves around the concept of creating an ensemble of weak learners, typically decision stumps (one-level decision trees) and subsequently enhancing their performance through an iterative process that introduces new learners, with a focus on correcting the misclassifications by their predecessors.

These new learners are assigned weights reflecting their accuracy, and the final prediction is derived through a weighted majority vote of the ensemble. Adaboost is valuable for improving the accuracy and reducing the variance of a single weak learner, particularly when that learner is elementary or prone to underfitting (Tao et al., 2019). Furthermore, Adaboost can handle both numerical and categorical variables while being proficient at capturing nonlinear relationships and interactions (Oyedele et al., 2021).

3.3. Performance metrics

To assess the effectiveness of the credit risk prediction models, we use a range of performance metrics: accuracy, area under the curve (AUC), precision, recall, and f1 score, which are all positive indicators.

[3] The bank that supplied the sample has not yet introduced a classification system for its loan customers. However, it is developing a plan to divide its loan customers into seven different tiers (e.g., AAA, AA, A, BBB, BB, B, and C).

In this context, higher values indicate superior model performance. Furthermore, this evaluation includes mean squared error (MSE) and mean squared logarithmic error (MSLE) as performance metrics, but they are reverse indicators, in which lower values indicate better model performance. By calculating and comparing the values of these performance metrics for each model, we can determine the most effective credit default prediction model among the alternatives, providing a comprehensive assessment of their predictive capabilities.

4. Data

4.1. Dataset

The dataset used in this paper comprises loans to 10,396 farmers over the period 2017 to 2020. This dataset pertains to rural household microfinance and comes from a rural commercial bank in China. The original dataset encompasses a spectrum of traditional credit risk prediction variables, encompassing factors such as farmers' characteristics (age, gender, marital status, and education level), household financial details (family size by membership, annual income, and debt-to-income ratio), loan amounts, and credit histories. These variables form the foundation of the dataset, which serves as the basis for the analysis conducted in this study.

In the loan application process, borrowers are required to complete a preloan questionnaire, which includes questions about their financial literacy divided into five dimensions:

Attention to financial information (*Fin_Information*): Borrowers are asked to describe their level of attention to financial information, with the following choices of response: (1) Not paying attention at all, (2) Rarely paying attention, (3) Paying occasional attention, (4) Paying average attention, and (5) Paying great attention; understanding of financial products (*Fin_Products*): Borrowers are prompted to assess their understanding of financial products, including stocks, funds, foreign exchange, and financial derivatives with the following choices of response: (1) No understanding of financial products, (2) Essentially no understanding of financial products, (3) Basic understanding of various financial products and their risks, (4) Substantial understanding of various financial products and their risks, and (5) Profound understanding of various financial products and their risks; investment experience (*Fin_Experence*): Borrowers are asked about their investment experience, with the following choices of response: (1) No investment experience, (2) Limited investment experience aside from savings, (3) Participation in financial products such as bank finance and insurance, (4) engagement in trading stocks, funds, and similar products, and (5) Engagement in trading warrants, futures, options, and other advanced financial instruments; risk preference (*Fin_Risk*): Borrowers are presented with a hypothetical investment scenario involving a principal of RMB 1 million and are asked to select their preferred investment opportunity among options with varying risk-reward profiles; investment duration (*Fin_Years*): Borrowers are asked to indicate how long they have been engaged in investment related to stocks, funds, foreign exchange, financial derivatives, and other risk-bearing financial products, with the following choices of response: (1) No experience, (2) Less than two years, (3) 2–5 years, (4) 5–10 years, and (5) More than 10 years.

Their responses to the questionnaire regarding financial literacy serve as essential variables in our analysis, enabling us to assess the impact of financial literacy on credit default prediction.

We thus construct three different credit default prediction models, each with different combinations of variables[4] to evaluate the impact of financial literacy variables on predictive performance. Table 1 summarizes the distinctions between these models.

- *Normal model*: This model incorporates only traditional predictive variables and does not include any financial literacy variables. It serves as the baseline model for selecting machine learning algorithms.
- *Normal-fin model*: In this model, we include both traditional predictive variables and financial literacy variables. The primary distinction of this model is the addition of financial literacy variables, allowing us to assess their contribution to predictive performance.
- *Fin model:* This model comprises only financial literacy variables, enabling us to evaluate the performance of financial literacy indicators independently.

In short, the Normal model is used as the basis for machine-learning model selection. To assess the influence of financial literacy variables, we compare the performance of the Normal model with that of the Normal-fin model, with the sole difference that the latter includes financial literacy variables. Furthermore, we analyze the performance of the Fin model, which only uses financial literacy variables. These comparisons enable us to clarify the impact and effectiveness of financial literacy indicators in credit default prediction.

4.2. Data preprocessing

In the data preprocessing stage, we take several essential steps to ensure data quality and suitability for analysis: normalizing the target variable, managing categorical features, addressing missing values, and dividing the data into two sets. A detailed breakdown of these steps is as follows.

4.2.1. Normalization of quantitative indicators

Based on the approach by Rozo et al. (2023), each quantitative indicator is normalized to constrain its value from 0 to 1. This normalization ensures that all the variables are on a consistent scale, facilitating practical model training and comparison.

4.2.2. Categorical features

Categorical features are managed through one-hot encoding, which translates categorical variables into a numerical format using dummy variables or numerical representations. This encoding strategy is crucial for incorporating categorical data into the final dataset, ensuring that it can be effectively used in the analysis.

4.2.3. Missing values

Variables with missing data accounting for less than 5 percent were imputed with their mean value, preserving their presence in the final dataset. Variables with missing values exceeding this threshold were omitted from the dataset.

4.2.4. Data splitting

The data are divided into training and testing sets, at a ratio of 7:3. This process ensures that the distribution of farmers' default rates in both sets remained similar to those in the original dataset. The training set is employed for model training and parameter selection, whereas the testing set is reserved for evaluating model performance.

Table 1
Analytical models.

Model	Traditional variables	Financial literacy variables
Normal model	Yes	No
Normal-Fin model	Yes	Yes
Fin model	No	Yes

[4] Traditional predictive variables: age, gender, marital status, education level, family size, total asset, annual income, debt-to-income ratio, loan amounts, and credit histories.Financial literacy variables: Attention to financial information, understanding of financial products, investment experience, risk preference, and investment duration.

Table 2
Performance comparison of different models.

Models	Accuracy	AUC	Precision	recall	f1-score	MSE	MSLE
k-Means + logistics	0.603	0.630	0.772	0.603	0.655	0.396	0.190
k-Means + DT	0.618	0.587	0.766	0.618	0.667	0.381	0.183
k-Means + SVM	0.610	0.632	0.775	0.610	0.661	0.389	0.187
k-Means + RF	0.821	0.579	0.753	0.821	0.774	0.178	0.085
k-Means + Bagging	0.810	0.562	0.749	0.810	**0.771**	0.189	0.090
k-Means + GBT	0.787	0.594	0.757	0.787	0.770	0.212	0.101
k-Means + Adaboost	**0.840**	**0.656**	**0.775**	**0.840**	0.769	**0.159**	**0.076**

Table 3
Comparison of Normal model and Normal-fin model.

Models	Accuracy	AUC	Precision	recall	f1-score	MSE	MSLE
Training set Normal model	0.82542	0.68187	0.78975	0.82542	0.74922	0.17457	0.08387
Normal-Fin model	0.85513	0.80836	0.83940	0.85514	0.82862	0.14486	0.06959
Testing set Normal model	0.84018	0.65659	0.77531	0.84018	0.76933	0.15982	0.07679
Normal-Fin model	0.86617	0.7928	0.84854	0.86617	0.84368	0.13382	0.06429

5. Results

5.1. Predictive performance and model selection

Table 2 summarizes the predictive results of the seven classifiers, giving a comprehensive overview of the performance metrics for each hybrid model. The key findings and model selection criteria are as follows.

- *Ensemble Models Outperform Single Learning Classifiers*: Table 2 demonstrates that the hybrid models, which combine k-means clustering with ensemble classifiers (i.e., RF, bagging, GBT, and Adaboost), have better predictive performance than single learning classifiers (e.g., LR, DA, and DT). This observation is consistent with previous research findings that highlight the advantages of ensemble models in credit default analysis (Chopra & Bhilare, 2018; de Castro Vieira et al., 2019).
- *k-Means + Adaboost Model is the Top Performer*: Among the hybrid models, the k-means + Adaboost model is the most effective, with the highest values for key performance metrics. It achieves the highest accuracy (0.840), AUC (0.656), precision (0.775), and recall (0.840) and the lowest MSE (0.159), MAE (0.159), and MSLE (0.076).

Because of these findings and a thorough comparison of the performance metrics, showing exceptional predictive capabilities, we use the k-means + Adaboost model for subsequent analysis.

5.2. Validation of financial literacy: Normal model vs. Normal-Fin model

In this section, we use the k-means + Adaboost model to assess the impact of financial literacy variables by comparing the performance of the Normal model and the Normal-fin model in predicting credit default by farmers. Table 3 presents the results of this comparative analysis. The results demonstrate that, for the training and testing sets, the Normal-fin model outperforms the Normal model across various validation metrics. For example, for the training set, the Normal-fin model achieves accuracy of 85.29 percent, exceeding the rate of the Normal model, 82.54 percent. Similarly, for the testing set, the Normal-fin model attains accuracy of 86.39 percent, exceeding the rate of the Normal model, 84.01 percent. These findings highlight that the inclusion of financial literacy variables significantly enhances the predictive performance of credit default models in both the training and testing sets when the Normal-fin model is applied.

Next, we compare these same two credit default prediction models, measuring their predictive performance evaluation based on the area under the curve (AUC), with a DeLong test to ascertain the statistical significance of the difference between the results from these models. Table 4 lists the AUC values for both models for the training and testing sets, along with the corresponding *p*-values from the DeLong test. The results in Table 4 demonstrate that the inclusion of financial literacy variables leads to significant improvement ($p < 0.001$) in AUC for both the training and testing sets, confirming that financial literacy adds incremental value to credit default prediction.

Fig. 3 illustrates the receiver operating characteristic (ROC) curves for the Normal model and the Normal-fin model for both the training and testing sets. These curves reiterate the superiority of the Normal-fin model, reinforcing that incorporating financial literacy variables enhances credit default prediction performance.

Table 4
The AUC of Normal model and Normal-Fin model for predicting loan default.

Models	Testing set			Training set		
	AUC	Delong test	*p*	AUC	Delong test	*p*
Normal model	0.657	7.491	<0.001	0.682	10.114	<0.001
Normal-Fin model	0.793			0.808		

5.3. Validation of financial literacy: the Fin model

Previous research has consistently highlighted the correlation between lower financial literacy and suboptimal investment decision-making, often resulting in less stable investment returns (Jappelli & Padula, 2013; Skimmyhorn, 2016). These less-effective financial decisions, in turn, are associated with a high risk of credit default (Baidoo et al., 2020; Chen et al., 2018). Table 5 gives a detailed comparison of financial literacy levels among various groups, revealing statistically significant distinctions between defaulters and nondefaulters across various financial literacy metrics. These findings further emphasize the importance of financial literacy in shaping financial behaviors and outcomes, particularly in the context of credit default risk.

Next, we explore the predictive potential of a model with only financial literacy variables, but no traditional credit-related variables. This approach is particularly important because, in many countries, a substantial share of the population is age 18 and above, especially young agricultural entrepreneurs, who lack extensive credit records or have

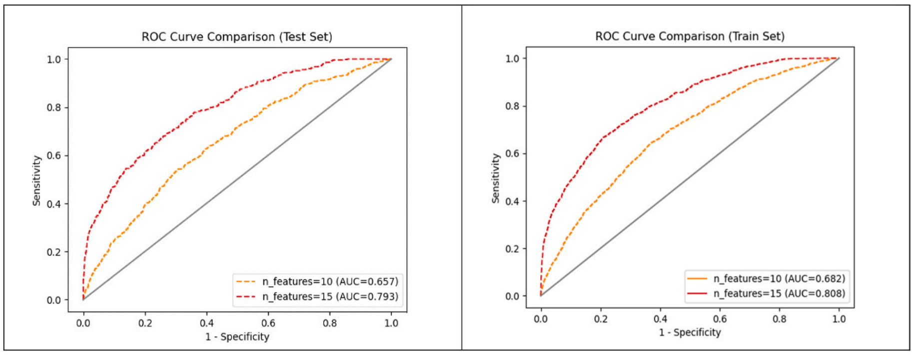

Fig. 3. ROC curves for the Normal model and Normal-Fin model.

Table 5
Independent samples' *t* test of variables for financial literacy.

Variables		value	*t*-value	*p*-value
Fin_Information	Default	2.676	5.158	<0.001
	No default	2.804		
Fin_Products	Default	2.491	8.213	<0.001
	No default	2.730		
Fin_Experence	Default	2.385	5.098	<0.001
	No default	2.246		
Fin_Risk	Default	1.933	0.049	0.960
	No default	1.932		
Fin_Years	Default	3.315	12.807	<0.001
	No default	2.943		

limited income and assets—which often impedes their ability to access credit. By evaluating the performance of a model that includes only financial literacy variables, we address the need to assess the credit of individuals with limited credit histories or modest financial resources. In this way, this investigation can help develop more inclusive and equitable credit evaluation systems that cater to a broader range of borrowers.

The models are trained exclusively using financial literacy variables, and a fivefold cross-validation approach is used to ensure the reliability and robustness of the results. Table 6 gives the performance metrics for the Fin model, showing that it has accuracy of 0.83039, an AUC of 0.68078, precision of 0.78236, recall of 0.83039, and f1 of 0.79115. These results indicate that the Fin model's predictive accuracy is comparable to that of models with both traditional credit predictors and financial literacy variables. Hence, this model has the potential to serve as a valuable tool for commercial banks in predicting loan defaults, mainly by borrowers who have limited credit histories or resources.

These results have significant implications for financial inclusion. First, the ability of this model to assess farmers' credit default risk based on their financial literacy, regardless of their family or income conditions and whether they have a prior credit record, would help them to obtain loans for business startup and growth. Second, commercial banks could potentially leverage this model to expand their customer base. Moreover, by relying solely on financial literacy and limited individual or household variables, banks can offer microcredit services to a broader range of customers to achieve greater financial inclusion.

6. Conclusion

This paper investigates the potential enhancement of credit default prediction models with a dataset on 10,396 Chinese farmers who received loans from a rural commercial bank. The study takes an innovative approach by introducing a hybrid machine-learning model that merges k-means clustering with an Adaboost classifier. Its primary focus is assessment of the predictive capabilities of financial literacy variables in contrast to conventional credit-related predictors. Through this thorough analysis, we show the accuracy and efficacy of incorporating financial literacy information to improve credit default prediction models.

Our findings robustly support the assertion that financial literacy is pivotal for forecasting loan defaults. This is highlighted by the substantial improvement in model performance after the inclusion of financial literacy variables. In the context of our research, farmers with higher financial literacy tend to exhibit prudent financial behavior, such as proactive financial planning, prudent budgeting practices, diversified income sources, debt avoidance, and proficiency in navigating financial challenges. Consequently, their behavior can significantly mitigate the likelihood of loan default, ultimately strengthening their creditworthiness. In essence, our results confirm the indispensable role of financial literacy as a powerful tool in mitigating the risk of loan default. Additional analysis also demonstrates that financial literacy variables, when employed in isolation, have a level of predictive precision. These financial variables offer an alternative path for assessing credit risk, supplementing models that depend on individual, household, and transactional factors. So, they have the potential to help banks and financial institutions in reducing obstacles to credit access for people who lack an extensive credit history or substantial assets, thus expanding financial inclusion.

Our results have profound implications. First, by including these financial literacy factors in their assessment of credit risk, banks and financial institutions can make more informed decisions regarding credit limits, interest rates, and repayment terms, as well as evaluate the creditworthiness of current and potential clients. This can be achieved by placing greater importance on financial literacy information and reducing reliance on other demographic characteristics such as income, assets, and past credit history. This approach effectively addresses the challenges of information asymmetry and collateral limitations that often hinder borrowers from accessing formal credit markets. Second, by using financial literacy data, banks and financial institutions can expand microcredit services to a broader range of clients, especially young agricultural entrepreneurs who lack business income and credit records. This initiative promotes financial inclusion and economic empowerment. Third, banks and financial institutions can develop more tailored

Table 6
The predictive performance of the Fin model.

	N	Accuracy	AUC	Precision	Recall	f1 score	MSE	MSLE
Training set	7277	0.82272	0.70523	0.77607	0.82273	0.77870	0.17727	0.08517
Testing set	3119	0.83039	0.68078	0.78236	0.83039	0.79115	0.16960	0.081487

and effective financial education programs aimed at enhancing the financial capabilities and behaviors of potential borrowers. Fourth, educational authorities should incorporate financial literacy education into the national curriculum. Finally, governments should invest in financial infrastructure and regulations to promote financial literacy education and oversee banks and financial institutions as they integrate financial literacy assessments into their credit risk evaluations. This will significantly advance the goal of promoting financial literacy as a national policy, leading to better financial decision-making and lower default rates overall.

Farmers in China often have financial requirements, income patterns, and risk profiles that are different from those in other occupations. So, this research contributes significantly to the literature on risk assessment in the agricultural sector. Agriculture is inherently exposed to many risks, due to weather-related uncertainty, market volatility, and policy shifts. Small farmers in particular are especially susceptible to financial hardship and credit default as a result of these variables. Thus, understanding the connection between financial literacy and credit default offers valuable insights into ways to mitigate this vulnerability.

Our paper establishes a direct correlation between financial literacy and credit default by farmers, with substantial policy implications for agricultural finance. Policy makers and agricultural lending institutions can leverage these findings to tailor financial education initiatives and lending practices to better cater to the particular needs of farmers. For example, governments, financial institutions, and agricultural organizations can use these findings as a basis for expanding and justifying financial education programs aimed at farmers. These programs should be targeted at enhancing financial literacy, crafting and providing financial products that align with the varying levels of financial literacy by farmers while managing the associated risks.

Although this study offers valuable insights into the role of financial literacy in predicting loan default by farmers, it has some limitations. Future research could address these limitations, yielding more robust and more widely applicable insights into the relationship between financial literacy and credit default prediction. First, the data used come from a single rural commercial bank in China, potentially limiting the generalizability of our findings to other regions or countries. Future research could benefit from collecting data from more diverse financial institutions to enhance the applicability of the results. Second, the financial literacy variables rely on self-reported survey responses, which can introduce measurement bias. Third, although the hybrid k-means and Adaboost model used in this study is effective, there is room for improvement. Subsequent research could also explore optimizing model parameters, using different clustering or ensemble methods, and incorporating more advanced machine-learning techniques, such as deep learning and neural networks, to enhance predictive performance and model interpretability.

Availability of data

The data that support the findings of this study are available from the corresponding author upon reasonable request.

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.

Acknowledgement

The authors appreciate the financial support from the National Social Science Fund of China (Project No: 20BJY158).

The authors deeply thank Professor Ali Kutan (the editor) and the anonymous reviewers for their invaluable comments, which enhanced the quality of our paper.

References

Abreu, M., & Mendes, V. (2010). Financial literacy and portfolio diversification. *Quantitative Finance, 10*(5), 515–528. https://doi.org/10.1080/14697680902878105

Ahlin, C., Lin, J., & Maio, M. (2011). Where does microfinance flourish? Microfinance institution performance in macroeconomic context. *Journal of Development Economics, 95*(2), 105–120. https://doi.org/10.1016/j.jdeveco.2010.04.004

Albaity, M., & Rahman, M. (2019). The intention to use Islamic banking: An exploratory study to measure Islamic financial literacy. *International Journal of Emerging Markets, 14*(5), 988–1012. https://doi.org/10.1108/ijoem-05-2018-0218

Awunyo-Vitor, D. (2018). Theoretical and conceptual framework of access to financial services by farmers in emerging economies: Implication for empirical analysis. *Acta Universitatis Sapientiae, Economics and Business, 6*(1), 43–59. https://doi.org/10.1515/auseb-2018-0003

Baidoo, S. T., Yusif, H., & Ayesu, E. K. (2020). Improving loan repayment in Ghana: Does financial literacy matter? *Cogent Economics & Finance, 8*(1), Article 1787693. https://doi.org/10.1080/23322039.2020.1787693

Björkegren, D., & Grissen, D. (2020). Behavior revealed in mobile phone usage predicts credit repayment. *The World Bank Economic Review, 34*(3), 618–634.

Boulesteix, A. L., Janitza, S., Kruppa, J., & König, I. R. (2012). Overview of random forest methodology and practical guidance with emphasis on computational biology and bioinformatics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2*(6), 493–507. https://doi.org/10.1002/widm.1072

Calvet, L. E., Campbell, J. Y., & Sodini, P. (2007). Down or out: Assessing the welfare costs of household investment mistakes. *Journal of Political Economy, 115*(5), 707–747. https://doi.org/10.1086/524204

Cao-Alvira, J. J., Novoa-Hoyos, A., & Núñez-Torres, A. (2021). On the financial literacy, indebtedness, and wealth of Colombian households. *Review of Development Economics, 25*(2), 978–993. https://doi.org/10.1111/rode.12739

Chang, D., Chen, W., Tai, X., & Si, Y. (2022). The impact of financial literacy on rural household self-employment: The mediating role of financial ability. *Emerging Markets Finance and Trade, 58*(11), 3297–3308.

Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends, 2*(1), 20–28. https://doi.org/10.38094/jastt20165

Chen, J., Jiang, J., & Liu, Y. J. (2018). Financial literacy and gender difference in loan performance. *Journal of Empirical Finance, 48*, 307–320. https://doi.org/10.1016/j.jempfin.2018.06.004

Chopra, A., & Bhilare, P. (2018). Application of ensemble models in credit scoring models. *Business Perspectives and Research, 6*(2), 129–141. https://doi.org/10.1177/2278533718765531

Crawford, G. S., Pavanini, N., & Schivardi, F. (2018). Asymmetric information and imperfect competition in lending markets. *The American Economic Review, 108*(7), 1659–1701. https://doi.org/10.1257/aer.20150487

De Castro Vieira, J. R., Barboza, F., Sobreiro, V. A., & Kimura, H. (2019). Machine learning models for credit analysis improvements: Predicting low-income families' default. *Applied Soft Computing, 83*, Article 105640. https://doi.org/10.1016/j.asoc.2019.105640

De Cnudde, S., Moeyersoms, J., Stankova, M., Tobback, E., Javaly, V., & Martens, D. (2015). *Who cares about your facebook friends? Credit scoring for microfinance. Working paper*.

Demirguc-Kunt, A., Klapper, L., & Singer, D. (2017). *Financial inclusion and inclusive growth: A review of recent empirical evidence*. World Bank. Retrieved from https://documents1.worldbank.org/curated/pt/403611493134249446/pdf/WPS8040.pdf.

Ding, C., Cao, X. J., & Næss, P. (2018). Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo. *Transportation Research Part A: Policy and Practice, 110*, 107–117. https://doi.org/10.1016/j.tra.2018.02.009

Disney, R., & Gathergood, J. (2013). Financial literacy and consumer credit portfolios. *Journal of Banking & Finance, 37*(7), 2246–2254.

Dlugosch, T. J., Klinger, B., Frese, M., & Klehe, U. C. (2018). Personality-based selection of entrepreneurial borrowers to reduce credit risk: Two studies on prediction models in low-and high-stakes settings in developing countries. *Journal of Organizational Behavior, 39*(5), 612–628. https://doi.org/10.1002/job.2236

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *The American Economic Review, 100*(3), 1238–1260. https://doi.org/10.1257/aer.100.3.1238

Domeher, D., & Abdulai, R. (2012). Access to credit in the developing world: Does land registration matter? *Third World Quarterly, 33*(1), 161–175. https://doi.org/10.1080/01436597.2012.627254

Drechsler, J., & Reiter, J. P. (2011). An empirical evaluation of easily implemented, nonparametric methods for generating synthetic datasets. *Computational Statistics & Data Analysis, 55*(12), 3232–3243. https://doi.org/10.1016/j.csda.2011.06.006

Fine, S. (2023). Banking on personality: Psychometrics and consumer creditworthiness. *The Journal of Credit Risk, 19*(2), 1–20. https://doi.org/10.21314/jcr.2023.001

Gao, Q., Lin, M., & Sias, R. (2023). Words matter: The role of readability, tone, and deception cues in online credit markets. *Journal of Financial and Quantitative Analysis, 58*(1), 1–28. https://doi.org/10.1017/S0022109022000850

Gaurav, S., & Singh, A. (2012). An inquiry into the financial literacy and cognitive ability of farmers: Evidence from rural India. *Oxford Development Studies, 40*(3), 358–380. https://doi.org/10.1080/13600818.2012.703319

Ge, R., Feng, J., Gu, B., & Zhang, P. (2017). Predicting and deterring default with social media information in peer-to-peer lending. *Journal of Management Information Systems, 34*(2), 401–424. https://doi.org/10.1080/07421222.2017.1334472

Grohmann, A., Klühs, T., & Menkhoff, L. (2018). Does financial literacy improve financial inclusion? Cross country evidence. *World Development, 111*, 84–96.

Gulati, P., Sharma, A., & Gupta, M. (2016). Theoretical study of decision tree algorithms to identify pivotal factors for performance improvement: A review. *International Journal of Computer Application, 141*(14), 19–25. https://doi.org/10.5120/ijca2016909926

Guo, S., & Liang, P. (2014). Social interactions, information channel and household stock market participation: An empirical study based on 2011 Chinese Household Finance Survey. *Economic Research Journal, 49*(S1), 116–131.

Hastings, J., & Tejeda-Ashton, L. (2008). *Financial literacy, information and demand elasticity: Survey and experimental evidence from Mexico*. NBER Working Paper No. 14538. Retrieved from https://www.nber.org/system/files/working_papers/w14538/w14538.pdf.

Hsiao, Y. H., Su, C. T., & Fu, P. C. (2020). Integrating MTS with bagging strategy for class imbalance problems. *International Journal of Machine Learning and Cybernetics, 11*, 1217–1230. https://doi.org/10.1007/s13042-019-01033-1

Isakson, S. R. (2015). Derivatives for development? Small-farmer vulnerability and the financialization of climate risk management. *Journal of Agrarian Change, 15*(4), 569–580. https://doi.org/10.1111/joac.12124

Jappelli, T., & Padula, M. (2013). Investment in financial literacy and saving decisions. *Journal of Banking & Finance, 37*(8), 2779–2792.

Kidwell, B., & Turrisi, R. (2004). An examination of college student money management tendencies. *Journal of Economic Psychology, 25*(5), 601–616.

Kim, M. J., & Kang, D. K. (2010). Ensemble with neural networks for bankruptcy prediction. *Expert Systems with Applications, 37*(4), 3373–3379. https://doi.org/10.1016/j.eswa.2009.10.012

Klapper, L., Lusardi, A., & Panos, G. A. (2013). Financial literacy and its consequences: Evidence from Russia during the financial crisis. *Journal of Banking & Finance, 37*(10), 3904–3923. https://doi.org/10.1016/j.jbankfin.2013.07.014

Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research, 259*(2), 689–702. https://doi.org/10.1016/j.ejor.2016.10.031

Krische, S. D. (2019). Investment experience, financial literacy, and investment-related judgments. *Contemporary Accounting Research, 36*(3), 1634–1668.

Kumar, P., Pillai, R., Kumar, N., & Tabash, M. I. (2023). The interplay of skills, digital financial literacy, capability, and autonomy in financial decision making and well-being. *Borsa Istanbul Review, 23*(1), 169–183. https://doi.org/10.1016/j.bir.2022.09.012

Kurowski, Ł. (2021). Household's overindebtedness during the COVID-19 crisis: The role of debt and financial literacy. *Risks, 9*(4), 62. https://doi.org/10.3390/risks9040062

Lessmann, S., Baesens, B., Seow, H. V., & Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research, 247*(1), 124–136. https://doi.org/10.1016/j.ejor.2015.05.030

Li, T. (2006). Social interaction, trust, and stock market participation. *Economic Research Journal, 1*, 34–45.

Li, J., Li, Q., & Wei, X. (2020). Financial literacy, household portfolio choice and investment return. *Pacific-Basin Finance Journal, 62*, Article 101370. https://doi.org/10.1016/j.pacfin.2020.101370

Liberati, C., & Camillo, F. (2018). Personal values and credit scoring: New insights in the financial prediction. *Journal of the Operational Research Society, 69*(12), 1994–2005. https://doi.org/10.1080/01605682.2017.1417684

Liu, S., He, J., & Xu, D. (2023). Understanding the relationship between financial literacy and Chinese rural households' entrepreneurship from the perspective of credit constraints and risk preference. *International Journal of Environmental Research and Public Health, 20*(6), 4981. https://doi.org/10.3390/ijerph20064981

Lu, T., Zhang, Y., & Li, B. (2023). Profit vs. equality? The case of financial risk assessment and a new perspective on alternative data. MIS quarterly. *Forthcoming*. https://doi.org/10.2139/ssrn.3758120

Luo, C. (2020). A comprehensive decision support approach for credit scoring. *Industrial Management & Data Systems, 120*(2), 280–290. https://doi.org/10.1108/imds-03-2019-0182

Lusardi, A. (2019). Financial literacy and the need for financial education: Evidence and implications. *Swiss Journal of Economics and Statistics, 155*(1), 1–8. https://doi.org/10.1186/s41937-019-0027-5

Machado, M. R., & Karray, S. (2022). Assessing credit risk of commercial customers using hybrid machine learning algorithms. *Expert Systems with Applications, 200*, Article 116889. https://doi.org/10.1016/j.eswa.2022.116889

Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings, 3*(1), 91–99. https://doi.org/10.1016/j.gltp.2022.04.020

Mandell, L. (2006). *Financial literacy: If it's so important, why isn't it improving? Policy brief No. 2006-NaN-8*. Networks Financial Institute. https://www.indstate.edu/business/sites/business.indstate.edu/files/Docs/2006-PB-08_Mandell.pdf.

Meoli, M., Rossi, A., & Vismara, S. (2022). Financial literacy and security-based crowdfunding. *Corporate Governance: An International Review, 30*(1), 27–54.

Mi, Y., Chen, X., & He, Q. (2022). Evaluating the effect of China's targeted poverty alleviation policy on credit default: A regression discontinuity approach. *Applied Economics, 54*(31), 3654–3667. https://doi.org/10.1080/00036846.2021.2016585

Netzer, O., Lemaire, A., & Herzenstein, M. (2019). When words sweat: Identifying signals for loan default in the text of loan applications. *Journal of Marketing Research, 56*(6), 960–980. https://doi.org/10.2139/ssrn.2865327

Nkundabanyanga, S. K., Kasozi, D., Nalukenge, I., & Tauringana, V. (2014). Lending terms, financial literacy and formal credit accessibility. *International Journal of Social Economics, 41*(5), 342–361. https://doi.org/10.1108/ijse-03-2013-0075

OECD/INFE. (2011). *Measuring financial literacy: Core questionnaire in measuring financial literacy. Questionnaire and guidance notes for conducting an internationally comparable survey of financial literacy*.

Owusu, G. M. Y., Korankye, G., Yankah, N. Y. M., & Donkor, J. B. A. (2023). *Financial risk tolerance and its determinants: The perspective of personnel from security services in Ghana. Borsa Istanbul Review*. https://doi.org/10.1016/j.bir.2023.02.005

Oyedele, A., Ajayi, A., Oyedele, L. O., Delgado, J. M. D., Akanbi, L., Akinade, O., & Bilal, M. (2021). Deep learning and boosted trees for injuries prediction in power infrastructure projects. *Applied Soft Computing, 110*, Article 107587. https://doi.org/10.1016/j.asoc.2021.107587

Pedro, J. S., Proserpio, D., & Oliver, N. (2015). MobiScore: Towards universal credit scoring from mobile phone data. In *International conference on user modeling, adaptation, and personalization* (pp. 195–207). Cham: Springer International.

Petrick, M. (2005). Empirical measurement of credit rationing in agriculture: A methodological survey. *Agricultural Economics, 33*(2), 191–203. https://doi.org/10.1111/j.1574-0862.2005.00384.x

Pietrzak, M. (2022). Can financial sector distress be detected early? *Borsa Istanbul Review, 22*(6), 1132–1144. https://doi.org/10.1016/j.bir.2022.08.002

Pitthan, F., & De Witte, K. (2021). Puzzles of insurance demand and its biases: A survey on the role of behavioural biases and financial literacy on insurance demand. *Journal of Behavioral and Experimental Finance, 30*, Article 100471.

Poon, K., & Weersink, A. (2011). Factors affecting variability in farm and off-farm income. *Agricultural Finance Review, 71*(3), 379–397. https://doi.org/10.1108/00021461111177639

Raza, A., Tong, G., Sikandar, F., Erokhin, V., & Tong, Z. (2023). Financial literacy and credit accessibility of rice farmers in Pakistan: Analysis for Central Punjab and Khyber Pakhtunkhwa regions. *Sustainability, 15*(4), 2963. https://doi.org/10.3390/su15042963

Rishehchi, F. M., Rasouli, M. R., & Amiri, B. (2021). A data-driven and network-aware approach for credit risk prediction in supply chain finance. *Industrial Management & Data Systems, 121*(4), 785–808. https://doi.org/10.1108/imds-01-2020-0052

Rozo, B. J. G., Crook, J., & Andreeva, G. (2023). The role of web browsing in credit risk prediction. *Decision Support Systems, 164*, Article 113879. https://doi.org/10.1016/j.dss.2022.113879

Santos, D. B., & Gallucci, N. H. (2020). Financial illiteracy and customer credit history. *Revista Brasileira de Gestão de Negócios, 22*, 421–436.

Sarfo, Y., Musshoff, O., & Weber, R. (2023). Farmers' awareness of digital credit: Does financial literacy matter? *Journal of International Development*. https://doi.org/10.1002/jid.3774

Schein, A. I., & Ungar, L. H. (2007). Active learning for logistic regression: An evaluation. *Machine Learning, 68*, 235–265. https://doi.org/10.1007/s10994-007-5019-5

Singh, U. (2014). Financial literacy and financial stability are two aspects of efficient economy. *Journal of Finance, Accounting and Management, 5*(2), 59. https://doi.org/10.1007/978-981-10-0360-8_7

Singh, K., Malik, D., & Sharma, N. (2011). Evolving limitations in K-means algorithm in data mining and their removal. *International Journal of Computational Engineering & Management, 12*(1), 105–109. https://doi.org/10.1109/icdm51629.2021.00114

Sivaramakrishnan, S., Srivastava, M., & Rastogi, A. (2017). Attitudinal factors, financial literacy, and stock market participation. *International Journal of Bank Marketing, 35*(5), 818–841. https://doi.org/10.1108/ijbm-01-2016-0012

Skees, J. R., Hartell, J., & Murphy, A. G. (2007). Using index-based risk transfer products to facilitate micro lending in Peru and Vietnam. *American Journal of Agricultural Economics, 89*(5), 1255–1261. https://doi.org/10.1111/j.1467-8276.2007.01093.x

Skimmyhorn, W. (2016). Assessing financial education: Evidence from boot camp. *American Economic Journal: Economic Policy, 8*(2), 322–343. https://doi.org/10.1257/pol.20140283

Tao, X., Li, Q., Guo, W., Ren, C., Li, C., Liu, R., & Zou, J. (2019). Self-adaptive cost weights-based support vector machine cost-sensitive ensemble for imbalanced data classification. *Information Sciences, 487*, 31–56. https://doi.org/10.1016/j.ins.2019.02.062

Van Rooij, M., Lusardi, A., & Alessie, R. (2011). Financial literacy and stock market participation. *Journal of Financial Economics, 101*(2), 449–472. https://doi.org/10.3386/w13565

Walker, K. W., & Jiang, Z. (2019). Application of adaptive boosting (AdaBoost) in demand-driven acquisition (DDA) prediction: A machine-learning approach. *The Journal of Academic Librarianship, 45*(3), 203–212. https://doi.org/10.1016/j.acalib.2019.02.013

Wang, H., Kou, G., & Peng, Y. (2021). Multi-class misclassification cost matrix for credit ratings in peer-to-peer lending. *Journal of the Operational Research Society, 72*(4), 923–934. https://doi.org/10.1080/01605682.2019.1705193

Xia, T., Wang, Z., & Li, K. (2014). Financial literacy overconfidence and stock market participation. *Social Indicators Research, 119*, 1233–1245. https://doi.org/10.1007/s11205-013-0555-9

Xu, H., Song, K., Li, Y., & Ankrah Twumasi, M. (2023). The relationship between financial literacy and income structure of rural farm households: Evidence from Jiangsu, China. *Agriculture, 13*(3), 711. https://doi.org/10.3390/agriculture13030711

Xu, S., Yang, Z., Ali, S. T., Li, Y., & Cui, J. (2022). Does financial literacy affect household financial behavior? The role of limited attention. *Frontiers in Psychology, 13*, Article 906153. https://doi.org/10.3389/fpsyg.2022.906153

Yu, S. S., Chu, S. W., Wang, C. M., Chan, Y. K., & Chang, T. C. (2018). Two improved k-means algorithms. *Applied Soft Computing, 68*, 747–755. https://doi.org/10.1016/j.asoc.2017.08.032

Yuesti, A., Ni, W. R., & Suryandari, N. N. A. (2020). Financial literacy in the COVID-19 pandemic: Pressure conditions in Indonesia. *Entrepreneurship and Sustainability Issues, 8*(1), 884–898. https://doi.org/10.9770/jesi.2020.8.1(59

Zhang, L., Cheng, W., Cheng, E., & Wu, B. (2020). Does land titling improve credit access? Quasi-Experimental evidence from rural China. *Applied Economics, 52*(2), 227–241. https://doi.org/10.1080/00036846.2019.1644446

Zhou, L., Lai, K. K., & Yu, L. (2010). Least squares support vector machines ensemble models for credit scoring. *Expert Systems with Applications, 37*(1), 127–133. https://doi.org/10.1016/j.eswa.2009.05.024