Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-024-02700-7 OPEN

# Credit risk assessment using the factorization machine model with feature interactions

Jing Quan[1]✉ & Xuelian Sun[1]

The accuracy of credit risk evaluation is crucial for the profitability of any financial institution. The factorization machine is a widely available model that can effectively be utilized for classification or regression through appropriate feature transformation. In this article, we apply the factorization machine model to the field of credit risk assessment. Since some features of the credit risk assessment data are not numerical, one-hot encoding is used, resulting in sparse training data. However, the computational complexity of the factorization machine is polynomial. To illustrate the effectiveness of the factorization machine credit risk assessment model and compare its performance with other classification approaches such as logical regression, support vector machine, *k*-nearest neighbors, and artificial neural network, we conduct numerical experiments on four real-world credit risk evaluation datasets. The experimental results demonstrate that the proposed factorization machine credit risk assessment model achieves higher accuracy compared to other machine-learning models on real-world datasets and is computationally more efficient. Therefore, the factorization machine model can be considered as a suitable candidate for credit risk assessment.

[1] School of Science, Chongqing University of Technology, Chongqing, China. ✉email: jquan@cqut.edu.cn

## Introduction

The development of modern society is intrinsically tied to the economy, and the economy heavily relies on credit. Credit plays a vital role in the financial transactions of corporations and financing institutions, as well as for consumers, especially in modern financial societies.

Credit risk is the probability that the borrower may default on their debt obligations as mutually agreed upon in the credit agreement after the lender provides credit to the borrower. It is typically considered when a financial institution has a loan relationship with a borrower. Banks typically provide credit to their customers in the form of corporate and consumer loans. In other words, the credit risk of an institution refers to the likelihood that the borrower may fail to pay their debts within the time and conditions stated in the credit agreement (Apostolik et al., 2009).

Credit risk, also known as counterpart risk, is the primary risk faced by financing institutions. Historical failures of financial institutions have been linked to credit exposure, such as the collapse of Bank Herstatt in 1974 (Jorion, 2003). In recent years, many financial institutions have faced significant losses due to rising counterparty defaults and bad loans. Tunisia is one of the countries most affected by credit risk, with the non-performing loan ratio of Tunisian banks rising from 13.2% in 2009 to 16.2% in 2014, reaching international high standards. The recent financial crisis and the new Basel II regulatory issues have caused widespread concern about credit risk analysis among financial institutions (Rayo Cantón et al., 2010). Leo (Leo et al., 2019) conducted a comprehensive literature review to analyze and evaluate machine-learning techniques applied in the field of banking risk management. The primary objective was to identify areas or problems within risk management that have received insufficient attention in previous research and have the potential to be explored further.

Credit risk assessment has traditionally involved classifying credit applicants into default and non-default categories based on their personal characteristics such as age, income, and employment status, as well as information about previous applicants and their performance. Credit risk rating or credit risk evaluation is widely used in various business areas to estimate credit risk and prevent harm from it (Huang et al., 2007; Huang and Wang, 2017). Credit risk assessment is particularly important for financial corporations, especially banks and credit card companies. Therefore, the ability to distinguish between default and non-default counterparties is crucial for credit-granting institutions. To this end, numerous credit risk assessment models have been proposed for the credit industry (Thomas, 2000, West, 2000).

Parametric statistical models, such as logistic regression (LR) and linear discriminant analysis (LDA), have been widely applied in credit risk assessment (Kleimeier, 2007; Rayo Cantón et al., 2010). Both LR and LDA models have the ability to estimate the probability of an instance belonging to a specific category, providing more than just a binary classification result. They can effectively handle datasets with a combination of categorical and continuous variables. However, both LR and LDA rely on linear assumptions, assuming that features are independent of each other. This can limit their ability to accurately capture the complex nonlinear relationships and feature interactions that are present in credit risk assessment. Bitetto et al. (Bitetto et al., 2021) compared the performance of a non-parametric approach using a historical random forest (HRF) model with a parametric approach using an ordered Probit model for estimating credit risk in small and mid-sized businesses. Their findings suggest that the non-parametric approach using the HRF model outperforms the parametric approach. Several studies have delved into credit risk assessment modeling, such as references (Henley, 1997; Rosenberg and Gleit, 1994; Thomas et al., 2002, 2005). However, these approaches may not fully meet the performance requirements to differentiate between default and non-default customers.

In recent years, artificial intelligence and machine-learning techniques have rapidly developed, leading to the extensive application of support vector machine (SVM) (Gestel et al., 2003; Schebesch and Stecking, 2005a, b), k-nearest neighbors (kNN) (Henley, 1997; Laha, 2007), decision trees (DT) (Davis et al., 1992), and artificial neural network (ANN) (Desai et al., 1996; Malhotra, 2002; West, 2000) in the field of credit assessment. These techniques have been proven to be more effective than statistical models and optimization techniques (Yu et al., 2008). Krivorotov (Krivorotov, 2023) constructed traditional risk models as well as ML-based profit models (ML: machine learning). The findings revealed that in the absence of risk guardrails, profit-based underwriting in card portfolios could potentially lead to an increase in riskiness. Guan (Guan et al., 2023) proposed the combined model, integrating machine learning and human expert rules, not only achieved comparable performance to a model trained on a larger dataset but also demonstrated improved decision-making capabilities. Among the various evaluation methods, the supervised learning algorithm SVM, first introduced by Vapnik (Vapnik, 1998, 1995), has been extensively applied in credit risk management and has achieved better performance compared to other classification techniques (Gestel et al., 2003; Schebesch and Stecking, 2005a, b; Wang et al., 2005). Some papers have focused on improving the intelligent optimization ability of credit scoring algorithms (Danėnas and Garsva, 2015; Harris, 2015; Jae Kim and Ahn, 2012).

In 2010, Steffen Rendle proposed the factorization machine (FM) model (Rendle, 2010), which combines the advantages of SVM and factorization models. Like SVM, FM is a universal predictor that uses any real-valued feature vector. However, unlike SVM, FM uses factorization parameters to model the interactions between all variables, enabling it to estimate interactions even when SVM fails. Additionally, FM model parameters can be trained straightforwardly, unlike nonlinear support vector machines that require dual forms. Steffen Rendle showed that the FM model parameters can be optimized in polynomial time, making the prediction ability of the FM model stronger. He also demonstrated the distinction between SVM and FM, particularly in parameter estimation in sparse settings.

Previous research in credit risk assessment has mainly focused on traditional machine-learning models such as LR, SVM, kNN, and ANN. While these models can capture linear relationships between features, they may struggle to capture complex interactions and nonlinear relationships. The research gap lies in exploring the use of the FM model, which is specifically designed to capture feature interactions, in the context of credit risk assessment.

Credit risk assessment is a critical task in the banking and finance industry, aiming to evaluate the creditworthiness of borrowers and make informed lending decisions. Traditional models may not fully exploit the potential of feature interactions, which can provide valuable insights into credit risk. The motivation behind this study is to investigate whether the FM model can improve credit risk assessment by capturing feature interactions more effectively. In this study, we propose building a credit risk assessment model using FM on real-world datasets to address the problem of no interactions between all features. Our proposal is different from previous practices as we consider the factorization machine model and real-world data.

The study contributes by applying the FM model to credit risk assessment, specifically focusing on capturing feature interactions. By leveraging the FM model's ability to model high-order feature

interactions, the study aims to enhance the predictive performance of credit risk assessment models. The research contributes to the existing literature by providing empirical evidence on the effectiveness of the FM model in credit risk assessment and comparing its performance with traditional models. Additionally, the study may explore novel feature engineering techniques or model enhancements to further improve the FM model's performance in credit risk assessment.

In summary, the research gap of this study lies in the application of the FM model with feature interactions to credit risk assessment, motivated by the potential to improve predictive performance compared to traditional models. The contribution lies in providing empirical evidence, comparing performance, and potentially introducing novel techniques to enhance the FM model's effectiveness in credit risk assessment. In the following, we first provide preliminaries and a brief review of the FM model. In the second stage, we present our research methodology, which includes credit risk assessment using the FM model for real-world datasets. In the following section, we discuss the experimental results and provide our analysis. Finally, we present our conclusions.

## Preliminaries and factorization machines model

The main objective of machine learning is to establish a function $y: R^n \to S$, which maps an $n$-dimensional real-valued feature vector $x \in R^n$ to a set $S$. If $S$ represents the set of all real numbers, it is referred to as regression. On the other hand, if $S$ represents sets such as $\{+,-\}$, $\{+1,-1\}$, or $\{yes, no\}$, it is known as classification.

In the field of machine learning, we assume the existence of a dataset $B = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \cdots, (x^{(N)}, y^{(N)})\}$, consisting of examples that are used to train the model's parameters. In real-world scenarios, many data features are described in non-numerical forms, making it challenging for most machine-learning methods to handle non-numeric characteristics. Converting character features into numbers using thermal coding is necessary to achieve better classification or regression performance. However, this often leads to highly sparse data where most components $x_k^{(i)}$ of the point $x^{(i)}$ are zero. In a credit risk assessment data system with $N$ points and $m$ features, the $i^{th}$ data point is represented by $(x^{(i)}, y^{(i)})$, where $y^{(i)} \in \{+1, -1\}$ ($+1$ and $-1$ indicate breach or non-compliance and compliance or non-default, respectively). In most machine-learning models, the logical function $\sigma(\psi) = \frac{1}{1+\exp(-\psi)}$ is used to model the likelihood of treaty violation.

Before delving into the proposed model, let us briefly review the linear regression model, logistic regression model, and support vector machine model. For a given eigenvector $(x_1, x_2, \cdots, x_m)^T$, the function used in the linear regression model is:

$$y(x) := w_0 + w_1x_1 + w_2x_2 + \cdots + w_mx_m = w_0 + \sum_{i=1}^{m} w_ix_i, \tag{2.1}$$

where $w_0$ and $w = (w_1, w_2, \cdots, w_m)^T$ are model parameters. The logistic regression model is

$$y(x) := \frac{1}{1 + e^{-(w^Tx+w_0)}}, \tag{2.2}$$

where $w^Tx + w_0 = w_1x_1 + w_2x_2 + \cdots + w_mx_m + w_0 = \sum_{i=1}^{m} w_ix_i + w_0, w = (w_1, w_2, \cdots, w_m)^T$ and $w_0$ are model parameters. While the support vector machine model is to find a hyperplane $w^Tx + w_0 = 0$, such that

$$\begin{aligned} &\min_{w,w_0} \tfrac{1}{2} w^Tw, \\ &s.t.\ y_i(wx_i + w_0) \geq 1, i = 1, 2, \cdots, l, \end{aligned} \tag{2.3}$$

where $w = (w_1, w_2, \cdots, w_m)^T$ and $w_0$ are model parameters, and $y_i \in \{+1, -1\}$. However, it is evident that the feature vectors are isolated from one another in the above models. The models only consider individual feature components, and there is no interaction between them. Typically, the parameter $w$ is also adapted by the FM model as a specific parameter of the corresponding model.

The factorization machine (FM) model was proposed by Steffen Rendle (Rendle, 2010), which is a common predictor that is similar to SVM. And it addresses the problem of isolated feature vectors in previous models. It is defined as follows:

$$y(x) := w_0 + \sum_{i=1}^{m} w_ix_i + \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} Interaction(i, j),$$

where the interaction between the $i^{\text{th}}$ vector and the $j^{\text{th}}$ vector is denoted by $Interaction(i, j)$. If the parameters $v_i$ and $v_j$ are the vector embeddings for the $i^{\text{th}}$ feature and the $j^{\text{th}}$ feature, respectively, then the interaction $Interaction(i, j)$ can be defined as $v_i^Tv_jx_ix_j$. The model formula for a second-degree FM is described as follows:

$$y(x) := w_0 + \sum_{i=1}^{m} w_ix_i + \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \langle v_i, v_j \rangle x_ix_j, \tag{2.4}$$

where the factorization machine model parameters $w_0 \in R$, $w = (w_1, w_2, \cdots, w_m)^T \in R^m$, and $V \in R^{m \times k}$ need to be estimated. The parameter $w_0$ is a scalar bias term, $w$ is a weight vector of size $m$ that is associated with the linear term, and $V$ is a matrix of size $m \times k$ that contains the latent vector representations of the feature interactions. The dot product of vectors $v_i$ and $v_j$, whose size is $k$, is denoted by $\langle v_i, v_j \rangle$.

$$\langle v_i, v_j \rangle := \sum_{f=1}^{k} v_{i,f} \cdot v_{j,f} \tag{2.5}$$

In the factorization machine model, a row $v_i$ of matrix $V$ represents the $i^{\text{th}}$ variable with $k$ factors. Here, $k \in N^+$ (positive integer) is a hyperparameter that determines the dimensionality of the factorization.

It is well-known that there exists a matrix $V$ that satisfies $W = V \cdot V^T$ when $W$ is positive definite, as long as $k$ is sufficiently large. This indicates that any interaction matrix $W$ can be represented in the FM model when $k$ is large enough. However, in a sparse environment, a small value of $k$ should be considered due to insufficient data to estimate the complex interaction matrix $W$. By restricting $k$, we can improve the expressiveness of the FM model in sparse environments. Moreover, large $k$ is not necessary in practice since the number of features is usually small for one data point.

The formula

$$\sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \langle v_i, v_j \rangle x_ix_j := \frac{1}{2} \sum_{f=1}^{k} \left( \left( \sum_{i=1}^{m} v_{i,f}x_i \right)^2 - \sum_{i=1}^{m} v_{i,f}^2x_i^2 \right)$$

indicates that the FM model can be trained successfully in polynomial time. The FM model can be utilized in various forecasting tasks, including regression and binary classification. In such scenarios, regularization terms such as $L_2$ are often incorporated into the optimization goal to prevent overfitting. For binary classification, the Hinge loss function $loss(\hat{y}, y) = \max\{0, 1 - y\hat{y}\}$, or the logit loss function $loss(\hat{y}, y) = -\ln \sigma(y\hat{y})$, are often applied to train the FM model's parameters. To solve the minimization problem of the loss function, several optimization methods such as the Stochastic Gradient Descent algorithm, Alternating Least Square method, and Markov Chain Monte Carlo method can be employed.

FM have the ability to estimate reliable parameters even in scenarios with high sparsity. FM leverages all interactions among

feature vectors to build the model, similar to the polynomial kernel in SVM. However, FM adopts factorized parameterizations instead of dense parameterizations. This approach enables the estimation of FM model parameters in polynomial time, depending only on a small number of parameters.

## Research methodology

This research work focuses on six aspects: credit dataset description, data cleaning, introduction of compared machine-learning models, experiment settings, evaluation method of model performance, and performance on real-world data sets. The experimental process is illustrated in the Route Fig. 1.

**Credit data sets description**. This section focuses on conducting numerical experiments to evaluate the efficacy of the FM model and compare its performance with other classification approaches, namely LR, SVM, KNN, and ANN. We use four real credit data sets from the UCI machine-learning repository for this purpose. Here, we provide a brief description of the four data sets used in our experiments.

*Bank marketing dataset*. The bank marketing dataset comprises a total of 45,211 samples, among which 39,922 are classified as good and 5289 as bad. This dataset is relevant to direct marketing activities carried out by Portuguese banking institutions, where marketing activities are conducted through phone calls. Often, multiple contacts are required with the same client before a product can be subscribed.

*Credit approval dataset*. The credit approval dataset pertains to credit card application processes and comprises 300 samples with 15 features and one class attribute. The attribute characteristics of this dataset include categorical, integer, and real values. To ensure the confidentiality of the data, all feature names and values have been replaced with meaningless characters. Additionally, this dataset contains some missing values, with some samples containing more than one missing value.

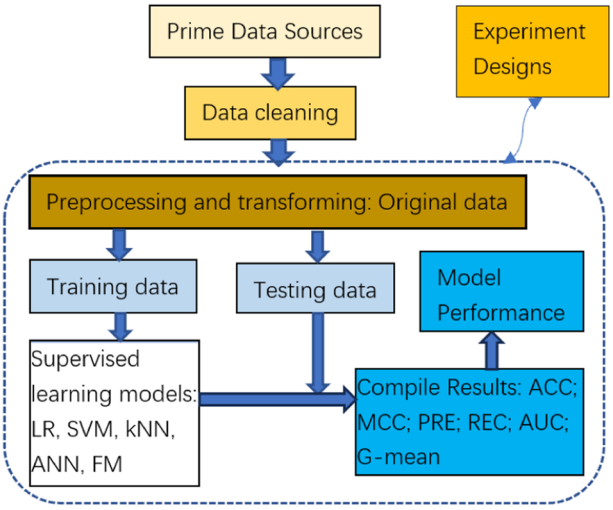


**Fig. 1 Flow chart of the experiment.** It contains the whole experimental process of the work. ACC, MCC, PRE, REC, AUC, G-mean are evaluation measures.

*German credit dataset*. The German credit dataset contains a total of 1000 instances, of which 700 are classified as non-defaulters and 300 as defaulters. Each instance is represented by 20 features or dimensions, comprising seven numerical features and 13 categorical features. The features are based on personal details such as age, employment situation, work, residence, credit record, bank account balance, amount of indebtedness, and use of proceeds, among others.

*Statlog (Australian credit approval) dataset*. The Statlog (Australian credit approval) dataset comprises 690 samples, including 307 bad samples and 383 good samples. It is represented by 14 dimensions, consisting of six consecutive features, eight categorical features, and one class attribute. This dataset is significant as it provides a good mix of attributes, including continuous attributes, nominal attributes with a small number of values, as well as nominal attributes with a large number of values.

Table 1 lists the details of the selected data sets, including the number of attributes or features, the number of instances, the number of bad samples, and the number of good samples for each dataset. The selected data sets comprise instances ranging from 690 to 45,211, with the number of attributes ranging from 15 to 24.

The real-world data sets have been divided into multiple groups using a specific procedure, with each group containing approximately three-quarters of the examples from the dataset. The remaining one-quarter is used as the training set and the testing set, respectively. In this process, a model algorithm is trained using the training set of each group and then tested with the testing set. This approach is commonly used to evaluate the performance of machine-learning algorithms.

**Data cleansing**. Data cleansing is typically focused on two main aspects: handling missing values and processing outliers. Various approaches, such as case deletion, missing data imputation, and model-based programs (García-Laencina et al., 2010), are used to deal with missing values. Often, missing values are processed based on experience. For instance, if 90% or more of the applicants do not fill out a particular feature, it is typically deleted, while missing values can be filled using the best possible values like mean interpolation and maximum likelihood estimation.

Processing outliers can improve model training performance (García et al., 2012). We verify the reasonableness of outliers based on statistical techniques or domain-specific techniques, such as $z$-scores, modified $z$-scores, box plots, or domain knowledge-based thresholds, and if they are rational, they are retained. Otherwise, they are replaced with missing values using the lower and upper values of the box plot. Furthermore, feature values are normalized to fit within a certain range. Non-

**Table 1 Overview of real-world dataset features and sizes.**

| Dataset | Features | Examples | Bad samples | Good samples |
|---|---|---|---|---|
| 1 Bank marketing data | 16 | 45211 | 5289 | 39,922 |
| 2 Credit approval data | 15 | 300 | 89 | 211 |
| 3 German credit data | 21 | 1000 | 300 | 700 |
| 4 Statlog (Australian credit approval) data | 14 | 690 | 307 | 383 |

numerical features can be transformed into numerical values using independent one-hot encoding.

**Introduction of compared machine models**. This section provides a brief introduction to the compared machine-learning models, which include LR, SVM, kNN, and ANN.

*Logistic regression*. Logistic Regression is a machine-learning method used to solve dichotomous and regression problems and has been the standard in the field of credit risk assessment (Lessmann et al., 2015). Unlike linear regression, LR's output is not a specific value but a probability. The LR model first seeks to find the regression function, which is then transformed through a logical function to obtain the prediction value. The LR model's formula is based on the conditional probability distribution:

$$p = P\{y = 1|x\} = \frac{\exp(\omega \cdot x)}{1 + \exp(\omega \cdot x)};$$
$$1 - p = P\{y = -1|x\} = \frac{1}{1 + \exp(\omega \cdot x)}.$$

The log odds or logical function of the event is:

$$\ln\left[\frac{p}{1-p}\right] := \omega \cdot x = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_m x_m = w_0 + \sum_{i=1}^{m} w_i x_i,$$

where $p$ represents the default probability and $w = (w_0, w_1, w_2, \cdots, w_m)^T$ is the model parameter that is related to the vectors $x_i, i = 1, 2, \cdots, m$. In LR models, the conditional probability of a certain sample belonging to a certain class can be predicted using maximum likelihood estimation.

*Support vector machine*. Support Vector Machine is a binary classification model first proposed by Vapnik (Vapnik, 1995). SVM is widely used in credit risk assessment due to its strong predictive power. The basic idea of SVM learning is to obtain a separated hyperplane that can correctly divide the training dataset while maximizing the geometric interval. The hyperplane $w^T x + b = 0$ satisfies (2.3) and is supported by the support vector. The maximum segmented hyperplane problem of the SVM model can be transformed to solve extremely small duality problems (3.6), expressed as the following constrained optimization problem:

$$\min_{\alpha} \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N} \alpha_i,$$
$$s.t. \sum_{i=1}^{N} \alpha_i y_i = 0, \alpha_i \geq 0, i = 1, 2, \cdots, N. \quad (3.6)$$

Based on the characteristics of the support vector, the labels of the testing samples can be estimated. In order to project the input data into a high-dimensional feature space, various kernel functions, such as polynomial kernel functions, radial basis functions, and sigmoid kernel functions (Zhou et al., 2010), can be chosen for nonlinear SVM.

*k-nearest neighbor*. The k-nearest neighbors model is a basic and simple classification algorithm (Cover and Hart, 1967) that has been successfully applied to credit risk assessment (Islam et al., 2007). In the kNN model, the classification of a new sample is determined by the k-training sample points closest to the new sample according to the classification decision rules. The kNN model has three basic elements: the distance measure, the selection of the $k$-value, and the classification decision rule. Euclidean distance is the most commonly used distance measure in the kNN model, and typically, the value of $k$ is less than 20. The classification decision rule of kNN is to make statistics of all samples in the neighborhood of new samples.

One advantage of the kNN model is its simplicity, as the algorithm only has one parameter $k$. A probabilistic strategy was established by Holmes and Adams (Holmes and Adams, 2002) for setting this parameter.

*Artificial neural network*. The artificial neural network model is a powerful machine-learning tool widely used in regression and classification problems (Bishop, 1997). Inspired by biological neurons, it mimics the human brain's mechanisms to cope with complex problems. ANN can establish and visualize a nonlinear equation with an input and output relation. Through reasonable network structure configuration, ANN can fit any nonlinear function.

One type of ANN is the multi-layer feedforward network, which comprises an input layer, hidden layers, and an output layer. In credit risk assessment, the ANN model transmits feature information to the input layer, transfers these characteristics through the hidden layer, and eventually produces the final results through the output layer. Weights are assigned based on the relative importance of each feature, and an activation function, such as sigmoid or tan-sigmoid, combines all the weighted vectors to produce the output (Malhotra and Malhotra, 2003). The weight adjustment processes are repeated in many cycles to minimize the errors between the real class and the estimated class.

**Experimental settings**. In this work, we first discuss the experimental settings from the perspective of the compared machine-learning models and FM. For LR, we select either L-BFGS or stochastic average gradient (SAG) Descent as the solver. For SVM, we choose the kernel function $ker(x_i, x_j) = \exp(-\frac{\|x_i - x_j\|^2}{\sigma^2})$, where the parameter $\sigma$ satisfies the equation $\sigma^2 = \frac{1}{N^2}\sum_{i,j=1}^{N} \| x_i - x_j\|^2$ ($N$ is the number of samples), or we select it from the set $\{10^{-3}, 10^{-2}, \cdots, 100, 1000, 10000\}$. For kNN, we seek the parameter k from 3 to 10. For ANN, we search for the hidden nodes from 10 to 50, and the logarithm is in the set of $\{0.001, 0.01, 0.1, 1\}$. For FM, its setting is similar to SVM.

Furthermore, we take the default values for the parameters not mentioned. We randomly choose 3/4 of the total samples as the training cases, and the remaining samples act as the testing set. We set $MaxIter = 1000$ as the maximal number of the iterations. The computational procedure is executed on Intel Core 2 Processors with 2.66 GHz, 8G RAM, Win10 system, and Python 3.8 environment.

**Evaluation method of model performance**. To evaluate the performance of credit risk assessment models on real datasets, we split the entire dataset into a training set and a testing set with a ratio of approximately 3:1. To validate the FM model for credit risk assessment, we adopt cross-validation methodology to train the model parameters.

We use several performance measures, such as accuracy (ACC), Matthews correlation coefficient (MCC) (Powers, 2011), precision (PRE), recall (REC), $F$-score, true-positive rate (sensitivity, TPR), true-negative rate (specificity, TNR), false-negative rate (type I error, FNR), false-positive rate (type II error, FPR), and the values of area under the ROC curve (AUC) and $G$-mean to assess the classification performance.

ACC shows the overall prediction availability of credit risk assessment models. When the real-world credit dataset is unbalanced, the deviation of ACC can be significant. MCC is

**Table 2 Confusion matrix adopted in this article.**

| True label | Predicted label Default | Predicted label Non-default |
|---|---|---|
| Default | True positive (TP) | False negative (FN) |
| Non-default | False positive(FP) | True negative (TN) |

usually applied to assess the performance of credit risk assessment models on unbalanced datasets (Kong and Yan, 2017). It can depict the confusion matrix more thoroughly than ACC. The MCC ranges from −1 to 1. If the MCC is greater than 0.6, it shows that the credit risk assessment model performs excellently.

PRE denotes the ratio of the number of correctly predicted defaults to the number of all predicted defaults, while REC is the ratio of the number of correctly predicted defaults to the total amount of actual defaults. The *F*-score value denotes the squared geometric mean of Precision and Recall divided by the arithmetic mean of Precision and Recall, that is, the harmonic mean of Precision and Recall. It can be formulated as $\frac{2*Precision*Recall}{Precision+Recall}$. The higher the *F*-score, the better the performance.

The TPR is the ratio of the number of correctly predicted defaults to the overall number of actual defaults. The TNR is the ratio of the number of correctly predicted non-defaults to the total number of actual non-defaults. The FNR is the ratio of the number of wrongly predicted non-defaults to the number of all actual non-defaults. The FPR is the ratio of the number of wrongly predicted defaults to the total amount of actual defaults.

The AUC is a synthetic assessment index that measures the area under the curve of the receiver operating characteristic (ROC) curve (Tom, 2006). AUC ranges from 0 to 1, and a higher AUC indicates better model performance. For binary classification, AUC can be calculated simply by the formula $AUC = \frac{1+TP_{rate}-FP_{rate}}{2}$ (Jin and Ling, 2005), where $TP_{\text{rate}}$ and $FP_{\text{rate}}$ are the percentage of correctly predicted default samples and the percentage of wrongly predicted non-default samples, respectively.

The *G*-mean index is the arithmetic square root of the product of TPR and FPR. It simultaneously considers Sensitivity and Specificity.

The evaluation measures mentioned above are computed based on the values in Table 2, where true positive (TP) denotes the number of samples whose predictions are defaults and the actual results are also defaults. False positive (FP) denotes the number of samples whose predictions are defaults, but the actual results are non-defaults. False negative (FN) denotes the number of samples whose predictions are non-defaults, but the actual results are defaults. Finally, true negative (TN) denotes the number of samples whose predictions are non-defaults, and the actual results are also non-defaults. The measurable indicators used in this article are listed in Table 3.

**Performance on real-world data sets**. In this subsection, we conducted a simulation study following the standard steps presented in Table 4.

Based on the model settings and evaluation criteria mentioned above, we calculated the ACC, MCC, PRE, REC, *F*-score, TPR, TNR, FNR, FPR, AUC, and *G*-mean (%) of the experimental results on the four real datasets using the compared machine-learning models. The values are presented in Tables 5–8, respectively. These evaluation criteria reflect different aspects of credit risk assessment performance.

**Table 3 The evaluation measures adopted in this article.**

| | |
|---|---|
| *Accuracy* | $ACC = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$ |
| *Matthews Correlation Coefficient* | $MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)*(TP+FN)*(TN+FP)*(TN+FN)}}$ |
| *Precition* | $PRE = \frac{TP}{TP+FP} \times 100\%$ |
| *Recall* | $REC = \frac{TP}{TP+FN} \times 100\%$ |
| *Fscore* | $Fscore = \frac{2*PRE*REC}{PRE+REC} \times 100\%$ |
| *True positive rate (sensitivity)* | $TPR = \frac{TP}{TP+FN} \times 100\%$ |
| *True negative rate (specificity)* | $TNR = \frac{TN}{TN+FP} \times 100\%$ |
| *False negative rate (type II error)* | $FNR = \frac{FN}{FN+TP} \times 100\%$ |
| *False positive rate (type I error)* | $FPR = \frac{FP}{FP+TN} \times 100\%$ |
| *The area under curve* | $AUC = \frac{1+TPR-FPR}{2} \times 100\%$ |
| *G-mean* | $G-mean = \sqrt{TPR*FPR} \times 100\%$ |

**Table 4 Standard steps for the experiments.**

Step 1: Load and preprocess the dataset, including data cleaning.
Step 2: Split the preprocessed dataset into training and testing sets with a ratio of 3:1.
Step 3: Build a credit risk assessment model using the factorization machine.
Step 4: Train the factorization machine model on the training set.
Step 5: Predict the outcomes of the testing set.

We prefer to use AUC to assess the performance of the models since it is more comprehensive than other criteria and can more effectively reflect the model performance on real-world datasets.

*Performance on bank marketing dataset*. Table 5 shows the prediction results of the compared machine-learning models on the testing set of the bank marketing data in terms of evaluation measures. The FM model had the best performance, with an ACC of 0.9021, MCC of 0.4922, REC of 0.5736, *F*-score of 0.5535, TPR of 0.5736, and *G*-mean of 0.7318. While the LR model had the best PRE (0.6428) and TNR (0.9754).

The FM model also had the best AUC (0.7343), while the SVM and ANN models also performed well, with AUCs of 0.7066, and 0.7022, respectively. It is worth noting that the FM and LR models outperformed other models in terms of FNR (0.4264) and FPR (0.0662), respectively. For the bank marketing dataset, model FM performs best followed by model LR.

*Performance on credit approval dataset*. Table 6 shows the prediction results of the compared machine-learning models on the testing set of the credit approval data in terms of evaluation measures. The FM model had the best performance, with an ACC of 0.9464, MCC of 0.8546, PRE of 0.9534, REC of 0.9761, *F*-score of 0.9647, TPR of 0.9761, FNR of 0.0238, FPR of 0.1428, AUC of 0.9053, and *G*-mean of 0.9147.

The FM model also had the best AUC (0.9053), while the SVM, LR, kNN, and ANN models also performed well, with AUCs of 0.8259, 0.8077, 0.7914, and 0.7226, respectively.

*Performance on South German credit dataset*. Table 7 presents the prediction results of the compared machine-learning models on the testing set of the South German Credit in terms of evaluation measures. The FM model had the best performance, with an ACC of 0.7996, MCC of 0.4725, PRE of 0.9329, *F*-score of 0.8399, TPR of 0.2083, AUC of 0.8165, and *G*-mean of 0.7776. While the SVM model had the best REC (0.9463), TPR (0.9463), and FNR of 0.0457.

**Table 5 The results(%) of the bank marketing dataset.**

| Models | ACC | MCC | PRE | REC | *F*-score | TPR | TNR | FNR | FPR | AUC | *G*-mean |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LR | 89.98 | 41.45 | **64.28** | 33.17 | 43.76 | 33.17 | **97.54** | 66.83 | **2.45** | 65.36 | 56.88 |
| SVM | 88.89 | 43.69 | 53.16 | 46.83 | 49.80 | 46.83 | 94.50 | 53.16 | 5.49 | 70.66 | 66.52 |
| kNN | 87.71 | 28.07 | 45.82 | 25.63 | 32.87 | 25.63 | 95.96 | 74.36 | 4.03 | 60.80 | 49.59 |
| ANN | 88.23 | 41.67 | 50.00 | 46.66 | 48.27 | 46.66 | 93.77 | 53.33 | 6.22 | 70.22 | 66.15 |
| FM | **90.21** | **49.22** | 53.48 | **57.36** | **55.35** | **57.36** | 93.37 | **42.64** | 6.62 | **73.43** | **73.18** |

The bold values highlight the optimal values of the evaluation indicators.

**Table 6 The results (%) of the credit approval dataset.**

| Models | ACC | MCC | PRE | REC | *F*-score | TPR | TNR | FNR | FPR | AUC | *G*-mean |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LR | 86.81 | 65.75 | 88.23 | 94.33 | 91.18 | 94.33 | 67.21 | 5.66 | 32.78 | 80.77 | 79.62 |
| SVM | 87.50 | 68.04 | 89.40 | 93.75 | 91.52 | 93.75 | 71.42 | 6.25 | 28.57 | 82.59 | 81.83 |
| kNN | 81.03 | 57.56 | 86.84 | 84.61 | 85.71 | 84.61 | 73.68 | 15.38 | 26.31 | 79.14 | 78.96 |
| ANN | 79.06 | 47.06 | 83.07 | 88.52 | 85.71 | 88.52 | 56.00 | 11.47 | 44.00 | 72.26 | 70.40 |
| FM | **94.64** | **85.46** | **95.34** | **97.61** | **96.47** | **97.61** | **85.71** | **2.38** | **14.28** | **90.53** | **91.47** |

The bold values highlight the optimal values of the evaluation indicators.

**Table 7 The results(%) of South German credit dataset.**

| Models | ACC | MCC | PRE | REC | *F*-score | TPR | TNR | FNR | FPR | AUC | *G*-mean |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LR | 76.88 | 38.83 | 80.42 | 89.47 | 83.70 | 89.47 | 45.20 | 10.52 | 54.79 | 67.33 | 63.59 |
| SVM | 76.09 | 45.26 | 75.00 | **94.63** | 83.68 | **94.63** | 41.98 | **4.57** | 58.02 | 68.30 | 63.03 |
| kNN | 76.76 | 46.69 | 79.45 | 87.87 | 83.45 | 87.87 | 55.88 | 12.12 | 44.11 | 71.88 | 70.07 |
| ANN | 68.70 | 28.19 | 73.05 | 81.88 | 77.22 | 81.88 | 44.44 | 17.09 | 55.56 | 63.16 | 63.03 |
| FM | **76.96** | **47.25** | **93.29** | 76.37 | **83.99** | 76.37 | **79.17** | 24.29 | **20.83** | **81.65** | **77.76** |

The bold values highlight the optimal values of the evaluation indicators.

**Table 8 The results(%) of Statlog (Australian credit approval) dataset.**

| Models | ACC | MCC | PRE | REC | *F*-score | TPR | TNR | FNR | FPR | AUC | *G*-mean |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LR | 87.68 | 74.06 | 83.33 | 84.90 | 84.11 | 84.90 | 89.41 | 15.09 | 10.58 | 87.15 | 87.13 |
| SVM | 86.51 | 73.52 | 81.14 | **91.24** | 85.90 | **91.24** | 82.64 | 8.75 | 17.35 | 86.94 | 86.83 |
| kNN | 80.43 | 59.14 | 73.21 | 77.35 | 75.23 | 77.35 | 82.35 | 22.64 | 17.64 | 79.85 | 79.81 |
| ANN | 80.79 | 61.89 | 73.88 | 84.61 | 77.98 | 84.61 | 77.98 | 15.38 | 22.01 | 81.30 | 81.23 |
| FM | **88.44** | **76.78** | **88.52** | 86.63 | **87.80** | 86.63 | **90.04** | **7.10** | **9.95** | **89.28** | **88.32** |

The bold values highlight the optimal values of the evaluation indicators.

The FM model also had the best AUC (0.8165), while the kNN, SVM, LR, and ANN models also performed well, with AUCs of 0.7188, 0.6830, 0.6733, and 0.6316, respectively. For the South German credit dataset, model FM performs best followed by model SVM.

*Performance on Statlog (Australian credit approval) dataset*. Table 8 shows the prediction results of the compared machine-learning models on the testing set of the Statlog (Australian credit approval) dataset in terms of evaluation measures. The FM model had the best performance, with an ACC of 0.8844, MCC of 0.7678, PRE of 0.8852, *F*-score of 0.8780, TNR of 0.9004, FNR of 0.0710, AUC of 0.8928, and *G*-mean of 0.8832. The SVM model had the best REC (0.9124) and TPR (0.9124).

The FM model also had the best AUC (0.8928), while the LR, SVM, ANN, and kNN models also performed well, with AUCs of 0.8715, 0.8694, 0.8130, and 0.7985, respectively. For the Statlog (Australian credit approval) dataset, model FM performs best followed by model SVM.

*Comprehensive performance*. Figure 2 displays the performance of LR, SVM, kNN, ANN, and FM models in terms of ACC, MCC, and *G*-mean measures. The FM model outperformed the other machine-learning models in terms of ACC, MCC, and *G*-mean on every real dataset. The ACC, MCC, and *G*-mean of the FM model were the highest compared to the other machine-learning models.

Figure 3 illustrates the performance of LR, SVM, kNN, ANN, and FM models in terms of *F*-score and AUC measures. The FM model achieved the highest *F*-score and AUC compared to the other machine-learning models on every real dataset.

Although the LR model achieved the highest PRE and TNR on the bank marketing data, the FM model had the highest PRE and TNR on the other three datasets. Moreover, the FM model

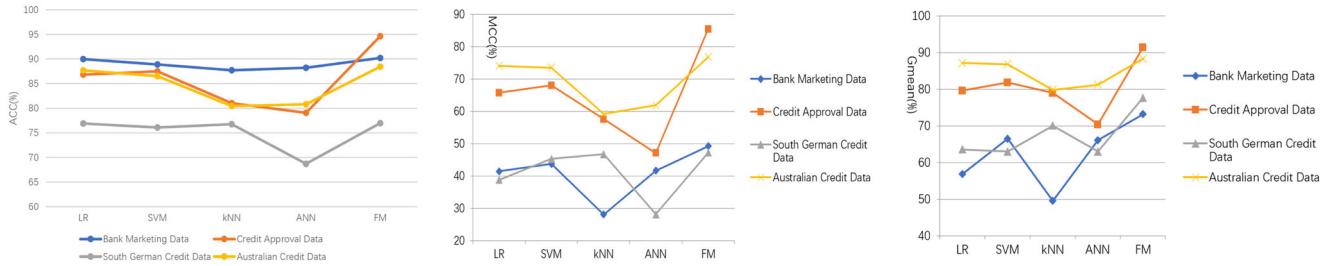


**Fig. 2 Comparison chart of classification performance metrics.** LR, SVM, kNN, ANN and FM are the compared models. Bank marketing dataset, Credit approval data, German credit data, Australian credit data are the used datasets. ACC (left), MCC (center) and Gmean (right) are the classification performance metrics.

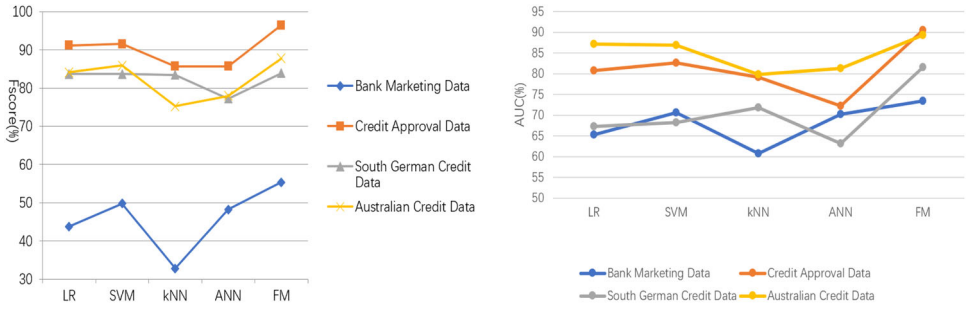


**Fig. 3 Comparison chart of classification performance metrics.** LR, SVM, kNN, ANN and FM are the compared models. Bank marketing dataset, Credit approval data, German credit data, Australian credit data are the used datasets. Fscore (left) and AUC (right) are the classification performance metrics.

achieved the best REC and TPR on the bank marketing data and credit approval data, while the SVM model had the best REC and TPR on the other two datasets. Although the SVM model had the lowest FNR on the South German credit data, the FM model had the best FNR on the other three datasets. Furthermore, the LR model had the best FPR on the bank marketing data, while the FM model had the best FPR on the other three datasets.

By utilizing independent datasets from various sources, countries, or time periods, we are able to evaluate the effectiveness and robustness of the FM model across different scenarios. Applying the FM model to datasets with diverse credit risk distributions enables us to assess its performance compared to other well-established models commonly used in credit risk assessment. Through this comparative analysis of different datasets, we are able to provide evidence of the FM model's generalizability and effectiveness, highlighting its superiority over alternative approaches.

**Why is the factorization machine better**. Factorization machine is a powerful machine-learning technique that excels in capturing interactions between variables, especially in scenarios with high-dimensional and sparse data. This characteristic makes it particularly suitable for credit risk prediction, where the relationships between various features can be complex and nonlinear. Compared to traditional machine-learning methods such as logistic regression, support vector machine, k-nearest neighbors, and artificial neural network, factorization machine offers several advantages. First, it can effectively handle high-dimensional feature spaces and mitigate the issue of overfitting, which is crucial in credit risk assessment where the number of variables can be large. FM achieves this by modeling feature interactions through factorized parameters, allowing it to capture complex patterns and dependencies in the data. Second, the factorization machine incorporates both linear and nonlinear effects, enabling it to capture both simple and complex relationships between variables. This flexibility enhances its predictive performance by effectively modeling the underlying credit risk factors. Furthermore, factorization machines have been successfully applied in various domains and have demonstrated competitive performance compared to other machine-learning methods. Its ability to model feature interactions has proven beneficial in tasks such as recommender systems, click-through rate prediction, and sentiment analysis.

## Experimental results and discussions

The experimental results presented in Tables 5–8 led to several conclusions:

- **(R1)** The overall accuracy of the FM method was found to be the best among all the compared methods, followed by LR, SVM, kNN, and ANN models, clearly demonstrating the effectiveness of the FM credit risk assessment method. Furthermore, the predicted default and non-default accuracies of the FM model were exceptional, indicating that the proposed FM model possesses superior classification ability for credit risk evaluation;
- **(R2)** The average values of ACC, MCC, PRE, *F*-score, AUC, and *G*-mean of the FM model were found to be higher than those of the LR, SVM, kNN, and ANN models, respectively. In addition, the average FNR and FPR of the FM model were lower than those of the compared models, providing clear evidence of the FM model's advantage in classifying datasets compared to LR, SVM, kNN, and ANN models;
- **(R3)** The superior performance of the FM method is attributed to its ability to use feature interactions, which reduces the computational complexity of the model, leading to improved computational efficiency;
- **(R4)** The FM model contributed to the best classification performance on all four real-world datasets, providing strong evidence of the superiority of the FM credit risk

assessment method compared to the other learning machines examined in this study;
- **(R5)** The results indicate that for some datasets, such as the bank marketing dataset, the AUC of the SVM model performed better than that of the FM model, and the PRE and TNR of the LR model were better than those of the FM model. This may be due to the large size of the training set or data imbalance. However, despite these occasional results, the overall findings of this study indicate that the FM credit risk assessment model is superior to the other compared machine-learning models across all datasets.

Although the FM credit risk assessment model demonstrated excellent overall performance across all datasets, there were some instances where its performance was not perfect, particularly in relation to the PRE, TNR, FPR, and REC or TPR measures on some datasets, such as the bank marketing data, South German credit data, and Statlog (Australian credit approval) data. Future improvements to the FM model could focus on optimizing the loss function further, given that the selected optimization techniques (i.e., stochastic gradient descent algorithm, Alternating least square method, and Markov Chain Monte Carlo method) produced varying results across different datasets. Additionally, other applications of FM models may warrant further investigation in future studies.

Open questions: can factorization machine models be adapted to handle online credit risk assessment, where new data arrives continuously and the model needs to be updated in real-time? What are the challenges and potential solutions for implementing factorization machine models in online credit risk assessment systems?

## Conclusions

In today's complex economic landscape, financial credit risk assessment is becoming increasingly important, as risk factors grow more complex and dynamic. This study focuses on credit risk assessment methods using a model called the factorization machine, which requires nested feature vector interactions. To assess the effectiveness of the FM credit risk assessment model, four real-world datasets were used to compare different credit risk assessment models, including LR, SVM, kNN, and ANN. Various criteria, such as ACC, MCC, PRE, REC, *F*-score, TPR, TNR, FNR, FPR, AUC, and *G*-mean, were used to evaluate the classification performance. The experimental results demonstrate that the FM credit risk assessment model outperformed the other machine-learning models on real-world datasets, suggesting its potential effectiveness in the credit risk assessment process. This method should be studied more intensively, as it has the potential to accurately predict realistic credit risk characteristics.

## Data availability

The data that support the findings of this study are openly available in the UCI Machine-Learning Repository at: https://archive.ics.uci.edu/dataset/222/bank+marketing; https://archive.ics.uci.edu/dataset/27/credit+approval; https://archive.ics.uci.edu/dataset/522/south+german+credit; https://archive.ics.uci.edu/dataset/143/statlog+australian+credit+approval.

Received: 13 October 2023; Accepted: 19 January 2024;
Published online: 08 February 2024

## References

Apostolik R, Donohue C, Went P (2009) Foundations of banking risk: an overview of banking, banking risks, and risk-based banking regulation. John Wiley & Sons
Bishop CM (1997) Neural networks for pattern recognition. J Am Stat Assoc 92:1642–1645
Bitetto A, Cerchiello P, Filomeni S, Tanda A, Tarantino B (2021) Machine learning and credit risk: empirical evidence from SMEs. DEM Working Papers Series
Cover TM, Hart PE (1967) Nearest neighbor pattern classification. IEEE Trans Inf Theory 13(1):21–27
Danėnas P, Garsva G (2015) Selection of support vector machines based classifiers for credit risk domain. Expert Syst Appl 42:3194–3204
Davis RH, Edelman DB, Gammerman A (1992) Machine-learning algorithms for credit-card applications. IMA J Manag Math 4:43–51
Desai VS, Crook JN, Overstreet GA (1996) A comparison of neural networks and linear scoring models in the credit union environment. Eur J Oper Res 95:24–37
García V, Marqués AI, Sánchez JS (2012) On the use of data filtering techniques for credit risk prediction with instance-based models. Expert Syst Appl 39(18):13267–13276
García-Laencina PJ, Sancho-Gómez J-L, Figueiras-Vidal AR (2010) Pattern classification with missing data: a review. Neural Comput Appl 19(2):263–282
Gestel TV, Baesens B, Garcia J, Dijcke PV (2003) A support vector machine approach to credit scoring. Bank-en Financiewezen 2:73–82
Guan C, Suryanto H, Mahidadia A, Bain M, Compton P (2023) Responsible credit risk assessment with machine learning and knowledge acquisition. Human-Cent Intell Syst 3:232–243
Harris T (2015) Credit scoring using the clustered support vector machine. Exp Syst Appl 42(2):741–750
Henley WE (1997) Construction of a k-nearest-neighbour credit-scoring system. IMA J Math Appl Bus Ind 8(4):305–321
Holmes CC, Adams NM (2002) A probabilistic nearest neighbour method for statistical pattern recognition. J R Stat Soc 64(2):295–306
Huang CL, Chen MC, Wang CJ (2007) Credit scoring with a data mining approach based on support vector machines. Exp Syst Appl 33(4):847–856
Huang J, Wang H (2017) A data analytics framework for key financial factors. J Modell Manag 12(2):178–189
Islam MJ, Wu QMJ, Ahmadi M, Sid-Ahmed MA (2007) Investigating the performance of naive-bayes classifiers and k-nearest neighbor classifiers. In: 2007 International Conference on Convergence Information Technology (ICCIT 2007). IEEE. pp. 1541–1546
Jae Kim K, Ahn H (2012) A corporate credit rating model using multi-class support vector machines with an ordinal pairwise partitioning approach. Comput Oper Res 39:1800–1811
Jin H, Ling CX (2005) Using AUC and accuracy in evaluating learning algorithms. IEEE Trans Knowl Data Eng 17(3):299–310
Jorion P (2003) Financial risk manager handbook, 2nd edn. John Wiley & Sons
Kleimeier DS (2007) Credit scoring model for Vietnam's retail banking market. Int Rev Financ Anal 16(5):471–495
Kong Y, Yan A (2017) Qsar models for predicting the bioactivity of polo-like kinase 1 inhibitors. Chemomet Intell Lab Syst 167:214–225
Krivorotov G (2023) Machine learning-based profit modeling for credit card under writing implications for credit risk. J Bank Financ 149:106785
Laha A (2007) Building contextual classifiers by integrating fuzzy rule based classification technique and k-nn method for credit scoring. Adv Eng Informatics 21:281–291
Leo M, Sharma S, Maddulety K (2019) Machine learning in banking risk management: a literature review. Risks 7(1):1–22
Lessmann S, Baesens B, Seow H-V, Thomas LC (2015) Benchmarking state-of-the-art classification algorithms for credit scoring: an update of research. Eur J Oper Resh 247(1):124–136
Malhotra R, Malhotra D (2003) Evaluating consumer loans using neural networks. Omega 31(2):83–96
Malhotra RMK (2002) Differentiating between good credits and bad credits using neuro-fuzzy systems. Eur J Oper Res 136(2):190–201
Powers DMW (2011) Evaluation: from precision, recall and f-measure to roc, informedness, markedness correlation. J Mach Learn Technol 2(1):37–63
Rayo Cantón S, Lara Rubio J, Camino Blasco D (2010) A credit scoring model for institutions of microfinance under the basel ii normative. J Econ Financ Adm Sci 15(28):89–124
Rendle S (2010) Factorization machines. 2010 IEEE International Conference on Data Mining. IEEE, pp. 995–1000
Rosenberg E, Gleit A (1994) Quantitative methods in credit management: a survey. Oper Res 42(4):589–613
Schebesch KB, Stecking R (2005a) Support vector machines for classifying and describing credit applicants: detecting typical and critical regions. J Oper Res Soc 56:1082–1088
Schebesch KB, Stecking R (2005b) Support vector machines for credit scoring: extension to non standard cases. In: Innovations in classification, data science, and information systems. pp. 498–505
Thomas L, Edelman D, Crook J (2002) Credit scoring and its applications. Society of Industrial and Applied Mathematics

Thomas LC (2000) A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. Int J Forcast 16(2):149–172
Thomas LC, Oliver RW, Hand DJ (2005) A survey of the issues in consumer credit modelling research. J Oper Res Soc 56(9):1006–1015
Tom F (2006) An introduction to roc analysis. Pattern Recognit Lett 27(8):861–874
Vapnik V (1998) The support vector method of function estimation. Springer US, Boston, MA. pp. 55–85
Vapnik VN (1995) The nature of statistical learning theory. Springer
Wang Y, Wang S, Lai KK (2005) A new fuzzy support vector machine to evaluate credit risk. IEEE Trans Fuzzy Syst 13(6):820–831
West D (2000) Neural network credit scoring models. Comput Oper Res 27:1131–1152
Yu L, Wang S, Lai KK (2008) Credit risk assessment with a multistage neural network ensemble learning approach. Expert Syst Appl 34(2):1434–1444
Zhou L, Lai KK, Yu L (2010) Least squares support vector machines ensemble models for credit scoring. Exp Syst Appl 37(1):127–133

## Acknowledgements

This work was partly supported by the Natural Science Foundation of Chongqing (No.cstc 2021jcyj-msxmX0388; No.cstb2023nscq-msx0374), by the Chongqing Education Commission Humanities and Social Sciences Research Project for 2022(Project Number: 22SKG H321), and Science Foundation of Chongqing Education Commission (Grant KJQN202101125).

## Author contributions

JQ played a pivotal role in developing the research idea, conceptualizing the study, and writing the manuscript. He also contributed to the interpretation of research findings, results, and the overall structure of the article. XS took charge of coding and implementing the factorization machine model, as well as conducting essential data preprocessing tasks. Both JQ and XS actively engaged in discussions, and reviewed and edited the manuscript, collectively enhancing its clarity, coherence, and overall quality.

## Competing interests

The authors declare no competing interests.

## Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

## Informed consent

No informed consent was needed as this study did not include human subjects.

## Additional information

**Correspondence** and requests for materials should be addressed to Jing Quan.

**Reprints and permission information** is available at http://www.nature.com/reprints

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.