

RESEARCH

Open Access



# Two-stage credit scoring using Bayesian approach

Sunghyon Kyeong<sup>1</sup> and Jinho Shin<sup>2\*</sup>

\*Correspondence:  
william.shin@lab.kakaobank.com

<sup>1</sup> Division of Big-Data Analytics, Kakaobank, 9F, 131, Bundangnaegok-Ro, Bundang-Gu, Sungnam-Si, Kyeonggi-Do, Korea

<sup>2</sup> Division of Research and Development, Kakaobank, 11F, 131, Bundangnaegok-Ro, Bundang-Gu, Sungnam-Si, Kyeonggi-Do, Korea

## Abstract

Commercial banks are required to explain the credit evaluation results to their customers. Therefore, banks attempt to improve the performance of their credit scoring models while ensuring the interpretability of the results. However, there is a tradeoff between the logistic regression model and machine learning-based techniques regarding interpretability and model performance because machine learning-based models are a black box. To deal with the tradeoff, in this study, we present a two-stage logistic regression method based on the Bayesian approach. In the first stage, we generate the derivative variables by linearly combining the original features with their explanatory powers based on the Bayesian inference. The second stage involves developing a credit scoring model through logistic regression using these derivative variables. Through this process, the explanatory power of a large number of original features can be utilized for default prediction, and the use of logistic regression maintains the model's interpretability. In the empirical analysis, the independent sample t-test reveals that our proposed approach significantly improves the model's performance compared to that based on the conventional single-stage approach, i.e., the baseline model. The Kolmogorov–Smirnov statistics show a 3.42 percentage points (%p) increase, and the area under the receiver operating characteristic shows a 2.61%p increase. Given that our two-stage modeling approach has the advantages of interpretability and enhanced performance of the credit scoring model, our proposed method is essential for those in charge of banking who must explain credit evaluation results and find ways to improve the performance of credit scoring models.

**Keywords:** Two-stage logistic regression, Credit scoring model, Bayesian approach, Machine learning

## Introduction

The primary business of commercial banks involves providing credit loans to individuals. This is one of essential tasks for measuring the default risk of loan applicants as accurately as possible [1]. To perform this task effectively, it is necessary to utilize a credit scoring model, which is usually developed through logistic regression analysis because logistic regression deals with the binary dependent variable and achieves satisfactory performance while using a small number of parameters [2]. In addition, regression analysis reveals a linear relationship between a dependent variable and explanatory variables. Therefore, it is intuitive, easy to understand, and can be used to perfectly respond

to customer requests to explain evaluation results [3]. A perfect explanation is a significant advantage of logistic regression analysis in practice. However, although machine learning-based models are currently attracting attention as credit scoring techniques for significantly improving model performance, their interpretability regarding the credit evaluation results has a few limitations because they are black box models in which there is difficulty determining the relationship between the explanatory variable and dependent variable [4]. To overcome this limitation, studies on improving the interpretability of machine learning-based credit rating models have emerged [5–9].

Machine learning-based models, which have been widely distributed because of recent technological advances, are currently being used increasingly to improve the performance of credit scoring models. The excellent performance of a machine learning-based model is prominent in areas where machine learning algorithms are suitable such as image recognition [10]. In the field of credit evaluation, it was determined that for several reasons, the predictive power of default risk was better than that of the logistic model, e.g., the assumption that the explanatory variables of logistic regression should be linearly independent [8, 11, 12]. Generally, approximately 10 explanatory variables must be finally adopted [3], thereby constraining the improvement of predictive performance. For example, even a newly discovered variable in a new information domain cannot be used as an explanatory variable if it has a considerable linear dependency on other variables. However, machine learning techniques can use many features because explanatory variables have no restrictions, which is one of the reasons for improving the credit scoring model performance. However, machine learning-based models often show improved performance, but in many cases, the improvement results from overfitting [13]. Therefore, considerable attention must be given to the overfitting issue when developing a machine learning-based model.

A practical benefit of the logistic model is that the process of variables changing the scoring result can be explicitly explained; thus, financial institutions still develop their main credit scoring models using logistic regression [4]. Therefore, how can the performance of the model be further improved while maintaining perfect interpretability? In this study, we present a two-stage logistic regression model to improve predictive performance by allowing additional features to be used as explanatory variables. In this method, the first step involves using the Bayesian approach to extract the explanatory power regarding the dependent variable, i.e., the default, contained in the features, after which we create a derivative variable by linearly combining the features and the extracted explanatory powers. In the second step, we develop the final credit scoring model using the derivative variables as explanatory variables.

In contrast to the previous studies, this study proposes a method for improving the performance of credit rating models by combining the Bayesian inference and two-step modeling approaches. Few previous studies analyzed whether Bayesian methods are effective in improving the performance of credit rating models. For example, the Bayesian approach showed higher predictive power than the standard maximum likelihood estimation approach in the bank's default prediction [14]. In addition, some studies show that the Naïve Bayesian algorithm or the Bayesian algorithm using a neural network can improve the performance of credit rating models [15, 16]. Furthermore, studies aimed at improving the interpretability and the performance of models using Bayesian techniques

have been conducted. For example, the Bayesian behavior scoring model helps to identify factors that reflect customers' behavior and affect default probability [17].

However, there is limited literature on two-stage credit scoring modeling. For example, as a first step, the multivariate adaptive regression splines (MARS) method was applied to identify significant variables [18]. As a second step, these variables were used as the input nodes of the neural network model. After applying the mortgage data, the performance was improved compared to that of conventional logistic regression and neural networks. Similarly, irrelevant and noisy features were eliminated in the first stage because these features would lower model performance. Subsequently, a neural network model was employed in the second stage of categorizing credit applicants [19]. In addition, principal component analysis was applied as a first step, followed by developing a model with logistic regression in the second step to consider the interactions between explanatory variables in regression analysis [3]. The performance of a credit scoring model was also improved by developing a two-stage additive model using a machine learning technique in the first stage and logistic regression in the second stage while simultaneously increasing the interpretability of the model's prediction results. However, the interpretability of these models has a few limitations because the part using machine learning still remains a black box [5, 20].

The novelty of this paper lies in filling the gap between research on improving the credit scoring performance and interpretability of credit evaluation results. This study presents a method using logistic regression analysis in two steps to provide excellent performance and interpretability of the model. Additionally, we present the rationale of the proposed method using the Bayesian approach. To the best of our knowledge, there is no study suggesting a method for developing a credit scoring model by combining a logistic regression analysis in two steps based on the Bayesian approach.

For empirical analysis, to verify whether our proposed method significantly improves the performance of commercial banks' credit rating models, we analyze the credit loan data of KakaoBank, a prominent internet bank in Korea with the biggest market share as of March 2022. The bank has a deposit balance of KRW 33 trillion and 15 million monthly active users, and thus, it is ranked first among all Korean financial institutions.

This study has academic and practical importance as it presents a logistic regression-based credit modeling methodology. Because our proposed model is an extension of the logistic regression approach that banks generally use for developing credit rating models, it is easy for practitioners to understand and use. This study also suggests a new method for efficiently improving the performance of credit rating models while ensuring perfect interpretability.

The remainder of this paper is organized as follows: In "[Related works](#)" section, we reviewed related works. In "[Two-stage model using Bayesian approach](#)" section, we present our theoretical two-stage model based on Bayesian inference. In "[Empirical analysis](#)" section, we build the baseline and two-stage models using KakaoBank data. We discuss the results in "[Discussion](#)" section and present the conclusions in "[Conclusion](#)" section.

## Related works

The research on improving credit scoring performance is divided into two parts: the utilization of new significant features and improvement of prediction algorithms.

### Utilization of big data and data mining techniques for credit scoring

The discovery of input features for credit scoring is made by finding or deriving data that contains information about the customer's default behavior. For developing a credit scoring model, past financial transaction and demographic data are generally used, and digital footprints have recently attracted attention [21]. For example, various activity records left by customers on a mobile app or internet homepage of a financial institution are beginning to be used as important features for credit evaluation [22]. Consistent with this research direction, this study also used system log data that records customers' activities on the mobile banking application.

Studies to improve the credit scoring performance by introducing significant input features using data mining techniques are ongoing. For example, data mining techniques, such as linear discriminant analysis, backpropagation neural networks, and support vector machine methods, have been applied to analyze the credit scoring performance for the Taiwanese credit card portfolio [23]. Additionally, there is a study using the genetic algorithm for feature selection to develop a credit card fraud prediction model [24].

### Modeling algorithms for credit scoring

The traditional method for developing a credit scoring model is logistic regression analysis [25, 26]. This method provides a good predictive value of the likelihood of default. Moreover, because it reveals how specific input features affect the scoring results, the credit evaluation results can be completely interpreted, which is an irreplaceable advantage. Therefore, logistic regression analysis is widely used in practice by financial institutions that have an obligation to explain the customer's credit evaluation results in an easy-to-understand manner [25, 26].

However, the machine learning algorithms, which are developing rapidly in recent years, show improved performance compared to that of logistic regression analysis in the field of credit scoring. For example, an adaptive neuro-fuzzy inference system and binary classifiers based on machine learning and deep learning were applied to develop credit scoring models [27, 28]. Additionally, neural networks have been applied for credit card fraud detection and several marketing tasks [29, 30]. To consider the time-series characteristics of input features, the long short-term memory algorithm has been used to develop both credit card delinquency prediction and fraud detection models [8, 9]. In addition, various machine learning algorithms, such as decision tree, support vector machine, random forest, and genetic algorithm, are widely used in fraud detection modeling [31–33].

In this study, we created a baseline model using machine learning algorithms for comparing the performance of our proposed model.

### Two-stage modeling for credit scoring

A few studies have been conducted on the two-stage hybrid credit scoring methodology, and most of these studies have consisted of a machine learning algorithm as the first stage and a different methodology as the second stage. For example, the two-stage credit scoring model to combine artificial neural networks and multivariate adaptive regression splines showed better performance than the credit scoring models using discriminant analysis, logistic regression, artificial neural networks, and MARS solely [18]. There is also a study that presented a two-stage credit scoring model by combining widely used classification models such as extreme gradient boosting, gradient boosting decision tree, support vector machine, random forest, and linear discriminant analysis [34]. Similarly, there is a study combining the use of an artificial neural network in the first stage and conditional inference using the case-based reasoning in the second stage [35].

However, rather than simply combining two models as in the aforementioned studies, recent studies have been conducted regarding extracting features particularly relevant to the target variable in the first stage and enhancing the prediction performance by using the explanatory power of the extracted information in the second stage. Specifically, irrelevant and noisy features are removed in the first stage, and then, the remaining significant features are combined with neural networks in the second stage [19].

The hybrid models in these previous studies are similar to our proposed model, which extracts features closely covariate with the target variable in the first stage and then predicts the default probability based on the significant features in the second stage. However, our study is differentiated from the previous studies because we suggest a robust approach to extracting information related to the target variable by introducing the Bayesian framework. The Bayesian approach utilizes significant information contained in data and is widely used in social sciences such as financial economics. Specifically, rationality in economics is often assumed to be Bayesian optimization that makes optimal decisions based on observed information [36]. To the best of our knowledge, this study is the first to present a two-stage credit scoring model using significant information extracted from original features by the Bayesian approach.

### Explainable artificial intelligence (AI) for credit scoring

As in the aforementioned section, machine learning algorithms have an advantage in improving the performance of credit scoring. However, compared to the traditional logistic regression analysis, the insufficient interpretability of the credit evaluation results is a significant limitation. To overcome the limitation, some studies have attempted to develop an explainable AI model for credit scoring. For example, combining extreme gradient boosting with a framework to provide various viewpoints of explanation would provide some level of interpretability [37]. Additionally, previous studies suggested a method to provide some degree of interpretability by using logistic regression analysis in the second stage, even if the machine learning technique was used in the first stage [3, 5, 38]. However, because these studies used principal

component analysis or machine learning techniques in the first stage, the effects of the input features on the final credit evaluation results remain black boxes. This would be a severe limitation for financial institutions that are obligated to explain the credit evaluation results to customers.

To the best of our knowledge, there is no study suggesting how to develop a credit scoring model that achieves both excellent performance similar to machine learning algorithms and perfect interpretability. In this study, by using logistic regression analysis in each of the two stages, our proposed model showed not only an enhanced performance by utilizing many more input features than conventional single-stage regression analysis but also perfect interpretability of credit evaluation results by revealing the influence of input features on the prediction completely.

### Contributions of our study

We found that many of the recent studies on credit scoring focus on improving credit scoring performance using machine learning algorithms or big data, and some studies try to compensate for the weak interpretability of machine learning with a two-step modeling method. However, to the best of our knowledge, there is no study suggesting how to develop a credit scoring model that achieves both excellent performance similar to machine learning algorithms and perfect interpretability.

Reflecting on these aspects, this study made three contributions. First, we propose a novel two-stage modeling structure using logistic regression analysis in each of the two stages. Our proposed model shows an enhanced performance by utilizing many more input features than conventional single-stage regression analysis. Second, in contrast to previous works, our proposed model provides perfect interpretability of credit evaluation results by revealing the influence of input features on the prediction completely. Third, in contrast to previous studies, where a rationale was not provided for the two-stage model, we introduce a theoretical model based on the Bayesian framework for our two-stage model to utilize significant information from many more features than a conventional one-stage model.

### Two-stage model using Bayesian approach

In this section, we discuss the process in which we developed our two-stage logistic regression model by extracting the explanatory power of the features using a Bayesian approach.

#### First stage model and information weights

Generally, when a financial institution develops a credit scoring model, it selects highly significant features for each information domain and uses them as the explanatory variables. The standard logistic function about the probability of default, i.e.,  $p(\text{default}) : \mathbb{R} \rightarrow (0, 1)$ , is defined as follows:

$$p(\text{default}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X)}}, \quad (1)$$

where  $\beta_0$  is a constant, and  $X$  and  $\beta_1$  are vectors of input features and their coefficients, respectively. By dividing both sides of the above equation by  $(1 - p(\text{default}))$  and taking the logarithm, the following equation can be obtained:

$$\ln\left(\frac{p(\text{default})}{1 - p(\text{default})}\right) = \beta_0 + \beta_1 X \tag{2}$$

The left-hand side of Eq. (2) is called 'log odds', and here, we denote it as  $Y$ . Estimating the above equation is to obtain the conditional expectation of  $Y$ , i.e.,  $E(Y|X)$ , using the feature vector  $X$ . The advantage of using Eq. (2) is that it can express the conditional expectation in a linear form that is easier to understand than Eq. (1). Therefore, Bayesian inference can be directly applied to Eq. (2). Particularly, we use the following linear form, the same as Eq. (2), for convenience in this section:

$$E(Y|X) = \beta_0 + \beta_1 X, \tag{3}$$

where  $Y = \ln\frac{p(\text{default})}{1-p(\text{default})}$ .

In the first stage, we extract the default-related information contained in each feature belonging to a specific information domain, and then we create a derivative variable by combining the features using this extracted information as a weight. Using the derivative variable may predict default behavior more accurately than using only the most significant feature.

The Bayesian inference is utilized in the first stage. Particularly, we calculate the weight using conditional expectations based on the Bayesian framework. We assume that a feature ( $s_{ki}$ ) belonging to a specific information domain ( $k$ ) is a type of a noisy signal, and it comprises the default-related information ( $f_{ki}$ ) and other noises ( $e_i$ ) as shown below [36]:

$$s_{ki} = f_{ki} + e_{ki}, \tag{4}$$

where  $e_{ki} \sim iid(\mu_k, \sigma_k^2)$ .

If random variables  $X$  and  $Y$  follow a jointly normal distribution and their standard deviations and the correlation coefficient are denoted as  $\sigma_X, \sigma_Y$ , and  $\rho$ , the conditional expectation on  $Y$  given  $X$  is as follows:

$$E(Y|X) = E(Y) + \rho \frac{\sigma_Y}{\sigma_X} \{X - E(X)\}. \tag{5}$$

Equation (5) can be written as follows by reflecting the fact that  $Cov(X, Y) = \rho\sigma_X\sigma_Y$ , where  $Cov$  represents covariance function:

$$E(Y|X) = E(Y) + \frac{Cov(X, Y)}{\sigma_X^2} \{X - E(X)\}. \tag{6}$$

Therefore, we define the first-stage model as a conditional default prediction model that uses the features ( $s_{ki}$ ) belonging to an information domain  $k$  as follows:

$$E(Y|s_{k1}, \dots, s_{kI}) = E(Y) + \sum_{i=1}^I w_{ki} s_{ki}, \tag{7}$$

where  $w_{ki} = \frac{Cov(Y, s_{ki})}{Var(s_{ki})}$ .  $Var$  represents variance function.  $E(Y)$  is an unconditional mean, and  $I$  represents the number of features belonging to information domain  $k$ . We call these coefficients ( $w_{ki}$  s) information weights, which can be expressed as follows:

$$w_{ki} = \frac{Cov(Y, s_{ki})}{Var(s_{ki})} = \frac{Cov(Y, f_{ki} + e_{ki})}{Var(f_{ki} + e_{ki})} = \frac{Cov(Y, f_{ki}) + Cov(Y, e_{ki})}{Var(f_{ki}) + Var(e_{ki})} = \frac{Cov(Y, f_{ki})}{Var(f_{ki}) + \sigma_k^2} \tag{8}$$

From the equation presented above, we determine two characteristics of information weights. First, the form of the information weight is the same as the estimated coefficient of linear regression because the numerator is the covariance between the dependent variable and a feature, and the denominator is a feature’s variance. Therefore, the information weight represents the explanatory power of a feature. Second, the larger the variance ( $\sigma_k^2$ ) of the noise, i.e., the lower the reliability of the default-related information, the smaller the information weight for the corresponding feature. Thus, the feature contributes less to default prediction.

**Second stage model**

The second-stage model uses the derivative variable ( $S_k$ ) generated from information domain  $k$ , which is a linear combination of the features and their information weights obtained in the first stage. The second-stage model is expressed as follows:

$$E(Y|S_1, S_2, \dots, S_K) = b_0 + \sum_{k=1}^K b_k S_k, \tag{9}$$

where  $S_k = \sum_{i=1}^I \widehat{w}_{ki} s_{ki}$ , which is a derivative variable obtained by linearly combining the original features ( $s_{ki}$ ) and their estimated coefficients ( $\widehat{w}_{ki}$ ), i.e., information weight in the first-stage model, and  $b_k = \frac{Cov(Y, S_k)}{Var(S_k)}$  is the covariance between the target variable and the derivative variable divided by the variance of the derivative variable, which is the same as the definition of the estimated coefficient by regression analysis.

The derivative variables would improve the predictive power of the model. For a single-stage logistic regression model using the original features, the number of selected features is generally approximately 10–15 because of the assumption of linear independence among the features. However, when developed as a two-stage model, each derivative variable comprises several features with linear independence and high statistical significance. Therefore, the final model can utilize the predictive information involving many more features.

The coefficient of the second-stage model can be expressed as follows:

$$b_k = \frac{Cov(Y, S_k)}{Var(S_k)} = \frac{Cov\left(Y, \sum_{i=1}^I \widehat{w}_{ki} s_{ki}\right)}{Var\left(\sum_{i=1}^I \widehat{w}_{ki} s_{ki}\right)} = \frac{\sum_{i=1}^I \widehat{w}_{ki} Cov(Y, s_{ki})}{\sum_{i=1}^I \widehat{w}_{ki}^2 Var(s_{ki}) + 2 \sum_{i=1}^I \sum_{j=1(i \neq j)}^I \widehat{w}_{ki} \widehat{w}_{kj} Cov(s_{ki}, s_{kj})} \tag{10}$$

The numerator is the sum of the covariances between the dependent variable and the original features. If  $S_k$  is composed of the features with higher covariance, i.e., high predictive power, with the dependent variable, the coefficient  $b_k$  of the corresponding derivative variable ( $S_k$ ) becomes larger, and thus, the variable would have a greater influence on the default prediction result. However, the denominator implies



that the coefficient decreases as the variance of the original features increases or as the covariances among the original features increase. Therefore, if the original features have larger noisy terms or higher interdependence, the derivative variable has a smaller coefficient, and consequently, its influence on default prediction is also reduced.

**Sensitivities of the original features**

The original features used in the first-stage model affect the final default probability prediction through two logistic regressions. Here, we introduce the sensitivities of the original features to the final default prediction in numerical form. As denoted,  $S_k = \sum_{i=1}^I \widehat{w}_{ki} s_{ki}$ , where an original feature ( $s_{ki}$ ) belongs to an information domain ( $k$ ). Let  $Z$  be  $Z = b_0 + \sum_{k=1}^K b_k S_k$  in the second-stage model. The final default probability is then calculated using logistic regression as follows:

$$p(Y|S_1, S_2, \dots, S_K) = \frac{e^Z}{1 + e^Z} = \left( e^{-Z} + 1 \right)^{-1}. \tag{11}$$

The sensitivity of the final default prediction with respect to the change in  $s_{ki}$  is in a closed form as follows:

$$\begin{aligned} \frac{\partial p}{\partial s_{ki}} &= \frac{\partial p}{\partial S_k} \times \frac{\partial S_k}{\partial s_{ki}} = -\left( e^{-Z} + 1 \right)^{-2} \left( -b_k e^{-Z} \right) \times \widehat{w}_{ki} \\ &= b_k \widehat{w}_{ki} \left( e^{-Z} + 1 \right)^{-2} e^{-Z} \end{aligned} \tag{12}$$

Using this equation, it is possible to understand the original features responsible for the calculation of or change in credit scoring results. We can obtain this formula because we only use logistic regression, which clearly reveals a functional relationship between the original features and a dependent variable, rather than the black box models such as machine learning algorithms.

**Empirical analysis**

**Datasets**

To develop the credit scoring model, we use KakaoBank’s system log data, which is the same dataset introduced in our previous study [22]. We prepare the training and test datasets for the development and out-of-sample validation of the credit scoring models. Each dataset comprises 100,000 randomly sampled unsecured loans booked during the third and fourth quarters of 2018. For the *binary dependent variable*, we define loans with an overdue payment of more than 60 days within 12 months after loan execution as bad and otherwise as good. The bad ratios of the training and test datasets are 1.28% and 1.24%, respectively.

We then create the original features by counting each event code from the system log data stored in the KakaoBank mobile application system over 6 months before the loan execution date. To generate the features for credit scoring models, we convert the counting-based numeric features into normalized weights of evidence (WoE), and then we apply the feature selection criteria as an information value (IV)  $\geq 0.2$ , which indicates

that the variable has at least a weak predictive power according to a credit scoring textbook [25]. Through this process, we obtained 66 unique event codes that survived as the candidate features. These features are categorized into ten information domains according to the types of user actions: registration, custom setting of the banking application, clicking the menu or tab, user authentication, transaction, management of account, selecting types and options of cards, login or logout, response to recommendations, and optical character recognition (OCR). We have obtained appropriate permission from the bank to use the datasets in this study, and a detailed description of these datasets can be found elsewhere [22].

### Development of the baseline model

To develop the baseline model, we use the conventional logistic regression because it is used widely to develop credit scoring models in practice. The following significant features remained in the baseline model after backward selection: changing the name of the account, balance change in the safe box, i.e., the parking deposit account, clicking the FAQ category in the guide tab, login, to-do card exposure count, OCR: automatic identification of customer identity, registration, and clicking the card use button of an on-demand account. Table 1 shows the estimated coefficients and significance levels of the fitted model. All the explanatory features are statistically significant and have positive signs because of WoE transformation.

### Two-stage model development

#### First-stage model and information weights

The first-stage logistic regression models, "first-stage models" are developed using candidate features associated with each category representing a specific information domain. As explained in "First stage model and information weights" section, these features are divided into ten information domains according to the types of user actions, such as registration, authentication, transaction, account activity, debit card, recommendation, OCR, menu/tab, login, and custom setting. Table 2 shows

**Table 1** Basic statistical properties of the features and the estimated results of the model

Variable	Basic Statistical Properties			Model Fit Results		
	Mean (SD)	Skewness	Kurtosis	C	Z-stat	P-value
Constant	–	–	–	–4.35	–150.67	<0.001
Change the name of account	–0.006 (0.125)	–7.68	62.46	1.00	2.81	0.005
Safe box balance change	–0.005 (0.109)	–3.36	9.40	1.05	3.22	0.001
click FAQ category in guide tab	–0.001 (0.049)	7.07	48.01	1.33	2.67	0.008
Login	–0.041 (0.286)	0.42	–0.81	0.86	8.27	<0.001
To-do card exposure count	–0.037 (0.280)	–0.43	–0.69	0.72	6.22	<0.001
Automatic recognition of customer identity	–0.005 (0.117)	–6.01	46.56	0.97	2.91	0.004
Registration	–0.007 (0.117)	2.95	6.71	1.32	6.32	<0.001
Click card use button of an on-demand account	–0.004 (0.085)	3.15	8.22	1.26	4.13	<0.001

C: coefficient; FAQ: frequently asked questions; OCR: optical character recognition; SD: standard deviation

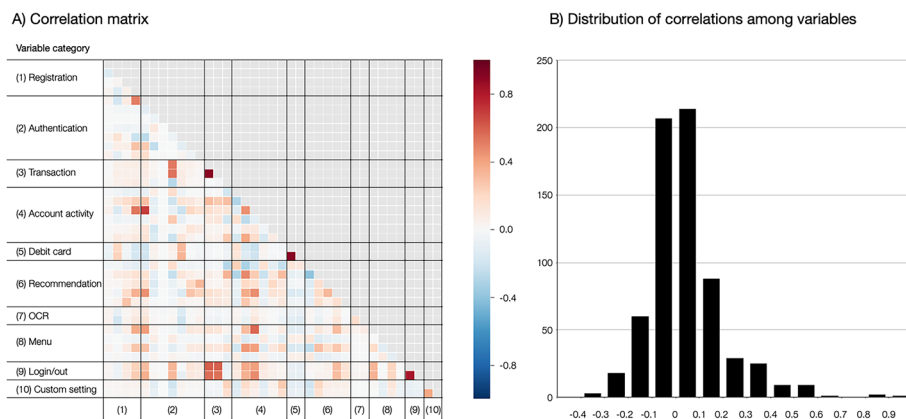
**Table 2** Basic statistical properties of the features and the estimation results of the first-stage models

Variable	Basic Statistical Properties			Model Fit Results		
	Mean (SD)	Skewness	Kurtosis	C	Z-stat	P-value
Category: Registration						
Constant	–	–	–	–4.35	–153.34	<0.001
Customer registration: connecting a social network service	–0.001 (0.050)	–8.86	76.58	1.32	1.88	0.060
Customer registration: completed	–0.007 (0.117)	2.95	6.71	1.26	5.98	<0.001
Product registration: acquiring mobile marketing information	–0.004 (0.101)	–4.09	21.36	0.87	2.57	0.010
Product registration: click application form	–0.011 (0.150)	–0.30	–1.23	1.13	5.72	<0.001
Category: Authentication						
Constant	–	–	–	–4.35	–150.65	<0.001
Confirmation of identification	–0.002 (0.070)	0.90	–1.19	0.79	1.84	0.066
Confirmation of personal mobile phone number	–0.008 (0.114)	6.87	47.77	1.23	6.83	<0.001
Customer verification	–0.001 (0.049)	2.13	36.05	1.14	2.21	0.027
PIN authentication	–0.013 (0.147)	4.98	22.78	0.95	6.80	<0.001
Completion of enhanced verification	–0.003 (0.081)	1.92	1.69	1.45	4.20	<0.001
Application of mobile scraping	–0.007 (0.129)	–2.03	2.10	1.49	5.56	<0.001
Application of PC scraping	–0.052 (0.399)	–3.34	9.57	0.99	6.98	<0.001
Category: Transaction						
Constant	–	–	–	–4.35	–152.49	<0.001
Click the transfer request button	–0.039 (0.273)	0.80	2.50	0.66	2.87	0.004
Transfer completed	–0.037 (0.260)	1.11	3.48	0.43	1.83	0.067
Safe box balance change	–0.005 (0.109)	–3.36	9.40	1.20	3.84	0.000
Category: Account activity						
Constant	–	–	–	–4.35	–152.89	<0.001
Click card use button of an on-demand account	–0.004 (0.085)	3.15	8.22	1.73	5.47	<0.001
Select my account	–0.009 (0.134)	0.25	–1.64	0.78	2.94	0.003
Produce information inquiry	–0.009 (0.130)	1.44	0.06	0.61	2.69	0.007
Click transaction history search icon	0.000 (0.018)	3.66	11.40	2.90	1.86	0.063
Completion of monthly rental agreement information	–0.008 (0.150)	–5.74	33.16	0.83	2.93	0.003
Completed of the produce cancellation	–0.013 (0.174)	–1.99	2.09	0.86	4.13	<0.001
Category: Debit card						
Constant	–	–	–	–4.35	–154.09	<0.001
Debit card: select types of card	–0.008 (0.122)	1.77	1.14	0.72	1.44	0.150
Debit card: check card application details	–0.007 (0.116)	2.04	2.17	0.32	0.62	0.535
Category: Recommendation						
Constant	–	–	–	–4.35	–151.72	<0.001
Click ad banner	0.000 (0.021)	2.46	4.04	5.15	3.58	<0.001
Click login-popup banner	–0.006 (0.111)	–1.28	–0.14	0.69	2.25	0.024
To-do card click	–0.032 (0.293)	–3.05	7.57	0.74	4.67	<0.001
To-do card exposure count	–0.037 (0.280)	–0.43	–0.69	0.77	6.69	<0.001
Click app-push	–0.013 (0.171)	–2.55	7.46	0.76	3.66	<0.001
Category: OCR						
Constant	–	–	–	–4.35	–154.18	<0.001

**Table 2** (continued)

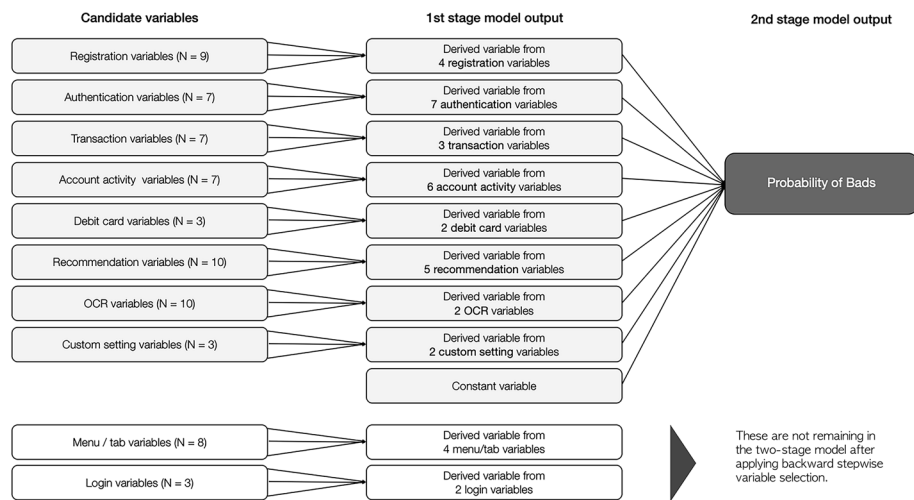
Variable	Basic Statistical Properties			Model Fit Results		
	Mean (SD)	Skewness	Kurtosis	C	Z-stat	P-value
Automatic recognition of customer identity	-0.005 (0.117)	-6.01	46.56	0.92	2.82	0.005
Taking photo of customer identification	-0.001 (0.053)	-3.94	13.56	0.63	1.06	0.288
Category: Menu						
Constant	--	-	-	-4.35	-153.56	<0.001
All tab: list menu	-0.012 (0.158)	0.04	-1.47	0.85	4.67	<0.001
All tab: profile area	-0.001 (0.051)	6.42	41.52	1.15	2.44	0.015
Click recommendation tab	-0.009 (0.137)	-1.48	5.37	0.90	3.91	<0.001
Guide tab: click FAQ category	-0.001 (0.049)	7.07	48.01	1.22	2.45	0.014
Category: Login/out						
Constant	--	-	-	-4.35	-152.75	<0.001
Login	-0.041 (0.286)	0.42	-0.81	0.69	4.00	<0.001
Run the application	-0.036 (0.262)	0.83	0.19	0.40	2.22	0.027
Category: Custom setting						
Constant	--	-	-	-4.35	-154.13	<0.001
Change the name of account	-0.006 (0.125)	-7.68	62.46	0.86	2.34	0.019
Change the color of account	-0.003 (0.089)	-9.43	86.92	0.63	1.27	0.205

C: coefficient; FAQ: frequently asked questions; OCR: optical character recognition; PC: personal computer; PIN: personal identification number



**Fig. 1** Correlation coefficients among features selected in the first stage models (A) and distribution of correlation values (B). The list of variables in the correlation matrix (A) is in the order of the variable names arranged in Table 2

the estimation results of the first-stage model. The features in each information domain are selected using the backward stepwise variable selection approach. Correlation analysis was performed on the selected variables (Fig. 1). Because most of the absolute correlation values were less than 0.5 (i.e., 98% among 66 feature pairs), almost all correlations values are not significant between the features that remained in the first stage model. The estimated coefficient represents the information weight for linearly combining the original features, as explained in "Related works" section. The combined variables constitute the explanatory variables for the second-stage model.



**Fig. 2** Schematic diagram of the two-stage logistic regression model

**Table 3** Estimation results of the second-stage model

Variable	Coefficient	Z-statistic	P-value
Constant	-17.2	-15.1	<0.001
Derivative variable from the registration category	1.8	3.3	0.001
Derivative variable from the authentication category	3.4	9.1	<0.001
Derivative variable from the transaction category	2.3	5.8	<0.001
Derivative variable from the account activity category	1.6	3.4	0.001
Derivative variable from the debit card category	5.3	6.1	<0.001
Derivative variable from the recommendation category	2.3	5.6	<0.001
Derivative variable from the OCR category	4.2	3.5	0.001
Derivative variable from the setting category	4.7	3.4	0.001

**Table 4** Correlation coefficients among the derivative variables

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Derivative variable from the registration category (1)	1.00							
Derivative variable from the authentication category (2)	0.14	1.00						
Derivative variable from the transaction category (3)	0.16	0.19	1.00					
Derivative variable from the account activity category (4)	0.26	0.13	0.27	1.00				
Derivative variable from the debit card category (5)	-0.03	0.06	-0.04	-0.05	1.00			
Derivative variable from the recommendation category (6)	0.30	0.33	0.19	0.33	-0.10	1.00		
Derivative variable from the OCR category (7)	0.01	0.01	-0.05	0.01	-0.10	0.06	1.00	
Derivative variable from the setting category (8)	0.06	-0.07	0.07	0.15	-0.11	0.07	0.00	1.00

**Second-stage model**

To develop the second-stage model, we use the derivative variable obtained from the first-stage model. This variable comprises an unconditional mean and the linear

combination using the information weights obtained through the first-stage model. Figure 2 shows a schematic diagram of the two-stage logistic regression model. We develop the second-stage logistic regression model called the “second-stage model” using the derivative variable for each first-stage model as the input variables.

Finally, the second-stage model comprises eight derivative variables obtained from the first-stage model, as presented in Table 3. This table shows that the derivative variables related to user actions appear statistically significant and have positive signs because of the WoE transformation. However, the two derivative variables regarding menu/tab and login do not remain in the fitted model after applying the backward stepwise variable selection method because of the insignificance of their coefficients.

Because user activities in the mobile application are interconnected, it is necessary to examine multicollinearity. We compute the correlation coefficients among the derivative variables fitted in the second-stage model, as presented in Table 4. The highest correlation coefficients appear to be 0.33 between authentication variables (2) and recommendation variables (6) and between account activity variables (4) and recommendation variables (6). Because the other correlations are less than 0.3, the correlations among the variables are determined to be low overall, thereby indicating that each derivative variable from the first-stage model has a unique explanatory power in the second-stage model.

To test the goodness of fit, we conduct the Hosmer–Lemeshow test for the second-stage model. We obtain a significantly small  $\chi^2$  statistic and significantly large  $P$ -value ( $\chi^2 = 0.123$ ;  $P$ -value = 0.872), which indicate that our model is fitted well [39].

**Comparison of models’ performance**

To evaluate the performance of the baseline and two-stage models, we use the Kolmogorov–Smirnov (K-S) statistics and the area under the receiver operating characteristic (AUROC) curve, which effectively measure and compare credit scoring models’ performances [11]. The AUROC curve evaluates the discriminatory power of a credit scoring model, which can be interpreted as the probability that “the Goods” receive better scores than “the Bads” [40]. Additionally, the K-S statistics measure the maximum difference between the two cumulative distributions of Goods and Bads. A higher K-S value indicates enhanced performance of the credit scoring model [40].

Table 5 shows that the K-S statistics and AUROC values of the two-stage model are 16.38% and 59.52%, respectively. Compared to the baseline model, the two-stage model shows significant improvements in credit scoring performance such as the K-S statistics by 3.42 percentage points (%p) and the AUROC by 2.61%p, respectively. According to the statistical test method proposed in the previous study [22], to evaluate whether these improvements in performance are statistically significant, we perform simulations

**Table 5** Credit scoring performance of the baseline and two-stage model

	Baseline model	Two-stage model	T-statistics	P-value
K-S statistics	12.96 (0.48)	16.38 (0.47)	72.02	< 0.0001
AUROC	56.91 (0.30)	59.52 (0.30)	87.05	< 0.0001

The numbers in parentheses represent standard deviations

**Table 6** Summary of the robustness test

	Single-stage approach (baseline model)	Two-stage approach (proposed model)	T-statistics	P-value
Alternative modeling techniques: Random Forest				
K-S statistics	12.94 (0.81)	14.88 (0.73)	17.73	<0.001
AUROC	57.47 (0.55)	58.18 (0.50)	9.52	<0.001
Alternative modeling techniques: Neural network				
K-S statistics	16.29 (0.63)	16.61 (0.61)	3.66	<0.001
AUROC	59.64 (0.40)	59.84 (0.41)	3.55	<0.001

to obtain the distributions of the K-S statistics and AUROC by iteratively extracting samples with the same size and developing the baseline and two-stage models. After 2000 iterations, we conduct the independent sample *t*-test to compare the differences in the sampled mean distributions of both the K-S statistics and AUROC between the two models. Consequently, the two-stage model showed significantly higher AUROC ( $P < 0.0001$ ) and K-S statistics ( $P < 0.0001$ ) values compared to the baseline model.

#### Robustness of the results

Although we developed a two-stage model using logistic regression with Bayesian-based formulation, the modeling algorithm could be replaced with other approaches. To verify the robustness of the two-stage modeling approach, we use machine learning algorithms, such as random forest and neural networks, in the first-stage model instead of logistic regression. Generally, these two algorithms perform well when applied in the development of credit scoring models [6, 7, 20, 41]. Table 6 shows that the two-stage model comprising first-stage modeling using random forest and neural network models, followed by second-stage modeling using logistic regression, shows significantly improved performance compared to the conventional single-stage machine learning-based model. The improvements in the K-S statistics and AUROC values are 1.94%p and 0.71%p for the random forest model and 0.32%p and 0.20%p for the neural network model with five hidden layers, respectively. Therefore, all the modeling techniques using a two-stage methodology enhance the performance of credit scoring models. Interestingly, neural network modeling in the first-stage model yields the best results. However, the performance of the final model is similar to that of our proposed model using logistic regression in the first stage.

#### Discussion

In this study, we propose a new method for utilizing the information contained in features by applying two-stage logistic regression based on the Bayesian approach instead of the conventional single-stage modeling approach. Previous studies on two-stage modeling attempted to improve the performance of models using machine learning as the first-stage modeling approach while maintaining interpretability to some extent using regression analysis as the second stage [5, 20]. In contrast, we used logistic regression for both stages in this study. The first reason is that it is sufficient for improving the model's

performance compared to other models utilizing machine learning techniques. Second, it is important from the practical perspective of financial institutions, which have an obligation to explain credit evaluation results to customers to ensure that they fully understand the influence of an original feature on credit scoring results.

**Improvement of the model's performance**

We created a baseline model using single-stage logistic regression, which is commonly used as a method for developing credit scoring models. We then developed the proposed model using two-stage logistic regression. As a result, the K-S statistics and the AUROC values of the proposed model were 16.38% and 59.52%, respectively, which is an improvement compared to the baseline model with 12.96% and 56.91%, respectively. Although the number of original features considered in both models was the same (66), the performance of the proposed model significantly improved by extracting correlated information with dependent variables through Bayesian inference in the first-stage model. Previous studies on two-stage modeling also reported the improved performance of credit scoring models [5, 20].

Contrary to our proposed model, they used machine learning techniques in the first stage. Therefore, we also examined the degree of performance improvement of the credit scoring model using machine learning techniques instead of logistic regression. As a result of verifying the final models' performance when the random forest and neural networks were applied in first-stage modeling, the models were significantly improved through the two-stage modeling approach in both cases. However, there is little difference in the performance of the models using various algorithms during two-stage modeling because the performance of the final models was found to be similar. Specifically, the AUROC values of the final model were 59.52% of our proposed model, 58.18% when using random forest, and 59.84% when using neural networks. However, the performance of the first single-stage models employing machine learning was higher than that of our proposed model. In particular, the neural networks showed significantly higher performance. The AUROC values of our first stage model, random forest, and neural network were 56.91%, 57.47%, and 59.64%, respectively. Consequently, when the two-stage modeling method is used, there is little difference in the performance depending on the modeling method.

Because the development process and datasets used in this study are different from the previous studies, there is a limitation in comparing the model performance with other studies. In particular, the bad ratio of the datasets of our study is less than 2%, whereas that of other studies [38] ranges from 22 to 48%, indicating that the binary class of our datasets is significantly imbalanced. Nevertheless, our two-stage model shows two

**Table 7** Comparison of performance improvements between the two-stage model proposed by this study and the hybrid two-phase model proposed by other study [38]

	Baseline Model	Proposed Model	Performance Improvement
This study	56.91%	59.52%	2.61%p
Other study	80.30%	80.89%	0.59%p

Average AUROC values are presented for the baseline and proposed two-stage models



advantages over the other type of two-stage hybrid model regarding performance and interpretability. First, Table 7 shows performance improvements between the two-stage logistic regression and the hybrid two-phase model composed of deep neural networks and logistic regression proposed by the other study [38]. The performance improvement of our two-stage model compared with the baseline model was 2.61%p whereas that of the hybrid two-phase model showed 0.59%p. Although we used highly imbalanced datasets, the performance improvement of our two-stage logistic regression model surpassed that of the existing model. Particularly, other studies of two-step hybrid models using deep learning or principal component analysis could not be included in Table 7 because they used performance indicators that were difficult to compare with the performance of our model [3, 5]. Second, because our model uses only logistic regression analysis, the functional relationship between the input features and the evaluation results is clearly revealed, making it possible to fully explain the credit evaluation results. However, because other models use deep learning algorithms, their interpretability is limited.

#### **Interpretability of the model**

Previous studies on two-stage credit scoring improved the performance of models using machine learning techniques in first-stage modeling [5, 20]. However, because machine learning models are a black box, it is not possible to establish exactly how the initial features affect the final credit scoring result. This would cause significant difficulties, particularly in practice, because banks are obligated to explain the credit evaluation results to their customers. When customers are interested in any change in their credit grades, they frequently ask the bank for an explanation of the reasons behind the change in their credit scores. For example, if a customer's credit score was bad in the past because of a history of delinquency but it has currently improved because of an increase in income, the bank must know exactly how these variables affect credit scoring results to explain why the credit score has changed. However, if the intermediate process of credit scoring is a black box, it is impossible to provide an exact and detailed explanation to the customer regarding the reason behind the change in the customer's credit evaluation result. This example helps us understand that the two-stage modeling approach using logistic regression employed in this study provides a practical advantage to financial institutions because it is a method that perfectly maintains the interpretability of the model.

#### **Financial inclusion and credit accessibility**

Our proposed method would be more effective in providing financial inclusion and credit accessibility rather than improving credit scoring using machine learning or big data. Previous literature reported that improving credit scoring performance provides financial inclusion for those with limited access to financial services, such as those who are young, have thin credit information, or in countries with constrained information sharing [22, 41]. However, if financial institutions obtain perfect interpretability and performance enhancement of a credit scoring model as our proposed algorithm, they would actively provide credit accessibility because the interpretability can be used to coach customers to improve their credit evaluation results. Financial coaching can effectively

contribute to credit improvement even for young people [42]. For example, the financial institution would advise the customer to correct their behaviors if the banking system identifies undesirable behaviors such as cherry-picking or abusing. Consequently, customers will get an improved credit evaluation by remedying their behaviors and use additional financial services such as unsecured loans.

## Conclusion

In this study, we established that our proposed two-stage logistic regression method demonstrated significantly better performance than conventional single-stage logistic regression analysis, which is used widely in developing credit scoring models. The conventional logistic model requires the assumption of linear independence between explanatory variables, and thus, only approximately 10–15 explanatory variables are used, whereas our proposed model can utilize many features through a two-stage modeling approach.

In the first stage of our proposed model, we extracted the explanatory power contained in the original features based on Bayesian inference for each information domain. We then created a new derivative variable by linearly combining the original features with their explanatory powers, which we called information weights. The second stage involved developing a credit scoring model using logistic regression with these derivative variables. Through this process, the explanatory power of numerous original features can be fully utilized for default prediction.

As a result of the empirical analysis, the performance of our proposed model significantly improved by 3.42%p for the K-S statistics and 2.61%p for the AUROC compared to the baseline model. Even when machine learning methods, such as random forest or neural networks, were used for the robustness analysis, the performance of the credit scoring model significantly improved through the proposed two-stage modeling approach.

Previous studies reported that using machine learning in the two-stage modeling approach improved the credit scoring model's performance compared to the conventional logistic regression method [5, 20]. We confirmed during the robustness analysis that these studies are consistent with the results of the conventional single-stage model. However, if we use the two-stage modeling approach that we suggested, the performance of logistic regression models does not differ much from that of the machine learning-based models. These results suggest that the two-stage logistic regression method based on the Bayesian framework is superior to machine learning techniques in terms of enhancing the model's performance while perfectly maintaining the interpretability of the credit scoring model. This implication is especially important for those in charge of banking who have a duty to explain credit evaluation results to customers and contemplate ways to improve the performance of credit scoring models.

## Abbreviations

MARS	Multivariate adaptive regression splines
AI	Artificial intelligence
Cov	Covariance function
Var	Variance function
WoE	Weights of evidence
IV	Information value

OCR	Optical character recognition
FAQ	Frequently asked questions
SD	Standard deviation
PC	Personal computer
PIN	Personal identification number
K-S	Kolmogorov–Smirnov statistics
AUROC	Area under the receiver operating characteristic

#### Acknowledgements

Not applicable.

#### Author contributions

SK designed the research outline and carried out empirical analysis and summarized analysis results. JS designed the research outline and the theoretical framework and interpreted analysis results. SK and JS participated in preparing the manuscript and finalizing this work. All authors read and approved the final manuscript.

#### Funding

No fund received.

#### Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

#### Declarations

##### Ethics approval and consent to participate

Not applicable.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare that they have no competing interests.

Received: 10 January 2022 Accepted: 20 October 2022

Published online: 14 November 2022

#### References

1. Khashei M, Mirahmadi A. A soft intelligent risk evaluation model for credit scoring classification. *Int J Financ Stud.* 2015;3:411–22.
2. Nurlybayeva K, Balakayeva G. Algorithmic scoring models. *Appl Math Sci.* 2013;7:571–86.
3. Walusala WS, Rimiru DR, Otieno DC. A hybrid machine learning approach for credit scoring using PCA and logistic regression. *Int J Comput.* 2017;27:84–102.
4. Dong G, Lai KK, Yen J. Credit scorecard based on logistic regression with random coefficients. *Procedia Comput Sci.* 2010;1:2463–8.
5. Chen C, Lin K, Rudin C, Shaposhnik Y, Wang S, Wang T. An interpretable model with globally consistent explanations for credit risk. *Comput Res Repos.* 2018;abs/1811.1. <http://dblp.uni-trier.de/db/journals/corr/corr1811.html#abs-1811-12615>
6. Dumitrescu E, Hué S, Hurlin C, Tokpavi S. Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects. *Eur J Oper Res.* 2022;297:1178–92.
7. Bussmann N, Giudici P, Marinelli D, Papenbrock J. Explainable machine learning in credit risk management. *Comput Econ.* 2021;57:203–16. <https://doi.org/10.1007/s10614-020-10042-0>.
8. Alaraj M, Abbod MF, Majdalawieh M. Modelling customers credit card behaviour using bidirectional LSTM neural networks. *J Big Data.* 2021;8:69. <https://doi.org/10.1186/s40537-021-00461-7>.
9. Benchaji I, Douzi S, El Ouahidi B, Jaafari J. Enhanced credit card fraud detection based on attention mechanism and LSTM deep model. *J Big Data.* 2021;8:151. <https://doi.org/10.1186/s40537-021-00541-8>.
10. Bishop CM. *Pattern recognition and machine learning.* New York: Springer; 2006.
11. Abdou HA, Pointon J. Credit scoring, statistical techniques and evaluation criteria: a review of the literature. *Intell Syst Acc Financ Manag.* 2011;18:59–88. <https://doi.org/10.1002/isaf.325>.
12. Gunnarsson BR, vanden Broucke S, Baesens B, Öskarsdóttir M, Lemahieu W. Deep learning for credit scoring: do or don't? *Eur J Oper Res.* 2021;295:292–305.
13. Genriha I, Voronova I. Methods for evaluating the creditworthiness of borrowers. *RTU Publ House.* 2012;22:42–9.
14. Löffler G, Posch PN, Schone C. Bayesian methods for improving credit scoring models. *SSRN.* 2005;
15. Chen H, Jiang M, Wang X. Bayesian ensemble assessment for credit scoring. *2017 4th Int Conf Ind Econ Syst Ind Secur Eng.* 2017;1–5.
16. Okesola OJ, Okokpuije KO, Adewale AA, John SN, Omoruyi O. An improved bank credit scoring model: a naïve Bayesian approach. *Int Conf Comput Sci Comput Intell.* 2017;2017:228–33.
17. Kao L-J, Lin F, Yu C-Y. Bayesian behavior scoring model. *J Data Sci.* 2013;11:433–50.

18. Lee T-S, Chen IF. A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Syst Appl.* 2005;28:743–52.
19. Tripathi D, Edla DR, Bablani A, Kupplli V. Two-stage credit scoring model based on evolutionary feature selection and ensemble neural networks. *Mach Learn Algorithms Appl.* 2021. <https://doi.org/10.1002/9781119769262.ch6>.
20. Munkhdalai L, Lee JY, Ryu KH. A hybrid credit scoring model using neural networks and logistic regression. *Adv Intell Inf Hiding Multimed Signal Process Smart Innov Syst Technol.* Singapore: Springer; 2019. p. 251–8.
21. Berg T, Burg V, Gombović A, Puri M. On the rise of FinTechs: credit scoring using digital footprints. *Rev Financ Stud.* 2020;33:2845–97. <https://doi.org/10.1093/rfs/hhz099>.
22. Kyeong S, Kim D, Shin J. Can system log data enhance the performance of credit scoring?—Evidence from an internet bank in Korea. *Sustainability.* 2022;14:130.
23. Hsieh H, Lee T, Lee T. Data mining in building behavioral scoring models. 2010 Int Conf Comput Intell Softw Eng. 2010. p. 1–4.
24. Illeberi E, Sun Y, Wang Z. A machine learning based credit card fraud detection using the GA algorithm for feature selection. *J Big Data.* 2022;9:24. <https://doi.org/10.1186/s40537-022-00573-8>.
25. Siddiqi N. *Credit risk scorecards: developing and implementing intelligent credit scoring.* Hoboken: Wiley; 2005.
26. Finlay S. *Credit scoring, response modelling and insurance rating.* London: Palgrave Macmillan; 2010.
27. Akkoç S. An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis: the case of Turkish credit card data. *Eur J Oper Res.* 2012;222:168–78.
28. Addo PM, Guegan D, Hassani B. Credit risk analysis using machine and deep learning models. *Risks.* 2018;6:38.
29. Alborzi M, Khanbabaei M. Using data mining and neural networks techniques to propose a new hybrid customer behaviour analysis and credit scoring model in banking services based on a developed RFM analysis method. *Int J Bus Inf Syst.* 2016;23:1–22. <https://doi.org/10.1504/IJBIS.2016.078020>.
30. Jurgovsky J, Granitzer M, Ziegler K, Calabretto S, Portier P-E, He-Guelton L, et al. Sequence classification for credit-card fraud detection. *Expert Syst Appl.* 2018;100:234–45.
31. Khare N, Sait SY. Credit card fraud detection using machine learning models and collating machine learning models. *Int J Pure Appl Math.* 2018;118:825–38.
32. Dornadula VN, Geetha S. Credit Card fraud detection using machine learning algorithms. *Procedia Comput Sci.* 2019;165:631–41.
33. Seera M, Lim CP, Kumar A, Dhamotharan L, Tan KH. An intelligent payment card fraud detection system. *Ann Oper Res.* 2021. <https://doi.org/10.1007/s10479-021-04149-2>.
34. Wei S, Yang D, Zhang W, Zhang S. A novel noise-adapted two-layer ensemble model for credit scoring based on backflow learning. *IEEE Access.* 2019;7:99217–30.
35. Chuang C-L, Huang S-T. A hybrid neural network approach for credit scoring. *Expert Syst.* 2011;28:185–96. <https://doi.org/10.1111/j.1468-0394.2010.00565.x>.
36. Daniel K, Hirshleifer D, Subrahmanyam A. Investor psychology and security market under- and overreactions. *J Financ.* 1998;53:1839–85. <https://doi.org/10.1111/0022-1082.00077>.
37. Demajo LM, Vella V, Dingli A. Explainable AI for interpretable credit scoring. 10th Int Conf Artif Intell Soft Comput Appl. London, United Kingdom; 2020. p. 3749. <https://ideas.repec.org/p/arx/papers/2012.03749.html%5C>
38. Munkhdalai L, Lee JY, Ryu KH. A hybrid credit scoring model using neural networks and logistic regression. In: Pan J-S, Li J, Tsai P-W, Jain LC, editors. *Adv Intell Inf Hiding Multimed Signal Process.* Singapore: Springer; 2020. p. 251–8.
39. Chi B-W, Hsu C-C. A hybrid approach to integrate genetic algorithm into dual scoring model in enhancing the performance of credit scoring model. *Expert Syst Appl.* 2012;39:2650–61.
40. Niu B, Ren J, Li X. Credit scoring using machine learning by combining social network information: evidence from peer-to-peer lending. *Information.* 2019;10:397.
41. Óskarsdóttir M, Bravo C, Sarraute C, Vanthienen J, Baesens B. The value of big data for credit scoring: Enhancing financial inclusion using mobile phone data and social network analytics. *Appl Soft Comput.* 2019;74:26–39.
42. Modestino AS, Sederberg R, Tuller L. Assessing the effectiveness of financial coaching: evidence from the Boston youth credit building initiative. *J Consum Aff.* 2019;53:1825–73. <https://doi.org/10.1111/joca.12265>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---