Journal of Organizational and End User Computing
Volume 34 • Issue 3

Research on the Risk of Social Stability of Enterprise Credit Supervision Mechanism Based on Big Data

Tao Meng, International Business College, Dongbei University of Finance and Economics, China
Qi Li, School of Business Administration, Dongbei University of Finance and Economics, China
Zheng Dong, School of Business Administration, Dongbei University of Finance and Economics, China
Feifei Zhao, School of Business Administration, Dongbei University of Finance and Economics, China

ABSTRACT

The study aims to establish a platform-based enterprise credit supervision mechanism and, combined with big data, accurately evaluate the credit assets of enterprises under the influence of social stability risk and improve the ability of enterprises to deal with risks. Using descriptive statistical methods, the study shows that most local enterprises exist in the form of micro loans, which promotes the development of local economy to a certain extent, but it is a vicious cycle of economic development. The overall prediction accuracy of the single enterprise risk assessment model under the influence of social stability risk is 65%. Compared with the single algorithm, the prediction accuracy of the integrated algorithm model is significantly improved, and the prediction accuracy can reach 83.5%; the standard deviation of data prediction is small, and the stability of the model is high.

KEYWORDS

Big Data Analysis, Credit Mechanism, Platform Enterprise, Social Stability Risks, Sustainable Consumption

1. INTRODUCTION

With the rapid development of economy, problems such as population expansion, excessive use of resources, and environmental pollution continue to emerge (Song 2019). With the increasing demand of human consumption, in order to meet the expanding consumption demand, there are predatory exploitation and destruction of natural resources, which have caused serious environmental pollution and ecological crisis (Zou 2019). Only when the inappropriate consumption of human beings is fully curbed, and people take the road of sustainable consumption, can the environment be protected and the ecological crisis can be gradually eliminated (Bengtsson 2018). At present, sustainable development has become a globally recognized economic development strategy (Barrow 2018). The two basic aspects of sustainable development are sustainable production and sustainable consumption (Lukman 2016). Among them, sustainable consumption means that contemporary people cannot exceed the limit of ecological environment carrying capacity when meeting the needs of their consumption development. Consumption should be conducive to environmental protection and ecological balance (Govindan 2018). It requires not only the optimal and sustainable utilization of resources, but also the minimum discharge of waste and the minimum pollution to the environment (Wang 2019). In order to realize the concept of sustainable development, the sharing economy model appears in the

DOI: 10.4018/JOEUC.289223

existing market. Sharing economy is a new economic model which optimizes resource allocation and efficient social governance. It is based on the Internet and other modern information technology support. The resource supplier will provide the temporarily idle resources to the resource demander with compensation through the technology platform. The demand side obtains the right to use resources, while the supply side gets the corresponding reward (Puschmann 2016). However, under the background of sharing economy, the alienation of platform social responsibility behavior has cast a shadow on the promotion of sustainable consumption, which is mainly reflected in the following aspects: the content of social responsibility of platform enterprises needs to be defined; the alienation of social responsibility needs to be addressed; the behavior of consumer dishonesty in the level of credit environment needs to be restricted; the overall credit environment needs to be improved; under the level of consumer rights and interests, the rights and interests of consumers in the sharing platform need to be protected; the governance system of social responsibility in platform enterprises needs to be established urgently (Vith 2019). Among them, the most important is to establish a new enterprise credit risk assessment model with high accuracy under the sharing economy (Richter 2017).

At present, there are few researches on enterprise credit risk assessment model, and most of them focus on the establishment of risk assessment system. Among them, established a risk index system by using the analytic hierarchy process (AHP) according to the characteristics of credit business process of small and medium-sized enterprises of commercial banks, and carried out risk analysis. On the basis of risk identification, the fuzzy evaluation model and judgment matrix were established to effectively reduce the enterprise credit risk (Shi 2016); in order to explore enterprise credit risk assessment, compared the application effect of several common neural network models in China's SME data sets, and found that probabilistic neural network has the minimum error rate, the highest area under curve (AUC) value, and has good robustness (Huang 2018); used the nonlinear least squares support vector machine (LS-SVM) model to analyze the credit risk indexes of supply chain enterprises, and combined with the index selection principle to determine the final index system, and constructed the enterprise online supply chain financial credit index. The results show that the classification accuracy of LS-SVM evaluation model is higher than that of Logistic regression model, and has strong generalization ability (Wang 2019); proposed an enterprise credit risk assessment method based on improved genetic algorithm, established a grid structure of longitude and latitude, improved and optimized longitude latitude grid genetic algorithm, and improved enterprise credit risk assessment. This method is better than traditional strategies and can be widely used (Yang 2020). It can be seen that for the analysis of enterprise credit rating, AHP method and traditional data analysis model are often used. This method has poor accuracy and is difficult to evaluate enterprise credit effectively.

Therefore, three kinds of single algorithm (Logistic regression model, decision tree and neural network model) are comprehensively compared. Moreover, the performance of single algorithm model is optimized by adjusting parameters and introducing cost matrix. In order to further improve the accuracy of the model prediction, the integrated algorithm and combined optimization algorithm model are constructed. Combined with the credit data analysis of 55 enterprises, the accuracy and stability of some models are compared and analyzed. Finally, combined with an example, the influence of enterprise credit on sustainable consumption is analyzed. This study can be used to effectively establish the platform enterprise credit mechanism and provide theoretical basis and practical value for sustainable consumption.

2. METHOD

2.1 Influencing Factors of Sustainable Consumption

Sustainable consumption is not only in line with the principle of intra generational justice, but also in line with the principle of intergenerational justice. It is a kind of consumption that ensures that human needs at all levels such as material consumption, spiritual consumption and ecological consumption

are met and constantly evolve from low level to high level (Wu 2016). The connotation of sustainable consumption can be concluded as follows. First, the concept of sustainable consumption is different from consumption in the sense of traditional economics. It is included in all aspects of consumption, such as waste discharge, sewage treatment and other ecological concepts. Second, the consumption subject of sustainable consumption includes not only contemporary people, but also future generations. It should not only meet the needs of contemporary people, but also cannot deprive future generations of their needs. Third, spiritual and cultural consumption has become an important part of sustainable consumption. With the improvement of the level of economic development, people's consumption level is also constantly improving, and material consumption has not become the whole of people's consumption. Fourth, sustainable consumption still aims at improving the quality of life, and does not deliberately require consumers to excessively control their own consumption in order to protect the environment and leave wealth resources for future generations (Vergragt 2016).

In short, sustainable consumption needs to grasp two key points. The first is consumption needs, that is, not only to meet the consumption needs of contemporary people, but also not harm the ability of future generations to meet their consumption needs; the second is the limit of consumption ecological environment. That is to say, once the tolerance limit of the earth's ecological environment and natural resources on which human beings rely for survival and development is broken, it will not only affect and restrict the survival and development of contemporary people, but also will endanger the survival and development ability of future generations. It can be seen that sustainable consumption is based on the concept of composite system, which not only considers the consumption demand of all aspects and levels of human beings, but also takes improving the quality of human life as the goal. At the same time, it also realizes the reasonable allocation and sustainable utilization of resources, and achieves the minimum pollution of the environment and the minimum discharge of waste. The influencce of sustainable consumption includes environmental factors, personal behavior control, enterprise credit, response capacity, environmental carrying capacity and consumption outcome cognition(Figure 1 shows Influencing factors of sustainable consumption). It further affects sustainable consumption behavior by influencing individual behavior network, or directly influencing sustainable consumption behavior (Geng 2017).

2.2 Logistic Regression

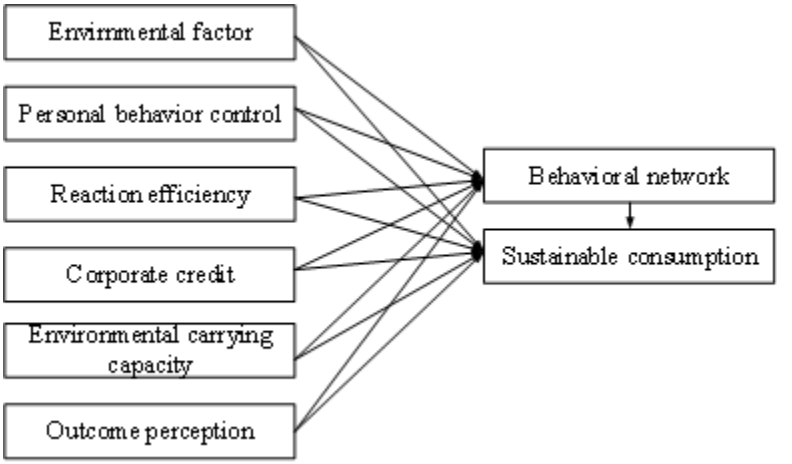
Logistic regression is a logarithmic probability model, one of the models of discrete choice method, belonging to the category of multivariate analysis, and is a common method for statistical empirical analysis such as sociology, biostatistics, clinical, quantitative psychology, econometrics, and marketing (Ranganathan 2017). The change trend of Logistic function image is smaller at both ends and larger in the middle. The value range of function value is (0,1). Figure 2 shows the specific structure. The function is defined as:

$$Logistic = y - \frac{1}{1+e^{-x}} \tag{1}$$

If there is a continuous reaction variable y_i^*, which indicates the possibility of the event, and its value range is $(-\infty, +\infty)$, when the value of the variable is greater than the critical value, it means that the event occurs. If the value is less than or equal to the critical value, it means that the event does not occur, that is:

$$\begin{cases} y_i = 1, y_i^* > 0 \\ y_i = 0, y_i^* \le 0 \end{cases} \tag{2}$$

Figure 1. Influencing factors of sustainable consumption



Where y_i is the observed real response variable. $y_i = 1$ indicates that the event occurs and $y_i = 0$ indicates that the event does not occur. It is assumed that there is a linear relationship between the explained variable y_i* and the explanatory variable x_i, that is:

$$y_i^* = \alpha + \beta x_i + \varepsilon \tag{3}$$

It can be further obtained that:

$$p(y_i = 1 \left| x_i \right.) = p[\alpha + \beta x_i + \varepsilon_i > 0] = p[\varepsilon_i > (-\alpha - \beta x_i)] \tag{4}$$

Assuming that random error ε_i obeys Logistic distribution, the cumulative distribution function can be obtained due to the symmetry of Logistic function distribution.

$$p(y_i = 1 \left| x_i \right.) = p[\varepsilon_i \le (\alpha + \beta x_i)] = F[\alpha + \beta x_i] \tag{5}$$

Where F is the cumulative distribution function of ε_i, and the function form depends on the distribution of ε_i. The standard Logistic function distribution is selected with the mean value of 0 and the standard deviation of $\pi^2 / 3$. Thus, the cumulative distribution function can be expressed in a simple form.

$$p(y_i = 1 \left| x_i \right.) = p[\varepsilon_i \le (\alpha + \beta x_i)] = \frac{1}{1 + e^{-(\alpha + \beta x_i)}} \tag{6}$$

When ε_i approaches positive infinity, function $p(y_i = 1 \left| x_i \right.)$ approaches 1. When ε_i approaches negative infinity, function $p(y_i = 1 \left| x_i \right.)$ approaches 0. It can be seen that the value range of any ε_i

is in the range of (0,1). When β is greater than 0, the function monotonically decreases, and when β is less than 0, the function monotonically increases. If $p_i = p(y_i = 1 \left| x_i \right.)$, the probability of occurrence of the event is:

$$p_i = \frac{1}{1 + e^{-(\alpha+\beta x_i)}} = \frac{e^{\alpha+\beta x_i}}{1 + e^{\alpha+\beta x_i}} \tag{7}$$

The probability that the event does not occur is as follows.

$$1\text{-}p_i = 1\text{-}\frac{1}{1 + e^{-(\alpha+\beta x_i)}} = \frac{1}{1 + e^{\alpha+\beta x_i}} \tag{8}$$

The probability ratio of event occurrence and non-occurrence is as follows.

$$\frac{1}{1\text{-}p_i} = e^{\alpha+\beta x_i} \tag{9}$$

For the convenience of calculation, it is transformed into a linear form.

$$\ln(\frac{1}{1\text{-}p_i}) = \alpha + \beta x_i \tag{10}$$

The probability ratio represents the ratio of the probability of an event occurring and not occurring. The greater the relative risk, the greater the probability of an event. This is a univariate Logistic model. For multiple independent variables, the Logistic model can be extended to multivariate logistic model.

$$p = \frac{1}{1 + e - (\beta_0 + \beta_1 x_1 + ... + \beta_n x_n)} \tag{11}$$

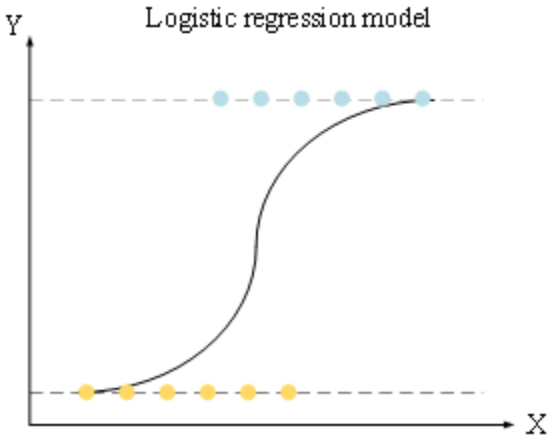
Where, p is the independent variable, taking (x1, x2, x3 xn). The linear form of multiple Logistic regression model is as follows.

$$Logistic(p) = \ln(\frac{p}{1-p}) = \beta_0 + \beta_1 x_1 + ... + \beta_n x_n \tag{12}$$

2.3 Decision Tree Algorithm

Decision tree algorithm is a method to approximate the value of discrete function. It is a typical classification method. In essence, decision tree is a process of classifying data through a series of rules (Tayefi 2017). The decision tree algorithm constructs a decision tree to find the classification rules contained in the data. Its structure is divided into two steps. The first step is the generation of decision tree, which is the process of generating decision tree from training data set. In the second step, the decision tree is pruned and modified. Figure 3 shows the structure.

Figure 2. Logistic regression model



The specific process of building decision tree is as follows. The first step is to regard all the training samples as a node; the second step is to select the best segmentation point, and compare each method of measuring the purity of variables to determine the segmentation standard and select the best segmentation point; the third step is to use the segmentation points selected in the second step, and the samples are divided into several sub nodes according to the attribute value, which are recorded as m_1, m_2,…m_3, respectively (k represents the number of attribute values of the node). The fourth step is that the first step and the second step are executed recursively on each sub node m_1,m_2… m_k until the purity of each node reaches a certain standard (Sarker 2020).

The decision tree can grow to the maximum extent possible. By selecting the required segmentation features, each sample is classified completely, so that the accuracy of the decision tree reaches 100%. However, if the decision tree growth is too large and too specific, the model will over fit the training data set, which will reduce the accuracy of the test data. Therefore, it is necessary to prune the generated decision tree to simplify the model, so that the model has better generalization ability (Yan 2016).

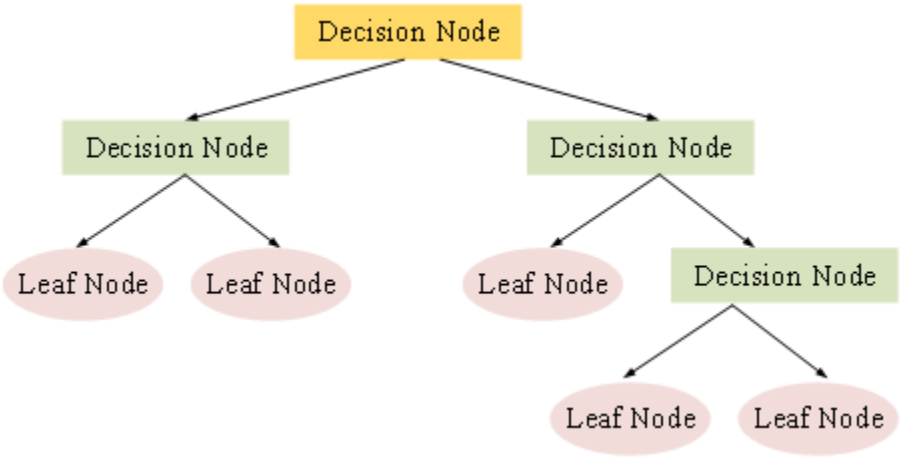
2.4 Neural Network Model

Artificial neural network (ANN) is a model established by simulating the actual neural network of human brain. There are nodes similar to human brain neurons in neural network, which are divided into input layer, hidden layer and output layer. The given values in the input layer are passed along the input layer, the hidden layer and the output layer in turn. Among them, Figure 4 shows the structure diagram of the neural network model (Van 2017).

In the neural network, after the input is given, it will pass along the arrow shown in the figure to reach the hidden layer. The hidden layer node will be activated according to the given input, and the activated hidden layer node will calculate the output value. Then, the calculation results are passed to the output layer. Next, the nodes of the input layer are activated. The activated node calculates the final output value and takes it as the final prediction result of the model (Carleo 2017).

$$net_j = \sum x_i wi_j + w_{ji0} \tag{13}$$

Figure 3. Structure of decision tree algorithm



The value passed to node j is called net activation and net_j is called the composite function. The value sent by the hidden layer node j to the output layer is the result of the net activation value processed by the activation function. One of the commonly used activation functions is sign function. According to different symbols, the output results are different. The other is the sigmoid function.

$$sign(net_j) = \begin{cases} 1, net_j \geq 0 \\ -1, net_j < 0 \end{cases} \tag{14}$$

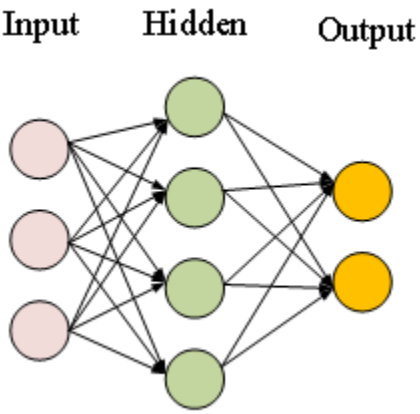
$$sigmoid(net_j) = \frac{1}{1 + e^{-x}} \tag{15}$$

The output value of sigmoid function is located in the interval (0,1), and the value of x is in a small interval. The output is sensitive to the input, and the function value changes greatly; the value of x is in a larger interval, the output is not sensitive to the change of input, and the function value changes little. The learning process of neural network is actually the process of adjusting the weighted value *w*. There is an input value *x* and an expected output value *y*, and the value is input to the input layer. Then, whether the output value of the model is consistent with the expected output value is judged. If it is different, the weighted value *w* should be adjusted appropriately.

There are two standards to reflect the difference between the model output value and the expected output value. One is sum Of Squared Error (SSE) and the other is information entropy. If the output value of node i in the output layer is $\overline{y_i}$ and the expected output value is y_i, the SSE equation is as follows:

$$SSE = \sum (y_i - \overline{y_i})^2 \tag{16}$$

Figure 4. Structure diagram of ANN neural network



If the difference is found, that is, the value of SSE is not zero, the weighted value is adjusted according to the reverse order (i.e. from the output node with difference to the hidden node connected with it, and then from the hidden node to the input node), so that the SSE equation is close to 0. After several adjustments, a more suitable weighting value is finally found. In the learning process of neural network model, if the weighted value is adjusted too appropriately for the given data, the model is prone to over fitting. In order to avoid over fitting phenomenon, the following methods are adopted. When the weighted value is corrected according to the back-propagation algorithm, the value in the range of (0,1) is multiplied by the weighted value, which is called "weight attenuation". The calculation equation is as follows:

$$w = w * (1 - \varepsilon) \tag{17}$$

2.5 Ensemble Learning Algorithm

Boosting is a family of algorithms that can promote weak learners to strong learners, and can be widely used in any machine learning algorithm. Its idea originated from Valiant's probably approxi mate correct (PAC) learning model (Dorogush 2018). The basic idea of Boosting is to train a base learner from the initial training set, and adjust the distribution of training samples according to the performance of the base learner, so that the training samples that the previous base learner did wrong receive more attention in the follow-up. Then, the next base learner is trained based on the adjusted sample distribution; It is repeated until the number of base learners reaches the preset value T. Finally, the T-base learners are weighted.

Random forest model is an integrated learning model. The basic classifier is composed of decision trees. These decision trees are obtained by Bagging ensemble learning technology. The output of a single decision tree is voted to determine the final classification result of the random forest model, which is similar to the voting form of the diversity expert group meeting (Brokamp 2018). In the random forest model, only a group of classifiers with good classification ability need to be trained first and then integrated. In this way, the time for finding the single best classifier is saved, and the single classification preference will not be dominant, so the over fitting phenomenon of the model can be reduced (Deng 2020). Random forest has a good filtering effect on noise and outliers, and can overcome the problem of over fitting. Especially in the classification of high-dimensional data, it shows good parallelism and scalability. The random forest model, driven by data, obtains classification rules

by learning and training specified samples, and does not need any prior knowledge of classification. It is a nonparametric classification model (Sinha 2019).

After the k-th decision tree is constructed according to the above method, the process is repeated continuously, and the combination of K decision trees is established, so that the random forest is obtained. When the samples to be classified are input into the random forest, the output results are voted according to the output results of the K decision trees. The final classification result with more votes is regarded as the output of random forest. The process of training each decision tree is the process of generating random forest. The training process of each decision tree is independent of each other, so the generation process of random forest can be realized by parallel processing technology, which greatly improves the efficiency of the model.

2.6 Data Source and Processing

(1) Data sources: the research purpose is to establish an enterprise credit risk assessment model by using relevant algorithms, so as for the model data set, the data of enterprise related credit business should be selected. According to the relevant laws and regulations of China, the enterprise information is generally not released to the public, and the enterprise information of other national banks is not disclosed to the public. Therefore, in terms of data, German bank data set and Australian bank data set are widely used by domestic and foreign scholars. German bank data set is obtained from a German credit institution. Professor Hans Hofmann of Hamburg University published it in 1994 (Koch 2017; Karout 2016). The credit data set contains 1000 enterprise loan cases and 20 data indexes. The Bank of Australia data set contains 690 loan cases and 14 data indexes. In contrast, the German bank data set contains more cases and more data indexes, so German bank data set is selected for enterprise credit risk assessment research. The data set contains 700 normal customers and 300 default customers.
(2) Data processing: first, the "A +" in the original data is removed, such as the variable of checking account. Second, the classification variables with more category values are merged. According to the proportion of each category in the variable, the smaller categories are merged. The category values of seven variables including checking deposit account, loan credit history, loan purpose, savings account, enterprise establishment period, guarantee enterprise situation and installment payment are combined. Then, the variables with larger order of magnitude in numerical variables are processed. In the data, only the loan amount variable has a large order of magnitude. Its value is divided by 1000. Finally, numerical variables with less attribute values are changed into typed variables, that is, numerical variables are discretized. The variables of debt ratio, living time of current address, current number of bank loans and number of people to support are discretized (Wang 2018).

For the missing value (the original data set is empty), the error value (the error value in the original data set is marked with X), variables with outliers, the mean or mode of the corresponding variable is used to repair. For example, the missing value in the loan term variable is filled with the mean value of the variable, and the error value and missing value in the loan target variable are filled with the mode of the variable. It is mainly evaluated from the prediction accuracy (ACC). This value represents the proportion of processed samples correctly divided into positive samples (Jabbarian 2017).

$$ACC = \frac{\sum_{u \in U} \left| R(u) \cap T(U) \right|}{\sum_{u \in U} \left| R(u) \right|} \tag{18}$$

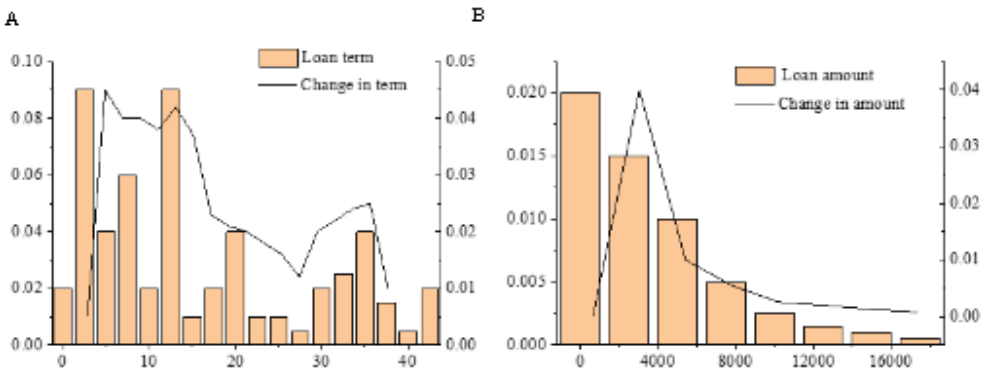
Among them, $R(u)$ is the correct forecast quantity, $T(U)$ is the actual quantity; $f(x^{(i)})$ is the correct forecast quantity, and $y^{(i)}$ is the wrong forecast quantity.

3. RESULTS AND DISCUSSION

3.1 Descriptive Statistics

After the original data are processed to a certain extent, the basic information of each variable in the data set is analyzed by descriptive statistics. Figure 5 shows the distribution of enterprise length of maturity. It can be seen that the length of maturity in this data set is mainly between 0 and 20 months, and the longest length is not more than 40 months. In the field of credit, length of maturity has two sides. The longer the length of maturity, the higher its repayment interest rate, the higher the bank's income. However, with the growth of length of maturity, the economic capacity of borrowing enterprises is also changing. Once the borrower's economic situation fluctuates, its default rate will also increase. Therefore, for the length of maturity, the bank needs to set the corresponding threshold for the enterprise as a limit. The data of loan amount mainly concentrate in (0, 5000DM) range, and the data distribution is right skewed, indicating that the small amount of loans in this data set is the majority.

Figure 5. Descriptive statistical analysis results



3.2 Empirical Analysis Based on Single Algorithm Model

Logistic model: the training data set and test data set are brought into the model for training. Table 1 shows the model results. In the training data set, the overall prediction accuracy rate of the sample is 80.1%, the default prediction accuracy rate is 61%, and the normal prediction accuracy rate is 86.5%; in the test data set, the overall prediction accuracy rate of the model is 70%, the default prediction accuracy rate is 69%, and the normal prediction accuracy rate is 71%. It can be seen that the effect of the model is good, but the prediction accuracy of the training data set is higher than that of the test data set, indicating that the model has a certain degree of over fitting phenomenon.

Decision tree model: Table 2 shows that in the training data set, the model correctly predicts 688 cases, with the overall prediction accuracy rate of 86%, the normal prediction accuracy rate of 90.2%, and the default prediction accuracy rate of 73.5%; in the test data set, the model correctly predicts 146 cases, with the overall prediction accuracy rate of 73%, the normal prediction accuracy rate of 83%,

Table 1. Prediction results of Logistic model

		Breach of contract	Normal	All	ACC
Training set	Breach of contract	122	78	200	0.61
	Normal	81	519	600	0.865
	All	203	597	800	0.801
Test set	Breach of contract	69	31	100	0.69
	Normal	29	71	100	0.71
	All	98	102	200	0.7

and the default prediction accuracy rate of 63%. It can be seen that the overall prediction accuracy of the model and the prediction accuracy of customers are high, but the prediction accuracy of default is poor, and the model has a certain degree of over fitting phenomenon.

Neural network model: according to Table 3, in the training data set, the neural network model correctly classifies 560 samples, and the overall prediction accuracy is 70%. Among them, 426 samples of normal customers are correctly classified, and the prediction accuracy rate is 71%; for 134 samples in default, the prediction accuracy rate is 67%. In the test data set, 128 samples are correctly classified, and the overall prediction accuracy is 64%. Among them, 65 are normal and correct, and the prediction accuracy is 65%; 63 defaults are classified correctly, and the prediction accuracy is 63%. It can be seen that the prediction accuracy of the model is about 65% ~ 70%, and the prediction effect of the model on the data is general.

Table 2. Prediction results of decision tree model

		Breach of contract	Normal	All	ACC
Training set	Breach of contract	147	53	200	0.735
	Normal	59	541	600	0.902
	All	206	594	800	0.86
Test set	Breach of contract	63	37	100	0.63
	Normal	17	83	100	0.83
	All	80	120	200	0.73

3.3 Empirical Analysis Based on Integration Model

Figure 6 shows that for the training data set, the Boosting integrated model correctly predicts 720 cases. The overall prediction accuracy of the model is 90%, and the error rate is 10%. The model has a better effect. Among them, the normal correct prediction rate is 93%, and the default prediction accuracy rate is 81%. Compared with the previous decision tree model, the prediction accuracy of the model is significantly improved. For the test data set, the Boosting integrated model correctly predicts 154 cases. The overall prediction accuracy rate of the model is 77%, the normal prediction accuracy rate is 84%, and the prediction accuracy rate of defaulting customers is 70%. The effect of the model is also significantly improved. However, through the comparative analysis of the prediction results of the Boosting integrated model on the training data set and the test data set, it can be seen that the

Table 3. Prediction results of neural network model

		Breach of contract	Normal	All	ACC
Training set	Breach of contract	134	66	200	0.67
	Normal	174	426	600	0.71
	All	308	492	800	0.7
Test set	Breach of contract	63	37	100	0.63
	Normal	35	65	100	0.65
	All	98	102	200	0.64

prediction accuracy of the model has been significantly improved. However, for the data over fitting phenomenon in the decision tree model, the improvement effect of Boosting integrated model is not obvious, and there are still some defects in the model.

Note: A1-A6 are the normal number of training sets, the number of default training sets, the total number of training sets, the normal number of test sets, the number of default test sets and the total number of test sets

The overall prediction accuracy of the random forest integrated model for the training data set is 79%, the default accuracy rate is 86%, and the normal accuracy rate is 76.7%; the overall prediction accuracy rate of the model for the test data set is 77%, the normal customer's prediction accuracy rate is 71%, and the default customer's prediction accuracy rate is 83%. It can be seen that the random forest algorithm basically overcomes the over fitting phenomenon of the model, and the prediction accuracy of the model is also high, and the performance of the model is good.

3.4 Comprehensive Performance Analysis of the Model

Figure 7 shows that from the overall prediction accuracy of the model, the combined optimization model, neural network model and Boosting model have higher accuracy, which are 83.5%, 78% and 77%, respectively. The accuracy rate of decision tree model is the lowest, only 65%; from the perspective of the normal prediction accuracy of the model, Boosting model is the highest, which is

Figure 6. Empirical analysis based on integrated model

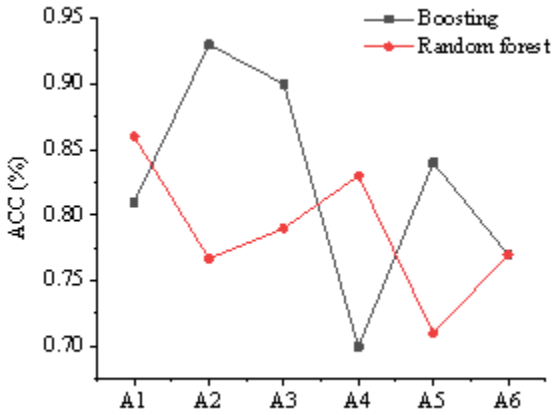
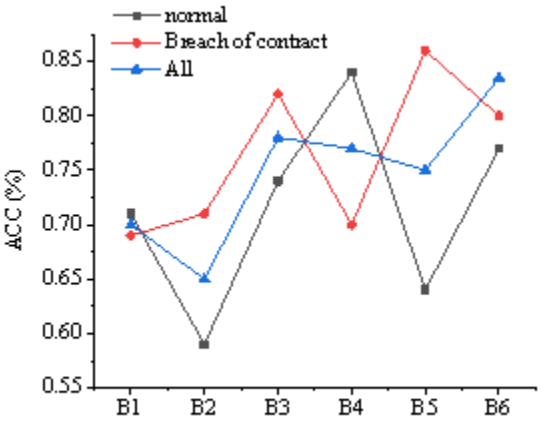


Figure 7. Analysis results of comprehensive performance of the model



84%, followed by the combined optimization model with 77%; from the perspective of the prediction accuracy rate of default, the combined optimization model, random forest model and neural network model have higher prediction accuracy, which are 90%, 86% and 82%, respectively. From the interpretation of the output results, the decision tree model and Logistic model are more explanatory, the combined optimization model is in the middle, and the neural network and random forest model are poor explanatory. From the stability of the model, the stability of combined optimization is the best, and the stability of neural network model is poor.

4. CONCLUSION

Based on the analysis of the disadvantages of the third-party platform enterprise credit evaluation mechanism under the current sharing economy mode, the enterprise risk assessment model based on data analysis is established by combining data mining technology with enterprise credit evaluation. The advantages and disadvantages of each model are comprehensively evaluated in terms of prediction accuracy, model complexity and model stability. Among them, in the three single algorithm models, the overall prediction accuracy of the models is good, while the overall prediction accuracy of the neural network model is the highest. Compared with the single algorithm model, the integrated model has higher prediction accuracy and less over fitting tendency. The combined optimization model has obvious improvement in the overall prediction accuracy, the prediction accuracy for default enterprises and the stability of the model, and the model has the best effect. Although data mining methods have been used for empirical analysis, there are still some deficiencies in the research process. First, in the study of enterprise credit risk. The relevant credit information of enterprises is only concerned from the micro level, but not from the macro level. Different economic and social structure and economic development level will have a certain influence on the credit status of enterprises, resulting in changes in the whole credit risk; second, simple data mining methods are used to analyze the influence on sustainable consumption from the perspective of enterprise credit mechanism, and it is also necessary to optimize the defects of the algorithm itself. The optimized algorithm can be used for data modeling, and deep learning algorithm can also be used for model construction. In order to provide more research ideas and methods for the influence of third-party enterprise credit on sustainable consumption in the sharing economy, in-depth research will be carried out in these two aspects.

ACKNOWLEDGMENT

General project of National Natural Science Foundation of China (Project No.72072026). The plan of developing talents in Liaoning Province“Research on the mechanism and promotion policy of high growth of enterprises in the era of digital economy” (Project No.XLYC1904017).

REFERENCES

Barrow, C. J. (2018). *Sustainable development*. The International Encyclopedia of Anthropology.

Bengtsson, M., Alfredsson, E., Cohen, M., Lorek, S., & Schroeder, P. (2018). Transforming systems of consumption and production for achieving the sustainable development goals: Moving beyond efficiency. *Sustainability Science*, *13*(6), 1533–1547. doi:10.1007/s11625-018-0582-1 PMID:30546486

Brokamp, C., Jandarov, R., Hossain, M., & Ryan, P. (2018). Predicting daily urban fine particulate matter concentrations using a random forest model. *Environmental Science & Technology*, *52*(7), 4173–4179. doi:10.1021/acs.est.7b05381 PMID:29537833

Carleo, G., & Troyer, M. (2017). Solving the quantum many-body problem with artificial neural networks. *Science*, *355*(6325), 602–606. doi:10.1126/science.aag2302 PMID:28183973

Deng, X., Liu, Z., Zhan, Y., Ni, K., Zhang, Y., Ma, W., Shao, S., Lv, X., Yuan, Y., & Rogers, K. M. (2020). Predictive geographical authentication of green tea with protected designation of origin using a random forest model. *Food Control*, *107*, 106807–106811. doi:10.1016/j.foodcont.2019.106807

Dorogush, A. V., Ershov, V., & Gulin, A. (2018). *CatBoost: gradient boosting with categorical features support*. arXiv preprint arXiv:181011363.

Geng, D., Liu, J., & Zhu, Q. (2017). Motivating sustainable consumption among Chinese adolescents: An empirical examination. *Journal of Cleaner Production*, *141*, 315–322. doi:10.1016/j.jclepro.2016.09.113

Govindan, K. (2018). Sustainable consumption and production in the food supply chain: A conceptual framework. *International Journal of Production Economics*, *195*, 419–431. doi:10.1016/j.ijpe.2017.03.003

Huang, X., Liu, X., & Ren, Y. (2018). Enterprise credit risk evaluation based on neural network algorithm. *Cognitive Systems Research*, *52*, 317–324. doi:10.1016/j.cogsys.2018.07.023

Jabbarian Amiri, B., Asgarian, A., & Sakieh, Y. (2017). Introducing landscape accuracy metric for spatial performance evaluation of land use/land cover change models. *Geocarto International*, *32*(11), 1171–1187. doi:10.1080/10106049.2016.1206628

Karout, M., Miesch, M., Geoffroy, P., Kraft, S., Hofmann, H. D., Mensah-Nyagan, A. G., & Kirsch, M. (2016). Novel analogs of allopregnanolone show improved efficiency and specificity in neuroprotection and stimulation of proliferation. *Journal of Neurochemistry*, *139*(5), 782–794. doi:10.1111/jnc.13693 PMID:27256158

Koch, H. M., Rüther, M., Schütze, A., Conrad, A., Pälmke, C., Apel, P., Brüning, T., & Kolossa-Gehring, M. (2017). Phthalate metabolites in 24-h urine samples of the German Environmental Specimen Bank (ESB) from 1988 to 2015 and a comparison with US NHANES data from 1999 to 2012. *International Journal of Hygiene and Environmental Health*, *220*(2), 130–141. doi:10.1016/j.ijheh.2016.11.003 PMID:27863804

Lukman, R. K., Glavič, P., Carpenter, A., & Virtič, P. (2016). Sustainable consumption and production–Research, experience, and development–The Europe we want. *Journal of Cleaner Production*, *138*, 139–147. doi:10.1016/j.jclepro.2016.08.049

Puschmann, T., & Alt, R. (2016). Sharing economy. *Business & Information Systems Engineering*, *58*(1), 93–99. doi:10.1007/s12599-015-0420-2

Ranganathan, P., Pramesh, C., & Aggarwal, R. (2017). Common pitfalls in statistical analysis: Logistic regression. *Perspectives in Clinical Research*, *8*(3), 148–153. PMID:28828311

Richter, C., Kraus, S., Brem, A., Durst, S., & Giselbrecht, C. (2017). Digital entrepreneurship: Innovative business models for the sharing economy. *Creativity and Innovation Management*, *26*(3), 300–310. doi:10.1111/caim.12227

Sarker, I. H., Colman, A., Han, J., Khan, A. I., Abushark, Y. B., & Salah, K. (2020). Behavdt: A behavioral decision tree learning to build user-centric context-aware predictive model. *Mobile Networks and Applications*, *25*(3), 1151–1161. doi:10.1007/s11036-019-01443-z

Shi, H. T., & Nan, X. W. (2016). Study on Operation Risk Management of Commercial Bank Small and Micro Enterprise Credit Business. DEStech Transactions on Engineering and Technology Research, 235-240.

Sinha, P., Gaughan, A. E., Stevens, F. R., Nieves, J. J., Sorichetta, A., & Tatem, A. J. (2019). Assessing the spatial sensitivity of a random forest model: Application in gridded population modeling. *Computers, Environment and Urban Systems*, *75*, 132–145. doi:10.1016/j.compenvurbsys.2019.01.006

Song, M., Fisher, R., & Kwoh, Y. (2019). Technological challenges of green innovation and sustainable resource management with large scale data. *Technological Forecasting and Social Change*, *144*, 361–368. doi:10.1016/j.techfore.2018.07.055

Tayefi, M., Tajfard, M., Saffar, S., Hanachi, P., Amirabadizadeh, A. R., Esmaeily, H., Taghipour, A., Ferns, G. A., Moohebati, M., & Ghayour-Mobarhan, M. (2017). hs-CRP is strongly associated with coronary heart disease (CHD): A data mining approach using decision tree algorithm. *Computer Methods and Programs in Biomedicine*, *141*, 105–109. doi:10.1016/j.cmpb.2017.02.001 PMID:28241960

Van Gerven, M., & Bohte, S. (2017). Artificial neural networks as models of neural information processing. *Frontiers in Computational Neuroscience*, *11*, 114–121. doi:10.3389/fncom.2017.00114 PMID:29311884

Vergragt, P. J., Dendler, L., de Jong, M., & Matus, K. (2016). Transitions to sustainable consumption and production in cities. *Journal of Cleaner Production*, *134*, 1–12. doi:10.1016/j.jclepro.2016.05.050

Vith, S., Oberg, A., Höllerer, M. A., & Meyer, R. E. (2019). Envisioning the 'sharing city': Governance strategies for the sharing economy. *Journal of Business Ethics*, *159*(4), 1023–1046. doi:10.1007/s10551-019-04242-4

Wang, F., Ding, L., Yu, H., & Zhao, Y. (2019). Big data analytics on enterprise credit risk evaluation of e-Business platform. *Information Systems and e-Business Management*, ●●●, 1–40. doi:10.1007/s10257-019-00414-x

Wang, Y., Xiang, D., Yang, Z., & Ma, S. S. (2019). Unraveling customer sustainable consumption behaviors in sharing economy: A socio-economic approach based on social exchange theory. *Journal of Cleaner Production*, *208*, 869–879. doi:10.1016/j.jclepro.2018.10.139

Wang, J.C. (2018). Technology, the nature of information, and Fintech Marketplace Lending. *Federal Reserve Bank of Boston Research Paper Series Current Policy Perspectives Paper, 18*(3), 88-92.

Wu, C. S., Zhou, X. X., & Song, M. (2016). Sustainable consumer behavior in China: An empirical analysis from the Midwest regions. *Journal of Cleaner Production*, *134*, 147–165. doi:10.1016/j.jclepro.2015.06.057

Yan, R., Ma, Z., Zhao, Y., & Kokogiannakis, G. (2016). A decision tree based data-driven diagnostic strategy for air handling units. *Energy and Building*, *133*, 37–45. doi:10.1016/j.enbuild.2016.09.039

Yang, J. (2020). Research on the Forecasting of Enterprise Credit Scoring Based on SVR Model. *Academic Journal of Engineering and Technology Science.*, *3*(1), 529–536.

Zou, L. (2019). The Research on Influence of Tourism Economy and Environment Based on Environmental Taxation. In *IOP Conference Series: Earth and Environmental Science*. IOP Publishing. doi:10.1088/1755-1315/252/4/042056

Tao Meng was born in Hanzhong, Shanxi, China, in 1975. Now, he is the Dean and Professor of the International Business College of Dongbei University of Finance and Economics. His research interest includes network organization, sharing economy.

Qi Li was born in Jinan, Shandong, China, in 1987. Now, he is a doctoral student in the School of Business Administration of Dongbei University of Finance and Economics. His research interest include network organization, sharing economy. He is the corresponding author.

Zheng Dong was born in Jinan, Shandong, China, in 1987. Now, he is a doctoral student in the School of Business Administration of Dongbei University of Finance and Economics. His research interest include network organization, sharing economy. School of Business Administration,

Feifei Zhao was born in Jining, Shandong, China, in1986. Now, he is a doctoral student in the School of Business Administration of Dongbei University of Finance and Economics. His research interest include network organization, sharing economy.