

## Research Article

# Toward an Efficient and Effective Credit Scorer for Cross-Border E-Commerce Enterprises

Chang Xu <sup>1</sup>, Ruize Guo,<sup>1</sup> Yulai Zhang,<sup>1</sup> and Xinyuan Luo<sup>2</sup>

<sup>1</sup>School of Information and Electronic Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China

<sup>2</sup>School of Computer Science and Technology, Zhejiang University, Zhejiang, Hangzhou 310027, China

Correspondence should be addressed to Chang Xu; xc@zust.edu.cn

Received 4 April 2022; Revised 3 December 2022; Accepted 17 December 2022; Published 12 January 2023

Academic Editor: Basilio B. Fraguela

Copyright © 2023 Chang Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Building an efficient and effective credit scorer for enterprises is an important and urgent demand in the cross-border e-commerce industry. In this paper, we present a framework to build a credit scorer using e-commerce data integrated from various sources. First, an improved dependency graph approach is proposed to recognize distinct records in the dataset. Then, we apply logistic regression using a prejudice remover regularizer to train the model, preceded by predictor preparation through binning and evaluating their information value. Lastly, we build the credit scorer according to the coefficients of the model. We implement our framework on a dataset from the official customs database and a large cross-border e-commerce platform. The empirical results demonstrate that the scorer built by our methodology can be used to effectively evaluate enterprises, while also removing prejudice against small and medium enterprises to a certain extent.

## 1. Introduction

E-commerce has become a new business mode with the rapid influence of the Internet and wireless communications technology. As the world's largest exporter and second-largest importer, there are hundreds of thousands of cross-border enterprises in China, making trades with foreign partners and customers on global e-commerce platforms such as Amazon and Alibaba. Unlike the traditional offline mode, the main body of e-commerce is composed of a large number of small- and medium-sized enterprises (SMEs), which render conventional credit assessment methods, such as authorized enterprise organization (AEO), inapplicable. Both the government and the e-commerce platforms need a new efficient credit score method to evaluate these enterprises according to the multifarious data produced from their trades.

Meanwhile, the COVID-19 pandemic has severely impacted global imports and exports. As the world's second-largest economy, China's import and export trade has also been greatly affected, and SMEs are the most affected. To ease the pressure on these SMEs and speed up the flow of

goods, the Chinese administration needs to establish an efficient, effective, and nondiscriminatory credit evaluation system.

Although various credit evaluation methods have been applied in many industries, there remains a lack of effective credit assessment approaches for cross-border e-commerce companies. There are a number of challenges posed here. First, the enterprises' trade-related data are scattered across multiple isolated data sources, each of which shapes the data in its own format. Integrating these heterogeneous data (in particular, distinguishing repeated records from distinct data) is not an easy job. Second, official Chinese policies advocate the development of SMEs, whereas huge, multinational corporations have overwhelming advantages in the cross-border trade business. Thus, an effective model should consciously restrain the underlying advantages of large-scale enterprises, which obviously exist in the traditional credit score evaluation. Third, the current research study on cross-border e-commerce credit methods mainly focuses on the internal business information of enterprises, by third-party research institutions, without considering the research from the perspective of supervision units [1, 2]. Lastly, we assert

that an applicable model here is not only effective and extensible but also explicable, making the model suitable for regulated e-commerce services.

In this paper, we propose an efficient framework to build a credit scorer for cross-border e-commerce enterprises. We first integrate the e-commerce data from various sources using an improved record reconciliation approach, with the aim of recognizing and refining the entities in the heterogeneous data. Then, we train a classifier model using logistic regression, preceded by predictor selection through binning and evaluating their information value (IV). In order to eliminate the model's prejudice against SMEs, we also implement a prejudice remover regularizer in the training. Lastly, we build the credit scorer from the model coefficients. We apply our framework to the data provided by the official customs database and a large Chinese cross-border e-commerce platform. The original dataset includes about 1,400,000 records on over 4000 companies. The results indicate that the proposed scorer is effective for evaluating enterprises and is relatively friendly to SMEs.

The remainder of the paper is organized as follows: in Section 2, we discuss related work; in Section 3, we describe the details of our framework; in Section 4, we present our evaluation and the corresponding results; and lastly, we conclude our work in Section 5.

## 2. Related Work

The credit scoring problem initially originated in the banking industry [3]. In earlier years, it relied on using questionnaire data and the applicant's credit history. With the popularization of information technology, people began to explore the truth about credit from transaction data using statistical learning methods. These approaches included logistic regression [4, 5], decision trees [6, 7], SVM [8–11], nearest neighbor [12], and neural networks [13–16]. However, most of the abovementioned studies are oriented toward personal credit issues in financial applications. In recent decades, on the basis of personal credit research, experts and scholars began to study how to evaluate enterprise credit from the perspective of enterprise operation information, from the initial expert system to the later classical machine learning methods and then to the recently popular deep learning methods [17]. Compared with several different credit evaluation methods, logistic regression performs well in most cases [18]. Therefore, this study will be mainly based on logistic regression.

In recent years, some researchers have paid considerable attention to the credit score problem in e-commerce, especially for the C2C e-commerce type [19–21]. We noted that these studies always focused on transaction data, while ignoring the data on other elements of the cross-border trade chain. For example, in reference [22], the author selected some objective values from the data of a single e-commerce website as indicators and admitted that the data for logistics were not sufficient for an evaluation. Therefore, a complete assessment for cross-border enterprises needs comprehensive data from various sources, which brings another problem, i.e., record reconciliation.

Record reconciliation is one of the major problems in data integration [23]. When combining data from multiple sources, we need to recognize the various references to the same real-world entity on an individual level. The approaches include merging/purging [24], record linkage [25], hardening soft databases [26], reference matching [27], deduplication [28], object identification [29], and identity uncertainty [30]. In reference [31], the authors proposed a novel method based on propagating reference similarity decisions in a dependency graph. The information is enriched and propagated in the process of calculating the nodes of the graph. This method is extremely suitable for the scenario of our research, but its performance exhibits a severe bottleneck in practice.

It is important in machine learning to choose an appropriate penalty term to prevent overfitting. One of the main problems is removing the prejudice against SMEs. The authors of reference [32, 33] focused on evaluating and reducing indirect prejudice through a statistical dependence between sensitive features and other information. A prejudice remover regularizer was proposed to enforce the determination's independence from sensitive information, which is considered to be the underlying cause of the disadvantages of SMEs.

## 3. Framework

The experiments described in this section are performed on a cloud server with the Linux operating system. In terms of hardware, the server memory is 8 GB and the CPU is an e5-2680 v4. In terms of software, the Sklearn 1.0.1 framework is used and the programming language is Python 3.7.

We propose a framework to build a credit scorer for cross-border e-commerce enterprises using their trade-related data, which are collected and integrated from multiple data sources. First, we refine the integrated dataset to eliminate the redundant records, which would otherwise distort the learning process. Then, we implement logistic regression using a prejudice remover regularizer to train the model, preceded by predictor preparation through binning and evaluating their information value. Lastly, we build the credit scorer according to the coefficients of the model we trained. The details of the abovementioned steps are described in the sections below.

### 3.1. Record Reconciliation

*3.1.1. Problem Description.* Record reconciliation is one of the major problems when integrating data from multiple sources. As different sources have their own data formats, there exist duplicate records referring to the same real-world entity. In our research, this entity may be trade records, transport records, check records, or administrative penalty records. Furthermore, there exist unavoidable inconsistencies in business systems, such as misspellings, heterogeneous use of abbreviations, discrepancies in particular attributes, and errors in data transforming/formatting. For example, in Figure 1(a), we list three trade records  $T_1$ ,  $T_2$ , and  $T_3$  with their reference enterprise records  $E_1$  and  $E_2$  and

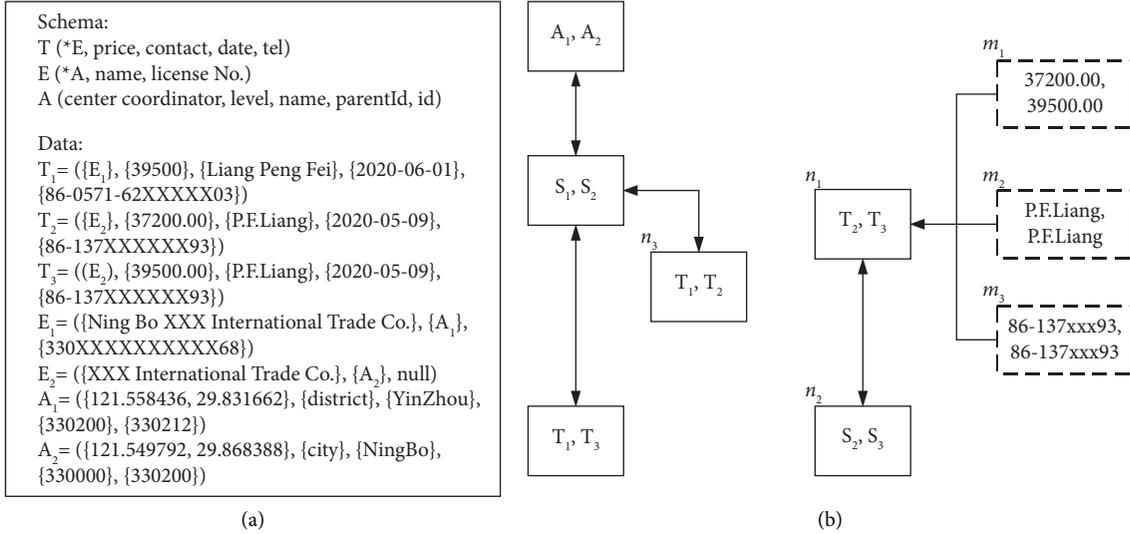


FIGURE 1: Example of dependency graph approach. (a) Three trade records  $T_1$ ,  $T_2$ , and  $T_3$  and their related records.  $T_1$  is from the official customs database;  $T_2$  and  $T_3$  are from an e-commerce platform database. (b) The dependency graph of the example. Due to space limitations, we only show the atomic attribute nodes of  $T_2$  and  $T_3$  in the graph.

area records  $A_1$  and  $A_2$ , where  $T_1$  is from the official customs database and  $T_2$  and  $T_3$  are from an e-commerce platform's database. While they seem to be three different trades of two enterprises, they are actually the same.

Obviously, this kind of data mistake will produce phantom data for subsequent model training, causing unpredictable errors. In order to improve data quality, we first need to combine these records. The dependency graph approach proposed in reference [27] is suitable for this problem. However, in practice, we found it to be extremely computationally expensive and memory-intensive. Thus, we modified the original approach for our case to improve its efficiency.

**3.1.2. The Original Dependency Graph Approach.** The approach begins by constructing the dependency graph of all records. One node in the graph marks a similarity between a pair of records or a pair of atomic attributes (whose values are of a simple type, such as string or numeric). There exists an edge from node  $n$  to  $m$  if, and only if, the similarity of  $m$  truly depends on the similarity of  $n$ . According to the different types of dependency, the edges can be classified as strong Boolean-valued, weak Boolean-valued, and real-valued. Figure 1(b) depicts a part of the dependency graph using the data in Figure 1(a).

The second step of the approach is to iteratively recompute the scores on the nodes to propagate similarity decisions in the graph. This step starts by computing the similarity scores of the atomic attribute nodes and then their neighbors, using different strategies according to the type of edge. During the propagation, when two nodes are merged, their neighbors need to be concurrently recomputed, thereby enriching the information about the record. After all record nodes are computed, we can then compute the transitive closure for the final reconciliation results. Due to

space limitations, we do not describe all details of the approach here; instead, the reader is referred to reference [27].

The dependency graph approach is novel and powerful as it not only employs the similarity between atomic attributes but also explores the similarity dependency in complex information spaces. For example, in the data of Figure 1(a), if we merge  $E_1$  and  $E_2$ , the similarity between  $T_1$  and  $T_2$  is certainly increased. We adopted this approach in our research. Unfortunately, we found the computational cost and memory consumption to be extensive when using our dataset. Thus, we optimize the original algorithm to improve its efficiency for our case, as described in Section 3.1.3.

**3.1.3. Improvement.** Using performance analytics, we identified two major performance bottlenecks in the original approach:

- (i) Gigantic graph volume: As described in Section 3.1.2, the approach needs to first construct a complete graph, in which each node represents a pair of records and their attributes. Therefore, theoretically, it needs to allocate at most  $(m + 1)C_n^2$  nodes for  $n$  records with  $m$  attributes. Although the original paper optimized this case by omitting the nodes with an initial similarity lower than an empirical threshold, it still consumed too much memory in practice. In addition to the serious computational burden, as the physical memory cannot accommodate the full graph structure, the huge mass of I/Os prevents the calculation from being completed in an acceptable time.
- (ii) Frequent computation on merge: When two records are merged, all nodes connected to them need to be checked, some need to be recomputed, and their neighbors need to be reconnected [27]. This involves

```

//Rb is the base set, while Rc is the candidate record set to be reconciled into Rb.
R'c ← Rc;
for each record r in Rb
    G ← construct dependency graph of r and each record in R'c;
    R' ← do reconciliation on G;
    R'c ← R'c - R';
end for
G' ← construct dependency graph of R'c;
R' ← do reconciliation on G'
return Rb ∪ R'

```

ALGORITHM 1: Reconcile ( $R_b, R_c$ ).

many search and update operations, which represent a heavy burden in a large graph structure.

The reason for these bottlenecks is that the original approach is symmetric, whereby all records need to be compared with each other and verified at each iteration. However, in reality, many datasets are asymmetric. In our scenario, the data from the official customs database can be considered exact. Thus, we can use these data as the *base set*, whereas data from other sources can be seen as *candidates* for reconciliation. Therefore, the original approach can be improved as outlined in Algorithm 1.

The key point of Algorithm 1 is that we change the original symmetric approach to a progressive one. In each loop, we pick one record from the base set and reconcile it with the remaining records in the candidate dataset. We need not consider the reconciliation between the records in the base set as we assume them to be exact. Then, we remove records recognized as repeated from the candidate set to reduce unnecessary computation in the next loop. Lastly, we reconcile the remaining records in the candidate dataset and then merge them into the final dataset. In this way, we reduce the upper bound size of the dependency graph to the volume of the candidate dataset, which greatly improves the efficiency of the original approach.

**3.2. Model Training.** We apply logistic regression for model training because of its efficiency and explicability. To choose the appropriate predictors for each feature, we first discretize the values into several bins and then calculate the weight of evidence (WOE) for each bin. Second, we calculate the information value (IV) of each feature to choose the most powerful ones. Lastly, we train the model by applying logistic regression using a prejudice remover regularizer.

**3.2.1. Binning and WOE.** For each feature, we choose the corresponding binning method as a matter of experience. The methods include equal-width, equal-frequency, and k-means clustering. It is unnecessary to consider more complicated binning methods, such as supervised binning, as most features are evenly distributed.

After binning, we obtain a series of discrete values for each feature; then, we need to calculate the WOE for each bin as follows:

TABLE 1: WOE for the “average age of the customers.” The values are binned using equal-frequency clustering and then normalized.

Bin	WOE
1: [0, 0.0028]	-1.4633
2: [0.0028, 0.0061]	-0.639
3: [0.0061, 0.0082]	-0.4933
4: [0.0082, 0.0149]	0.0276
5: [0.0149, 0.0215]	0.5489
6: [0.0215, 0.048]	0.5802
7: [0.048, 0.0607]	0.8021
8: [0.0607, 1]	0.9527

$$\text{WOE}(i) = \ln \frac{G_i/G_{\text{total}}}{B_i/B_{\text{total}}}, \quad (1)$$

where  $G_i$  represents the number of good enterprises in each bin and  $G_{\text{total}}$  represents the total number of good enterprises. Accordingly,  $B_i$  represents the number of bad enterprises in each bin and  $B_{\text{total}}$  represents the total number of bad enterprises. Thus, the WOE of the  $i$ th bin denotes the discrepancy in the proportion of good/bad enterprises in this bin, which represents its distinguishability.

Table 1 lists the WOE of each bin for the “average age of the customers”.

**3.2.2. Information Value.** The information value (IV) of a feature is one of the most widely used indicators to pick predictors in credit score models. The formula for IV is as follows:

$$\text{IV} = \sum_i^n \left( \frac{B_i}{B_{\text{total}}} - \frac{G_i}{G_{\text{total}}} \right) \times \text{WOE}(i). \quad (2)$$

We can see that IV is simply the weighted sum of WOE values. A larger IV denotes that the feature is more likely to be picked as a predictor. Table 2 lists the IV values for some of the features.

**3.2.3. Prejudice Remover Regularizer.** As mentioned in Section 1, our aim was to eliminate the prejudice against SMEs. The authors of reference [28] proposed the prejudice index (PI) to quantify the degree of indirect prejudice, which

TABLE 2: IV coefficient values and  $P$  value for some features.

Feature	IV	Coefficient value in final model	$P$ value
Inspection hit rate of goods	0.925	1.2453	<0.01
Percentage of related tiny transport enterprises	0.794	1.2921	<0.01
High amount from the same payer	0.766	1.1201	<0.01
Record of execution on dishonesty	0.701	0.6463	0.0512
Average age of the buyer	0.563	0.6998	0.0295
Consistency of the buyer and payer	0.343	0.6525	0.0242
Abnormal contact details	0.004	-1.0202	0.0139
Related transport enterprise has a record of dishonesty	0.121	1.4309	0.0332

$P$  value is the significance level of the features calculated according to the actual statistics, and a value lower than 0.05 indicates a significant difference.

is straightforwardly defined as the mutual information between  $Y$  and  $X_a$ :

$$PI = \sum_{Y, X_a} P(Y, X_a) \ln \frac{P(Y, X_a)}{P(X_a)P(Y)}, \quad (3)$$

where  $Y$  represents the target result and  $X_a$  represents the value of feature  $a$ . Obviously, a larger PI indicates that  $a$  is a more sensitive feature. Thus, the regularizer can be formulated as follows:

$$\min(L[f(X), Y] + \eta \times PI), \quad (4)$$

where  $f(X)$  is the loss function and  $\eta$  is used to control the degree of prejudice removal. Accordingly, this regularizer can help to weaken the prejudice toward SMEs in the logistic regression. After training, we obtain the coefficient value of each predictor, as shown in Table 2.

**3.3. The Credit Card Scorer.** The last step is to build a credit card scorer from the model. We define the probability that an enterprise’s credit score is bad as  $p$ . Then,  $\ln(p/(1-p))$  is the relative probability (**RS**) that the enterprise’s credit score is bad. The credit card score can be defined as the linear expression of **RS**. Subsequently, we can calculate the scale for each bin by substituting some known scores obtained from experts.

Table 3 lists the scale for “the average age of the buyer.”

## 4. Evaluation and Results

**4.1. Evaluation.** The experimental data for our research were obtained from the official customs database and a large Chinese e-commerce platform. We selected more than 4000 active enterprises on the platform and collected their related data from both the platform’s database and the official customs data cloud, which consists of different government databases, including the Tax Bureau, Statistics Bureau, and Ministry of Industry and Information Technology. The data we chose were created between July, 2020, and December, 2020. For each enterprise, we obtained at least 35 transaction data records and other records related to transport information, payment information, etc. After integration, the number of features was more than 500. We classified them into different categories as shown in Table 4.

The indicators in the paper are classified as discrete and continuous, both of which are based on discussions with

TABLE 3: Final scale on “the average age of the buyers.”

Bin	Scale
1: [0, 0.0028]	97
2: [0.0028, 0.0061]	56
3: [0.0061, 0.0082]	52
4: [0.0082, 0.0149]	21
5: [0.0149, 0.0215]	-2
6: [0.0215, 0.048]	-11
7: [0.048, 0.0607]	-12
8: [0.0607, 1]	-32

experts who have hands-on case-handling experience within the supervision unit.

As can be seen from Table 4, customs clearance information, transportation information, partners, customs database remarks, and platform remarks are all new credit indicator categories constructed in this work which meet the needs of supervision units.

We evaluated our framework from three perspectives. First, we measured the efficiency and results of our improved record reconciliation approach. Then, we evaluated the classification quality of the scorer. Lastly, we conducted experiments to verify the effect of the prejudice remover regularizer.

## 4.2. Results

**4.2.1. Improved Record Reconciliation Approach.** As described in Section 3.1, we improved the original dependency graph approach to meet our dataset’s requirements. The motivation of our work is that the original approach could not be completed in an acceptable time when the dataset contained more than 100,000 records. In our improved approach, the algorithm becomes progressive and deals with only part of data in each iteration, thereby reducing both the computational cost and the memory consumption.

To measure the performance improvement of our approach, we applied the original approach and our improved one to datasets of various sizes. The full dataset consisted of about 1,400,000 records on over 4000 enterprises. In this step, we needed to characterize the transaction data, payment data, transport data, and partner data. For comparison, we divided the full dataset into various subsets, as described in Table 5.

TABLE 4: Feature categories of the dataset.

Feature category	Description	Proportion (%)
Enterprise information	The static information of the enterprises	6
Transaction information	Data about the trade	20
Transport information	Data about the transport	14
Clearance information	Data about the clearance process in customs, e.g., declaration and examination	14
Credit history	Identification of any historical bad credit records	12
Cooperators	Relations to other entities	10
Comments in customs database	Comments in the official customs database	8
Comments on platform	Comments on the e-commerce website	8
Others	Others	8

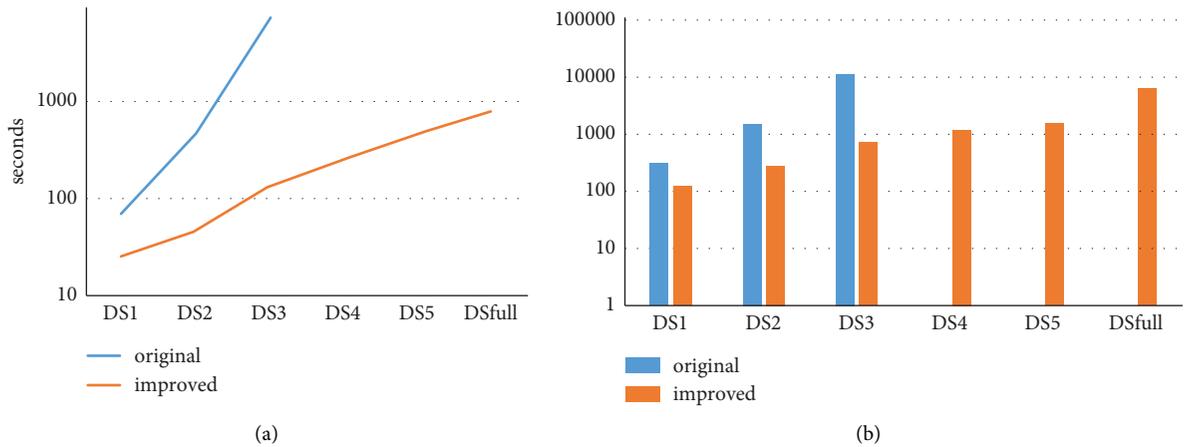


FIGURE 2: Computation cost and memory consumption comparison for original approach and improved approach. (a) Computational cost comparisons on different data sets. (b) Memory consumption comparison on different data sets.

Figure 2 depicts the computational cost and memory consumption for each dataset using the *original* approach and our *improved* approach. We can see that the original approach exhibited an exponential increase in computational cost with the increase in dataset size, as each pair of records needed to be theoretically compared. In fact, we could not obtain a valid output for datasets larger than  $DS_3$  due to surpassing the memory requirement. In comparison, our improved approach exhibited good scalability with the dataset size, as it only computed the similarity of candidate records to one record in each iteration, thus linearizing the complexity of the algorithm.

Although the performance of the approach was considerably improved, we were still concerned about the effect of the reconciliation. We evaluated the effects on  $DS_1$ ,  $DS_2$ , and  $DS_3$  as a function of precision, recall, and the reported  $F$ -measure, defined as follows:

$$F - \text{measure} = \frac{(2 \times \text{precision} \times \text{recall})}{(\text{precision} + \text{recall})}, \quad (5)$$

Table 6 lists the average precision/recall and  $F$ -measure for all classes of records involved in the reconciliation. We found that the improved approach outperformed the original approach for all three datasets. Specifically, upon increasing the size of the dataset, this advantage was also

increased. For example, the recall for  $DS_1$  was improved by 5.4%, while it was improved by 17.8% for  $DS_3$ . This is because the improved approach avoids unnecessary similarity computations between records from the exact base set, which may result in incorrect reconciliation and a decrease in recall.

**4.2.2. Quality of Scorer.** Referring to the FICO credit score rule and experts' suggestions from customs, we mapped the credit score of cross-border e-commerce enterprises within a range of 200 to 950, categorized into five levels, as defined in Table 7.

We selected and integrated records of 500 enterprises from the customs database and platform as the validation dataset. For comparison, we trained the model using datasets  $DS_1$  to  $DS_{\text{full}}$ , and then, we built the scorers individually using the method described in Sections 3.2 and 3.3. Then, we scored the enterprises and determined their credit level. Lastly, we compared the results to those manually crafted by experts to obtain the AUC value.

Figure 3 describes the learning curves of our model with/without new indicators on training datasets of different sizes. We can see that the AUC increases with the size of the data set, indicating our method's correctness and effectiveness. When the number of records exceeded 0.8 million, the quality of the

TABLE 5: Various datasets for performance evaluation.

Data set	# of enterprises	# of records
DS <sub>1</sub>	100	37,699
DS <sub>2</sub>	200	72,804
DS <sub>3</sub>	500	194,206
DS <sub>4</sub>	1000	402,152
DS <sub>5</sub>	2000	800,422
DS <sub>full</sub>	4025	1,424,228

TABLE 6: Comparison of the original and improved approach.

Data set	Original		Improved	
	Precision/recall	F measure	Precision/recall	F measure
DS <sub>1</sub>	0.954/0.921	0.937	0.999/0.975	0.989
DS <sub>2</sub>	0.922/0.905	0.913	0.988/0.970	0.979
DS <sub>3</sub>	0.892/0.745	0.812	0.989/0.923	0.955

TABLE 7: Credit levels and their corresponding score intervals.

Credit level	Highest score	Lowest score
BB	347	200
BBB	497	348
A	653	498
AA	801	654
AAA	950	802

A score lower than 200 is scaled to 200 and a score greater than 950 is scaled to 950.

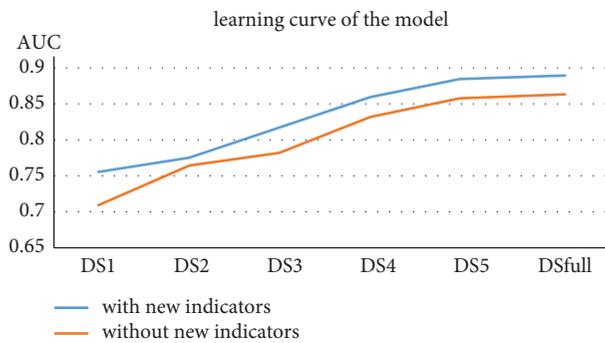


FIGURE 3: The learning curve on training datasets of various sizes.

TABLE 8: Qualified/unqualified results for enterprises on different levels.

Credit level	Qualified	Unqualified	Qualified ratio (%)
BB	11	39	22
BBB	29	21	58
A	40	10	80
AA	48	2	96
AAA	50	0	100

classification stabilized at an AUC value of 0.89, representing an empirically qualified score to deploy in production. The AUC value of the classification quality of the model without the new indicators is only 0.82, which indicates that the new index added here improves the quality of the model.

In view of the ineluctable discrepancy in human recognition that exists in manual grading, we conducted another validation test by selecting 50 enterprises on each level from the results of the validation dataset. Consequently, the experts assigned each case a binary remark (qualified/unqualified) according to their behavior. The results are summarized in Table 8. We can see that our model score and the qualified ratio from the experts exhibited a strong convergence, again indicating the correctness of our model.

**4.2.3. Effect of the Prejudice Remover Regularizer.** In order to evaluate the effect of the prejudice remover regularizer, we used the disparate impact (DI) [28], which is defined as follows:

$$DI(X) = \frac{P(Y = B | X = 0)}{P(Y = B | X = 1)}. \quad (6)$$

DI is calculated by dividing the conditional probability of the positive result given a sensitive  $X$  value by that given a nonsensitive  $X$  value. A DI value nearer to 1 denotes that the predictor is more insensitive to the score result.

Several potentially sensitive predictors were chosen according to the experts' advice and classified into three categories: penalty, capital, and profit. Figure 4(a) demonstrates the DI for the three categories of predictors with/without the regularizer. We can see that the DI with the regularizer was nearer to 1 than that without the regularizer, indicating its ability to remove certain prejudices originally existing toward SMEs. In the three categories, the effect on

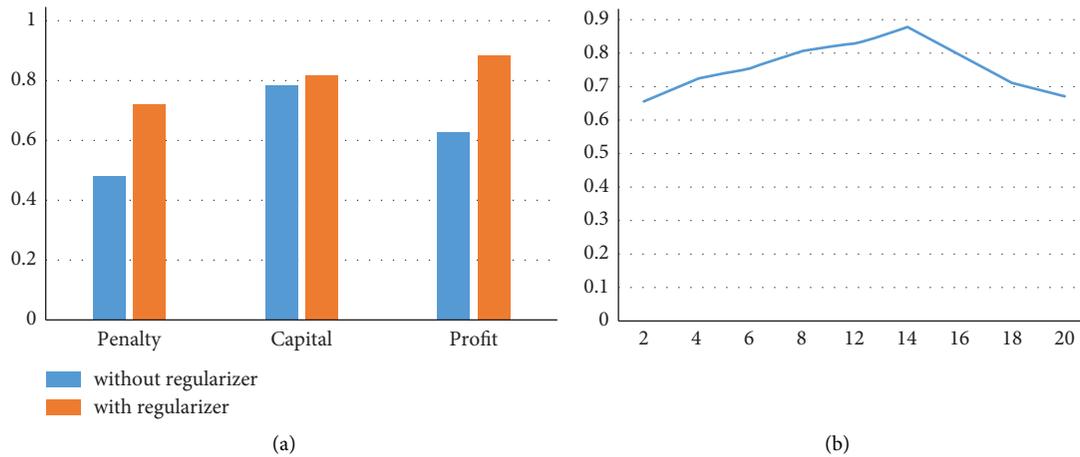


FIGURE 4: DI evaluation on (a) different categories of predictors and (b) various values of  $\eta$  for category profit.

*capital* was the weakest, suggesting a potentially strong correlation between capital-related features and credit score in cross-border e-commerce transactions.

Figure 4(b) depicts the effect of prejudice removal on various values of  $\eta$  as calculated in equation (4). We can see that when the value of  $\eta$  was equal to 14, the effect of prejudice removal was best. The category we used in this experiment was *profit*.

## 5. Conclusions

In this paper, we propose a framework for building an effective credit scorer for cross-border e-commerce enterprises. The contributions of this paper are as follows: first, we introduced a better dependency graph approach to reconcile the redundancy of records from multiple data sources, which improved the data quality for consequent model training. Compared with the original algorithm, our method could save considerable computational cost and memory consumption, making it applicable for large datasets. Second, we used logistic regression for data training, preceded by predictor preparation through binning and evaluating the information value, which was maturely adopted to personal credit applications. The advantage of this approach is its efficiency and explicability. Third, we used a prejudice remover regularizer to weaken the excessive advantage of large enterprises, in response to the official policy to advocate the development of small and medium e-commerce enterprises. Experiments on real datasets demonstrated both the effectiveness and efficiency of our framework.

In the future, our aim is to try other kinds of regularizers for prejudice removal, which would help us better understand the underlying clues related to the credit score of cross-border e-commerce enterprises. Furthermore, our aim is to integrate additional data from other sources, especially related to enterprises' daily activities, to improve the model's precision. Building a credit score system is extremely important for cross-border e-commerce; accordingly, we anticipate further challenges in the foreseeable future.

## Data Availability

The data in this study came from the Big Data Department of Hangzhou Customs, Zhejiang Province, China. Due to data security, the data and code of the study cannot be provided.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] Z. Deng, M. Yan, and X. Xiao, "An early risk warning of cross-border e-commerce using BP neural network," *Mobile Information Systems*, vol. 2021, Article ID 5518424, 8 pages, 2021.
- [2] Y. Tao and W. Zhang, "Establishment of cross-border e-commerce credit evaluation system based on big data," *Management & Engineering*, vol. 24, no. 3, 2016.
- [3] L. Thomas, J. Crook, and D. Edelman, *Credit Scoring and its Applications*, pp. 1–10, SIAM Journals Online, Philadelphia, PA, USA, 2017.
- [4] J. C. A. Wiginton, "A note on the comparison of logit and discriminant models of consumer credit b," *Journal of Financial and Quantitative Analysis*, vol. 15, no. 3, pp. 757–770, 1980.
- [5] N. M. I. B. Lund and I. Consultant, *Weight of Evidence Coding and Binning of Predictors in Logistic Regression*, MidWest SAS Users Group, Cincinnati, OH, USA, 2016.
- [6] P. Makowski, "Credit scoring branches out," *Credit World*, vol. 75, pp. 30–37, 1985.
- [7] D. Zhang, X. Zhou, S. C. H. Leung, and J. Zheng, "Vertical bagging decision trees model for credit scoring," *Expert Systems with Applications*, vol. 37, no. 12, pp. 7838–7843, 2010.
- [8] T. van Gestel, B. Baesens, J. Garcia, and P. van Dijke, "A support vector machine approach to credit scoring," *Bank Financ*, vol. 2, pp. 73–82, 2003.
- [9] C. L. Huang, M. C. Chen, and C. J. Wang, "Credit scoring with a data mining approach based on support vector machines," *Expert Systems with Applications*, vol. 33, no. 4, pp. 847–856, 2007.
- [10] R. Y. Goh and L. S. Lee, "Credit scoring: a review on support vector machines and metaheuristic approaches," *Advances in*

- Operations Research*, vol. 2019, Article ID 1974794, 30 pages, 2019.
- [11] S.-T. Li, W. Shiue, and M. H. Huang, "The evaluation of consumer loans using support vector machines," *Expert Systems with Applications*, vol. 30, no. 4, pp. 772–782, 2006.
- [12] W. E. Henley and D. J. Hand, "Construction of a k-nearest-neighbour credit-scoring system," *IMA Journal of Management Mathematics*, vol. 8, no. 4, pp. 305–321, 1997.
- [13] D. West, S. Dellana, and J. Qian, "Neural network ensemble strategies for financial decision applications," *Computers & Operations Research*, vol. 32, no. 10, pp. 2543–2559, 2005.
- [14] L. Yu, S. Wang, and K. K. Lai, "Credit risk assessment with a multistage neural network ensemble learning approach," *Expert Systems with Applications*, vol. 34, no. 2, pp. 1434–1444, 2008.
- [15] H. Abdou, J. Pointon, and A. El-Masry, "Neural nets versus conventional techniques in credit scoring in Egyptian banking," *Expert Systems with Applications*, vol. 35, no. 3, pp. 1275–1292, 2008.
- [16] Y. M. Huang, C. M. Hung, and H. C. Jiau, "Evaluation of neural networks and data mining methods on a credit assessment task for class imbalance problem," *Nonlinear Analysis: Real World Applications*, vol. 7, no. 4, pp. 720–747, 2006.
- [17] D. West, "Neural network credit scoring models," *Computers & Operations Research*, vol. 27, no. 11-12, pp. 1131–1152, 2000.
- [18] I. Brown and C. Mues, "An experimental comparison of classification algorithms for imbalanced credit scoring data sets," *Expert Systems with Applications*, vol. 39, no. 3, pp. 3446–3453, 2012.
- [19] M. Xiang and S. S. Zhang, "Study about China's C2C e-commerce credit management," in *Proceedings of the 2010 International Conference on E-Business and E-Government*, pp. 2249–2252, Guangzhou, China, May 2010.
- [20] C. Piao, J. An, and M. Fang, "Study on credit evaluation model and algorithm for C2C E-commerce," in *Proceedings of the IEEE International Conference on E-Business Engineering*, pp. 392–395, Hong Kong, China, October 2007.
- [21] J. Pang and Z. Liu, "A study on credit appraisal system of C2C E-commerce &lt;br/&gt;—based on the analysis of t," *American Journal of Industrial and Business Management*, vol. 04, no. 12, pp. 721–727, 2014.
- [22] F. Liang, R. Zou, and Q. He, "E-commerce seller credit comprehensive evaluation mode," in *Proceedings of the International Conference of Computer Science and Service System*, pp. 1246–1249, Nanjing, China, June 2011.
- [23] H. B. Newcombe, J. M. Kennedy, S. J. Axford, and A. P. James, "Automatic linkage of vital records," *Science*, vol. 130, no. 3381, pp. 954–959, 1959.
- [24] M. A. Hernandez and S. J. Stolfo, "The merge/purge problem for large databases," *ACM SIGMOD Record*, vol. 24, no. 2, pp. 127–138, 1995.
- [25] W. E. Winkler, "The state of record linkage and current research problems," Technical Report, U.S. Bureau of the Census, Washington, DC, USA, 1999.
- [26] W. W. Cohen, H. Kautz, and D. McAllester, "Hardening soft information sources," in *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 255–259, Boston, MA, USA, August 2000.
- [27] A. McCallum, K. Nigam, and L. H. Ungar, "Efficient clustering of high-dimensional data sets with application to reference matching," in *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 169–178, Boston, MA, USA, August 2000.
- [28] S. Sarawagi and A. Bhamidipaty, "Interactive deduplication using active learning," in *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 269–278, Edmonton, Canada, July 2002.
- [29] S. Tejada, C. A. Knobloc, and S. Minton, "Learning domain-independent string transformation weights for high accuracy object identification," in *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 350–359, Edmonton, Canada, July 2002.
- [30] A. McCallum and B. Wellner, "Toward conditional models of identity uncertainty with application to proper noun coreference," in *Proceedings of the IJCAI 2003 Workshop on Information Integration on the Web*, pp. 79–86, Acapulco, MX, USA, August 2003.
- [31] X. Dong, A. Halevy, and J. Madhavan, "Reference reconciliation in complex information spaces," in *Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data*, pp. 85–96, Baltimore, MD, USA, June 2005.
- [32] T. Kamishima, S. Akaho, H. Asoh, and J. Sakuma, "Fairness-aware classifier with prejudice remover regularizer," in *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Database*, Bristol, UK, September 2012.
- [33] M. Xiang and S. S. Zhang, "Study about China's C2C e-Commerce Credit Management," in *Proceedings of the 2010 International Conference on E-Business and E-Government*, pp. 2249–2252, Guangzhou, China, May 2010.