Information Resources Management Journal
Volume 37 • Issue 1

An Exploration of the Computer Big Data Mining Service Model Under Resource Sharing

WeiWei Hu, Henan Kaifeng College of Science Technology and Communication, China
Lina Sun, Henan University School of Software, China*
Lijie Li, Henan Kaifeng College of Science Technology and Communication, China

ABSTRACT

In order to meet the diverse needs of users for data mining services and improve resource utilization and enterprise competitiveness, this article aims to construct a Big Data Analytics (BDA) data mining service model based on resource sharing mechanisms. This article designs a customized data mining service model for BDA based on its characteristics. In this model, the authors apply the improved Apriori algorithm to determine the optimization plan and improve the ant colony optimization algorithm to improve the efficiency and accuracy of data mining. By analyzing the experimental results, the scientificity and rationality of the proposed data mining service model for BDA were demonstrated, and the implementation strategy of the data mining model was improved. These research findings provide important references for BDA's data mining service model based on response surface modeling and also provide guidance for enterprises on how to better utilize resources and improve competitiveness when facing big data.

KEYWORDS

Artificial Neural Network (ANN), Big Data Alliance (BDA), Big Data Mining, Logistic Regression Model (LRM), Support Vector Machine (SVM)

INTRODUCTION

With the advent of the information age, big data mining has become an indispensable part of enterprise decision-making and competition. In order to meet user's diverse needs for data mining services and to improve resource utilization and enterprise competitiveness, researchers have begun to focus on computer big data mining service models based on resource sharing mechanisms. In this context, this article aims to construct a data mining service model for big data analytics (BDA) based on resource sharing mechanisms. First, this article presents in-depth research on data mining services and introduces the concept and characteristics of BDA. Subsequently, based on the characteristics of BDA, we share the design of a customized data mining service model for BDA. In this model,

DOI: 10.4018/IRMJ.340032
*Corresponding Author

we applied the improved Apriori algorithm to determine the optimization plan and improve the ant colony optimization algorithm to, in turn, improve the efficiency and accuracy of data mining. In addition, we have established a mathematical model based on actual data mining results to evaluate and validate the scientificity and feasibility of the proposed service model. Finally, by analyzing the experimental results, we demonstrate the scientificity and rationality of the proposed BDA-oriented data mining service model and improved the implementation strategy of the data mining model. These research findings provide important references for response surface model-based data mining service models for BDA and also provide guidance for enterprises on how to better utilize resources and improve competitiveness when facing big data.

LITERATURE REVIEW

Data technology has given data high application and research values in the internet era. Thus, the big data research industry came into being. Traditionally, data mastered by different departments often showed an island effect; it revealed great correlation but lacked full utilization (Saura et al., 2021). With the active promotion of national policies, big data service platforms continue to emerge. They lay a strong foundation for obtaining the required resources in the data ocean and excavating the insidious value behind the data. In the big data ecosystem (BDE), data resources are gaining much larger weight in everyday life and industrial productions. The BDE's knowledge and valuable information inject momentum into the world economy, market activities, and enterprise development. Thus, comprehensively using information resources has become a research hotspot (Gastermann et al., 2015).

Extensive research has been carried out in related fields. Xiao et al. (2019) used virtualization and related technologies to build a big data platform based on Hadoop and proposed an improved Apriori parallel algorithm to improve data mining (DM) efficiency. Pranata and Utomo (2020) analyzed the application requirements, the application scope, and the potential application value of big data with distributed and mobile technical characteristics. A big data-oriented classification and mining framework was proposed, with high classification, mining effect, and distributed computing. An optimization result could be obtained while balancing memory occupation and network communication costs between nodes. Karpagam (2022) constructed an adaptive parallel mining algorithm to overcome the high error rate of image segmentation through adaptive control. Zhang et al. (2018) discovered that media sensors extracted vast amounts of information and assisted medical diagnosis by generating text, audio, and video images (media content). It used k-means clustering (KMC) method to study the patient's disease data. The data extraction method transformed the heterogeneous information into useful quality information for decision-making, contributing to the functional technology. Shi and Liu (2021) observed that intelligent system integration mainly included applying intelligent technologies such as artificial intelligence (AI) and computational intelligence (CI) methods to different levels of the system. They introduced the basic concepts of DM, AI, machine learning (ML), statistical analysis, fuzzy logic, pattern recognition, and artificial neural networks (ANN). The general DM algorithm's structure was analyzed to classify the DM technology into over ten subsets like decision tree (DT), neural network (NN), rough set, and fuzzy set technology. Finally, the research directions of DM in AI, e-commerce applications, and mobile computing were discussed. Mobile communication computing was discussed. At present, the research on stream DM can be roughly divided into two directions. One direction studies the time model of streaming DM and proposes snapshot, landmark, and sliding window models. The stream data DM methods include sampling, histogram representation, load discarding, stream data outline, stream data aggregation, and a series of technical methods. The other direction of research addresses streaming DM-related problems, including streaming data clustering analysis, streaming data classification, streaming data frequent item statistics, streaming time series analysis (TSA), and streaming data frequent pattern mining. Although there has been extensive research on big data mining and analytics (BDMA), most studies are oriented toward a single enterprise. Nevertheless, given limited capability, single enterprises cannot meet users' diverse DM service needs. Therefore, Big Data Alliance (BDA) integrates the membership resource, technology, and capability advantages. The aim is to provide users with multi-faceted, multi-level, and high-quality DM services. Accordingly, this work builds a BDA-oriented DM service model to meet users' multi-faceted needs. It provides effective management solutions for data enterprises to give full play to their respective advantages through alliance cooperation. Finally, it improves the data resource utilization rate (RUR) and enterprise competitiveness.

RELATED MATERIALS AND METHODS

Data Mining Service

The information and data generated are growing exponentially with the increasing frequency and diversity of human social activities. However, people have difficulty retrieving desired information from large amounts of miscellaneous data. Against this backdrop, DM technology came into being (Feng et al., 2020). DM, sometimes referred to as knowledge discovery in databases(KDD),represents a process of obtaining value from data, formally expressed as data+tool+method+goal+action=val ue. DM focuses mainly on finding the relationship between the model and massive amounts of data. Technically, DM is the process of extracting hidden, unknown, insidious, and helpful information and knowledge from numerous random, noisy, fuzzy, and incomplete practical application data. DM involves statistical analysis, sequential pattern discovery, and information mining (Zhou et al., 2021). In particular, this work defines DM as a deep-seated data analysis method, an exploratory process of analyzing large amounts of data and information to reveal unknown and implicit knowledge.

With the advancement of DM technology, the standardization of the DM process is very important. However, the standardization process of DM has not been unified in academia and industry. DM process models mainly include the Fayyad model; sample, explore, modify, model, and assess (SEMMA); and the cross industry standard process for data mining (CRISP-DM) (Watada et al., 2020). CRISP-DM is a universal data mining process model developed by European Union funded organizations such as Statistical Package for the Social Sciences(SPSS) between 1996 and 2000, aiming to provide standardized methods and processes for data mining projects(Yu, 2021). The CRISP-DM model describes DM tasks' whole process from understanding business requirements, seeking business solutions, and accepting practical tests (Rani et al., 2023). The six main stages of the CRISP-DM model are as follows:

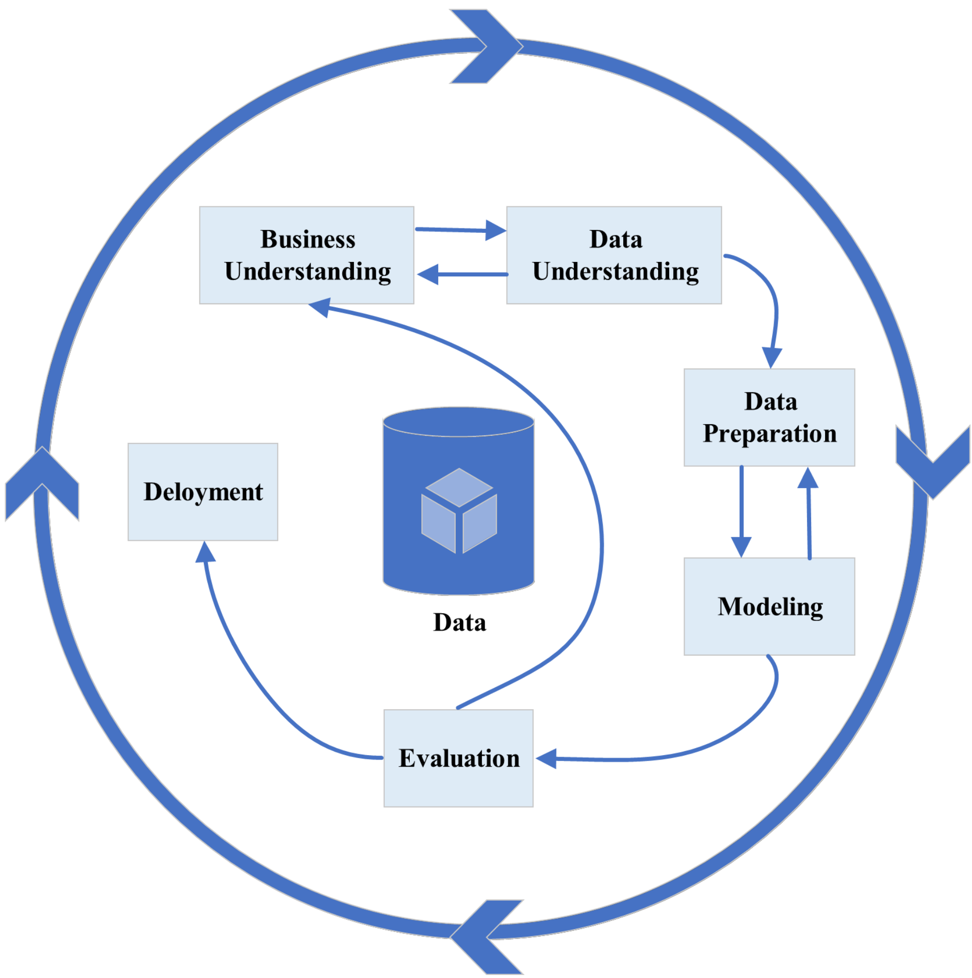
1. Business understanding stage: In this stage, the project's goals, scope, resources, and constraints are determined and understanding of business requirements and the purpose of data mining is gained. This requires communication and negotiation with the business side to ensure that project objectives are aligned with business objectives.
2. Data understanding stage: In this stage, the quality, completeness, availability, suitability, and relevance of available data is understood and evaluate. This also includes preliminary exploration and analysis of data to assist in further data preprocessing and feature engineering.
3. Data preparation stage: In this stage, preprocessing work, such as data cleaning, transformation, integration, and selection, is carried out to prepare for modeling. This also includes data processing steps such as feature extraction, variable selection, and sample partitioning.
4. Modeling stage: In this stage, various modeling techniques, such as classification, regression, clustering, and association rules, are used to construct and evaluate models to discover patterns and patterns in the data.
5. Evaluation stage: In this stage, the modeling results are evaluated and validated to ensure the accuracy, reliability, and generalization ability of the model. This includes steps such as model selection, parameter adjustment, cross validation, and testing.

6. Deployment stage: In this stage, the mining results are transformed into commercial value and applied to actual business operations. This requires collaboration with business personnel to ensure the interpretability and operability of the results in order to better support decision-making and innovation.

The CRISP-DM model is a universal data mining process model that can be applied to various data mining projects, such as marketing, financial risk control, and medical diagnosis. It provides a standardized method and process, making data mining projects more standardized and controllable. The sequence of stages mainly depends on the results of specific tasks in each stage. If the results of one stage are necessary for the next stage, the sequence cannot be changed. The process model is shown in Figure 1 (da Rocha & de Sousa, 2010).

The CRISP-DM process model in Figure 1 will be explained in detail. Firstly, business understanding represents understanding the needs of the task, transforming the business problem into the problem of data analysis, and forming a preliminary plan for the goal of the DM task at this stage (Xiao & Wang, 2020). Secondly, data understanding mainly collects data according to the results of business understanding. This involves perceiving the correlation of each field in the data set, tables,

Figure 1. A CRISP-DM Process Model



and data. Meanwhile, it is necessary to analyze the quality of the collected data set (Cui & Yan, 2020)). Thirdly, data preparation involves organizing, cleaning, and converting the collected data to meet modeling requirements. Lastly, the primary function of modeling is to obtain knowledge from data that is convenient for users to use and understand. It also mines reports or realizes a relatively complex and reusable DM process (Xu et al., 2020).

Enterprise Big Data Analysis

Enterprise big data analysis (BDA) refers to the use of big data technology and methods by enterprises to collect, store, process, and analyze large-scale data in order to obtain valuable insights and decision support. The following are some important measures that enterprises can take in BDA:

- Developing a data strategy: Enterprises should clarify their data strategy and determine the goals and key areas of data analysis. This requires a comprehensive evaluation of the business needs and data assets of the enterprise to develop suitable data strategies.
- Building data infrastructure: Enterprises need to build comprehensive data infrastructure, including data warehouses, data lakes, cloud computing platforms, and so on. This will provide enterprises with efficient data storage, processing, and management capabilities to support large-scale data analysis.
- Integrating and cleaning data: Enterprises should integrate and clean various data sources to ensure the accuracy and consistency of data. This includes data cleaning, data integration, data quality management, and other work to improve the credibility and availability of data.
- Applying data mining technology: Enterprises can adopt various data mining technologies, such as machine learning, data visualization, and natural language processing, to discover patterns, trends, and association rules in data to support decision-making and business innovation.
- Establishing a data-driven culture: Enterprises should establish a data-driven culture to encourage employees to fully utilize data in decision-making and business processes. This includes training employees in data skills, developing incentive mechanisms to encourage data-driven behavior, and so on.
- Protecting data security and privacy: Enterprises need to take corresponding measures to protect data security and privacy. This includes developing data security policies, strengthening data access controls, and complying with relevant regulations and standards.
- Engaging in continuous improvement and optimization: Enterprises should continuously improve and optimize their BDA capabilities and improve the effectiveness and value of data analysis through continuous feedback and learning.

In summary, by taking the above measures, enterprises can achieve more effective data analysis and decision support, as well as enhance their competitiveness and innovation capabilities.

Core competence theory holds that the sustainability and availability of enterprise core competence (ECC) impact competitive enterprise advantage. Therefore, forming and acquiring unique core competence is the focus of enterprises. In order to obtain market competitiveness, enterprises must improve their core competencies through market transactions, research and development (R&D), and alliance sharing (Fu et al., 2021). According to the characteristics of big data enterprises, this work constructs the ECC architecture of BDA memberships with data resources, technology, and service capability. Figure 2 describes data resource capability, data technology capability, and data service capabilities. The first capability refers to the ability of enterprises to acquire and integrate rich business data resources. Data technology capability is the ability of enterprises to have robust BDT. Generally, it entails big data infrastructure, data security, data acquisition, DM, and data analysis technology. With data technology capabilities, enterprises can provide technical support for their peers (Ghoshal et al., 2020). Data service capability is service-oriented and provided externally based on data resource and technology capability. Its services include infrastructure, data collection and acquisition, data transactions, analysis, and mining services (Xu, 2020). Establishing a BDA is strategically significant for improving ECC in data resources and services. Table 1 explains the main dimensions.

Data Mining Service Modeling

Once the service structure is determined, designers can learn according to the designed structure. A perfect organization system is needed to complete DM services in the context of Cloud Computing (CC). In other words, it establishes the corresponding relationships in various DM forms and model components, as shown below. First, it defines services and collects as much data information as

Figure 2. BDA Membership ECC Architecture

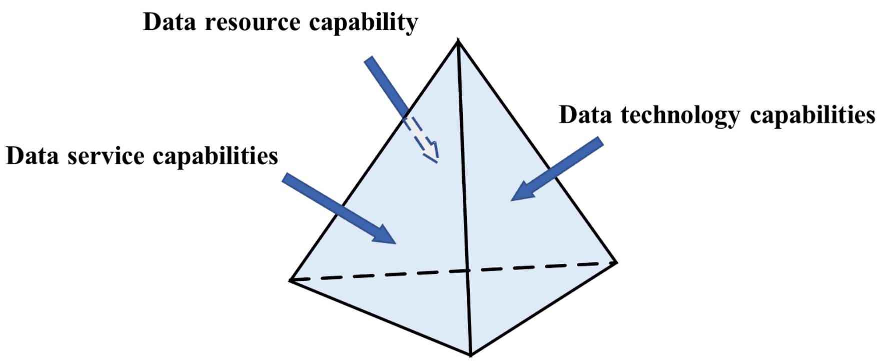


Table 1. Strategic Significance of Establishing Enterprise BDA

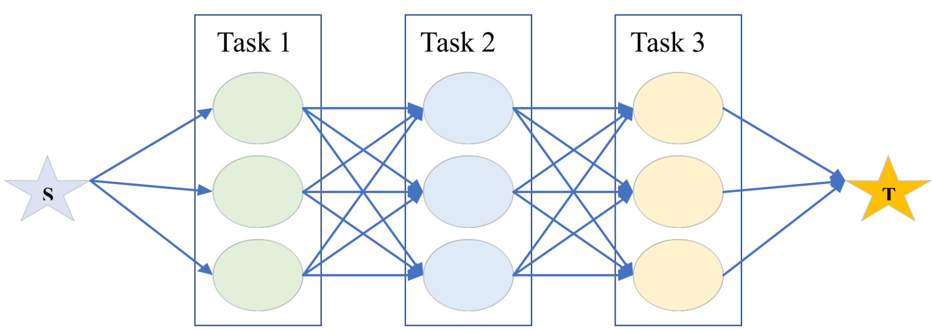
Dimension	Significance	Details
Data resources	Optimize internal resources	BDA has changed the traditional way of obtaining resources. Memberships learn from each other based on resource sharing to realize the integration effect of high added value to optimize their internal resources and improve their market competitiveness.
	Get external resources	BDA is initially established because enterprises strive to acquire scarce and valuable resources. BDA can integrate top resources, add value to data resources, and improve the ECC. Eventually, the competence of the BDA will be strengthened.
Data technology	Complementary technological advantages	By establishing a BDA and gaining strong technical resources, members can complement each other's technical advantages. Thereby, BDA avoids singular effort in technology R&D and alleviates repeated labor and resource wastes within the industry.
	Promote technological innovation	BDA gathers members' technology and knowledge resources to obtain more effective knowledge spillover effects. On this basis, members carry out the continuous innovation of relevant technologies, enhance the rate of technological innovation, and minimize the cost of technological innovation.
	Form technical standards	All member enterprises and relevant institutions participate in constructing technical standards and continuously improving them. BDA promotes the licensing and implementation of technical standards in the big data industry. It raises the technical standards at the industrial level and forms the technical standardization of the big data industry.
Data services	Reduce data service costs	BDA gathers enterprises with different core competencies in different big data industry chain positions. BDA greatly reduces the cost of data search and data transaction in service realization, thus reducing the service cost and improving the service competitiveness.
	Improve data service quality and efficiency	BDA complements the advantages of data, technical, and service resources. The service combination form saves the service cost and improves the service quality and efficiency.

possible that needs DM services. Then, it analyzes the specific core content of DM services, uses data support definitions to produce services, collects the data service description, and finally forms service candidate content and the service content in the service directory. Second, data services are bound. The data functions are formed after defining services to improve the DM services comprehensively. The information in the candidate and key services are built based on the service definition. Doing so forms countless links to provide comprehensive information for the DM services and obtains the associated data binding information in using the service information provisions. The third is to build component services. In the process of DM services, serving the bound data is also a way of defining services. Component services must complete the packaging services to form independent and complete service components.

In the context of CC, DM services must build a perfect data service architecture to ensure the DM service quality. The service architecture can be divided into infrastructure, virtualization, platform, and application. First, the infrastructure can provide a platform for computing resources, storing DM services, sharing network resources, and virtualizing DM services of terminal interfaces. Second, regarding virtualization, resources are usually distributed in the cloud background, and the virtualized resources are aggregated and uniformly assembled into a transparent and standardized form to lay the foundation for the centralized logic of service resources. Finally, management resources are formed to help encapsulated resources to be applied and developed in the platform. Third, the platform is the core service level in the context of CC. It undertakes the functions and services of analyzing and completing DM services. Fourth, the interface layer and terminal layer are unified at the application level to construct the application layer, complete DM services in the context of CC, and provide users with convenient access and use functions. At present, analyzing the DM service process in the CC environment can roughly indicate that the DM service is expressed through software. For example, the DM service process can be divided into three stages: data analysis, design, and development. The CC-based DM services begin by analyzing the information obtained. This process involves analyzing customers' actual needs and proposing related customer services. The second stage is the design stage, completing the corresponding design and improvement process according to user needs.

User needs must be formalized to understand users' personalized needs better and carry out targeted customized DM service solutions. When users submit personalized demand tasks, the BDA-oriented DM service platform will split the submitted service demand tasks according to the DM process. The tasks are divided into data collection subtasks, data processing subtasks, data analysis subtasks, and data interpretation subtasks. A task team of data resource-based, technology-based, and data application enterprises with matching service capabilities completes these tasks. This study uses the ant colony algorithm (ACA) for reference, employing its powerful search and solution ability to form task teams through the ability vector of alliance members. The core idea of ACA is to simulate the formation process of a DM customized service task team through the behavior of ants' food foraging, as shown in Figure 3.

Figure 3. A Process Model of the Customized Service Task Force



In Figure 3, the user needs are regarded as the starting point S for ants to find food. Several alliance members eligible for subtasks are combined into services (the formation process of task group A) to jointly complete the general task T of customized DM services. In this way, the DM customized service is to select eligible alliance members from S to the target T. These enterprises must have good non-functional attributes of services.

The basic idea of Apriori is to find out all frequency sets first, and the frequency of occurrence of these itemsets is at least the same as the predefined minimum support. Then, strong association rules are generated from the frequency set, which must meet the minimum support and minimum reliability. The above analysis suggests that the Apriori algorithm may generate a large number of candidate sets and may need to scan the database repeatedly, which are two major disadvantages of the Apriori algorithm. In this work, the Apriori algorithm is improved to avoid the shortcomings of the algorithm itself. The first improvement is based on the idea of partition, preprocessing and mining the data in different areas. In this way, the library does not need to scan all the data every time, and only the relevant data that meets the conditions are scanned, which greatly reduces the scanning time and speeds up the access speed. The second is to use similar search to classify the data. The results of the similar search can improve the association degree of the objects of the association algorithm and meet the requirements of the actual operation.

Zhonglong and Hongliang (2021), based on big data technology and aimed at the characteristics of data mining services, expanded and changed the traditional model and proposed the big data alliance data mining service process model. Using intelligent decision theory and the knowledge reasoning method, an intelligent service model with rapid response and reusable service was constructed to realize the scalability of data mining services (Al-Shourbaji et al., 2022). In this work, m is the total number of ants in the ant colony, $b_i\left(t\right)$ represents the number of ants located at element i at time t, $\tau_{ij}\left(t\right)$ is the amount of information on the path $\left(i,j\right)$ at time t, and $d_{ij}\left(i,j=1,2,\cdots,n\right)$ represents the distance between cities i and j. At the initial time, the amount of information on each path is equal, let $\tau_{ij}\left(0\right)=C$ (C is constant). During the movement of ant $k\left(k=1,2,\cdots,m\right)$, the transfer direction is determined according to the information on each path; $p_{ij}^k\left(t\right)$ represents the state transition probability of ant k from city i to city j at time t and has the relationship shown as Equation (1):

$$p_{ij}^k\left(t\right)=\begin{cases}\dfrac{\left[\tau_{ij}^k\left(t\right)\right]^\alpha\times\left[\eta_{ik}\left(t\right)\right]^\beta}{\sum_{s\subset\text{allowed}_k}\left[\tau_{is}\left(t\right)\right]^\alpha\times\left[\eta_{is}\left(t\right)\right]^\beta}, j\in\text{allowed}_k\\ 0\end{cases} \tag{1}$$

where $\text{allowed}_k=\left\{0,1,\cdots,n-1\right\}$-tabu k represents the set of cities that ant k is allowed to select in the next step. Unlike the real ant colony, the artificial ant colony system has a memory function. Experiments prove the proposed BDA-oriented DM service (Wu, 2020). $tabuk\left(k=1,2,\cdots,m\right)$ is used to record the cities that ant k has passed through currently, and the set $tabuk$ is dynamically adjusted with the evolution process. After n times, ants can walk all the cities and complete a cycle. Each ant's path is a solution. At this time, the amount of information on each path should be updated according to Equations (2) and (3):

$$\tau_{ij}\left(t+n\right)=\left(1-\rho\right)\times\tau_{ij}\left(t\right)+\Delta\tau_{ij}\left(t\right) \tag{2}$$

$$\Delta\tau_{ij}\left(t\right) = \sum\nolimits_{k=1}^{m} \Delta\tau_{ij}^{k}\left(t\right) \tag{3}$$

where $\rho \subset$ is the volatilization factor, and $1-\rho$ denotes an information residual factor.

$\Delta\tau_{ij}^{k}\left(t\right)$ stands for the amount of information left by ant *k* between cities *i* and *j* in this cycle. Its calculation is determined according to the calculation model. The most commonly used ant circle system model is shown in Equation (4).

$$\Delta\tau_{ij}^{k}\left(t\right) = \begin{cases} \dfrac{Q}{L_k} \\ 0 \end{cases} \tag{4}$$

Using the results of data mining, an alternative tabu table is formed. After each ant selection, the tabu table is modified to form a new tabu table until a new optimization scheme is formed.

The path selection strategy in the basic ant colony algorithm is adjusted. According to the confidence degree of the mining association rules as the probability of relevant alternative paths, the tabu table of alternative nodes is listed and the probability of each alternative path is calculated τ_{ij}^{k}. By selecting the path strategy of ant colony algorithm, a tabu table is formed to select and determine the next node to arrive. It can be seen that there is no randomness in the path selection of ants after adjustment.

Evaluation Index System for Data Mining Models

After model selection by the data technology-based enterprise, the service task group will get several effective DM models. According to the principles of comprehensiveness, balance, and applicability, this work constructs the evaluation index system of the BDA-oriented DM model. The index system has a two-tier structure: an overview layer and a detailed layer from the correctness, value, and cost of the model (He & Yin, 2021). Figure 4 presents the index details.

Then, the Accuracy, Precision, Recall, and F-value of clustering results are calculated by Equations (5), (6), (7), and (8):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

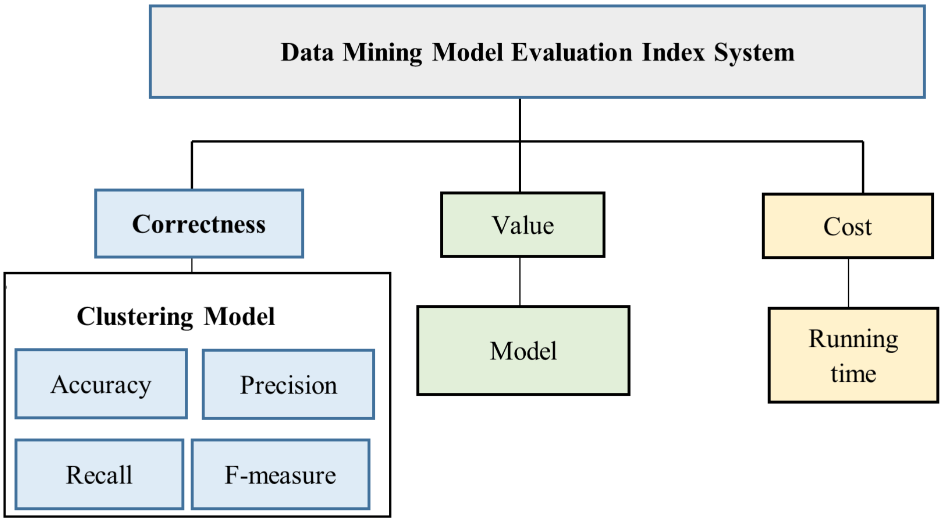
$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$F1 = \frac{2P \times R}{P + R} \tag{8}$$

Here, TP indicates that the positive class is determined positively. TN means the negative class is determined negatively. FP implies that the negative class is determined positively. FN suggests the positive class is determined negatively. P and R represent Accuracy and Recall, respectively (Du, 2020).

The promotion of the model is mainly evaluated from the promotion effect of detailed individual factors. Summing the promotion of detailed factors gets the model promotion while the response time evaluates the model's running time. Overall, the model's comprehensive evaluation can be counted by Equation (9):

Figure 4. BDA-Oriented DM Model's EIS



$$p = \alpha \text{Correct} + \beta \text{Value} + \lambda \text{Cost} \quad (9)$$

In (9), α, β, and λ are the weights of the parameters: Correct, Value, and Cost. This work sets α, β, and λ to 0.6, 0.3, and 0.1, respectively.

Table 2 lists the model parameter settings.

RESULTS AND ANALYSIS

Analysis of Experimental Results

The users of this experiment select a P2P (peer-to-peer) online credit company whose user groups are small and medium-sized firms (SMFs) and individuals (Guo & Liu, 2020). The company involves the long-term business of large, medium, and small online credit. Through the BDA-oriented customized

Table 2. Model Parameter Setting

Models	Logistic Regression Model (LRM)	Support Vector Machine (SVM) Model	Artificial Neural Network (ANN)
Parameters	penalty='l2', tol=0.0001, C=1.0, fit_intercept=True, intercept_scaling=1, class_weight=None, random_state=None, solver='warn', max_iter=100, multi_clas='warn', verbose=0, warm_state=False, n_jobs=None	C=1, loss='hinge', max_iter=10000	model_dir=None, n_classes=2, weight_column=None, label_vocabulary=None, optimizer='Adagrad', activation_fn=tf.nn.relu, dropout=None, input_layer_ partitioner=None, config=None, warm_start_from=None, loss_reduction=losses. Reduction.SUM

DM service, the company collects, mines, analyses, and forecasts the relevant information of the borrower. Professional software is used to comprehensively and scientifically assess the borrower's credit risk in a simple and convenient form. The LRM, SVM, and ANN models are selected in the DM model base to predict the borrower's credit risk. The effect of the model needs to be evaluated from Accuracy, Precision, Recall, and F-value. The confusion matrix of the three network models is shown in Figure 5.

This section evaluates the model's performance from the aspects of Accuracy, Precision, Recall, F-value, model improvement, and the total time of model operation. The evaluation results of each index are shown in Figures 6, 7, and 8.

Figures 6, 7, and 8 present the evaluation results of each index of the three models. Figure 6 compares Accuracy, Precision, Recall, and F-value; Figure 7 compares model improvement; and Figure 8 compares model response time.

Figures 6, 7, 8 show that the evaluation results of the Accuracy, Precision, Recall, and F-value of the LRM are 95.9, 97.7, 95.6, and 96.5, respectively. The improvement degree of the model is 20.52, and the running time is 332s. Therefore, according to the user's customized DM service task, the LRM is selected to predict the borrower's credit risk through model training and model performance comparison.

The above experimental results suggest that service realization should also build a cost control supervision mechanism to strengthen the control of service costs. The continuous deepening of BDA-oriented DM services has increased operation costs, collaboration, and communication. The increased cost factors are often considered in the final service product pricing, increasing transaction costs (Sun et al., 2020). Therefore, a service cost-controlling supervision team should be established in service realization. The cost composition and product pricing must be reviewed to improve the product pricing mechanism.

Figure 5. Confusion Matrix of Three Models

Note. a represents the LRM confusion matrix, b is the SVM confusion matrix, and c is the ANN confusion matrix.

Confusion Matrix		Predictive value	
		0	1
True value	0	5938	262
	1	148	3982

a

Confusion Matrix		Predictive value	
		0	1
True value	0	5879	321
	1	253	3877

b

Confusion Matrix		Predictive value	
		0	1
True value	0	5308	892
	1	115	4015

c

Figure 6. Comparison of Accuracy, Precision, Recall, and F-Value

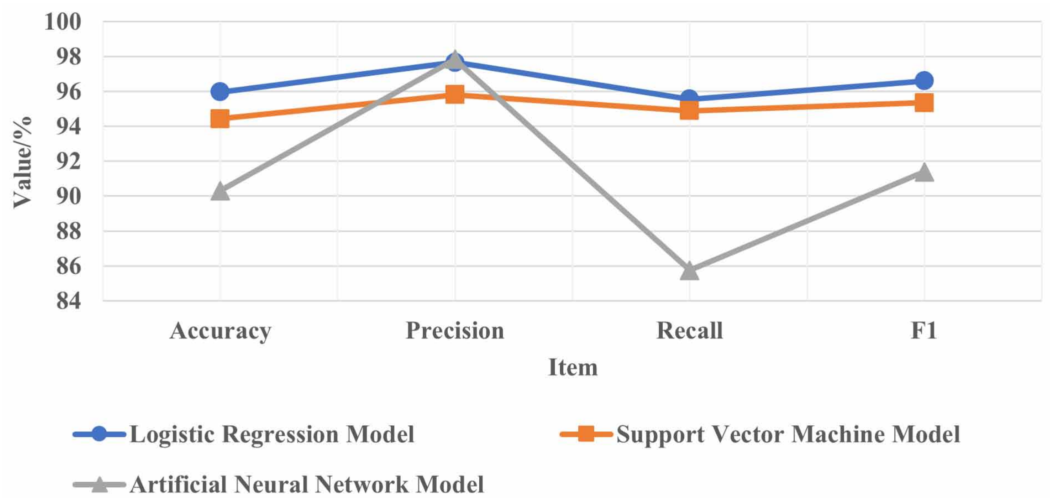
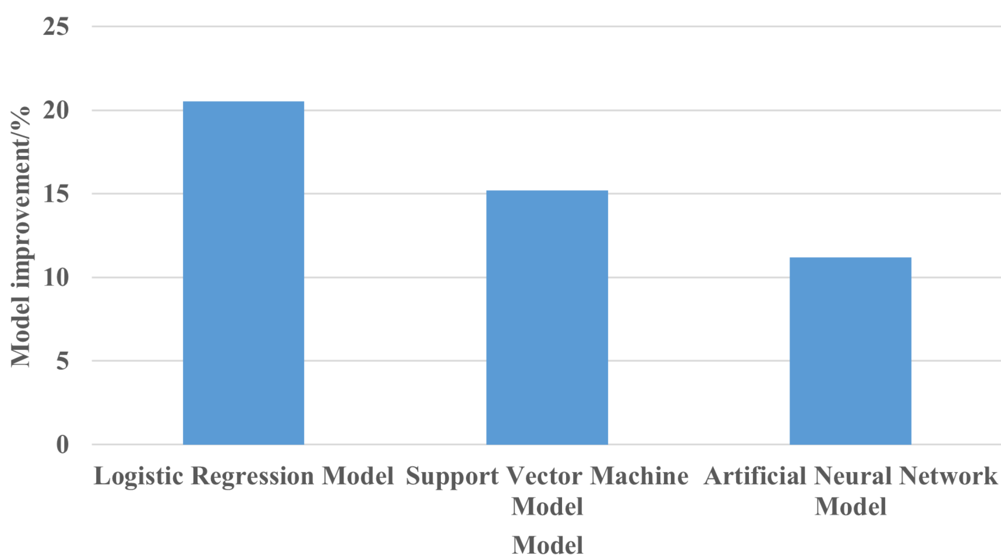


Figure 7. Comparison of Model Improvement

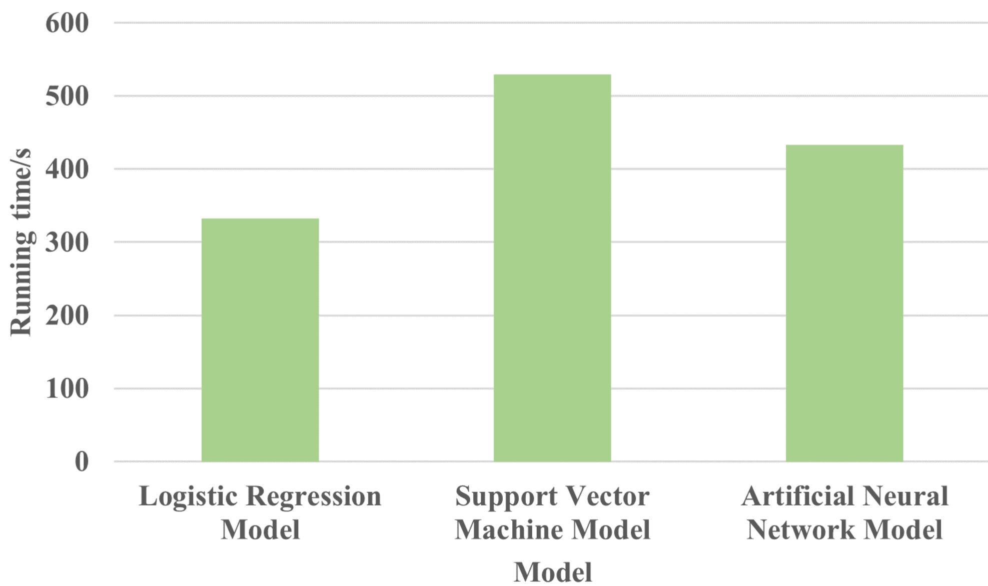


Analysis of Practical Applications

Although this study has achieved certain results, there are still some limitations that need to be considered. These limitations may affect the applicability and reliability of research results; therefore further in-depth exploration is needed in future research. Below, we will provide a detailed description of the limitations of this study in order to better understand the scope and applicability of the research findings.

- Data sample limitation: The data sample used in this study is limited and may not fully represent the real big data environment. Future research might consider increasing data samples to improve the universality and reliability of research results.

Figure 8. Comparison of Model Response Time



- Limitations of algorithm improvement: The improved algorithm applied in this article is designed based on specific data mining tasks and requirements and may have limited adaptability to other fields or tasks. Future research can further optimize and improve algorithms to meet the needs of different fields and tasks.
- Lack of practical application verification: The research in this article is mainly based on theory and simulation experiments, lacking practical application verification. Future research can combine practical scenarios to conduct practical application verification, in order to verify the effectiveness and feasibility of the proposed model in practical situations.
- Limitations of scalability: This study mainly focuses on data mining services for a single enterprise or organization, and there is no in-depth exploration of data sharing and mining for multiple enterprises or cross organizations. Future research can further investigate how to achieve resource sharing and scalability of data mining services among multiple enterprises or organizations.

The research results of this article are of great significance for practical applications. In the era of big data, enterprises need to obtain data quickly and accurately and conduct in-depth mining and analysis to optimize business decisions and improve efficiency. The resource sharing based data mining service model for BDA proposed in this article can help enterprises better utilize existing resources, improve data mining efficiency and accuracy, and further enhance their competitiveness. This model has a wide range of applications and can be applied to industries such as finance, healthcare, e-commerce, and logistics, meeting the diverse needs of different users. For example, in the financial field, this model can help banks identify potential risks and improve the accuracy of credit assessments; in the medical field, this model can help hospitals analyze case data, improve diagnostic efficiency and treatment effectiveness; in the field of e-commerce, this model can help e-commerce enterprises conduct user behavior analysis, improve sales revenue and user stickiness; and in the field of logistics, this model can assist logistics enterprises in route planning and resource scheduling, improving delivery efficiency and reducing costs. Therefore, the resource sharing based data mining service model for BDA proposed in this article has important practical applications, which can provide better data mining services for enterprises and further promote the development of big data applications. In the future, regarding the development of resource sharing based data mining service models for BDA, this article provides the following opinions:

- Intelligent optimization: In the future, more artificial intelligence technologies such as deep learning and reinforcement learning can be introduced to achieve intelligent optimization of data mining processes. By automating feature engineering, model selection, and parameter tuning, the efficiency and accuracy of data mining can be further improved.
- Privacy protection: With the increasing awareness of data privacy protection, future research can explore how to protect user privacy in data sharing and mining processes, such as differential privacy technology and federated learning, to balance the relationship between data sharing and privacy protection.
- Cross organizational cooperation: In the future, further research can be conducted on data sharing and mining cooperation models among multiple organizations, achieving resource sharing and scalability of data mining services in cross organizational scenarios and promoting broader data cooperation and innovation.
- Adaptive model: Based on the needs of different industries and application scenarios, future research and development of data mining models with adaptability can be carried out, which can be flexibly adjusted according to changes in different environments and tasks, improving the model's generalization ability and adaptability.
- Practical application verification: Future research can strengthen the integration with practical scenarios, conduct more practical application verification, verify the effectiveness and feasibility of the model in actual business environments, and promote the implementation and application of research results.

In summary, the future development directions include intelligent optimization, privacy protection, cross organizational cooperation, adaptive models, and practical application verification, in order to further promote the development and application of resource sharing based data mining service models for BDA.

CONCLUSION

This article aims to construct a data mining service model for BDA based on resource sharing mechanisms in order to meet the diverse needs of users and improve resource utilization and enterprise competitiveness. Through research on data mining services and the introduction of BDA, we have designed a customized data mining service model for BDA and applied an improved Apriori algorithm and an ant colony optimization algorithm for optimization. The experimental results indicate that the proposed data mining service model is scientifically reasonable and can effectively improve the efficiency and accuracy of data mining. By establishing a mathematical model based on actual data mining results, we have verified the scientificity and feasibility of the model. At the same time, we have improved the implementation strategy of the data mining pattern, further enhancing the application value of the model. This study is of great significance for the big data mining service model based on resource sharing. It provides guidance for enterprises on how to better utilize resources, meet user needs, and improve competitiveness when facing big data. In addition, the methods and models proposed in the study also provide references for the study of BDA-oriented data mining service models based on response surface models. The research in this article mainly focuses on data mining services for a single enterprise or organization, and there is no in-depth exploration of data sharing and mining for multiple enterprises or cross-organizations.

Future research can further investigate how to achieve resource sharing and the scalability of data mining services among multiple enterprises or organizations.

DATA AVAILABILITY

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

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING STATEMENT

This work was not supported by any funds.

ACKNOWLEDGMENT

The authors would like to show sincere thanks to those who have contributed to this research.

REFERENCES

Al-Shourbaji, I., Helian, N., Sun, Y., Alshathri, S., & Abd Elaziz, M. (2022). Boosting ant colony optimization with reptile search algorithm for churn prediction. *Mathematics*, *10*(7), 1031. Advance online publication. doi:10.3390/math10071031

Cui, Z., & Yan, C. (2020). Deep integration of health information service system and data mining analysis technology. *Applied Mathematics and Nonlinear Sciences*, *5*(2), 443–452. doi:10.2478/amns.2020.2.00063

da Rocha, B. C., & de Sousa, R. T. (2010). Identifying bank frauds using CRISP-DM and decision trees. *AIRCC's International Journal of Computer Science and Information Technologies*, *2*(5), 162–169. doi:10.5121/ijcsit.2010.2512

Du, Z. (2020). Energy analysis of Internet of things data mining algorithm for smart green communication networks. *Computer Communications*, *152*, 223–231. doi:10.1016/j.comcom.2020.01.046

Feng, Z., Huang, J., Tang, W. H., & Shahidehpour, M. (2020). Data mining for abnormal power consumption pattern detection based on local matrix reconstruction. *International Journal of Electrical Power & Energy Systems*, *123*, 106315. Advance online publication. doi:10.1016/j.ijepes.2020.106315

Fu, D., Hu, S., Zhang, L., He, S., & Qiu, J. (2021). An intelligent cloud computing of trunk logistics alliance based on blockchain and big data. *The Journal of Supercomputing*, *77*(12), 13863–13878. doi:10.1007/s11227-021-03800-w

Gastermann, B., Stopper, M., Kossik, A., & Katalinic, B. (2015). Secure implementation of an on-premises cloud storage service for small and medium-sized enterprises. *Procedia Engineering*, *100*, 574–583. doi:10.1016/j.proeng.2015.01.407

Ghoshal, A., Kumar, S., & Mookerjee, V. (2020). Dilemma of data sharing alliance: When do competing personalizing and non-personalizing firms share data. *Production and Operations Management*, *29*(8), 1918–1936. doi:10.1111/poms.12959

Guo, X., & Liu, Y. (2020). Intelligent traffic cloud computing system based on ant colony algorithm. *Journal of Intelligent & Fuzzy Systems*, *39*(4), 4947–4958. doi:10.3233/JIFS-179980

He, B., & Yin, L. (2021). Prediction modelling of cold chain logistics demand based on data mining algorithm. *Mathematical Problems in Engineering*, *2021*, 3421478. Advance online publication. doi:10.1155/2021/3421478

Karpagam, K. (2022). An improved k-means clustering algorithm for pattern discovery in data mining. *Journal of Positive School Psychology*, *6*(6), 2387–2395.

Pranata, B. S., & Utomo, D. P. (2020). Penerapan data mining algoritma FP-growth untuk persediaan sparepart pada bengkel motor (study kasus Bengkel Sinar Service). *Bulletin of Information Technology*, *1*(2), 83–91.

Rani, S., Gupta, D., Herencsar, N., & Srivastava, G. (2023). Blockchain-enabled cooperative computing strategy for resource sharing in fog networks. *Internet of Things : Engineering Cyber Physical Human Systems*, *21*, 100672. Advance online publication. doi:10.1016/j.iot.2022.100672

Saura, J. R., Ribeiro-Soriano, D., & Palacios-Marqués, D. (2021). Setting privacy "by default" in social IoT: Theorizing the challenges and directions in big data research. *Big Data Research*, *25*, 100245. Advance online publication. doi:10.1016/j.bdr.2021.100245

Shi, X., & Liu, Y. (2021). Sample contribution pattern based big data mining optimization algorithms. *IEEE Access : Practical Innovations, Open Solutions*, *9*, 32734–32746. doi:10.1109/ACCESS.2021.3060785

Sun, W., Tang, M., Zhang, L., Huo, Z., & Shu, L. (2020). A survey of using swarm intelligence algorithms in IoT. *Sensors (Basel)*, *20*(5), 1420. Advance online publication. doi:10.3390/s20051420 PMID:32150912

Watada, J., Roy, A., & Vasant, P. (2020). Preference identification based on big data mining for customer responsibility management. *International Journal of Intelligent Technologies & Applied Statistics*, *13*(1), 1–24. doi:10.6148/IJITAS.202003_13(1).0001

Wu, H. (2020). Data mining model based on improved ant colony algorithm. *Journal of Physics: Conference Series*, *1693*(1), 012103. Advance online publication. doi:10.1088/1742-6596/1693/1/012103

Xiao, S., & Wang, L. (2020). Analyzing problems of distribution transformer service areas based on customer-side data mining. *IOP Conference Series. Materials Science and Engineering*, *68*(6), 062033. Advance online publication. doi:10.1088/1757-899X/768/6/062033

Xiao, Z., Fu, X., Zhang, L., & Goh, R. S. M. (2019). Traffic pattern mining and forecasting technologies in maritime traffic service networks: A comprehensive survey. *IEEE Transactions on Intelligent Transportation Systems*, *21*(5), 1796–1825. doi:10.1109/TITS.2019.2908191

Xu, S. (2020). Research on anomaly detection method for hybrid big data subarea based on ant colony algorithm. *International Journal of Information and Communication Technology*, *17*(2), 164–177. doi:10.1504/IJICT.2020.108965

Xu, Z., Zhang, T., Zheng, R., Li, X., & Li, D. (2020). Knowledge discovery and analysis of ICT customer service data relying on multi-frequency mode data mining. *Journal of Physics: Conference Series*, *1648*(4), 042033. Advance online publication. doi:10.1088/1742-6596/1648/4/042033

Yu, Y. (2021). Design of mobile learning service platform based on data mining. *Journal of Physics: Conference Series*, *1757*(1), 012134. Advance online publication. doi:10.1088/1742-6596/1757/1/012134

Zhang, J., Williams, S. O., & Wang, H. (2018). Intelligent computing system based on pattern recognition and data mining algorithms. *Sustainable Computing : Informatics and Systems*, *20*, 192–202. doi:10.1016/j.suscom.2017.10.010

Zhonglong, Z., & Hongliang, W. (2021). The analysis of internet commercial judicial based on big data alliance and mining service process model. *Complexity*, *2021*, 1–17. doi:10.1155/2021/2897823

Zhou, W., Xia, X., Zhang, Z., & Wang, L. (2021). Association rules mining between service demands and remanufacturing services. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, *35*(2), 240–250. doi:10.1017/S0890060420000396