Journal of King Saud University – Science 34 (2022) 102041

Contents lists available at ScienceDirect

Journal of King Saud University – Science

journal homepage: www.sciencedirect.com

Original article

Establishment of big data evaluation model for green and sustainable development of enterprises

Dong Meiyou [a,b,*], Yao Ye [a]

[a] School of management, Wuhan University of Technology 430070, PR China
[b] Xianning Vocational and Technical College, Xianning 437100, Hubei province, PR China

ARTICLE INFO

Article history:
Received 10 January 2022
Revised 16 March 2022
Accepted 14 April 2022
Available online 22 April 2022

Keywords:
Big Data
Sustainable Development
Green Environment
Enterprises and Supply chain Management

ABSTRACT

Big data analytic is described as the complex process of investigating big data to reveal hidden data such as patterns, market trends, and correlation, and customer preferences. This might help to take important decisions. To analyze the big data from the industries we proposed a new optimized green supply chain management approach. This also mitigates the inherent risk that occurs due to hazardous materials. This includes emission of carbons and economic cost. Here we consider three scenarios. The former one is used to deal with the emission of carbon and mitigates the risk. The second scenario can be used to reduce both emission and risk and minimize the overall cost. The latter scenario can be used to reduce all three factors simultaneously. The findings and discussion are performed to showcase the advantages of the proposed method.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Big Data technology is a software tool designed to analyse, process, and extract information from very complex and massive data sets that standard data processing software is incapable of handling. Big data technology have already used by several enterprises to create business innovation and to promote future sustainable development of an enterprise [Zhang et al., 2021]. The application of big data technology in enterprises can promote it to a more efficient and scientific direction.Table 1.Table 2..

Big data is defined as "a collection of methodologies and technologies that requires specialized forms of integration in order to identify substantial hidden values from diverse, intricate, and large datasets." Big data is being used to help businesses improve their operations. Advanced big data technology enhances enterprise positioning accuracy. Big data enables businesses to share data more openly, rapidly, and comprehensively, and diversified data may be supplied directly to positions requiring information. Enterprises will examine and aggregate data generated by the integration of big data resources, and this data will be used as a one-of-a-kind resource in the company innovation process.

A sustainable environment mainly satisfies the needs of the present generation without sacrificing the needs of the future generation [Palstam et al., 2021]. Big data will have a greater impact on any company's long-term growth [Hassani et al., 2021]. It offers for a better understanding of a variety of situations. Big data has the potential to alter how long organisations – particularly those with the biggest environmental impact but having access to vast quantities of data – can take sustainable action.

Understanding how business affects the natural environment can lead to fresh ideas for achieving long-term success in a company. Without accurate information it is difficult to predict the growth of enterprise. Establishing big data evaluation will give an opportunity to gain a deeper understanding about the factors which are important for the sustainable development of enterprise [Kar, and Dwivedi, 2020]. Green enterprise encompasses a wide range of products, services, and occupations aimed at lowering greenhouse gas emissions, assuring sustainability, and postponing climate change [Rahman et al., 2020]. Renewable energy, transportation, construction, agriculture, biotechnology, recycling, forestry, and waste management are among the industries covered by the Green enterprise. Any organization's "greening" or adoption of environmentally sensitive practises is sometimes referred to as a "green enterprise.".

China has showed vast improvement in the field of enterprise environmental management after the acceptance of the ISO14001

* Corresponding author.
E-mail address: dongmy2021@163.com (D. Meiyou).

Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.jksus.2022.102041
1018-3647/© 2022 The Author(s). Published by Elsevier B.V. on behalf of King Saud University.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Table 1
Indices.

Indices	Explanation
DC	distribution center set
RC	recycling center
ML	Municipal landfill
MK	Key manufacturer
S	Suppliers set
PC	Potential consumer set
PI	Incineration plant

Table 2
parameter settings.

Input parameters	Explanation
D_k	The demand of raw materials for the key manufacturing in kg
D_c	Consumer demand's in kg
$T_{ff'}$	Cost for transporting the unit from f to f'
C_r	Cost for purchasing the raw materials from the supplier
f'	$f' \in \{S, DC, RC, PI, ML\}$
UCP	Processing cost per unit $f \in \{S, DC, RC, PI, ML\}$
CF	Capacity that included in facility $f \in \{S, DC, RC, PI, ML\}$
TC	Transportation capacity
I_{RCa}	Risk control investment while the level of the risk is a
I_{CEl}	Carbon emissions are reduced when the emission level is equal to l

environmental standard[Wang etal., 2020]. To optimize the supply chain (SC), more enterprises have adopted for the green and sustainable approaches to enhance the enterprise operations. Mainly government agencies focus on the energy output rather than the effective use of energy. It is important to adopt big data evaluation and green practices in the enterprises for the green and sustainable growth. Integrating big data in the enterprise plays an important role in solving asymetrical information and insufficient management abilities in the enterprise (Dani et al., 2021; Li 2021). The main contributions of this work is delineated as follows:

- The modified k-means clustering algorithm is used to identify the critical points prone to failure, increased energy consumption, and energy leakages. This methodology is mainly aims to enhance the energy efficiency and offer fault tolerance by identifying the enterprises that consumes more energy.
- In the enterprise, big data is used to execute tasks such as data extraction, pre-processing, modeling, analysis, and the application of machine learning approaches to large data collected from businesses.
- The use of big data in enterprises advances the field of enterprise technology while also assisting in the creation of a green and sustainable development of businesses by reducing energy consumption and carbon emissions.

The rest of this paper is arranged accordingly. Section 2 presents the litearature review and section 3 elaborates the background of big data storage and green computing. Section 4 presents the result findings and both the advantages and limitations of this technique is discussed. Section 5 concludes the paper.

2. Literature review

Muhammad Waqas et al. [2021] evaluated the influence of Big Data Analytics (BDA) on boosting Competitive Advantage (CA), Green Innovation (GI), an EP in the Chinese manufacturing industry. According to the conclusions of this study, Big Data Analytics aids in attaining CA and EP. Xuehong Zhu et al. [8] developed a theoretical framework to explain how the governance transformation model affects green process innovation (GPI) in Chinese iron and steel companies (CISEs). The goal of this research was to see how environmental regulation (ER) and governance changes affected the TFEGE of key CISEs. The outcomes of the study indicated that ER and TFEGE have a U-shaped relationship, that GPI and Governance reform can considerably improve TFEGE.

Jinhua Dong et al. [2021] devised a new technique to evaluate and choose state-owned firms in China with sustained high-quality development capability in a linguistic distribution assessment (LDA) environment (SHQDC). The feasibility of the technique was demonstrated by comparing the sustained high-quality development capabilities of three different organizations owned by china.Raut et al. [2021] investigated the role of Big Data Analytics (BDA) as a mediator between' sustainable supply chain business performance' and other key factors such as social practices, lean practises, organisational practices, etc. The study addressed challenges in LARG practice management in order to contribute to a sustainable supply chain. Yubing Yu et al. [2021] developed a conceptual model to explain how corporate green management, supply chain relationship capital, and financial performance are linked. According to the findings of the study, supply chain relationship capital has a positive influence on corporate green management, which enhances financial performance for enterprises undergoing Industry 4.0′s digital transformation.Sachin et al. [2021] investigated the role of "Big Data Analytics" in mediating the relationship between "Project Performance" and nine variables, including environmental technologies, green purchasing, project operational capabilities, project knowledge management, collaboration, and explorative learning, top management, social responsibility, project complexity, and project success. According to the study's findings, project operational capabilities, green buying, and project knowledge management all necessitate the use of big data analytics as a bridge-builder. Big Data Analytics adoption has a positive influence on project performance in the industrial unit.

3. Theoretical background

3.1. Green and sustainable companies in china industry

Green IT is made up of two words: green (which means "environmentally friendly") and IT (which stands for information technology) (IT). Green IT is defined as an environmentally beneficial solution to IT problems, such as resolving environmental challenges through the recycling of IT equipment trash. Green IT 1.0 is the name given to it. It is also acknowledged as a means for IT to aid in the settlement of ecological problems. The phrase for it is Green IT 2.0. The Green data centers, IT energy management, storage capacity optimization, localized cooling, managed printing services, IT asset disposal, and recycling services are the applications that belong to Green IT 1.0. The Smart Grid, Intelligent Transport System, Remote working, paperless working, Teleconference, Eco-Friendly supply chain management, and Building Energy Management System.

3.2. Data acquisition

Big data acquisition is based on four parameters such as variety, volume, value, and velocity. Moreover, the preprocessing of real-time data encloses live streams of data. The processing of real-time data is made on its arrival or sometimes it takes a small period of time for buffering and thus provides real-time analysis.

3.3. Data preprocessing

It involves three steps namely identification, extraction, and cleansing. Identification is very important to do further operations. Meanwhile, extraction is performed to convert the intricate data into a simple and structured form. This will help for rapid analysis. Besides, cleaning is performed to remove irrelevant documents from the big data. This might have helped to extract the required features from the big data.

3.4. Storage and management technology for the big data

For the management of big data and its transference, it requires memory storage and the establishment of the respective database. Thus importance has been given to the management and processing of technology for the unstructured, semi-structured, and complex structured big data. Hence it is essential to overcome some issues like big data presentation, reliability, storage, transmission, and affective processing [Acharjya, and Ahmed, 2016]. Therefore the designing of the flexible distributed file system (DFS), computing of storage, redundancy elimination, optimized energy-efficient storage, effective big data storage technique, data fusion technology of heterogeneous type, breakage of distributed uncorrelated big data management and processing methodology, organization of big data technology, modeling of research data, indexing techniques, and so on.

The database technology [Selinger, 1987] is of three types namely non-relational databases, relational databases, and database caching systems. relational databases are of two types that is a conventional relational database system and a NewSQL database. Meanwhile, the non-relational database is termed as NoSQL database and categorized into picture storage database, key database, document database, and column database.

Moreover, the development of security in big data technology is also necessary. This can be performed by transparent encryption, data destruction technology, data auditing, and data access control. Besides, the enhancement of data verification, privacy protection and inferential management, data authenticity detection, and forensics were also made.

3.5. Data analysis and mining technology for energy consumption

The big data from the industrial environment receives data based on the emission of energy level which can be sued to derive the potentially useful data along with the knowledge. However, these data are noisy, implicit, fuzzy, random, and incomplete. Based on the applications the data mining is categorized as data summary, predictive models, association rules discovery, clustering, the discovery of exceptions, and discovery of sequential patterns. Moreover, based on the objects of mining they are classified as, spatial databases, relational databases, text data, heritage databases, multimedia databases, temporal databases. However, by following the methods of mining they are categorized as statistical methods, neural networks, machine learning approaches, and database methods. Statistical methods are of type's discriminant analyses, regression analyses, exploratory analyses, and cluster analyses. The neural network includes self-organizing neural networks, forward neural networks. Machine learning approaches are of type's genetic algorithms models, inductive learning approaches, and case study approaches.

3.6. Modified K-means clustering for the estimation of industrial energy efficiency

The internal connections in the unstructured dataset can be analyzed by the clustering approach which classifies the same featured data. K-means algorithm is considered as the most emphasized and unsupervised clustering approach and is of different types. Here we utilize a modified algorithm to analyze the energy efficiency of industrial big data. Some of the types of modified k-means clustering algorithms are spherical means clustering [Duwairi, and Abu-Rahmeh,2015], fuzzy c-means clustering [Wetzels et al., 2009], k-medoids [Bezdek,1984], and so on.

3.7. Model hypothesis

Our proposed supply chain network contains raw inventory management, distribution centers, and enterprises for key manufacturing. Also, general disposal centers such as recycling, landfill, and incineration plants are presented in the life cycle of our approach. Fig. 1, illustrates the full life cycle proposed green supply chain network. The simplification of our work can be made with the help of certain hypotheses as shown below,

- Find the position and capacity of the suppliers of raw materials, distribution centers, and key manufactures and also the route for transporting it.
- Consider the supply chain is for single use and not stored for future use.
- The products that are manufactured and refurnished are considered with the same priority.
- Equal priority is given to all suppliers irrespective of the quality, brand, and so on.

3.8. Survey and data collection

Our research focused on the role of Big Data Evaluation in achieving green and long-term organizational performance. The survey's participants are from businesses that use Big Data technology and Green practices. Big data technology and green business were familiar to the responders and they choose the convenient sampling strategy. Using Gail, a total of 520 people were contacted and invited them to take part in the poll. The study's sample was made up of employees in the China industry. For the study, a questionnaire was created. The goal of the research was mentioned in the cover booklet that came with the questionnaire. Usable return questionnaire for this study is 210.

3.9. Measures

The components of the conceptual model are formulated using multiple scale items. In addition to the demographic variables, the conceptual study model includes 51 components. Green supply chain cooperation improves supplier selection by including environmental issues and involving suppliers in product development and production. Green product innovation includes environmentally friendly materials, eco-labelling, and materials that are simple to discard, reuse, and decompose. Green process innovation promotes the use of fewer natural resources in the manufacture of commodities. It also involves reusing, recycling and remanufacturing resources. As per environmental criteria redesigning of product and services. Environmental performance involves air emission decrement, hazardous waste decrement, decrement in consumption of fuel, environmental obiedience.

3.10. Parameters settings with their definitions

The parameter settings are made in this step as shown below. The respective indices are shown in table 1. However, the parameters for the above indices are shown below, table 2.

Fig. 1. full life cycle of the proposed green supply chain network.

3.11. Inherent risk mitigation

The green supply chain is composed of associate risks such as casual risks, environmental pollution risks, loss of property risk due to the hazardous materials [Park, and Jun 2009]. These risks are caused due to accidents in the green supply chain [Zhao et al., 2017] and can be given as,

$$W_i = \mathrm{Prob}_i \sum_{p=1}^{3} M_{ip} \quad (1)$$

The risk factors in the supply facility are given as W_i and the failure events probability can be given as Prob_i and the consequences occur due to the risk of p can be given asM_{ip}. The risk of transportation during the green supply chain can be given as,

$$W_{ij} = \mathrm{Prob}_{ij} \sum_{p=1}^{3} M_{ijp} \quad (2)$$

$$\mathrm{Prob}_{ij} = Xr_{ij}.K_{ij}.R_{ij} \quad (3)$$

The path length from I to j is given as K_{ij} and the transported products can be indicated as R_{ij} and the rate of accident occurs during the transportation is given asXr_{ij}.

Meanwhile the mitigation of inherent risk by using the objective function as,

$$O = \min(W_f + W_t) \quad (4)$$

Here risk in all the facilities that are from suppliers to the disposal center is given asW_f. And the transportation risk is given asW_t.

4. Findings and discussion

The system's performance should be assessed by determining if it will provide proper, dependable operations throughout peak working hours. Following the test, a performance evaluation will be conducted to analyze the system's current performance, predict future performance, and identify the system's bottleneck, among other things. Testing tools are used to assess performance by allowing several users to use the system at once and monitoring the server's response time. To analyze the performance of the big data precision system, Apache JMeter is utilized as a testing tool. The results of this Apache JMeter recording of the server's response to several concurrent users are displayed below.

Fig. 2(a) & (b) shows the test results recorded by the Apache JMeter testing tool. On obtaining the test results we can come to know that the Big Data Enterprise system can be applied in real-time enterprises. The big data enterprise functions and performance meet both the enterprise application requirements and business requirements. The functions and applications are tested and the results are given as screenshots. The result related to the performance is shown in the graphs. Finally, we conclude that the system proved to meet the business applications requirements of the enterprise.

Based on the methods [Szucs, D. and Ioannidis, 2017], and [Wetzels,etal 2009] the relation between the external CSR and respective performance of big data analytic is illustrated in Fig. 3. When the big data analytic capability will move from low to high and it is represented by sharp values.

Table 3 presents the optimal option in terms of risk factors, carbon emissions, and overall cost for the selected scenarios. All scenarios have pros and cons in terms of the appropriate green SC network. Scenario 1 has the lowest risk factor in the chart, but the cost is higher. In the second scenario, we can notice a decrease in carbon emissions and an increase in the risk factor. As a result, selecting the appropriate scenario is crucial for increasing stakeholder satisfaction.Table 4..

To analyze the environmental impact the authors utilize monitoring, emergency rescue, cost of emergency treatment, and so on. however, the impact of the green supply chain based on the monetary values is estimated and shown in table 4.

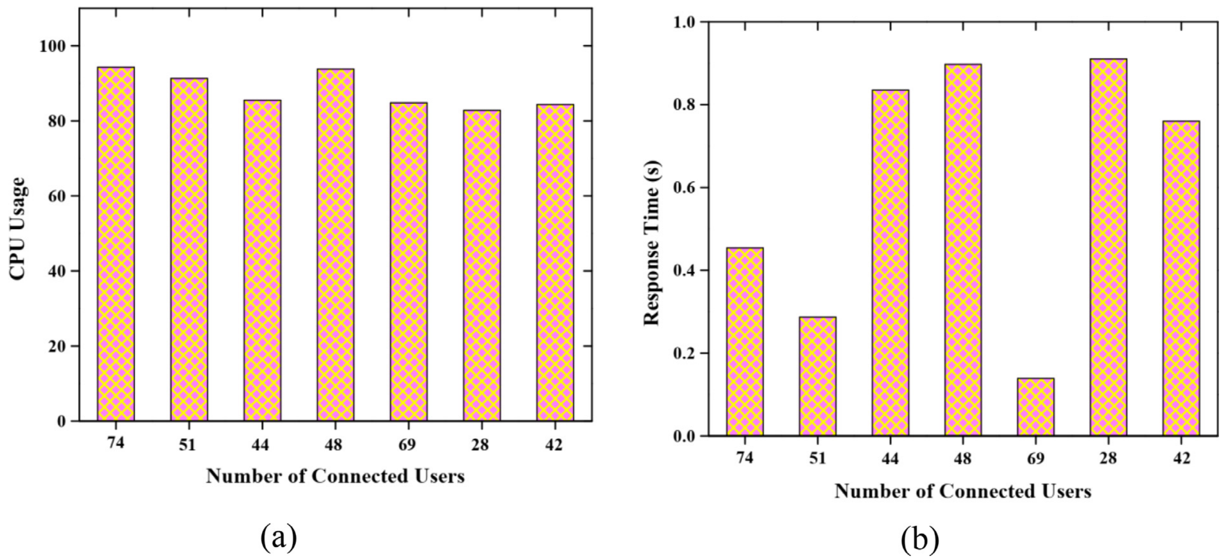

Fig. 2. Performance evaluation using CPU usage and response time for different numbers of users. (a) CPU Usage and (b) Response Time.

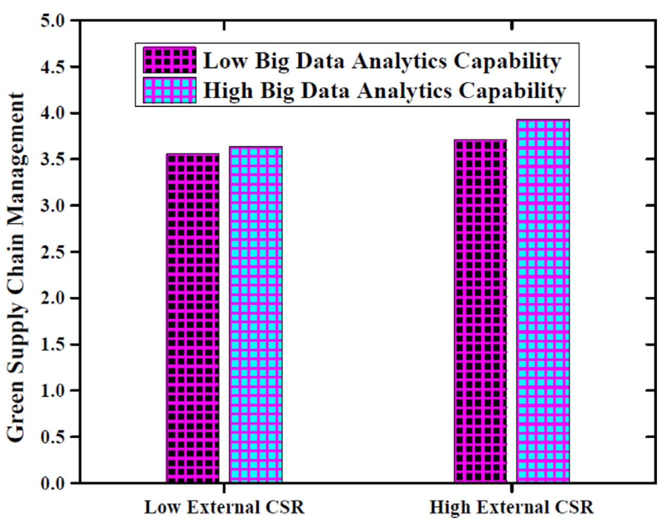

Fig. 3. performance analysis based on the external corporate social response and the green supply chain management.

Table 3
Objective function results correlated to various scenarios.

Scenarios	Carbon emissions	Inherent risk	Total cost
1	41.67	3498.23	1.19
2	55.37	3512.45	0.48
3	55.37	4178.34	0.31

Table 4
Evaluation of enviromental impact using monetary values.

Facility	Consequence
Key manufacturer A	**0.58**
Incineration plant	9.12
Supplier D	0.27
Supplier B	5.12
Recycling center	3.25
Supplier C	0.97
Transportation	0.67
Landfill	10.23

5. Conclusion

The article in this paper is based on the multi-objective optimized work which utilizes the modified k-means clustering to achieve energy efficiency in the industrial green supply chain. The optimized model hypothesis was made to circumvent the inherent risk caused due to the hazarduous materials. These hazarduous materials were acquired due to the carbon emission in the atmosphere. Environmental pollutions also cause an inherent risk. Three scenarios were considered to analyze the three factors such as economic cost, risk, and carbon emission. These were analyzed and listed in the findings and discussion sections.

Conflict of Interest statement

The authors declared no conflict in this manuscript and publications.

References

Acharjya, D.P., Ahmed, K., 2016. A survey on big data analytics: challenges, open research issues and tools. International Journal of Advanced Computer Science and Applications 7 (2), 511–518.
Dani, R., Juyal, D., Rawal, Y.S., 2021. A Critical Analysis of the Restaurant Industry's Effect on Environment Sustainability. SPR. 1 (4), 435–442.
Duwairi, R., Abu-Rahmeh, M., 2015. A novel approach for initializing the spherical K-means clustering algorithm. Simulation Modelling Practice and Theory 54, 49–63.
Hassani, H., Huang, X., MacFeely, S., Entezarian, M.R., 2021. Big data and the united nations sustainable development goals (UN SDGs) at a glance. Big Data and Cognitive Computing 5 (3), 28.
Kar, A.K., Dwivedi, Y.K., 2020. Theory building with big data-driven research–Moving away from the "What" towards the "Why". International Journal of Information Management 54, 102205.
Li, Z., 2021. Treatment and Technology of Domestic Sewage for Improvement of Rural Environment in China-Jiangsu: A Research. SPR. 1 (4), 355–364.
Palstam, A., Andersson, M., Lange, E., Grenholm, A., 2021. A Call to Include a Perspective of Sustainable Development in Physical Therapy Research. Physical Therapy 101 (3), p.pzaa228.
Park, H.S., Jun, C.H., 2009. A simple and fast algorithm for K-medoids clustering. Expert systems with applications 36 (2), 3336–3341.
Rahman, M., Billah, M.M., Hack-Polay, D., Alam, A., 2020. The use of biotechnologies in textile processing and environmental sustainability: An emerging market context. Technological Forecasting and Social Change 159, 120204.
Selinger, P.G., 1987. Database technology. *IBM systems journal* 26 (1), 96–106.
Szucs, D., Ioannidis, J., 2017. When null hypothesis significance testing is unsuitable for research: a reassessment. Frontiers in human neuroscience 11, 390.

Waqas, M., Honggang, X., Ahmad, N., Khan, S.A.R., Iqbal, M., 2021. Big data analytics as a roadmap towards green innovation, competitive advantage and environmental performance. Journal of Cleaner Production 323, 128998.
Wetzels, M., Odekerken-Schröder, G., Van Oppen, C., 2009. Using PLS path modeling for assessing hierarchical construct models: Guidelines and empirical illustration. MIS quarterly, 177–195.
Yu, Y., Zhang, J.Z., Cao, Y., Kazancoglu, Y., 2021. Intelligent transformation of the manufacturing industry for Industry 4.0: Seizing financial benefits from supply chain relationship capital through enterprise green management. Technological Forecasting and Social Change 172, 120999.
Zhang, W., Yang, C., Cheng, Y., Chen, H. and Wang, W., 2021. Research on the Mechanism of the Sustainable Development Model of Enterprises Based on Big Data Analysis Model. *Mobile Information Systems*, *2021*.
Zhao, R., Liu, Y., Zhang, N., Huang, T., 2017. An optimization model for green supply chain management by using a big data analytic approach. Journal of Cleaner Production 142, 1085–1097.