Big Data Challenges in Retail Sector: Perspective from Data Envelopment Analysis

Praveen M. Kulkarni, Prayag Gokhale, and Padma S. Dandannavar

1 Introduction

Technology advancement has provided an opportunity to the organizations to collect data related to customers for providing better services to the customers [1]. The data generated through various methods of technology supports in creating competitive advantage for the organization [2].

Research conducted by Aktas Emel and Meng Yuwei, 2017 [3], mentions that there would be an increase in the application of big data in the retail sector by 60% by 2025.

Although big data management implementation would increase in the retail sector, there are challenges in implementing big data applications in the retail sector [4].

Studies related to Indian retail sector show that Indian retail market is expected to reach $1.3 trillion by 2025; this demands for implementation of latest technology such as big data management in understanding the customer's [5].

This is evitable as Indian retail sector is experiencing tremendous growth in both demographically and economically in India [5]. Further, the online retail market has

P. M. Kulkarni (✉)
KLS Institute of Management Education and Research, Hindwadi, Belagavi, India

P. Gokhale
KLE Dr. M. S. Sheshgiri College of Engineering and Technology, Department of MBA, Udyambag, Belgaum, India

P. S. Dandannavar
KLS Gogte Institute of Technology, Department of Computer Science and Engineering, Udyambag, Belgaum, India
e-mail: padmad@git.edu

A. Haldorai et al. (eds.), *5th EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing*, EAI/Springer Innovations in Communication and Computing, https://doi.org/10.1007/978-3-031-28324-6_8

grown from $6 billion in the year 2015 to $70 billion in the year 2019, this indicates that offline retailing and online retailing has increased in India [5].

However, there are challenges related to implementing big data in retail sector; they are (a) data privacy, (b) data credibility, (c) data analysis manpower, (d) management culture of data management, (e) data security, (f) top management support, (g) data decision making, (h) affordability and operational cost, (i) technology infrastructure, and (j) data related to specific customers [2, 6, 7].

In relation to the research with regard to the challenges of implementing big data in retail sector in India, there is limited information to the body of knowledge related to big data management in retail sector of India [8, 9]; hence, this study is undertaken to understand the challenges witnessed by the retail sector in India in implementing big data technology.

The present study has applied data envelopment analysis method which is a linear programming methodology which numerically shows the efficiency of different entities in the study [10, 11]. Hence, the present study has adopted this method of analysis to understand the various level of challenges with regard to implementation of big data management in the retail sector of India.

2 Literature Review

In this section, the study presents the information with regard to the literature review from the perspective of big data in retail management.

The present literature review would consist of two sections: firstly, information related to big data in the retail sector, and secondly, challenges of implanting big data in the retail sector of India.

2.1 Big Data in the Retail Sector

Big data and retailing are almost related, as we find data related to stores which sell thousands of products and have billions of financial transactions [12]. For instance, Walmart has a business operation in more than 10,000 stores in 20 countries and more and have service capacity to offer products to more than 30 million customers on everyday basis through its network of stores across the globe [13].

Further, from the perspective of data from the customers in the retail sector show that large amount of data is produced with regard to purchase, buying behavior, and volume of transactions [14].

These data augment information related to inventory management, supply chain management, and providing offers and discounts to the customers.

For instance, predictive analytics provides information on the real-time data for preparing the stock of products in the retail stores, which can reduce the cost of inventory in the stores [15].

Still, there are considerable areas of development in the retail sector in relation to big data, even though this sector is on forefront for creating and implementing big data in the sector. However, there are areas of development for implementing this technology in retail sector. For instance, there are fewer organizations who have the capability to gather and analyses the data and take complete benefit of data analyzed through big data technology for retail sector.

Further, larger organizations are not eager to invest at the level that would be appropriate to benefit of big data and there are scuffle to gain actionable consumer insights from the increasing data availability from the perspective of retailers, customers, and supply chain management [16].

These explanations propose that retailing, for both practitioners and academicians, is at the epicenter of a storm of big data prospects and challenges, which demands additional work on how to derive extra value from big data.

2.2 Challenges of Implementing Big Data in Retail Sector

Development in the area of information technology and reduced cost of data collection have provided an opportunity to the organizations to integrate data for business development, especially for the retail sector.

This abundance of data supports the retail sector to develop competitive advantage by understanding the retail buying behavior and mapping the future business development in the retail sector.

Big data technology can collect data and also have the capability to trace the customers buying behavior, however the major concern level of data privacy and integration of data for the gaining meaningful insights about the customers.

Another, challenge with regard to implementation of big data technology in the retail sector is the data credibility; this related to with regard to quality of data which is applied for the process of data analysis.

Another, factor associated with big data analysis is the lack of shortage of talented workforce to conduct the analysis of the data and provide meaningful information for decision makers in the retail sector.

One of the most critical observations is with regard to the development of management culture to integrate big data culture for improving customer satisfaction and management practices in the retail sector.

The study is from the perspective of academic research; however, future research from the marketing and big data in the retail sector needs deeper understanding and provide greater quantitative analysis. The study analysis would benefit the business practices and enhance higher customer satisfaction in retail sector of India.

3 Research Methods

The Data Envelopment Analysis (DEA) procedure presented by Abraham Charnes and contemporaries estimate an efficiency frontier by bearing in mind the top performance observations (extreme points) which "envelop" the residual observations by means of mathematical programming techniques. The notion of efficiency can be demarcated as a ratio of outputs to inputs:

$$\text{Efficiency} = \frac{\text{Outputs}}{\text{Inputs}} \tag{1}$$

With the intention that an inefficient unit can turn out to be efficient by increasing products (output) maintaining the equivalent level of employed resources, or by decreasing the used resources and retaining the equivalent production level, or by a blend of both $j = 1, 2, 3, m$, DMUs through $x_i \mid i = 1, 2, 3, n$ inputs to result in $y_r \mid r = 1, 2, 3,$ outputs and multipliers v_i and u_r allied with inputs and outputs, the efficiency expression presented in (1) can be validated as the ratio between weighted outputs to inputs:

$$\text{Efficiency} = \frac{\sum_{r=1}^{s} u_r y_{jr}}{\sum_{i=1}^{n} v_i x_{ji}} \tag{2}$$

In Charnes et al. [17], the degree for the technical efficiency and the multipliers, for an explicit DMU, is appraised by explaining the fractional programming:

$$\max \frac{\sum_{r=1}^{s} u_r y_{jr}}{\sum_{i=1}^{n} v_i x_{ji}} \mid \sum_{r=1}^{s} u_r y_{jr} - \sum_{i=1}^{n} v_i x_{ji} \leq 0 \tag{3}$$

With i, j, r, and positive v_i, u_r. The problem denominates the CCR "constant return to scale input-oriented model," which in contrast is equal to explaining the subsequent linear programming:

$$\min(\theta) \mid \sum_{j=1}^{m} z_j x_{ji} \leq \theta x_{oi}; \sum_{j=1}^{m} z_j y_{jr} \geq y_{or}; \sum_{j=1}^{m} z_j = 1; z_j \geq 0 \tag{4}$$

Consequently, an efficiency score θ varies from 0 to 1 entitling the efficiency for respective DMU. Peripheral influence of each input and output can obtained in the "Multiplier model of (3)," the peers of efficiency and particular weights in the envelopment form of (4) and similarly the probable for enhancements and slacks in an extension form of (4).

4 Results

Data envelopment analysis can be applied to a linear programming-based procedure and optimizing to evaluate the efficiency of respective unit. By means of refining the efficiency of respective unit, a reference set for an inefficient unit is obtained and the efficiency of several units can be equated to the efficiency frontier.

4.1 Project Specifications

In this study, data privacy, data credibility, data security, technology infrastructure, customer data, and cost of data are the part of decision-making unit (DMU), the was evaluated with respect to organizational culture as input variable and decision-making process and sales as output variables. The DEA form adopted in this study is the model basic radial grounded on the model constant return to scale.

4.2 Efficiency

The efficiency value found by the defined model is shown in Table 1. Furthermore, to the value of efficiency, its type is also be made known in Table 1.

4.3 Reference Set

In every DEA, the resulting method attempts to enhance the efficiency of the target unit to the maximum. This exploration process will end when either of efficiency of target unit or one or more units is = 1. Hence, for an ineffective unit minimum one other unit should have the efficiency = 1, with the identical weightages of the target unit are attained from the result of the model. Therefore, these efficient units are identified as the peer group for the inefficient unit. The peers are illustrated in Table 2.

Table 1 Efficiency analysis through DEA

Variables	Efficiency	Result
Data privacy	0.311	Inefficient
Data credibility	0.137	Inefficient
Data security	1	Efficient
Technology infrastructure	0.479	Inefficient
Customer data	1	Efficient
Cost of data	1	Efficient

Table 2 References

Parameters	Peer1	Peer2	Peer3
Data privacy	Data security	Cost of data	–
Data credibility	Data security	Customer data	Cost of data
Data security	Data security	–	–
Technology infrastructure	Data security	Cost of data	–
Customer data	Customer data	–	–
Cost of data	Cost of data	–	–

Table 3 Peer frequencies

Parameters	Frequencies
Data security	4
Customer data	2
Cost of data	4

Table 4 Lambdas

	Data privacy	Data credibility	Data security	Technology infrastructure	Customer data	Cost of data
Data privacy	0	0	0.143	0	0	0.448
Data credibility	0	0	0.026	0	0.031	0.165
Data security	0	0	1	0	0	0
Technology infrastructure	0	0	0.318	0	0	0.223
Customer data	0	0	0	0	1	0
Cost of data	0	0	0	0	0	1

Table 3 also shows the number of the repeated peer units.

4.4 λ (Weights for Peer Units)

Altering the value of each input and output in such a mode that the deliberated unit is traced on the efficiency frontier, then the hypothetical unit located on the efficiency frontier can be regarded as the virtual unit. λ denotes the grouping of the peer units utilized to build each virtual unit. The values of λ are shown in Table 4.

4.5 Weights (Values of the Variables for the Primary Model)

Tables 5 and 6 show the values of the variables for the primary model, which v_i is coefficient or weight assigned by DEA to input and u_r is coefficient or weight assigned by DEA to output.

Table 5 Input weights

Parameters	Organization culture	Manpower
Data privacy	0.001	0
Data credibility	0.001	0
Data security	0.003	0.001
Technology infrastructure	0.002	0
Customer data	0.001	0.002
Cost of data	0.002	0

Table 6 Output weights

Parameters	Decision making	Sales
Data privacy	0	0.001
Data credibility	0	0.001
Data security	0	0.002
Technology infrastructure	0	0.002
Customer data	0.001	0
Cost of data	0	0.002

Table 7 Input slacks

Parameters	Organization culture	Manpower
Data privacy	0	0
Data credibility	0	0
Data security	0	0
Technology infrastructure	0	0
Customer data	0	0
Cost of data	0	0

Table 8 Output slacks

Parameters	Decision making	Sales
Data privacy	8.17	0
Data credibility	0	0
Data security	0	0
Technology infrastructure	41.052	0
Customer data	0	0
Cost of data	0	0

4.6 Input and Output Slacks

The slacks related to respective units are shown respectively in Tables 7 and 8.

4.7 Target Values

Table 9 presents the actual and target values of each input.

Table 10 presents the actual and target values of each output.

Table 9 Inputs and target inputs

Parameters	Organization culture	Manpower
Data privacy	825 → 256.226	555 → 172.37
Data credibility	675 → 92.678	427 → 58.628
Data security	217 → 217	668 → 668
Technology infrastructure	378 → 181.008	523 → 250.443
Customer data	127 → 127	404 → 404
Cost of data	503 → 503	172 → 172

Table 10 Outputs and target outputs

Parameters	Decision making	Sales
Data privacy	384 → 392.17	346 → 346
Data credibility	169 → 169	121 → 121
Data security	178 → 178	413 → 413
Technology infrastructure	198 → 239.052	274 → 274
Customer data	936 → 936	139 → 139
Cost of data	819 → 819	641 → 641

5 Discussion

As per Table 1, it can be observed that Data security in adopting big data customer data and cost of data are the variables which are efficient in a Retail Set-up; this indicates that the adoption of big data in the retail sector would be positively supported by the variables which are 100% efficient. Whereas on other hand, if we look at the efficiency of Privacy of the data, it is just 31.11%, which indicates that the issue with data privacy of the retail customer is significantly inefficient and will have a negative impact on the implementation of big data in the retail sector. Similarly, the credibility of the data is highly inefficient with just 13.7% and may cause instability in the model as the data collected might not be credible to support the results predicted by the defined model. The technology infrastructure available in the retail sector in also found to be inefficient with 47.9%.

6 Conclusion

The retail sector is one of the fastest growing sectors and will dominate the economies of the world. So, most of the retailers are aiming to make optimum use of the customers data to deploy predictive models and induce the customers to purchase added products or persuade the consumers for higher unplanned purchases. The analysis done in this paper indicates that the deployment of big data model or analysis in the retail sector would be highly beneficial to the retailers, with variables like data security, customer data, and cost of the data being highly efficient in successfully implementing the defined model, but on the other hand, the issues like credibility of the data, privacy of the data, and the technology infrastructure available in the retail sector may significantly hamper the effective implementation of the model. Therefore, though the implementation of the big data model in the retail sector carries a lot of benefits in terms of predictability, one should be vigilant while using the data for the effective deployment of the defined model.

References

1. C. Raddats, P. Naik, A.Z. Bigdeli, Creating value in servitization through digital service innovations. Ind. Mark. Manag. **104**, 1–13 (2022)
2. M. Roe, K. Spanaki, A. Ioannou, E.D. Zamani, M. Giannakis, Drivers and challenges of internet of things diffusion in smart stores: A field exploration. Technol. Forecast. Soc. Chang. **178**, 121593 (2022)
3. E. Aktas, Y. Meng, An exploration of big data practices in retail sector. Logistics **1**(2), 12 (2017)
4. M. Naeem, T. Jamal, J. Diaz-Martinez, S.A. Butt, N. Montesano, M.I. Tariq, E. De-la-Hoz-Franco, E. De-La-Hoz-Valdiris, Trends and future perspective challenges in big data, in *Advances in Intelligent Data Analysis and Applications*, (Springer, Singapore, 2022), pp. 309–325
5. N. Kumar, An overview of emerging technologies in the Indian retail industry, in *Global Challenges and Strategic Disruptors in Asian Businesses and Economies*, (IGI Global, Hershey, 2021), pp. 247–256
6. M.A.E.A. Youssef, R. Eid, G. Agag, Cross-national differences in big data analytics adoption in the retail industry. J. Retail. Consum. Serv. **64**, 102827 (2022)
7. M.H. Sazu, S.A. Jahan, How big data analytics impacts the retail management on the European and American markets? CECCAR Bus. Rev. **3**(6), 62–72 (2022)
8. S.A. Gawankar, A. Gunasekaran, S. Kamble, A study on investments in the big data-driven supply chain, performance measures and organisational performance in Indian retail 4.0 context. Int. J. Prod. Res. **58**(5), 1574–1593 (2020)
9. N. Verma, J. Singh, An intelligent approach to big data analytics for sustainable retail environment using Apriori-MapReduce framework. Ind. Manag. Data Syst. **117**, 1503 (2017)
10. S. Brajesh, Big data analytics in retail supply chain, in *Big Data: Concepts, Methodologies, Tools, and Applications*, (IGI Global, Hershey, 2016), pp. 1473–1494
11. W.W. Cooper, L.M. Seiford, K. Tone, *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, 2nd edn. (Springer, New York, 2007)
12. M.G. Dekimpe, Retailing and retailing research in the age of big data analytics. Int. J. Res. Mark. **37**(1), 3–14 (2020)
13. F. Xiong, M. Xie, L. Zhao, C. Li, X. Fan, Recognition and evaluation of data as intangible assets. SAGE Open **12**(2), 21582440221094600 (2022)
14. L.L. Har, U.K. Rashid, L. Te Chuan, S.C. Sen, L.Y. Xia, Revolution of retail industry: From perspective of retail 1.0 to 4.0. Procedia Comput. Sci. **200**, 1615–1625 (2022)
15. E. Nilsson, Changes in the retail landscape-a Nordic perspective. Int. Rev. Retail Distrib. Consum. Res. **32**(4), 349–350 (2022)
16. X. Lyu, F. Jia, B. Zhao, Impact of big data and cloud-driven learning technologies in healthy and smart cities on marketing automation. Soft. Comput., 1–14 (2022). https://doi.org/10.1007/s00500-022-07031-w
17. A. Charnes, W.W. Cooper, E. Rhodes, Measuring the efficiency of decision making units. Eur. J. Oper. Res. **2**(6), 429–444 (1978)