

## Study on Big data analytics in E-commerce

<sup>1</sup>GAMPA KALPANA, <sup>2</sup>V. MADHUKAR

<sup>1</sup>PhD Scholar, Dept. of Computer Science ,Chaitanya Deemed to be University

<sup>2</sup>Assistant Professor, Dept. of Computer Science ,Chaitanya Deemed to be University

**Abstract:** *Big data is a large volume of both structured and unstructured data that is massive and it is difficult to process using traditional database and software techniques. In most enterprise scenarios the volume of data is too big or it moves too fast or it exceeds current processing capacity. Now-a-days, the tendency of shopping and purchasing products from vendors are totally changed with respect to the development of online shopping services. The logic is briefly called as E-Commerce, in which this port allows the user to register the identity and start purchasing the products based on their needs. This system attracts many corporate and commercial organizations to change their business strategy and start selling their products via online. In this paper, providing a brief review on the bigdata analysis of E-commerce. And the application of big data analytics in e-commerce is analysed and the increase in performance of dataset and the scalability problem is discussed and studied. In addition, the application of big data analytics in the e-commerce and the various technologies that makes analytics of consumer data possible is discussed. It also reviews the challenges faced by these e-commerce vendors while implementing big data analytics.*

**Keywords:** *Big Data, E-commerce, Classification, prediction, Data Analytics.*

### I. INTRODUCTION

One of the most rapidly expanding Big Data Analytics (BDA) segments is e-commerce. Due to their need to stay on top of their game, e-commerce companies are one of the fastest BDA adopters. Both structured and unstructured data are dealt with by e-commerce companies. Unstructured data includes clicks, likes, links, tweets, voices, and other unstructured data, whereas structured data concentrates on demographic information

such as name, age, gender, date of birth, address, and preferences. They are referred to as Big Data because of the variety, velocity, and volume of data they contain. Data is collected over time utilizing customer browsing and transactional points in ecommerce to track consumer purchase behaviour and personalize offers. This section discusses the many types of big data and how they influence e-commerce.

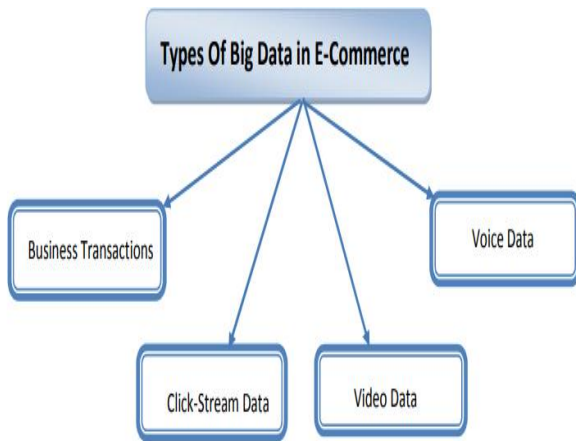


Fig.1 Types of Big Data in E-commerce

Fig. 1 illustrates four types of big data in e-commerce. E-commerce vendors can use big data to provide customized products, dynamic pricing, as well as targeting the right customer. Furthermore, with the Internet of Things (IoT) capabilities, smart shopping platforms are developed to support shopping customization and recommendations. Google, Amazon, eBay, Facebook and other big e-commerce companies have adopted BDA and experienced exponential growth

Big data types used in e-commerce:

**a) Data on transactions or business activity:**

- Data on transactions or business activity evolves over time as a result of interactions between the client and the company.
- These data are structured and come from a variety of sources, including customer relationship programmes (e.g., company-

maintained customer profiles, the occurrence of customer complaints) and sales transactions.

**b) Click-stream data:**

• In today's connected world, social media and online marketing, such as the use of click-stream data, play a crucial part in a company's ongoing promotional strategy, allowing management to make informed strategic and tactical decisions.

- Clickstream data is gathered from e-commerce websites and online advertisements, as well as social media content including tweets, blogs, and Facebook wall postings.

**c) Video data:**

• When paired with image analysis software, e-commerce organisations are eager to exploit not only clickstream data or transaction data, but also video data.

- The term "video data" refers to data obtained through the capture of live views. E-commerce firms are capable of analysing extremely unstructured data, such as video or audio.

• This data and the ability to be valuable to e-commerce businesses.

**d) Voice data:**

- Another type of data connected with the big data family is voice data, or data

originating from phone conversations, contact centres, or customer service.

- Voice data is useful for evaluating consumer purchasing patterns and attracting new clients. After analysing four types of user behaviour: click, collect, add to cart, and purchase. The researcher discovered that clicks received the highest proportion because customers view items by clicking on them but do not always add them to a list or cart. In addition to the Recency, Frequency, and Monetary (RFM) model, clients were classified as (i)VIP, (ii)loyal, (iii)significant customer, and (iv)most likely to leave. These solutions assist e-commerce businesses in communicating with and retaining customers.

## **BIG DATA FOR E-COMMERCE**

Online buyers and sellers are making use of big data for better shopping, selling the goods and gaining good customer relationship, giving better services and customer satisfaction in order to generate more sales. This is how big data benefits e-commerce companies:

- Distribute something more Valuable.
- Distribute More Personalized Interactions.
- Build Accurate Predictions.

- Decrease the shopping cart rejection Rate.

- Supply Customers with a better experience

## **II. REVIEW OF LITERATURE**

The related work in this section helps in understanding the role of big data in the e-commerce market. As e-commerce companies deal with structured data (e.g., personal information) and unstructured data (e.g., clicks, voice, images, tweets, etc.), they were the first in adopting BDA. With BDA, e-commerce companies enhance their operations and generate new strategies to gain benefits and increase customer value.

Ayman Abdalmajeed Alsmadiet al. [2023]The results of this paper show that many researchers in the e-commerce area focused on and applied data analytical solutions to fight the COVID-19 disease and establish preventive actions against it in various innovative manners. In addition, BDA and innovation in e-commerce is an interdisciplinary research field that could be explored from different perspectives and approaches, such as technology, business, commerce, finance, sociology, and economics. Moreover, the research findings are considered an invitation to those data analysts and innovators to contribute more to the body of the

literature through high-impact industry-oriented research which can improve the adoption process of big data analytics and innovation in organizations.

Hicham Kalkha et al. [2023] The purpose of this study is to investigate the impact of digitalization on trade logistics in e-commerce, emphasizing the significance of smart logistics for the e-commerce industry. We reviewed 288 articles published in the last decade in the Scopus database to assess the maturity of research in this area. For researchers, this study provides a better understanding of smart e-commerce logistics and identifies research gaps in the literature. For e-commerce professionals, it can help them adopt the latest technological trends in their logistics. Through a systematic literature review and network analysis approach, the study has contributed by identifying 5 clusters related to ICT application fields in e-commerce and 5 clusters related to important ICT enablers in smart logistics. We also identified several research gaps and areas for future study, including the underutilization of computer vision technology and the need for further research on product quality inspection and accessibility for people with disabilities.

Sukhendu et al. [2023] This paper presents applications of big data in various fields such as healthcare systems, social media

data, e-commerce applications, agriculture application, smart city application, and intelligent transport system. The paper also tries to focus on the characteristics, storage technology of using big data in these applications. This survey provides a clear view of the state-of-the-art research areas on big data technologies and its applications in recent past.

Nikita Gaikwad et al. [2022] Unlike any other time in history, this one is seeing a significant increase in the amount of data generated and captured. Data growth has seen a revival, fueled by ever-lower computing power and the internet's pervasiveness. The E-commerce industry has undergone a paradigm shift as a result of this. Big data analytics is gaining traction in the e-commerce world (BDA). Experts and scholars are eager to investigate the influence of this new analytics tool on corporate values and difficulties. It is, nevertheless, underdeveloped as a notion, preventing theoretical and practical progress. One of the major hurdles to e-commerce as a result of the information revolution is the massive amount of data that must be processed and reviewed to enjoy the benefits. Big Data Analytics (BDA) strives to improve decision-making by analysing and comprehending large amounts of data, such as messages and social media posts.

The impact of big data analytics on e-commerce is discussed in this study.

Sarah s. Alrumiah et al. [2021] this paper aims to study the values of implementing BDA in e-commerce to both vendors and consumers. Fifteen papers are selected to analyse the impacts of analyzing big data in e-commerce. Electronic vendors (E-vendors) use BDA to gain the competitive advantages they need to understand consumer behaviour and increase their income by improving customer loyalty. Besides, recommendation systems derived from BDA personalize the searching and shopping experience of the customers. However, there are some negative effects derived from applying BDA in ecommerce, such as shopping addiction. In addition, e-vendors have to deal with expensive BDA tools and professionals. In conclusion, even though BDA enhances consumers' and vendors' electronic shopping experience, the rapid growth of data is still challenging.

John Yeung et al. [2019] This paper presents the technical challenges faced by e-commerce players and how cloud computing has been designed and developed to handle these technical challenges. Once these technical challenges have been handled, another challenge of building data analytics on cloud platforms come out. This paper also

highlights the main reasons of driving organizations to build data analytics on cloud. Besides, this paper exhibits how to integrate machine learning models to the data analytic processes, for making more sophisticated analysis for e-commerce activities. The platform Amazon SageMaker is adopted to illustrate how machine learning models were integrated in the data analytic processes. The public cloud platform Amazon Web Services is used to present how machine learning models were integrated in the practices with a real-life e-commerce case.

Rayner Alfred et al. [2016] This paper addresses the rise of machine learning for big data analytics. First, machine learning and several terms related to machine learning are defined and explained in details and these terms include artificial intelligence, data mining, data science, data analytics and knowledge discovery, statistics and Business Intelligence. These definitions will show how these terms are inter-related to each other. Then, the definition of big data is outlined based on three terms: Volume, Velocity and Variety. Implementing a good big data strategy is very crucial in order to guarantee the success of applying machine learning for learning big data. As a result, the trending in Big Data is also illustrated and defined based on the landscape of big data;

Infrastructure, Analytics, Applications, Cross-Infrastructures/Analytics, Open Sources, Data Sources and API, Incubators and Schools.

Aamod Khatiwada et al. [2020] This research is the addition to the steps of making online business more user-friendly, interactive and output-oriented. The online marketing and sales of the products increase significantly if the opinion of the public for the product is analyzed intermittently. As the trend in today's date is that people update their response to the products faced by them in day-to-day life immediately in the social media, the developed system provides the platform for the large-scale producers to inspect how the consumers are responding to their products. The concept of Big Data Analysis is used for data collection, pre-processing and data analysis. A model is obtained by training the available Data using Deep Learning, which is used for the determination of sentiment values of the collected comments.

Peilu Feng et al. [2019] In the era of big data, while providing massive information, it also challenges the development of related activities in the overall environment. In the context of the rapid development of e-commerce, the opportunities of the development of the Internet of things technology are

analysed from the aspects of logistics distribution, quality control and facilities promotion. Electronic commerce is a new form of trade under the development of modern information technology, while cloud computing and the Internet of Things provide related services. Under the exertion of their related functions, the revolutionary improvement of e-commerce mode has been realized, and to a certain extent, it has promoted the development and operation of modern market economy. This article analyzes the development strategy of e-commerce based on Internet of things and cloud computing under the overall environment of big data era.

Yizhi Liet al. [2020] In order to overcome the traditional logistics system problems, such as poor interaction, low security, difficult to achieve remote monitoring, slow information transmission, this paper proposes an intelligent logistics system based on big data. The system comprehensively combines a variety of advanced technologies, such as big data technology, artificial intelligence technology, cloud computing technology, etc., and comprehensively realizes the integration of different advanced technologies. Based on this, the system can also realize remote monitoring and real-time monitoring, which provides great convenience for logistics staff. The

experimental results show that the system can optimize the traditional logistics system, and improve the efficiency of the logistics system and customer satisfaction.

Lili Wang et al. [2022] This paper proposes an economic forecasting method based on artificial intelligence methods combined with big data analytics. In our model, we consider the economic statistics, equilibrium, and future prediction with the big data. Through the artificial intelligence method based on deep learning, the possible political factors, human activity factors, and social environmental factors in actual economic activities are effectively combined to form the main analysis subject affecting the economy. The results show that our model can be used as a basic model for economic statistics, economic analysis, economic decision-making, economic self-regulation, and other functions under the current development trend of the data economy.

Biresh Kumar et al. [2021] This paper identifies out various attributes for estimating the usability and security aspects of e-commerce websites. A comparison study has been conducted on the various usability and evaluation models which help in the identification of an efficient model for assessing and evaluating the usability and security of e-commerce websites.

Rongrui Yu et al. [2021] This article starts with the analysis of the existing electronic commerce system, summarizes its characteristics, and analyses and solves its existing problems. Firstly, the characteristics of the relational database My Structured Query Language (MySQL) and the distributed database HBase are analysed, their respective advantages and disadvantages are summarized, and the advantages and disadvantages of each are taken into account when storing data. My SQL is used to store structured business data in the system, while HBase is used to store unstructured data such as pictures. These two storage mechanisms together constitute a data storage subsystem. Secondly, considering the large amount of data in the e-commerce system and the complex calculation of the data mining algorithm, this paper uses MapReduce to realize the parallelization of the data mining algorithm and builds a Hadoop-based commodity recommendation subsystem on this basis.

Diana Teresa et al. [2021] This paper aims to assess information and communication technologies (ICT) policies for digital transformation in Colombia to determine their effectiveness in technology readiness for Internet of Things (IoT) adoption in small and medium enterprises (SMEs) in the trading sector.

Lin Li et al. [2019] With the advent of the Internet age, the sales industry has also launched a new networked sales channel through the information network platform. In the network sales platform where all kinds of e-commerce are gathered, the mastery and analysis of various information data can provide accurate data decision analysis for e-commerce. At present, information data analysis systems based on big data and artificial intelligence technology provide data collection, analysis and decision-making to help e-commerce provide more efficient and accurate information services. Based on the e-commerce platform's big data intelligent data processing decision system, the product recommendation and effect of e-commerce concern in the paper are studied, and the collaborative filtering algorithm is adopted to recommend the product for the e-commerce user interest preference.

Isaac Kofi Nti et al. [2022] The availability of digital technology in the hands of every citizenry worldwide makes an available unprecedented massive amount of data. The capability to process these gigantic amounts of data in real-time with Big Data Analytics (BDA) tools and Machine Learning (ML) algorithms carries many paybacks. However, the high number of free BDA tools, platforms, and data

mining tools makes it challenging to select the appropriate one for the right task. This paper presents a comprehensive mini-literature review of ML in BDA, using a keyword search; a total of 1512 published articles was identified. The articles were screened to 140 based on the study proposed novel taxonomy.

### III. CHALLENGES IN E-COMMERCE USING BIG DATA

Online Shopping System has the irrelevant information of the products that are stored in servers. Massive amount of data is to be managed by clustering. The Information used in clustering mainly uses product ID, product category and property set of a product. The Map Reduce operation is simple and it uses mapper function and reducer function in two phases. In the map process, it filters the product by its category as the key and the processed product as the value. In the reduce process, the data in the same category will be write into a file. If an individual file is generated for a category, then the number of individual files will be too large to be supported by the Hadoop file management system. Hence, we can adjust the number of files to a specific range. Map reduce framework makes the system to process in fast and efficient manner. Hadoop Distributed File System supports and



stores the enormous amount of data. This System has the function such as service comparison, recommendation system, user register, login, commodity search, commodity management, order management, system management, shopping order and user management, these functions are performed by using hadoop environment in an efficient manner.

#### **TECHNIQUES TO OVERCOME THE LIMITATIONS OF E-COMMERCE.**

Due to the explosive growth in e-commerce system, the enormous number of products are added in a single website by crawling the product dataset from two or three websites. The dataset is processed by Hadoop environment by using HDFS storage; here data is scalable using map reduce, a parallel programming model it processes the dataset in a parallel manner. The Cluster formation is used in this model which separates the relevant information of the product as product information and features. Collaborative filtering helps the recommendation system to be efficient by analysing the taste of customer in choosing the products online. Hadoop environment is for fast processing of data. HDFS is used for dataset storage. Manipulation of data, data storage and processing of data is done using big data; which makes the system more efficient and scalable by adding the more number of

products in future. It consumes less time consumption. This makes the system to process in a fast and effective manner.

#### **IV. CONCLUSION**

In this research, we have analysed the BDA impacts on the e-commerce experience of vendors and customers. Authors conclude that applying BDA capabilities in e-commerce projects improves the online shopping experience and increases vendor's revenues. And the techniques which are used in current e-commerce websites which has large volume of data are studied and its limitations are analyzed. The study of the e-commerce websites, its storage in HDFS and the scalability of products in hadoop environment is analyzed and studied. An online shopping system which produces an integrated dataset from various websites can be proposed to support the scalability problem and to reduce the time consumption which processes the system in an efficient manner.

#### **REFERENCES**

1. D. Chrimes and H. Zamani, Using distributed data over HBase in big data analytics platform for clinical services, *Comput. Math. Methods Med.*, vol. 2017, p. 6120820, 2017.

2. I. Kotenko, I. Saenko, and A. Branitskiy, Framework for mobile internet of things security monitoring based on big data processing and machine learning, *IEEE Access*, vol. 6, pp. 72714–72723, 2018.
3. A. Khan, I. Ibrahim, M. I. Uddin, M. Zubair, S. Ahmad, M. D. Al Firdausi, and M. Zaindin, Machine learning approach for answer detection in discussion forums: An application of big data analytics, *Sci. Program.*, vol. 2020, p. 4621196, 2020
4. Nikita Gaikwad, Dr. Harshali Patil, 2023, “An Overview of Big Data Analytics in Electronic Commerce”
5. Hicham Kalkha; Azeddine Khiat; Ayoub Bahnasse, 2023, “The Rising Trends of Smart E-Commerce Logistics”, Page(s): 33839 – 33857.
6. Sukhendu S. Mondal, Somen Mondal, Sudip Kumar Adhikari, "Applications of Big Data in Various Fields: A Survey", *Recent Trends in Intelligence Enabled Research*, vol.1446, pp.221, 2023
7. SARAH S. ALRUMIAH AND MOHAMMED HADWAN, 2021, “Implementing Big Data Analytics in E-Commerce: Vendor and Customer View”, pp.37281-37286.
8. Ayman Abdalmajeed Alsmadi, Ahmed Shuhaiber, Manaf Al-Okaily, “Big data analytics and innovation in e-commerce: current insights and future directions”, pp.
9. Rayner Alfred et al., 2016, “The rise of machine learning for big data analytics”,
10. Aamod Khatiwada and Sandip Agrahari, 2019, “Big Data Analytics and Deep Learning Based Sentiment Analysis System for Sales Prediction”, pp.1-6
11. Peilu Feng, 2019, “Big Data Analysis of E-Commerce Based on the Internet of Things”, pp.345-347.
12. Yizhi Li, "Intelligent Logistics System Based on Big Data", *2020 International Conference on Robots & Intelligent System (ICRIS)*, pp.299-302, 2020.
13. Lili Wang, Liuyang Zhao, "Digital Economy Meets Artificial Intelligence: Forecasting Economic Conditions Based on Big Data Analytics", *Mobile Information Systems*, vol.2022, pp.1, 2022.
14. Biresh Kumar, Sharmistha Roy, "An Empirical Study on Usability and Security of E-Commerce Websites", *Research in Intelligent and Computing in Engineering*, vol.1254, pp.735, 2021.
15. Rongrui Yu, Chunqiong Wu, Bingwen Yan, Baoqin Yu, Xiukao Zhou, Yanliang Yu, Na Chen, "Analysis of the Impact of Big Data on E-

- Commerce in Cloud Computing Environment", *Complexity*, vol.2021, pp.1, 2021.
16. Diana Teresa Parra-Sánchez, Leonardo Hernan Talero-Sarmiento, Cesar Dario Guerrero, "Assessment of ICT policies for digital transformation in Colombia: technology readiness for IoT adoption in SMEs in the trading sector", *Digital Policy, Regulation and Governance*, vol.23, no.4, pp.412, 2021.
17. Isaac Kofi Nti, Juanita Ahia Quarcoo, Justice Aning, 2022, "A Mini-Review of Machine Learning in Big Data Analytics: Applications, Challenges, and Prospects", pp.81-97.
18. S. Puttinaovarat and P. Horkaew, Flood forecasting system based on integrated big and crowdsource data by using machine learning techniques, *IEEE Access*, vol. 8, pp. 5885–5905, 2020