

# Big Data Analytics for E-Commerce

**Dr. Shweta and Ms. Hemlata**

Assistant Professor in Computer Science,  
Saraswati Mahila Mahavidyalaya, Palwal, India

**Abstract:** Big data refers to a substantial amount of data, both organised and unorganised, that is extremely large and poses challenges when it comes to processing using conventional database and software methods. In the majority of enterprise scenarios, the data is either excessively large, rapidly changing, or surpasses the current processing capabilities. In recent times, the practice of shopping and buying goods from sellers has undergone a complete transformation due to the advancement of online shopping services. The concept is commonly referred to as E-Commerce, where this platform enables users to register their identity and initiate product purchases according to their requirements. This system entices numerous corporate and commercial entities to alter their business strategy and commence selling their products through online channels. This paper presents a concise overview of the analysis of big data in the field of E-commerce. The analysis of big data analytics in e-commerce is examined, focussing on the improvement of dataset performance and the study of scalability issues. Furthermore, this text explores the utilisation of big data analytics in the field of e-commerce, as well as the diverse technologies that enable the analysis of consumer data. Additionally, it examines the difficulties encountered by these e-commerce merchants when implementing big data analytics.

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

## I. INTRODUCTION

E-commerce is one of the fastest-growing segments in the field of Big Data Analytics (BDA). E-commerce companies are among the fastest adopters of Big Data Analytics (BDA) due to their imperative to maintain a competitive edge. E-commerce companies handle both structured and unstructured data. Unstructured data encompasses various types of information such as clicks, likes, links, tweets, voices, and other forms of unorganised data. In contrast, structured data focusses on specific demographic details like name, age, gender, date of birth, address, and preferences. The term "Big Data" is used to describe data sets that are characterised by their variety, velocity, and volume. Data is gathered over a period of time by using customer browsing and transactional data in ecommerce to monitor consumer purchasing patterns and customise promotional offers. This section examines the various categories of big data and their impact on e-commerce.

Fig.1 Types of Big Data in E-commerce

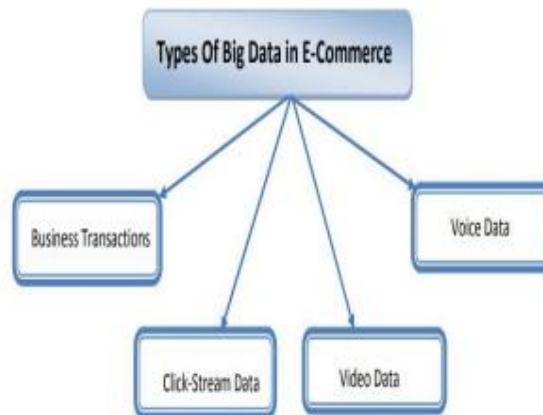


Figure 1 depicts four distinct categories of big data in the field of electronic commerce. E-commerce vendors can leverage big data to offer personalised products, implement dynamic pricing, and effectively target their desired customer base. In addition, the development of smart shopping platforms with Internet of Things (IoT) capabilities enables support for personalised shopping experiences and recommendations. Google, Amazon, eBay, Facebook, and other prominent e-commerce companies have implemented Big Data Analytics (BDA) and witnessed significant and rapid expansion.

**Big data types used in e-commerce:**

**A. Data on transactions or business activity:**

- The data regarding transactions or business activity undergoes changes over time due to the interactions between the client and the company.
- The data provided are organized and obtained from various sources, such as customer relationship programs (e.g., company-maintained customer profiles, customer complaints) and sales transactions.

**B. Click-stream data:**

- Social media and online marketing, including the analysis of click-stream data, are essential components of a company's promotional strategy in today's interconnected world. They enable management to make well-informed strategic and tactical decisions.
- Click stream data is collected from e-commerce websites and online advertisements, as well as social media content such as tweets, blogs, and Face book wall posts.

**C. Video data:**

- E-commerce organizations are keen to leverage video data in addition to click stream data or transaction data when combined with image analysis software.
- This data possesses the potential to be highly valuable to e-commerce businesses.
- The term "video data" pertains to data acquired by capturing real-time visual perspectives. E-commerce companies have the ability to analyse highly unstructured data, such as video or audio.

**D. Voice data**

- Another category of data associated with the big data domain is voice data, also known as data.
- Voice data is valuable for assessing information derived from telephone conversations, contact centers, or customer service interactions. Consumer buying behaviour and customer acquisition strategies. Upon analyzing four distinct user behaviours, namely click, collect; add to cart, and purchase. The researcher found that clicks accounted for the largest proportion, as customers view items by clicking on them but do not always add them to a list or cart. Furthermore, clients were categorized based on the Regency, Frequency, and Monetary (RFM) model into four groups: (i) VIP, (ii) loyal, (iii) significant customer, and (iv) most likely to leave. These solutions facilitate communication and customer retention for e-commerce businesses.

**II. BIG DATA FOR E-COMMERCE**

- Online consumers and merchants are utilizing big data to enhance their shopping experience, optimize sales of goods, establish strong customer relationships, provide superior services, and ensure high levels of customer satisfaction, ultimately resulting in increased sales. Big data provides significant advantages to e-commerce companies:
  - Enhance the dissemination of tailored interactions
  - Disseminate something of greater worth
  - Develop precise forecasts
  - Enhance customer satisfaction by providing an improved experience
  - Reduce the rate of shopping cart rejections.

### III. REVIEW OF LITERATURE

The literature review in this section provides insights into the impact of big data on the e-commerce industry. E-commerce companies were the pioneers in adopting Big Data Analytics (BDA) as they handle both structured data (such as personal information) and unstructured data (such as clicks, voice, images, tweets, etc.). BDA enables e-commerce companies to optimise their operations and devise novel strategies to maximize advantages and enhance customer value.

Ayman Abdalmajeed Alsmadi. The year 2023. The findings of this study indicate that numerous researchers in the field of e-commerce have concentrated on and implemented data analytical solutions to combat the COVID-19 disease and establish preventive measures against it in diverse and inventive ways. Furthermore, the study of Big Data Analytics (BDA) and innovation in e-commerce is a multidisciplinary research area that can be examined from various viewpoints and methodologies, including technology, business, commerce, finance, sociology, and economics. Furthermore, the research findings serve as a call for data analysts and innovators to make further contributions to the existing body of literature by conducting industry-focused research with significant impact. This research can enhance the implementation of big data analytics and innovation in organisations.

The authors of the publication are Hicham Kalkha and colleagues, and the publication was released in 2023. This study aims to examine the influence of digitalization on trade logistics in the e-commerce sector, with a particular focus on the importance of smart logistics for the e-commerce industry. We conducted a comprehensive analysis of 288 articles that were published within the past ten years in the Scopus database. Our objective was to evaluate the level of advancement and development in this particular field of research. This study enhances researchers' comprehension of smart e-commerce logistics and identifies areas of research that have not been adequately addressed in the existing literature. E-commerce professionals can benefit from adopting the latest technological trends in their logistics. The study has made a valuable contribution by using a systematic literature review and network analysis to identify 5 clusters associated with ICT application fields in e-commerce and 5 clusters associated with significant ICT enablers in smart logistics. In addition, we have identified various areas for future study and research gaps. These include the insufficient utilization of computer vision technology and the necessity for further investigation into product quality inspection and accessibility for individuals with disabilities.

Sukhendu and colleagues [2023] This paper explores the utilization of big data in diverse domains including healthcare systems, social media data, e-commerce applications, agriculture application, smart city application, and intelligent transport system. The paper also aims to examine the attributes and storage technology associated with the utilization of big data in these applications. This survey offers a comprehensive overview of the cutting-edge research on big data technologies and their recent applications. The authors of the publication are Nikita Gaikwad and others, and the publication was released in 2022. Unlike any previous era, the present time is witnessing a substantial surge in the volume of data being produced and collected. The proliferation of the internet and the continuous decrease in computing power has led to resurgence in data growth. This has caused a significant transformation in the E-commerce sector. Big data analytics, also known as BDA, is becoming increasingly popular in the field of e-commerce. Experts and scholars are keen to examine the impact of this novel analytics tool on corporate values and challenges.

However, the concept is still not fully developed, which hinders both theoretical and practical advancements. A significant challenge posed by the information revolution to e-commerce is the substantial volume of data that needs to be processed and analyzed in order to reap its advantages. Big Data Analytics (BDA) aims to enhance decision-making by analyzing and understanding vast quantities of data, including messages and social media posts. This study examines the influence of big data analytics on e-commerce. The authors of the study are Sarah S. Alrumiah and colleagues. The year is 2021. This paper seeks to examine the benefits of incorporating Big Data Analytics (BDA) in the field of e-commerce for both sellers and buyers. A total of fifteen papers have been chosen to examine the effects of analyzing large volumes of data in the field of electronic commerce. E-vendors utilize Big Data Analytics (BDA) to acquire the competitive edge necessary for comprehending consumer behaviour and enhancing their revenue through the enhancement of customer loyalty. In addition, recommendation systems derived from big data analytics (BDA) customize the searching and shopping experience of customers. Nevertheless, the implementation of BDA in e-commerce can lead to adverse consequences, including the development of shopping addiction. Furthermore, electronic vendors are required to manage costly Big Data Analytics (BDA) tools and employ skilled professionals. To

summarise, while BDA improves the electronic shopping experience for both consumers and vendors, the exponential increase in data remains a significant challenge. John Yeung and others. The year is 2019. This paper outlines the technical obstacles encountered by e-commerce participants and how cloud computing has been specifically designed and developed to address these challenges. After successfully addressing these technical obstacles, another challenge arises in the construction of data analytics on cloud platforms.

This paper also emphasizes the primary factors that motivate organisations to implement data analytics on cloud platforms. In addition, this paper demonstrates the integration of machine learning models into data analytic processes to enhance the analysis of e-commerce activities. The Amazon SageMaker platform was utilized to demonstrate the integration of machine learning models into data analytic processes. Amazon Web Services (AWS) showcased the integration of machine learning models in real-life e-commerce scenarios on their public cloud platform. Rayner Alfred and colleagues [2016] This paper discusses the increasing utilization of machine learning in the field of big data analytics. The text provides comprehensive definitions and explanations of machine learning and its associated terms, such as artificial intelligence, data mining, data science, data analytics, knowledge discovery, statistics, and business intelligence. These definitions will demonstrate the interconnectedness of these terms. Next, the definition of big data is elucidated by considering three key factors: Volume, Velocity, and Variety. Having a well-executed big data strategy is essential to ensure the successful implementation of machine learning for processing large amounts of data. Consequently, the current trends in Big Data are depicted and categorised according to the various aspects of big data: Infrastructure, Analytics, Applications, Cross-Infrastructures/Analytics, Open Sources, Data Sources and API, Incubators and Schools. The authors of the publication are Aamod Khatiwada et al. [2020]. This research aims to enhance the process of making online businesses more user-friendly, interactive, and focused on producing desired outcomes. The online marketing and sales of the products experience a substantial increase when the public's perception of the product is periodically analysed. In today's society, it is common for people to promptly share their opinions about products they encounter in their daily lives on social media. The developed system offers a platform for major producers to analyse the feedback and reactions of consumers towards their products on a large scale. Big Data Analysis involves the gathering, initial preparation, and examination of data. A model is created through the process of training the available data using Deep Learning. This model is then utilised to determine the sentiment values of the collected comments. The authors of the study are Peilu Feng et al. and the study was published in 2019. In the age of big data, although it offers a wealth of information, it also poses challenges to the development of related activities within the broader context. Within the context of the fast-paced growth of e-commerce, this analysis examines the potential of Internet of Things (IoT) technology in the areas of logistics distribution, quality control, and facilities promotion. Electronic commerce is a contemporary trade method that is evolving with the advancement of modern information technology. It is supported by services such as cloud computing and the Internet of Things. The revolutionary advancement of e-commerce has been achieved through the implementation of its various functions. This advancement has played a significant role in promoting the growth and functioning of the modern market economy. This article examines the growth strategy of e-commerce using Internet of Things and cloud computing within the broader context of the big data era. Yizhi Li et al. The year 2020. This paper suggests an intelligent logistics system based on big data to address common issues in traditional logistics systems, including inadequate communication, low security, challenges in remote monitoring, and slow information transmission. The system effectively integrates a range of advanced technologies, including big data, artificial intelligence, and cloud computing, to achieve a comprehensive integration of different advanced technologies. In addition, the system has the capability to perform remote monitoring and real-time monitoring, offering significant convenience for logistics personnel. The experimental results demonstrate that the system has the capability to enhance the efficiency of the traditional logistics system and increase customer satisfaction. The publication titled "Lili Wang et al. [2022]" This paper presents a novel approach to economic forecasting by integrating artificial intelligence techniques with big data analytics. Our model incorporates economic statistics, equilibrium, and future prediction using big data. The artificial intelligence method based on deep learning effectively combines the possible political factors, human activity factors, and social environmental factors in actual economic activities to form the main analysis subject that impacts the economy. The findings demonstrate that our model is suitable as a fundamental framework for economic statistics, economic analysis, economic decision-making, economic self-regulation, and other related functions within the current trajectory of the data-driven economy. Biresh Kumar and

colleagues [2021]. This paper delineates multiple attributes for assessing the usability and security aspects of e-commerce websites. A comparative study has been conducted on different usability and evaluation models to identify an effective model for assessing and evaluating the usability and security of e-commerce websites.

The authors of this publication are Rongrui Yu et al. and the publication was released in 2021. This article commences by examining the current electronic commerce system, outlining its features, and analysing and resolving any prevailing issues. Firstly, this analysis examines the characteristics of the relational database MySQL and the distributed database HBase. It summarises their respective advantages and disadvantages, and takes into account the advantages and disadvantages of each when storing data. MySQL is utilized for the storage of organized business data within the system, whereas HBase is employed for the storage of unorganized data, such as images. These two storage mechanisms collectively form a data storage subsystem. Furthermore, due to the extensive volume of data within the e-commerce system and the intricate calculations involved in the data mining algorithm, this study employs Map Reduce to achieve parallelisation of the data mining algorithm. Additionally, a commodity recommendation subsystem based on Hadoop is constructed upon this foundation. Diana Teresa and others [2021] This paper seeks to evaluate the efficacy of information and communication technologies (ICT) policies in Colombia for facilitating digital transformation. Specifically, it aims to determine how these policies contribute to the technology readiness of small and medium enterprises (SMEs) in the trading sector for adopting Internet of Things (IoT) technologies. In the era of the Internet, the sales industry has introduced a new sales channel that operates through an information network platform. On the network sales platform, which brings together various types of e-commerce, the ability to master and analyze diverse information data can enable precise decision-making analysis for e-commerce. Currently, big data and artificial intelligence technology-based information data analysis systems offer data collection, analysis, and decision-making capabilities to enhance the efficiency and accuracy of information services in e-commerce. This study focuses on the product recommendation and its impact on e-commerce using an advanced data processing decision system based on big data intelligence. The collaborative filtering algorithm is utilized to recommend products based on the user's preferences in e-commerce. The authors of the publication are Isaac Kofi Nti and others, and the publication was released in 2022. The widespread access to digital technology globally has resulted in an unprecedented abundance of data. The ability to analyze vast quantities of data in real-time using Big Data Analytics (BDA) tools and Machine Learning (ML) algorithms offers numerous benefits. Nevertheless, the abundance of free BDA tools, platforms, and data mining tools poses a difficulty in choosing the most suitable one for a specific task. This paper provides a thorough review of the existing literature on machine learning in big data analytics. The review was conducted by performing a keyword search, which resulted in the identification of a total of 1512 published articles. The articles were filtered down to 140 based on the study's innovative taxonomy.

#### **IV. CHALLENGES IN E-COMMERCE USING BIG DATA**

The Online Shopping System contains extraneous data regarding the products that are stored within servers. Clustering is used to manage a substantial volume of data. Clustering primarily relies on product ID, product category, and property set information for analysis. The Map Reduce operation is a straightforward process that involves the use of a mapped functional reducer function in two distinct phases. During the mapping process, the product is sorted by its category and assigned as the key, while the processed product is assigned as the value. During the reduction process, data belonging to the same category will be written into a file. If a category generates an individual file, the quantity of individual files will exceed the capacity of the Hadoop file management system. Therefore, we have the ability to modify the quantity of files within a designated range. The map reduce framework enhances the system's processing speed and efficiency. The Hadoop Distributed File System (HDFS) is capable of accommodating and managing vast quantities of data. This system utilizes the Hadoop environment to efficiently perform various functions such as service comparison, recommendation system, user registration, login, commodity search, commodity management, order management, system management, shopping order, and user management.

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

As a result of the rapid expansion of the e-commerce system, a large quantity of products is incorporated into a single website by extracting the product dataset from two or three other websites. The dataset is processed in the Hadoop

environment using HDFS storage. The data is made scalable through the use of map reduce, a parallel programming model that allows for parallel processing of the dataset. This model utilizes Cluster formation to distinguish and isolate the pertinent details of the product, such as product information and features. Collaborative filtering enhances the efficiency of the recommendation system by analyzing customers' preferences in selecting products online. The Hadoop environment is designed to facilitate efficient data processing. HDFS is utilized for the purpose of storing datasets. The manipulation, storage, and processing of data are accomplished through the utilization of big data, which enhances system efficiency and scalability by accommodating an increased number of products in the future. It requires less time. This enables the system to operate efficiently and expediently.

## VI. CONCLUSION

This study examines the effects of Big Data Analytics (BDA) on the e-commerce experience of both vendors and customers. The authors' conclusion is that implementing Business Data Analytics (BDA) capabilities in e-commerce projects enhances the online shopping experience and boosts the revenues of vendors. The study focuses on analyzing the techniques employed in current e-commerce websites with a substantial amount of data, as well as identifying their limitations. An analysis and study are conducted on the study of e-commerce websites, the storage of data in HDFS, and the scalability of products in a Hadoop environment. A proposed solution to address the scalability issue and improve the efficiency of an online shopping system is to develop an integrated dataset by gathering information from multiple websites. This approach aims to reduce the time required to process the system's operations.

## REFERENCES

- [1]. 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.
- [2]. 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.
- [3]. 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.
- [4]. 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
- [5]. Nikita Gaikwad, Dr. Harshali Patil, 2023, "An Overview of Big Data Analytics in Electronic Commerce"
- [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 Khatriwada 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]. Hicham Kalkha; Azeddine Khia; Ayoub Bahnasse, 2023, "The Rising Trends of Smart E-Commerce Logistics", Page(s): 33839 – 33857.
- [13]. Yizhi Li, "Intelligent Logistics System Based on Big Data", 2020 International Conference on Robots & Intelligent System (ICRIS), pp.299-302, 2020.
- [14]. 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.