Journal of Information Systems Engineering and Management
2023, 8(3), 22638
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Unleashing the Power of Big Data: Designing a Robust Business Intelligence Framework for E-commerce Data Analytics

Xubo Ye[1], Mababa Jonilo[2*]

[1] Ph.D candidate, Graduate School, Angeles University Foundation, Angeles, Philippines
[2] Professor, Graduate School, Angeles University Foundation, Angeles, Philippines
* **Corresponding Author:** mababa.jonilo@auf.edu.ph

Citation: Ye, X., & Jonilo, M. (2023). Unleashing the Power of Big Data: Designing a Robust Business Intelligence Framework for E-commerce Data Analytics. *Journal of Information Systems Engineering and Management*, 8(3), 22638. https://doi.org/10.55267/iadt.07.13946

ARTICLE INFO

Received: 19 May 2023
Accepted: 28 July 2023

ABSTRACT

E-commerce companies are struggling to make use of the enormous amounts of data collected from diverse sources in the big data era. Designing and implementing a strong business intelligence (BI) framework that makes use of big data analytics is essential to overcoming this difficulty. The study examines how cloud computing affects mobile e-commerce, stressing its benefits for real-time data processing, scalability, and analysis. which boosts the competitiveness of Chinese e-commerce businesses. This study's goal is to deploy a comprehensive BI platform designed especially for e-commerce research to unleash the power of big data. Furthermore, the deployment of precision marketing techniques based on the RFM model and historical data analysis increases client segmentation, leading to targeted marketing efforts, greater customer happiness, and higher conversion rates. The major goal of this research is to equip e-commerce companies with the tools they need to take advantage of big data's potential and make decisions that will give them a competitive edge. Data storage, retrieval, and data mining are made possible by the integration of big data technologies, such as relational and distributed databases, along with parallelization via MapReduce. Ordinary, the findings of this newsletter spotlight the importance of embracing massive information technologies and methodologies in the e-trade sector. Leveraging cloud computing, records mining, and enterprise intelligence strategies can free up the capability of tremendous records assets, permitting enterprises to make informed choices, drive innovation, and gain a competitive facet. The paper highlights the value of security safeguards and risk assessment models in e-commerce systems, offering suggestions for spotting and reducing potential dangers and preserving the integrity of the system.

Keywords: E-Commerce, Big Data, Business Intelligence, Security Safeguards, Risk Assessment Models.

INTRODUCTION

A significant amount of data is being produced in the fourth industrial revolution era from a variety of sources, including GPS, sensors, websites, applications, and social media platforms. The massive amounts of information produced by organizations are continuously stored on these data servers (Pancić et al., 2023). Data from websites, social media interactions, tracking devices, Internet of Things (IoT) applications, sensors, and online news items are all included in this. The ability to gather heterogeneous data from various sources, including both organized and unstructured, complicated and simple information, has improved because of developments in computing and communication technology. Businesses are currently making a sizable profit from the analysis of unstructured data (Li et al., 2023). This analysis offers insightful information that enables businesses to improve their operational procedures and general productivity. Additionally, analyzing unstructured data has profound effects on industries including education, security, healthcare, and manufacturing.

Big data analytics, artificial intelligence, and efficient data management can all be used to achieve business intelligence and better decision-making. Chandrakar and Hulipalled (2022) stated that big data analytics is significant for technique improvement. Business intelligence is the umbrella term for all technology, tools, systems, and applications that make it easier to compile, analyse, integrate, and deliver business data so that proactive decision-making is possible. This method offers important help in analyzing, regulating,

and generating insights from data that can be used to create corporate processes and procedures (Rosas et al., 2023; Ma et al., 2022; Langerhuizen et al., 2021).

Business intelligence (BI) refers to an organization's capacity to gain actionable insights from the data that is routinely gathered as a result of its regular business operations and activities. Decision-makers rely on BI to help them by giving them insightful data that helps them be more productive and make decisions more quickly. Additionally, BI strengthens and supports operational rules' effectiveness and their influence on corporate-level decision-making, administration, budgeting, and financial reporting. Better strategic options are thus produced in dynamic company contexts as a result of this. In addition to its many other advantages, BI helps organizations function better by seeing new possibilities, providing business insights, highlighting hazards, and increasing decision-making procedures (Shamim et al., 2021; Lawani et al., 2019). The retail industry has witnessed substantial transformations in the era of big data, primarily attributed to the exponential growth of online shopping websites, commonly known as "e-tailers."

Modern consumers are lured more and more to these online platforms, especially in nations like India, because of things like convenient and affordable Internet connection. The competition has been further heightened by telecom firms like Reliance Jio offering free unlimited internet data usage for a limited time as part of their membership programs. The popularity of online buying is further influenced by enticing incentives like cashback offers and hassle-free returns without shipping charges provided by portals like PayTm, as well as by features like cash on delivery from E-Commerce websites like Flipkart, Amazon, and others (Rosas et al., 2023).

Finding the best e-commerce site to fulfill a customer's unique purchasing needs is not always simple, though. Customers primarily use conventional search engines like Google and Bing to look for e-commerce websites. Despite the use of sophisticated algorithms, even well-known search engines frequently return comparable results when different users conduct identical E-Commerce queries. This suggests that the customer's personal preferences and background may not have a big impact on the search results. In a research of Gomez-Nieto et al. (2014), this topic is discussed. Furthermore, as stated by (Pancić et al., 2023), the issue might get much trickier if the query is vague or imprecise. The significance of business intelligence in the context of e-commerce is highlighted in this case. Businesses must use data analytics and AI-driven technology to understand client preferences, personalize search results, and offer specialized recommendations in order to improve the online buying experience (Choi et al., 2022).

E-commerce can efficiently analyze client searches, enhance search engine algorithms, and deliver more accurate and relevant results by leveraging business intelligence tools and methodologies. This, in turn, contributes to the growth and profitability of e-commerce enterprises, as it improves consumer satisfaction, increases the likelihood of successful conversions, and enhances client retention (Song et al., 2022). Despite the substantial amount of data generated by online transactions and the rapid growth of e-commerce in China, there might still be a lack of comprehensive and robust frameworks for commercial business intelligence tailored specifically for e-trade evaluation in the Chinese environment. While research on business intelligence and big data analytics in e-trade exists, there remains a need for studies that focus specifically on the unique characteristics and challenging circumstances of the Chinese e-commerce market.

Current frameworks and approaches might not sufficiently address the unique challenges of reading e-commerce data in China, which includes accounting for cultural variables, regulatory environments, and market dynamics. The special needs and complexity of this context necessitate further research and tailored methodologies (Zhuang et al., 2021; Yu et al., 2021; Shittu & Weigelt, 2022). The focus of this study is on the absence of a dependable business intelligence framework that effectively utilizes vast data for analyzing e-commerce in China. Currently, the lack of specialized approaches and tools capable of fully leveraging the enormous amount, breadth, and speed of data generated through online transactions characterizes the current state of research and practice in this field.

The capacity of e-commerce groups to obtain valuable information, make informed decisions, and gain a competitive edge in the Chinese market is hindered by this gap. Therefore, there is a crucial need to design a comprehensive and customized business intelligence framework that addresses the unique demands and necessities of analyzing e-commerce statistics in China. Such a framework would enable businesses to unlock the full potential of massive data for strategic decision-making and performance improvement (Himeur et al., 2023). The study aims to address the specific challenges and requirements of analyzing e-commerce data in China. Its objective is to design a tailored business intelligence framework that effectively utilizes extensive data for e-commerce evaluation within the Chinese market. The study will evaluate the framework's performance, provide practical recommendations, and contribute to improved decision-making and overall performance of e-commerce businesses in China (Maroufkhani et al., 2023). Business intelligence is essential for making wise decisions, but the current frameworks might not be able to meet the unique needs of the Chinese setting, which includes cultural influences, legal complexities, and changing market situations. To fully utilize huge data for strategic decision-making and performance enhancement for Chinese e-commerce companies.

The scale and growth of China's e-commerce market, along with the enormous volumes of data produced, have made it necessary to address the issue of developing a strong business intelligence framework for e-commerce data analytics. With a sizable consumer base and a robust digital environment, China boasts one of the largest e-commerce markets in the world (Du, 2022). With the use of big data technologies, this study seeks to provide a solid business intelligence (BI) foundation for e-commerce data analytics in China. The study emphasizes real-time analytics, cross-channel integration, AI-driven business intelligence, and

personalization. In order to improve the effectiveness of e-commerce, it also examines sustainability, sentiment analysis, and privacy issues(Jahani et al., 2023; Muchenje & Seppänen, 2023).

The study intends to fill a research gap that has been found regarding the shortcomings of the present business intelligence frameworks in efficiently analyzing e-commerce data in the Chinese industry. The study will pay particular attention to how current strategies might not adequately account for the special difficulties brought on by cultural variables, legislative complexities, and dynamic market movements in China. The research intends to offer e-commerce enterprises expanded data analytics skills for better decision-making and performance enhancement in the Chinese environment by building a complete and personalized business intelligence platform.

This study will make a contribution to the present academic literature via addressing the research hole within the discipline of e-commerce evaluation in China. It will offer a complete and tailored business intelligence framework specially designed to handle the unique challenges and characteristics of the Chinese language e-trade market. Via integrating ideas from business intelligence, big records analytics, and e-trade, the look will make contributions to the development of expertise in those domains, offering insights into the powerful usage of massive records for strategic decision-making in the e-trade sector. The structure of this study is as follows: The first section defines the background, problem, and contribution of the study. The second section demonstrates the literature review of different constructs and the third part of this paper is methodology and techniques. The fourth part of this study explains the results and findings and the last section is about the conclusion and future recommendations.

LITERATURE REVIEW

E-commerce businesses are confronted with an unprecedented influx of enormous and diversified data from many sources in the quickly changing digital landscape. Businesses must successfully utilize this plethora of information through advanced analytics and data-driven decision-making to achieve a competitive edge. In order to realize the full potential of e-commerce data, the current literature study focuses on the creation and application of a strong business intelligence (BI) architecture that makes use of big data technologies and cloud computing (Madhura & Niyaz Panakaje, 2022; Sri, 2022). The emergence of big data has completely changed the e-commerce sector by offering insightful information about consumer behavior, market trends, and overall company performance. In this context, cloud computing is essential since it allows for real-time data processing, scalability, and analysis. Studies have emphasized the advantages of cloud computing for mobile e-commerce since it enables companies to manage enormous amounts of data effectively and affordably (Gawusu et al., 2022). Duplicate data increases processing time and can cause overfitting problems (Abdelhafez et al., 2023).

It takes a solid business intelligence framework to turn unprocessed data into insights that can be put to use. This framework should cover relational and distributed databases, as well as MapReduce parallelization, as well as various data storage, retrieval, and mining techniques. E-commerce enterprises may manage enormous databases and identify insightful patterns and trends by integrating these technologies (Williams, 2016). Precision marketing strategies must be used by e-commerce enterprises to increase client engagement and satisfaction. Businesses can segment their clients depending on their behavior and preferences by using the Recency, Frequency, and Monetary (RFM) model in conjunction with historical data analysis. This segmentation allows for more focused marketing campaigns, individualized advice, and higher conversion rates (Williams, 2016; Jukka, 2023).

Any successful e-commerce business must base its judgments on data. Businesses that include analytics in their decision-making processes can make informed judgments that lead to increased operational effectiveness and optimum e-commerce performance. Studies show that companies that employ data-driven tactics outperform their competitors and have higher levels of customer satisfaction. Despite the benefits of big data and business analytics, e-commerce enterprises confront significant security challenges. The literature highlights how important it is to create robust security mechanisms and risk assessment models in order to protect sensitive client data and keep the e-commerce system's integrity. Understanding potential risks in-depth and adopting precautionary measures are key to establishing and maintaining client trust (Pancić et al., 2023).

Business intelligence (BI) and analytics are fields which have advanced over the years. Even as "commercial business intelligence" sprang to prominence in the Nineties, the term "intelligence" changed into to start with related to synthetic intelligence studies. Later, the analytical element of BI—"business analytics"—emerged. The exam of sizeable and complex information sets has lately been known as "big data analytics" and "large information" respectively. BI&A is used right here to refer to business intelligence, analytics, and big data analytics together as one phrase. The current state, which draws from well-established database management practices, is represented by BI&A 1.0. It is dependent on structured data held in commercial relational database management systems (RDBMS) that was gathered through legacy systems. The analytical methods used in BI&A 1.0 have their roots in data mining methods from the 1980s and statistical approaches created in the 1970s (Rosas et al., 2023; Du, 2022; Choi et al., 2022).

The core elements of BI&A 1.0 are data management and warehousing, which include creating data marts and putting extraction, transformation, and load (ETL) procedures into practice. Data exploration is facilitated by tools for querying, online analytical processing (OLAP), and reporting, while business performance management (BPM) uses scorecards and dashboards to display performance measures. For a variety of business applications, including association analysis, clustering, classification, anomaly detection, and predictive modeling, BI&A 1.0 also contains statistical analysis and data mining techniques. Notably, top

commercial BI platforms provided by significant IT suppliers now include these capabilities (Zhuang et al., 2021; Jahani et al., 2023).

Business intelligence and analytics (BI&A) 1.0, the first generation of commercial enterprise intelligence structures, emerged in the early levels of the technology of data-driven selection-making. Enterprises might use some of BI&A 1. Zero technologies, methods, and practices to collect, procedure, and examine information are a good way to generate actionable insights and assist strategic selection-making. It played a key position in turning unstructured information into understanding and building the inspiration for later advances in big data analytics (Duan et al., 2020).

The centralization of data into data warehouses, reporting and dashboard capabilities, OLAP for multidimensional data analysis, and fundamental data visualization techniques were some of the main components of BI&A 1.0. With the aid of these capabilities, organizations were able to have a unified view of their operations, conveniently access historical and operational data, track key performance indicators (KPIs), and investigate data from numerous angles (Yao et al., 2017).

BI&A 1.0 offered organizations a number of advantages. Delivering timely and pertinent information, aided data-driven decision-making and allowed organizations to develop strategies based on solid evidence. Because less time and effort were needed for data retrieval and analysis, business processes were made more efficient and simplified as a result of data integration. Organizations were able to assess their performance and make progress towards their goals thanks to BI&A 1.0's assistance in measuring performance against predetermined targets. Additionally, it offered perceptions into consumer behavior, assisting companies in efficiently customizing their goods and services to match consumer wants (Smink et al., 2020).

BI&A 1.0 had significant drawbacks despite its benefits. It had trouble with the enormous numbers and many unstructured data types that were prevalent in the big data era, such as social media data or sensor data. Its weakness in real-time data processing made it difficult to analyze data as it was being generated. Performance limitations resulted from BI&A 1.0's inability to scale as data volumes increased rapidly. Furthermore, it lacked sophisticated predictive analytics skills and mostly concentrated on historical data analysis (Riegger et al., 2021).

Particularly in the context of e-commerce data analytics, big data analytics plays a crucial role in combining with BI&A 1.0 to overcome the issues given by unstructured data. The large volumes of unstructured data produced by e-commerce platforms, such as social networking data, customer reviews, and clickstream data, are beyond the capabilities of BI&A 1.0, which excels at handling structured data and doing conventional data analysis. By increasing its capacity to efficiently handle and extract insights from this unstructured data, big data analytics enhances BI&A 1.0 (Krafft et al., 2021). One of the key issues in e-commerce information analytics is coping with unstructured information, which is basically more complex and diverse than based records. Large statistics analytics answers like Apache Hadoop and Apache Spark may be used to address such unstructured information on a huge scale. By way of merging big data analytics with BI&A 1. Zero, e-commerce organizations might also seize, save, and analyze sizeable amounts of unstructured information in real time, allowing greater considerable and correct evaluation (Karim et al., 2021).

The generation of huge information unfolds, and conventional business intelligence (BI) answers, with their information garage, evaluation, and real-time processing skills, show insufficient for dealing with the complexity of unstructured data resources (Langerhuizen et al., 2021). To address this challenge, it is crucial to leverage existing BI and big data technologies effectively. While traditional BI primarily relies on internal operational and management systems for data, big data predominantly originates from sources such as microblogs, web pages, and other forms of data sharing on the internet (Muchenje & Seppänen, 2023). These data sources differ significantly in terms of collection, processing, storage, and potential applications. Consequently, a new platform architecture is required to accommodate these divergent factors, as illustrated in **Figure 1**.

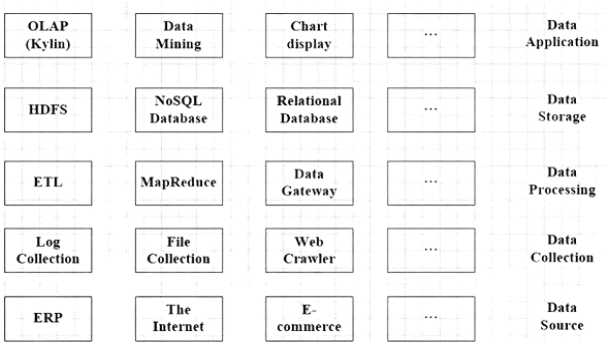

Figure 1. Business Intelligence and Big Data Platform

Resources on statistics for businesses encompass both internal and external data statistics. Information from the OA system, ERP system, financial statements system, as well as other related defined data, is included in internal statistics (Bose & Bhattacharjee, 2022). Unstructured information from the internet, such as hypertext, photos, and videos, makes up external records. Records acquisition adds a fresh approach to the traditional method of gathering data from the internet. For organized and unstructured facts, different processing techniques are used. Traditional records are still saved in relational databases while unstructured documents are converted into based information and stored in distributed databases (Chen et al., 2018; Williams, 2016). Distributed file systems (HDFS) and NoSQL databases are specifically used to store large amounts of data. The final data is mostly used for online processing, data mining, record visualization, and other related tasks (Akter & Wamba, 2016; Chen et al., 2012). (see **Figure 2** for details).

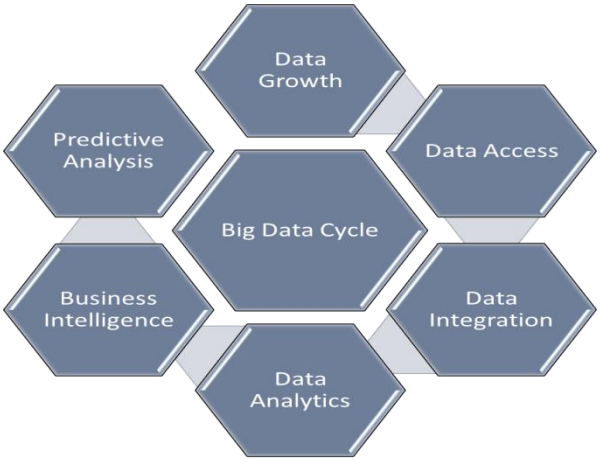

Figure 2. Big Data Cycle

METHODOLOGY

System log collection, network data collection, and data interface collecting are the three basic categories into which data collection methods for massive data in China can be divided. Each technique has a particular function in gathering information for analysis and decision-making. System Log Collection: In Chinese, the term "system log collection" refers to the process of gathering log messages produced by the logging subsystem of various hardware or software. These log messages offer insightful information on how the system functions, how users interact with it, and numerous occurrences. Organizations can acquire a better knowledge of their operations and spot trends or abnormalities in the data by extracting and analyzing system logs. In China, gathering network data particularly entails having access to and acquiring information from online resources using net crawlers technology. HTTP protocols are utilized by web crawler era to imitate browser behavior and hook up with net servers. Net crawlers navigate internet pages with the aid of using well-known aid locator (URL) addresses and retrieve pertinent statistics, after which ship it again for additional processing. Enterprises can use this approach to reap information from the diffusion of net sources, which include information portals, social networking websites, and e-commerce websites (Yu et al., 2021). Utilizing the utility programming interfaces (APIs) offered by corporate data assets is required for statistics interface collecting. Common business information APIs in China work with relaxing (Representational country switch) APIs to let businesses access and retrieve data. Agencies can get real-time or historical records relevant to their studies or business needs by integrating with such APIs. This method offers a tried-and-true method for gathering data from diverse sources.

The major participants in the study are e-commerce businesses that serve the Chinese market, from which data is gathered. Industry representation, a sizable market presence, data accessibility, a range of business sizes, geographic diversity, a variety of business models, adherence to data ethics and permission rules, and business model diversity were the criteria used to choose these companies. Access to their transactional data, customer behavior data, and marketing data is made available by the chosen businesses. To improve the performance of e-commerce and decision-making in the Chinese market, this data is used to create a thorough business intelligence framework and carry out big data analytics. The study ensures the privacy and anonymity of the data from the participating companies.

Research Framework

Utilizing big data technologies and cloud computing to improve decision-making and e-commerce performance, the framework intends to develop a solid business intelligence (BI) platform for e-commerce data analytics. In order to better target marketing efforts and increase consumer happiness, the framework also incorporates the application of precision marketing strategies based on the RFM model and historical data analysis. It also discusses risk assessment models and security measures to safeguard the e-commerce system's integrity. Here is a diagram showing how the framework looks in **Figure 3**.

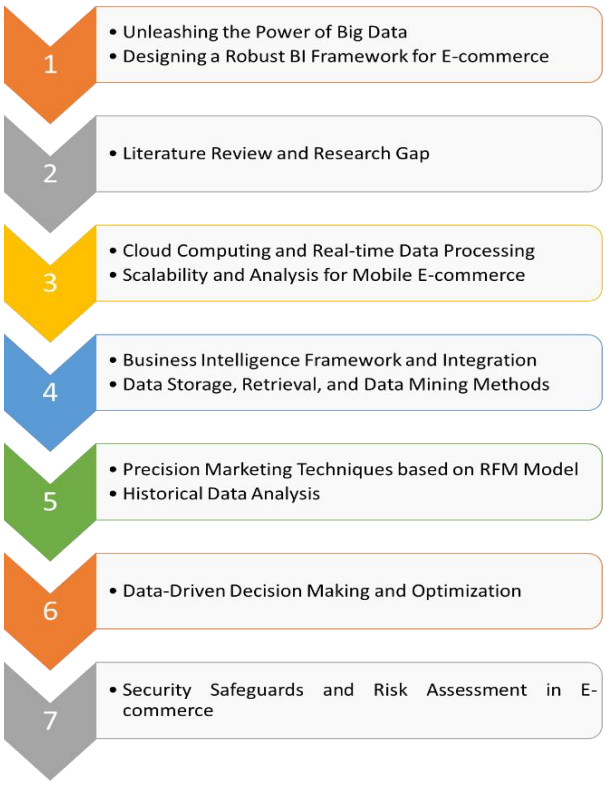

Figure 3. Research Process

This framework places a focus on how various elements must work together in order to create and deploy a successful business intelligence platform for e-commerce data analytics. Each phase builds on the one before it, helping to achieve the overall objective of enabling e-commerce businesses to make use of big data's potential for establishing a competitive edge.

DATA ANALYSIS AND FINDINGS

The analysis emphasizes the study's suggested solutions' practical ramifications for e-commerce players. It emphasizes the use of relational databases, parallel data mining, JavaEE

technology, targeted marketing, and cybersecurity precautions. With the use of these insights, businesses can strengthen security, improve consumer experiences, and optimize their platforms, giving them a competitive advantage in the fast-paced e-commerce environment.

K-mean Algorithm for Applying E-Commerce Big Data

The K-means technique is a widely used clustering technique that aims to divide hard and fast D entities into N clusters. The intention is to maximize the similarities inside every cluster while minimizing the similarities among specific clusters (Zhuang et al., 2021). The set of rules follows those steps:

To begin with, N data points are randomly decided on as centroids.

The gap from each statistics factor in the dataset to the centroids is calculated, and the statistics points are grouped into N clusters for that reason.

Based on the N units of information elements obtained in step 2, a new centroid is iteratively calculated.

Steps 2 and 3 are repeated till the gap between the final centroid and the previous centroid becomes small sufficient, indicating convergence.

In the k-means clustering procedure, all observations are carefully examined, and every remark is assigned to the cluster that has the closest centroid. The centroid and distance are essential criteria within the ok-manner algorithm. The centroid represents the middle of a cluster, which may be seen as a pattern or records point A in the dataset, reflecting its similarities with other information points. The choice of centroids notably influences the clustering consequences. Initially, the algorithm randomly selects centroids, resulting in an arbitrary partitioning of the information. To approximate the favored clustering outcomes, the centroids go through numerous iterations: statistics factors with similarities are grouped together, sharing the equal centroid. Due to the random nature of preliminary centroid choice, the very last outcomes might not be predictable, requiring more than one iteration to converge.

Distance, then again, serves as a measure of similarity in various applications and is utilized with specific formulas. Common distance measures encompass the New York distance, Euclidean distance, Minkowski distance, and Chebyshev distance. In clustering analysis, the Euclidean distance is extensively used due to its simplicity, intuitiveness, and ability to account for rotation and migration within the coordinate area (Gao, 2021). While the usage of the Euclidean distance, the initial similarity among items can nevertheless be determined based totally on their genuine similarities, as the gap value stays unchanged. If we denote the gap among item a and object b as D(x, y), then the gap characteristic d(x, y) must fulfill the following 3 situations:

Non-negativity: The distance between two objects can't be bad; its miles are usually equal to or more than 0.

Symmetry: The space between item a and object b is the same as the gap between object b and item a.

Triangle Inequality: The gap between object a and object c is by no means more than the sum of the distances between object a and item b, and between item b and item c.

Those attributes make certain that the gap metric maintains consistency and reliability in measuring the similarity or dissimilarity between gadgets. By means of applying suitable distance formulation, researchers can efficaciously determine the similarity and dissimilarity relationships among objects, allowing various analytical strategies including clustering, class, and similarity-based totally seek.

In conclusion, the K-means algorithm is a potent and popular clustering technique that allows for the effective clustering of data points based on similarities. The algorithm's iterative design aids in pinpointing more precise cluster centers, and the distance metric provides accurate comparisons of the similarity or dissimilarity of data points. Researchers can successfully analyze the data and get important insights into the underlying patterns and correlations by using appropriate distance calculations and centroid selection.

Applying the E-Commerce in Business Intelligence

In the era of big data, the fee of unbiased records itself is limited. The focal point lies in leveraging information for forecasting future tendencies and uncovering hidden insights. Chinese language natural decoction companies have embraced technological improvements, establishing online stores for promoting their merchandise. As a result, a giant amount of client intake information has been accumulated. Studying these data permits the segmentation of purchasers into distinct corporations primarily based on their purchasing conduct, permitting personalized advertising strategies. Consumer type helps the supply of tailored offerings and well-timed adjustment of business strategies in response to marketplace and customer modifications.

The RFM model, a broadly used patron classification technique, consists of a couple of factors. R (Recency) represents the time elapsed for the reason that the client's most current transaction, F (Frequency) denotes the number of interactions or purchases inside a selected period, and M (financial) indicates the whole fee of purchaser transactions inside a described time frame. By using assessing these three dimensions, RFM measures the patron's value primarily based on their real financial contribution (Malhotra & Rishi, 2021).

To conduct the evaluation, relevant records from a Chinese herbal remedy decoction e-trade internet site were crawled. The collected information was then processed, making use of precise information processing standards to smooth and prepare the facts. The resulting dataset includes patron facts, containing 3000 entries. Particularly, R represents the time c programming language for the reason that closing the purchase of Chinese language natural drugs, F indicates the frequency of buying decoction pieces, and M represents the total amount eaten up at the platform. Relevant information samples are offered in **Table 1**.

Table 1. Customer/Consumer Data

Customer Serial #	R	F	M
1	30	3	332.46
2	16	3	757.44
3	20	10	1381.64
4	6	23	3279.02
5	23	6	152.9
6	17	10	499.63
7	21	3	1720.18
8	16	7	105.43
9	13	20	971.61
10	11	7	762.8
11	17	4	1249.65
12	23	21	921.53
13	29	23	1009.43
14	15	22	1845.86
15	25	7	1667.71
16	25	22	327.21
17	11	24	752.19
18	17	3	1376.39
19	11	16	3273.77
20	24	5	147.65
21	21	22	494.38
22	19	13	1714.93
23	32	4	100.18
24	6	16	966.36
25	29	23	757.55
26	17	20	1244.4
27	6	22	916.28
28	30	12	1004.18
29	29	19	1840.61
30	17	24	1662.46

Table 2. Consumer Data After Data Normalization

Customer Serial #	R	F	M
1	30	3	330.83
2	16	3	755.81
3	20	10	1380.01
4	6	23	3277.39
5	23	6	151.27
6	17	10	498
7	21	3	1718.55
8	16	7	103.8
9	13	20	969.98
10	11	7	761.17
11	17	4	1248.02
12	23	21	919.9
13	29	23	1007.8
14	15	22	1844.23
15	25	7	1666.08
16	25	22	325.58
17	11	24	750.56
18	17	3	1374.76
19	11	16	3272.14
20	24	5	146.02
21	21	22	492.75
22	19	13	1713.3
23	32	4	98.55
24	6	16	964.73
25	29	23	755.92
26	17	20	1242.77
27	6	22	914.65
28	30	12	1002.55
29	29	19	1838.98
30	17	24	1660.83

To account for versions in numerical scales and devices amongst one-of-a-kind record attributes, they cannot be at once used for calculations. As an example, the total amount of product M bought via clients may also have a big numerical value in the variety of loads, while the frequency of product purchases within a specific term can be distinctly small and have minimum impact in comparison to the general consumption volume. To cope with this trouble and ensure compatibility between attributes and their respective scales, it's essential to normalize the data. This normalization process permits direct calculations using those standardized values in subsequent tiers. In our experiment, we applied a normalization method to technique the records, resulting in a dataset of three thousand instances. The normalized information can now be applied for addition analysis and calculations (**Table 2**).

In our implementation of the k-mean approach algorithm, we aimed to create three clusters by undertaking a maximum of three iterations. The Euclidean distance was used because of the degree of dissimilarity between records points. Because of the random initialization of centroids, the particular results obtained for every cluster might also vary. But, after conducting multiple experiments, we found that the clustering outcomes remained regular. This shows that the identified clusters may be utilized to research the cluster behavior of users and put in force targeted marketing strategies for specific agencies. Three iterations of the k-means method were used to cluster the customer data into three groups. Consistent results using Euclidean distance allowed for the analysis of user cluster behavior and the development of targeted marketing plans for each group (referred to as agencies 1, 2, and 3). The clusters offers the consumer company insightful information. The generated

clusters, labeled as agency one (**Figure 3**), two (**Figure 4**), and three (**Figure 5**), are displayed for the consumer company.

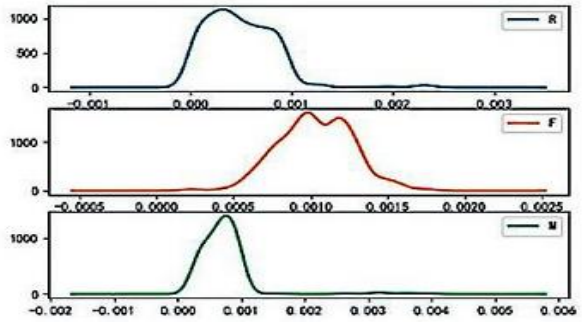

Figure 3. Consumer Group 1

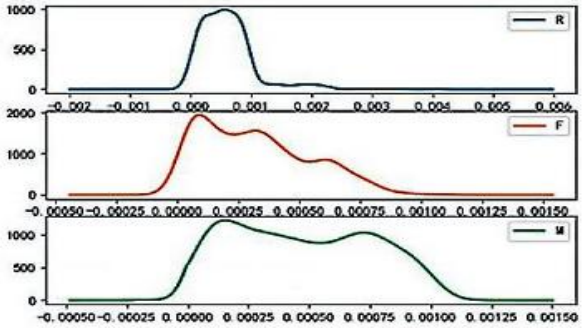

Figure 4. Consumer Group 2

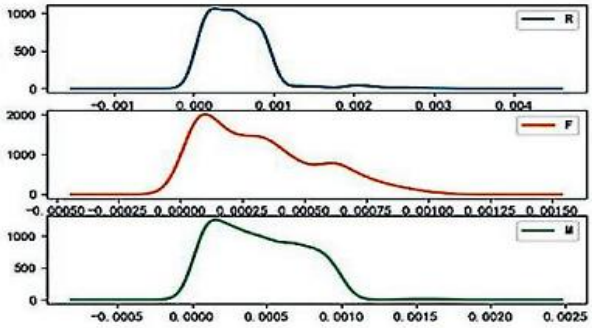

Figure 5. Consumer Group 3

The consumer segmentation analysis found 3 wonderful groups. Group of customer 1 includes fairly treasured and capable clients who spend a shorter time c programming language (R) on our platform and show off a better general consumption (M). Although they constitute a small share, they contribute significantly to our business. Priority should be given to aid allocation and one-to-one advertising and marketing to decorate their loyalty and pride. In contrast, group customer 2 demonstrates mild buy frequency (F), shorter time intervals (R), and slight general intake (M). Their consumption behavior is uncertain, requiring well-timed engagement and proactive monitoring to expect adjustments. Sooner or later, the institution of group 3 represents ordinary clients with widespread buy frequency (F), slight durations between transactions (R), and lower general consumption (M). They're mainly interested in discounted traditional Chinese time-based medicinal drug decoction products. By means of information about those patron companies and imposing tailor-made advertising strategies, we will optimize customer relationships, expand customer lifecycles, and maximize commercial enterprise effects.

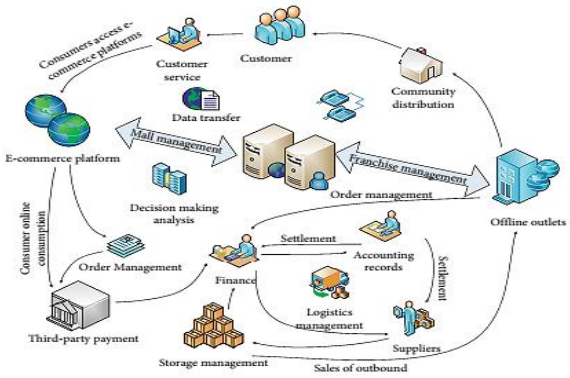

Figure 6. Process of Business Transaction of Electronic Commerce

The transaction enterprise system of the e-trade system primarily based on large facts is illustrated in **Figure 6**. The device is composed of 3 layers: the cloud service layer, digital aid layer, and bodily resource layer. The network architecture is depicted. The cloud provider layer encompasses Infrastructure-as-a-service (IaaS), Platform-as-a-carrier (PaaS), and software program-as-a-provider (SaaS) fashions. The virtual useful resource layer enables resource integration inside the cloud computing surroundings, forming logical virtual aid pools that share computing, garage, and network assets. The physical aid layer connects the cloud infrastructure, which includes servers, software program resources, and alertness modules, to the community through the cloud terminal era to help hardware necessities for implementing cloud offerings.

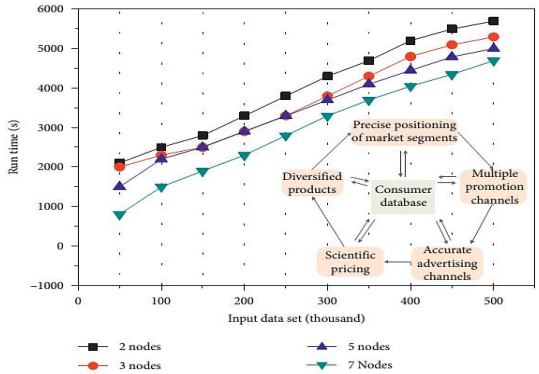

Figure 7. Marketing-Specific Processing and Recommendation Efficiency Test

The concrete procedure of precision advertising and marketing is depicted in **Figure 7**. The process starts off evolving with the established order of a client database, which serves as the foundation for enforcing precise advertising strategies. Within the context of e-commerce,

client records are measured, calculated, and analyzed to develop greater correct marketing strategies. By way of leveraging the consumer database, organizations can behavior in-depth analysis and provide specific insights on pricing, merchandise, distribution channels, and promotional activities. This allows them to attain greater correct communication, centered marketing, personalized product services, and fee-added services. **Figure 7** illustrates the experimental evaluation of algorithmic performance, in particular focusing on the running pace of the set of rules. The set of rules for jogging time serves because of the assessment metric, in which shorter walking times indicate higher computational strength and extra advice performance. The test entails exclusive statistics sets of various sizes, along with from 100,000 to 500,000 statistics records units. The set of rules is accomplished on specific information Node nodes, along with 2 nodes, 3 nodes, 5 nodes, and 7 nodes, respectively. By way of evaluating the walking time of the algorithm across one-of-a-kind nodes, the experimental outcomes show the set of rules performance. The figure above presents the experimental outcomes, indicating that the progressed set of rules proposed within the paper shows faster going-for-walk instances and higher recommendation performance in comparison to conventional techniques. This suggests that the recommended set of rules offers superior computational strength, enabling greater efficient and effective recommendation structures.

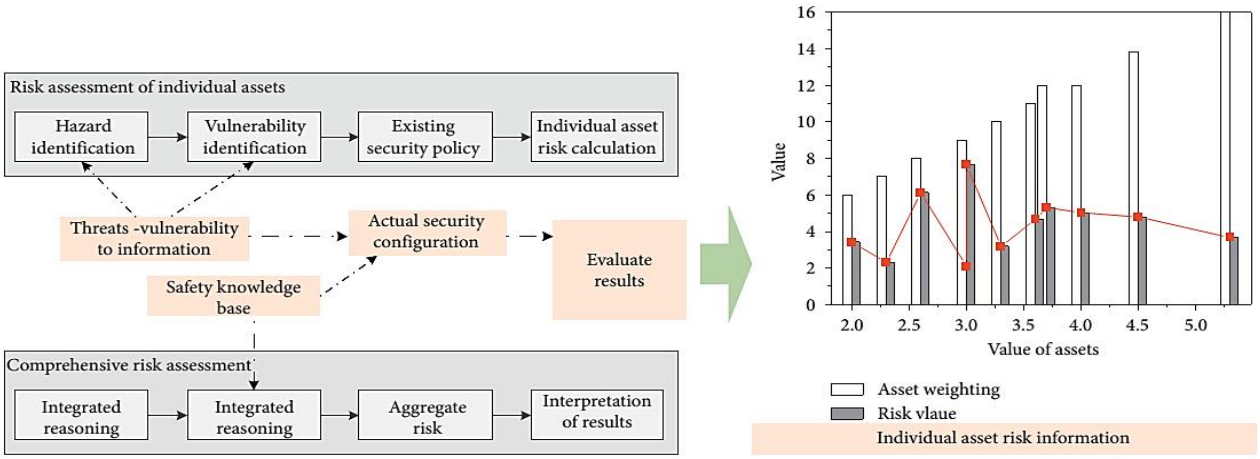

Figure 8. Distribution Points with Demands and Routes

This paper addresses risk assessment within the context of an e-commerce device based totally on cloud computing. A corresponding hazard assessment theoretical version is proposed, which identifies the threats and vulnerabilities associated with information belongings. Threats refer to external factors that can pose risks to the system, together with machine screw-ups, environmental dangers, or human-associated troubles. Vulnerabilities, on the other hand, represent inner weaknesses within the gadget, along with the assessment of cloud service modes and the safety scale of the service issuer. **Figure 8** illustrates the risk evaluation model, which considers information on property, threats, and vulnerabilities. Via a mixture of questionnaire survey outcomes and expert ratings, every danger detail is assigned a numerical value. This allows the calculation of chance statistics for individual assets. Based totally on a danger grading classification trend, the chance values are labeled into different levels, which include low, medium, high, and additional excessive-threat degrees. The evaluation suggests that the gadget risks, particularly those associated with hardware and the e-trade platform, are at a medium degree. Areas of problem encompass physical protection, gadgets admission to, data backup control, and vulnerabilities in cloud facts security and safety management configuration. In reaction to those recognized risks, the paper proposes corresponding development techniques. Those consist of enhancing the physical surroundings of the PC room, enforcing access management policies with better safety levels, optimizing cloud information storage strategies through negotiations with service vendors, strengthening the technical skills of personnel, and enhancing the undertaking of responsibilities. By means of conducting a complete research of statistics assets within the e-commerce machine and comparing the present-day protection hazard level with the use of a numerical method, this look provides appropriate management techniques for boosting the machine's protection overall performance. Additionally, it offers valuable insights for selecting appropriate cloud carrier vendors. **Figure 8** offers the risk evaluation version, however without the particular content mentioned earlier, illustrating the technique of comparing dangers related to data assets in the e-commerce gadget.

CONCLUSION

This article analyzed and addressed the challenges faced with the aid of current e-trade structures, identifying their characteristics and proposing solutions. Key improvements were made in information storage, parallelization of information mining algorithms the use of MapReduce, and the improvement of a complete web mall device based on Hadoop and HBase. The mixing of those technologies and structures turned into a finished through Web Service telecommunication technology, resulting in a cohesive and practical e-trade atmosphere. Furthermore, the object

explored the impact of cloud computing on mobile e-commerce, delving into relevant theories, service modes, architectures, core technology, and applications. Leveraging the benefits of cloud computing for statistics garage and processing, the object emphasized the significance of timely and efficient analysis and processing of huge-scale information in traditional e-commerce structures. It highlighted the need for e-commerce businesses to capitalize on the possibilities supplied through massive facts, utilizing fact-mining techniques to extract precious insights, regulate development techniques, and beautify precision advertising efforts. Through doing so, businesses can provide higher-first-rate offerings, improve client delight, increase monetary benefits, and beautify their typical competitiveness.

Additionally, the object targeted the application of huge information and traditional commercial enterprise intelligence in the information evaluation of digital business organizations, with a particular emphasis on the k-mean approach algorithm for clustering evaluation. Through efficiently mining patron intake information from e-commerce websites, cluster analysis enabled the segmentation of customers into awesome groups. This segmentation, based on the characteristics of different consumer companies, facilitated focused marketing initiatives and the success of differentiated advertising goals.

Ordinary, the findings of this newsletter spotlight the importance of embracing massive information technologies and methodologies in the e-trade sector. Leveraging cloud computing, records mining, and enterprise intelligence strategies can free up the capability of tremendous records assets, permitting enterprises to make informed choices, drive innovation, and gain a competitive facet. These advancements are in particular relevant within the Chinese context, in which the e-commerce industry is hastily developing and evolving.

As e-commerce firms in China harness the energy of large records, they're poised to seize marketplace opportunities, improve customer pleasure, and pressure monetary boom. The integration of data mining and business intelligence provides a strong foundation for reading customer conduct, tailoring marketing techniques, and handing over personalized studies. With the aid of leveraging the insights derived from massive statistics analytics, establishments can adapt to changing market dynamics, enhance their center skills, and contribute to the continued development of the e-commerce industry in China.

Despite the substantial developments covered in the article, there are still certain limitations and opportunities for future development in the use of big data technology in the e-commerce industry. One restriction is the potential difficulty of data privacy and security issues, especially when working with sizable amounts of sensitive customer data. Future studies should concentrate on creating effective data protection mechanisms and ensuring compliance with pertinent rules to address this issue. The difficulty and expense of creating and maintaining big data infrastructure is another restriction. It is crucial to solve scalability challenges and find affordable ways to store and process data effectively as the volume of data keeps increasing. Future suggestions include researching developments in cloud computing and distributed computing methods to handle the constantly growing amount of data. The k-means technique was predominantly employed in the article's clustering study. The use of more advanced machine learning and artificial intelligence approaches to glean deeper insights from massive data may be investigated in the future. Techniques like neural networks, deep learning, or natural language processing could be among them. These could provide a more thorough and accurate analysis of consumer behavior.

IMPLICATIONS

Practical Implications

Beyond the creation of e-commerce systems, this research explores the practical implications for many players in the digital business world. The suggested methods, including the usage of relational and distributed databases, parallelization of data mining algorithms through MapReduce, and exploitation of JavaEE technology, can assist e-commerce companies all over the world in creating and refining their online platforms. Organizations may create a seamless and user-friendly purchasing experience by putting these innovations into practice, which will boost customer happiness and loyalty.

The research's findings are also useful for marketing experts, notably for e-commerce precision marketing. Marketers can create personalized plans suited to various client segments by using the RFM model and historical data analysis. Delivering highly targeted promos, ads, and product recommendations is the strategy's key step in improving consumer engagement and boosting conversion rates for global e-commerce enterprises.

For cybersecurity specialists working in the field of e-commerce platforms, the research also offers useful information. These experts can efficiently assess and reduce risks by using the risk assessment model and detecting potential threats and vulnerabilities. The suggested security enhancement solutions, which incorporate better physical security, access control procedures, optimized cloud data storage, and improved technical skill sets, provide helpful advice for boosting the security posture of e-commerce platforms around the world.

Theoretical Implications

Integration of big statistics and commercial enterprise Intelligence: This research contributes to the theoretical knowledge of the mixing of massive records and commercial enterprise intelligence in e-trade evaluation. By leveraging massive records technology and methodologies, organizations can free up treasured insights from big-scale data sets, permitting informed selection-making, precise marketing, and advanced overall performance. The look expands the theoretical know-how base by way of showcasing the utility of data mining algorithms, clustering analysis, and recommendation structures in the context of e-trade organizations in China.

Cloud Computing and Cellular E-trade: The theoretical

implications of this study lie inside the exploration of the impact of cloud computing on mobile e-commerce. With the aid of inspecting the theories, service modes, architectures, and center technology, the observation affords insights into the advantages and packages of cloud computing in processing and analyzing huge-scale records in actual time. This contributes to the knowledge of the position of cloud computing in facilitating the growth and success of mobile e-trade in China (Lutfi et al., 2023).

Threat evaluation and protection fashions: The studies contribute to the theoretical information of danger assessment and protection models in e-commerce structures. By way of growing a threat assessment model that identifies threats and vulnerabilities, they have a look at affords a theoretical framework for comparing and dealing with dangers in the e-trade domain. This theoretical contribution enhances the expertise on the significance of incorporating safety features and risk evaluation practices within the layout and operation of e-trade structures (Himeur et al., 2023).

Standard, the sensible implications of this study offer guidance to e-trade businesses in China for gadget design, precision advertising, and danger management, whilst the theoretical implications strengthen the information of the integration of big information, cloud computing, and security models in the e-commerce domain.

REFERENCES

Abdelhafez, A. A., Ismael, O., & Elkady, H. (2023). IoT-Based Data Size Minimization Using Cluster-Based-Similarity- Elimination. International Journal of Communication Networks and Information Security (IJCNIS), 15(2), 34-50. https://doi.org/10.17762/ijcnis.v15i2.6151

Akter, S., & Wamba, S. F. (2016). Big data analytics in E-commerce: a systematic review and agenda for future research. Electronic Markets, 26(2), 173-194. https://doi.org/10.1007/s12525-016-0219-0

Bose, S., & Bhattacharjee, S. (2022). Big Data , Data Analytics and Artificial Intelligence in Accounting : An Overview. In S. Akter, & S. F. Wamba (Eds.), Handbook of Big Data Research Methods (pp. 32-51). Edward Elgar Publishing. https://doi.org/10.4337/9781800888555.00007

Chandrakar, I., & Hulipalled, V. R. (2022). Improved technique for preserving privacy while mining real time big data. International Journal of Communication Networks and Information Security, 14(1), 86-92. https://doi.org/10.17762/ijcnis.v14i1.5187

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS quarterly, 1165-1188. https://doi.org/10.2307/41703503

Chen, H., H.L.Chiang, R., & C. Storey, V. (2018). Business Intelligence and Analytics: From Big Data To Big Impact. MIS Quarterly, 36(4), 1165-1188. https://doi.org/10.2307/41703503

Choi, L. K., Panjaitan, A. S., & Apriliasari, D. (2022). The Effectiveness of Business Intelligence Management Implementation in Industry 4.0. Startupreneur Business Digital (SABDA Journal), 1(2), 115-125. https://doi.org/10.34306/sabda.v1i2.106

Du, J. (2022). Project Attribute Information and Initiator's Commitment in Crowdfunding Message Strategy: An Empirical Investigation of Financing Performance in Agri-Food Crowdfunding. Sustainability (Switzerland), 14(19). https://doi.org/10.3390/SU141912702

Duan, Y., Cao, G., & Edwards, J. S. (2020). Understanding the impact of business analytics on innovation. European Journal of Operational Research, 281(3), 673-686. https://doi.org/10.1016/j.ejor.2018.06.021

Gawusu, S., Zhang, X., Jamatutu, S. A., Ahmed, A., Amadu, A. A., & Djam Miensah, E. (2022). The dynamics of green supply chain management within the framework of renewable energy. International Journal of Energy Research, 46(2), 684-711. https://doi.org/10.1002/er.7278

Himeur, Y., Elnour, M., Fadli, F., Meskin, N., Petri, I., Rezgui, Y., ... & Amira, A. (2023). AI-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives. Artificial Intelligence Review, 56(6), 4929-5021. https://doi.org/10.1007/s10462-022-10286-2

Jahani, H., Jain, R., & Ivanov, D. (2023). Data science and big data analytics: A systematic review of methodologies used in the supply chain and logistics research. Annals of Operations Research, 1-58. https://doi.org/10.1007/s10479-023-05390-7

Jukka, T. (2023). Does business strategy and management control system fit determine performance?. International Journal of Productivity and Performance Management, 72(3), 659-678. https://doi.org/10.1108/IJPPM-11-2020-0584

Madhura, K., & Niyaz Panakaje, D. (2022). Customer Perception in Relationship Between Social-Media and Purchasing Behavior of Fashion Products. International Journal of Case Studies in Business, IT, and Education. https://doi.org/10.47992/ijcsbe.2581.6942.0185

Karim, M. W., Chowdhury, M. A. M., Al Masud, M. A., & Arifuzzaman, M. (2021). Analysis of factors influencing impulse buying behavior towards e-tailing sites. Contemporary Management Research, 17(2), 97-126. https://doi.org/10.7903/CMR.20457

Krafft, M., Kumar, V., Harmeling, C., Singh, S., Zhu, T., Chen, J., Duncan, T., Fortin, W., & Rosa, E. (2021). Insight is power: Understanding the terms of the consumer-firm data exchange. Journal of Retailing, 97(1), 133-149. https://doi.org/10.1016/j.jretai.2020.11.001

Langerhuizen, D. W. G., Brown, L. E., Doornberg, J. N., Ring, D., Kerkhoffs, G. M. M. J., & Janssen, S. J. (2021). Analysis of Online Reviews of Orthopaedic Surgeons and Orthopaedic Practices Using Natural Language Processing. The Journal of the American Academy of Orthopaedic Surgeons, 29(8), 337-344. https://doi.org/10.5435/JAAOS-D-20-00288

Lawani, A., Reed, M. R., Mark, T., & Zheng, Y. (2019). Reviews and price on online platforms: Evidence from sentiment analysis of Airbnb reviews in Boston. Regional Science and Urban Economics, 75, 22-34. https://doi.org/10.1016/j.regsciurbeco.2018.11.003

Li, Y., Xu, Z. S., & Wang, X. X. (2023). Methods and applications of sentiment analysis with online reviews. Kongzhi Yu Juece/Control and Decision, 38(2), 304-317. https://doi.org/10.13195/J.KZYJC.2022.1788

Lutfi, A., Alrawad, M., Alsyouf, A., Almaiah, M. A., Al-Khasawneh, A., Al-Khasawneh, A. L., ... & Ibrahim, N. (2023). Drivers and impact of big data analytic adoption in the retail industry: A quantitative investigation applying structural equation modeling. Journal of Retailing and Consumer Services, 70, 103129. https://doi.org/10.1016/j.jretconser.2022.103129

Ma, G., Ma, J., Li, H., Wang, Y., Wang, Z., & Zhang, B. (2022). Customer behavior in purchasing energy-saving products: Big data analytics from online reviews of e-commerce. Energy Policy, 165, 112960. https://doi.org/10.1016/J.ENPOL.2022.112960

Malhotra, D., & Rishi, O. (2021). An intelligent approach to design of E-Commerce metasearch and ranking system using next-generation big data analytics. Journal of King Saud University-Computer and Information Sciences, 33(2), 183-194. https://doi.org/10.1016/j.jksuci.2018.02.015

Maroufkhani, P., Iranmanesh, M., & Ghobakhloo, M. (2023). Determinants of big data analytics adoption in small and medium-sized enterprises (SMEs). Industrial Management & Data Systems, 123(1), 278-301. https://doi.org/10.1108/IMDS-11-2021-0695

Muchenje, G., & Seppänen, M. (2023). Unpacking task-technology fit to explore the business value of big data analytics. International Journal of Information Management, 69, 102619. https://doi.org/10.1016/j.ijinfomgt.2022.102619

Pancić, M., Ćućić, D., & Serdarušić, H. (2023). Business Intelligence (BI) in Firm Performance: Role of Big Data Analytics and Blockchain Technology. Economies, 11(3). https://doi.org/10.3390/economies11030099

Riegger, A. S., Klein, J. F., Merfeld, K., & Henkel, S. (2021). Technology-enabled personalization in retail stores: Understanding drivers and barriers. Journal of Business Research, 123, 140-155. https://doi.org/10.1016/j.jbusres.2020.09.039

Rosas, N. G. T., Paredes, S. A. V., Gonzales, J. L. R., Kaviska, R. S., Cano, A. A. C., & Taherdoost, H. (2023). Business Intelligence as a Key Driver of E-business Enterprises; Prospects and Challenges. In 2023 International Conference on Inventive Computation Technologies (ICICT) (pp. 1000-1006). IEEE. https://doi.org/10.1109/ICICT57646.2023.10134229

Shamim, A., Qureshi, M. A., Jabeen, F., Liaqat, M., Bilal, M., Jembre, Y. Z., & Attique, M. (2021). Multi-attribute online decision-making driven by opinion mining. Mathematics, 9(8). https://doi.org/10.3390/MATH9080833

Shittu, E., & Weigelt, C. (2022). Accessibility in sustainability transitions: U.S. electric utilities' deployment of solar. Energy Policy, 165. https://doi.org/10.1016/j.enpol.2022.112942

Smink, A. R., Van Reijmersdal, E. A., Van Noort, G., & Neijens, P. C. (2020). Shopping in augmented reality: The effects of spatial presence, personalization and intrusiveness on app and brand responses. Journal of Business Research, 118, 474-485. https://doi.org/10.1016/j.jbusres.2020.07.018

Song, J., Xia, S., Vrontis, D., Sukumar, A., Liao, B., Li, Q., Tian, K., & Yao, N. (2022). The Source of SMEs' Competitive Performance in COVID-19: Matching Big Data Analytics Capability to Business Models. Information Systems Frontiers, 24(4), 1167-1187. https://doi.org/10.1007/S10796-022-10287-0

Sri, V. D. S., & Vemuru, S. (2022). A Framework for Uncertain Cloud Data Security and Recovery Based on Hybrid Multi-User Medical Decision Learning Patterns. International Journal of Communication Networks and Information Security, 14(1s), 136-152. Retrieved from https://ijcnis.org/index.php/ijcnis/article/view/5637

Williams, S. (2016). Business intelligence strategy and big data analytics: a general management perspective. Morgan Kaufmann. https://doi.org/10.1016/C2015-0-01169-8

Yao, X., Zhou, J., Zhang, J., & Boër, C. R. (2017). From intelligent manufacturing to smart manufacturing for industry 4.0 driven by next generation artificial intelligence and further on. In 2017 5th international conference on enterprise systems (ES) (pp. 311-318). IEEE. https://doi.org/10.1109/ES.2017.58

Yu, R., Wu, C., Yan, B., Yu, B., Zhou, X., Yu, Y., & Chen, N. (2021). Analysis of the Impact of Big Data on E-Commerce in Cloud Computing Environment. Complexity, 2021. https://doi.org/10.1155/2021/5613599

Zhuang, W., Wang, M. C., Nakamoto, I., & Jiang, M. (2021). Big Data Analytics in E-commerce for the US and China Through Literature Reviewing. Journal of Systems Science and Information, 9(1), 16-44. https://doi.org/10.21078/JSSI-2021-016-29