Journal of Information Systems Engineering and Management
2025, 10(11s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Enhancing Search Functionality-based Customer Analysis Expending Big Data in E-commerce

Pudi Sreelakshmi[1], R.Vanidhasri[2], Jose.P[3], Priyadharsshini.C[4], B.Vidhya[5]

[1]*Associate Professor, Department of Electronics and Communication Engineering, Audisankara College of Engineering and Technology, Gudur. sree1612@gmail.com*

[2]*Assistant Professor, Computer Science and Business Systems, Panimalar Engineering College, Chennai. vanidhasris@gmail.com*

[3]*Associate Professor, Department of CSE, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai. drjosep@veltech.edu.in*

[4]*Assistant Professor, CSE Department, Dhanalakshmi Srinivasan University, School of Engineering and Technology, Perambalur. priyacs516@gmail.com*

[5]*Assistant Professor, Department of Computer Science and Engineering, Rajiv Gandhi College of Engineering and Technology, Puducherry. vidhyarajiv19@gmail.com*

ARTICLE INFO

Received: 06 Nov 2024
Revised: 24 Dec 2024
Accepted: 05 Jan 2025

ABSTRACT

Customer behavior analytics has been included in many systems, but the field is still young and undeveloped, with much room for improvement. Big data is one of the newest and most promising technological breakthroughs. It can fundamentally alter how businesses analyze and turn consumer behavior into insightful knowledge. It is possible to analyze data effectively even using decision trees. This research attempts to investigate the benefits of implementing BDA in e-commerce to vendors and customers based on the significance of big data analysis and its advantages to e-commerce operations. By enhancing client loyalty, BDA gives electronic sellers, or "E-vendors," the competitive edge they need to comprehend consumer behavior and boost revenue. Additionally, recommendation systems built on BDA personalize users' search and buying experiences. Nevertheless, there are drawbacks to using BDA in e-commerce, namely the development of a shopping addiction. In addition, orientations that leverage BDA and the application of quantum computing to address current e-commerce issues are investigated to improve the evidence for decision-making that may prove useful in subsequent studies.

Keywords: Big Data Analytics, E-vendors, businesses analysis, Customer Behavior Analysis.

INTRODUCTION

This Big data is an enormous volume of unstructured data that is collected from many sources, such as the web and commercial organisations, and arrives at us quickly. Because it is unstructured, handling it with typical database management methods is difficult and time-consuming. We can refer to it as an expanding torrent. Thus, storage, search, distribution, transfer, analysis, and visualization are the main challenging problems in large data processing [1]. Formerly, the word "analytics" denoted the examination of current data to discover possible trends and to assess the results of particular choices or occurrences that can be employed to obtain a variety of insightful business intelligence insights. Finding all the hidden information in the massive amount of data gathered from a wide range of sources is today's greatest difficulty. Big Data analytics then enters the scene. Customer analytics, often known as customer behaviour analysis, is one of them.

By enabling businesses to forecast consumer behaviour, customer analytics helps transform large data into significant value. This is achieved by enhancing sales, inventory management, fraud detection, and many other applications. There are many different strategies that may be employed, but one that sticks out is the categorisation process that uses decision trees and is effectively applied in consumer analytics. Furthermore, the increased use of the internet led to the initiation of Electronic Commerce (e-commerce) endeavours. E-commerce refers to the online process of buying, selling, and exchanging goods, services, and information. E-commerce businesses have benefited from big data analysis by streamlining their operations and boosting sales at the same time.

Big Data analytics, or BDA, is the practice of examining large amounts of data to offer a picture for business decision-making [2]. BDA has a wide range of effects on e-commerce activities, including bettering business revenue and gaining insight into user behaviour and interests. For instance, the user's sentiment may be inferred from mouse cursor movements.

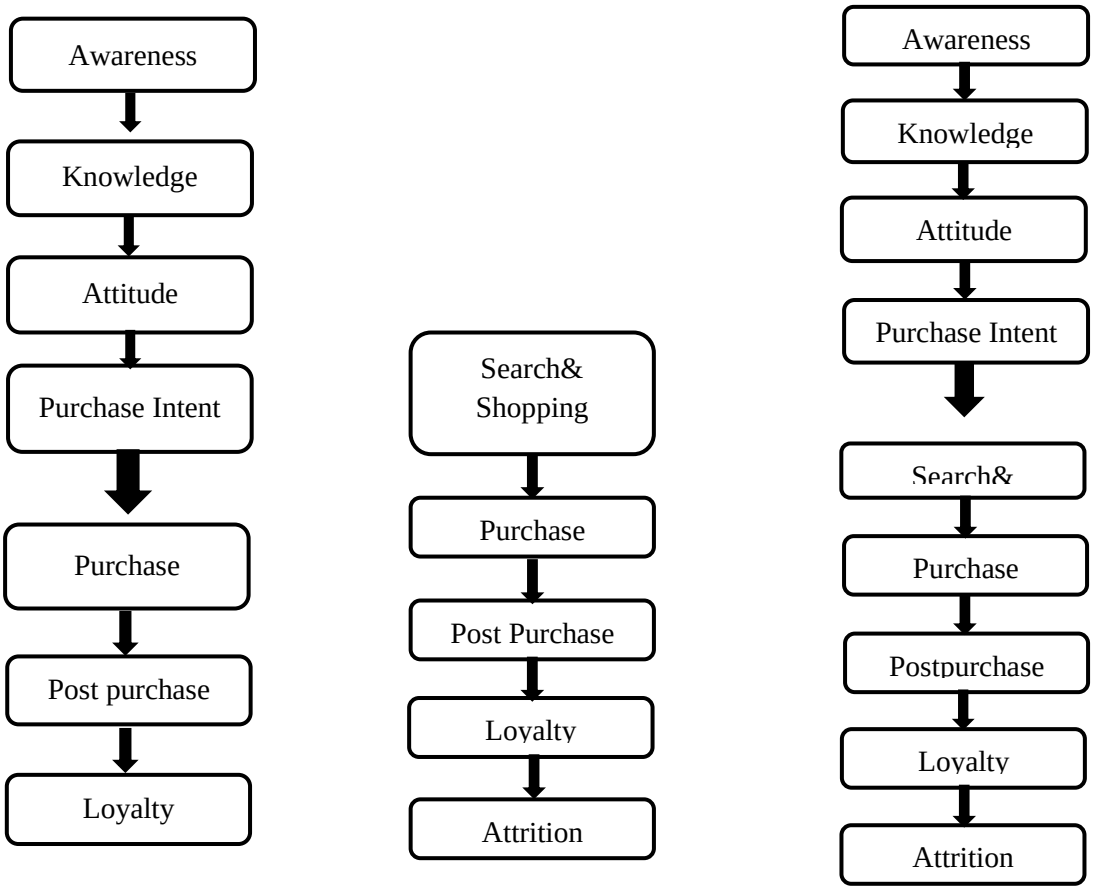

Figure 1.1: The Customer Purchasing Process: Offline and Online

The first flow chart on Figure 1.1 depicts the conduct of typical offline clients. Customers understand their requirements and desires above all else. Following that, buyers begin to look up as much information as they can about the products from various sources (friends, the internet, etc.) on features and price. Following that, consumers would assess the options in light of their personal preferences. At last, the choice to buy is made. Given that online buying is rapidly shifting to mobile devices, 54.8% of China's 410 million online shoppers did their purchasing on a mobile device at the close of 2015. Online shoppers finish the entire transaction by looking, enquiring, buying, paying, and monitoring delivery, among other things, online.

Big data has a variety of effects on e-commerce operations, some of which may present new chances for e-commerce development. This study will examine the various effects of big data on the e-commerce experience for vendors and users, or customers [3], based on relevant research and reviews. The current study will assist researchers in enhancing the use of BDA in e-commerce through an analysis of the experiences of vendors and consumers.

- To determine big data analytics definitional viewpoints
- To discern the features of big data in relation to e-commerce
- To investigate the many forms of big data in e-commerce
- To demonstrate big data's commercial potential in e-commerce
- To offer recommendations for addressing the difficulties associated with using big data in e-commerce

The overall goal of this study is to give a comprehensive representation of what big data means in the context of e-commerce. This paper is divided into five main sections. First, we provide the findings of our systematic review and describe the methodological gestalt in section 2. After gathering this data, we define the function of big data in e-commerce and point out competing definitional viewpoints in section 3. Second, we examine the unique characteristics and categories of big data in e-commerce in section 4. Thirdly, we list some forms of business value that can be obtained in the e-commerce industry using BDA in section 5. Lastly, we list the issues and offer ways to resolve them so that BDA can continue to expand in e-commerce.

RELATED WORKS

Data analysis and demonstrated that the proposed big data analytics service-oriented architecture (BASOA) is feasible for enhancing BI and enterprise information systems. Over the past 20 years [4], business intelligence and analytics (BI&A) and the related field of big data analytics have grown in importance in both the academic and business communities. Business analytics was first introduced as the primary analytical component in the late 2000s. Later, in 2005, architecture for enhanced business intelligence was developed, consisting of two infrastructure types: business integration and information integration. The goal of this architecture was to increase value and reduce action time through real-time analytics.

Looking at business processes from a strategic management angle, BDA and IoT capabilities can provide a lot of value if they are backed by high-quality data, which will improve competitive advantage [5]. One new paradigm in the manufacturing sector is smart manufacturing, which is the fourth revolution in the field. Since the Internet of Things is still in its early stages of development and presents many obstacles to achieving a more intelligent life, many academics have been drawn to the field. Research on the scientific measurement of the Internet of Things indicates that application, communication protocols, operating systems, and other related topics are currently hot research topics.

Big data analytics are obviously essential for strategic innovation; however opinions among academics over the precise definition of this phrase are divided. Numerous studies have looked into the capabilities of big data analytics, and this has led to some theoretical and empirical discussion about problems related to innovation, technology, and strategic management (among other things) [6], particularly in the context of digital networks. The current study aimed to determine how big data analytics-related innovation processes differ from the traditional innovation pathways for e-commerce research in an effort to broaden the body of existing literature.

Since business informatics was one of the organization's main tools for digital development, it had a special function [7]. Policymakers should support an innovative business model since it is becoming more and more essential for effective institution improvement. It is important to note that the writers have seen not only corporate representatives but also customers reacting in an unavoidable way and changing their behavior. The business community, government services, and consumers are some of the segments that are examined in this article. There is varying degrees of digitalisation among them, as well as complete absence of it. A detailed analysis on the path and the impact of the COVID-19 crisis on e-commerce can be found in the article.

Even though data is crucial to the success or failure of e-commerce businesses, not enough study has been done on efficient ways to use it to provide insightful data for e-commerce management and strategy. The goal of this special issue is to draw attention to the necessity of doing more methodical study and research in this crucial field [8]. Many other forms of data should be included in the research, such as text, audio, and video data; stocks of archived data that already exists and lows of current data streams; transaction-based, opinion-related, and temporally-changing data.

The research provided a useful suggestion. A preference matrix was created to record user preferences based on product attributes since people often communicate their opinions informally. We also use weighted association rules to find these patterns, which improves the quality of recommendations [9]. The development of a user behavior-based recommendation system that determines users' preferences based on implicit product attribute ratings is the most important contribution. Moreover, a sequential buying pattern is utilized in this strategy to enhance the quality of offers. This study ignores the other issues with the previously described RSs in favor of focusing on fixing sparsely and using the product's attributes to gauge their similarity.

Growing their client base and turning a profit are key components of e-commerce firms' success in order to remain competitive. Data is essential for success since it is relevant, accurate, and helps make wise decisions [10]. Although there is universal agreement about the importance of data, there are few general findings about what factors lead to BDA adoption—particularly for e-commerce firms. Analytics and its variations have been injected to comprehend the present and anticipate the future by the evolving business environment. Startups in the e-commerce industry operate in a dynamic business climate, which calls for careful evaluation. Analytics have shown to be an essential tool in the e-commerce start-up sector, which has also changed into a competitive advantage-seeking battlefield.

METHODS AND MATERIALS

3.1 E-commerce

Artificial intelligence (AI)-driven commercials are a common sight in our daily lives; they are utilized for purchasing, working, and other purposes. AI changes marketing to an automation system seen in social media.

E-commerce enhances the resources available to marketers for online buying and selling. Modern trade has developed into a powerful tool that connects buyers and sellers. The global economy is being greatly impacted by the internet, which has also increased the rate at which foreign currencies are exchanged. In recent years, AI has matured into marketing. It now assists in promoting marketing plans and frees up marketers' time by summarising their manual duties. There are numerous ways to handle e-commerce, and one of them is customer lifetime value, which is based on how devoted a consumer is to a particular brand and how they will stand behind it [11].

E-commerce simplifies work strategies and supports individuals in making decisions by enabling them to do so more quickly and efficiently. It also creates intelligent seals, which enable the sealer to identify the products that customers prefer from those that are often overlooked.

3.2 Big data types employed in e-commerce

Online sales of goods and services, such as those conducted on Amazon, Zappos, eBay, Expedia, and other sites, or through recurring transactions, such as those conducted on Netflix, Match.com, LinkedIn, and other sites, are referred to as e-commerce. Online retailers such as Amazon and Netflix collect a variety of data (such as purchases, shopping baskets, trips, consumers, sending links, phrases, catalogue browsing, and social data [12]), which can be generally categorized into four categories:

(a) Click-stream data;
(b) Transaction or business activity data
(c) Audio data;
(d) Visual data. In e-commerce, data—which are gathered over time utilising customer browsing and purchase points—are essential for tracking consumer shopping behaviour and personalizing offerings [13]. The various forms of big data are covered in this section, along with how they affect e-commerce.
Given their imperative to remain competitive, e-commerce companies are among the quickest to adopt BDA. Online retailers typically handle both organised and unorganised data. The main focus of structured data is demographic information such as name, age, gender, date of birth, address, and preferences; unstructured data, on the other hand, is information such as clicks, likes, links, tweets, voices, and so on. In the context of business-to-business analytics, the task at hand involves managing both forms of data to produce significant insights and boost conversion rates.

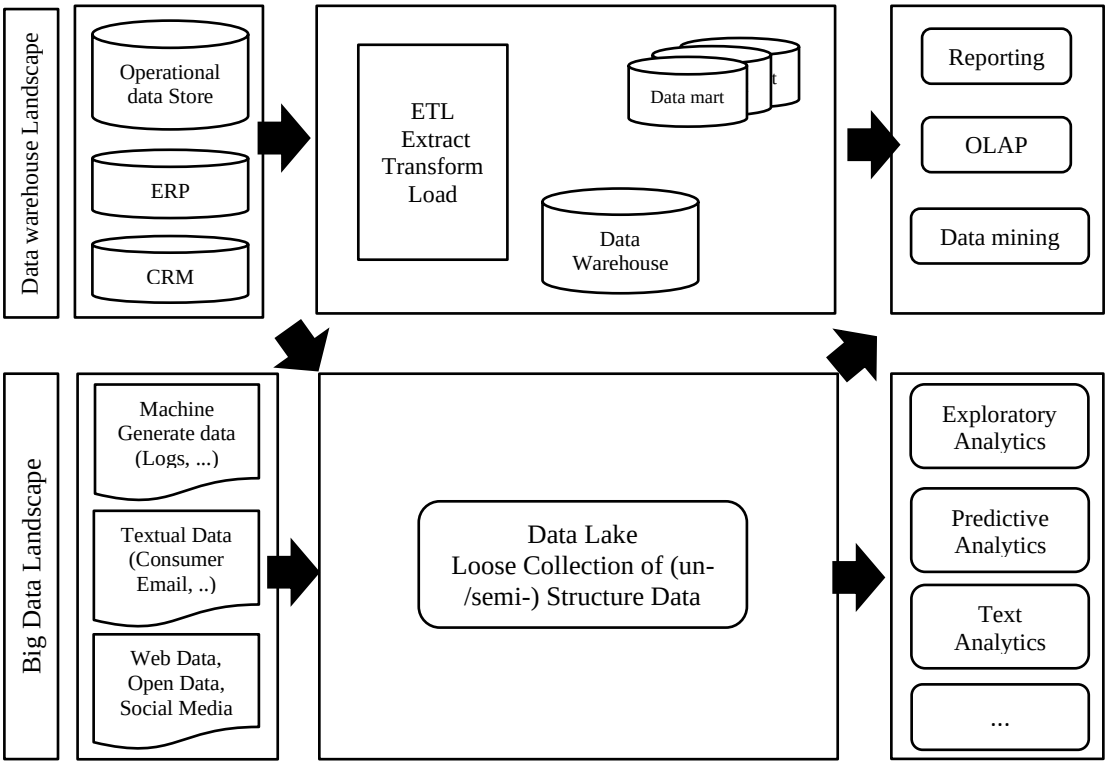

Figure 3.1: The expanding field of big data analytics

Governments, society, and shareholders have put pressure on the e-commerce industry to increase its sustainability performance over the past ten years in Figure 3.1. Customers have found that there are several benefits to using e-commerce, including practicality, convenience, utility, and habit, as well as advantages over environmental concerns and traffic or weather inconveniences. Due to this pressure, environmental, social, and economic issues are being included in the e-commerce operations. As a result, in recent years, the sustainability concern related to e-commerce and consumer behaviour has gained relevance.

Sustainability depends on three pillars [14]: the economy, society, and the environment. In contrast, e-commerce refers to business-to-business, consumer-to-consumer, and other types of electronic commercial transactions conducted through the Internet.

A complete awareness of customer behaviour as well as sustainability in e-commerce requires a comprehensive examination of findings from studies. There is a growing interest in this subject. Collaboration consumption, for example, is a growing notion that revolves around sharing, lending, borrowing, renting, and gifting. The rise of modern technologies as e-commerce has led to a rise in this kind of consumption. It enhances resource allocation and increases sustainable resource consumption as a result. This particular topic is relevant to the sharing economy. Researchers have an interest in the sharing economy because it enables people to exchange underutilised material possessions with each other. Airbnb is a prime example of the sharing economy. By analysing the importance of customer drives for and against peer-to-peer sharing it developed a fascinating investigation on this topic and reached the conclusion that ecological sustainability was one of the most significant variables influencing the use of peer-to-peer sharing platforms.

Furthermore, from the circular economy and waste-free perspective, other academics like Arman and Mark-Herbert noted collaborative consumption, in which the consumer participates in consumer-to-consumer transactions as both the buyer and the supplier. A circular economy focusses on cutting waste and carbon emissions while extending the useful life and life of products. The authors conducted comprehensive interviews with people who had experience trading used products on Facebook in order to gain insight into the consumer perspective from

the perspective of pro-environmental self-identity. The findings showed that buying, using, and selling used goods while participating in the circular economy are examples of how consumers' pro-environmental self-identity is manifested in their daily behaviour.

Another new area of study in e-commerce that has to do with sustainability and customer behaviour is city logistics. City logistics can improve economic growth and efficiency, lower transaction costs, and have a favourable environmental impact. Villa and Monzón created a related study. The authors believed that putting in place a delivery service that uses metro stations to pick up packages in lockers will lower air pollution, noise pollution, and greenhouse gas emissions. Therefore, the event that more information regarding the influence on social and environmental sustainability became available, e-commerce buyers would alter their preference for last-mile delivery. Their findings imply that showing the environmental and social effect information of last-mile delivery affects e-commerce customers' preferences and causes them to select a more sustainable choice. Furthermore, in a study designed to comprehend the influence of e-grocery by taking into account consumers' shopping visits and freight movements from the distribution centres to consumers came to the conclusion that home deliveries would be more environmentally friendly due to their increased efficiency.

Numerous writers have taken several different manners to determine how customer habits and the environment in online shopping are interconnected, which ended in the presentation of many research papers. The objective of this bibliometric analysis is to discover new understandings into consumer habits and sustainability in e-commerce by classifying and synthesising the body available existing literature.

3.3 Categories of Big Data

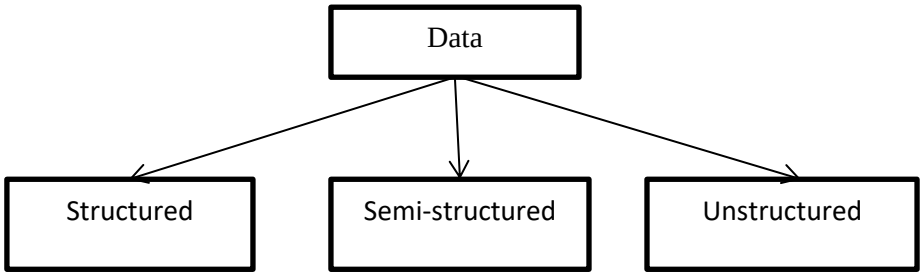

Figure 3.2: Types of big data

1. Organized Data

- The data contained in a specified field within a record is a rough definition of structured data in Figure 3.2.
- We are most accustomed to this type of data in our daily lives. For example: address, birthday
- All of the data has the same set of attributes since it is bound by a specific schema. Relational data is another name for structured data. By generating a single record to represent an entity, it is divided across several tables to improve the data's integrity. Table constraints are applied to enforce relationships.
- The ability of an organization to use its current systems and procedures for analysis is what gives structured data its business value.

2. Semi-Organized Information

- There are no strict guidelines for handling and storing semi-structured data. In contrast to a spreadsheet, the data is not in relational format and is not arranged neatly into rows and columns. Nonetheless, certain characteristics, like as key-value pairs, aid in distinguishing one entity from another.
- Semi-structured data is often referred to as NoSQL data since it does not require a structured query language.
- Systems that may potentially have different underlying infrastructures can share semi-structured data using a data serialisation language.
- Semi-structured information can include files with computer program machine instructions, but it is also frequently used to hold metadata about business processes.

- Usually, external sources like social networking sites or other web-based data feeds provide this kind of information.

3. Unstructured Information

- Data that doesn't fit into any specific format or set of guidelines is referred to as unstructured data. It is arranged haphazardly and without planning.
- In general, log files, text documents, videos, and photographs can all be categorised as unstructured data in Figure 3.3. The data being handled is unstructured, even though the information that goes with a picture or a video may be semi-structured.
- Furthermore, because unstructured data cannot be analysed without the right software tools, it is often referred to as "dark data."

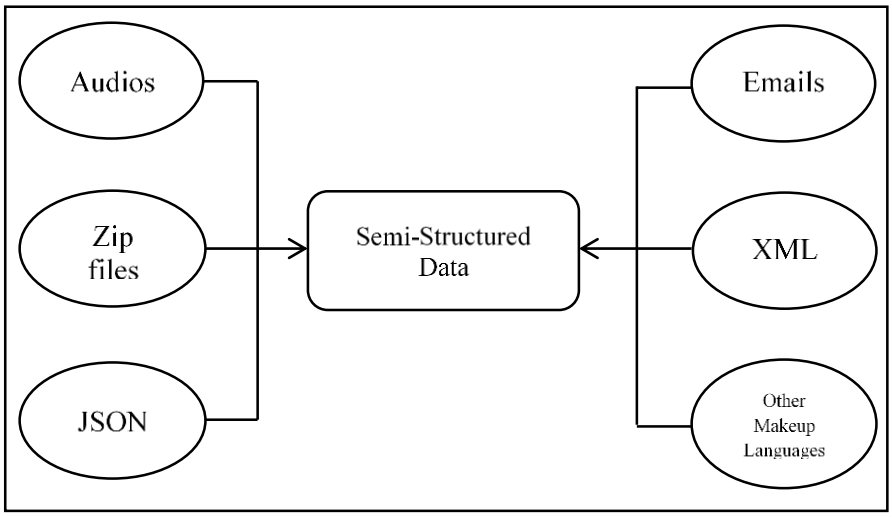

Figure 3.3: An Example of Unstructured Information

IMPLEMENTATION AND EXPERIMENTAL RESULTS

We assessed our strategy in two stages because the suggested abstraction method is complicated. Our initial approach was centred on using Latent Dirichlet Allocation to abstract items and events. A domain expert supplied explicit categories, and the latent categories derived via events abstraction were compared (indirectly) using the pattern recognition test. The patterns recognition measures of confidence and support were used to compare these two methods.

A behaviour abstraction was the main emphasis of the second phase. With the customer gender prediction job, we assessed our three parallel pattern mining techniques (sequence patterns, n-grams, and association rules).

4.1 Pre-processing and the dataset

We contrasted the sequence patterns based on the expert and latent categories, just like in association rules mining. Stated differently, the abstraction led to a greater number of generally accepted rules. Using the top 10 pertinent patterns, we also compared expert categories and latent categories in the second stage (Table 1) [15]. Sequence patterns with higher support are produced by latent categories derived from our item abstraction, as predicted; these patterns are "stronger" than those derived from expert categories. Furthermore, because the latent representation obscures the specifics of the product representation and hierarchy, these patterns are more resilient. Expert categories, on the other hand, produce more intricate patterns.

Table 1: A comparison of the best appropriate sequence patterns' support based on expert or latent categories

Rule position	**Expert**	**Latent**
1	0.028	0.096

2	0.828	0.067
3	0.021	0.036
4	0.021	0.035
5	0.021	0.035
6	0.019	0.035
7	0.019	0.029
8	0.019	0.024
9	0.018	0.019
10	0.017	0.019

We investigated the suggested method using a supervised classification problem as well, namely a gender prediction (binary classification) task, in addition to the unsupervised machine learning work. We picked a subset of users based on their demographics from the available data; yielding approximately 42k samples (15k women and 27k men). To get classes of comparable sizes, we then balanced the sample (by lowering the major class) an 80:20 ratio was used to divide the dataset into the train and test sets.

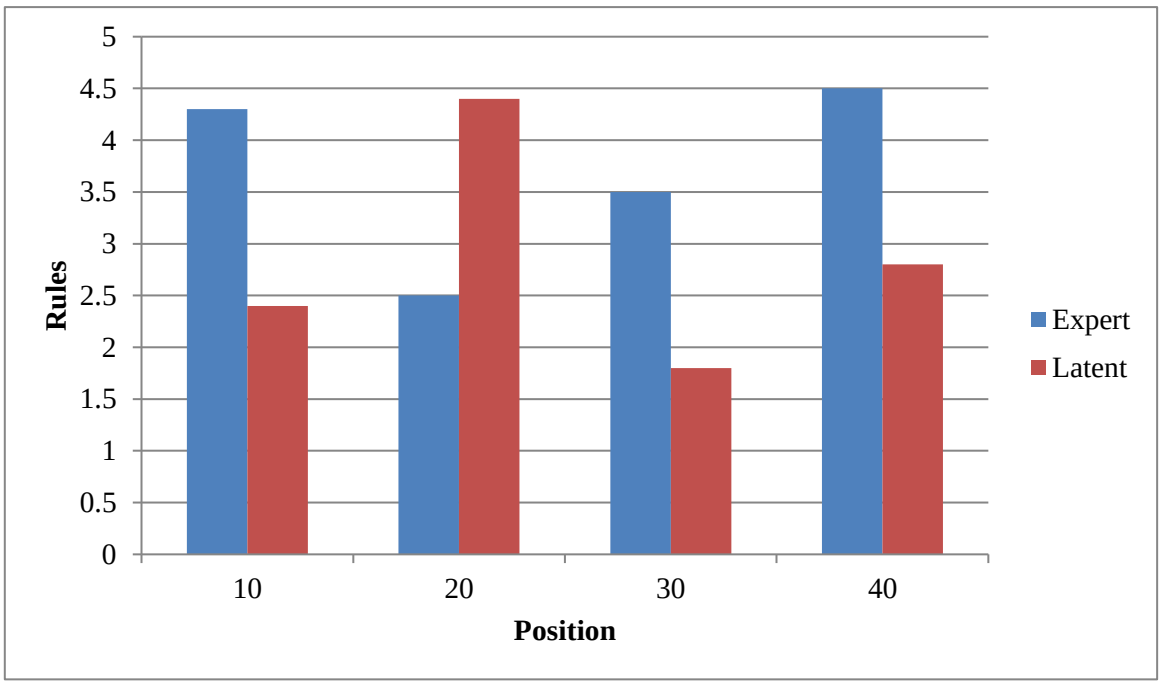

Figure 4.1: Symbolised by the sequence pattern mining and association rules

In order to investigate every combination (association rules, sequence pattern, and N-grams), we created seven datasets with features produced by each method (or a combination of them). The feature selection process was then carried out using the information gain metric. According to the experimental findings (see Figure 4.1 and 4.2), N-grams and sequence patterns performed comparably in the demographic prediction task; both strategies' recall rates were 67% (Table 2).

Table 2: The binary classifier's findings

Dataset	**Generated Features**	**Selected features**	**precision**	**Recall**	**FI**
AR	3122	246	0.65	0.61	0.62
SP	3553	169	0.73	0.69	0.80
NG	2373	142	0.73	0.69	0.70
AR + SP	5412	129	0.73	0.71	0.71

SP + NG	5922	127	0.75	0.72	0.71
AR + NG	4232	99	0.72	0.70	0.71
AR + NG + SP	7781	166	0.72	0.68	0.69

The behaviour abstraction strategy that combined sequence patterns and N-grams produced the greatest results (recall = 70%). There is an improvement in the behaviour abstraction case when comparing the gender prediction results based on behaviour with item and events abstraction (both metrics – precision and recall). This is actually a favourable outcome that shows how beneficial the behaviour abstraction is [16].

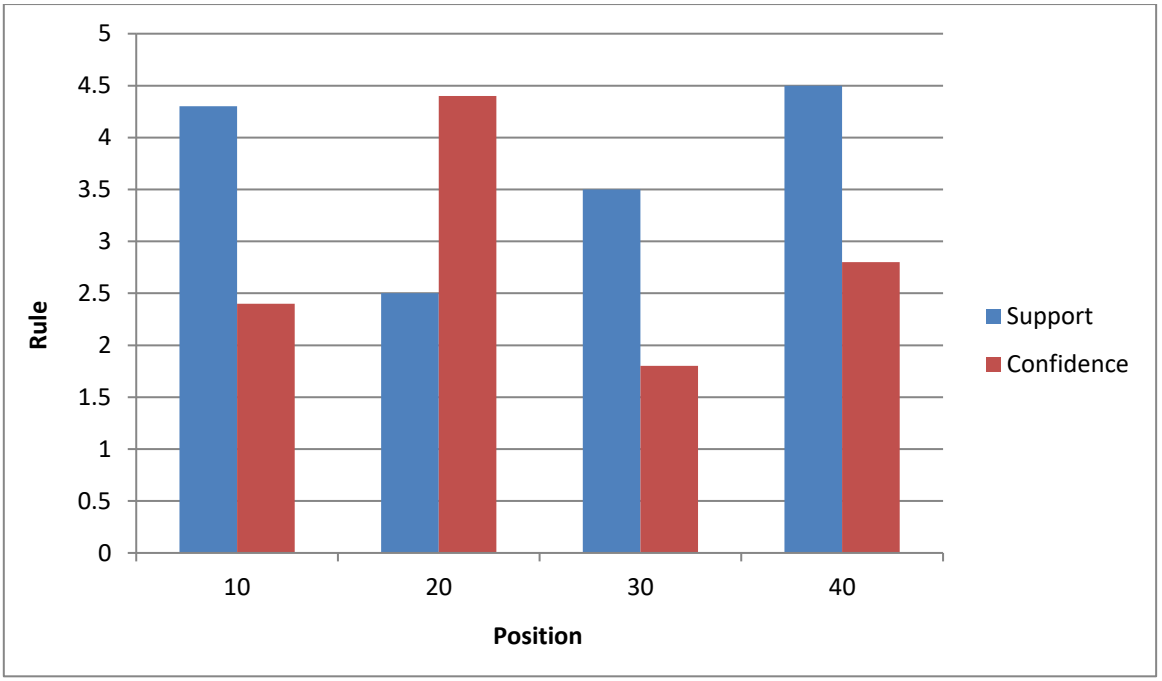

Figure 4.2: Comparing to the most significant characteristic produced by expert categories

The sequence pattern characteristics and N-grams features are close, as predicted; six out of ten identified patterns are the same. The primary distinction is how significant the best aspects are. Overall, N-gram features received higher relevance ratings. The association rules produced patterns that were less significant but more complexes. The most sophisticated features were frequently produced by combining different pattern mining techniques than by using only one. When sequence patterns and N-grams are combined, more complex top patterns are produced than when N-grams and sequence pattern mining are done independently. Conversely, in the context of association rules mining, the complexity of top features was reduced.

CONCLUSION

In this research, we examined how BDA affects vendors' and customers' online shopping experiences. The application of BDA capabilities in e-commerce initiatives enhances online purchasing and boosts vendor profits, according to the authors' conclusion. Additionally, by offering customised services and goods, businesses can draw in clients by knowing the demands and behaviours of their clientele. Both vendors and buyers had a better online shopping experience thanks to BDA. But even with this tremendous expansion in data, big data analysis remains difficult. This study served as a useful springboard for investigating BDA's potential uses in the e-commerce sector. The authors invite scholars to investigate the difficulties BDA faces in e-commerce, including data security and accuracy, and to help provide instruments and solutions to address existing issues.

The evaluation and suggested approach have several drawbacks as well. Examining how the behaviour abstraction affects different demography prediction problems would be fascinating. Moreover, the suggested abstractions can be useful for other jobs, such as churn rate or customised recommendations. As we mentioned, one advantage of the suggested strategy is model resilience. Extensive studies conducted across multiple time frames are required to investigate this topic. The investigation of the necessary amount of data is an additional intriguing avenue. For

example, the number of users who withdraw their data or provide false information will have an impact on any application task (like gender prediction).

This can be applied to the interpretation of the results in domains where users are unwilling to allow the storage of their actions (because of privacy concerns). These tests can mimic how the model performs when it comes to different kinds of consumer privacy concerns. We think that the suggested abstraction is useful for marketing or business purposes as well. These days, the field standard is the user segmentation. We found that using the abstraction helped with unsupervised machine learning tasks (represented by user segmentation), therefore we anticipate similar good outcomes.

REFERENCES

[1] Khade, A. A. (2016). Performing customer behavior analysis using big data analytics. Procedia computer science, 79, 986-992.
[2] Alrumiah, S. S., & Hadwan, M. (2021). Implementing big data analytics in e-commerce: Vendor and customer view. Ieee Access, 9, 37281-37286.
[3] Li, Q., Xing, J., Liu, O., & Chong, W. (2017). The Impact of Big Data Analytics on Customers‟ Online Behaviour. In Proceedings of the International MultiConference of Engineers and Computer Scientists (Vol. 2, pp. 15-17).
[4] 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.
[5] Pan, C. L., Liu, Y., & Pan, Y. C. (2022). Research on the status of e-commerce development based on big data and Internet technology. International Journal of Electronic Commerce Studies, 13(2), 027-048.
[6] Alsmadi, A. A., Shuhaiber, A., Al-Okaily, M., Al-Gasaymeh, A., & Alrawashdeh, N. (2023). Big data analytics and innovation in e-commerce: current insights and future directions. Journal of Financial Services Marketing, 1-18.
[7] Fedushko, S., & Ustyianovych, T. (2022). E-commerce customers behavior research using cohort analysis: A case study of COVID-19. Journal of Open Innovation: Technology, Market, and Complexity, 8(1), 12.
[8] Kauffman, R. J., Srivastava, J., & Vayghan, J. (2012). Business and data analytics: New innovations in the management of e-commerce. Electronic Commerce Research and Applications, 11(2), 85.
[9] Abdul Hussien, F. T., Rahma, A. M. S., & Abdulwahab, H. B. (2021). An e-commerce recommendation system based on dynamic analysis of customer behavior. Sustainability, 13(19), 10786.
[10] Behl, A., Dutta, P., Lessmann, S., Dwivedi, Y. K., & Kar, S. (2019). A conceptual framework for the adoption of big data analytics by e-commerce startups: a case-based approach. Information systems and e-business management, 17, 285-318.
[11] Mohammed, Z. N., & Kadhem, S. M. (2021). A study about e-commerce based on customer behaviors. Engineering and Technology Journal, 39(7), 1060-1068.
[12] Urbancokova, V., Kompan, M., Trebulova, Z., & Bielikova, M. (2020). Behavior-based customer demography prediction in E-commerce. Journal of Electronic Commerce Research, 21(2), 96-112.
[13] Rajesh, G., Preethi, S. P., Priya, R. S., Rajesh, L., & Raajini, X. M. (2021). Customer Behavior Prediction for E-Commerce Sites Using Machine Learning Techniques: An Investigation. Reinventing Manufacturing and Business Processes Through Artificial Intelligence, 79-93.
[14] Rajesh, G., Preethi, S. P., Priya, R. S., Rajesh, L., & Raajini, X. M. (2021). Customer Behavior Prediction for E-Commerce Sites Using Machine Learning Techniques: An Investigation. Reinventing Manufacturing and Business Processes Through Artificial Intelligence, 79-93.
[15] Reddy, V. M., & Nalla, L. N. (2022). Enhancing Search Functionality in E-commerce with Elasticsearch and Big Data. International Journal of Advanced Engineering Technologies and Innovations, 1(2), 37-53.
[16] Mou, J., Cui, Y., & Kurcz, K. (2019). Bibliometric and visualized analysis of research on major e-commerce journals using Citespace. Journal of electronic commerce research, 20(4), 219-237.