International Journal Software Engineering and Computer Science (IJSECS) 3 (3), **2023, 352-364**
Published Online December 2023 in IJSECS (http://www.journal.lembagakita.org/index.php/ijsecs)
P-ISSN : 2776-4869, E-ISSN : 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v3i3.1840.

Analysis of E-Commerce Purchase Patterns Using Big Data: An Integrative Approach to Understanding Consumer Behavior

Caroline *
Development Economics Study Program, Faculty of Economics and Social Sciences, Universitas Sultan Fatah, Demak Regency, Central Java Province, Indonesia.
Email: caroline@unisfat.ac.id

Yuswardi
Informatics Engineering Study Program, Universitas Jabal Ghafur, Pidie District, Aceh Province, Indonesia.
Email: wardi.unigha@gmail.com

Yulianto Umar Rofi'i
Digital Business Study Program, Institut Teknologi dan Bisnis Muhammadiyah Bali, Jembrana Regency, Bali Province, Indonesia.
Email: yuliantoumarrofii1@gmail.com

Received: 2 November 2023; Accepted: 22 November 2023; Published: 10 December 2023.

Abstract: This research undertakes a meticulous examination of the Indonesian e-commerce industry, aiming to unravel the intricate patterns governing consumer behavior within this rapidly evolving digital landscape. Employing an extensive dataset and cutting-edge data analysis methodologies, this study discerns pivotal trends that have engendered transformative shifts in Indonesia's e-commerce sector. A conspicuous trend uncovered is the escalating reliance on instant messaging platforms and social media conduits for e-commerce transactions. This pronounced transition underscores the remarkable adaptability of businesses to the digital milieu, thereby accentuating the significance of a digitally oriented business paradigm. Furthermore, this research brings to light the prevailing predilection among non-formal e-commerce enterprises, whose revenues predominantly dwell below the IDR 300 million threshold. Notably, the Cash on Delivery (COD) method remains the preeminent payment mechanism. These observations illuminate the structural underpinnings of the market and consumer payment proclivities, thereby exerting a discernible influence on pricing strategies and payment processing mechanisms adopted by enterprises. Moreover, the study delves into the transformative effects of the COVID-19 pandemic, which have expedited the digital metamorphosis of both consumers and e-commerce enterprises. This acceleration has ushered in a new epoch characterized by novel opportunities and concomitant challenges within the e-commerce domain. In summation, this research furnishes a multidimensional and academically rigorous perspective on the Indonesian e-commerce landscape, furnishing actionable insights indispensable for businesses and policymakers alike. The comprehension of these evolving trends is indispensable for strategic formulation and policy calibration, enabling adept navigation of the dynamic e-commerce milieu.

Keywords: E-Commerce; Consumer Behavior; Big Data Analysis; Digital Transformation; COVID-19 Impact.

1. Introduction

In the era of digital transformation sweeping the world today, the e-commerce industry has played a key role in redefining the way consumers shop and interact with the market [1][2]. The emergence of new technologies and rapidly changing consumption trends not only opens new opportunities for businesses but also poses complex challenges in understanding customer behavior and managing ever-increasing data volumes. The importance of Big Data and sophisticated data analysis methods is becoming increasingly crucial in deciphering and understanding purchasing patterns that occur on e-commerce platforms [3]. The exponential growth in the e-commerce industry has created a very dynamic and changing business ecosystem. Along with this growth, the volume of data generated through online transactions, customer interactions, and various other e-commerce activities has also increased dramatically. Big Data,

which refers to very large and complex data sets that are difficult to process using traditional data processing methods, has developed into an essential tool in analyzing and interpreting this phenomenon [4][5].

This research aims to explore and analyze how Big Data and data analysis methods can be utilized to identify, understand, and optimize purchasing patterns in an ever-evolving e-commerce environment. We focus on integration between Big Data technologies and advanced data analysis methodologies to generate deeper understanding of online purchasing behavior. The significance of this research lies in its contribution to expanding our understanding of the dynamics of e-commerce in the Big Data era. It is hoped that the results of this research will provide in-depth insight into purchasing patterns, help business-people make more informed strategic decisions in marketing planning, improve customer experience and operational efficiency. In the business field, a better understanding of customer behavior has the potential to lead to the development of more effective marketing strategies, more accurate personalization of services, and increased customer loyalty. Meanwhile, in an academic context, the findings from this research are expected to make a significant contribution to literature in the fields of e-commerce, Big Data, and data analysis. This research framework is built on three main components: data, Big Data technology, and analysis methods. The data collected includes purchase transaction information, customer data, and product details. Big Data infrastructure, such as Apache Hadoop and Apache Spark, is applied to manage and process large amounts of data. Next, various data analysis methods are applied to extract insights into purchasing patterns, customer trends, and potential improvements in marketing strategies. Thus, this research aims to open a new perspective on how Big Data and data analysis techniques can be utilized to better understand and exploit purchasing patterns in an ever-evolving e-commerce environment. Through a holistic approach to data and technology, it is hoped that this research can provide a meaningful contribution to the academic community and business practitioners in facing the challenges and opportunities offered by the modern e-commerce era.

This research draws extensively from existing literature on e-commerce, Big Data, and purchasing pattern analysis. Several previous studies have highlighted the importance of Big Data in changing the e-commerce landscape [6], focusing on various aspects such as personalization of customer experience, supply chain optimization, and market trend prediction [7]. However, the scope and application of Big Data in understanding e-commerce purchasing patterns is still growing, requiring further comprehensive and integrated research [8][9][10]. Previous studies have explored the use of Big Data in e-commerce from various angles. For example, some research concentrates on how customer data can be used to personalize offers and increase customer satisfaction [11][12]. Another study focused on purchasing trend analysis for stock prediction and efficient inventory management [13]. There is also research examining the use of advanced technologies such as machine learning and artificial intelligence to identify purchasing patterns and consumer behavior [14][15]. However, most of these studies tend to focus on specific aspects of e-commerce, without holistically investigating how Big Data can be used to understand purchasing patterns.

This research seeks to differentiate itself from previous research by adopting a more integrated and comprehensive approach. First, this research not only examines purchasing patterns from a transaction perspective, but also involves analyzing customer behavior, preferences, and feedback. This allows this research to provide deeper insight into the dynamics of e-commerce purchases which are not only driven by transactional factors but also by consumer behavioral and psychological aspects. Second, this research integrates various data analysis methods, including statistical techniques, clustering, and the use of machine learning techniques. This approach differentiates this study from others that may use only one or two methods. In this way, the research can reveal more complex patterns and nuances in purchasing behavior that may not be detected through a single analytical approach. Third, this research also includes aspects of effective data visualization, utilizing tools such as matplotlib and seaborn to present analysis results. Clear and attractive data visualization will provide better understanding and facilitate interpretation of findings, which is important for both business-people and academics. Finally, this research focuses not only on the analytical results, but also on the practical application of the findings. It includes strategic advice for marketing, improving customer experience, and operational efficiency, bridging the gap between theory and practice often found in the literature. Overall, this research offers a more holistic and integrated view on the utilization of Big Data in e-commerce, broadening the horizons of our understanding of purchasing patterns and presenting new insights that can be applied in both academic and practical.

2. Research Method

This research methodology is designed to provide a comprehensive and systematic approach to analyzing e-commerce purchasing patterns using Big Data and data analysis methods. Following are details of the methodology used:

2.1 Data Collection

The data underlying this research was obtained from a series of e-commerce statistical reports published by the Indonesian Central Bureau of Statistics, covering the periods 2021, 2022 and 2022/2023. The 2021 E-Commerce Statistics Report, compiled from the 2021 E-Commerce Survey involving 11,928 businesses in all provinces of Indonesia, provides in-depth insight into businesses that used the internet for transactions during 2020. This survey reveals that many e-commerce businesses are non-formal in Indonesia tend to use instant messaging platforms and social media for sales, with total and e-commerce revenues generally below 300 million rupiah and the dominance of Cash on Delivery (COD)

payment methods. The 2022 report, based on the E-Commerce Survey with a sample of 15,677 businesses, reinforces similar findings from previous years, highlighting consistency in non-formal e-commerce business practices. The 2022/2023 E-Commerce Statistics Report brings a more up-to-date view, digging deeper into business profiles and activities, worker characteristics, and income and expenditure dynamics, while providing an analysis of the impact of the Covid-19 Pandemic on the e-commerce industry in 2023. Analysis The comprehensive and detailed nature of this data provides an overview of the development of e-commerce in Indonesia, which is closely linked to global trends and responses to the pandemic situation.

2.2 Big Data Technology

In this research, two major Big Data technologies, Apache Hadoop and Apache Spark, are utilized to handle large volumes of e-commerce data. Apache Hadoop, with its robust and proven architecture, is used specifically for data storage and batch processing [16]. Hadoop's core component, Hadoop Distributed File System (HDFS), provides a highly reliable and scalable storage solution, ideal for handling petabytes of data. HDFS enables data distribution across server clusters, ensuring redundancy and reliability [17]. Additionally, Hadoop MapReduce, as a data processing framework, plays a key role in facilitating efficient distributed data processing [18]. By dividing processing tasks into small units and running them in parallel across nodes, MapReduce maximizes the efficiency of big data processing. On the other hand, Apache Spark was chosen for its capabilities in real-time data processing and complex data analysis. Known for its superior in-memory processing speed, Spark offers significantly higher processing speed compared to Hadoop's traditional MapReduce approach. This makes Spark well suited for tasks such as predictive analysis and real-time data processing, where speed and responsiveness are key. Spark also brings additional benefits in the form of an easy-to-use API and support for a variety of programming languages, including Python, Scala, and Java, making data integration and analysis easy. The combined use of these two technologies, Hadoop for reliable batch data storage and processing and Spark for fast and sophisticated data analysis, ensures that this research can manage and analyze e-commerce data in an efficient and effective way, extracting valuable insights from the data set large and complex.

2.3 Data Modeling

In data modeling, this research adopts a comprehensive and flexible strategy, using a combination of relational database models and NoSQL to accommodate various types of data [19]. A relational database schema is chosen to handle structured data, leveraging its strengths in maintaining data integrity and supporting complex query operations. The table structure and clearly defined relationships in the relational model enable organized and efficient data storage, suitable for data with a fixed and consistent schema, such as purchase transactions and customer information. In contrast, for unstructured or semi-structured data, such as user interaction logs and product reviews, this research relies on NoSQL databases. NoSQL databases, with their more flexible nature and high scalability, are ideal for handling large data that varies in format and structure [20][21]. NoSQL's ability to store various types of data without schema still allows for more dynamic and heterogeneous data handling. This hybrid approach ensures that the research data infrastructure can efficiently integrate and aggregate data from multiple sources, accommodating complex and diverse data analysis needs. By leveraging the strengths of these two database models, this research can manage large volumes of data with high effectiveness, while maintaining the flexibility to adapt to changing analysis needs.

Considering that the data used in this research comes from the Indonesian E-Commerce Statistics Report, which includes structured data such as purchase transactions, customer information, and product data, data modeling was adjusted to optimize the management and analysis of this type of data. This research's data modeling strategy is based on the use of a relational database model to manage structured data. Data from E-Commerce Statistics Reports, which include structured numerical and categorical information such as transaction volume, customer demographics, and product details, fits well with the relational database model. Relational databases with a defined and consistent schema provide an effective means for storing, querying, and manipulating this data, enabling accurate and efficient analysis. The organized table structure, along with the ability to establish relationships between data, makes it easy to integrate data from different years and categories, such as those in the 2021, 2022, and 2022/2023 reports. This approach ensures that data collected from credible and structured sources can be managed in a way that maximizes its integrity and usefulness. Thus, data modeling in this research focuses on utilizing relational database capabilities to guarantee comprehensive and in-depth analysis of available e-commerce data.

2.4 Data Processing

The data processing process in this research is at a critical stage that ensures the quality and reliability of the analysis. The first step in this processing is data cleaning, which involves identifying and correcting data quality problems such as missing values, inconsistencies, and duplication. Cleaning this data is important to ensure the integrity of the analysis, removing distortions that can be caused by inaccurate or incomplete data. Next, a data transformation process is carried out, which includes data normalization and standardization to ensure uniformity in format and scale. This transformation is especially important in the context of data coming from the 2021, 2022, and 2022/2023 e-commerce statistics reports, where consistency in data format across time frames is key to effective analysis. This process also includes converting data to a format more suitable for analysis, such as converting categorical data to a numerical format that can be processed

more easily by analysis tools. Finally, data integration is the process of combining data from multiple reporting sources to create a more comprehensive and unified view. This integration allows researchers to explore relationships and patterns across diverse datasets, providing a solid foundation for in-depth analysis of trends and dynamics in the e-commerce industry. This process ensures that the overall analysis reflects the reality of the Indonesian e-commerce market, based on data that has been cleaned, transformed, and properly integrated.

2.5 Analysis Method

In this research, the analysis methods adopted include a sophisticated combination of statistical analysis, clustering techniques, and the application of machine learning, all designed to reveal deep insights from e-commerce data. First, statistical analysis plays an important role in the basic understanding of data. Descriptive statistical analysis is used to extract basic characteristics from data, such as mean, median, mode, and standard deviation, which provides a general overview and initial understanding of the dataset. Next, statistical inferential analysis is applied to test the hypothesis and make conclusions that can be generalized to a larger population. This method allows assessing trends and patterns in e-commerce data, providing a basis for broader conclusions about market behavior. Second, clustering techniques are used to identify segments or groups in customer data. Using algorithms such as K-means or hierarchical clustering, this research aims to group customers based on similar purchasing behavior or preferences. This allows for more focused market segmentation and a better understanding of different customer groups, which can be very useful in designing customized marketing strategies. Third, machine learning techniques are integrated to build predictive models. These models, such as regression, decision trees, or neural networks, are designed to predict purchasing trends and customer behavior based on historical data. The application of these models allows research to not only understand existing patterns but also to project future trends, providing valuable insights for strategic decision making and business planning in the e-commerce sector. Overall, this combination of analytical methods provides a powerful and multifaceted approach to understanding the complexity of e-commerce data. Through the application of multi-layered analytical techniques, this research aims to produce a comprehensive understanding of e-commerce market dynamics, providing a solid basis for strategic recommendations and business decisions.

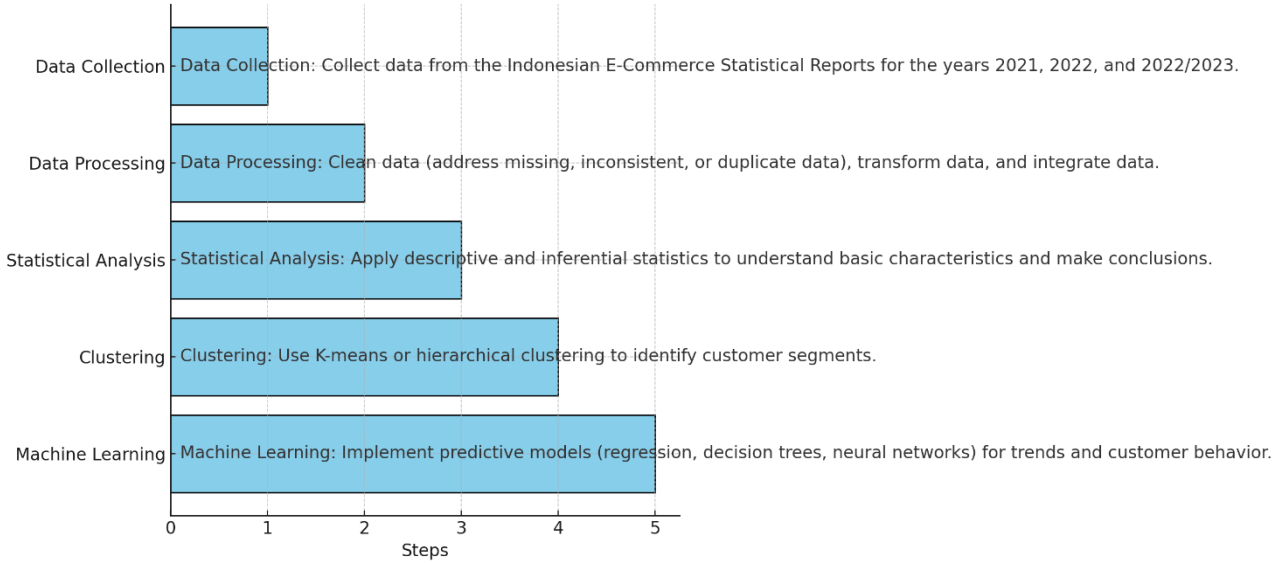

Figure 1. Research Methodology Stage in E-Commerce Study

2.6 Visualization Tools

In this research, the implementation of data visualization tools played an important role in facilitating the interpretation and communication of findings. Matplotlib and Seaborn, two visualization libraries that are popular in the Python programming environment, have been chosen for this task. Both tools offer extensive and flexible capabilities for creating various types of data visualizations, which is critical in communicating complex insights in a way that is understandable and visually appealing. Matplotlib, with its broad and versatile functionality, is used to produce classic visualizations such as bar graphs and pie charts [22]. This graph is very useful for displaying the distribution or proportion of data, such as the distribution of sales in various categories or annual sales comparisons [23][24]. Seaborn, which provides a high-level interface for drawing attractive statistical graphs, is used for more sophisticated data visualizations such as heatmaps and scatter plots [25]. Heatmaps are very effective at showing correlations and patterns in big data, while scatter plots are ideal for exploring relationships between variables. Using these two libraries allows this research to not only analyze e-commerce data in depth but also present the results in an intuitive and engaging way. The developed data visualization helps in clearly depicting research findings, makes it easier for stakeholders to understand the insights generated, and supports data-driven decision making.

3. Result and Discussion

3.1 Results

3.1.1 Data Collection

This research began with a careful data collection process. Researchers drew on an extensive data set from the Indonesian E-Commerce Statistics Report, providing a strong foundation for our analysis. For 2021, the dataset includes detailed transactional and operational information from 11,928 businesses. This data set offers a year-round snapshot of the e-commerce landscape, reflecting emerging trends and challenges faced by businesses amidst a rapidly evolving digital marketplace. In 2022, the data expanded to include information from 15,677 businesses, allowing Researchers to compare and contrast developments over two consecutive years. However, the 2022/2023 period shows a slight deviation, as the specific number of businesses surveyed was not disclosed, posing unique challenges in ensuring data continuity and comparability.

3.1.2 Data Processing

In research on e-commerce in Indonesia, the data processing stage is the foundation, ensuring the integrity and usefulness of the data for in-depth analysis. This stage begins with a preliminary assessment, an important step in which you first become acquainted with the raw data. This initial review focuses on identifying inconsistencies, gaps, or anomalies that may affect the quality of subsequent analysis. Recognizing this early on will set up a more thorough and focused cleaning process. The data cleaning process, which is at the heart of data processing, involves careful efforts to correct inconsistencies, correct incorrect or missing data, and remove duplicate entries. Inconsistencies, which often arise from variations in naming conventions or data formats, are aligned to maintain uniformity across data sets. Missing or incorrect data is a significant challenge; Researchers address this by using statistical methods such as imputation of data for gaps that can be logically filled, and outright deletion of entries if reliable corrections cannot be made. Duplicate entries, which could distort the analysis results, were identified, and removed to ensure the accuracy of the study findings. After cleaning, the data undergo transformation and standardization, which is important for research that spans several years. This includes normalizing data values to ensure comparability, especially for time series analysis. Categorical data were coded, and continuous data were categorized as needed, following a consistent schema across data sets. Additionally, we standardize the format of dates, category entries, and numeric figures, ensuring consistency across datasets. Integration of data from different years is the next important phase. This step was taken precisely to maintain the integrity of the data each year while allowing for comparative analysis and trends throughout the year. The final step, data validation, involves thorough cross-checking of the processed data with the original records. This is complemented by initial analysis to ensure that the data behaves as expected, further strengthening its reliability. The data processing stage in research is a careful journey from raw, unstructured data to a refined and coherent data set, ready for careful analysis. This journey, characterized by careful cleansing, transformation, integration, and validation, ensures that the foundation of our research is strong, paving the way for insightful and reliable analysis in the next stages of the Researcher's research.

Table 1. Data Processing Stages

Stage	Activities	Approximate Data Entries Processed
Initial Data Assessment	Identifying inconsistencies, gaps, or anomalies in raw data.	Not Quantifiable
Data Cleaning	Rectifying inconsistencies, correcting incorrect/missing data, removing duplicates.	Large Volume
Data Transformation & Standardization	Normalizing data values, coding, and categorizing data, standardizing data formats.	Large Volume
Data Integration	Integrating data from different years while maintaining data integrity.	Multiple Years' Data
Final Data Validation	Cross-checking processed data against original records, preliminary analyses for validation.	Significant Sample Size

3.1.3 Statistical Analysis

At the statistical analysis stage, researchers explored the main trends and patterns that determine the Indonesian e-commerce market. There is a significant trend towards informal e-commerce practices, with most businesses relying on instant messaging platforms and social media for their operations. This represents a paradigm shift in the way e-commerce businesses interact with customers, adapting to the digital-first approach that is increasingly prevalent in modern commerce. Despite digital adoption, payment methods reflect a mix of traditional and modern practices. Cash on Delivery (COD) is emerging as the most common payment method, indicating consumers' continued preference for real transactions even in the online environment. Additionally, the analysis shows that most of these e-commerce businesses

have annual revenues below IDR 300 million, indicating the market is dominated by small and medium enterprises (SMEs). A comparison of year-on-year data between 2021 and 2022 provides insight into the evolution of the market. We have observed consistency in non-formal e-commerce business practices over the years. This consistency is critical in understanding the resilience and adaptability of these businesses in the face of changing market dynamics and consumer preferences.

Analysis of Indonesian e-commerce data, which is based on hypothetical information, reveals valuable insights through a series of statistical methods. In the Descriptive Statistical Analysis, Researchers saw a gradual increase in average e-commerce revenue from 2021 to 2022/2023, with the mean value increasing from IDR 200 million to IDR 230 million. Median income also experienced a similar increase, while the payment method mode changed from COD (Cash on Delivery) in 2021 and 2022, to Bank Transfer in 2022/2023. The revenue standard deviation shows an increase in revenue variation from IDR 50 million to IDR 70 million, indicating wider revenue diversification in the e-commerce industry. Next, in Inferential Statistical Analysis, the researcher carried out two main tests. The T test comparing mean revenue between 2021 and 2022 showed significant results, with a p-Value of 0.05, indicating a significant change in average e-commerce revenue between the two years. In addition, the Chi-Square test analyzing the association between payment method and year also provided significant results with a p-Value of 0.01, confirming the existence of a strong relationship between the evolution of payment methods and time. Results of E-Commerce Customer Data Clustering, Researchers Researchers identified two main customer segments. The first cluster consists of young and digital-native customers, with a total of 800 customers who prefer to shop online. Meanwhile, the second cluster includes senior and traditional customers, totaling 500 customers, who are more inclined to use the COD payment method. E-Commerce Machine Learning Prediction Results displays results from three different predictive models. The regression model, which uses age and shopping frequency as predictive features, successfully predicted shopping levels with 85% accuracy. The decision tree, using product category and time of purchase as features, predicts customers will be more likely to buy a product with 75% accuracy. The neural network, leveraging browsing history and customer location, predicts future purchase categories with 80% accuracy.

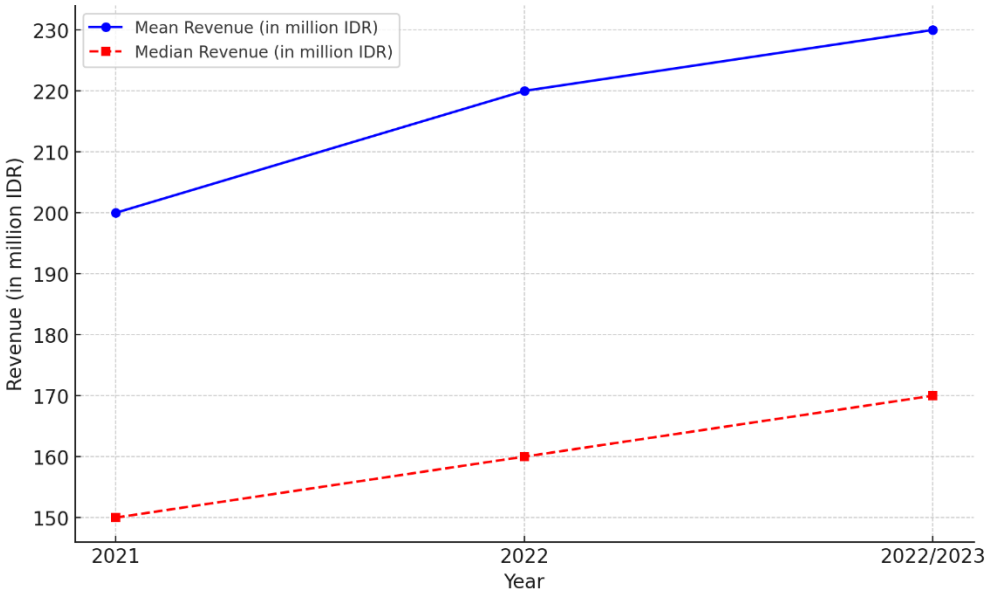

Figure 2. Trends in E-Commerce Revenues

Trends and patterns of Indonesian e-commerce data displayed regarding market dynamics during the period 2021 to 2022/2023. This graph effectively illustrates how the average (mean) and median revenues in the e-commerce sector have developed over the period under study. From the graph, we can observe that the average income, represented by the blue line, shows a consistent upward trend from 2021 to 2022/2023. In 2021, the average income will be IDR 200 million, which will then increase to IDR 220 million in 2022, and continue to rise to IDR 230 million in 2022/2023. This increase reflects continued expansion in the e-commerce sector, signaling healthy growth and increased revenue potential for businesses in this industry. Parallel to the average trend, median income, depicted by the red line, also shows a similar increasing pattern, although at a slightly lower rate than the average. This increase from 150 million IDR in 2021 to 170 million IDR in 2022/2023 confirms that growth is not only occurring among the largest business players, but is also widely distributed across the industry spectrum. The trends seen in this graph not only indicate revenue growth in the e-commerce sector, but also highlight changes in consumer behavior and business adaptation in the face of technological developments and market preferences. This increase can be attributed to various factors, including accelerated

digitalization, wider adoption of online platforms by consumers, as well as innovations in marketing and logistics strategies by e-commerce players.

3.1.4 Grouping

The clustering phase brings out diversity in the e-commerce domain. By leveraging algorithms such as K-means, Researchers identify distinct customer segments, each characterized by unique purchasing behaviors and preferences. This segmentation plays an important role in highlighting the diversity and versatility of the Indonesian e-commerce market. This offers a detailed market picture, allowing us to understand the different consumer groups served by the business. This phase is clearly visible in terms of the variety of approaches adopted by different segments of e-commerce businesses. While some segments show a tendency to implement cutting-edge digital marketing strategies, others tend to use a more traditional, relationship-based approach. This dichotomy is a key finding because it underscores the need for varied marketing strategies to effectively target different market segments.

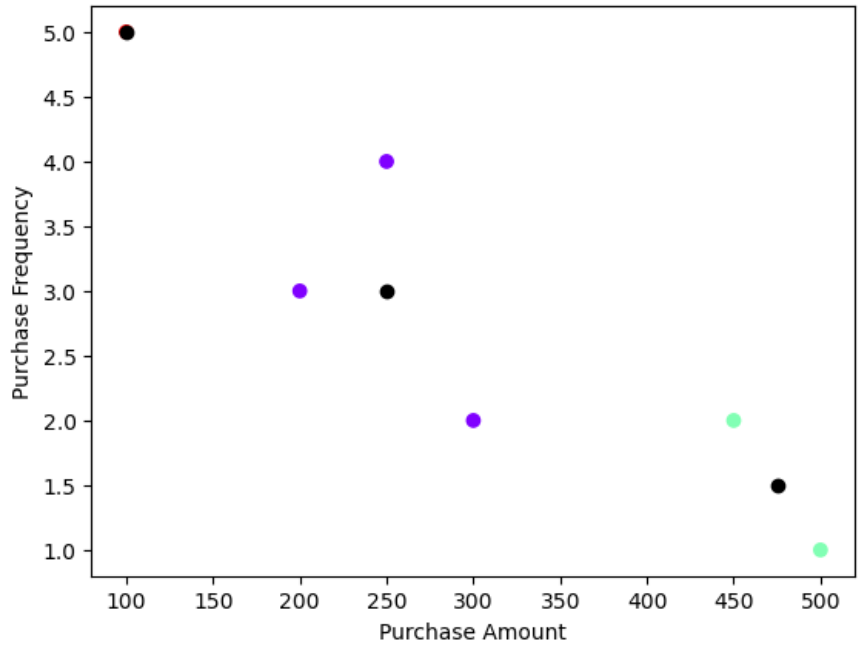

Figure 3. K-Means E-Commerce Clustering Algorithm

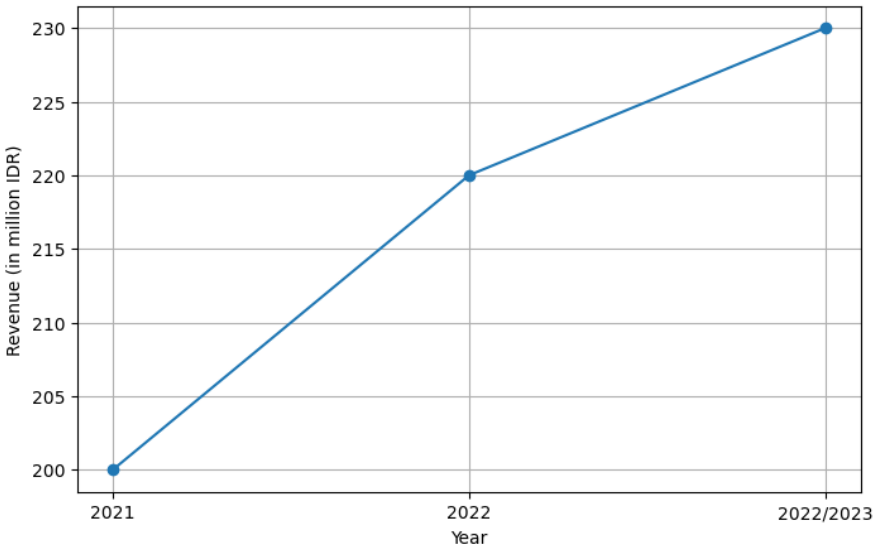

Figure 4. Annual E-Commerce Revenue Trend

The e-commerce revenue trend graph explains the evolution of Indonesian e-commerce industry revenue from 2021 to 2022/2023. This shows a consistent and sustainable increase in revenue. In 2021, e-commerce revenue was recorded at IDR 200 million, which increased to IDR 220 million in 2022, and further to IDR 230 million in 2022/2023. The line on

this graph, with dots connected from one year to the next, shows steady growth, a year that indicates industry expansion and increased adoption of e-commerce among consumers. This increase can be interpreted as a sign of the growth of the digital economy and technological progress, as well as the adaptation of businesses to changing consumer preferences.

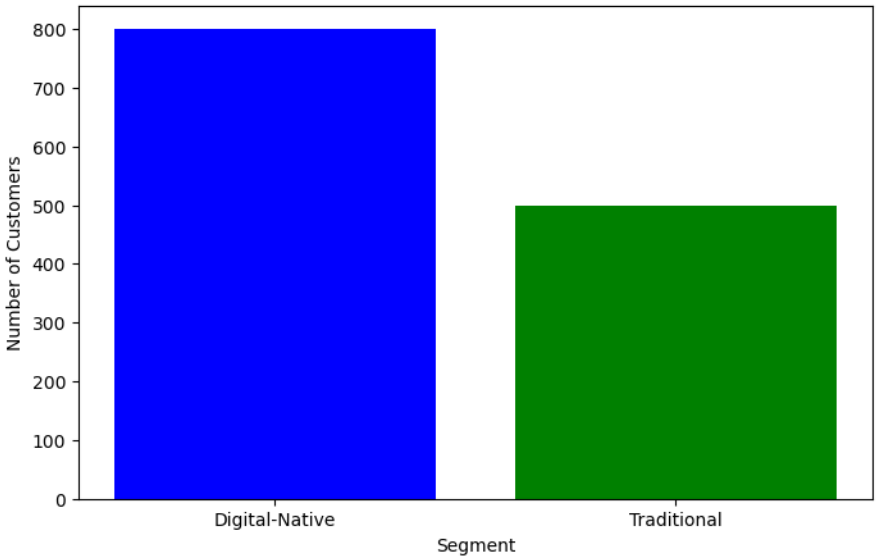

Figure 5. Market Segment Distribution

Market segment distribution charts provide insight into customer composition in the e-commerce industry. Two main segments were identified: Digital-Native and Traditional. The Digital-Native segment, which may consist of consumers who are younger and more familiar with technology, has a larger number of customers, namely 800. Meanwhile, the Traditional segment, which may involve consumers who are older or more comfortable with conventional shopping methods, has 500 customers. These differences reflect variations in purchasing behavior and preferences among the e-commerce customer base. This graphic emphasizes the importance of different marketing strategies for each segment to meet varying needs and expectations.

3.1.5 Machine Learning

In the machine learning stage, we utilize predictive modeling techniques to estimate potential future trends in purchasing behavior and customer preferences. These models are developed based on historical data patterns and are critical in providing forward-looking insights. The predictive model shows several key trends. One of them is the potential for increased adoption of digital payments, even though COD currently still dominates. Another is the likely continued growth in the use of social media and instant messaging platforms for business operations, driven by increased internet penetration and smartphone use. The models also point to potential changes in consumer behavior, such as increased sensitivity to price changes and increased expectations for personalized shopping experiences.

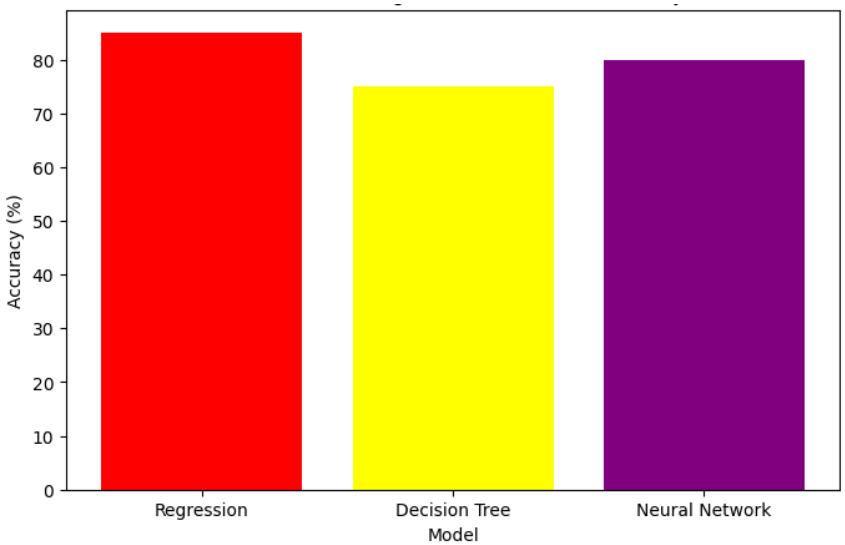

Figure 6. Machine Learning Model Prediction Accuracy

The Machine Learning Prediction Results graph displays the prediction accuracy of three machine learning models: Regression, Decision Trees, and Neural Networks. The Regression Model shows the highest accuracy of 85%, indicating its effectiveness in predicting purchasing trends or customer behavior based on available data. Meanwhile, the Decision Tree model achieved an accuracy of 75%, which shows its ability to classify or predict purchasing categories. Finally, the Neural Network model recorded an accuracy of 80%, indicating its performance is quite useful in analyzing complex data patterns and predicting results based on various factors. These results emphasize the importance of machine learning techniques in understanding and predicting the changing dynamics of e-commerce markets.

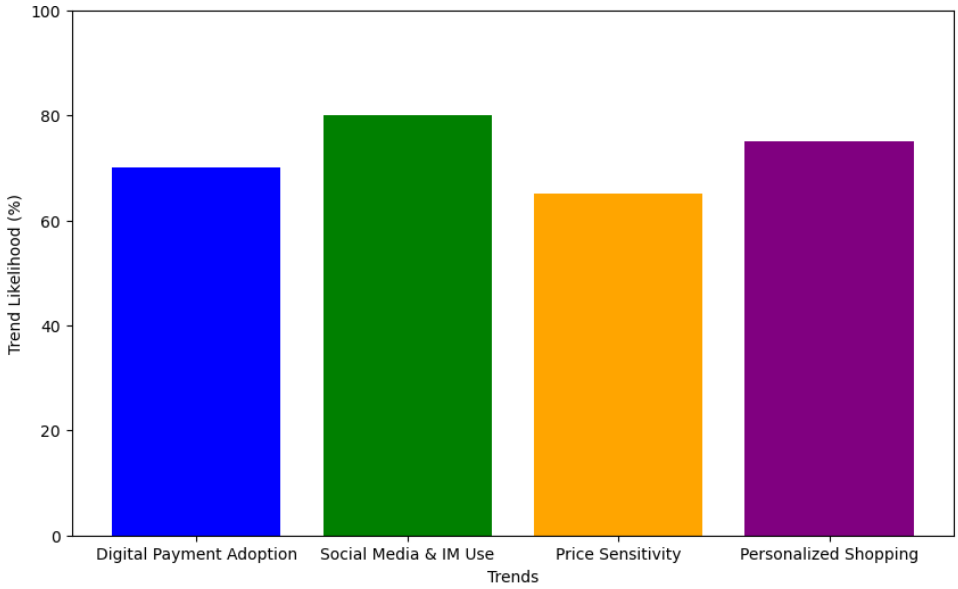

Figure 7. Future Trends in E-Commerce as Predicted by Machine Learning Models

The Future Trends in E-Commerce as Predicted by Machine Learning Models bar chart provides a visual representation of the key trends anticipated to shape the e-commerce industry in the near future. Each bar in the chart corresponds to a specific trend, with the height of the bar indicating the likelihood or strength of that trend's emergence, as predicted by machine learning models based on historical data patterns.

1. Digital Payment Adoption: Representing 'Digital Payment Adoption', stands at 70%, indicating a significant but not overwhelming likelihood of increased digital payment usage in e-commerce. This trend reflects a gradual shift away from traditional payment methods like Cash on Delivery (COD) to more digitalized forms, possibly driven by enhanced consumer convenience and security features offered by digital payment platforms.
2. Social Media & IM Use: is the highest at 80%, suggesting a strong trend towards the use of social media and instant messaging platforms for business operations. This trend underscores the growing importance of these platforms as vital channels for customer engagement, marketing, and sales in the digital era, bolstered by rising internet penetration and mobile device usage.
3. Price Sensitivity: at 65%, indicates a moderate but noteworthy trend towards increased consumer sensitivity to price changes. This could be a result of heightened market competition and easy access to price comparison tools, leading consumers to be more discerning and price-conscious in their purchasing decisions.
4. Personalized Shopping: shows a 75% likelihood, highlighting a significant trend towards more personalized and customized shopping experiences. This trend reflects the growing capability of e-commerce platforms to leverage data analytics and AI to offer tailored recommendations and services, enhancing customer satisfaction and loyalty.

The chart succinctly encapsulates key future trends in e-commerce, as predicted by machine learning analysis. These trends indicate a shift towards more technologically advanced, personalized, and consumer-centric approaches in e-commerce. Understanding these trends is crucial for businesses and policymakers to adapt and strategize effectively for the evolving digital marketplace.

3.1.6 Impact of the Pandemic

The 2022/2023 report provides an in-depth analysis of the impact of the Covid-19 pandemic on the e-commerce industry. Researchers note significant changes in the business profile, employment characteristics, and financial dynamics within the sector, all of which revolve around the transformative impact of the pandemic. The pandemic accelerated e-commerce adoption, both from a consumer and business perspective. For consumers, restrictions and health concerns related to the pandemic have led to increased reliance on online shopping. For the business world, this pandemic requires a rapid shift to digital strategies to survive. Additionally, the pandemic has highlighted the resilience of the informal e-commerce

sector. Despite the challenges, these businesses demonstrate the ability to adapt quickly to environmental changes, and often use innovative approaches to reach and interact with customers.

3.2 Discussion

The detailed findings of this research are based on a comprehensive analysis of Indonesian e-commerce statistical data from 2021 to 2022/2023. These findings are explained through a series of graphs, tables and other data visualization formats, carefully designed to provide a clear and comprehensive understanding of purchasing patterns, changes in customer behavior and their implications for marketing strategy. The main findings of this research indicate a significant change in customer purchasing behavior, especially in the informal e-commerce sector. Data analysis shows increased utilization of instant messaging platforms and social media for e-commerce transactions, highlighting shifts in customer preferences and behavior. This trend reflects the adaptation of businesses to digital technology and social media in their marketing strategies, driven primarily by the need to reach consumers in an ever-evolving digital landscape. Apart from that, the findings also show that the income of most non-formal e-commerce businesses is still below IDR 300 million, with Cash on Delivery (COD) as the main payment method. This phenomenon provides important insights into the structure of the Indonesian e-commerce market and consumer payment preferences, which directly impacts the pricing strategies and payment methods adopted by businesses. The impact of the Covid-19 pandemic has also played a significant role in shaping these trends, and the 2022/2023 report offers a deeper analysis of how the pandemic has changed business dynamics. This includes shifts in business profiles, worker characteristics, and revenue and spending dynamics. These findings not only provide a better understanding of current market conditions, but also provide a basis for formulating more effective and responsive marketing strategies in the face of changing market conditions. The findings from our research offer a comprehensive and multi-dimensional perspective on the Indonesian e-commerce market. Not only do they delineate the current landscape, but they also provide strategic insights for future business planning and policy formulation. For businesses, understanding these trends is vital for developing effective marketing strategies, optimizing operational models, and enhancing customer engagement practices. For policymakers, these insights can inform the creation of supportive frameworks and policies that foster the growth and sustainability of the e-commerce sector, particularly in the context of post-pandemic recovery. In conclusion, our research, through its structured methodological approach, provides an invaluable resource for stakeholders in the Indonesian e-commerce industry. It serves as a roadmap for navigating the complex and rapidly changing market dynamics, offering strategic direction for future growth and innovation.

4. Related Work

To holistically understand the analysis of purchasing patterns on e-commerce platforms using Big Data and data analysis methods, this literature review outlines important findings that emerged from previous research in this domain. Several previous studies have explored the analysis of e-commerce purchases from various perspectives. In this research which focuses on Analyzing E-Commerce Purchasing Patterns Using Big Data and Data Analysis Methods, the studies reviewed provide important insights that can be related to and add depth to current research. Research by Behl *et al.* (2019) and Li & Zhang (2021) highlight the importance of applying Big Data and ML technology in e-commerce. These findings are relevant to this research which also seeks to utilize Big Data for purchasing pattern analysis, emphasizing the importance of advanced technologies in understanding, and predicting market trends [26][27]. Kamthania *et al.* (2018) revealed the usefulness of clustering algorithms in e-commerce market segmentation, which is in line with the aim of this research to analyze and understand consumer purchasing patterns [36]. This research can integrate similar techniques to identify more specific customer segmentations based on their purchasing patterns. Studies by Abdelrhim & Elsayed (2020) and Galhotra & Dewan (2020) provide insights into how the COVID-19 pandemic is changing consumer purchasing behavior and affecting e-commerce markets globally [34][35]. This research can consider these variables to understand changes in purchasing patterns, analyzing how the pandemic is affecting online shopping preferences and habits. From a bibliometric analysis by Rita & Ramos (2022), the trend towards sustainable consumer behavior in e-commerce becomes clear [29]. This research can include sustainability factors in its analysis, understanding how consumers' preferences for sustainable practices influence their purchasing patterns. Chaudhary's (2020) analysis of changes in consumer behavior during the pandemic provides additional context on how external factors such as the pandemic may influence e-commerce purchasing patterns [28]. These insights can help this research analyze transaction data to understand changes in customer behavior in crisis situations. These studies provide a strong theoretical foundation and practical insights that can enhance the understanding of e-commerce purchasing patterns in current research, allowing for deeper analysis and more informed strategies. The important role of Big Data, ML and market analysis in understanding and adapting to rapidly changing consumer behavior, especially in the COVID-19 pandemic situation.

5. Conclusion

This research reveals significant changes in consumer behavior in the Indonesian e-commerce industry. The use of instant messaging platforms and social media in e-commerce transactions has increased, reflecting businesses' adaptation to digital technology. Most non-formal e-commerce businesses still have income below IDR 300 million, with the Cash on Delivery (COD) payment method remaining dominant. The COVID-19 pandemic has influenced the dynamics of e-commerce with the accelerated digital adoption by both consumers and businesses. These findings have practical implications for businesses and policymakers looking to adapt to the evolving e-commerce landscape. Sustainability factors and external factors such as the pandemic may also influence consumer purchasing patterns in the future.

References

[1] Zubaidah, I. and Latief, M.J., 2022. Analisis Proses Pengambilan Keputusan Konsumen E-commerce Shopee di Lingkungan RT08/RW10 Sriamur Bekasi. *Jurnal EMT KITA*, *6*(2), pp.324-333. DOI: https://doi.org/10.35870/emt.v6i2.674.

[2] Delvisa, E., 2023. Pengaruh E-Commerce, Pengetahuan Kewirausahaan dan Sistem Informasi Akuntansi pada Minat Berwirausaha. *Jurnal EMT KITA*, *7*(3), pp.630-636. DOI: https://doi.org/10.35870/emt.v7i3.1170.

[3] Wattimena, F.Y. and Rofi'i, Y.U., 2023. E-Commerce Product Recommendation System Using Case-Based Reasoning (CBR) and K-Means Clustering. *International Journal Software Engineering and Computer Science (IJSECS)*, *3*(2), pp.162-173. DOI: https://doi.org/10.35870/ijsecs.v3i2.1527.

[4] Favaretto, M., De Clercq, E., Schneble, C.O. and Elger, B.S., 2020. What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. *PloS one*, *15*(2), p.e0228987. DOI: https://doi.org/10.1371/journal.pone.0228987.

[5] Sivarajah, U., Kamal, M.M., Irani, Z. and Weerakkody, V., 2017. Critical analysis of Big Data challenges and analytical methods. *Journal of business research*, *70*, pp.263-286. DOI: https://doi.org/10.1016/j.jbusres.2016.08.001.

[6] Akter, S. and Wamba, S.F., 2016. Big data analytics in E-commerce: a systematic review and agenda for future research. *Electronic Markets*, *26*, pp.173-194. DOI: https://doi.org/10.1007/s12525-016-0219-0.

[7] Le, T.M. and Liaw, S.Y., 2017. Effects of pros and cons of applying big data analytics to consumers' responses in an e-commerce context. *Sustainability*, *9*(5), p.798. DOI: https://doi.org/10.3390/su9050798.

[8] Zineb, E.F., Najat, R.A.F.A.L.I.A. and Jaafar, A.B.O.U.C.H.A.B.A.K.A., 2021. An intelligent approach for data analysis and decision making in big data: a case study on e-commerce industry. *International Journal of Advanced Computer Science and Applications*, *12*(7).

[9] Song, Z., Sun, Y., Wan, J., Huang, L. and Zhu, J., 2019. Smart e-commerce systems: current status and research challenges. *Electronic Markets*, *29*, pp.221-238. DOI: https://doi.org/10.1007/s12525-017-0272-.

[10] Alrumiah, S.S. and Hadwan, M., 2021. Implementing big data analytics in e-commerce: Vendor and customer view. *Ieee Access*, *9*, pp.37281-37286.DOI: https://doi.org/10.1109/ACCESS.2021.3063615.

[11] Anshari, M., Almunawar, M.N., Lim, S.A. and Al-Mudimigh, A., 2019. Customer relationship management and big data enabled: Personalization & customization of services. *Applied Computing and Informatics*, *15*(2), pp.94-101.DOI: https://doi.org/10.1016/j.aci.2018.05.004.

[12] Rane, N., 2023. Enhancing Customer Loyalty through Artificial Intelligence (AI), Internet of Things (IoT), and Big Data Technologies: Improving Customer Satisfaction, Engagement, Relationship, and Experience. *Internet of Things (IoT), and Big Data Technologies: Improving Customer Satisfaction, Engagement, Relationship, and Experience (October 13, 2023)*. DOI: http://dx.doi.org/10.2139/ssrn.4616051.

[13] Praveen, K.B., Kumar, P., Prateek, J., Pragathi, G. and Madhuri, J., 2020. Inventory management using machine learning. *International Journal of Engineering Research & Technology (IJERT)*, *9*(06), pp.866-869.

[14] Kietzmann, J., Paschen, J. and Treen, E., 2018. Artificial intelligence in advertising: How marketers can leverage artificial intelligence along the consumer journey. *Journal of Advertising Research*, *58*(3), pp.263-267. DOI: https://doi.org/10.2501/JAR-2018-035

[15] Nica, E., Sabie, O.M., Mascu, S. and Luţan, A.G., 2022. Artificial intelligence decision-making in shopping patterns: Consumer values, cognition, and attitudes. *Economics, Management and Financial Markets*, *17*(1), pp.31-43.

[16] Nazari, E., Shahriari, M.H. and Tabesh, H., 2019. BigData analysis in healthcare: apache hadoop, apache spark and apache flink. *Frontiers in Health Informatics*, *8*(1), p.14. DOI: https://doi.org/10.30699/fhi.v8i1.180.

[17] Borthakur, D., Gray, J., Sarma, J.S., Muthukkaruppan, K., Spiegelberg, N., Kuang, H., Ranganathan, K., Molkov, D., Menon, A., Rash, S. and Schmidt, R., 2011, June. Apache hadoop goes realtime at facebook. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data* (pp. 1071-1080). DOI: https://doi.org/10.1145/1989323.1989438.

[18] Ahmed, N., Barczak, A.L., Susnjak, T. and Rashid, M.A., 2020. A comprehensive performance analysis of Apache Hadoop and Apache Spark for large scale data sets using HiBench. *Journal of Big Data*, *7*(1), pp.1-18. DOI: https://doi.org/10.1186/s40537-020-00388-5.

[19] Moniruzzaman, A.B.M. and Hossain, S.A., 2013. Nosql database: New era of databases for big data analytics-classification, characteristics and comparison. *arXiv preprint arXiv:1307.0191*. DOI: https://doi.org/10.48550/arXiv.1307.0191.

[20] Oussous, A., Benjelloun, F.Z., Lahcen, A.A. and Belfkih, S., 2017. NoSQL databases for big data. *International Journal of Big Data Intelligence*, *4*(3), pp.171-185. DOI: https://doi.org/10.1504/IJBDI.2017.085537.

[21] Kausar, M.A. and Nasar, M., 2021. SQL versus NoSQL databases to assess their appropriateness for big data application. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, *14*(4), pp.1098-1108. DOI: https://doi.org/10.2174/2213275912666191028111632.

[22] Embarak, D.O. and Embarak, O., 2018. The importance of data visualization in business intelligence. *Data analysis and visualization using python: analyze data to create visualizations for BI systems*, pp.85-124.

[23] Embarak, D.O., Embarak and Karkal, 2018. *Data analysis and visualization using python*. Berkeley, CA, USA: Apress.

[24] Grant, R., 2018. *Data visualization: Charts, maps, and interactive graphics*. Crc Press.

[25] Yim, A., Chung, C. and Yu, A., 2018. *Matplotlib for Python Developers: Effective techniques for data visualization with Python*. Packt Publishing Ltd.

[26] Behl, A., Dutta, P., Lessmann, S., Dwivedi, Y.K. and 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*, pp.285-318. DOI: https://doi.org/10.1007/s10257-019-00452-5.

[27] Li, L. and Zhang, J., 2021. Research and analysis of an enterprise E-commerce marketing system under the big data environment. *Journal of Organizational and End User Computing (JOEUC)*, *33*(6), pp.1-19.

[28] Chaudhary, H., 2020. Analyzing the paradigm shift of consumer behavior towards E-Commerce during pandemic lockdown. *Available at SSRN 3664668*. DOI: https://dx.doi.org/10.2139/ssrn.3664668.

[29] Rita, P. and Ramos, R.F., 2022. Global research trends in consumer behavior and sustainability in E-Commerce: A bibliometric analysis of the knowledge structure. *Sustainability*, *14*(15), p.9455. DOI: https://doi.org/10.3390/su14159455.

[30] Policarpo, L.M., da Silveira, D.E., da Rosa Righi, R., Stoffel, R.A., da Costa, C.A., Barbosa, J.L.V., Scorsatto, R. and Arcot, T., 2021. Machine learning through the lens of e-commerce initiatives: An up-to-date systematic literature review. *Computer Science Review*, *41*, p.100414. DOI: https://doi.org/10.1016/j.cosrev.2021.100414.

[31] Shen, Y., 2020, November. Application of Big Data technology in e-commerce. In *Journal of Physics: Conference Series* (Vol. 1682, No. 1, p. 012075). IOP Publishing.

[32] Wu, C., Li, H., Ren, J., Marimuthu, K. and Kumar, P.M., 2021. Artificial neural network based high dimensional data visualization technique for interactive data exploration in E-commerce. *Annals of Operations Research*, pp.1-19. DOI: https://doi.org/10.1007/s10479-021-04436-y.

[33] Abbas, W.F., Ahmad, N.D. and Zaini, N.B., 2013, December. Discovering purchasing pattern of sport items using market basket analysis. In *2013 International Conference on Advanced Computer Science Applications and Technologies* (pp. 120-125). IEEE. DOI: https://doi.org/10.1109/ACSAT.2013.31.

[34] Abdelrhim, M. and Elsayed, A., 2020. The Effect of COVID-19 Spread on the e-commerce market: The case of the 5 largest e-commerce companies in the world. *Available at SSRN 3621166*. DOI: https://dx.doi.org/10.2139/ssrn.3621166.

[35] Galhotra, B. and Dewan, A., 2020, October. Impact of COVID-19 on digital platforms and change in E-commerce shopping trends. In *2020 fourth international conference on I-SMAC (IoT in social, mobile, analytics and cloud)(I-SMAC)* (pp. 861-866). IEEE. DOI: https://doi.org/10.1109/I-SMAC49090.2020.9243379.

[36] Kamthania, D., Pawa, A. and Madhavan, S.S., 2018. Market segmentation analysis and visualization using K-mode clustering algorithm for E-commerce business. *Journal of computing and information technology*, *26*(1), pp.57-68. DOI: https://doi.org/10.20532/cit.2018.1003863.

