MATICS : Jurnal Ilmu Komputer dan Teknologi Informasi ISSN : 1978-161X(p); 2477-2550(e)
Volume 17, No. 1(2025), pp 1-4 DOI : 10.18860/mat.v17i1.24274
Received : November 19th 2023; **Accepted** : March 17th 2025 ; **Avalaible Online** : March 28th 2025

Marketing Optimization: Purchase Data-Based Customer Segmentation Decision Support System

Vinncent Alexander Wong, Muhammad Althaaf Fadhillah, Dzikry Aji Santoso, Luthfia Rahmi Setyorini, Andi Ahyar Almuhajir, Yusi Tyroni Mursityo

***Abstract*—In the era of technological development and changes in shopping culture, e-commerce is increasingly dominating the market, and customer purchase data is becoming a valuable source of information for companies. To address the challenges of inappropriate targeting, customer retention, customer satisfaction, and measuring the effectiveness of marketing campaigns, this research aims to design a decision support system for customer segmentation based on purchase data, identify the optimal parameters of clustering algorithms, and develop appropriate marketing strategies for each group of customers generated from clustering. By using tools such as Matplotlib, Numpy, and Pandas, this research is expected to provide valuable guidance for companies in optimizing their marketing strategies in the competitive e-commerce market.**

***Index Terms*—Customer Segmentation,Decision Support System, E-commerce, Market Trend**

I. INTRODUCTION

The current technological advancements have driven transformations across various sectors, including the creative economy. Changes in shopping culture are also evident due to technological progress, shifting from traditional brick-and-mortar stores to online platforms as a means of commerce. E-commerce has witnessed a substantial increase as both sellers and buyers prefer the convenience and efficiency it offers in terms of time and effort.

In the digital era, companies gather valuable customer purchase data, serving as a crucial source of information to understand customer behavior, product preferences, and market trends. By comprehending this data, companies can identify customers, measure satisfaction, and develop more precise marketing strategies. Customer segmentation is essential in data processing, categorizing customers into groups with similar characteristics. Clustering methods, within the context of data mining and Customer Relationship Management (CRM), play a vital role in identifying customer segments with similar purchasing patterns, allowing companies to formulate focused and effective marketing strategies for each segment.

This paper addresses the integration of CRM and analytical structures like data mining, employing techniques such as decision trees, clustering algorithms, genetic algorithms, and association rule algorithms. These techniques have been utilized in various fields, including commerce, to solve customer-related issues and formulate new strategies.

Customer purchase data serves as a valuable source of information for businesses. Understanding and analyzing this data enable companies to identify customers, measure satisfaction, and develop marketing strategies to enhance sales. However, based on existing data from Amazon, it is observed that many Amazon Sellers lack a profound understanding of marketing strategies based on weekly sales trend segmentation. This knowledge gap can negatively impact sellers' revenue, leading to decreased sales, increased costs, and diminished customer satisfaction.

Amazon Sellers can utilize data classification to comprehend weekly purchasing trends based on Amazon's sales data. Given Amazon's prominence in the e-commerce sector, this understanding can guide Sellers in taking appropriate steps to improve their business based on the classification results and

Manuscript received November 19, 2023. This work was supported in part by Information System Department of Brawijaya University.

Vinncent Alexander Wong is with the Information System Departement of Brawijaya University, Malang, Indonesia (email vincentva44@student.ub.ac.id)

Dzikry Aji Santoso is with the Information System Departement of Brawijaya University, Malang, Indonesia (email dzikryaji@student.ub.ac.id)

Luthfia Rahmi Setyorini is with the Information System Departement of Brawijaya University, Malang, Indonesia (email luthfiarahmi@student.ub.ac.id)

Muhammad Althaaf Fadhiilah is with the Information System Departement of Brawijaya University, Malang, Indonesia (email althaaf02@student.ub.ac.id)

Andi Ahyar Almuhajir Amrani is with the Information System Departement of Brawijaya University, Malang, Indonesia (email andiahyar@student.ub.ac.id)

Yusi Tyroni Mursityo is with the Information System Departement of Brawijaya University, Malang, Indonesia (email yusi_tyro@ub.ac.id)

implement relevant marketing strategies.

The primary research question in this study is how to determine relevant marketing techniques for Amazon Sellers through segmentation based on purchase data. By segmenting Amazon's data, Sellers can develop more effective marketing strategies, offer relevant products, and enhance customer retention. Therefore, this research aims to analyze customer segmentation based on purchase data on Amazon to determine relevant marketing techniques for Sellers on the platform.

This study employs Amazon as the reference e-commerce platform, using its data for analysis. The integration of customer segmentation, purchase data analysis, and marketing techniques aims to provide insights to Amazon Sellers. With a deeper understanding of customer segments, Sellers can optimize marketing strategies and achieve higher customer satisfaction.

II. Research Methodology

This research will employ the following methods:

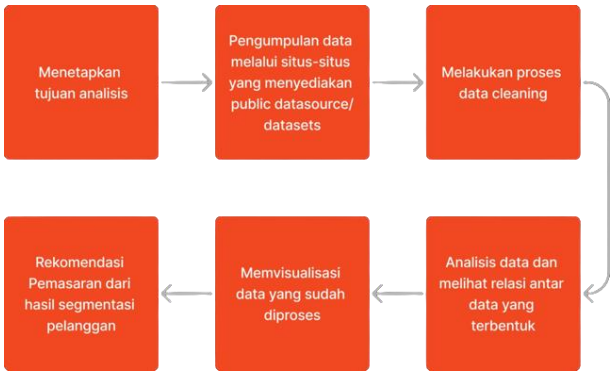

Fig. 1. Research Flowchart

This section outlines the research methodology employed in this study. The initial step involves clearly defining the analysis objectives, aiming to extract relevant patterns and insights from the studied data. Data retrieval from publicly accessible sources, such as websites offering datasets, is the second step, with Kaggle being the selected platform for data collection. The gathered information serves as the starting point for further investigation. The following paragraphs will discuss the details of each step in the research flow based on Figure 1.

Data cleaning, a crucial stage in this research, follows the data collection phase. Considered an effective approach to enhancing data quality, data cleaning minimizes errors and inconsistencies, ensuring data integrity. This process involves a series of minimal operations on raw data to meet specified integrity constraints. The cleaned data is then prepared for subsequent analysis, eliminating unnecessary information and improving data quality.

The next stage is data analysis, aiming to uncover relationships and correlations among variables within the dataset. The use of the K-means Clustering algorithm is pivotal in this phase, as it assists in grouping the cleaned data, facilitating the identification of trends and relevant information for the study. The researchers will develop a Python program for data cleaning and analysis to streamline these processes.

Subsequently, data visualization becomes essential, using Matplotlib, a Python plotting library. This step visualizes the analyzed data to ease the process of recommending marketing techniques in the next stage. The visualized data helps in providing insights and facilitating the identification of patterns.

The final stage involves recommendations. After clustering the data and visualizing the results using Matplotlib, marketing recommendations are derived from the insights gained. Clusters with high sales receive marketing strategies aimed at maintaining or increasing their sales values. On the other hand, clusters with low sales are targeted with marketing techniques to gradually boost their sales.

This research methodology, encompassing data definition, collection, cleaning, analysis, visualization, and recommendation, provides a systematic and comprehensive approach to understanding customer segmentation and deriving relevant marketing strategies from purchase data on Amazon.

III. Result and Discussion

A. Definition of Analysis Objectives

Analysis conducted in this research aims to determine marketing strategies based on the distribution of sales data visualized using Python with the pandas and Matplotlib library.

B. Data Collection

The data utilized in this study was obtained from publicly accessible sources, specifically Kaggle. The dataset was downloaded in .csv format and comprises sales data for various clothing categories on the Amazon e-commerce platform. The included categories are kurta, western dress, top, set, saree, ethnic dress, dupatta, bottom, and blouse. The dataset serves as a valuable resource for analyzing sales patterns and customer preferences within the specified clothing categories on Amazon. The choice of Kaggle as the data source ensures access to a comprehensive and well-structured dataset for the research analysis.

C. Data Cleaning & Data Analysis

This section discusses the data cleaning process carried out using the Python programming language. The process commences with the importation of the pandas and matplotlib libraries. The objective of data cleaning is to eliminate inaccurate data within the dataset, encompassing empty cells, duplicated data, and incorrectly formatted data. Subsequently, an analysis process is undertaken to obtain accurate visualization results. The raw data and clean data may be seen on Figure 2 and Figure 3.

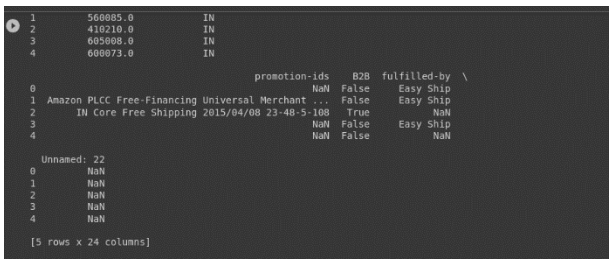

Fig. 2. Before data cleaning

Fig. 3. After *data cleaning*

D. Data Visualization

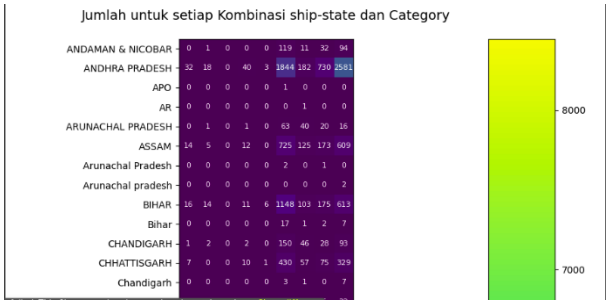

Fig. 4. Data Visualization

The cleaned data is then visualized in the form of a pivot table to enhance data comprehension. However, due to the substantial volume of processed data, the researcher opted to visualize the data using cells to represent a set of data with similar classifications. The results of the data visualization can be observed in Figure 4.

E. Marketing Recommendations

This section will explain further the marketing recommendation based on data visualization that has been applied on Figure 4. The data visualization in this study aims to discern the distribution of sales data. The dataset utilized pertains to the sales of clothing through the Amazon e-commerce platform in the region of India. The x-axis represents clothing categories, while the y-axis signifies regional names. Based on the outcomes of the conducted data visualization, it can be inferred that the sales levels of Saree, Ethnic dress, Dupatta, Bottom, and Blouse, which broadly constitute traditional Indian women's attire, are relatively low. As a measure to enhance sales, the researcher recommends marketing strategies employing the 5P method, encompassing Product, Place, Price, People, and Promotion.

1) Product

Indian traditional attire should be designed with meticulous attention to distinctive traditional patterns, while still considering contemporary fashion trends. Emphasis should be placed on the quality of materials and craftsmanship.

2) Place

Amazon provides a user-friendly platform for sellers to market and sell products. With this ease of use, sellers can swiftly upload products, manage inventory, and interact with customers. Amazon facilitates sellers in distributing their products globally, offering opportunities to reach a broader market share and diversify their customer base.

3) Price

In determining product pricing, sellers must consider market prices to ensure that the set price is neither too cheap nor too expensive. If the set price is too low, potential buyers may harbor doubts regarding the quality of the provided goods. Conversely, if the designated price is too high, buyers may be reluctant to purchase the product.

4) People

The target market for traditional Indian clothing products comprises the Indian community utilizing Amazon as a platform for purchasing goods. The researcher assumes that the target market for these products is individuals aged 18-40 who are actively engaged with gadgets and social media, necessitating all promotional activities to be conducted through digital platforms such as Instagram.

5) Promotion

Sellers can employ professionally shot photographs for each product, enhancing the visual appeal of the displayed product images. To cultivate a sense of pride among the Indian community regarding their regional attire, sellers may engage in endorsements through social media influencers, particularly on Instagram, to broaden the product's reach. In order to heighten brand awareness, sellers can craft a brand logo and store name that reflect the nature of the products being sold.

IV. Conclusion

This research aims to determine marketing strategies to increase sales volume. In order to identify categories with low sales levels, the researcher employed the K-Means Clustering algorithm to process the data available in the dataset. After data processing, visualization was conducted using a pivot table. Based on the visualization results, it can be concluded that the category with low sales is specifically Indian traditional clothing, particularly for women. Therefore, the researcher recommends a marketing strategy utilizing the 5P marketing mix: Product, Place, Price, People, and Promotion. The researcher suggests sellers to consistently ensure the quality, pricing, and presentation of the products are attractive. Additionally, expanding market reach is advised, and sellers can achieve this by endorsing products through influencers on social media. Creating a distinctive logo and store name that reflects the products is also recommended to enhance brand recall for potential buyers.

ACKNOWLEDGMENT

With the completion of this research, the author would like to thank:

Ir. Yusi Tyroni Mursityo, S.Kom., M.AB., as the supervisor of the final project and the head of the Information System Department who has been pleased to guide and support the author in the work of the final project.

REFERENCES

[1] "An overview of Electronic Commerce (e-commerce)," Journal of Contemporary Issues in Business and Government, vol. 27, no. 3, 2021. doi:10.47750/cibg.2021.27.03.090

[2] R.-S. Wu and P.-H. Chou, "Customer segmentation of multiple category data in e-commerce using a soft-clustering approach," Electronic Commerce Research and Applications, vol. 10, no. 3, pp. 331–341, 2011. doi:10.1016/j.elerap.2010.11.002

[3] A. Agrawal, P. Kaur, and M. Singh, "Customer segmentation model using K-means clustering on e-commerce," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), 2023. doi:10.1109/icscds56580.2023.10105070

[4] [M. T. Ballestar, P. Grau-Carles, and J. Sainz, "Customer segmentation in e-commerce: Applications to the Cashback Business Model," Journal of Business Research, vol. 88, pp. 407–414, 2018. doi:10.1016/j.jbusres.2017.11.047

[5] S. Koul and T. M. Philip, "Customer segmentation techniques on e-commerce," 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2021. doi:10.1109/icacite51222.2021.9404659

[6] M. Hosseini and M. Shabani, "New approach to customer segmentation based on changes in customer value," Journal of Marketing Analytics, vol. 3, no. 3, pp. 110–121, 2015. doi:10.1057/jma.2015.10

[7] T. Kansal, S. Bahuguna, V. Singh, and T. Choudhury, "Customer segmentation using K-means clustering," 2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS), 2018. doi:10.1109/ctems.2018.8769171

[8] G. Liadeli, F. Sotgiu, and P. W. J. Verlegh, "A meta-analysis of the effects of brands' owned social media on social media engagement and sales," Journal of Marketing, vol. 87, no. 3, pp. 406–427, 2022. doi:10.1177/00222429221123250

[9] R. Ström, M. Vendel, and J. Bredican, "Mobile marketing: A literature review on its value for consumers and retailers," Journal of Retailing and Consumer Services, vol. 21, no. 6, pp. 1001–1012, 2014. doi:10.1016/j.jretconser.2013.12.003

[10] C. Chunjiang, "E-commerce search engine marketing mechanism analysis and Optimization," 2020 13th International Conference on Intelligent Computation Technology and Automation (ICICTA), 2020. doi:10.1109/icicta51737.2020.00028

[11] X. Ding et al., "IOT data cleaning techniques: A survey," Intelligent and Converged Networks, vol. 3, no. 4, pp. 325–339, 2022. doi:10.23919/icn.2022.0026

[12] H. Koehler and S. Link, "Possibilistic data cleaning," IEEE Transactions on Knowledge and Data Engineering, vol. 34, no. 12, pp. 5939–5950, 2022. doi:10.1109/tkde.2021.3062318

[13] [A. Alamsyah et al., "Customer segmentation using the integration of the Recency Frequency Monetary Model and the K-means cluster algorithm," Scientific Journal of Informatics, vol. 9, no. 2, pp. 189–196, 2022. doi:10.15294/sji.v9i2.39437

[14] G. Das, "Factors affecting Indian shoppers′ attitude and purchase intention: An empirical check," Journal of Retailing and Consumer Services, vol. 21, no. 4, pp. 561–569, 2014. doi:10.1016/j.jretconser.2014.04.005

[15] S. B. Ahmad and A. Rohman, "Analisis Strategi pemasaran Dalam Meningkatkan omset Penjualan Produk Tanaman Hias ditinjau Dari perpektif Ekonomi Islam," Ekonomi, Keuangan, Investasi dan Syariah (EKUITAS), vol. 4, no. 3, pp. 824–831, 2023. doi:10.47065/ekuitas.v4i3.2684