Online version at https://journal.lenterailmu.com/index.php/jafotik

JAFOTIK

E-ISSN: 3031-2698, DOI prefix: 10.70356

PT. Lentera Ilmu Publisher

Jurnal Sistem Informasi dan Teknik Informatika

E-Commerce Activity Clustering Based on Online Shopping and Income per Capita in Indonesian Cities

Tri Prasetyo[1], Pandu Wiliantoro[1], Aldi Feridanto[a], Iqbal Julyansyah[1], Liza Sania Khulatus Syafaah[1]*

[1]*Faculty of Computer Science, Universitas Pamulang, South Tangerang, Indonesia*

ARTICLE INFO

Article history:
Received 01 August 2026
Revised 02 September 2026
Accepted 12 September 2026
Published 15 September 2026

Keywords:
E-Commerce,
K-Means Clustering,
Online shopping,
Income per capita,
Data mining,
Regional clustering

ABSTRACT

The rapid growth of digital technology and internet penetration has contributed to the expansion of e-commerce activity across cities and regencies in Indonesia. However, the intensity of online shopping does not necessarily correspond directly to regional income per capita. This study aims to examine the relationship between online shopping activity and income per capita and to classify Indonesian cities and regencies based on these characteristics using the K-Means Clustering algorithm. This research employs a quantitative approach with a descriptive-exploratory design. Secondary data were collected from statistical and e-commerce sources and processed through data cleaning, transformation, and normalization. The K-Means Clustering algorithm was subsequently applied to group regions according to their levels of online shopping activity and income per capita. The Elbow Method was used to determine the optimal number of clusters. The results identified three clusters with distinct levels of e-commerce activity. Cluster 0 represented regions with the highest level of activity, followed by Cluster 1 and Cluster 2. Several regions demonstrated relatively high levels of e-commerce activity, including Papua, Central Java, West Java, Riau, and DKI Jakarta. These findings provide a regional overview of e-commerce activity and income characteristics that can support digital business strategies and contribute to the development of more inclusive and equitable digital economic policies.

1. Introduction

The development of information and communication technology has brought significant changes to various aspects of human life, particularly in the way people conduct economic activities. One of the most visible changes is the growth of online shopping through e-commerce. The increasing penetration of the internet and the widespread use of smartphones have encouraged people to shift from conventional purchasing activities toward digital transactions. In Indonesia, the growth of e-commerce has become

* *Corresponding author:* Tri Prasetyo
E-mail address: dosen02669@unpam.ac.id

increasingly significant, making online shopping an important component of the country's digital economy.

The growth of e-commerce is supported by various conveniences offered to consumers, including easy access to products, faster transactions, product variety, and promotional programs provided by digital commerce platforms. The COVID-19 pandemic also accelerated digital transformation in the trade sector, as restrictions on community activities encouraged people to use online transactions to fulfill their daily needs. These developments have resulted in increasingly diverse digital transaction activities across different regions in Indonesia [1].

However, the development of e-commerce is not evenly distributed across all cities in Indonesia. Large cities such as Jakarta, Surabaya, Bandung, and Medan tend to demonstrate higher levels of online shopping activity compared with medium-sized and smaller cities. This difference may be associated with various economic and technological factors, including the purchasing power of the population. Income per capita is commonly used as an indicator of economic welfare and may provide an indication of the purchasing capacity of people in a particular region [2].

The relationship between online shopping activity and income per capita is not necessarily uniform across cities. A city with relatively high income per capita does not always demonstrate a high level of online shopping activity, as some communities may still prefer conventional transactions. Conversely, cities with moderate income levels may exhibit relatively high online shopping activity due to technological adoption, lifestyle changes, and intensive e-commerce promotions. This phenomenon indicates the need for an analytical approach capable of identifying groups of cities with similar characteristics based on online shopping activity and income per capita [3].

A data-based mapping of cities can provide a clearer understanding of the distribution of digital commerce activity in Indonesia. Such mapping can identify cities with high digital consumption and relatively low economic capacity, cities with high purchasing power but relatively limited online shopping activity, and cities in which purchasing power and digital consumption show relatively similar patterns. This information is relevant for understanding differences in the development of the digital economy and the potential existence of a digital divide among regions in Indonesia.

To address this issue, data mining can be applied to process and analyze data containing different characteristics of cities. One of the techniques that can be used is clustering, which groups observations based on similarities among their characteristics. In this study, the K-Means Clustering algorithm is selected to group cities in Indonesia based on online shopping activity and income per capita. The algorithm allows cities with relatively similar characteristics to be assigned to the same cluster, thereby providing a structured representation of the distribution of digital consumption and economic conditions [4].

This study uses a quantitative approach with a descriptive-exploratory design. The research focuses on clustering cities in Indonesia using two main variables: online shopping activity and income per capita. Online shopping activity is represented by the number or index of e-commerce transactions, while income per capita represents the average income of the population in each city and serves as an indicator of purchasing power and economic welfare. The population consists of cities in Indonesia, while the analyzed sample is determined based on the availability and completeness of data for both variables. The data are collected using a documentation method from secondary sources. Income per capita data are obtained from the Badan Pusat Statistik (BPS), while online shopping information is obtained from e-commerce industry reports, official publications, marketplace reports, and other relevant public sources. The collected data are subjected to preprocessing, including data cleaning and normalization, to ensure that incomplete or inconsistent observations are handled appropriately and that the variables have comparable scales before the clustering process is performed [5].

The K-Means Clustering process consists of determining the optimal number of clusters, initializing the cluster centers or centroids, calculating the distance between observations and centroids using Euclidean Distance, assigning observations to the nearest cluster, and updating the centroids based on the mean values of observations within each cluster. This process is repeated until the cluster centers become stable. The optimal number of clusters can be considered using the Elbow Method or Silhouette Coefficient, while the quality and characteristics of the resulting clusters are evaluated through quantitative and descriptive analysis [6].

The main objective of this study is to analyze the relationship between online shopping activity and income per capita across cities in Indonesia and to apply K-Means Clustering to group cities according to their similarities in digital consumption and economic conditions. The study also aims to provide information that can support business actors, government institutions, and academics in understanding the characteristics and potential of different city clusters within Indonesia's digital economy. The resulting classification can be used as a basis for identifying areas with high digital market potential as well as areas where the utilization of e-commerce remains relatively limited [7]. Therefore, this research is expected to contribute to a more structured understanding of the distribution of e-commerce activity in Indonesia by combining indicators of online shopping and income per capita within a clustering framework. The results are expected to provide a data-driven overview of the characteristics of Indonesian cities and support further research related to digital consumer behavior, regional economic conditions, and the development of a more inclusive digital economy [8].

2. Method

2.1 Research Approach

This study employs a quantitative approach combined with data mining to cluster cities in Indonesia based on online shopping activity and income per capita. The main objective is to group cities with similar characteristics in terms of digital consumption and economic conditions. The resulting clusters are used to provide an overview of the relationship between online shopping activity and the economic conditions of each city.

2.2 Data Collection

The study uses secondary data collected through a documentation method [9]. The data are obtained from relevant official and public sources. Income per capita data are obtained from the Badan Pusat Statistik (BPS), while information related to online shopping activity is collected from e-commerce industry reports, market research reports, marketplace publications, and other relevant government or research institution publications. The main data required for the analysis consist of online shopping activity and income per capita for each city or regency [10].

2.3 Research Variables

Two main variables are used in the clustering process. The first variable is online shopping activity, represented by the number or index of e-commerce transactions that reflects the intensity of digital consumption in each city. The second variable is income per capita, represented by the average income of the population in each city and used as an indicator of purchasing power and economic welfare (Table 1).

Table 1 - Research Variables

Variable	Description	Measurement
Online Shopping Activity	Represents the intensity of digital consumption in each city	Number or index of e-commerce transactions
Income per Capita	Represents purchasing power and economic welfare	Average income per city

2.4 Data Preprocessing

Before the clustering process, the collected data are subjected to data preprocessing to improve data quality and ensure that the variables can be appropriately analyzed. The preprocessing consists of data cleaning, data transformation, and normalization. Data cleaning is performed by removing incomplete, duplicate, or inconsistent observations. Data transformation is used to adjust the format and scale of the variables, while normalization is applied to ensure that online shopping activity and income per capita have comparable scales in the clustering process [11].

2.5 K-Means Clustering

The main analytical method used in this study is K-Means Clustering. The algorithm groups cities into several clusters based on the similarity of their online shopping activity and income per capita. The clustering process begins by determining the optimal number of clusters, followed by the initialization of cluster centroids. The distance between each observation and the centroids is then calculated using Euclidean Distance [12]. Each city is assigned to the nearest cluster, after which the position of each centroid is recalculated based on the mean values of the observations within the cluster. This process is repeated until the cluster centers become stable or the predefined iteration condition is reached [13].

The optimal number of clusters is determined using the Elbow Method and Silhouette Coefficient. These approaches are used to identify an appropriate cluster configuration before the final clustering analysis is performed.

2.6 Cluster Evaluation

The resulting clusters are evaluated using the Silhouette Score and descriptive analysis. The Silhouette Score is used to assess the quality and internal consistency of the clusters, while descriptive analysis is conducted to identify the characteristics of each cluster. The evaluation focuses on whether the resulting groups provide a meaningful representation of cities with similar patterns of online shopping activity and income per capita [14].

2.7 Visualization and Interpretation

The clustering results are presented through tables, graphs, and city-level cluster visualizations. The visualization process is used to facilitate the interpretation of the characteristics and distribution of the resulting clusters. The analysis particularly considers cities with high online shopping activity but relatively low income per capita, cities with high income per capita but relatively low online shopping activity, and cities showing a relatively balanced relationship between purchasing power and digital consumption [15].

2.8 Research Flow

The overall research process consists of seven stages: literature review, data collection, data preprocessing, K-Means Clustering, cluster evaluation, visualization and interpretation, and final report preparation (Figure 1). The research flow is designed to provide a systematic sequence from data collection to the interpretation of clustering results.

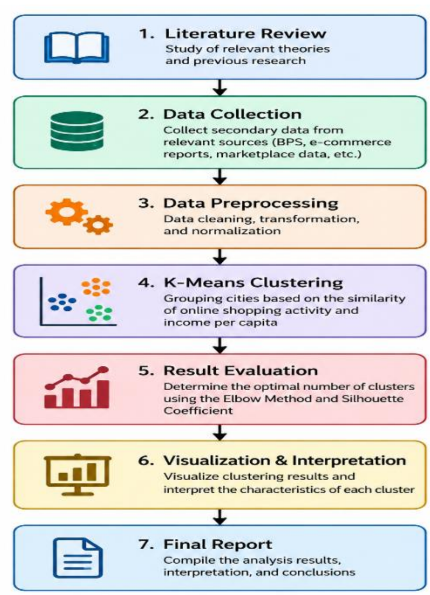

Figure 1 - Research Flow Chart

3. Result and Discussion

3.1 Dataset and Data Analysis

The analysis was conducted using Google Colab as the computational environment for processing and visualizing the e-commerce dataset. The dataset was processed to identify patterns of online shopping activity and income per capita across cities or regencies. The analysis included data visualization, clustering, and evaluation of the resulting groups. The dataset used in the analysis contains several variables related to e-commerce transactions, including transaction quantity, product weight, returned quantity, discount, product categories, shipping costs, total payment, and income per capita. The report states that the dataset consisted of 20,848 observations.

3.2 Online Shopping Activity

The visualization of online shopping activity was performed to identify cities or regencies with relatively high levels of e-commerce activity. The analysis produced a ranking of the locations with the highest shopping activity. Based on the reported results, Kabupaten Tangerang occupied the first position, followed by Kabupaten Bogor and Kota Tangerang. The reported income-per-capita values for these locations were above 101,244,178 for Kabupaten Tangerang, above 91,306,900 for Kabupaten Bogor, and above 41,326,237 for Kota Tangerang (Figure 2).

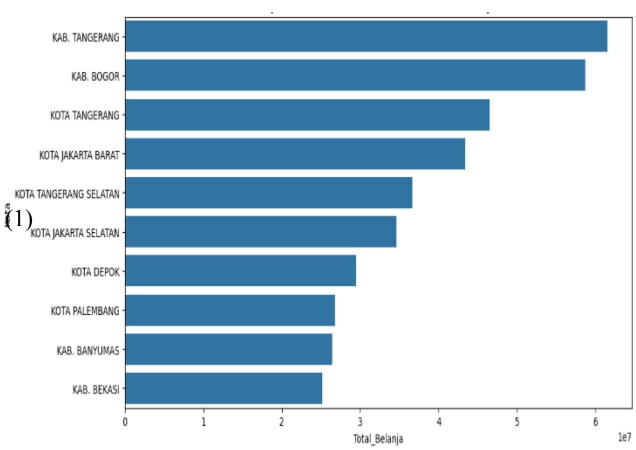

Figure 2 - Top 10 Cities or Regencies Based on Total E-Commerce Spending

3.3 K-Means Clustering Results

The K-Means Clustering analysis was applied to group the observations according to their characteristics. The results produced three clusters, namely Cluster 0, Cluster 1, and Cluster 2. Among the three groups, Cluster 0 was reported as having the highest e-commerce activity, followed by Cluster 1 and Cluster 2.

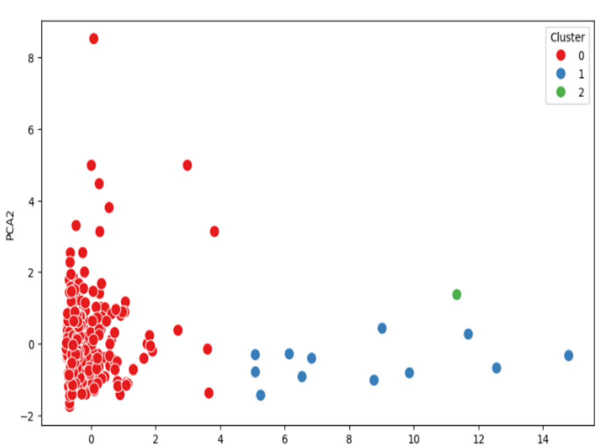

Figure 3 - K-Means Clustering Results

The formation of three clusters indicates that the observed data can be divided into groups with different levels of e-commerce activity. Cluster 0 represents the group with the highest activity, while Cluster 2 represents the group with the lowest activity among the three clusters. This classification provides a more structured view of the distribution of online shopping activity than a simple ranking of individual cities or regencies (Figure 3).

3.4 Determination of the Number of Clusters

The Elbow Method was used to support the determination of the appropriate number of clusters. This method evaluates the pattern of clustering performance across different numbers of clusters and helps identify a suitable point at which adding additional clusters provides diminishing improvement. The report presents the Elbow Method visualization as part of the process of selecting the clustering configuration (Figure 4).

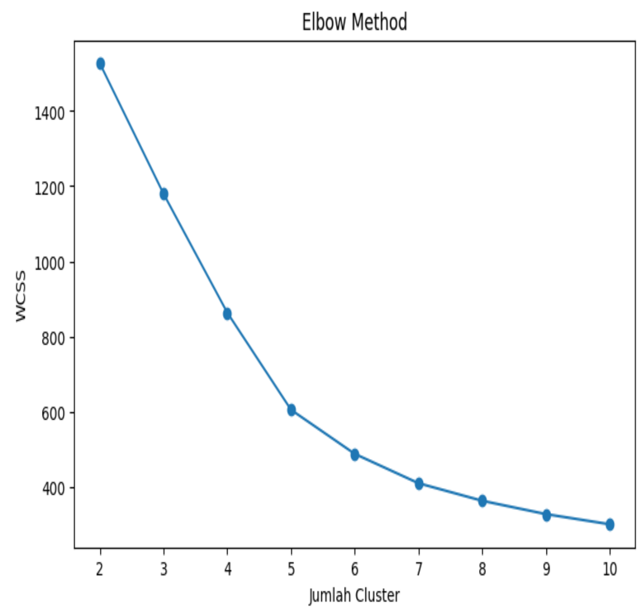

Figure 4 - Elbow Method for Cluster Selection

The Elbow Method visualization supports the use of three clusters in the subsequent analysis. The selected configuration provides a manageable grouping of the observations while maintaining differences among the resulting clusters. The use of this approach is consistent with the research objective of identifying groups of cities or regencies with similar characteristics.

3.5 Distribution of Active E-Commerce Areas

Further visualization was conducted to identify the regions with the highest active e-commerce activity. The five locations reported as having the highest activity were Papua, Jawa Tengah, Jawa Barat, Riau, and DKI Jakarta. Papua was reported as having more than 1,600 active shopping activities, followed by Jawa Tengah with more than 1,200, Jawa Barat with more than 1,000, Riau with approximately 600, and DKI Jakarta with more than 400 (Figure 5).

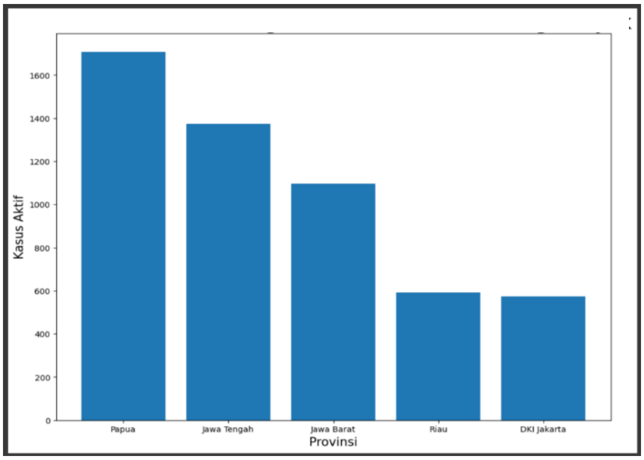

Figure 5 - Five Cities or Regencies with the Highest Active E-Commerce Activity

The distribution indicates that high e-commerce activity is not concentrated exclusively in the largest metropolitan areas. The presence of Papua, Jawa Tengah, Jawa Barat, Riau, and DKI Jakarta among the reported highest-activity regions suggests differences in digital consumption patterns across regions. These differences provide an important basis for interpreting the resulting clusters in relation to regional economic characteristics.

3.6 Discussion

The results demonstrate that K-Means Clustering can be used to organize e-commerce observations into groups according to their characteristics. The identification of three clusters provides a simplified representation of differences in online shopping activity among the observed locations. The results also show that individual locations with high activity do not necessarily represent the characteristics of all regions, making a clustering approach useful for identifying broader patterns. The analysis further indicates that online shopping activity varies substantially among regions. The ranking results show that several locations have considerably higher activity than others, while the active-region analysis reveals differences across geographical areas. These findings are relevant to the research objective of mapping digital consumption patterns and examining their relationship with income per capita.

From a practical perspective, the resulting clusters can provide information for businesses when determining potential e-commerce markets and developing location-specific marketing strategies. For government institutions, the clustering results can support the identification of regions with different levels of digital economic activity and can contribute to considerations related to more equitable digital economic development. The results can also provide a basis for further research examining other factors that may influence online shopping activity.

4. Conclusion

This study applied the K-Means Clustering algorithm to analyze and group cities or regencies in Indonesia based on E-Commerce activity and income per capita. The results show that online shopping activity varies across regions and does not always correspond directly to income levels. The clustering process produced three clusters with different levels of E-Commerce activity, with Cluster 0 representing the highest activity, followed by Cluster 1 and Cluster 2. The Elbow Method was used to determine the appropriate number of clusters, while visualization helped identify differences in regional E-Commerce activity. The analysis also showed that several regions had particularly high levels of active E-Commerce activity, including Papua, Central Java, West Java, Riau, and DKI Jakarta. Overall, the results demonstrate that K-Means Clustering can be used to provide a clearer regional mapping of E-Commerce activity and its relationship with economic conditions. The resulting clusters can support businesses, digital SMEs, and policymakers in developing more targeted strategies for digital economic development and more equitable E-Commerce adoption across regions.

Acknowledgements

The authors would like to express their sincere gratitude to Universitas Pamulang for providing support and permission to conduct this research. The authors also extend their appreciation to all parties who provided guidance, encouragement, and support throughout the research process. Special thanks are given to the authors' family for their continuous prayers, motivation, and support during the completion of this research.

References

[1] M. Sinaga, S. Winda Hardiyanti Damanik, R. Surianto Zalukhu, R. Piter Sony Hutauruk, and D. Collyn, "Pengaruh Pendidikan, Pendapatan Per Kapita Dan Pengangguran Terhadap Kemiskinan Di Kepulauan Nias," *J. Ekuilnomi*, vol. 5, no. 1, 2023, doi: https://doi.org/10.36985/a4tg4s88.

[2] Z. Alamin, R. Missouri, S. Sutriawan, F. Fathir, and K. Khairunnas, "Perkembangan E-commerce: Analisis Dominasi Shopee sebagai Primadona Marketplace di Indonesia," *J-ESA (Jurnal Ekon. Syariah)*, vol. 6, no. 2, 2023, doi: https://doi.org/10.52266/jesa.v6i2.2484.

[3] M. Sinaga, S. W. H. Damanik, R. S. Zalukhu, R. P. S. Hutauruk, and D. Collyn, "Pengaruh Pendidikan, Pendapatan Per Kapita Dan Pengangguran Terhadap Kemiskinan Di Kepulauan Nias," *J. Ekuilnomi*, vol. 5, no. 1, 2023.

[4] T. H. Mardzuki, R. Lubis, and F. F. Adiwijaya, "Penerapan Algoritma K-Means Clustering pada Sistem Prediksi Kelulusan Tepat Waktu," *Komputika J. Sist. Komput.*, vol. 13, no. 2, 2024, https://doi.org/10.34010/komputika.v13i2.14097.

[5] E. Gari and R. Sari, "Analisis Segmentasi Wilayah Penjualan menggunakan Algoritma K-Means Clustering," *J. Infortech*, vol. 7, no. 2, 2025, doi: https://doi.org/10.31294/infortech.v7i2.11385.

[6] D. Irawan, G. Wijaya, and T. T. Warisaji, "Penerapan Algoritma K-Means Clustering untuk Segmentasi Nasabah Bank," *BIOS J. Teknol. Inf. dan Rekayasa Komput.*, vol. 6, no. 1, 2025.

[7] A. N. Khormarudin *et al.*, "Teknik Data Mining: Algoritma K-Means Clustering," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 1, no. 2, 2016.

[8] R. Supardi and I. Kanedi, "Implementasi Metode Algoritma K-

Means Clustering pada Toko Eidelweis," *J. Teknol. Inf.*, vol. 4, no. 2, 2020, https://jurnal.una.ac.id/index.php/jurti/article/view/1444.

[9] A. P. Bagustio, A. I. Purnamasari, and Irfan Ali, "ANALISIS DATA PENJUALAN MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING PADA TOKO KECANTIKAN PUTRI," *PROSISKO J. Pengemb. Ris. dan Obs. Sist. Komput.*, vol. 11, no. 2, 2024.

[10] P. sari and R. Syahri, "ALGORITMA K-MEANS CLUSTERING: SEBUAH STUDI LITERATUR," *J. Inform.*, vol. 1, no. 01, 2024.

[11] A. A. Alya Putri and S. A. Rahmah, "IMPLEMENTASI DATA MINING DENGAN ALGORITMA K-MEANS CLUSTERING UNTUK ANALISIS BISNIS PADA PERUSAHAAN ASURANSI," *Djtechno J. Teknol. Inf.*, vol. 5, no. 1, 2024.

[12] N. Kholidin, N. Suarna, and W. Prihartono, "Segmentasi Gaya Hidup Mahasiswa Menggunakan Algoritma K-Means Clustering," *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 12, no. 3, 2024.

[13] V. Alvianatinova, I. Ali, N. Rahaningsih, and A. Bahtiar, "PENERAPAN ALGORITMA K-MEANS CLUSTERING DALAM PENGELOMPOKAN DATA PENJUALAN SUPERMARKET BERDASARKAN CABANG (BRANCH)," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, 2024.

[14] A. Nur Khormarudin *et al.*, "Teknik Data Mining: Algoritma K-Means Clustering," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 1, no. 2, 2022.

[15] R. S. Lisudatu, M. Marchelin, and L. K. Wibisono, "Klasifikasi UMKM Berdasarkan Kinerja Keuangan Menggunakan Algoritma K-Means Clustering," *J. Bisnis Mhs.*, vol. 5, no. 2, 2025.