REGIONAL TYPOLOGY OF E-COMMERCE BUSINESS CONSTRAINTS IN INDONESIA: A MACHINE LEARNING APPROACH

Fathur Rachman[1]*, Harun Al Azies[2,3]
[1]Universitas Airlangga (Indonesia)
[2]Universitas Dian Nuswantoro (Indonesia)
[3]Institut Teknologi Bandung (Indonesia)
*) email: fathur.rachman-2024@vokasi.unair.ac.id

Abstract

This study analyzes regional disparities in e-commerce business constraints in Indonesia using an unsupervised machine learning approach. Using province-level data from *Statistik E-Commerce 2024* published by Statistics Indonesia, the analysis covers 38 provinces. It examines seven major constraints, including funding limitations, skilled labor shortages, limited internet access, fraud, marketing challenges, delivery constraints, and other operational barriers. K-Means clustering with z-score standardization is applied to identify regional typologies of e-commerce business constraints. The optimal number of clusters is determined using the elbow method, the silhouette score, the Davies-Bouldin index, and the Calinski-Harabasz index. The results reveal five distinct regional clusters with different combinations of constraints. The findings show that provinces in Java and Bali are mainly constrained by capital and marketing pressures despite relatively advanced digital infrastructure. Several regions outside Java face balanced structural constraints involving multiple interrelated obstacles, while capital-heavy constraints dominate others. In contrast, Papua Pegunungan and Papua Tengah exhibit severe digital infrastructure constraints, indicating persistent digital divides. This study contributes by providing a province-level typology of e-commerce business constraints using official statistics and machine learning, offering a data-driven basis for designing region-specific strategies to support inclusive e-commerce development in Indonesia.

Keywords: E-commerce, machine learning, clustering analysis, regional typology.

1. INTRODUCTION

The rapid expansion of the digital economy has positioned e-commerce as a key driver of business transformation and regional economic development (Ahi et al., 2022). In Indonesia, the growth of e-commerce has been supported by rising internet penetration, the expansion of digital payment systems, and the growth of logistics platforms (Sosyal Araştırmalar, 2025). For businesses, particularly small and medium-sized enterprises, e-commerce provides opportunities to expand market access and improve operational efficiency (Xu et al., 2022). Nevertheless, the adoption and sustainability of e-commerce activities remain uneven across regions (Criveanu, 2023).

Recent national data indicate that only 42.02 percent of businesses in Indonesia have adopted e-commerce, while the majority remain non-e-commerce. This disparity suggests that digital transformation is not determined solely by technological availability but is also shaped by local economic conditions, human capital,

infrastructure, and business environments (Al Azies & Herowati, 2023). Consequently, e-commerce businesses across Indonesian provinces face heterogeneous constraints that affect their sustainability and growth potential.

Previous studies have examined e-commerce development in Indonesia from different perspectives. Asikin (2024) employs cluster analysis on firm-level microdata from the national e-commerce survey to identify diverse e-commerce business models based on business characteristics, innovation activities, and performance. While this study highlights heterogeneity among firms, its analysis is limited to the firm level and does not address regional disparities or spatial patterns of business constraints(Asikin, 2024). Other studies have applied machine learning techniques to e-commerce data for predictive purposes. For example, Wijaya et al. (2022) demonstrate that e-commerce transaction data can be used to estimate city-level poverty rates, illustrating the potential of e-commerce data as a proxy for socio-economic indicators. However, such studies focus on outcome prediction rather than on identifying the structural constraints faced by e-commerce businesses across regions(Wijaya et al., 2020).

In addition, research on e-commerce and geographical challenges in Indonesia's retail sector has largely relied on qualitative approaches and literature reviews. Bakri et al. (2024) emphasize the role of logistics, market penetration strategies, and firm-level innovation in overcoming geographical barriers in Indonesia's archipelagic context(Bakri et al., 2024). Although this line of research underscores the importance of geography and infrastructure, it does not provide a systematic, data-driven classification of regional e-commerce business constraints using official statistics. Existing studies either focus on firm-level typologies, predictive modeling, or qualitative assessments of geographical challenges. Limited attention has been given to developing a province-level typology of e-commerce business constraints that integrates financial, human capital, logistical, and digital infrastructure dimensions within a single analytical framework. This gap motivates the present study.
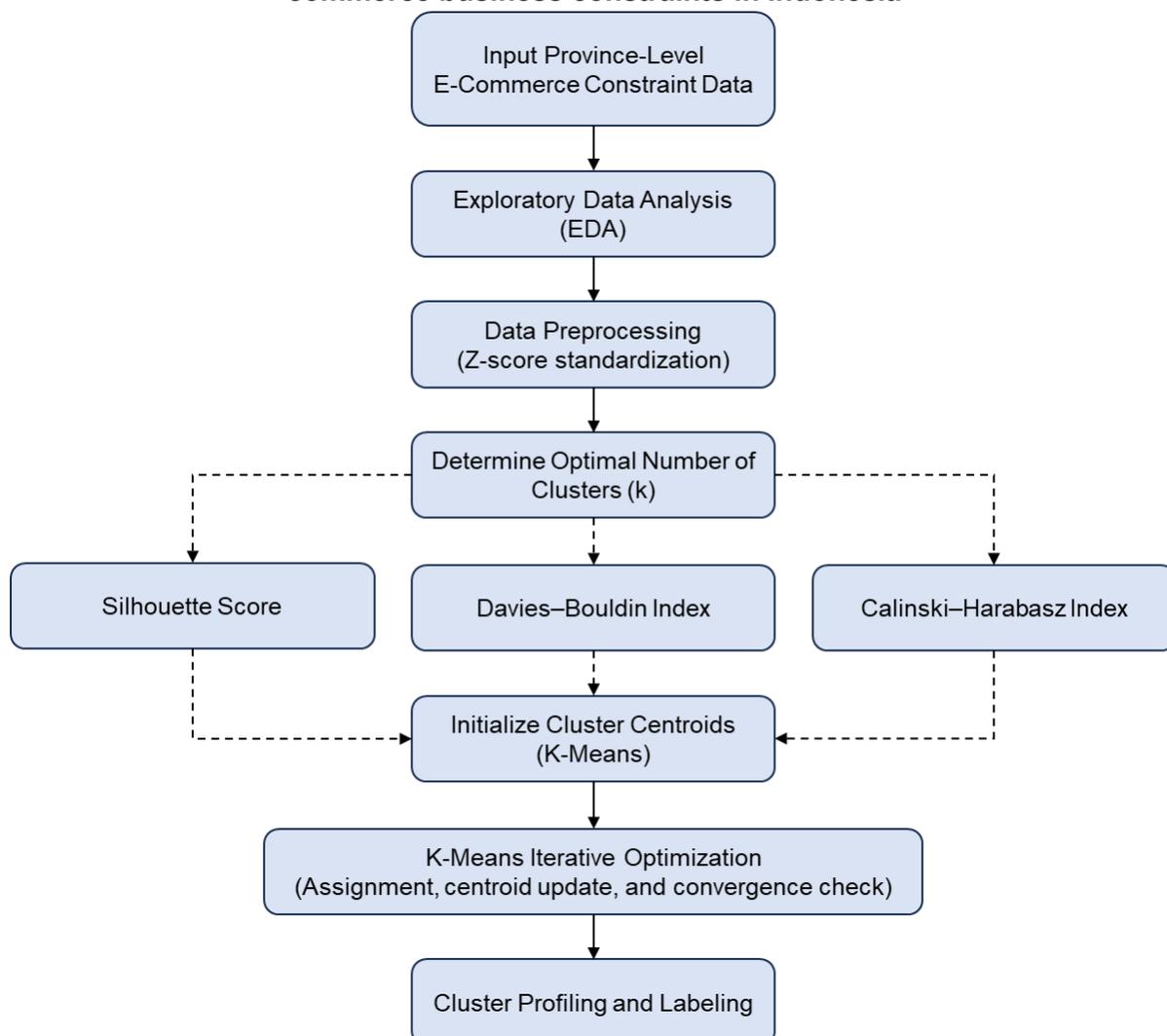
Statistics Indonesia, through the Statistik E-Commerce 2024 publication, provides comprehensive province-level data on the major obstacles faced by e-commerce businesses, including a lack of funding, a shortage of skilled workers, limited internet access, fraud in buying and selling processes, product marketing constraints, delivery service limitations, and other operational challenges. These indicators capture multidimensional risks that directly affect e-commerce sustainability across regions. Given the multidimensional and regionally heterogeneous nature of these constraints, an exploratory and pattern-oriented analytical approach is required(Dadashpoor et al., 2022; Li et al., 2026). Unsupervised machine learning methods, particularly clustering techniques, are well-suited to uncover latent regional structures without imposing predefined outcome variables (Azies & Rositawati, 2021; Baligodugula, 2023; Monica et al., 2021). This approach allows regional characteristics to emerge directly from the data and is especially relevant for policy-oriented analysis.

This study aims to develop a regional typology of e-commerce business constraints in Indonesia using a machine learning–based clustering approach. By applying K-Means clustering to standardized provincial data, this study identifies distinct groups of provinces with similar constraint profiles. The resulting typology provides a structured overview of regional challenges and offers empirical insights to support region-specific strategies for sustainable e-commerce development.

2. METHODOLOGY

This study adopts a quantitative and exploratory research design using an unsupervised machine learning approach. The methodological workflow is summarized in Figure 1 and consists of sequential stages ranging from data preparation to the development of a regional typology of e-commerce business constraints in Indonesia.

Figure 1. Research methodology workflow for identifying regional typologies of e-commerce business constraints in Indonesia

2.1 Data Source and Unit of Analysis

The study utilizes secondary data obtained from *Statistik E-Commerce 2024* published by Statistics Indonesia. The dataset provides province-level information on the major constraints faced by e-commerce businesses across Indonesia. The unit of analysis is the province, covering all 38 provinces. Using provincial-level data enables the analysis to capture spatial heterogeneity in e-commerce business conditions within a national context.

Seven variables are employed to represent major e-commerce business constraints, namely lack of funding, lack of skilled workers, limited internet access, fraud in buying and selling activities, product marketing constraints, delivery service limitations, and other operational challenges. Each variable is measured as a percentage, reflecting the proportion of e-commerce businesses in each province reporting a specific constraint.

2.2 Exploratory Data Analysis

As illustrated in the first analytical stage of the flowchart, an exploratory data analysis is conducted to examine the basic characteristics of the data. This stage includes descriptive statistical analysis to summarize the distribution of each constraint across provinces, as well as visual inspection of variable distributions(Cooksey, 2020; Mishra et al., 2019). In addition, Pearson correlation analysis is applied to assess the relationships among constraint variables and to identify potential redundancy (Edelmann et al., 2021; Liu et al., 2020). The exploratory results reveal substantial regional heterogeneity and generally low to moderate correlations, supporting the inclusion of all variables in the clustering analysis.

2.3 Data Preprocessing

Following the exploratory stage, data preprocessing is performed as shown in the flowchart. All variables are standardized using z-score normalization (Luengo et al., 2020). This step is necessary because K-Means clustering relies on distance-based calculations and is sensitive to differences in variable scales. Standardization ensures that each constraint contributes equally to the clustering process and prevents variables with larger numerical ranges from dominating the results.

2.4 Determination of the Optimal Number of Clusters

The next stage of the methodology focuses on determining the optimal number of clusters (k), as depicted in the flowchart (Onumanyi et al., 2022; SchubertErich, 2023; Sowan et al., 2023). Several internal validation techniques are applied, including the elbow method, silhouette score, Davies-Bouldin index, and Calinski-Harabasz index (Ros et al., 2023). These metrics evaluate cluster compactness, separation, and overall structure. By jointly considering these criteria, a five-cluster solution is selected as the most balanced configuration in terms of statistical performance and interpretability (Aik et al., 2023; Lima & Cruz, 2020).

2.5 K-Means Clustering Procedure

After determining the optimal number of clusters, K-Means clustering is applied to the standardized data (Gratsos et al., 2023; Inkongngarm et al., 2024; Sinaga & Yang, 2020). As indicated in the flowchart, the clustering process begins with the initialization of cluster centroids (Abbas et al., 2020; Nirmal, 2008). The algorithm then proceeds through a K-Means iterative optimization stage, which includes assignment of provinces to the nearest centroid using Euclidean distance, updating centroid positions, and checking for convergence (Ahmed et al., 2019). This iterative process continues until cluster membership stabilizes.

### 2.6 Cluster Profiling and Labeling

The final stage of the methodology involves cluster profiling and labeling. Cluster centroids are transformed back to their original scale to facilitate interpretation. Each cluster is then characterized based on the dominant constraints reflected in its centroid profile. This process results in the identification of distinct regional typologies of e-commerce business constraints in Indonesia, which form the basis for the results and discussion presented in the subsequent section.

---

3. FINDINGS AND DISCUSSION

3.1 Exploratory Data Analysis Results

Figure 2 presents the distribution of major e-commerce business constraints across Indonesian provinces based on exploratory data analysis. The results reveal substantial variation in the intensity and dispersion of constraints, indicating pronounced regional heterogeneity.

Figure 2. Exploratory distribution of e-commerce business constraints by province
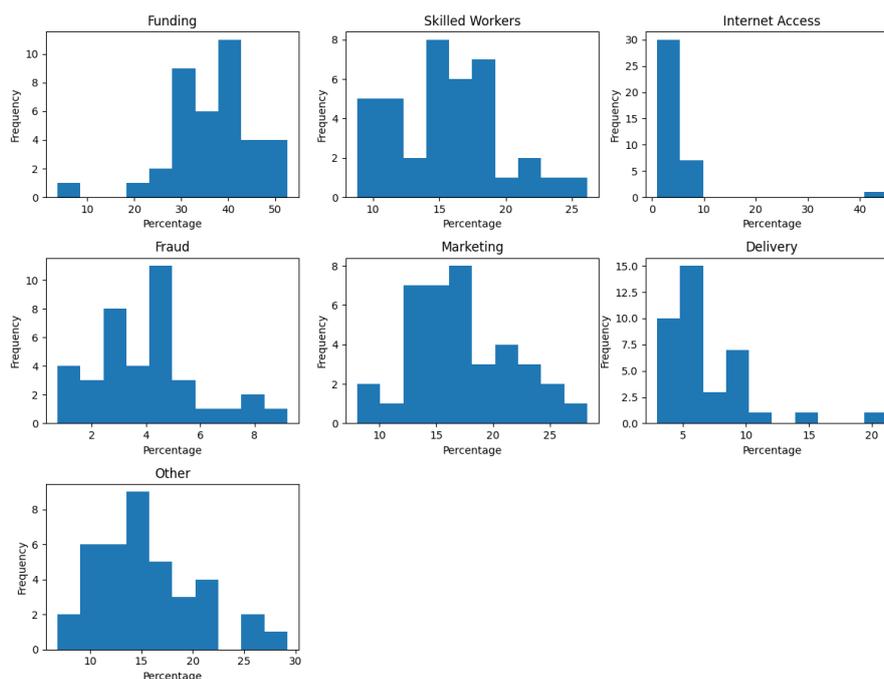
Funding-related constraints dominate across most provinces, with values concentrated at relatively high percentages, confirming that access to capital remains the most prevalent challenge for e-commerce businesses at the national level. Marketing constraints also show a wide distribution, suggesting uneven market access and differences in competitive intensity across regions.

In contrast, internet access constraints exhibit a highly skewed distribution. While most provinces report relatively low values, a small number of provinces display extremely high percentages, indicating severe digital infrastructure limitations. Constraints related to skilled workers, fraud, and delivery services show moderate dispersion, suggesting that operational and human capital challenges vary across regions but are less extreme than digital infrastructure gaps.

3.2 Correlation Analysis of E-Commerce Business Constraints

Figure 3 presents the correlation matrix of major e-commerce business constraints across Indonesian provinces. The results indicate that most variables exhibit low to moderate correlations, suggesting that each constraint captures a distinct dimension of regional challenges.

Figure 3. Correlation matrix of e-commerce business constraints
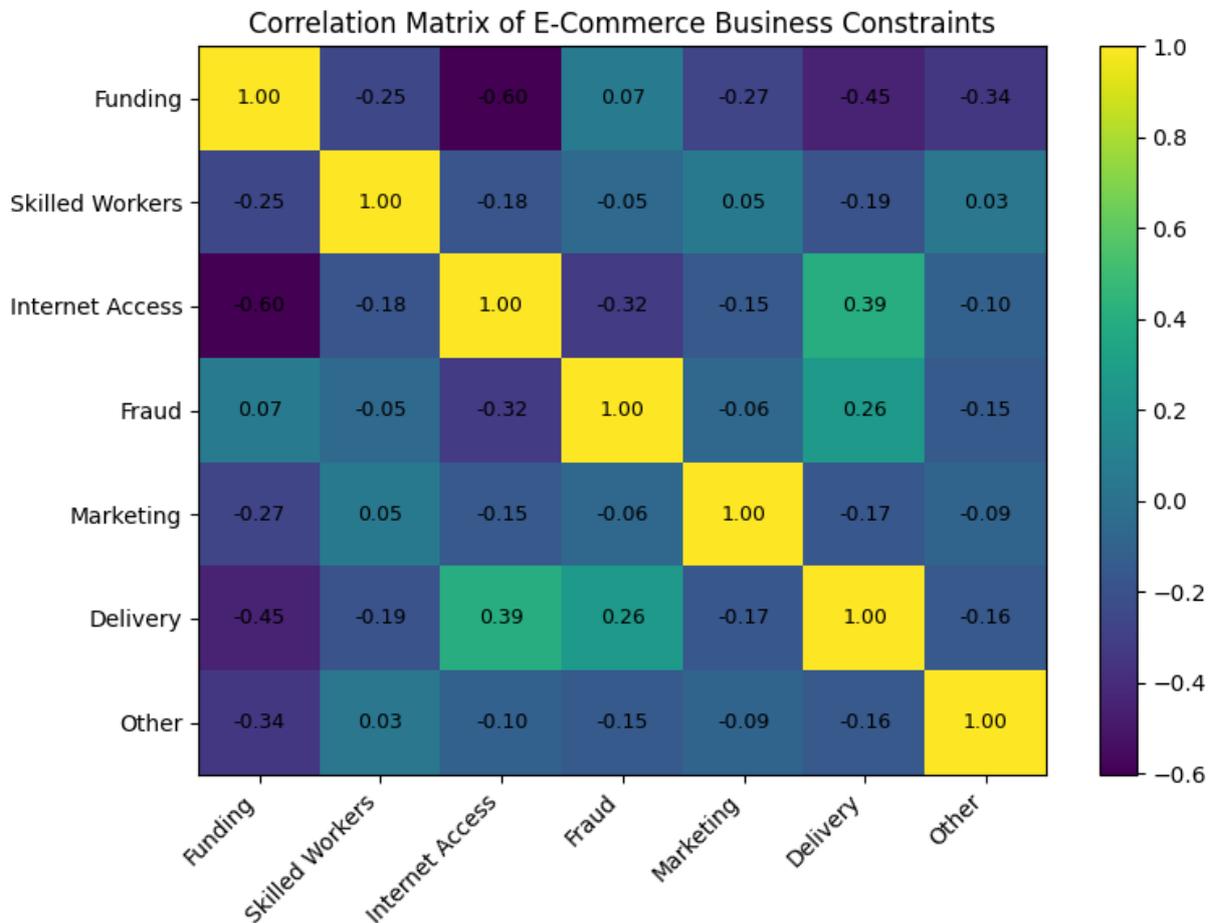
This pattern supports the use of all variables in the clustering analysis without concerns of severe multicollinearity. The strongest relationship is observed between funding constraints and internet access constraints, with a correlation coefficient of −0.60. This negative correlation indicates that provinces experiencing severe limitations in digital infrastructure also tend to face higher funding constraints. This finding suggests a structural linkage between digital access and financial capacity, where limited infrastructure may reduce business viability and access to external financing. A moderate positive correlation is identified between internet access constraints and delivery service limitations (r = 0.39). This relationship implies that provinces with inadequate digital infrastructure are also more likely to experience logistical challenges, reflecting interconnected weaknesses in both digital and physical support systems for e-commerce activities. In addition, delivery service limitations show a positive association with fraud-related constraints (r = 0.26), indicating that logistical inefficiencies may increase transaction risks in certain regions. Most other relationships remain weak. For instance, skilled worker constraints exhibit low correlations with other variables, suggesting that human capital limitations operate relatively independently from financial, infrastructural, and logistical constraints. Similarly, marketing-related constraints display only weak associations with other factors, indicating that market access challenges are not necessarily driven by infrastructure or funding conditions alone.

3.3   Clustering Results and Cluster Validation

To determine the optimal number of clusters, several internal validation metrics are evaluated, including the silhouette score, Davies-Bouldin index, and Calinski-Harabasz index. Table 1 summarizes the clustering performance for different numbers of clusters ranging from three to nine.

Table 1. Internal Cluster Validation Metrics for Different Numbers of Clusters

Number of Clusters (k)	Silhouette Score	Davies-Bouldin Index	Calinski-Harabasz Index
3	0.143	1.387	10.031
4	0.169	1.065	9.859
5*	0.173	0.999	10.826
6	0.135	1.051	9.879
7	0.148	1.087	9.444
8	0.165	0.987	9.973
9	0.129	1.003	9.656

Note:*)Higher Silhouette and Calinski–Harabasz values indicate better cluster separation and structure, while lower Davies–Bouldin values indicate more compact clusters.

The results show that the silhouette score reaches its highest value at 5 clusters (0.173), indicating better separation among clusters than in other configurations. Although the absolute value of the silhouette score remains moderate, this outcome

is common in regional socio-economic data where group boundaries are not sharply defined. The Davies-Bouldin index also supports the five-cluster solution, with a value close to its minimum (0.999), suggesting improved cluster compactness and separation. In contrast, higher values of the Davies-Bouldin index observed at lower cluster numbers, such as 3 clusters (1.387), indicate weaker clustering. Furthermore, the Calinski-Harabasz index attains its maximum value at 5 clusters (10.826), indicating that this configuration provides the strongest overall cluster structure among the tested alternatives. Although a slightly lower Davies-Bouldin index is observed at eight clusters, the corresponding silhouette and Calinski-Harabasz values are lower than those obtained for the five-cluster solution. This suggests that increasing the number of clusters beyond five may lead to over-fragmentation without substantial gains in interpretability. Based on the joint consideration of all three evaluation metrics, a five-cluster solution is selected as the final model. This configuration offers the best balance among cluster separation, compactness, and interpretability. The selected clustering solution forms the basis for the identification and discussion of regional typologies of e-commerce business constraints in Indonesia in the following subsection.

Figure 4. Spatial distribution of provincial e-commerce business constraint clusters in Indonesia identified using K-Means clustering (k = 5).

Powered by Bing
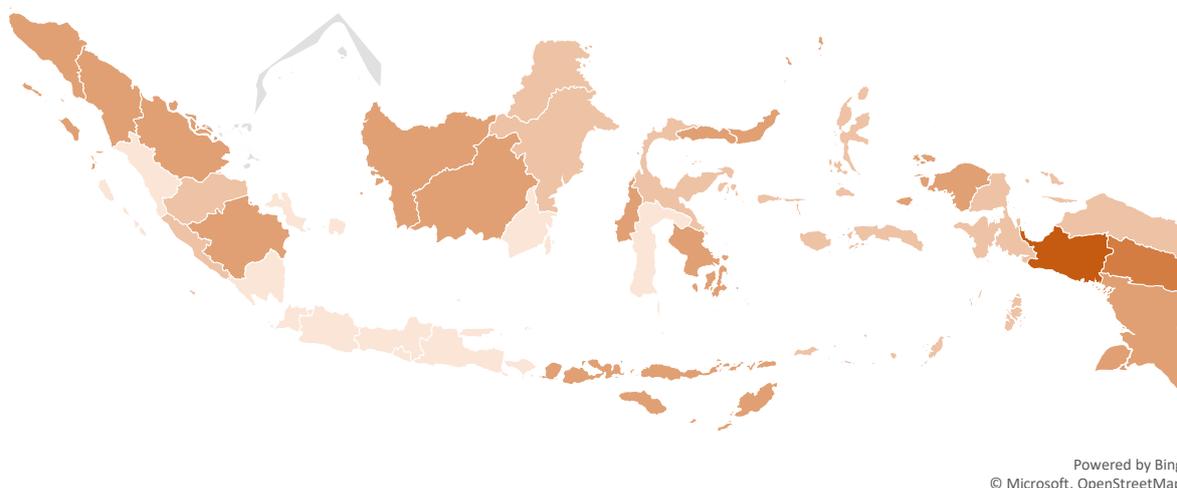© Microsoft, OpenStreetMap

Figure 4 illustrates the spatial distribution of e-commerce business constraint clusters across Indonesian provinces. The mapping reveals clear geographic patterns, indicating that provinces with similar constraint profiles tend to be spatially concentrated. Provinces in Java and Bali are predominantly grouped into clusters characterized by capital and market-related constraints, reflecting relatively advanced digital infrastructure but persistent financial and competitive pressures. In contrast, several provinces outside Java, particularly in eastern Indonesia, are associated with

clusters characterized by logistical challenges and digital infrastructure limitations. Extreme digital constraint clusters are geographically concentrated in Papua regions, highlighting persistent spatial disparities in digital readiness. These spatial patterns reinforce the findings of the clustering analysis and demonstrate that regional differences in e-commerce sustainability are closely linked to geographic and infrastructural contexts.

3.4 Regional Typologies of E-Commerce Business Constraints

Figure 5 presents the centroid profiles of the five identified clusters of e-commerce business constraints across Indonesian provinces, while Table 1 and the spatial mapping illustrate their geographic distribution. Each cluster represents a distinct regional typology characterized by a dominant combination of constraints, rather than differences in constraint intensity alone. The results reveal that these typologies are spatially concentrated and closely aligned with regional economic and infrastructural conditions.

Figure 5. Centroid profiles of five clusters of e-commerce business constraints across Indonesian provinces.

Centroid Profiles of Five Clusters of E-Commerce Business Constraints
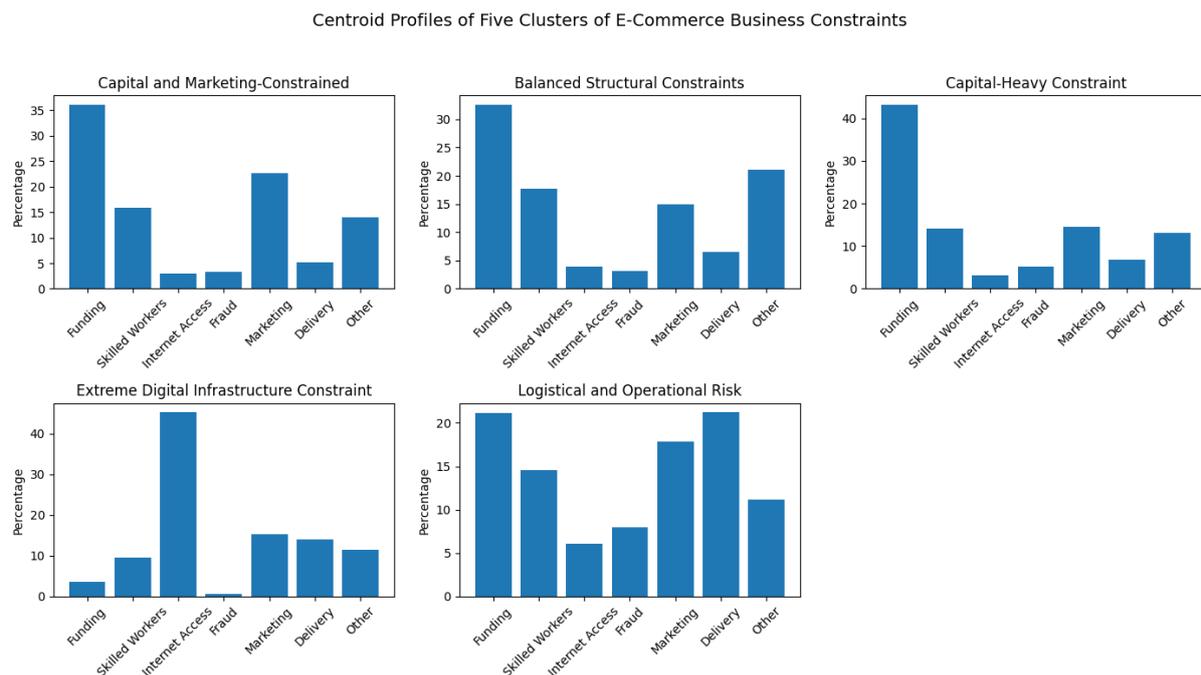
The first cluster (Cluster 0: Capital and Marketing-Constrained) is predominantly concentrated in Java and Bali, as well as several economically active provinces outside Java. Provinces in this cluster include Sumatra Barat, Lampung, Kepulauan Bangka Belitung, DKI Jakarta, Jawa Barat, Jawa Tengah, D.I. Yogyakarta, Jawa Timur, Banten, Bali, Kalimantan Selatan, and Sulawesi Selatan. These regions generally benefit from relatively advanced digital infrastructure and market access. However, intense competition, high marketing costs, and persistent funding constraints limit business expansion and long-term sustainability. In this typology, the

primary challenges are associated with scaling and market penetration rather than basic digital readiness.

The second cluster (Cluster 1: Balanced Structural Constraints) is mainly distributed across parts of Sumatra, Kalimantan, Sulawesi, Maluku, and western Papua. Provinces in this group include Jambi, Bengkulu, Kepulauan Riau, Kalimantan Timur, Kalimantan Utara, Sulawesi Tengah, Maluku, Maluku Utara, Papua Barat, and Papua. This cluster exhibits a relatively even distribution of constraints across funding, skilled labor, marketing, logistics, and other operational factors. The absence of a single dominant obstacle indicates that e-commerce sustainability in these regions is constrained by multiple interrelated structural challenges, requiring coordinated and integrated development strategies.

The third cluster (Cluster 2: Capital-Heavy Constraint) comprises regions where funding constraints are the dominant bottleneck, despite the presence of basic digital infrastructure. This cluster is largely observed in Sumatra, Nusa Tenggara, Kalimantan, and several parts of eastern Indonesia. Provinces classified in this group include Aceh, Sumatra Utara, Riau, Sumatra Selatan, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Sulawesi Utara, Sulawesi Tenggara, Gorontalo, Sulawesi Barat, Papua Barat Daya, and Papua Selatan. In these regions, insufficient financial capacity restricts business continuity and scaling, making access to capital the most critical leverage point for improving e-commerce sustainability.

The fourth cluster (Cluster 3: Extreme Digital Infrastructure Constraint) is geographically concentrated in Papua Pegunungan. This cluster is characterized by exceptionally high internet access constraints, which far exceed other obstacles. In this typology, inadequate digital infrastructure constitutes a fundamental structural barrier to e-commerce activities. Improvements in financing, marketing, or logistics are unlikely to be effective without substantial investments in digital connectivity.

The fifth cluster (Cluster 4: Extreme Digital Infrastructure Constraint – Severe) consists solely of Papua Tengah, which emerges as a distinct outlier with the most severe digital infrastructure limitations among all provinces. This cluster represents the end of the digital divide, where e-commerce sustainability is structurally infeasible under current conditions. The separation of this province into its own cluster highlights the depth of regional disparities in digital readiness across Indonesia.

4. CONCLUSION

This study identifies distinct regional typologies of e-commerce business constraints in Indonesia using a K-Means clustering approach based on provincial data from Statistik E-Commerce 2024. The results show that e-commerce sustainability challenges are unevenly distributed across regions and shaped by different combinations of financial, infrastructural, and operational constraints. Five

clusters are identified, ranging from regions constrained mainly by capital and marketing pressures to regions facing extreme digital infrastructure limitations. These findings indicate that regional disparities in e-commerce development are driven not only by the severity of constraints but also by distinct constraint profiles. Consequently, uniform national strategies are unlikely to be effective. Instead, region-specific approaches are required to support more inclusive and sustainable e-commerce development across Indonesian provinces.

REFERENCES

Abbas, S. A., Aslam, A., Rehman, A. U., Abbasi, W. A., Arif, S., & Kazmi, S. Z. H. (2020). K-Means and K-Medoids: Cluster Analysis on Birth Data Collected in City Muzaffarabad, Kashmir. *IEEE Access*, *8*, 151847–151855. https://doi.org/10.1109/ACCESS.2020.3014021

Ahi, A. A., Sinkovics, N., & Sinkovics, R. R. (2022). E-commerce Policy and the Global Economy: A Path to More Inclusive Development? *Management International Review 2022 63:1*, *63*(1), 27–56. https://doi.org/10.1007/s11575-022-00490-1

Ahmed, S. R. A., Al-Barazanchi, I., Jaaz, Z. A., & Abdulshaheed, H. R. (2019). Clustering algorithms subjected to K-mean and gaussian mixture model on multidimensional data set. *Periodicals of Engineering and Natural Sciences*, *7*(2), 448–457. https://doi.org/10.21533/PEN.V7I2.484

Aik, L. E., Choon, T. W., & Abu, M. S. (2023). K-means Algorithm Based on Flower Pollination Algorithm and Calinski-Harabasz Index. *Journal of Physics: Conference Series*, *2643*(1), 012019. https://doi.org/10.1088/1742-6596/2643/1/012019

Al Azies, H., & Herowati, W. (2023). Unravelling Income Inequality in Indonesia. *Jurnal Riset Ilmu Ekonomi*, *3*(2), 89–100. https://doi.org/10.23969/JRIE.V3I2.63

Asikin, Z. (2024). Diverse E-Commerce Business Models In Indonesia: A Cluster Analysis From The National E-Commerce Survey. *Business Review and Case Studies*, *5*(2), 319–319. https://doi.org/10.17358/brcs.5.2.319

Azies, H. Al, & Rositawati, A. F. D. (2021). Mapping of the Reading Literacy Activity Index in East Java Province, Indonesia: an Unsupervised Learning Approach. *Proceedings of The International Conference on Data Science and Official Statistics*, *2021*(1), 211–223. https://doi.org/10.34123/ICDSOS.V2021I1.128

Bakri, Rochmah, A. A. N., Safitri, E. A., Indriani, K., & Erlina, S. R. A. D. (2024). E-commerce and Market Penetration Strategies in Overcoming Geographical Challenges in Indonesia's Retail Industry. *Journal of Contemporary Administration and Management (ADMAN)*, *2*(2), 539–546. https://doi.org/10.61100/adman.v2i2.197

Baligodugula, V. (2023). Unsupervised-based Distributed Machine Learning for Efficient Data Clustering and Prediction. *Browse All Theses and Dissertations*. https://corescholar.libraries.wright.edu/etd_all/2791

Cooksey, R. W. (2020). Descriptive Statistics for Summarising Data. *Illustrating Statistical Procedures: Finding Meaning in Quantitative Data*, 61–139. https://doi.org/10.1007/978-981-15-2537-7_5

Criveanu, M. M. (2023). Investigating Digital Intensity and E-Commerce as Drivers for Sustainability and Economic Growth in the EU Countries. *Electronics 2023, Vol. 12, 12*(10). https://doi.org/10.3390/electronics12102318

Dadashpoor, H., Malekzadeh, N., & Saeidishirvan, S. (2022). A typology of metropolitan spatial structure: a systematic review. *Environment, Development and Sustainability 2022 25:12*, *25*(12), 13667–13693. https://doi.org/10.1007/s10668-022-02641-8

Edelmann, D., Móri, T. F., & Székely, G. J. (2021). On relationships between the Pearson and the distance correlation coefficients. *Statistics & Probability Letters*, *169*, 108960. https://doi.org/10.1016/J.SPL.2020.108960

Gratsos, K., Ougiaroglou, S., & Margaris, D. (2023). A Web Tool for K-means Clustering. *Lecture Notes in Networks and Systems*, *783 LNNS*, 91–101. https://doi.org/10.1007/978-3-031-44097-7_9

Inkongngarm, A., Bootkrajang, J., Somhom, S., Trongratsameethong, A., & Luekhong, P. (2024). Enhancing Educational Strategy Through K-Means Clustering: A Study on Academic Departments. *Proceedings - 21st International Joint Conference on Computer Science and Software Engineering, JCSSE 2024*, 310–315. https://doi.org/10.1109/JCSSE61278.2024.10613649

Li, Kaiming, Wang, L., Yue, L., & Li, Kaishun. (2026). Spatial Heterogeneity and Gradient Governance of Idle Rural Homesteads in Megacities: Evidence from Shanghai. *Land 2026, Vol. 15, 15*(2), 246. https://doi.org/10.3390/land15020246

Lima, S. P., & Cruz, M. D. (2020). A genetic algorithm using Calinski-Harabasz index for automatic clustering problem. *Revista Brasileira de Computação Aplicada*, *12*(3), 97–106. https://doi.org/10.5335/RBCA.V12I3.11117

Liu, Y., Mu, Y., Chen, K., Li, Y., & Guo, J. (2020). Daily Activity Feature Selection in Smart Homes Based on Pearson Correlation Coefficient. *Neural Processing Letters*, *51*(2), 1771–1787. https://doi.org/10.1007/S11063-019-10185-8/FIGURES/13

Luengo, J., García-Gil, D., Ramírez-Gallego, S., García, S., & Herrera, F. (2020). Big Data Preprocessing: Enabling Smart Data. *Big Data Preprocessing: Enabling Smart Data*, 1–186. https://doi.org/10.1007/978-3-030-39105-8/COVER

Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive Statistics and Normality Tests for Statistical Data. *Annals of Cardiac Anaesthesia*, *22*(1), 67. https://doi.org/10.4103/ACA.ACA_157_18

Monica, M., Ayuningtiyas, N. U., Al Azies, H., Riefky, M., Khusna, H., & Rahayu, S. P. (2021). Unsupervised Learning Approach for Evaluating the Impact of COVID-19 on Economic Growth in Indonesia. *Communications in Computer and Information Science*, *1489 CCIS*, 54–70. https://doi.org/10.1007/978-981-16-7334-4_5/COVER

Nirmal, S. (2008). Comparative Study between K-Means and K-Medoids Clustering Algorithms. *International Research Journal of Engineering and Technology*, *839*. www.irjet.net

Onumanyi, A. J., Molokomme, D. N., Isaac, S. J., & Abu-Mahfouz, A. M. (2022). AutoElbow: An Automatic Elbow Detection Method for Estimating the Number of Clusters in a Dataset. *Applied Sciences 2022, Vol. 12, Page 7515*, *12*(15), 7515. https://doi.org/10.3390/APP12157515

Ros, F., Riad, R., & Guillaume, S. (2023). PDBI: A partitioning Davies-Bouldin index for clustering evaluation. *Neurocomputing*, *528*, 178–199. https://doi.org/10.1016/J.NEUCOM.2023.01.043

SchubertErich. (2023). Stop using the elbow criterion for k-means and how to choose the number of clusters instead. *ACM SIGKDD Explorations Newsletter*, *25*(1), 36–42. https://doi.org/10.1145/3606274.3606278

Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, *8*, 80716–80727. https://doi.org/10.1109/ACCESS.2020.2988796

Sosyal Araştırmalar, A. (2025). Sustainable e-Commerce: Transformation in Environmental, Economic, and Social Dimensions. *Akademic Social Studies*, *9*(31), 261–290. https://doi.org/10.31455/asya.1622024

Sowan, B., Hong, T. P., Al-Qerem, A., Alauthman, M., & Matar, N. (2023). Ensembling validation indices to estimate the optimal number of clusters. *Applied Intelligence*, *53*(9), 9933–9957. https://doi.org/10.1007/S10489-022-03939-W/FIGURES/11

Wijaya, D. R., Paramita, N. L. P. S. P., Uluwiyah, A., Rheza, M., Zahara, A., & Puspita, D. R. (2020). Estimating city-level poverty rate based on e-commerce data with machine learning. *Electronic Commerce Research 2020 22:1*, *22*(1), 195–221. https://doi.org/10.1007/s10660-020-09424-1

Xu, G., Zhao, T., & Wang, R. (2022). Research on the Efficiency Measurement and Spatial Spillover Effect of China's Regional E-Commerce Poverty Alleviation from the Perspective of Sustainable Development. *Sustainability 2022, Vol. 14, 14*(14). https://doi.org/10.3390/su14148456