Indonesian Journal of Advanced Research (IJAR)
Vol. 4, No. 9, 2025: 1991-2002

Segmentation Analysis of Countries Based on Human Development Index and Artificial Intelligence Readiness Using Unsupervised Learning Methods: Principal Component Analysis and K-Means Clustering

Alexander J.P. Sibarani[1]*, Preddy Marpaung[2]
Universitas Mahkota Tricom Unggul
Corresponding Author: Alexander J.P. Sibarani
alexanderjpsibarani@mtu.ac.id

A R T I C L E I N F O

Keywords: Human Development Index, Artificial Intelligence Readiness, PCA, K-Means Clustering, Country Segmentation

Received : 7, August
Revised : 21, August
Accepted: 23, September

A B S T R A C T

This study explores the segmentation of over 120 countries based on indicators of human development and digital readiness using unsupervised learning methods. By applying Principal Component Analysis (PCA) for dimensionality reduction and K-Means Clustering for grouping, countries were categorized into four clusters reflecting different levels of development and AI preparedness. Indonesia was positioned within the developing cluster, highlighting both its digital growth potential and structural challenges in innovation and R&D. The analysis provides strategic insights for policy formulation and emphasizes the importance of digital inclusion, research investment, and international collaboration. The clustering approach offers an intuitive visual framework to understand global patterns in the era of technological transformation.

INTRODUCTION

Human development is a fundamental pillar of a nation's progress. Beyond macroeconomic indicators, it reflects the quality of life through access to education, healthcare, and decent income. As development paradigms evolve, international organizations such as the United Nations Development Programme (UNDP) have established the Human Development Index (HDI) as a quantitative tool to measure a country's achievements in improving human well-being. The HDI has been widely used as a reference in public policy formulation, budget allocation, and as a comparative framework for evaluating progress among nations (UNDP (United Nations Development Programme), 2024), (Roy, 2025).

In the era of the Fourth Industrial Revolution (Industry 4.0), the challenges of human development are becoming increasingly complex. Digital transformation, automation, and Artificial Intelligence (AI) are not only reshaping industrial landscapes but also significantly impacting employment structures, educational systems, and public services. Countries that are able to adapt and implement inclusive digital policies tend to demonstrate better human development outcomes. As a result, indicators such as the AI Readiness Index, internet accessibility, and national innovation capacity have become essential complements in assessing a country's preparedness for the digital era (Hankins et al., 2023) (World Intellectual Property Organization (WIPO), 2023) (Nasution et al., 2024).

Indonesia, as a developing country and the fourth most populous nation in the world, faces both vast opportunities and significant challenges. The Indonesian government, through its Medium-Term National Development Plan (RPJMN) 2020–2024, has prioritized the development of high-quality human capital, digital transformation, and innovation enhancement. National programs such as *Indonesia Emas 2045*, *100 Smart Cities*, and the National AI Strategy (Badan Pengkajian dan Penerapan Teknologi (BPPT), 2020) reflect Indonesia's commitment to a technology-driven future. Nonetheless, digital inequality, limited infrastructure, and low investment in Research and Development (R&D) remain major obstacles to achieving this vision.

Cross-country comparative studies are crucial for assessing Indonesia's position in the global landscape. By comparing HDI and digital readiness indicators across countries, Indonesia can identify clusters of nations with similar characteristics and formulate more contextual and data-driven development strategies.

One effective method for such segmentation is K-Means Clustering, an unsupervised learning algorithm commonly used for partitioning observations into distinct clusters based on similarity. Recent studies have applied this technique to analyze multidimensional indicators in global development contexts (Saraiva & Caiado, 2025) (Naeem et al., 2023). To complement clustering, Principal Component Analysis (PCA) is used to reduce dimensionality while retaining the essential structure of the data, enabling effective visualization of country groupings (Jolliffe & Cadima, 2016).

This study aims to visualize and segment countries around the world based on human development and AI readiness indicators. Through a data mining and multidimensional visualization approach, it seeks not only to map Indonesia's position globally but also to provide strategic insights for policymakers and stakeholders in strengthening an inclusive and sustainable digital transformation.

Thus, the findings of this research are expected to contribute meaningfully to academic literature, while also serving as evidence-based input for governments and related institutions in designing future strategies for human development and technology in the digital age.

LITERATURE REVIEW

The Human Development Index (HDI) is a composite indicator developed by the United Nations Development Programme (UNDP) to measure long-term achievements in three fundamental dimensions of human development: a long and healthy life (life expectancy), knowledge (education index), and a decent standard of living (gross national income per capita) (UNDP (United Nations Development Programme), 2024). HDI serves as an essential tool to evaluate a country's development performance not only from an economic standpoint but also in terms of the overall quality of life of its population. HDI rankings have been used in numerous cross-country studies to identify global disparities in social development. Countries with high HDI scores typically enjoy broad access to healthcare services, quality education, and strong social systems. In contrast, countries with lower HDI values continue to face structural limitations that hinder both social and economic advancement.

As artificial intelligence (AI) technologies continue to evolve, many countries are adopting AI in their public policies and services. To measure a country's readiness to implement such technologies, Oxford Insights developed the Government AI Readiness Index, which evaluates the government's capacity to absorb, utilize, and manage AI in public service delivery (Hankins et al., 2023). This index considers dimensions such as national AI strategy and vision, digital ecosystem, data infrastructure, and digital human resource competencies. Countries with high AI readiness scores often have clear national strategies, significant R&D investments, and active innovation ecosystems (Calvino & Fontanelli, 2023). On the other hand, developing countries like Indonesia still face challenges such as limited AI talent, regional disparities in digital infrastructure, and low levels of technological investment from both the public and private sectors. Internet access plays a foundational role in the digital ecosystem. According to World Bank data, internet penetration is highly correlated with economic productivity and the efficiency of public services (World Bank, 2024). Additionally, the Global Innovation Index (GII) published by WIPO assesses a country's ability to generate and implement innovations using indicators such as the number of scientific publications, patents, R&D expenditures, and international collaborations (World Intellectual Property Organization (WIPO), 2023). These two indicators are important complements for understanding a country's digital readiness.

Cluster analysis is a statistical method used to group objects based on the similarity of specific characteristics. One of the most widely used techniques in data mining is K-Means Clustering, which organizes data into *k* clusters by minimizing the Euclidean distance between data points and their respective centroids (MacQueen, 1967). This method has been extensively applied in various comparative studies to map clusters of countries, regions, or organizations based on social, economic, or technological indicators (Tan et al., 2019). However, one challenge in applying K-Means is determining the optimal number of clusters. To address this, techniques such as the Elbow Method or Silhouette Score are often used to evaluate the quality of clustering results. Before applying clustering—especially in multidimensional datasets—dimension reduction techniques such as Principal Component Analysis (PCA) are often necessary. PCA is a statistical technique used to reduce the number of variables while retaining the maximum variance within the data (Jolliffe & Cadima, 2016). PCA facilitates the visualization of clustering results in two or three dimensions, making patterns in the data easier to interpret. Previous studies, such as that conducted by Saraiva & Caiado (2025), have successfully used a combination of PCA and K-Means to classify countries based on economic and social indicators, demonstrating the effectiveness of this approach in a global context.

METHODOLOGY

This research adopts a quantitative approach with exploratory visual analysis to identify and cluster countries based on human development and digital readiness indicators. The data processing integrates unsupervised machine learning techniques, specifically K-Means Clustering and Principal Component Analysis (PCA) for dimensionality reduction. This approach is chosen for its capability to group objects (in this case, countries) into clusters based on multivariate similarity without requiring labelled or target variables. By visualizing the results of PCA, global patterns in multidimensional data can be presented more intuitively.

The dataset is compiled from multiple credible international sources as follows:

Table 1. Dataset Source

Indicator	Source
Human Development Index (HDI)	UNDP
Internet Access (% of population)	World Bank
AI Readiness Index	Oxford Insights
Gross National Income (GNI) per capita	UNDP / World Bank
R&D Expenditure (% of GDP)	World Bank / WIPO
Global Innovation Index	WIPO

The dataset comprises data from 120 countries and spans the most recent available years between 2022 and 2024, depending on the latest published data for each indicator. To ensure cross-country comparability, priority was given to globally recognized and standardized sources. The dataset was compiled manually from multiple trusted sources and includes the following six variables:

1. HDI (X1): Human Development Index, a composite index ranging from 0 to 1 that measures average achievement in key dimensions of human development: a long and healthy life, being knowledgeable, and having a decent standard of living. Source from United Nations Development Programme (UNDP) Human Development Report 2023/2024.
2. Internet Access (X2): The percentage of the population with access to the Internet. Source from International Telecommunication Union (ITU) and World Bank Open Data (latest 2022–2023 data).
3. AI Readiness (X3): The country’s readiness for AI adoption and governance, with a score between 0 and 100. Source from Oxford Insights – *Government AI Readiness Index 2023*.
4. GNI per capita (X4): Gross National Income per capita (in USD), reflecting the average income of a country's citizens. Source from World Bank and UNDP (latest data from 2022 or 2023 depending on availability).
5. R&D Expenditure (X5): Gross domestic expenditure on research and development, expressed as a percentage of GDP. Source from UNESCO Institute for Statistics and World Bank (latest data between 2020 and 2023, depending on reporting country).
6. Global Innovation Index (X6): A composite indicator that reflects a country's innovation performance, scored between 0 and 100. Source from World Intellectual Property Organization (WIPO) – *Global Innovation Index 2023*.

All variables are numeric and will be normalized to ensure equal scaling during the clustering process.

Due to data unavailability for some countries (especially low-income or small island nations), several values were estimated to maintain sample completeness for unsupervised learning. For missing values, estimations were made using regional averages, income group averages, or neighboring country proxies. In rare cases, linear extrapolation from previous years or correlated indicators (e.g., estimating R&D from GII or GNI level) was used. These estimations were made conservatively to prevent distortion of clustering results, ensuring the general distributional integrity of the dataset while maintaining sufficient sample size (n = 120).

Data Normalization

Numerical data is normalized using the Min-Max Scaling method with the formula:

$$x' = \frac{X - Xmin}{Xmax - Xmin}$$

This ensures that variables with larger scales, such as GNI, do not dominate the clustering process.

After normalization, PCA is applied to reduce dimensionality. PCA serves to: simplify data visualization and identify the principal components that contribute most to variance among countries. PCA does not eliminate original data but transforms it into a linear combination of the original variables into fewer principal components.

The K-Means algorithm is used to group countries into k clusters, involving: initialization of k random centroids, calculation of Euclidean distance from each data point to each centroid, assignment of data points to the nearest centroid, recalculation of new centroids based on the average of each cluster, and iterative refinement until convergence.

Euclidean Distance Formula:

$$d(x,c) = \sqrt{\sum_{i=1}^{n}(xi - ci)^2}$$

The optimal number of clusters is determined using the Elbow Method, which involves plotting inertia values for different k values. Inertia measures the total within-cluster squared distance. The "elbow point" on the graph indicates the optimal number of clusters.

The data processing is conducted using Python 3.10 with Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, and plotly. Additional visualization using Microsoft Excel.

To assess clustering quality, the Silhouette Score metric is applied. This measures how well an object fits within its assigned cluster compared to other clusters. A silhouette scores close to 1 indicates strong clustering. Summary of Workflow:

1. Collect and clean data from official sources
2. Normalize variables
3. Apply PCA for dimensionality reduction
4. Determine optimal number of clusters (k)
5. Perform K-Means clustering
6. Visualize results
7. Analyze Indonesia's position and other clusters

Through this methodology, the research aims to produce valid, interpretable, and policy-relevant segmentation of countries—especially Indonesia—in mapping their global position within the context of digital transformation and human development.

RESEARCH RESULT

This study identified four main clusters from over 120 countries based on six indicators: Human Development Index (HDI), Internet Access, AI Readiness, Gross National Income (GNI) per capita, R&D expenditure, and the Global Innovation Index. The number of clusters was determined using the Elbow Method, where the inertia value showed stabilization at k = 4.

The following table summarizes the average scores of each indicator within the four clusters:

Table 2. Average Scores

Indicator	Cluster 0 (Least developed)	Cluster 1 (Mid-range developing countries)	Cluster 2 (Fast-developing economies)	Cluster 3 (Highly developed countries)
Number of Countries	30	30	30	30
HDI (0–1)	0.585	0.727	0.876	0.940
Internet Access (%)	31.4%	67.0%	90.3%	95.0%
AI Readiness (0–100)	27.2	44.6	70.1	86.2
GNI per Capita (USD)	1485	3902	37157	66943
R&D Expenditure (% of GDP)	0.145%	0.328%	1.003%	2.613%
Global Innovation Index	22.8	32.1	42.9	58.5

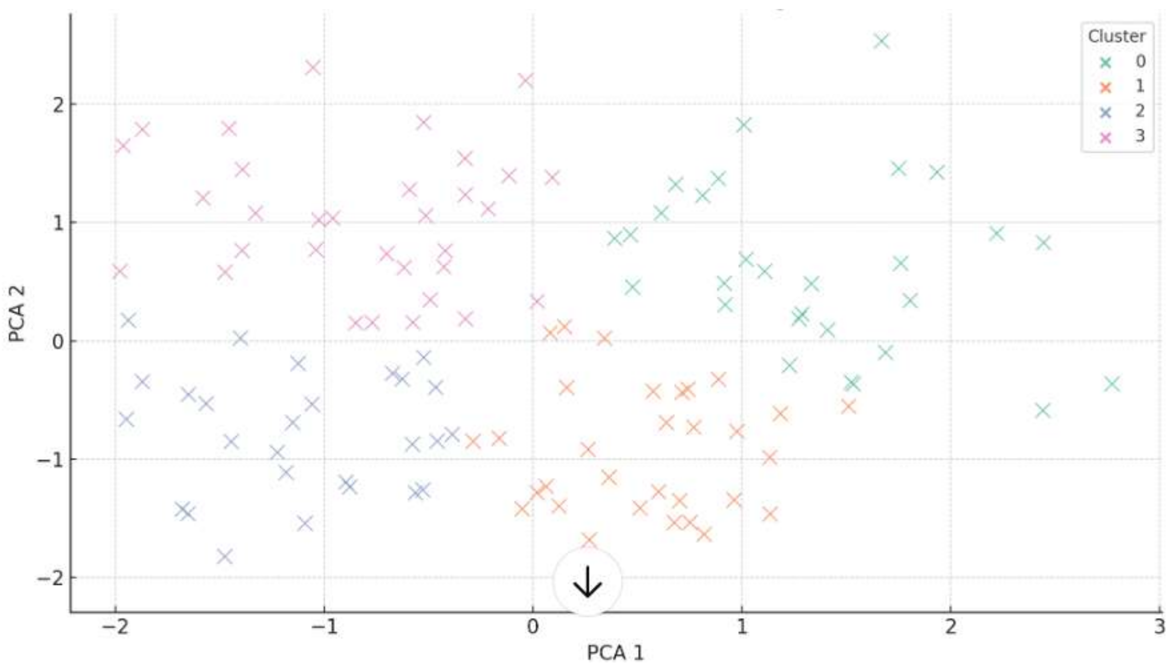

Figure 1. PCA Visualization

List of Countries in Each Cluster:

Table 3. Countries in Each Cluster

Cluster	Countries
Cluster 0	Afghanistan, Angola, Bangladesh, Benin, Burkina Faso, Cambodia, Cameroon, Chad, Congo (Dem. Rep.), Côte d'Ivoire, Ethiopia, Gambia, Guinea, Haiti, Lao PDR, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mozambique, Myanmar, Niger, Rwanda, Senegal, Sierra Leone, Sudan, Tanzania, Togo
Cluster 1	Argentina, Brazil, Colombia, Costa Rica, Dominican Republic, Ecuador, Egypt, El Salvador, Fiji, Ghana, Guatemala, Honduras, India, Indonesia, Jamaica, Jordan, Kenya, Lebanon, Morocco, Namibia, Nepal, Nigeria, Pakistan, Paraguay, Peru, Philippines, Sri Lanka, Tunisia, Uzbekistan, Vietnam
Cluster 2	Albania, Armenia, Belarus, Bosnia and Herzegovina, Bulgaria, Croatia, Georgia, Greece, Hungary, Kazakhstan, Kyrgyzstan, Malaysia, Maldives, Mauritius, Mexico, Moldova, Mongolia, Montenegro, North Macedonia, Panama, Qatar, Romania, Russia, Saudi Arabia, Serbia, Slovakia, South Africa, Suriname, Thailand, Turkey
Cluster 3	Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Iceland, Ireland, Israel,

Cluster	Countries
	Italy, Japan, Netherlands, New Zealand, Norway, Singapore, Slovenia, South Korea, Spain, Sweden, Switzerland, United Arab Emirates, United Kingdom, United States, Czech Republic, Estonia, Latvia, Lithuania, Poland

DISCUSSION

Cluster 0 consists of nations with low HDI and low internet penetration. Their GDP per capita is low, and literacy levels tend to be below global averages. These countries are in the early stage of digital transformation and cannot support AI or innovation ecosystems meaningfully. Most nations here face structural issues such as conflict, poor governance, or chronic underinvestment. International support and deep structural reform are necessary to improve both human development and digital readiness.

Cluster 1 includes developing countries that are undergoing significant transformation. While their HDI is moderately high, these nations are still catching up in terms of GDP per capita, R&D investment, and internet accessibility. They often display growing digital economies and some early integration of AI and innovation policies. Countries in this cluster, such as Indonesia, have strong demographic momentum and an expanding digital consumer base. However, structural bottlenecks in infrastructure, education, and institutional capacity limit their full potential.

Cluster 2 comprises nations that are still in the early or middle stages of socio-economic and technological development. These countries show moderate HDI, low GDP per capita, and inconsistent digital access. Their internet and urbanization levels are improving, but slowly, and they face resource limitations that restrict investment in innovation and education. Countries in this cluster need targeted investment in public services, education, and digital infrastructure to move upward.

Cluster 3 represents countries that have achieved an optimal combination of high Human Development Index (HDI), high GDP per capita, and advanced digital and innovation infrastructure. These nations have mature governance systems, widespread digital literacy, and exceptional internet accessibility. They have completed most stages of industrial and digital transition, and are now global front-runners in digital policy, AI governance, and innovation ecosystems. Countries in this cluster typically invest significantly in education, technology, and research, which fuels both economic and social progress. Their high median ages reflect stable population structures and well-developed health and education systems.

Indonesia's position within Cluster 1 reflects its status as a developing country with both significant potential and critical challenges. Based on the available data, Indonesia records an HDI score of approximately 0.73, placing it in the upper-middle category. Internet access is around 65%, but remains uneven across regions. Its AI Readiness Index stands at about 52—still lagging behind neighboring countries such as Malaysia and Thailand. The country's Gross National Income (GNI) per capita is around USD 4,800, while R&D expenditure constitutes only about 0.24% of GDP. The Global Innovation Index (GII) for Indonesia is estimated between 35 and 40.

Compared to other ASEAN countries, Malaysia and Thailand perform better in terms of AI readiness and R&D spending. Vietnam, while close to Indonesia in terms of innovation index, demonstrates stronger initiatives in digital education and MSME (micro, small, and medium enterprise) digital transformation strategies.

Indonesia's placement in the middle cluster highlights a dual challenge. On one hand, it has considerable demographic potential and momentum in digital growth. On the other, it lacks substantial acceleration in innovation ecosystems and AI policy development. These findings can inform the formulation of a national roadmap to strengthen AI readiness, the expansion of R&D investment through fiscal incentives, and the integration of AI and innovation literacy into the national education curriculum.

CONCLUSIONS AND RECOMMENDATIONS

This study aimed to visualize and segment countries around the world based on indicators of human development and their readiness to face the era of artificial intelligence (AI). By applying unsupervised learning methods, particularly K-Means Clustering preceded by Principal Component Analysis (PCA), more than 120 countries were classified into four main clusters based on six key indicators: the Human Development Index (HDI), internet access, AI readiness, gross national income (GNI) per capita, R&D expenditure, and the Global Innovation Index.

The results reveal that countries can be grouped into four broad clusters. Cluster 3 comprises advanced economies that exhibit consistently high scores across all measured variables, reflecting robust digital infrastructure, strong human development, and a mature innovation ecosystem. Cluster 2 includes upper-middle-income countries that are undergoing a transitional phase of development; while they demonstrate relatively high HDI and digital access, their investment in AI readiness and R&D remains limited. Cluster 1 consists of nations with moderate performance, often constrained by structural challenges such as low innovation input, limited research investment, and uneven access to technological infrastructure. Lastly, Cluster 0 includes underdeveloped countries that face significant deficits across most indicators, including low HDI, minimal internet penetration, and negligible investment in research and innovation. These findings provide a basis for understanding global disparities in technological readiness and highlight the differentiated policy approaches required to foster inclusive digital development. Indonesia is positioned within the second cluster, alongside countries such as Brazil, India, and Vietnam. This classification indicates that while Indonesia has achieved a relatively good HDI and shows promising digital growth, it has yet to match the level of innovation infrastructure, research investment, and AI readiness demonstrated by more advanced nations.

The PCA-based visualization clearly distinguishes the clusters, though some countries appear near the boundaries between clusters. This reflects the dynamic and non-linear nature of digital transformation processes across nations.

The clustering and data visualization approach employed in this study has proven effective in providing a comprehensive and intuitive understanding of a country's strategic position in the global context of human development and technological transformation.

ADVANCED RESEARCH

Based on the findings of this study, several recommendations can be made, both from academic and practical perspectives. For the Government of Indonesia, it is crucial to strengthen the national AI strategy through a holistic and data-driven approach. This should go beyond mere technological adoption and include the development of human capital and an inclusive innovation ecosystem. Increasing the national budget allocation for research and development—across both public and private sectors—should be encouraged through fiscal incentives and collaborative funding schemes. Expanding internet access and improving digital literacy, especially in frontier, outermost, and underdeveloped regions, should be prioritized as a foundation for enhancing national AI readiness. Furthermore, Indonesia should actively participate in regional and international collaborations focused on inclusive and ethical AI development to ensure it is not left behind in the global digital transformation.

For researchers and academics, future studies should explore the application of spatiotemporal clustering models to monitor dynamic changes in human development and digital readiness over time. Integrating qualitative data—such as policy strategies, regulatory maturity, and organizational culture related to technology use—would enrich the analytical depth and policy relevance of such models. Moreover, the inclusion of primary data sources and up-to-date statistical inputs is highly recommended, as these can significantly enhance the accuracy, contextual relevance, and robustness of clustering outcomes. Utilizing real-time or recently published datasets will not only ensure better reflection of current development trajectories but also improve the responsiveness of policy recommendations derived from such analyses.

In terms of national technology development, promoting AI adoption beyond large-scale industries is essential. Efforts should target micro, small, and medium enterprises (MSMEs), the education sector, and public services. Interdisciplinary research that combines information technology, public policy, development economics, and social sciences is needed to shape a national AI strategy that is both effective and sustainable.

REFERENCES

Badan Pengkajian dan Penerapan Teknologi (BPPT). (2020). *Strategi Nasional Kecerdasan Artifisial Indonesia 2020-2045*.

Calvino, F., & Fontanelli, L. (2023). *A portrait of AI adopters across countries: Firm characteristics, assets' complementarities and productivity* (OECD Science, Technology and Industry Working Papers). https://doi.org/10.1787/0fb79bb9-en

Hankins, E., Nettel, P. F., Martinescu, L., Grau, G., & Rahim, S. (2023). *Government AI Readiness Index 2023*.

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

MacQueen, J. (1967). *Some methods for classification and analysis of multivariate observations*. https://api.semanticscholar.org/CorpusID:6278891

Naeem, S., Ali, A., Anam, S., & Ahmed, M. M. (2023). An Unsupervised Machine Learning Algorithms: Comprehensive Review. *International Journal of Computing and Digital Systems*, *13*(1), 911–921. https://doi.org/10.12785/ijcds/130172

Nasution, M. K. M., Elveny, M., Pamucar, D., Popovic, M., & Andrić Gušavac, B. (2024). Uncovering the Hidden Insights of the Government AI Readiness Index: Application of Fuzzy LMAW and Schweizer-Sklar Weighted Framework. *Decision Making: Applications in Management and Engineering*, *7*(2), 443–468. https://doi.org/10.31181/dmame7220241221

Roy, A. (2025). Global Research on Human Development Index: Bibliometric Analysis and Policy Insights. In *A Basic Handbook of Science, Technology and Innovation for Inclusive Development [Volume: 2]* (pp. 1–20). International Academic Publishing House (IAPH). https://doi.org/10.52756/bhstiid.2025.e02.001

Saraiva, C., & Caiado, J. (2025). Global development patterns: A clustering analysis of economic, social and environmental indicators. *Sustainable Futures*, *10*, 100907. https://doi.org/https://doi.org/10.1016/j.sftr.2025.100907

Tan, P.-N., Karpatne, A., Steinbach, M., & Kumar, V. (2019). *Introduction to Data Mining Global Edition*. Pearson Deutschland. https://elibrary.pearson.de/book/99.150005/9780273775324

UNDP (United Nations Development Programme). (2024). *Human Development Report 2023-24*. http://report2023-24.hdr.undp.org.s3-website-us-east-1.amazonaws.com/

World Bank. (2024). *Digital Progress and Trends Report 2023*.

World Intellectual Property Organization (WIPO). (2023). *Global Innovation Index 2023 : Executive Summary*.