DOI: https://doi.org/10.70656/ijcse.v3i01.910

MSME Data Integration and Visualization Dashboard for Commodity Distribution Analysis Using ETL Approach (Case Study: Bank Indonesia Representative Office of Bengkulu Province)

Annisa Rizky Maulida
Information System Study Program
Department of Engineering
University of Bengkulu
Bengkulu, Indonesia
annisarizkymaulida14@gmail.com

Tiara Eka Putri
Information System Study Program
Department of Engineering
University of Bengkulu
Bengkulu, Indonesia
tiaraekaputri@unib.ac.id

Dyah Retnaningrum
Assistant Manager for MSME
Development Implementation Function
Bank Indonesia Representative Office
of Bengkulu Province
Bengkulu, Indonesia
dyah_r@bi.go.id

Abstract*— Bank Indonesia Representative Office of Bengkulu Province manages MSME data collected from multiple heterogeneous sources, including internal databases, G-Form e-catalogue responses, and internal systems. However, this data remains unintegrated and stored in fragmented formats, hindering effective monitoring and data-driven decision-making. The lack of a structured integration pipeline and visualization tool further limits the unit's ability to analyze MSME commodity distribution across Bengkulu Province. This study aims to (1) integrate MSME data from three heterogeneous sources into a unified dataset, (2) perform data transformation and cleaning from 82 raw records into 53 valid MSME entries, and (3) develop an interactive Power BI dashboard for analyzing commodity distribution, regional spread, and featured products. The ETL (Extract, Transform, Load) approach was implemented using Python and the Pandas library in Google Colab for data extraction and transformation, while Microsoft Power BI was used for data loading and interactive visualization, covering 53 MSMEs across 9 districts/cities in Bengkulu Province. The integration process successfully unified data from three sources, reducing 82 raw records to 53 validated entries through duplicate removal, incomplete data filtering, and irrelevant record elimination. The dashboard reveals that 66% of MSMEs are concentrated in the food and beverage sector, 54.7% are located in Kota Bengkulu, and coffee processing dominates as the leading commodity. This study contributes a replicable ETL-based framework for multi-source MSME data integration and visualization to support Bank Indonesia Bengkulu in more efficient, data-driven MSME development and policy decisions.

CCS CONCEPTS • Information Systems • Data Integration • Data Visualization

Keywords- data integration, data visualization, ETL, MSME, power BI

I. INTRODUCTION

Micro, Small, and Medium Enterprises (MSMEs) play a critical role in Indonesia's national economy, contributing more than 60% of the Gross Domestic Product (GDP) and absorbing approximately 97% of the total workforce [1]. As one of the institutions responsible for MSME development, Bank Indonesia Representative Office of Bengkulu Province actively monitors and develops MSME partners through its MSME Development and Implementation Function (FPPU). However, the management of MSME data at this unit faces a significant challenge: data is collected from multiple heterogeneous sources, including internal Excel databases, G-Form e-catalogue responses in CSV format, and internal systems that require manual extraction, resulting in fragmented and unintegrated data that is difficult to monitor comprehensively. This condition hinders the unit's ability to conduct effective analysis and make data-driven decisions regarding MSME development across Bengkulu Province.

To address this challenge, this study proposes the implementation of an ETL (Extract, Transform, Load) approach to integrate MSME data from the three heterogeneous sources into a unified, structured dataset, followed by the development of an interactive visualization dashboard using Microsoft Power BI. The ETL process was carried out using Python and the Pandas library in Google Colab, enabling systematic data extraction, transformation, and cleaning from 82 raw records into 53 validated MSME entries. The resulting dashboard provides interactive visualizations of commodity distribution, regional spread, and featured products of MSME partners, offering a comprehensive and accessible overview of MSME conditions across Bengkulu Province.

Data integration through the ETL approach has been widely recognized as a fundamental process in transforming fragmented, multi-source data into actionable insights. Khan et al. [2] emphasized that the ETL process is crucial in data warehousing, as it bridges the gap between unprocessed data and useful insights by addressing challenges including data quality, volume, and source heterogeneity. Similarly, Razali et al. [3] demonstrated that implementing quality scoring techniques in ETL processes significantly improves data reliability in analytics contexts. On the visualization side, the use of business intelligence tools such as Power BI has proven effective in supporting decision-making processes across various sectors. Campante et al. [4] showed that integrated dashboard visualizations built on Power BI enable quick, concise, and easy-to-interpret graphical analyses that directly support organizational decision-making. Furthermore, a study on agricultural foreign trade by Reyes et al. [5] demonstrated that Power BI dashboards successfully bridge the gap between

complex datasets and non-technical decision-makers, making data more accessible for strategic purposes.

Based on the background described above, this study has three main objectives: (1) to integrate MSME data from three heterogeneous internal sources at Bank Indonesia Bengkulu into a unified dataset using the ETL approach, (2) to perform data transformation and cleaning processes resulting in 53 validated MSME records from 82 initial raw data entries, and (3) to develop an interactive Power BI dashboard that visualizes commodity distribution, regional spread, and featured products of MSME partners to support data-driven decision-making at Bank Indonesia Representative Office of Bengkulu Province.

A. Data Visualization Using Power BI

Data visualization is the graphical representation of information and data using visual elements such as charts, graphs, and maps, enabling users to identify trends, patterns, and outliers in data more effectively than through traditional tabular formats [4]. In the context of business intelligence, Microsoft Power BI has emerged as one of the most widely adopted visualization platforms due to its scalability, cost-effectiveness, and ability to connect to diverse data sources. While other BI tools such as Tableau and Looker Studio also offer powerful visualization capabilities, Power BI distinguishes itself through its seamless integration with Microsoft Office ecosystems, lower licensing cost for organizational use, and built-in Power Query feature for multi-source data transformation making it particularly suitable for institutional settings such as Bank Indonesia Bengkulu. Campante et al. [4] demonstrated that Power BI enables the implementation of key performance indicators (KPIs) and the creation of concise, easy-to-interpret dashboards that directly support organizational decision-making processes. The platform's capability to handle multi-source data integration through its built-in Power Query feature further strengthens its suitability for this study.

Power BI has been applied across various domains to support data-driven decision-making. In the agricultural sector, Reyes et al. [5] successfully developed a Power BI dashboard to analyze export and import trade data, demonstrating that the platform can effectively bridge complex datasets and non-technical users. In the context of SME data management, Arifin et al. [6] highlighted that effective data visualization through business intelligence tools can consolidate disparate data points, enabling organizations to assess operational statuses more comprehensively and improve competitiveness. These findings support the use of Power BI as the primary visualization tool in this study for presenting MSME data in an accessible and interactive format.

Furthermore, Puklavec et al. [7] emphasized that business intelligence systems have evolved to become accessible and cost-effective solutions specifically designed for SMEs, enabling them to enhance decision-making processes and competitive advantage through data-driven insights.

B. ETL (Extract, Transform, Load) Approach

The ETL (Extract, Transform, Load) process is a fundamental data integration methodology that involves extracting data from multiple source systems, transforming it into a consistent and clean format, and loading it into a target system for analysis and reporting [2]. As a core component of data warehousing and business intelligence pipelines, ETL has become increasingly important in managing the growing volume, variety, and velocity of data generated across organizations. Khan et al. [2] noted that the ETL process fills the critical knowledge gap between unprocessed raw data and actionable insights, addressing challenges such as data quality inconsistencies, schema heterogeneity, and source diversity. In this study, the ETL process was implemented using Python and the Pandas library in Google Colab, which has been recognized as the de-facto tool for tabular data processing and ETL contexts in Python [8].

The transformation phase of ETL is particularly critical in ensuring data quality and reliability. Razali et al. [3] proposed a quality scoring technique within ETL processes that demonstrated measurable improvements in data reliability for downstream analytics. In practice, transformation activities include duplicate removal, handling of missing values, standardization of data formats, and elimination of irrelevant records all of which were performed in this study to reduce 82 raw MSME records from three heterogeneous sources to 53 validated entries. This cleaning process aligns with findings from a systematic review by Taleb et al. [9], which identified duplicate detection, missing value handling, outlier removal, and irrelevant data elimination as the four primary cleaning issues in data processing pipelines.

Furthermore, Yohannis et al. [10] confirmed that integrating heterogeneous data from diverse sources remains a key challenge requiring systematic approaches to ensure data consistency and interoperability. This is further supported by Durmuş et al. [11], who identified data format inconsistency across heterogeneous sources as a primary barrier in multi-source integration pipelines.

II. RESEARCH METHOD

This study employs the ETL (Extract, Transform, Load) approach as the primary methodology for integrating MSME data from multiple heterogeneous sources into a unified and structured dataset, followed by interactive visualization using Microsoft Power BI. The overall research workflow is illustrated in Figure 1. Data were collected from three internal sources at Bank Indonesia Representative Office of Bengkulu Province, covering 53 MSMEs across 9 districts/cities in Bengkulu Province. The following subsections describe each phase of the methodology in detail.

DOI: https://doi.org/10.70656/ijcse.v3i01.910

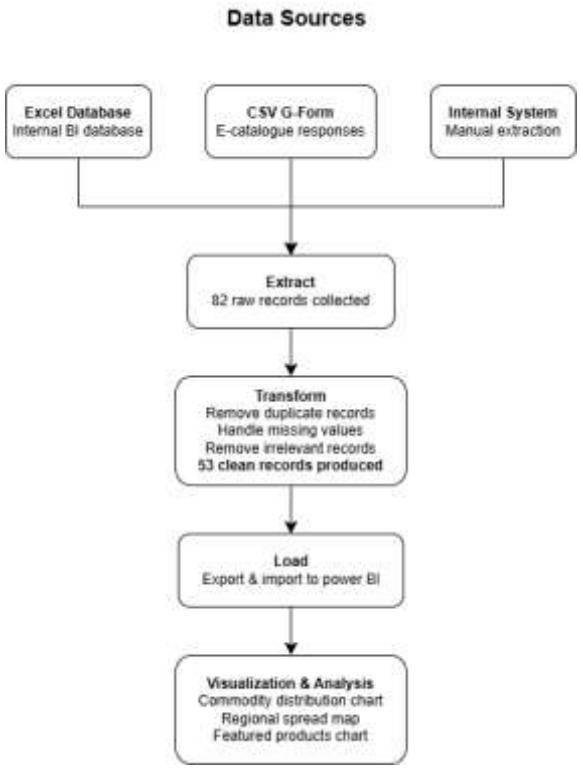


Fig. 1. ETL workflow for MSME data integration and visualization at Bank Indonesia Bengkulu

A. *Extract*

The extraction phase involved collecting MSME data from three heterogeneous internal sources at Bank Indonesia Bengkulu. The first source was an Excel-based internal database maintained by the FPPU unit, containing records of MSME partners registered under Bank Indonesia's development program. The second source was a CSV file generated from G-Form e-catalogue responses, which contained self-reported information submitted directly by MSME owners who registered their businesses through the e-catalogue system. The third source was Bank Indonesia's internal system, from which data were extracted manually due to restricted direct access. Data from this source were carefully read and re-entered to ensure accuracy. The combined extraction from these three sources yielded a total of 82 raw records, which served as the input for the subsequent transformation phase. All extraction activities were performed using Python and the Pandas library in Google Colab.

B. *Transform*

The transformation phase focused on cleaning and standardizing the 82 raw records obtained from the extraction phase to produce a high-quality, analysis-ready dataset. Three main cleaning operations were performed. First, duplicate records were identified and removed these occurred as a result of the same MSME appearing in more than one data source. Second, records with missing or incomplete values in critical fields such as MSME name, commodity type, region, and featured product were filtered out. Third, records that were deemed irrelevant including entries that did not meet the criteria for active MSME partners under Bank Indonesia Bengkulu's development program were eliminated. These transformation activities were carried out using a combination of manual verification and Python-based processing with the Pandas library in Google Colab, including functions such as drop_duplicates() for duplicate removal and dropna() for handling missing values. Following the transformation process, the dataset was reduced from 82 raw records to 53 validated and clean MSME records, ready for the loading phase.

C. *Load*

In the load phase, the cleaned dataset of 53 MSME records was exported from Google Colab into a structured CSV file and subsequently imported into Microsoft Power BI for visualization and analysis. The dataset contains six fields: sequential number, MSME name, brand name, district/city, commodity type, and featured product. Power BI's built-in data connector was used to load the CSV file directly into the Power BI Desktop environment, where it served as the primary data source for dashboard development.

D. *Visualization and Analysis*

The final phase involved developing an interactive dashboard in Microsoft Power BI to visualize and analyze the integrated MSME data. The dashboard was designed to address three analytical objectives: (1) commodity distribution analysis, presenting the proportion of MSMEs across eight commodity categories using a pie chart and bar chart; (2) regional spread analysis, displaying the distribution of MSMEs across 9 districts/cities in Bengkulu Province; and (3) featured product analysis, highlighting the most prominent products offered by MSME partners. The dashboard enables interactive filtering by commodity type and region, allowing users to explore the data dynamically and derive actionable insights to support data-driven decision-making at Bank Indonesia Bengkulu.

III. RESULT AND DISCUSSION

This section presents the results of the ETL-based MSME data integration process and the analytical findings derived from the Power BI visualization dashboard. The results are organized into three main discussions: commodity distribution, regional spread, and featured products of MSME partners binaan Bank Indonesia Representative Office of Bengkulu Province.

A. *ETL Process Results*

The ETL process successfully integrated MSME data from three heterogeneous internal sources at Bank Indonesia Bengkulu. During the extraction phase, a total of 82 raw records were collected from the Excel-based internal database, CSV G-Form e-catalogue responses, and the internal system through manual extraction. The transformation phase reduced this dataset to 53 validated MSME records through the removal of duplicate entries, records with missing or incomplete values, and irrelevant records that did not meet the criteria for active MSME partners. The cleaned dataset was subsequently loaded into Microsoft Power BI, as shown in Figure 2, which confirms that the dashboard successfully consolidates data from 53 MSME partners across Bengkulu Province. This result demonstrates that the ETL approach is effective in addressing the challenge of fragmented multi-source data management, consistent with findings by Khan et al. [2], who emphasized that ETL processes are critical in bridging the gap between unprocessed raw data and actionable insights.

DOI: https://doi.org/10.70656/ijcse.v3i01.910

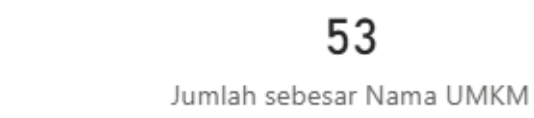


Fig. 2. Total validate MSME partners after ETL process

B. MSME Commodity Distribution

Figure 3 presents the distribution of 53 MSMEs across eight commodity categories in Bengkulu Province. The analysis reveals a significant dominance of the Food and Beverage sector, which accounts for 35 MSMEs (66.04% of the total). This is followed by the Creative Industry sector with 8 MSMEs (15.09%), Plantation Crops with 3 MSMEs (5.66%), and Pharmacy with 2 MSMEs (3.77%). The remaining four sectors Food Crops, Horticulture, Livestock, and Fertilizer each account for 2 or fewer MSMEs, representing a combined 9.44% of the total.

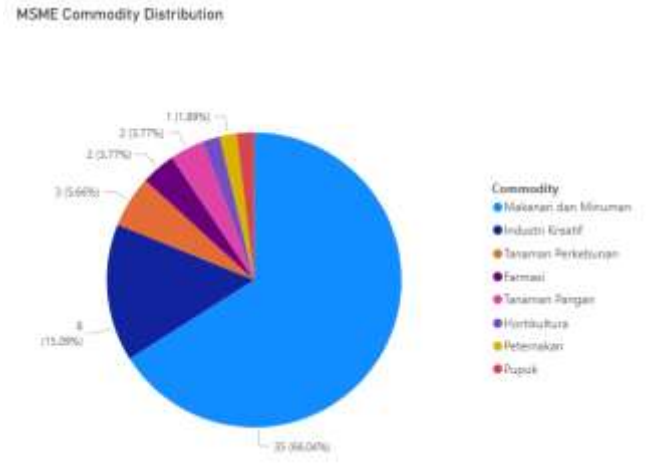


Fig. 3. MSME commodity distribution across Bengkulu Province

The overwhelming dominance of the Food and Beverage sector indicates a high concentration of MSME activities in a single commodity category, which poses a risk of sectoral over-dependence. This finding is significant for Bank Indonesia Bengkulu as it suggests the need for more targeted development programs to encourage diversification into underrepresented sectors such as Horticulture and Livestock, which have strong potential given Bengkulu Province's natural resources. This aligns with the argument by Adiningrat et al. [8] that data-driven insights from MSME platforms can directly inform strategic decisions for MSME development and policy formulation.

C. MSME Regional Distribution

Figure 4 illustrates the distribution of MSME partners across 9 districts/cities in Bengkulu Province. The results show a stark geographic concentration, with Kota Bengkulu accounting for 29 MSMEs representing 54.72% of the total MSME partners. Kabupaten Rejang Lebong ranks second with 5 MSMEs (9.43%), followed by Kabupaten Bengkulu Utara with 4 MSMEs (7.55%). The remaining six districts Kabupaten Bengkulu Selatan, Kabupaten Bengkulu Tengah, Kabupaten Kepahiang, Kabupaten Kaur, Kabupaten Lebong, and Kabupaten Seluma each have 2–3 MSMEs, collectively representing 28.30% of the total.

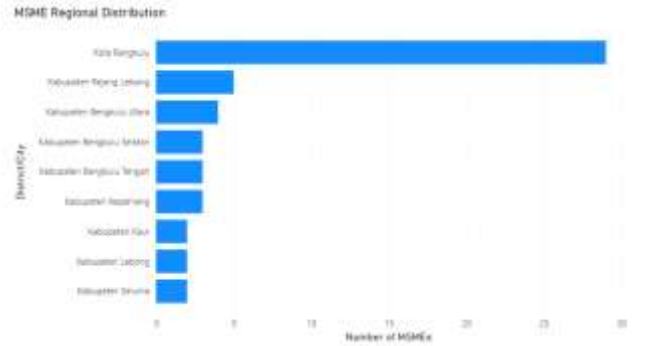


Fig. 4. MSME regional distribution by district/city

This uneven geographic distribution highlights a significant disparity in MSME development across Bengkulu Province, where more than half of all registered MSME partners are concentrated in the provincial capital. This finding suggests that Bank Indonesia Bengkulu's outreach and development programs may benefit from greater focus on rural districts to ensure more equitable MSME growth across the province. The visualization dashboard provides a clear and accessible basis for this kind of regional analysis, supporting the argument by Campante et al. [4] that integrated dashboards enable quick and comprehensive organizational decision-making.

D. MSME Featured Products

Figure 5 presents the distribution of featured products offered by the 53 MSME partners. The analysis identifies Olahan Kopi (Processed Coffee) as the most prominent featured product, represented by 6 MSMEs the highest count among all product categories. This is followed by Biji Kopi (Coffee Beans), Kain Batik Basurek, and Sirup Kalamansi, each represented by 3 MSMEs. A total of 8 products are each represented by 2 MSMEs, while the majority of the remaining products 28 in total are each represented by only 1 MSME, indicating a high diversity of product offerings among MSME partners.

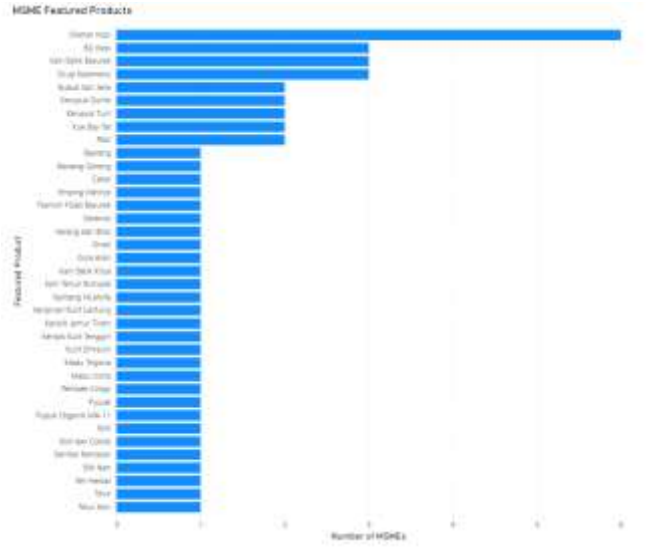


Fig. 5. MSME featured products distribution

The dominance of coffee-related products (Olahan Kopi and Biji Kopi combined: 9 MSMEs, 16.98%) reflects

Bengkulu Province's established role as one of Indonesia's five main coffee-producing provinces, alongside Aceh, Lampung, North Sumatra, and South Sumatra [12]. The province is particularly recognized for its Robusta coffee production, with smallholder farmer productivity ranging from 0.78 to 0.94 tons per hectare, and climate suitability projections indicating that Bengkulu remains one of the most promising regions for future Robusta coffee expansion in Indonesia [13], particularly known for its Robusta coffee. This finding suggests significant potential for coffee-based MSME clusters that could be further developed and promoted through Bank Indonesia's programs. Meanwhile, the presence of Kain Batik Basurek as one of the top featured products reinforces the cultural significance of this traditional textile as a key creative industry product from Bengkulu. These insights demonstrate the value of the visualization dashboard in surfacing actionable product-level intelligence that would otherwise remain hidden in fragmented spreadsheet data, consistent with findings by Reyes et al. [5] who demonstrated that Power BI dashboards effectively bridge complex datasets and strategic decision-making needs.

This is further supported by Alarcón-del-Amo et al. [14], who found that real-time data visualization is a key resource for MSME decision-making, providing immediate insights that guide both strategic and operational decisions.

IV. CONCLUSION

This study successfully implemented an ETL-based approach to integrate and visualize MSME data from three heterogeneous internal sources at Bank Indonesia Representative Office of Bengkulu Province. The ETL process effectively consolidated 82 raw records from an Excel-based internal database, CSV G-Form e-catalogue responses, and manually extracted internal system data into 53 validated MSME records through systematic data cleaning and transformation using Python and the Pandas library in Google Colab. The resulting Power BI dashboard provided actionable insights into MSME commodity distribution, regional spread, and featured products, revealing that 66.04% of MSMEs are concentrated in the Food and Beverage sector, 54.72% are located in Kota Bengkulu, and Olahan Kopi (Processed Coffee) emerges as the most prominent featured product. These findings demonstrate that the ETL-based data integration and visualization framework developed in this study is effective in transforming fragmented multi-source MSME data into comprehensive, accessible, and decision-ready information to support Bank Indonesia Bengkulu's MSME development programs.

This study has several limitations that should be acknowledged. The dataset is limited to 53 MSME partners currently registered under Bank Indonesia Bengkulu's development program and may not represent the full population of MSMEs in Bengkulu Province. Additionally, the data extraction from the internal system relied on manual input, which introduces a potential risk of human error. Future studies are recommended to explore the automation of the ETL pipeline through direct database connectivity, expand the dataset to include a broader range of MSME attributes such as revenue, number of employees, and years of operation, and investigate the application of more advanced analytical techniques such as clustering or predictive modeling to generate deeper insights for MSME policy development.

ACKNOWLEDGMENTS

The authors would like to express sincere gratitude to Bank Indonesia Representative Office of Bengkulu Province, particularly the MSME Development and Implementation Function (FPPU) unit, for providing access to internal MSME data and supporting the implementation of this study during the internship program. The authors also thank the Department of Information Systems, University of Bengkulu, for the academic support throughout this research.

REFERENCES

[1] I. Rahayu, Siti Kurnia; Budiarti, “Digitalization and Informal MSME: Digital Financial Inclusion for MSME Development in The Formal Economy,” vol. 10, no. 1, 2023.

[2] B. Khan, S. Jan, W. Khan, and M. I. Chughtai, “An Overview of ETL Techniques , Tools , Processes and Evaluations in Data Warehousing,” 2024, doi: 10.32604/jbd.2023.046223.

[3] N. Famiera, A. Razali, S. Baharom, S. Abdullah, and N. Indriaty, “Improving Data Reliability Assessment in ETL Processes through Quality Scoring Technique in Data Analytics,” vol. 8, no. December, pp. 2195–2202, 2024.

[4] T. Gonçalves and M. Jos, “Developing Integrated Performance Dashboards Visualisations Using Power BI as a Platform,” 2023.

[5] K. G. Sanabia-lizarraga, B. Carballo-mend, A. Arellano-gonz, and A. Bueno-solano, “Business Intelligence for Agricultural Foreign Trade : Design and Application of Power BI Dashboard,” 2024.

[6] F. N. Arifin, A. A. Cahyarani, and M. A. Furqon, “Small Medium Enterprise Data Visualization Analysis as Strategic Business Development,” vol. 18, pp. 1–7, 2023.

[7] K. Ragazou, I. Passas, A. Garefalakis, and C. Zopounidis, “Business intelligence model empowering SMEs to make better decisions and enhance their competitive advantage,” Discov. Anal., 2023, doi: 10.1007/s44257-022-00002-3.

[8] S. Raschka, J. Patterson, and C. Nolet, “Machine Learning in Python : Main Developments and Technology Trends in Data Science , Machine Learning , and Artificial Intelligence,” 2020, doi: 10.3390/info11040193.

[9] O. Alotaibi and E. Pardede, “Cleaning Big Data Streams : A Systematic Literature Review,” pp. 1–24, 2023.

[10] I. M. Putrama and P. Martinek, “Heterogeneous data integration : Challenges and opportunities,” Data Br., vol. 56, p. 110853, 2024, doi: 10.1016/j.dib.2024.110853.

[11] T. Horak, P. Strelec, M. Kebisek, P. Tanuska, and A. Vaclavova, “Purposes of Analysis within Industry 4 . 0 Concept,” 2022.

[12] F. Ashardiono and A. Trihartono, “Optimizing the potential of Indonesian coffee : a dual market approach,” Cogent Soc. Sci., vol. 10, no. 1, p., 2024, doi: 10.1080/23311886.2024.2340206.

[13] Y. Sarvina, T. June, S. H. Sutjahjo, R. Nurmalina, and E. Surmaini, “International Journal of Sustainable Development and Planning Projection of Robusta Coffee ’ s Climate Suitability for Sustainable Indonesian Coffee Production,” vol. 18, no. 4, pp. 1069–1078, 2023.

[14] G. García-vidal, A. Sánchez-rodríguez, L. Guzmán-vilar, and R. Martínez-vivar, “Exploring MSME Owners ’ Expectations of Data-Driven Approaches to Business Process Management,” 2025.