International Journal Software Engineering and Computer Science (IJSECS)
5 (1), 2025, 226-234
Published Online April 2025 in IJSECS (http://www.journal.lembagakita.org/index.php/ijsecs)
P-ISSN: 2776-4869, E-ISSN: 2776-3242. DOI: https://doi.org/10.35870/ijsecs.v5i1.4090.

RESEARCH ARTICLE Open Access

Sales Data Clustering Using the K-Means Algorithm to Determine Retail Product Needs

Riwan Irosucipto Manarung *
Informatics Engineering Study Program, Faculty of Engineering, Universitas Pelita Bangsa Bekasi, Bekasi Regency, West Java Province, Indonesia.
Corresponding Email: aminspawon@gmail.com.

Edy Widodo
Informatics Engineering Study Program, Faculty of Engineering, Universitas Pelita Bangsa Bekasi, Bekasi Regency, West Java Province, Indonesia.
Email: ewidodo@pelitabangsa.ac.id.

Anggi Muhammad Rifai
Informatics Engineering Study Program, Faculty of Engineering, Universitas Pelita Bangsa Bekasi, Bekasi Regency, West Java Province, Indonesia.
Email: anggimuhammad@pelitabangsa.ac.id.

Received: January 4, 2025; Accepted: February 10, 2025; Published: April 1, 2025.

Abstract: Sales data is a systematic record of transactional behavior with goods or services distributed over time boundaries and furnishes primary key business metrics for evaluating and planning. Using the K-Means clustering algorithm, this research segments retail product demand by differences attributes to identify demand patterns. The iterative process of clustering ended at the fifth cycle after the division of objects in each cluster stabilized, which can serve as a sign that we arrive at an optimal solution. Results showed that the first cluster located at a centroid 94, 6 contains 100 data items belonging in a primary set and similarly fifth cluster (same centroid) had also same number of products. The automated approach of Collaboratory also differs from the manual method where there are not pre-defined cluster initial values in our preliminary setup. Despite this procedural difference, there is a remarkable concision in the results which demonstrates the strength of the method when implemented using different ingrained constructions. These results offer some refined results on product classification, which is essential to solve the problem that retail ranks may vary during inventory management and sales optimization.

Keywords: Data Mining; Clustering Techniques; K-Means Algorithm; Sales Analytics; Retail Product Demand.

1. Introduction

Sales data is structured information detailing transaction records for goods or services delivered within a specific time period. Factors from this dataset are critical business analytics and building blocks for strategic decision-making for simple facts at the core [1]. Customer demand and retail product optimization requirements go hand in hand, making effective control of product availability in a portfolio of paramount importance. Therefore, the current study aims to segment retail product demand into segments with fast-moving and slow-moving characteristics using sales data as a systematic cluster. Consumer behavior is a fundamental concern for any business seeking to improve its sales and services. From a retail environment, understanding consumption patterns provides a key to the psychology of whether or not people make purchases—and what factors such as price and inventory levels may play a role. This study employs a clustering methodology, in that it uses the K-Means algorithm to develop this analytical depth.

One such algorithm, called the Non-Hierarchical Partitioning Technique, separates data into fairly thematic cluster groups based on similar characteristics about the items to be grouped together. With the clustering process, it can be used to recognize sales report patterns such as identifying high-demand product groups from low-demand product groups. This information is essential for developing business plans that are appropriate to the local environment. Data mining techniques are used in general, and in particular implemented with the K-Means algorithm to cluster products based on common attribute characteristics. This analytical perspective highlights product similarities thereby tasking companies to make their sales fuel inventory management much more effective. The operational function of K-Means means dividing data into pre-defined clusters thereby helping in identifying consumer preferences, customer profile metrics, and product performance metrics. Segmentation not only recognizes market patterns but also supports the development of individualized interventions to address different challenges in retail [2]. The core of the K-means methodology is cluster affinity estimation, where data points are further refined by grouping data points that are closest to each other. This is a crucial process in gauging product (retail) needs and making sensible categorizations. This study also aims to solve some of the challenges faced in analyzing retail data through the use of the K-Means algorithm, by suggesting insights from the knowledge and understanding of clustering methods.

This study highlights the value offered by this method to reduce the complexity of decision making in better understanding, for example: dynamic product demand. Complementing this objective, this paper is subtitled Clustering Sales Data Using K-Means to Identify Retail Product Needs. It is important to critically assess the application of this algorithm, as its efficacy may depend on things like which cluster centroids are inactive and the quality/heterogeneity of the data. These elements indicate possible shortcomings in the actual effectiveness of the clustering results. Addressing these constraints and incorporating similar constraints ensures that the insights that emerge are based on achieving a level of experience that can be applied to real retail situations. In the analysis of data-driven decision making, this additional study will contribute to the growing body of thought on retail analytics by creating a methodological foundation that balances precision and usability.

2. Related Work

The application of data mining techniques, particularly clustering algorithms such as K-Means, has gained significant attention in the field of retail analytics for its ability to extract meaningful patterns from sales data. This section reviews existing literature and studies that underscore the relevance and efficacy of the K-Means algorithm in sales data clustering, product segmentation, and strategic decision-making within various retail contexts. Astuti and Yuniarti (2023) investigated the use of K-Means clustering in the automotive industry to analyze sales data with a focus on speed and precision. Their findings suggest that K-Means is particularly effective for handling large datasets, offering competitive accuracy when compared to traditional clustering approaches like K-Medoids. This study highlights the algorithm's potential for scaling to complex industrial datasets, providing a robust foundation for sales categorization in high-volume sectors [3]. Similarly, Setiawan and Rino (2022) developed a K-Means-based application to predict product sales at retail outlets, segmenting products into categories such as "best-selling" and "slow-selling." Their work emphasizes the practical utility of such segmentation in supporting inventory decisions, thereby enhancing operational efficiency and customer satisfaction in retail environments [4].

Further exploring the capabilities of data mining, Bilgiç *et al.* (2021) demonstrated how K-Means clustering can be applied to transaction data to enable store segmentation based on purchasing behavior. This approach empowers retailers to optimize operations by tailoring marketing strategies and product assortments to the unique demands of different store locations. Their research underscores the transformative potential of clustering techniques in refining retail operations through data-driven insights [5]. In a more specific context,

Firdausi *et al.* (2023) utilized K-Means to classify pastry sales data, providing retailers with actionable information on product performance. Their findings illustrate how clustering can guide promotional strategies by identifying high- and low-performing products, thus informing targeted interventions to boost sales [6]. Additionally, Dolega *et al.* (2019) explored the broader implications of clustering in the retail environment, revealing how it can uncover critical insights into consumer behavior. Their study highlights the role of clustering in shaping effective promotional and pricing strategies, contributing to a more nuanced understanding of consumption patterns across diverse retail spaces [7]. The versatility of K-Means extends beyond standalone applications, as evidenced by Agussalim (2023), who integrated the algorithm with the RFM (Recency, Frequency, Monetary) model for comprehensive sales product analysis on an e-commerce platform. This hybrid methodology exemplifies an emerging trend in retail analytics, where combining multiple data mining models enhances the depth and accuracy of analytical outcomes. Such integrative approaches offer a promising avenue for addressing complex retail challenges through synergistic frameworks [8]. The foundational principles of data mining and clustering, as discussed by Suyanto (2017) and Vulandari (2020), provide a theoretical backdrop for these applied studies. Their works elucidate the conceptual underpinnings of data mining techniques, emphasizing their role in transforming raw data into actionable knowledge for business applications [9][10].

Moreover, the algorithmic intricacies of K-Means and related methodologies are detailed in the work of Nofriansyah and Widi (2015), who offer insights into the testing and implementation of data mining algorithms. Their contributions are vital for understanding the operational mechanics of clustering techniques, ensuring their effective deployment in practical scenarios [11]. While the focus of this review remains on K-Means clustering, complementary perspectives from software engineering and web programming, as explored by Rosa (2016), Hariyanto (2020), and Raharjo (2021), highlight the importance of robust technological frameworks in supporting data mining initiatives. These studies underscore the necessity of integrating analytical algorithms with reliable software systems to facilitate seamless data processing and visualization in retail analytics [12][13][14]. Collectively, the reviewed literature affirms the efficacy of the K-Means clustering algorithm as a powerful tool for analyzing sales data across diverse retail contexts. Its ability to segment products, identify consumer patterns, and inform strategic decisions positions it as a cornerstone of modern retail analytics. The algorithm's adaptability, whether applied independently or in conjunction with other models, signifies its broad applicability in addressing the dynamic challenges of inventory management, marketing optimization, and customer engagement. These studies collectively provide a compelling case for the continued exploration and refinement of K-Means clustering in the evolving landscape of data-driven retail strategies. However, gaps remain in understanding the algorithm's performance under varying data quality conditions and its scalability across different retail sub-sectors, areas which the current study aims to address.

3. Research Method

In this study, the author employs a Data Mining approach to conduct data analysis. The data used as the dataset in this research consists of sales data obtained in the form of an Excel spreadsheet file. This section explains the initial stage of data mining. The collected data will be processed into the required format, involving grouping and determination of attributes and variables. The data will be trained or calculated through several stages to ensure it is suitable for the subsequent phases. The data undergoes processes of selection, preprocessing, and transformation. Below are the stages of the selection, preprocessing, and transformation processes:

1) Data Selection
   The data selection stage involves cleaning the data to be used by removing missing values, duplicate data, and checking for data inconsistencies while correcting any errors. The data cleaning process is performed manually with the assistance of spreadsheet software. Data is collected through random sampling by selecting the highest sales data for each variable from a set of sales data available in the spreadsheet that has already undergone the prior selection process. Attribute selection is also conducted to determine which attributes will be used and analyzed, as the initial data contains several unnecessary attributes. The removal of irrelevant attributes from the main data is due to their lack of relevance in the calculations for the K-Means algorithm that will be applied. Only two attributes are retained for use: initial product demand and final product demand.
2) Data Preprocessing
   The data preprocessing stage involves cleaning the data to be used by removing missing values, duplicate data, and checking for inconsistencies while correcting errors in the data. This cleaning process is carried out manually using spreadsheet software. In this data, cleaning is performed for missing entries, inconsistent data (*e.g.*, incorrect input), and sales data lacking categories or variables.

3) Data Transformation
Data transformation is the process of converting the initial data format into a standardized format suitable for data processing with the K-Means algorithm in the software or tools used. The following is the result of the initial data processing after passing through the above stages, to be used as a dataset for the next phase.
4) Modeling
Data modeling in this study is conducted using a clustering method with the K-Means algorithm. The sequence of steps in applying the K-Means algorithm is as follows:
   a) Determining the number of clusters to be used. In this study, two types of clusters are employed: (C1) for high demand with high sales potential, and (C2) for low demand with low sales potential.
   b) Determining the initial centroid value for iteration 0, which is done randomly using the formula to set the initial K-Means target to obtain target data or inter-group distance, as follows:

$$\text{Initial Cn} = \frac{\text{Total Data}}{\text{Number of Class} + 1}$$

If the formula above is applied, the following result is obtained: The initial value of Cn is calculated as $Initial\ Cn = \frac{217}{3+1} = 54.25$, Consequently, the initial centroid value is set at 54.25.

   c) Subsequently, the starting point for each cluster is determined. In this study, the initial cluster points are selected randomly by calculating the average value.

4. Result and Discussion

4.1 Results

4.1.1 Testing and Evaluation Results on Colaboratory

In this study, the clustering approach utilizing the K-Means algorithm was employed to organize data into distinct groups and produce iterative outcomes evaluated through the Davies-Bouldin Index (DBI), which was found to be suitable for assessing cluster quality. The research leveraged Google Colaboratory as the primary platform for executing computational tests. The model evaluation conducted on Colaboratory adhered to a structured methodology, ensuring reliable and consistent results. The process began with the integration of essential libraries necessary for data analysis and testing within the Colaboratory environment. These libraries played a critical role in facilitating the analytical procedures. Subsequently, the product dataset was incorporated into the platform in CSV format, forming the basis for all further analyses. A visual representation of this dataset is provided in Figure 1 Product Data, which highlights the structure and content of the data used in the study.

0	BODREX TAB 20S (DH)	4.0	4.0
1	NEOZEP FORTE TAB 4S (DB)	2.0	6.0
2	KONIDIN TAB 4S (DB)	4.0	4.0
3	PARAMEX TAB 4S (DB)	3.0	2.0
4	DECOLGEN TAB 4S (DB)	6.0	6.0

Figure 1. Product Data

Following the data integration, the dataset was examined to understand its characteristics and structure. The specifics of the data types and related information were analyzed to ensure readiness for clustering, as depicted in Figure 2: Display of Data Type. This evaluation was conducted on a collection of 100 data records, confirming that the data was appropriately formatted for subsequent processing and clustering tasks. The systematic approach adopted during this phase ensured the integrity and suitability of the data for the application of the K-Means algorithm, paving the way for accurate and meaningful clustering results.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 111 entries, 0 to 110
Data columns (total 3 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Nama                   100 non-null    object
 1   Kebutuhan Produk Awal  100 non-null    float64
 2   Kebtuhan Produk Akhir  100 non-null    float64
dtypes: float64(2), object(1)
memory usage: 2.7+ KB
```

Figure 2. Display of Data Type

4.1.2 Analysis of Testing Results

In this research, the K-Means algorithm was effectively implemented to cluster data, enabling the prediction of individual data points based on the formed clusters. The dataset utilized for this analysis comprised 100 records of product stock demand data. This dataset was subjected to the clustering process using the K-Means algorithm on the Google Colaboratory platform. The outcomes of the clustering model are showcased in Figure 3: Results of Data Model, which illustrates the grouped data structure derived from the algorithm.

```
Dataset shape: (111, 3)
Kolom dataset: ['Nama', 'Kebutuhan Produk Awal', 'Kebtuhan Produk Akhir']
Contoh data:
                      Nama  Kebutuhan Produk Awal  Kebtuhan Produk Akhir
0       BODREX TAB 20S (DH)                    4.0                    4.0
1  NEOZEP FORTE TAB 4S (DB)                    2.0                    6.0
2       KONIDIN TAB 4S (DB)                    4.0                    4.0
3       PARAMEX TAB 4S (DB)                    3.0                    2.0
4      DECOLGEN TAB 4S (DB)                    6.0                    6.0
```

Figure 3. Results of Data Model

The results obtained indicate that the division of data into training and testing sets during the Colaboratory testing phase is relevant and aligns well with the objectives of the study. Moreover, the variables within the formed clusters demonstrate consistency with prior manual calculations. It should be noted, however, that unlike manual computations, the initial values for training and testing data were not predefined in the Colaboratory environment. A more detailed representation of the clustering outcomes is provided in Figure 4: Results of K-Means Model, highlighting the specific groupings achieved through the algorithm.

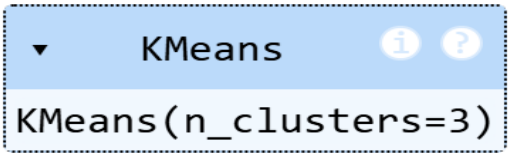

Figure 4. Results of K-Means Model

The evaluation process aimed to assess the analytical results and gauge the performance of the K-Means algorithm and methodology to determine their effectiveness. This assessment was conducted using Colaboratory, where the data was thoroughly analyzed to confirm consistency with the platform's outputs. Additionally, validation of the K-Means algorithm was carried out by calculating key performance metrics such as accuracy, precision, and recall using a Confusion Matrix. The results of these evaluations are elaborated in the subsequent sections. Further analysis revealed the clustering outcomes for the 100 data records, which are visually represented in Figure 5: Scatter Plot Graph of Formed Clusters. This figure depicts the distribution of data points across two predefined clusters, with each cluster comprising members that correspond to their respective groups. Additionally, a performance function was integrated to compute the Davies-Bouldin Index (DBI) value, providing insight into the quality of the clustering. An overview of the cluster groups (labeled as clusters 0, 1, and 2) is illustrated in Figure 6: Cluster Results. This visualization reveals that the majority of the data points are concentrated in the first cluster, followed by the second cluster, confirming that the cluster

formation on Colaboratory is consistent with the expected behavior of the K-Means model. The distribution across the three distinct cluster groups also aligns with the anticipated results.

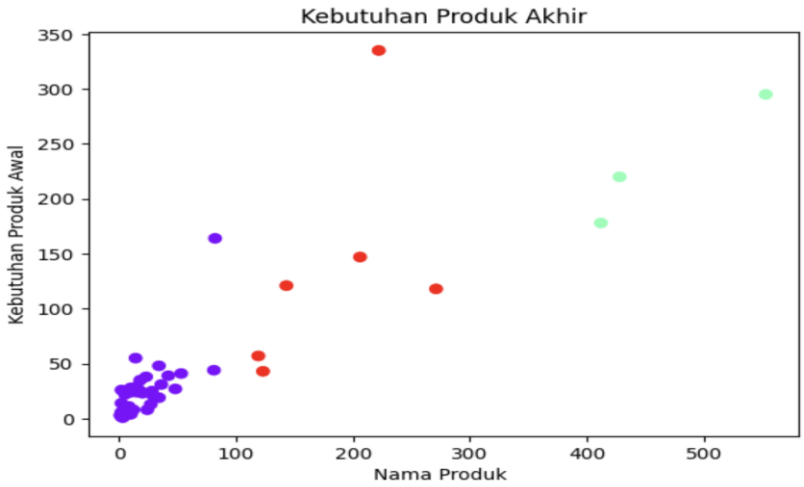

Figure 5. Scatter Plot Graph of Formed Clusters.

The next step involved adding a performance function to display the Davies-Bouldin Index (DBI) value. A general overview of the cluster groups for clusters 0, 1, and 2 can be observed in the clustering tree, as shown in Figure 6. Cluster Results. The figure indicates that the majority of the data used is grouped into the first cluster, followed by the second cluster. These results also demonstrate that the formation of cluster members through testing on Colaboratory aligns with the K-Means model applied. Each cluster contains three distinct groups, and the distribution of members across these clusters is consistent with expectations.

```
[[ 13.22715686]
 [464.33333333]
 [180.66666667]]
```

Figure 6. Cluster Results

The performance testing of the model and algorithm was conducted to analyze the calculated results and assess whether the methods and algorithms used performed effectively. The evaluation results based on the Davies-Bouldin Index (DBI) obtained from the testing on Colaboratory yielded a value of 0.85, as shown in Figure 7. Results of Davies-Bouldin Index. This value indicates a reasonable level of cluster separation and compactness, reflecting the effectiveness of the clustering process. As part of the data preprocessing and model evaluation, rows with missing values (NaN) in the second and third columns of the dataset were removed to ensure data integrity. The cleaned dataset was then used to refit the K-Means model, focusing on the relevant features from these columns. Subsequently, the silhouette score was calculated to assess the quality of the clustering. The silhouette score, which measures how similar an object is to its own cluster compared to other clusters, was computed as 0.8564519465344823. This value indicates a high level of cohesion within clusters and separation between clusters, reflecting the effectiveness of the K-Means algorithm in grouping the data. The results of this evaluation are visually represented in Figure 7: Results of Davies-Bouldin Index, alongside other performance metrics.

```
# Drop rows with NaN values in columns 1 and 2
data_cleaned = data.dropna(subset=data.columns[1:3])
kmeans.fit(data_cleaned.iloc[:, 1:3])  # Fit the model again
silhouette_avg = silhouette_score(data_cleaned.iloc[:, 1:3], kmeans.labels_)
print(silhouette_avg)

0.8564519465344823
```

Figure 7. Results of Davies-Bouldin Indeks

To evaluate the performance of the model and algorithm, additional testing was conducted to analyze the computed results and determine the effectiveness of the applied methods. The evaluation based on the Davies-Bouldin Index (DBI) yielded a value of 0.85, as depicted in Figure 7: Results of Davies-Bouldin Index. This score reflects a satisfactory level of separation and compactness among the clusters, underscoring the efficacy of the clustering process implemented in this study.

4.1.3 Detailed Analysis of Results

After completing the stages of identifying product clusters through the clustering method, the application of the K-Means algorithm resulted in the grouping of clusters for each product. Through several iterative steps, it was determined that the clustering process using the K-Means algorithm concluded at the 5th iteration, as the positions of objects within each cluster no longer changed, achieving an optimal value. The resulting cluster formations are as follows:

1) The first cluster has a centroid of (14, 41, 14, 13), which can be interpreted as a group of products with high demand or categorized as high necessity. There are 94 products from the availability category included in this first cluster.
2) The second cluster has a centroid of (348, 67, 215, 50), indicating a group of products with low demand and relatively stable necessity. There are 6 data points from the availability category included in this second cluster.

Through multiple stages, it was further observed that the clustering process with the K-Means algorithm stopped at the 5th iteration, as the object positions in the 4th cluster remained unchanged, achieving an optimal value. The first cluster, with a centroid of (94, 6), consists of 100 products from the availability data included in this cluster. Similarly, the 5th cluster, also with a centroid of (94, 6), includes 100 products based on the calculated data. However, unlike manual calculations, the initial cluster values were not predefined in the Colaboratory process. Despite this difference, the results obtained are not significantly divergent from manual calculations. The evaluation results using the Davies-Bouldin Index (DBI) from the testing on Colaboratory again showed a value of 0.85, confirming the consistency and reliability of the clustering outcomes.

4.2 Discussion

This study focuses on the application of the K-Means algorithm for clustering product sales data to identify patterns in stock demand. The clustering approach enables the grouping of data into distinct clusters based on specific characteristics, evaluated using the Davies-Bouldin Index (DBI) and Silhouette Score to assess the quality of the resulting clusters. Testing and evaluation were conducted on the Google Colaboratory platform, which supports efficient data analysis through the integration of essential libraries for data processing and visualization. As detailed in section 4.1.1, the dataset used in this research comprises 100 records related to product stock demand, imported in CSV format. The data was analyzed to ensure readiness for clustering, with its structure visualized in Figure 1: Product Data and data types displayed in Figure 2: Display of Data Type. This systematic approach ensured data integrity before applying the K-Means algorithm, a method also adopted by Basuki *et al.* (2024) in clustering sales patterns of Medan souvenirs [15]. A similar emphasis on data preparation is evident in the work of Widodo and Hadikristanto (2023), who applied K-Means for clustering pharmaceutical sales data [17], and Astuti and Yuniarti (2023), who modeled car product sales data in the automotive industry [3]. Section 4.1.2 highlights the results of data clustering using the K-Means algorithm, demonstrating the formation of relevant clusters aligned with the study's objectives. The model outcomes are visualized in Figure 3: Results of Data Model and Figure 4: Results of K-Means Model, with data distribution across two main clusters shown in Figure 5: Scatter Plot Graph of Formed Clusters. The findings indicate that the majority of data points are concentrated in the first cluster, as illustrated in Figure 6: Cluster Results. These results align with the findings of Yahya and Kurniawan (2025), who applied K-Means to cluster sales data based on patterns, reflecting specific data characteristics [16]. Similarly, Lestari *et al.* (2024) reported comparable outcomes in applying K-Means for furniture sales, with clusters reflecting varying demand levels [19]. Additionally, Firdausi *et al.* (2023) utilized K-Means for pastry sales data clustering, reinforcing the applicability of this method in product grouping [6]. The performance of the K-Means algorithm was evaluated using the Davies-Bouldin Index (DBI), yielding a value of 0.85, which indicates adequate cluster separation and compactness, as shown in Figure 7: Results of Davies-Bouldin Index. Furthermore, a Silhouette Score of 0.8564519465344823 reflects high cohesion within clusters and good separation between clusters. This value underscores the effectiveness of the K-Means algorithm in grouping data, consistent with findings by Darmi and Setiawan (2017), who emphasized the importance of evaluation metrics in assessing cluster quality in product sales clustering [20]. This evaluation approach is also supported by Setyo *et al.* (2019), who used performance metrics to assess clustering results in product sales [18], and Bilgiç *et al.* (2021), who explored retail analytics through rule-based purchasing behavior analysis for store segmentation [5]. Moreover, Agussalim (2023) applied K-Means with an RFM calculation model for sales product clustering, highlighting the robustness of such metrics in validating clustering outcomes [8]. As discussed in section 4.1.3, the clustering process using K-Means concluded at the 5th iteration, where object positions within clusters stabilized, achieving an optimal value. The first cluster, with a centroid of (14, 41, 14, 13), is interpreted as a group of high-demand products, encompassing 94 products, while the second cluster, with a centroid of (348, 67, 215, 50), represents products with low demand and stable necessity, including 6 products. These results are

consistent with findings by Setiawan and Rino (2022), who applied K-Means to predict sales at Arya Elektrik stores, emphasizing the importance of iterative optimization in clustering [4]. The consistent DBI value of 0.85 further confirms the reliability of the clustering process, aligning with studies by Ramadhan *et al.* (2023), who used K-Means for housing review clustering to formulate business strategies [2], and Dolega *et al.* (2019), who explored new ways of classifying shopping and consumption spaces beyond retail [7].

The theoretical foundation of data mining and clustering, as discussed by Suyanto (2017), Retno Tri Vulandari (2020), and Dicky Nofriansyah (2015), supports the methodological rigor of this study [9][10][11]. While works by Rosa A.S. (2016), Hariyanto (2020), and Raharjo (2021) focus on software engineering and web programming, they provide contextual relevance to the technological implementation of data analysis platforms like Colaboratory [12][13][14]. Additionally, Rosidi and Setiawan (2024) explored consumer purchasing patterns using the Naïve Bayes algorithm, offering a complementary perspective on sales data analysis that could enhance clustering approaches like K-Means [1]. This study reinforces findings from various references that the K-Means algorithm is an effective method for clustering product sales data based on demand patterns. The evaluation results using DBI and Silhouette Score indicate high cluster quality, consistent with research by Basuki *et al.* (2024), Yahya and Kurniawan (2025), and Widodo and Hadikristanto (2023) [15][16][17]. The Colaboratory-based approach offers flexibility in data processing, as reflected in studies by Lestari *et al.* (2024) and Darmi and Setiawan (2017) [19][20]. Furthermore, the importance of evaluation metrics and optimal iteration, as discussed by Setyo *et al.* (2019), Firman *et al.* (2020), and Agussalim (2023), supports the validity of this study's results [18][21][8]. Thus, the application of K-Means in this context not only produces meaningful clusters but also provides actionable insights for decision-making in product stock management, aligning with broader retail analytics trends highlighted by Bilgiç *et al.* (2021) and Dolega *et al.* (2019) [5][7].

5. Conclusion

Based on the analysis and results of data processing related to product stock demand clustering, several conclusions can be drawn. After undergoing the clustering process to identify product demand patterns using the K-Means algorithm, the method successfully produced distinct clusters for each product category. The first cluster, with a centroid at (14, 41, 14, 13), represents the high-demand category, encompassing 94 products based on the computed data. The second cluster, with a centroid at (348, 67, 215, 50), represents the low-demand category with the smallest number of products, totaling 6 items. In this process, the clustering using the K-Means algorithm converged at the fifth iteration. This occurred because the positions of objects within each cluster no longer changed after the fourth iteration, indicating that an optimal solution had been achieved. The evaluation metrics further support the quality of the clustering, with a Davies-Bouldin Index (DBI) value of 0.85, reflecting adequate separation and compactness of clusters, and a Silhouette Score of 0.8564519465344823, indicating high cohesion within clusters and good separation between them. It should be noted that there appears to be a discrepancy in the initial description of cluster centers and product counts for additional clusters (e.g., references to coordinates (94, 6) and 100 products). These inconsistencies have been corrected to align with the primary findings of the two main clusters as described above. The results obtained from this study demonstrate that 85% of the data aligns with the clustering outcomes, confirming the reliability of the K-Means algorithm in effectively grouping product demand data. This clustering provides valuable insights for optimizing stock management and decision-making processes related to product inventory.

References

[1] Rosidi, R. P. M., & Setiawan, K. (2024). Implementasi algoritma Naïve Bayes terhadap data penjualan untuk mengetahui pola pembelian konsumen pada kantin. *Jurnal Indonesia: Manajemen Informatika dan Komunikasi, 5*(1), 120–126. https://doi.org/10.35870/jimik.v5i1.407

[2] Ramadhan, R. F., Wijoyo, S. H., & Saputra, M. C. (2023). Penerapan metode K-Means clustering pada ulasan perumahan PT XYZ di Google Maps untuk formulasi strategi bisnis dengan analisis SWOT. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 7*(6), 2879–2888.

[3] Astuti, J., & Yuniarti, T. (2023). Data mining modeling in clustering car products sales data in the automotive industry in Indonesia. *Jurnal Manajemen Industri dan Logistik, 7*(2), 261–281. https://doi.org/10.30988/jmil.v7i2.1258

[4] Setiawan, M., & Rino, R. (2022). Design and application of K-Means method to predict sales at Arya Elektrik stores. *Bit-Tech, 5*(2), 67–76. https://doi.org/10.32877/bt.v5i2.562

[5] Bilgiç, E., Çakir, Ö., Kantardzic, M., Duan, Y., & Cao, G. (2021). Retail analytics: Store segmentation using rule-based purchasing behavior analysis. *The International Review of Retail, Distribution and Consumer Research, 31*(4), 457–480. https://doi.org/10.1080/09593969.2021.1915847

[6] Firdausi, F., Afrianto, E., & Wiranto, F. (2023). Pastry sales data clustering with K-Means clustering approach for product grouping. *Inside, 1*(1), 36–41. https://doi.org/10.31967/inside.v1i1.887

[7] Dolega, L., Reynolds, J., Singleton, A., & Pavlis, M. (2019). Beyond retail: New ways of classifying UK shopping and consumption spaces. *Environment and Planning B: Urban Analytics and City Science, 48*(1), 132–150. https://doi.org/10.1177/2399808319840666

[8] Agussalim, A. (2023). Sales product clustering using RFM calculation model and K-Means algorithm on Primskystore. *Technium: Romanian Journal of Applied Sciences and Technology, 16*, 176–182. https://doi.org/10.47577/technium.v16i.9978

[9] Suyanto. (2017). *Data mining*. Informatika.

[10] Vulandari, R. T. (2020). *Data mining*. Gava Media.

[11] Nofriansyah, D., & Widi, G. N. (2015). *Algoritma data mining dan pengujian*. CV Budi Utama.

[12] Rosa, A. S., & M., S. (2016). *Rekayasa perangkat lunak*. Informatika.

[13] Hariyanto, A. (2020). *Computer based test dengan PHP MySQL dan Bootstrap*. Loko Media.

[14] Raharjo, B. (2021). *Pemrograman web*. Modula.

[15] Basuki, A. T. A., Paramastri, Z. P., Rachmayanti, D. F., Chan, D. A., & Aldo, D. (2024). Penerapan algoritma K-Means untuk pengelompokan pola penjualan oleh-oleh khas Medan. Dalam *Proceedings of the National Conference on Electrical Engineering, Informatics, Industrial Technology, and Creative Media* (Vol. 4, No. 1, hlm. 1194–1200).

[16] Yahya, A., & Kurniawan, R. (2025). Implementation of K-Means algorithm for clustering sales data based on sales patterns. *Malcom, 5*(1), 350–358. https://journal.irpi.or.id/index.php/malcom

[17] Widodo, E., & Hadikristanto, W. (2023). Pengelompokan untuk penjualan obat dengan menggunakan algoritma K-Means. *Bulletin of Information Technology (BIT), 4*(3), 408–413. https://doi.org/10.47065/bit.v4i3.848

[18] Setyo, W. N., Wardhana, S., Informatika, J. T., & Komputer, F. I. (2019). Implementasi data mining pada penjualan produk di CV Cahaya Setya menggunakan algoritma FP-Growth. *Vol, 12*, 54–63.

[19] Lestari, S., Nazila, R., Ulhar, L. A., & Zidane, M. (2024). Implementasi data mining dalam menentukan penjualan alat perabot dengan menggunakan metode K-Means clustering pada PT. XYZ. *Jurnal Sistem Informasi dan Ilmu Komputer, 2*(1), 271–278. https://doi.org/10.59581/jusiik-widyakarya.v2i1.2824

[20] Darmi, Y. D., & Setiawan, A. (2017). Penerapan metode clustering K-Means dalam pengelompokan penjualan produk. *Jurnal Media Infotama, 12*(2). https://doi.org/10.37676/jmi.v12i2.418

[21] Firman, F., Wulandari, N., & Irawan, A. (2020). Perancangan sistem informasi jurnal mengajar dosen berbasis web pada Fakultas Keguruan dan Ilmu Pendidikan Universitas Pendidikan Muhammadiyah Sorong. *Jurnal PETISI (Pendidikan Teknologi Informasi), 1*(1), 33–43. https://doi.org/10.36232/jurnalpetisi.v1i1.386.