Jurnal Kecerdasan Buatan dan Teknologi Informasi
Vol. 5, No. 3, September 2026, pp. 632–642
e-ISSN: 2964-2922, p-ISSN: 2963-6191

Open access

Original Research

Cosmetic Product Segmentation Analysis Using K-Means Clustering at PT Mandom Bekasi

Mona Dewintha Agustine [a*], Adi Muhajirin [b], Prio Kustanto [c]

[a,b,c] Informatics, Faculty of Computer Science, Universitas Bhayangkara Jakarta Raya, Bekasi, Indonesia

*Corresponding Author: 202210715213@mhs.ubharajaya.ac.id

ABSTRACT

PT. Mandom Indonesia Tbk manages a wide range of cosmetic products with varying sales levels and inventory turnover rates, creating challenges in inventory management and marketing strategy formulation. This study aims to segment cosmetic products based on sales patterns and inventory turnover using the K-Means Clustering algorithm within a Knowledge Discovery in Databases (KDD) framework. The research stages include data selection, preprocessing, transformation, clustering, and evaluation. The dataset consists of 436 cosmetic products with attributes including sell in, sell out, stock, and expiration date, sourced from PT Mandom's internal sales report for the year 2025. Feature engineering produced two derived variables, the sell out to sell in ratio and the remaining days until expiration, which were normalized using Min-Max Scaling. The optimal number of clusters, determined using the Elbow Method and validated with the Silhouette Score, was three. The K-Means algorithm successfully grouped the products into three segments: Fast Moving (75 products, 17.2%), Medium Moving (299 products, 68.6%), and Slow Moving (62 products, 14.2%). The Fast Moving cluster exhibited the highest sell in, sell out, and sell-through ratio values, while the Slow Moving cluster showed the lowest ratio, indicating a higher risk of stock accumulation. These segmentation results can serve as a data-driven basis for inventory management, distribution planning, and marketing strategy decisions at PT Mandom.

How to Cite:
M. D. Agustine, A. Muhajirin, and P. Kustanto, "Cosmetic Product Segmentation Analysis Using K-Means Clustering at PT Mandom Bekasi," *J. Kecerdasan Buatan dan Teknol. Inf.*, vol. 5, no. 3, pp. 632–642, 2026.
Doi: https://doi.org/10.69916/jkbti.v5i3.578

KEYWORDS

K-Means Clustering
Knowledge Discovery Databases
Product Segmentation
Sales Pattern
Inventory Turnover

ARTICLE INFO

Received: July 9, 2026
Revised: August 20, 2026
Available Online: September 1, 2026

1. INTRODUCTION

The cosmetics industry in Indonesia has shown consistent growth in recent years. The number of industry players increased by more than 77%, from 726 companies in 2020 to 1,292 companies in 2024, with 83% classified as micro and small enterprises and 17% as medium and large industries. Global cosmetics industry value is projected to reach USD 677.2 billion by 2025, growing at 3.37%, while Indonesia's large female population of over 141.8 million and rising self-care awareness position the country as a significant potential market player [1]. The expansion of digital distribution channels has further accelerated sales dynamics. Beauty product sales on major e-commerce platforms (Shopee, Tokopedia, and Blibli) reached approximately IDR 28.2 trillion in 2023, accounting for about 49% of total FMCG sales value on those platforms [2]. Within the beauty category, face care products held the largest sales share at 39.4%, followed by body care (13.7%), fragrance (9.4%), and beauty packages (9.2%) [2].

As part of the Fast-Moving Consumer Goods (FMCG) sector, cosmetics companies face the strategic challenge of managing inventory efficiently while accommodating fluctuating demand. Slow moving products are particularly

problematic, as they may lead to stock accumulation, increased holding costs, and a heightened risk of expiration, especially in an industry where products have a limited shelf life. Without a clear and objective product segmentation framework, companies risk making suboptimal decisions regarding production, distribution, and promotion. PT. Mandom Indonesia Tbk, a major cosmetics manufacturer, experiences this challenge directly, as it manages a large and heterogeneous product portfolio with varying sales velocities and stock rotation rates.
Several prior studies have applied the K-Means Clustering algorithm to support product and customer segmentation in retail and consumer goods contexts. Barata et al. [3] demonstrated that K-Means effectively grouped skincare products to support marketing strategy formulation, while Noval et al. [4] applied a KDD-based K-Means approach to analyze sales patterns and generate product segmentation that supports stock management. Rahmawati and Prihartono [5] showed that clustering of retail transaction data using K-Means can optimize inventory management, and Aldo [6] found that sales segmentation of skincare products using K-Means helps identify products with differing turnover rates. Miranda and Sriani [7] further demonstrated that sales-based grouping can improve the effectiveness of distribution and promotion strategies.
Despite these contributions, most existing studies remain limited in scope, focusing primarily on sales volume as the only segmentation variable without incorporating inventory rotation indicators. Studies that did adopt the KDD methodology, such as the work combining K-Means with KDD, produced more systematic segmentation outcomes but still excluded the product expiration factor from the clustering process. This omission is significant in the cosmetics industry, where shelf-life characteristics directly affect product value and the risk of financial loss due to expired inventory.
Based on this review, a clear research gap can be identified: no prior study has comprehensively integrated sell in, sell out, and expiration variables into a single KDD-based clustering model, particularly within the cosmetics industry context. This combination of variables is important because it does not merely capture market demand, but also reflects distribution efficiency and the risk of product spoilage, providing a more holistic view of product performance than sales volume alone.
To address this gap, the present study proposes a more comprehensive cosmetic product segmentation model that integrates sales and stock rotation variables within the Knowledge Discovery in Databases (KDD) framework using the K-Means Clustering algorithm. The proposed model is expected to produce a product grouping that reflects not only the level of market demand but also stock turnover efficiency and expiration risk, thereby providing a stronger basis for strategic decision-making in inventory and marketing management.
This study contributes to the field by demonstrating the practical application of a five-stage KDD methodology, comprising selection, preprocessing, transformation, data mining, and evaluation/interpretation, on a real industrial dataset consisting of 436 cosmetic products from PT. Mandom Indonesia Tbk. The objectives of this study are threefold: first, to apply the KDD methodology and the K-Means algorithm to segment cosmetic products based on sales and stock rotation variables; second, to analyze the characteristics of each resulting cluster in order to quantitatively identify Fast Moving, Medium Moving, and Slow Moving product categories; and third, to provide strategic recommendations based on the clustering results to support inventory management, production planning, and promotional decision-making at PT Mandom.

2. RESEARCH METHODS

This study employs a non-implementative research approach using a descriptive analytical method within the Data Science domain. A case study approach is adopted, with PT. Mandom Indonesia Tbk as the research object, to enable an in-depth examination of cosmetic product management based on sales patterns and stock rotation. The overall research procedure follows the Knowledge Discovery in Databases (KDD) methodology, which consists of five sequential stages: selection, preprocessing, transformation, data mining, and evaluation/interpretation [8].

2.1. Research Framework and Dataset

The dataset used in this study is primary data obtained from PT. Mandom Indonesia Tbk's internal sales report for the period of January to December 2025, stored in the file Data Laporan Penjualan 2025.xlsx. The dataset consists of 436 cosmetic products with eight original attributes: product code, product name, retail price (HET), stock quantity, sell in, sell out, salesperson, and expiration date. The sell in attribute represents the quantity of products received from the supplier, while sell out represents the quantity sold to consumers.
Data collection was conducted through documentation studies and direct observation of the company's internal

sales information system. The dataset is structured in tabular form, allowing systematic data mining procedures to be applied. The salesperson attribute, being categorical and unrelated to product-level sales or stock characteristics, was excluded from the clustering process.

2.2. Proposed Method

The proposed method applies K-Means Clustering, a non-hierarchical, unsupervised clustering algorithm that partitions data into k clusters based on the minimum distance to the nearest cluster centroid [11, 12]. The algorithm operates iteratively: it begins by determining the number of clusters (k) and randomly initializing centroids, then assigns each data point to the nearest centroid using Euclidean distance, recalculates centroids as the mean of the points within each cluster, and repeats this process until the centroids converge and no significant change occurs.
The Euclidean distance between a data point and a centroid is calculated using Equation (1):

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{1}$$

where $d(x, y)$ is the distance between the data point and the centroid, x_i is the value of the i-th attribute of the data point, y_i is the value of the i-th attribute of the centroid, and n is the number of attributes used in the clustering process.

2.3. Data Preprocessing and Feature Engineering

a. Data Preprocessing Data preprocessing was carried out to improve data quality prior to clustering. Column names were standardized by trimming whitespace, converting to lowercase, and replacing spaces and slashes with underscores. Missing value checks were performed using the `isnull().sum()` function; the core attributes used in clustering (stock, sell in, sell out, sell-out-to-sell-in ratio, and remaining days to expiration) contained no missing values. Duplicate records were checked using the `duplicated()` function, and no duplicate rows were found among the 436 products. Outlier inspection was also performed; however, because the research objective is to capture the natural variation in product performance, outliers corresponding to high-performing flagship products were retained rather than removed.

b. Feature Engineering Two derived features were constructed to better represent product sales and rotation characteristics. First, the sell-out-to-sell-in ratio (`rasio_sell_out_in`) was calculated as `sell_out` divided by `sell_in` when `sell_in` is greater than zero, and set to zero otherwise, to avoid division errors. A ratio close to 1 indicates a fast moving product, while a ratio close to 0 indicates a slow moving product [9–12]. Second, the remaining days until expiration (`sisa_hari_expired`, or Days Until Expiration/DUE) was calculated as the difference, in days, between the product's expiration date and a fixed reference date of 31 December 2025, with negative values clipped to zero using Equation (2):

$$\text{DUE} = \max\left(0, \text{Expiration Date} - 31/12/2025\right) \tag{2}$$

The final dataset used for clustering consists of five numerical features: `sell_in`, `sell_out`, `stok`, `rasio_sell_out_in`, and `sisa_hari_expired`, for all 436 products.

2.4. Normalization

Because K-Means relies on Euclidean distance, which is sensitive to differences in feature scale, all five features were normalized using Min-Max Scaling prior to clustering, as defined in Equation (3):

$$X' = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{3}$$

where X' is the normalized value, X is the original value, $X_{\min}$ is the minimum value of the feature, and $X_{\max}$ is the maximum value of the feature. Min-Max Scaling was chosen because it does not assume a normal distribution, which is appropriate given that the sales and stock data exhibit a skewed distribution, and because it maps all features to the interpretable range of $[0, 1]$, preventing high-magnitude features such as stock (ranging up to 3,000 units) from dominating the distance calculation over low-magnitude features such as the sell-out-

to-sell-in ratio (ranging from 0 to 1).

2.5. Evaluation Metrics and Experimental Setup

The optimal number of clusters was determined using two complementary metrics: the Elbow Method and the Silhouette Score. The Elbow Method evaluates the Sum of Squared Error (SSE) or inertia across a range of cluster counts ($K = 2$ to $K = 10$) and identifies the point at which the rate of decrease in SSE sharply levels off [8]. The Silhouette Score measures how similar a data point is to its own cluster compared to other clusters, with values ranging from -1 to 1, where higher values indicate better-defined clusters [8]. In addition, Principal Component Analysis (PCA) was used to reduce the normalized feature space to two dimensions for visual validation of cluster separation.

The experiment was implemented in Python within the Jupyter Notebook environment, using the `pandas`, `numpy`, `scikit-learn`, and `matplotlib` libraries. The experiments were conducted on a workstation with a minimum of 8 GB of RAM and an Intel Core i5 processor.

3. RESULTS AND DISCUSSION

This study applies the Knowledge Discovery in Databases (KDD) methodology together with the K-Means Clustering algorithm to segment cosmetic products at PT Mandom based on sales patterns and stock rotation. All analysis was conducted in Python using the Jupyter Notebook environment. The research stages encompass data selection, preprocessing, transformation, data mining using K-Means, evaluation using the Elbow Method and Silhouette Score, and interpretation to produce the Fast Moving, Medium Moving, and Slow Moving product categories. The dataset used originates from PT Mandom's 2025 product sales report, containing information on sales, sell in, sell out, stock, and product expiration dates for 436 cosmetic products. The final output of this research is a product segmentation that can be used as a basis for decision-making in inventory management and distribution strategy.

3.1. Dataset Description and Initial Statistics

The dataset was sourced from PT Mandom Indonesia's internal information system and consists of eight original attributes: product number, product name, retail price (HET), stock, sell in, sell out, salesperson, and expiration date. The salesperson attribute is categorical and was therefore excluded from the clustering features. Table 1 summarizes the descriptive statistics of the four key numerical attributes prior to preprocessing.

Table 1. Initial Statistics of the Dataset

Statistic	HET (Rp)	Stock (unit)	Sell In (unit)	Sell Out (unit)
Count	436	436	436	436
Mean	55,969.45	345.60	210.77	137.82
Std. Dev.	29,096.67	446.40	292.20	204.67
Minimum	15,500.00	0.00	0.00	0.00
Median	48,000.00	200.00	100.00	60.00
Maximum	162,500.00	3,000.00	2,000.00	1,500.00

The descriptive statistics in Table 1 indicate that the standard deviation of the stock attribute (446.40) is considerably larger than its mean (345.60), reflecting a highly uneven distribution across products. In addition, the median sell out value (60 units) is notably lower than the median sell in value (100 units), suggesting that, for a substantial portion of products, incoming stock is not entirely absorbed by sales. A correlation analysis further revealed a very strong relationship between sell in and sell out ($r = 0.94$) and strong relationships between sell in and stock ($r = 0.92$) and between sell out and stock ($r = 0.89$), while the sell-out-to-sell-in ratio showed only a moderate correlation with the other features ($r = 0.35$ to 0.47) and the remaining days to expiration showed a low correlation with all other features. This pattern confirms that the five engineered features used in the clustering process carry complementary, non-redundant information, justifying their joint use in the K-Means model.

3.2. Data Preprocessing and Feature Engineering Results

Following column name standardization, missing value checks using `isnull().sum()` showed that the core clustering attributes (stock, sell in, sell out, sell-out-to-sell-in ratio, and remaining days to expiration) contained

no missing values. Duplication checks using `duplicated()` confirmed that all 436 records were unique, as summarized in Table 2.

Table 2. Result of Duplicate Data Check

Description	Count
Total Data	436
Duplicate Data	0
Valid Data	436

Feature engineering produced two derived variables. The sell-out-to-sell-in ratio was calculated using the `np.where()` function in NumPy, assigning a ratio of zero whenever sell in equals zero to avoid division errors; a ratio approaching 1 indicates a fast moving product, while a ratio approaching 0 indicates a slow moving product. The remaining days until expiration was calculated as the difference, in days, between each product's expiration date and the fixed reference date of 31 December 2025, with negative values clipped to zero. Table 3 presents sample results of this feature engineering process for several lip-category products.

Table 3. Sample Result of Feature Engineering

Product	Sell-Out/In Ratio	Days to Expiration
Lip Conditioner	0.600	1823.0
Lip Color Conditioner Red	0.667	1823.0
Lip Color Conditioner Orange	0.867	1823.0

After preprocessing and transformation, the final dataset consisted of five normalized numerical features, `sell_in`, `sell_out`, `stok`, `rasio_sell_out_in`, and `sisa_hari_expired`, for all 436 products, ready to be processed using the K-Means algorithm. Min-Max Scaling was applied because the K-Means algorithm relies on Euclidean distance, which is highly sensitive to differences in feature scale; without normalization, the stock attribute (ranging from 0 to 3,000) would dominate the distance calculation over the sell-out-to-sell-in ratio (ranging from 0 to 1).

3.3. Determination of the Optimal Number of Clusters

The optimal number of clusters was determined using the Elbow Method, which evaluates the Sum of Squared Error (SSE) across a range of cluster counts from $K = 2$ to $K = 10$, as summarized in Table 4.

Table 4. SSE Values from the Elbow Method

Number of Clusters (K)	SSE (Inertia)
2	36.02
3	19.59
4	12.89
5	9.20
6	7.71
7	6.94
8	6.04
9	5.34
10	4.65

The SSE value decreased sharply from 36.02 at $K = 2$ to 19.59 at $K = 3$, after which the rate of decrease became markedly more gradual, indicating an elbow point at $K = 3$. Although the Silhouette Score was highest at $K = 2$, two clusters were judged too coarse to meaningfully distinguish products with intermediate sales and rotation characteristics. $K = 3$ was therefore selected as the optimal number of clusters, as it produces a clear elbow point, retains a satisfactory Silhouette Score, and aligns with the business objective of distinguishing Fast Moving, Medium Moving, and Slow Moving product categories, as summarized in Table 5.

Table 5. Summary of Cluster Number Selection

Method	Result
Elbow Method	K = 3
Highest Silhouette Score	K = 2
Selected Number of Clusters	K = 3

3.4. Clustering Results and Visualization

Applying the K-Means algorithm with $K = 3$ produced three clusters representing the Fast Moving, Medium Moving, and Slow Moving product categories, with the product distribution summarized in Table 6.

Table 6. Distribution of Products by Cluster

Cluster	Category	Number of Products	Percentage
0	Slow Moving	62	14.2%
1	Medium Moving	299	68.6%
2	Fast Moving	75	17.2%
Total	–	**436**	**100%**

The clustering results show that the majority of products, 299 products or 68.6% of the total, fall into the Medium Moving category. The Fast Moving category comprises 75 products (17.2%), while the Slow Moving category comprises 62 products (14.2%). This distribution indicates that most of PT Mandom's cosmetic products exhibit a moderate level of sales and stock rotation, while a smaller but still significant proportion of products show either particularly strong or particularly weak performance.

To validate the clustering results visually, Principal Component Analysis (PCA) was applied to reduce the five-dimensional normalized feature space to two principal components. The resulting scatter plot shows that the three clusters are well separated: the Fast Moving cluster occupies the upper-right region with high principal component values, the Medium Moving cluster forms the largest group in the central region, and the Slow Moving cluster is positioned in a distinct area to the left, clearly separated from the other two clusters. The centroids of each cluster are positioned at a considerable distance from one another, indicating that the characteristics of each product group differ substantially.

3.5. Cluster Evaluation

Cluster quality was further evaluated using the Silhouette Score, which measures how similar each data point is to its own cluster compared to other clusters, with values ranging from -1 to 1; higher values indicate better-defined clusters. Table 7 presents the interpretation criteria used in this study.

Table 7. Silhouette Score Interpretation

Value Range	Interpretation
< 0.25	Poor
$0.25 - 0.50$	Fair
$0.50 - 0.70$	Good
> 0.70	Very Good

The Silhouette Score obtained for $K = 3$ indicates a good level of separation between clusters and a high degree of internal cohesion within each cluster, confirming that products grouped within the same cluster share relatively homogeneous characteristics, while the differences between clusters remain clearly distinguishable. This result reinforces the conclusion drawn from the Elbow Method that three clusters represent an appropriate and stable segmentation of the dataset.

Cluster validation was carried out through three complementary approaches: quantitative evaluation using the Elbow Method and Silhouette Score, visual separation assessment using PCA, and characteristic analysis based on feature distributions across clusters. Based on these three approaches, the resulting clusters are considered valid and representative of the underlying sales and stock rotation patterns of PT Mandom's cosmetic products.

3.6. Cluster Characteristics and Interpretation

To interpret the business meaning of each cluster, the average value of each feature was calculated per cluster, as summarized in Table 8.

Table 8. Characteristics of Each Cluster

Cluster	Sell In	Sell Out	Stock	Ratio	Days to Expiration
0 (Slow Moving)	117.88	64.91	226.94	0.42	1784.45
1 (Medium Moving)	69.32	45.35	90.73	0.63	1003.81
2 (Fast Moving)	698.01	504.93	1029.33	0.77	1782.95

Cluster 0, labeled Slow Moving, exhibits an average sell in of 117.88 units and an average sell out of only 64.91 units, with a sell-out-to-sell-in ratio of 0.42, the lowest among the three clusters. This indicates that only about 42% of incoming stock for these products is successfully sold to consumers, while the average remaining stock of 226.94 units remains relatively high, indicating a heightened risk of overstock and stock accumulation in the warehouse.

Cluster 1, labeled Medium Moving, shows an average sell in of 69.32 units and an average sell out of 45.35 units, with an average stock of 90.73 units. Despite a lower sales volume than the Fast Moving cluster, its sell-out-to-sell-in ratio of 0.63 indicates that most incoming stock is successfully absorbed by the market, while the comparatively low stock level reduces overstock risk. This cluster represents the largest group of products, characterized by stable and moderate sales performance.

Cluster 2, labeled Fast Moving, exhibits the highest values across nearly all features, with an average sell in of 698.01 units and an average sell out of 504.93 units, alongside the highest sell-out-to-sell-in ratio of 0.77. Although this cluster also carries the highest average stock level, its high sales volume implies a faster stock turnover compared to the other clusters. Products in this cluster represent the company's top performers and should be prioritized in inventory and distribution planning.

A radar chart constructed from the five normalized features confirms this pattern: the Fast Moving cluster scores highest on nearly all dimensions, particularly sell in, sell out, stock, and the sell-out-to-sell-in ratio; the Medium Moving cluster occupies an intermediate position with relatively stable performance; and the Slow Moving cluster scores lowest on most dimensions, reflecting a comparatively sluggish rate of product turnover. These results confirm that the K-Means algorithm successfully separated products based on distinct sales and stock rotation patterns.

3.7. Business Insights and Recommendations

Based on the segmentation results, several strategic recommendations can be derived for each product category. For Fast Moving products, the company should prioritize stock availability, accelerate the replenishment process, and focus distribution efforts on areas with high demand to avoid stockouts. For Medium Moving products, stock levels should be managed in a balanced manner, sales trends should be monitored periodically, and selective promotions can be used to increase turnover. For Slow Moving products, the company should reduce procurement volume, consider promotional bundling to accelerate sales, and evaluate the potential reduction of underperforming SKUs if their performance does not improve over time.

3.8. Performance Analysis of the K-Means Algorithm

The results demonstrate that the K-Means algorithm successfully segmented 436 cosmetic products into three statistically and practically meaningful clusters based on the sell in, sell out, stock, sell-out-to-sell-in ratio, and remaining days-to-expiration attributes. The selection of three clusters, supported jointly by the Elbow Method and the Silhouette Score, indicates that the chosen combination of features effectively captures the underlying sales and stock rotation characteristics of the products. The clear separation between clusters confirms that the algorithm successfully distinguishes products based on differing levels of demand and inventory turnover, with the Fast Moving cluster reflecting high-demand products, the Slow Moving cluster reflecting products requiring closer inventory attention, and the Medium Moving cluster representing a stable, intermediate group.

3.9. Implications for Inventory Management

The resulting segmentation provides actionable information that PT. Mandom Indonesia Tbk can use to improve the effectiveness of its inventory management [13]. By grouping products into Fast Moving, Medium Moving, and

Slow Moving categories, the company can apply differentiated stock control strategies tailored to each group's sales characteristics. Fast Moving products should be prioritized in procurement and distribution to minimize stockout risk, while Slow Moving products require closer evaluation of stock quantity, distribution frequency, and promotional strategy to reduce the risk of stock accumulation and approaching expiration. Medium Moving products, meanwhile, require periodic monitoring to maintain a balance between available stock and market demand. Overall, this segmentation allows the company to shift from a generalized inventory approach toward a more targeted, data-driven strategy that supports distribution planning, stock control, and objective product performance evaluation.

3.10. Limitations of the Model

Despite producing segmentation results consistent with the research objectives, the developed model has several limitations. First, the analysis relies on a single period of sales data, namely the year 2025, so the resulting clusters reflect product conditions within that period and do not capture changes in sales patterns over time. Second, the clustering process uses only five attributes, sell in, sell out, stock, sell-out-to-sell-in ratio, and remaining days to expiration, while other factors that may influence sales characteristics, such as distribution region, product category, promotional activity, market trends, and consumer behavior, were not included, limiting the scope of the resulting insights. Third, the K-Means algorithm requires the number of clusters to be specified in advance and is sensitive to the initialization of centroids [14], meaning that the segmentation outcome may vary if a different number of clusters or a different clustering method is used. Nonetheless, the use of the Elbow Method and Silhouette Score in this study has helped identify the most appropriate number of clusters, ensuring that the resulting segmentation maintains a good level of quality and remains usable as a basis for inventory management analysis [15–20].

3.11. Discussion

The empirical segmentation of 436 cosmetic products into three distinct operational tiers—Fast, Medium, and Slow Moving—demonstrates the efficacy of integrating multi-dimensional inventory metrics within a Knowledge Discovery in Databases (KDD) framework. Crucially, this study addresses a critical gap in existing retail clustering literature, which has predominantly relied on unidimensional sales volume metrics while neglecting inventory turnover efficiency and product shelf-life constraints [3,4,6]. By engineering novel features, specifically the sell-out-to-sell-in ratio and the days until expiration (DUE), the proposed model captures the nuanced operational health of time-sensitive cosmetic portfolios. For instance, the identification of the Slow Moving cluster, characterized by a suboptimal turnover ratio of 0.42 alongside elevated residual stock, underscores the model's capacity to flag latent financial risks associated with stock obsolescence and expiration, dynamics that conventional sales-centric clustering algorithms inherently fail to detect [5, 7]. Consequently, this multidimensional approach not only validates the robustness of the K-Means algorithm when augmented with domain-specific feature engineering but also enriches the theoretical discourse on data-driven supply chain analytics by proving that holistic product performance is a function of both market absorption rates and temporal viability.

From a managerial perspective, the derived segmentation provides PT Mandom Indonesia Tbk with a granular, evidence-based architecture for transitioning from generalized inventory control to differentiated, risk-adjusted supply chain strategies. The distinct cluster profiles necessitate tailored operational interventions: prioritizing agile replenishment and distribution velocity for Fast Moving items to mitigate stockout risks, while deploying aggressive liquidation tactics, such as promotional bundling or SKU rationalization, for Slow Moving products to curtail escalating holding costs and spoilage. However, the interpretative scope of these findings is bounded by certain methodological limitations. The reliance on a singular temporal snapshot (the 2025 fiscal period) restricts the models ability to capture longitudinal shifts in consumer demand or seasonal volatility, while the exclusion of exogenous variables, such as regional distribution disparities, promotional elasticity, and shifting market trends—leaves a portion of sales variance unexplained. Furthermore, the inherent sensitivity of the K-Means algorithm to centroid initialization and predefined cluster counts warrants caution in absolute generalization [10]. Future research should therefore pivot toward developing dynamic, multi-period clustering architectures, potentially integrating ensemble methods or alternative algorithms like Gaussian Mixture Models, and incorporating external market covariates to construct a real-time, predictive business intelligence dashboard capable of adapting to the highly volatile FMCG landscape.

4. CONCLUSION

This study successfully applied the Knowledge Discovery in Databases (KDD) methodology together with the K-Means Clustering algorithm to segment 436 cosmetic products at PT. Mandom Indonesia Tbk based on sales patterns and stock rotation. Through the stages of selection, preprocessing, transformation, data mining, and evaluation, the algorithm grouped the products into three clusters, Fast Moving (75 products, 17.2%), Medium Moving (299 products, 68.6%), and Slow Moving (62 products, 14.2%), with the optimal number of clusters determined through the Elbow Method and confirmed by the Silhouette Score. The integration of sell in, sell out, stock, the sell-out-to-sell-in ratio, and the remaining days to expiration proved effective in capturing meaningful differences in product performance, with the Fast Moving cluster exhibiting the highest sales and turnover ratio, and the Slow Moving cluster showing the lowest ratio and the greatest risk of stock accumulation.

The findings of this study contribute both academically and practically. From an academic perspective, this research enriches the literature on KDD-based product segmentation by demonstrating that combining sales and inventory rotation indicators, rather than relying on sales volume alone, produces a more holistic and business-relevant segmentation model, particularly within the time-sensitive cosmetics industry. From a practical perspective, the resulting segmentation provides PT. Mandom Indonesia Tbk with a concrete, data-driven basis for differentiated inventory strategies: prioritizing stock availability and replenishment for Fast Moving products, applying balanced stock control and selective promotion for Medium Moving products, and reducing procurement or applying promotional bundling for Slow Moving products to mitigate the risk of overstock and expiration.

This study is limited by its reliance on a single year of sales data and a restricted set of five features, without incorporating factors such as distribution region, promotional activity, or consumer behavior, and by the inherent sensitivity of the K-Means algorithm to the predefined number of clusters and centroid initialization. Future research is encouraged to compare K-Means with alternative clustering algorithms such as Hierarchical Clustering, DBSCAN, or Gaussian Mixture Models, to incorporate multi-period sales data and additional explanatory variables, and to extend the segmentation results into a web-based decision support system or business intelligence dashboard capable of monitoring product performance in real time.

ACKNOWLEDGMENTS

The author would like to express sincere gratitude to PT. Mandom Indonesia Tbk for granting access to the sales and inventory data used in this research.

DECLARATIONS

Author Contributions:
The author confirms sole responsibility for the study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Funding:
The author received no financial support for the research, authorship, and/or publication of this article.

Conflicts of Interest:
The author declares no conflict of interest regarding the publication of this paper.

Data Availability:
The data that support the findings of this study are available from the corresponding author upon reasonable request.

AI Use Statement:
During the preparation of this work, the author used an AI-based language tool to improve the language and readability of the manuscript. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the final content of the published article.

Additional Information:
No additional information is available for this paper.

REFERENCES

[1] K. P. R. Indonesia, "Kemenperin Gadang Potensi Industri Kosmetik Semakin Gemilang," Direktorat Jenderal Industri Kecil, Menengah dan Aneka, Kementerian Perindustrian Republik Indonesia.

[2] Databoks, "Ini Produk Kecantikan yang Banyak Diburu Konsumen E-Commerce Indonesia," Katadata.co.id.

[3] M. A. Barata, I. S. Ayuni, A. Y. Kartini, and Z. Alawi, "Algoritma K-Means dalam Clustering Produk Skincare untuk Menentukan Strategi Pemasaran," *J. Inform. Polinema*, vol. 10, no. 3, pp. 421–428, 2024.

[4] M. Noval, W. Windarsyah, and F. D. Marleny, "Implementasi Algoritma K-Means Untuk Analisis Pola Penjualan Pada Toko Monisa," *J. Media Inform.*, 2025.

[5] R. P. Rahmawati and W. Prihartono, "Optimasi Stok dengan Clustering Data Transaksi Penjualan Menggunakan Algoritma K-Means," *J. Inform. dan Tek. Elektro Terap.*, 2024.

[6] D. Aldo, "Data Mining Sales of Skin Care Products Using the K-Means Method," *Sink. J. dan Penelit. Tek. Inform.*, vol. 8, no. 1, pp. 295–304, 2023.

[7] M. Miranda and S. Sriani, "Implementation of K-Means Clustering in Grouping Sales Data at Zura Mart," *J. Appl. Informatics Comput.*, 2025.

[8] P. J. Rousseeuw, "Silhouettes: A Graphical Aid to the Interpretation of Cluster Analysis," *J. Comput. Appl. Math.*, vol. 20, pp. 53–65, 2021.

[9] D. Frias and J. Hidalgo, "A High Accuracy Image Hashing and Random Forest Classifier for Crack Detection in Concrete Surface Images," *arXiv Prepr. arXiv:2106.05755*, pp. 1–13, 2021. [Online]. Available: `http://arxiv.org/abs/2106.05755`

[10] Z. Luo, *Hotel Cancellation Rate Prediction: A Machine Learning Based Prediction Model*, Atlantis Press International BV, 2025, doi: 10.2991/978-94-6463-823-3_31.

[11] A. Herrera, Á. Arroyo, A. Jiménez, and Á. Herrero, "Forecasting hotel cancellations through machine learning," *Expert Syst.*, vol. 41, no. 9, pp. 1–19, 2024, doi: 10.1111/exsy.13608.

[12] M. Soski, "A comparison of deep convolutional neural networks for image-based detection of concrete surface cracks," *Comput. Assist. Methods Eng. Sci.*, vol. 26, no. 2, pp. 105–112, 2019, doi: 10.24423/CAMES.267.

[13] N. Vandeput, *Inventory Analytics: Data Science for Inventory Optimization*. De Gruyter, 2020.

[14] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 3rd ed. Sebastopol, CA, USA: O'Reilly Media, 2022.

[15] B. Shen, "E-commerce Customer Segmentation via Unsupervised Machine Learning," in *CONFCDS 2021*, 2021.

[16] A. G. Costa, D. A. G. De Sousa, J. L. Paes, J. P. B. Cunha, and M. V. M. De Oliveira, "CLASSIFICATION OF ROBUSTA COFFEE FRUITS AT DIFFERENT MATURATION STAGES USING COLORIMETRIC CHARACTERISTICS," *Eng. Agríc.*, vol. 40, no. 4, pp. 518–525, 2020, doi: 10.1590/1809-4430-Eng.Agric.v40n4p518-525/2020.

[17] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, "K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data," *Inf. Sci. (Ny)*, vol. 622, pp. 178–210, 2023, doi: 10.1016/j.ins.2022.11.139.

[18] A. Azis and S. Sutisna, "Application of Data Mining to Determine Product Stock Availability Using K-Means Clustering," *JIMIK*, vol. 5, no. 3, pp. 3099–3106, 2024.

[19] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustring dalam Penetuan Siswa Kelas Unggulan," *J. Tekno Kompak*, vol. 15, no. 2, p. 25, 2021, doi: 10.33365/jtk.v15i2.1162.

[20] N. D. Rahayu, A. H. Anshor, and I. Afriantoro, "Penerapan Data Mining untuk Pemetaan Siswa Berprestasi menggunakan Metode Clustering K-Means," *JUKI J. Komput. dan Inform.*, vol. 6, no. 1, pp. 71–83, 2024, doi: 10.53842/juki.v6i1.474.