Mantik Journal, 7 (4) (2024), ISSN 2685-4236 (Online)

IOCSCIENCE

Published by:Institute of Computer Science (IOCS)

Jurnal Mantik

Journal homepage: www.iocscience.org/ejournal/index.php/mantik

Strategic insights from clustering analysis of essential oil sales in UMKM: A comprehensive study on product types, sizes, couriers, and distribution across Indonesian Provinces

Novianti Madhona Faizah[1], Desyi Erawati [2], Shirlyani [3], Tiwuk Wahyuli Prihandayani [4], Luky Fabrianto[5]

[1]Departement of Computer Science, Universitas Tama Jagakarsa, Jakarta, Indonesia
[2]Departement of Accounting, Universitas Tama Jagakarsa, Jakarta, Indonesia
[3]Departement of Information System, Universitas Tama Jagakarsa, Jakarta, Indonesia
[4]Departement of Information System, Universitas Mercu Buana, Jakarta, Indonesia
[5]Departement of Digital Business, Universitas Nusa Mandiri, Jakarta, Indonesia

ARTICLE INFO

Article history:

Received Jan 07, 2024
Revised Jan 09, 2024
Accepted Jan 20, 2024

Keywords:

Clustering;
Courier;
Essential oil;
Principal Component Analysis;
Silhouette Coefficient.

ABSTRACT

This study explores the sales transactions of a Micro, Small, and Medium Enterprise (UMKM) that sells over 40 types of essential oils, totaling 2305 items sold in 2023. The products, packaged in small bottles (10-50 ml), were distributed to almost every province in Indonesia. The main objective is to cluster the data based on variables such as oil type, bottle size, courier company, and destination province. The elbow method determined an optimal number of clusters (k=4), and the Silhouette Coefficient validated the effectiveness of the clustering (0.7614). To simplify the complex clustering results, Principal Component Analysis (PCA) was used for visualization, providing a clear representation of 5 variables and 4 clusters. This study offers valuable insights for informed decision-making in UMKM's service enhancement and development.

Corresponding Author:

Luky Fabrianto,
Departement of Digital Business,
Universitas Nusa Mandiri,
Jl. Margonda Raya, No. 545, Depok, Jawa Barat, Indonesia, 16424
luky.lfb@nusamandiri.ac.id

1. INTRODUCTION

In Indonesia, there are 40 types of plants that produce essential oils, but only a portion of them is commercially utilized as sources of essential oils (Habib Romadhon & Tanti Kustiari, 2020). On the other hand, Small and Medium Enterprises (UMKM) in Indonesia represent a swift transaction cycle, where their products directly cater to the needs of the community(Pratama et al., 2022). An UMKM operating in the essential oil sales sector offers more than 40 types of oils, providing delivery services to almost every province in Indonesia(Mantik et al., 2023).

In the era of information explosion, data mining approaches have become increasingly relevant to facilitate decision-making in business development(Teknik et al., 2023). Clustering methods stand out as one such approach that can be implemented in this industry, allowing for the categorization of oil types, grouping provinces of purchase, and more, based on their shared characteristics(imron et al., 2020).

From the sales data held by the UMKM, new insights can be derived using data mining methods. This research employs clustering methods with the aim of unearthing valuable information through data analysis(Suwaryo et al., 2023), such as: province clusters related to the demand for preferred types of essential oils, clusters of courier companies based on delivery destinations, and other fresh insights derived from the clustering method. The interpretation of emerging patterns essentially contributes positively to the development of broader market penetration strategies and efficiency in delivery processes(Teichgraeber & Brandt, 2019).

The raw sales data from UMKM presents numerous dimensions even after simplification, such as preprocessing the shipping address based on the province names and adding 'jarak_km' attribute in kilometers (originating from the Jabodetabek area). This situation calls for PCA (Principal Component Analysis) to reduce the dimensionality of the clustered dataset(Bandyopadhyay et al., 2021). This is done to provide a visualization of the clustering results that is more visually accessible(Geladi & Linderholm, 2020).

This research is urgent as it addresses contemporary challenges and opportunities in the essential oil industry, aligning with the need for informed decision-making in the dynamic business landscape of UMKM in Indonesia and contributes to enhancing market penetration strategies and operational efficiency.

2. RESEARCH METHOD

The data used in this study pertains to sales transactions throughout the year 2023. The attributes in the dataset include transaction ID, customer name, transaction date, the type of oil purchased, and others. The UMKM willingly shares its sales data for the purpose of business development or, at the very least, operational efficiency by gaining new insights generated from this research. One of these insights includes the segmentation of destination provinces, selection of courier companies, or other novel insights. The dataset comprises 2305 sold items (rows) and 10 variables (columns). The description of these variables is outlined below: (a) id_penjualan : The unique identifier for each sales transaction. (b) tanggal : The date when the transaction occurred. (c) nama_pelanggan : The name of the buyer. (d) qty : The number of items purchased. (e) bahan : The type of essential oil purchased. (f) vol_ml : The amount of oil bought, measured in milliliters. (g) tot_vol : The overall volume of a specific item purchased. (h) tot_bayar : The total payment made by the buyer. (i) propinsi : The shipping address, utilizing only the province (excluding Jabodetabek). (j) nama_kurir : The name of the shipping company used for the transaction.

Clustering is an unsupervised learning method that groups objects based on the similarity of their characteristic attributes or variables(Abdulhafedh, 2021). It can also be described as a technique used to identify patterns(Shah, n.d.). This research employs a clustering algorithm, followed by dimensionality reduction using Principal Component Analysis (PCA) to identify obtained clusters for easier visualization(*Practical Guide To Principal Component Methods in R: PCA, M(CA), FAMD, MFA ... - Alboukadel KASSAMBARA - Google Books*, n.d.). Broadly speaking, the stages of this research include: (a) Data Collection: Utilizing sales data from the year 2023 (b) Data Preprocessing: Involving the transformation of nominal-valued attributes into numeric ones, converting address attributes into destination provinces, and adding a distance attribute for shipping (in kilometers) (Mehta et al., 2019). (c)Exploratory Data Analysis (EDA): Enhancing the understanding of the data. (d) Determining Optimal Cluster Number: Employing the Elbow method. (f) Implementation of K-Means Algorithm, (g) Clustering Validation: Measuring the proximity between objects using Silhouette score. (h) Visualization of Cluster Results: Leveraging Principal Component Analysis (PCA) to identify clusters. (h) Analysis of Cluster Results: Providing new insights for the UMKM.

3. RESULT AND DISCUSSION

3.1. Data Transformation

The initial stage after collecting the data involves transforming nominal values into numeric ones. This is done to facilitate calculations for the methods to be employed. Table 1 illustrates the transformations utilized.

Tabel 1. Transformasi Data

Material	Id_Material	Courier	Id_Courier	Province	id_ Province	distance _km
Agarwood	1	ANTERAJA	1	Bali	1	1200
Almond	2	GOJEK	2	Bangka Belitung	2	900
Alpinia Galangal	3	GRAB	3	Banten	3	100
Argan	4	IDEXPRESS	4	Bengkulu	4	850
Basil	5	J&T	5	DIY	5	570
Bergamot	6	JNE	6	Gorontalo	6	3000
Black Pepper	7	JOB	7	Jabodetabek	7	0
Cananga	8	LAZADA	8	Jambi	8	810
Cardamom	9	LIONPARSEL	9	Jawa Barat	9	180
Carrier	10	PAXEL	10	Jawa Tengah	10	470
Cedarwood	11	SHOPEEEXP	11	Jawa Timur	11	750
Chamomile	12	SICEPAT	12	Kalimantan Barat	12	1720
Cinnamon	13			Kalimantan Selatan	13	1690
Citronella	14			Kalimantan Tengah	14	1350
Clove Bud	15			Kalimantan Timur	15	2330
Eucalyptus	16			Kalimantan Utara	16	2780
Fennel	17			Kepulauan Riau	17	900
Frangipani	18			Lampung	18	350
Frankincense	19			Maluku	19	2600
Geranium	20			Maluku Utara	20	2500
Ginger	21			NAD	21	2320
Grapefruit	22			NTB	22	1530
Jasmine Sambac	23			NTT	23	2180
Juniper Bery Org	24			Papua	24	3500
Kaafir Lime	25			Papua Barat	25	3000
Lavender	26			Riau	26	1220
Lemon	27			Sulawesi Barat	27	1970
Lemongrass	28			Sulawesi Selatan	28	1830
Magnolia	29			Sulawesi Tengah	29	1710
Mint	30			Sulawesi Tenggara	30	2240
Myrrh	31			Sulawesi Utara	31	3210
Nutmeg	32			Sumatera Barat	32	1280
Orange	33			Sumatera Selatan	33	620
Patchouli	34			Sumatera Utara	34	1230
Rose	35					
Rosemary	36					
Sandalwood	37					
Tea Tree	38					
Turmeric	39					
Veytver	40					

3.2. Exploratory Data Analysis (EDA)

To better understand the data, this research employs Exploratory Data Analysis (EDA)(Sahoo et al., 2019). The stages involved are as follows:

Statistical Summary: Primary summary derived from the data, including minimum and maximum values, quartiles, and standard deviation. This can be observed in Table 2.

Tabel 2. Data descriptions

	id_penjualan	kode_bahan	tot_vol	id_propinsi	jarak_km	id_kurir	cluster
count	2305	2305	2305	2305	2305	2305	2305
mean	202304262236.68	22.70889	14.62689	9.495010	414.9587	4.98047	1.11670
std	2776620.33	9.405405	14.75537	5.931839	541.2902	4.32943	0.637243
min	202301016597	1	5	1	0	1	0
25%	202302016840	15	10	7	0	1	1
50%	202303018442	26	10	9	180	4	1
75%	202306018672	30	20	10	570	11	1
max	202311018001	40	250	29	3210	12	3

Excluding the 'id_penjualan' Attribute: This variable has minimal impact and tends to complicate calculations.

Correlation Matrix: A matrix aiding the visualization of correlations among different variables, as depicted in Figure 1.

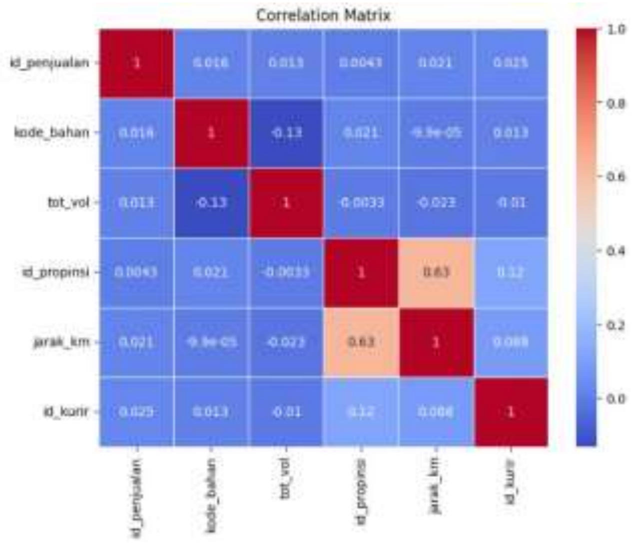

Figure 1. The correlation matrix reveals a relatively strong correlation between the' jarak_km' and 'id_propinsi' variables. This is particularly helpful in dimensionality reduction using Principal Component Analysis (PCA).

3.3. Elbow Method

The method used to assist in determining the optimal number of clusters in the k-means algorithm is known as the Elbow Method(Umargono et al., 2020). Its objective is to identify a point where an increase in the number of clusters no longer provides a significant improvement in variance or a significant decrease in inertia(Nainggolan et al., 2019). Figure 2 depicts a visualization to determine this optimal value. The mathematical formula for inertia (sum squared distance) is as follows.

$$I = \sum_{i=1}^{n} \sum_{j-i}^{k} \left\| x_i - C_j \right\|^2 \quad (1)$$

Where

n = number of data points

k = number of cluster

x_i = data point number i

$\left\| x_i - C_j \right\|^2$ = squared Euclidean distance between x_i and C_j

The formula for Euclidean distance is, $d(a,b) = \sqrt{\sum_{i=1}^{n}(a_i - c_j)^2}$ (2)
Where
n = number of dimensions
a_i = the *i*-th element of vector a
b_i = the *i*-th element of vector b

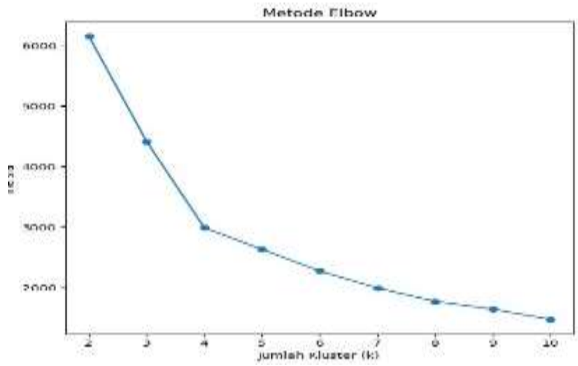

Figure 2. The optimal k value is 4

Once the optimal value for k is obtained, in this study k=4, the next step involves normalizing the data. Normalization transforms the values of variables in the dataset, ensuring that these variables have a similar scale.

3.4. K-Means Clustering

The primary goal of this algorithm is to group data into clusters in such a way that points within one cluster have high similarity, while points between clusters have lower similarity(Surohman et al., 2021). The K-means formula with Euclidean distance can be seen in formula number (2) in this paper (Breiding et al., 2022). The steps in this algorithm are as follows: (a) Initialize Centroids: Choose k points as the initial centroids of clusters. (b) Assign Data to Clusters: Calculate the distance between each data point and centroids. (c) Update Centroids: Compute the average of each cluster and determine the new centroids. (d) Repeat Steps 2 and 3: Repeat the assignment and centroid update steps until there is no change. (e) End: When the centroids no longer change or the maximum number of iterations is reached, the algorithm is considered complete.

After obtaining the optimal k value using the Elbow method (k=4), the K-means algorithm is executed. Figure 3 shows the distribution results for each cluster.

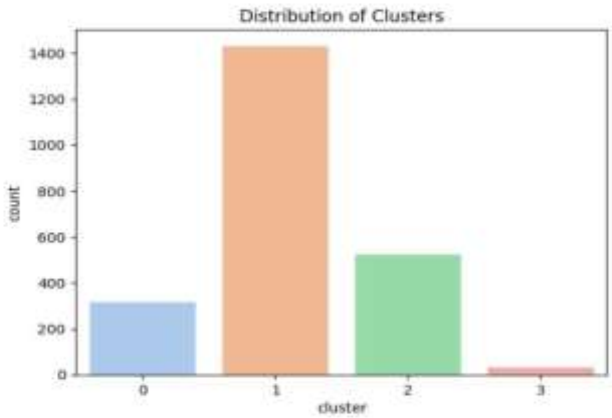

Figure 3. Distribution of each cluster

3.5. The Validation of K-means Clustering

The Silhouette Coefficient is an evaluation metric used to measure how well an object has been placed in a cluster based on its proximity to other objects within the same cluster and how separated it is from objects in neighboring clusters(Shahapure & Nicholas, 2020)(SAPUTRA et al., 2020). Figure 4 visualizes the results of the Silhouette Coefficient, and the obtained Silhouette Coefficient value is 0.7614.

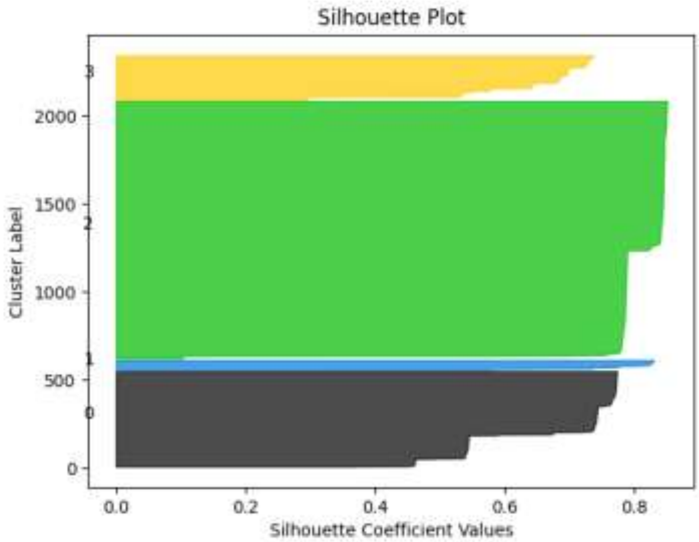

Figure 4. Silhouette Coefficient visualization between clusters

3.6. Visualization of cluster results using PCA

After obtaining the clustering results for each object, the visualization was performed using Principal Component Analysis (PCA). PCA is a statistical technique that can be used for dimensionality reduction, feature extraction, and data visualization(Gewers et al., 2021). It can analyze data, identify patterns, and reduce the dimension of the dataset with minimal loss of information(Abid et al., n.d.). PCA provides an effective and efficient solution for data visualization(Mohammed et al., 2021). In this study, PCA was utilized for two-dimensional visualization, with the clustered dataset transformed into two Principal Components.

In Figure 1, the correlation value between the variables 'id_propinsi' and 'jarak_km' indicates a covariance of 63%, while Table 3 displays the Eigenvalues and weights obtained from each attribute. Additionally, Table 4 presents the autovector matrix from PCA. In this study, the proportion of variance for PC1 is 0.3317, and for PC2, it is 0.2260.

Tabel 3. Eigenvalue and weight

Fitur	PC1 Bobot	PC2 Bobot	Eigen Value
Kode Bahan	0.0337	-0.7055	1.6597
Total Volume	-0.0376	0.7038	0.3730
ID Propinsi	0.692	0.0432	1.1308
Jarak (km)	0.6871	0.0467	0.8702
ID Kurir	0.2158	0.0544	0.9685

Tabel 4. Autovector Matrix

	PC1				
PC2	0.0337	-0.0293	0.7055	0.7068	0.2158
	-0.0293	-0.0287	-0.7038	0.7056	0.0678
	0.7055	-0.7038	-0.0432	0.0347	-0.1254
	0.7068	0.7056	-0.0467	-0.0218	-0.1746

-0.0267	0.0678	-0.1254	-0.1746	0.9739

Principal Component 1 (PC1) depicts variables that contribute significantly and positively to PC1, including 'ID Province', 'Distance (km)', and 'ID Courier'. PC1 reflects patterns related to 'province', 'distance', and 'courier'. PC2 depicts that the variable 'Total Volume' contributes significantly and positively to PC2. PC2 reflects patterns related to 'volume' and 'material type'. Positive or negative weights indicate the direction of variable contribution to the principal component. The larger the absolute weight value, the greater the contribution.

In PCA, it is also necessary to calculate the eigenvectors and eigenvalues of the covariance matrix(Tony Cai et al., 2020). Eigenvectors indicate the direction with maximum variance, while eigenvalues provide information about the amount of variance explained by each eigenvector(*Principal Component Analysis PROF XIAOHUI XIE SPRING 2019 CS 273P Machine Learning and Data Mining*, n.d.). Eigenvectors are vectors that show the direction of each principal component, and the length of the eigenvector is not very relevant; what matters is the direction(Björklund, 2019). Eigenvectors depict the linear combinations of the original features that form each principal component. Figure 13 is a heatmap of the eigenvector matrix, while Figure 14 is a Biplot that illustrates the direction of variables.

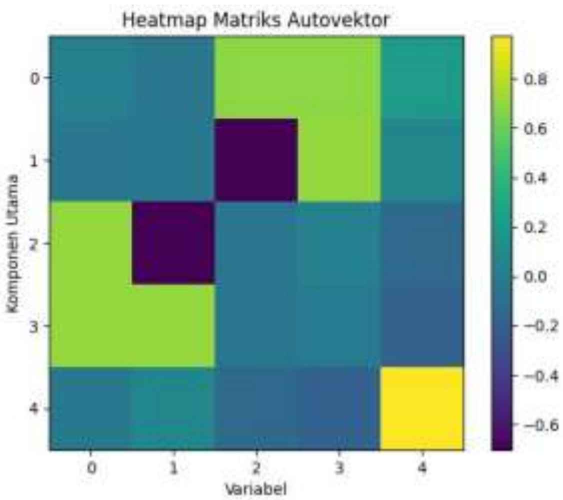

Figure 5. Heatmap Autovektor matrix

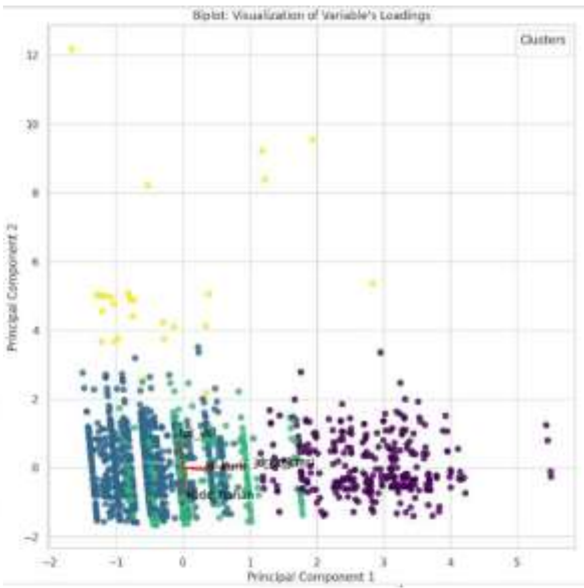

Figure 6. Biplots

From Figure 5, it can be observed that variables 'id_kurir' and 'jarak_km' point in the same direction, while variables 'tot_vol' and 'id_bahan' point in the opposite direction. Arrows approaching the same direction indicate a positive relationship between the variable and PC, while arrows approaching the opposite direction indicate a negative relationship.

Based on the explanations above, the author considers PCA to be quite effective for visualizing the clustering results, which consist of k=4 clusters and involve 5 variables. Figure 7 presents the clustered results displayed with PCA.

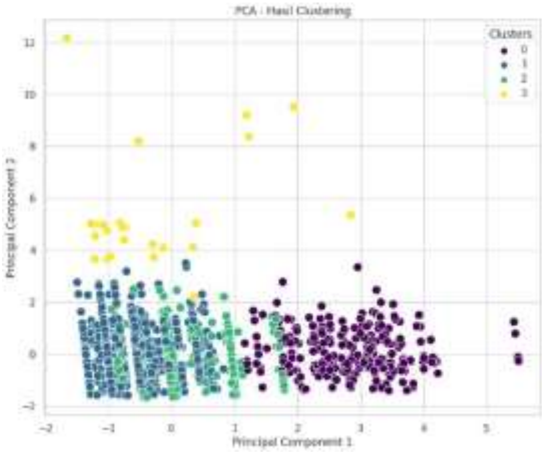

Figure 7. Visualization of cluster results

3.7. Analysis of cluster results

Here are the characteristics of each cluster (a) Cluster 0 consists of 1430 objects, with the majority of shipments going to provinces in Java. There are only 40 shipments to Bali and Lampung. (b) Cluster 1 consists of 319 objects, with all shipments going to provinces outside of Java, and the total volume of oil not exceeding 50 ml. (c) Cluster 2 consists of 528 objects, with all shipments going to provinces in Java and Sumatra, using less popular couriers in this study, such as JOB, LAZADA, SHOPEEEXP. (d) Cluster 3 consists of 32 objects, shipped to almost all provinces, with total volumes ranging from 80 to 250 ml for each object.

4. CONCLUSIONS

Clustering using the K-Means algorithm is effective in segmenting various variables, but in this study, the variables Province and Delivery Distance have a significant covariance, impacting the clustering results.

The use of PCA is highly effective for visualizing clustering results that involve 5 variables (Material, Total Volume, Province, Distance, and Courier). However, in Biplots, only 2 variables (Courier and Distance) have vectors in the same direction, while Total Volume and Material have vectors in opposite directions. The Elbow method is also useful for determining the optimal number of clusters, as evidenced by the excellent Silhouette Coefficient value.

Overall, the research provides practical guidance for UMKM and similar enterprises looking to leverage data-driven approaches for business development and customer relationship management. The significant covariance observed between Province and Delivery Distance may introduce bias in clustering results. Future research could explore advanced techniques or alternative clustering algorithms that account for such dependencies. To address these limitations, future research could explore advanced clustering methodologies, conduct in-depth cluster profiling, incorporate additional relevant variables, and consider external contextual factors for more holistic perspective on UMKM in the essential oil industry.

REFERENCES

Abdulhafedh, A. (2021). Incorporating K-means, Hierarchical Clustering and PCA in Customer Segmentation. *Journal of City and Development*, *3*(1), 12–30. https://doi.org/10.12691/jcd-3-1-3

Abid, A., Zhang, M. J., Bagaria, V. K., & Zou, J. (n.d.). *Exploring patterns enriched in a dataset with* *contrastive principal component analysis*. https://doi.org/10.1038/s41467-018-04608-8

Bandyopadhyay, S., Thakur, S. S., & Mandal, J. K. (2021). Product recommendation for e-commerce business by applying principal component analysis (PCA) and K-means clustering: benefit for the society. *Innovations in Systems and Software Engineering*, *17*(1), 45–52. https://doi.org/10.1007/S11334-020-00372-5/METRICS

Björklund, M. (2019). Be careful with your principal components. *Evolution*, *73*(10), 2151–2158. https://doi.org/10.1111/EVO.13835

Breiding, P., Sottile, · F, Woodcock, · J, Breiding, B. P., Sottile, F., & Woodcock, J. (2022). Euclidean Distance Degree and Mixed Volume. *Foundations of Computational Mathematics*, *22*, 1743–1765. https://doi.org/10.1007/s10208-021-09534-8

Geladi, P., & Linderholm, J. (2020). 2.03 - Principal Component Analysis. *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis, Second Edition: Four Volume Set*, *2*, 17–37. https://doi.org/10.1016/B978-0-12-409547-2.14892-9

Gewers, F. L., Ferreira, G. R., De Arruda, H. F., Comin, C. H., Amancio, D. R., Da, L., & Costa, F. (2021). *Principal Component Analysis: A Natural Approach to Data Exploration*. https://doi.org/10.1145/3447755

Habib Romadhon, C., & Tanti Kustiari, dan. (2020). PENGEMBANGAN USAHA MINYAK ATSIRI KABUPATEN JEMBER DENGAN METODE DECISION SUPPORT SYSTEM (DSS). *Prosiding Seminar Nasional Terapan Riset Inovatif (SENTRINOV)*, *6*(2), 122–130. https://proceeding.isas.or.id/index.php/sentrinov/article/view/430

imron, M., Hasanah, U., & Humaidi, B. (2020). Analysis of Data Mining Using K-Means Clustering Algorithm for Product Grouping. *International Journal of Informatics and Information Systems*, *3*(1), 12–22. http://ijiis.org/index.php/IJIIS/article/view/3

Mantik, J., Fabrianto, L., Prasetyo, J. H., Faizah, N. M., & Solichatun, S. (2023). Inventory management for essential oil UMKM: enhancing business performance with data mining. *Jurnal Mantik*, *7*(2), 702–711. https://doi.org/10.35335/MANTIK.V7I2.3909

Mehta, H., Kanani, P., & Lande, P. (2019). Google Maps. *Google Maps Article in International Journal of Computer Applications*, *178*(8), 975–8887. https://doi.org/10.5120/ijca2019918791

Mohammed, B., Hasan, S., & Mohsin Abdulazeez, A. (2021). A Review of Principal Component Analysis Algorithm for Dimensionality Reduction. *Journal of Soft Computing and Data Mining*, *2*(1), 20–30. https://doi.org/10.30880/jscdm.2021.02.01.003

Nainggolan, R., Perangin-Angin, R., Simarmata, E., & Tarigan, A. F. (2019). Improved the Performance of the K-Means Cluster Using the Sum of Squared Error (SSE) optimized by using the Elbow Method. *Journal of Physics: Conference Series*, *1361*(1). https://doi.org/10.1088/1742-6596/1361/1/012015

Practical Guide To Principal Component Methods in R: PCA, M(CA), FAMD, MFA ... - Alboukadel KASSAMBARA - Google Books. (n.d.). Retrieved January 4, 2024, from https://books.google.co.id/books?hl=en&lr=&id=eFEyDwAAQBAJ&oi=fnd&pg=PR5&dq=Practical+Guide+to+Principal+Component+Methods+in+R&ots=reV_kNjDPB&sig=jmZmoxWflxkjURdEgZ6S6bLp-OM&redir_esc=y#v=onepage&q=Practical Guide to Principal Component Methods in R&f=false

Pratama, R. A. R. J., Ningtiyas, A. F., Mauludin, H., Larosa, M. C., & Thaariq, K. A. (2022). Meningkatkan Potensi Desa Pagerwangi Melalui Produktivitas dengan Pengembangan UMKM sebagai Wujud Ekonomi Kreatif Desa. *Jurnal Manajemen Dan Bisnis Performa*, *19*(2), 92–98. https://doi.org/10.29313/performa.v19i2.10703

Principal Component Analysis PROF XIAOHUI XIE SPRING 2019 CS 273P Machine Learning and Data Mining. (n.d.).

Sahoo, K., Samal, A. K., Pramanik, J., & Pani, S. K. (2019). Exploratory data analysis using python. *International Journal of Innovative Technology and Exploring Engineering*, *8*(12), 4727–4735. https://doi.org/10.35940/ijitee.L3591.1081219

SAPUTRA, D. M., SAPUTRA, D., & OSWARI, L. D. (2020). *Effect of Distance Metrics in Determining K-Value in K-Means Clustering Using Elbow and Silhouette Method*. *172*, 341–346. https://doi.org/10.2991/AISR.K.200424.051

Shah, C. (n.d.). *A Hands-On Introduction to Data Science Preface About the Author A ckn owledgm ents*.

Shahapure, K. R., & Nicholas, C. (2020). Cluster quality analysis using silhouette score. *Proceedings - 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA 2020*, 747–748. https://doi.org/10.1109/DSAA49011.2020.00096

Surohman, S., Fabrianto, L., Riza, F., & Faizah, N. M. (2021). Korelasi Antara Profil dan Nilai

Akademis Siswa dengan Menggunakan Algoritma K-Means. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *8*(4), 845. https://doi.org/10.25126/jtiik.2021843034
Suwaryo, N., Rahman, A., Marini, D., Atmaja, U., & Basri, A. (2023). Klasterisasi Stok Produk Retail Untuk Menetukan Pergerakan Kebutuhan Konsumen Dengan Algoritma K-Means. *Bulletin of Information Technology (BIT)*, *4*(3), 306–312. https://doi.org/10.47065/BIT.V4I3.736
Teichgraeber, H., & Brandt, A. R. (2019). Clustering methods to find representative periods for the optimization of energy systems: An initial framework and comparison. *Applied Energy*, *239*, 1283–1293. https://doi.org/10.1016/J.APENERGY.2019.02.012
Teknik, J., Medicom, I. C. I. T., Ramadani, S., & Hidayat, S. (2023). *Application of data mining on inventory grouping using clustering method*. *15*(5), 228–239.
Tony Cai, T., Han, X., & Pan, G. (2020). Limiting laws for divergent spiked eigenvalues and largest nonspiked eigenvalue of sample covariance matrices. *Https://Doi.Org/10.1214/18-AOS1798*, *48*(3), 1255–1280. https://doi.org/10.1214/18-AOS1798
Umargono, E., Suseno, J. E., & Gunawan, S. . V. (2020). *K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula*. 121–129. https://doi.org/10.2991/ASSEHR.K.201010.019