Published *online* on the journal's web page : http://innovatics.unsil.ac.id

Innovation in Research of Informatics (INNOVATICS)

| ISSN (Online) 2656-8993 |

Unveiling Culinary Patterns: Implementation of K-Means Clustering Algorithm on Food Products in Cafes

Lasmi Lasmini Gumelar[1], Ruuhwan[2], Missi Hikmatyar[3]

[1,2,3]Department of Informatics, Perjuangan University, Tasikmalaya 46115, Indonesia
[1]lasmigumelar21@gmail.com, [2]ruuhwan@unper.ac.id, [3]missi@unper.ac.id

ARTICLE INFORMATION

Article History:
Received: September 30, 2023
Last Revision: November 13, 2023
Published *Online*: November 15, 2023

KEYWORDS

Barcode,
Clustering,
K-Means,
Rapidminer,
Food Products,
Data Mining

CORRESPONDENCE

Phone: +6285283926098

E-mail: lasmigumelar21@gmail.com

ABSTRACT

Barcode Se'i and Coffee is one of the cafes on JL. Major Utarya, No. 48, Empangsari, District. Tawang, Tasikmalaya. Barcode Se'i and Coffee is quite famous because the concept of the place is nice, comfortable, and instagrammable. Not only that, but the Barcode café was also the first to create cow sei in Tasikmalaya. By analyzing the cafe menu groupings, information can be found regarding the level of menu sales. This type of analysis, capable of assessing sales levels, involves the use of data mining techniques such as clustering. Data mining is a data processing stage that aims to identify and extract patterns from a certain set of data. One of the methods included in data mining is the clustering technique. Reclassification techniques are used to group objects into several groups based on observed indicators, ensuring that all objects have a significant level of similarity compared to objects placed in different groups. With Rapidminer software and using the k-means algorithm with sales data for 11 months with the calculations carried out producing 5 clusters. Based on the comparison results of 3 K-Means algorithms with different K values, namely 3, 4, 5, the result from Davies Bouldin with a value close to 0 is a value with K 5, with the result from Davies Bouldin being - 0.912.

1. INTRODUCTION

Barcode Se'i and Coffee is quite famous in Tasikmalaya because the concept of the place is nice, comfortable, and instagrammable. Not only that, but Barcode Café was also the first to create cow sei in Tasikmalaya. This cafe offers a wide selection of food and drink menus. However, the drink menu at this café has more variants than the food menu, so sales of the drink menu at this café are superior. So, the problem with Café Barcode lies in the gap between the marketing strategy and the inappropriate consumer profile. This is because the Barcode café does not yet understand the characteristics of its consumers. This challenge can be seen from transaction data for one year which experienced fluctuations. Thus, excessive stockpiling of food ingredients often occurs.

Information regarding menu sales levels can be found through analysis of menu groupings in a café [1]. This analysis process uses data mining grouping methods, especially the clustering method. Data mining is the process of processing data to recognize and reveal patterns in a group of data [2]. One of the techniques in data mining is the clustering technique [3]. This method aims to categorize entities into several groups based on observed indicators. This allows objects in similar groups to show significant similarities, compared to entities in other groups. Clustering analysis refers to the approach used to group data sets into groups based on the degree of similarity between them. An example of one well-known clustering algorithm is the K-means algorithm, which is known for its ease of use [4].

In this research, the method used is K-Means Clustering Algorithm. It serves as a dynamic tool for uncovering patterns, optimizing menus, and enhancing the overall efficiency of the business. As the cafe industry continues to evolve, leveraging advanced analytical techniques like K-Means clustering becomes increasingly essential for staying competitive and meeting the ever-changing demands of customers.

2. RELATED WORK

Hotel "X" has a poor marketing strategy that affects its occupancy rate. A study was conducted to determine the marketing strategy using historical data on gender, room type, travel companions, length of stay, price, area of origin, and booking method. The study [5] obtained three clusters: online-low customer, medium customer, and offline-manual customer. The results were analyzed using the K-means clustering method with the tools of SPSS.

Another study [6] explain the Leppangeng Cell Counter sells various digital products, but sales data is not well structured and cannot be used for marketing strategies. Data mining using K-Means method can be applied to determine sales of products. The method groups digital product stock data and produces 114 products that sell, 5 underperforming, and 14 not sold. Rapidminer uses this method to determine which products have high and low demand. Research [7] was conducted on food stalls, restaurants, and cafes in Batubara Regency using the K-Means Clustering algorithm to collect data on the number of places to eat in each sub-district in 2016, 2017, and 2021. The goal is to help tourists find culinary places and where to eat based on tourist attractions in the district. The research was based on practical work activities held for a month and used to see the economic development and culinary places in the area.

Data mining using the K-Means method is necessary to determine which clothing sales are best-selling, bestselling, and not selling [8]. The Helai store [9] specializes in selling hijab clothing, but the data on sales and unexpected expenses is not organized properly. The method involves grouping stock data and selecting 3 clusters randomly as the initial centroid. The data is then processed and formed K- means by inserting the stock data of the product and connecting it to Rapidminer Tools [10].

Another research [11] describe e-commerce which rapidly developing in various industries, including business. Online shops use E-Commerce for marketing or trading operations. To know consumer interest in a product, sales transactions can be collected. A study found that the most optimal number of clusters for frozen food product data is two. The study aims to apply the k- means clustering method to group frozen food sales based on consumer interest. The application of the K-Means clustering method to frozen food sales transactions stands as a pivotal initiative in unraveling the intricate dynamics of consumer interest [12].

Based on research related the versatility of the K-Means Algorithm extends beyond these examples, finding applications in fields as diverse as image processing, social network analysis, and environmental monitoring. Researchers and practitioners alike continue to explore and adapt the algorithm to address emerging challenges and capitalize on its strengths in uncovering hidden patterns within complex datasets. As the landscape of data mining evolves, the K-Means Algorithm remains a cornerstone in the toolkit of machine learning and data science professionals. Its adaptability and efficiency make it an asset for gaining actionable insights from large and intricate datasets across a spectrum of disciplines, contributing to advancements in knowledge discovery and decision-making processes.

3. METHODOLOGY

This research has flow stages, as shown in Figure 1 below for this reaserch.

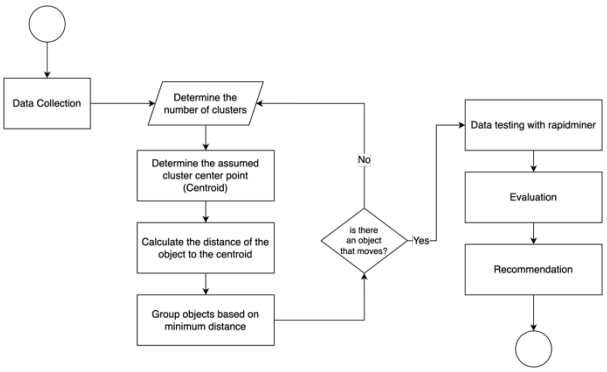

Figure 1. Methodology

3.1 Data Collection

The process of searching for data in the field that will be used to answer research problems by collecting research data obtained from the cashier system at Café Barcode in the form of sales data. The data obtained also contains several data attributes that will influence the clustering process using the k-means algorithm such as price, taste, and ingredients.

3.2 Determine the number of Clusters

The process for determining the number of clusters is carried out by comparing several centroids with the Davies Bouldin Index value, namely for Centroid 1 the number k=3, Centroid 2 the number k=4, Centroid 3 the number k=5. Then take the smallest dbi value among the clustering's to become Calculation reference using Excel.

3.3 Determine The Cluster Center Point (Centroid)

To determine the cluster center point (centorid) it is chosen randomly. the random selection of cluster center points (centroids) in the K-Means Clustering Algorithm serves as a catalyst for exploration and adaptability. This element of randomness contributes to the algorithm's ability to discover meaningful patterns within diverse datasets, making it a powerful tool for unsupervised machine learning in various fields.

3.4 Calculate The Distance Centroid

In simple terms, distance calculation for the K-means algorithm starts from the following stage:

a. Select the cluster center point (centroid).
b. Calculate the distance between the data and the centroid using the formula from Euclidean Distance.

$$\boldsymbol{D\,(a,b) = \sqrt{\sum_{k=1}^{n}(a_k - b_k)^n; i = 1,2,3,\ldots n}}$$

c. Update the centroid center value.
d. Do steps 2 and 3 repeatedly until the value of the centroid center point does not change any more.

3.5 Data Testing with Rapidminer

The data testing process uses the Rapidminer application. By importing data into the Rapidminer application and selecting the data type role according to the attributes to be processed.

3.6 Evaluation

The purpose of evaluating results is to use program findings as a basis for the choices to be made. This procedure attempts to assess the effectiveness of an activity or program and determine the success of a program.

3.7 Recommendation

After going through the data mining stages, the research results will be recommended to the owner as a reference for market segmentation which aims to assist in meeting customer needs, increasing attractiveness for customers, helping to organize the products provided by the company, and opening opportunities for business growth.

4. RESULT AND DISCUSSION

The research highlights the transformative impact of implementing the K-Means Clustering Algorithm on food products in cafes. Here below describe the point of the research result.

4.1 Data Collection

Barcode cafe owners were interviewed for data collection for this research. Sales data from December 2021 to October 2022 will be used as shown in Table 1 and 2, and the K-means method will be used to group the data.

TABLE 1. DATA FOOD TYPE AND PRODUCT

Food Type	Product	Qty	Price
Ayam Bakar Paha	Angkringan	680	IDR. 10.000
Ayam Bakar Sayap	Angkringan	620	IDR. 7.000
Ayam Taliwang	Angkringan	169	IDR. 46.000
Brulee Bomb	Barcode	1282	IDR. 15.000
Ceker Mercon	Angkringan	330	IDR. 7.000
…	…	…	…
Siomay Barcode	Barcode	1181	IDR. 22.000
Sosis Merah	Angkringan	153	IDR. 5.000
Sphagetti Brulee	Barcode	237	IDR. 20.000

TABLE 2. DETERMINATION OF CLUSTER POINT

Centroid	Food	P	K	J	H	R	B
C1	Churros	2	3	6	2	7	2
C2	Roasted Rice	1	1	5	1	3	2
C3	Liver Satay	1	1	5	1	1	1
C4	Sei Chicken	2	1	6	2	4	1
C5	Spaghetti Brulee	2	2	2	3	2	2

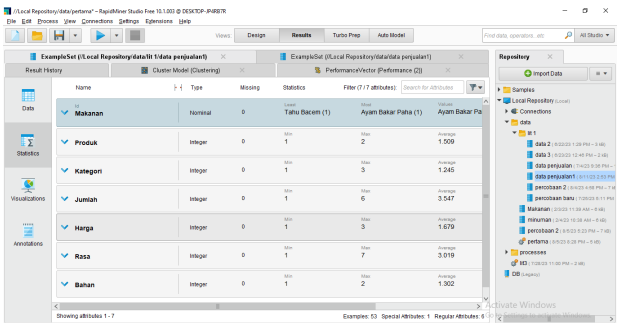

FIGURE 2. DATA ANALYSIS

This data analysis in Figure 2 step aims to remove incorrect or empty entries, resolve missing data, and verify each record for duplication. Inconsistent data tends to have interference (noise).

4.2 Determination of the Number of Clusters

After comparing the Davies Bouldin Index centroid values, the smallest Davies Bouldin Index value is at centroid 3-5 with the number k = 3-5 as shown in the Figure 3-5.

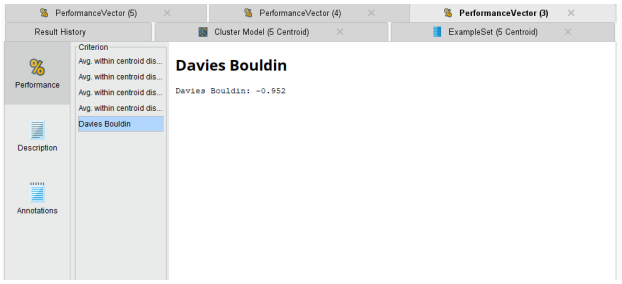

FIGURE 3. DBI CENTROID 3 VALUE

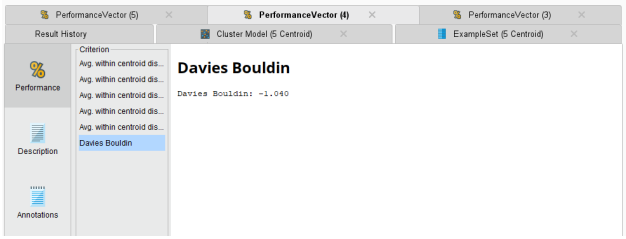

FIGURE 4. DBI CENTROID 4 VALUE

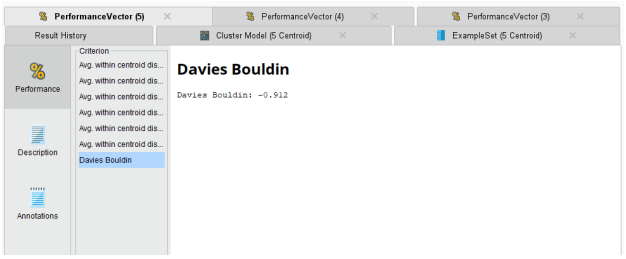

FIGURE 5. DBI CENTROID 5 VALUE

4.3 Data Testing with RapidMiner

Data testing was carried out with the RapidMiner application. The following steps are carried out during testing using the RapidMiner application: Data import is the stage of inputting a dataset from the specified excel file into Rapidminer. This data is data that has previously been carried out in the preprocessing stage and is ready to be processed using the k-means clustering algorithm.

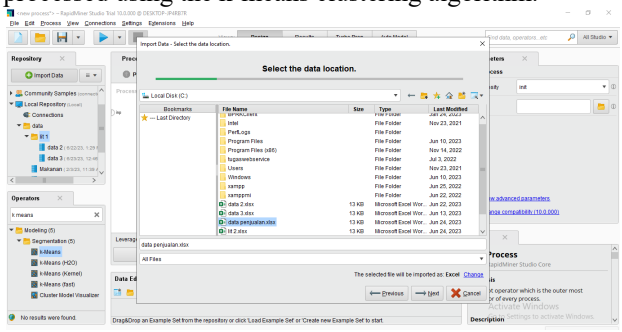

FIGURE 6. DATA TESTING

After the import process, the column format process then selects roles and data types that match the attributes to be processed. In this research, the food column changes to ID, while the number attribute data type becomes integer.

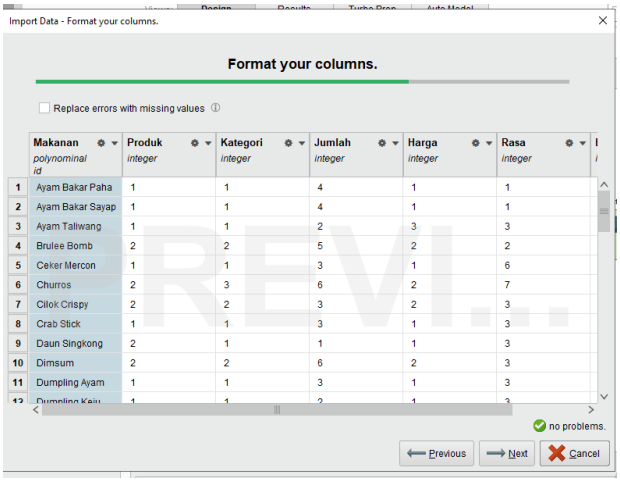

FIGURE 7. FORMAT COLUMNS

As shown in the following graphic, the following step is to drag and drop the parameters you want to process as shown in the image below:

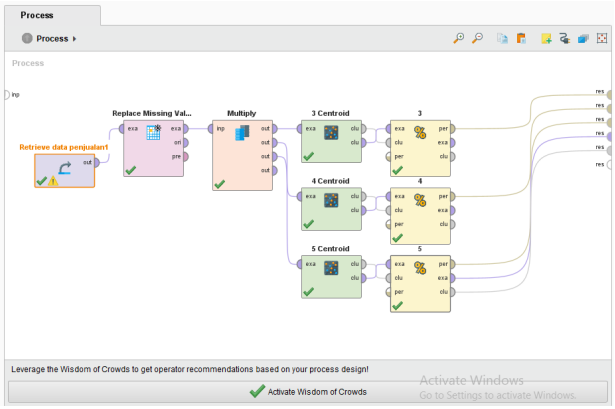

FIGURE 8. DETAILS PARAMETER DISPLAY

This stage involves the following processes:

1. Read Excel
   The process of inputting a dataset in the form of a file with the extension .xls into Rapidminer. This data has gone through the previous data preprocessing stage and is ready to be processed by the k-means clustering algorithm.
2. Replace Missing Values
   Replace missing values by replacing or arranging missing values such as using average values or other appropriate values.
3. Multiply
   This operator is used to multiply two or more columns of numbers in a dataset. For example, if you have a dataset with columns containing numbers, the multiply operator can be used to multiply the values in these columns.
4. Clustering
   The process of grouping data into similar groups based on similar attributes. Clustering can help identify patterns or groups that may not be apparent on initial viewing of the data. Following are the clustering parameters:

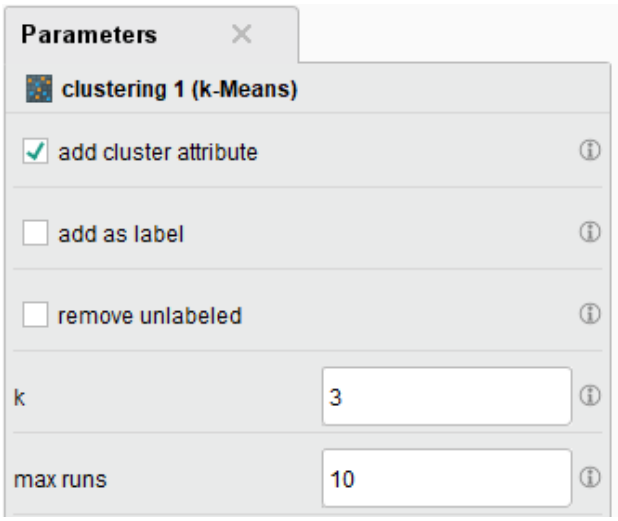

FIGURE 9. CLUSTERING PARAMETER 1 WITH THE NUMBER K=3

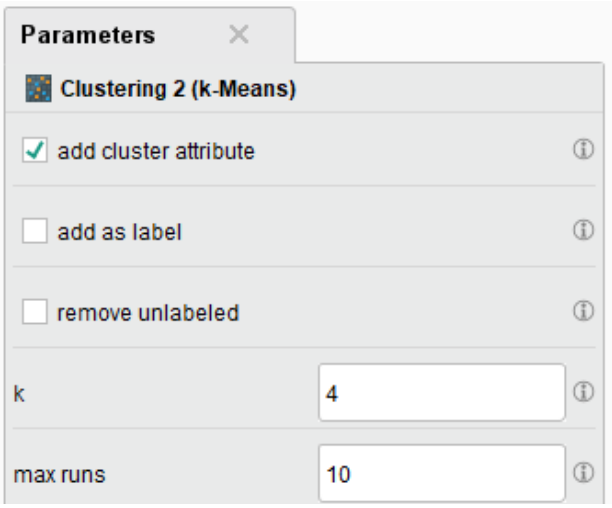

FIGURE 10. CLUSTERING PARAMETER 2 WITH THE NUMBER K=4

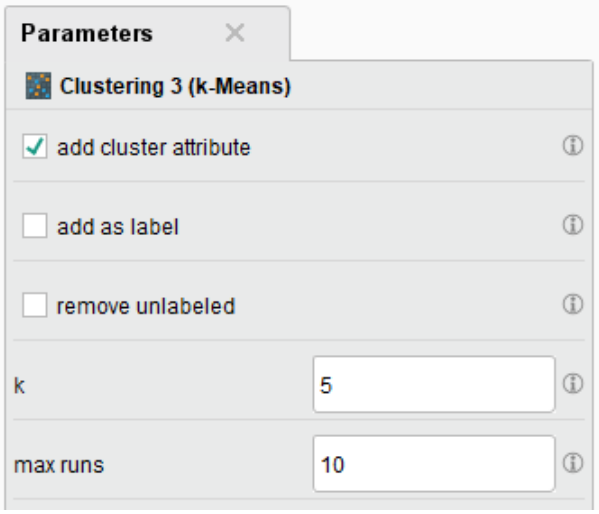

FIGURE 11. CLUSTERING PARAMETER 3 WITH THE NUMBER K=5

5. Performance
   It is the process of measuring the extent to which the model or data analysis process created works well and meets the objectives.

4.4 Evaluation

After completing all procedures, the results are displayed as follows:

1. Example Set

This result displays the cluster results in the form of a data display. The data view describes all the data that has been entered.

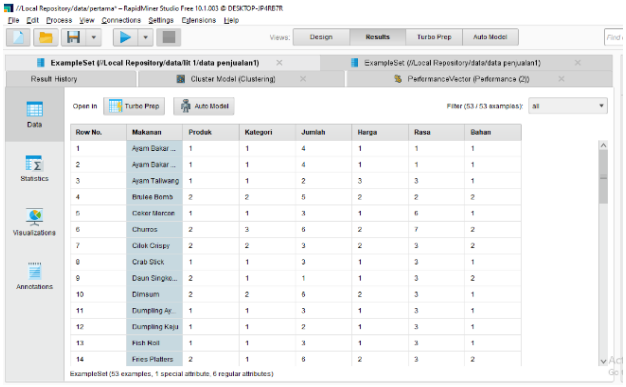

FIGURE 11. EXAMPLE SET

2. Scatter 3D

3D scatter is a type of data visualization that is used to display 3-dimensional data in a 3D scatter graph. Each point represents one entity or data point in 3-dimensional space. The X, Y and Z axes represent 3 different dimensions and the location of each point on the graph shows the values. from that dimension. Sactter 3D is used to visualize between 3 variables in one view.

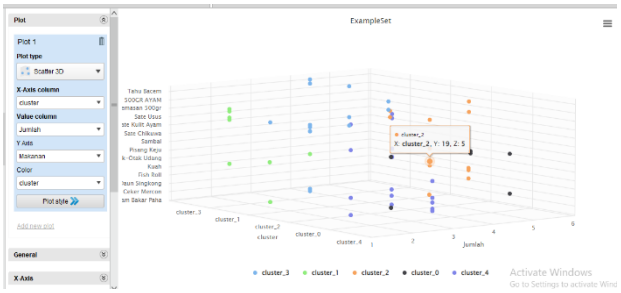

FIGURE 12. SCATTER 3D

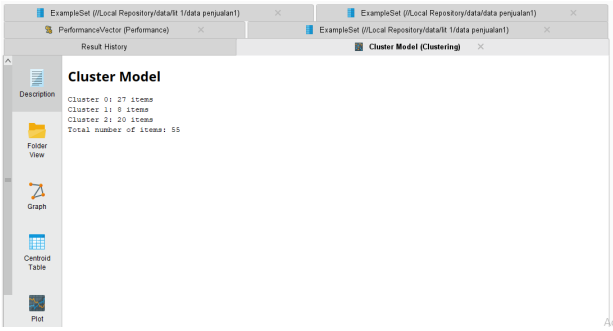

FIGURE 13. TEST VIEW CLUSTER MODEL

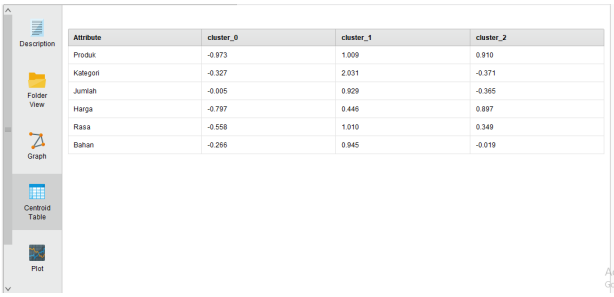

FIGURE 14. TABLE CENTROID

The centroid table contains the central point values in each cluster. These values are used as a basis for calculating each data set's distance from each center by measuring clusters.

3. Davies Bouldin Index

One method for assessing clustering performance is the Davies Bouldin index. The parameters obtained from this evaluation are in the form of values, where a smaller number indicates a better cluster. The Davies Bouldin index values from this research are as follows:

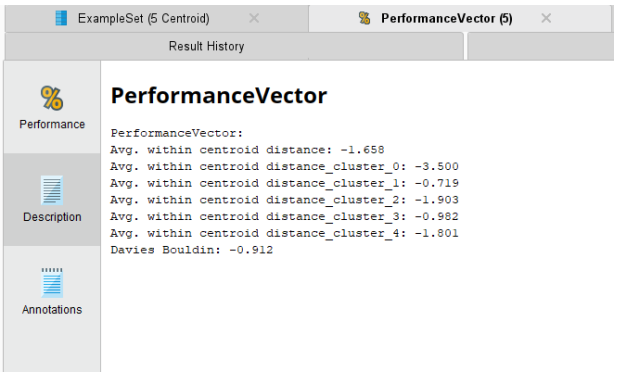

FIGURE 15. DAVIES BOULDIN INDEX

After carrying out the clustering and testing process using the RapidMiner tool, it can be concluded that of the five clusters produced, there are: cluster 0 with 6 items, cluster 1 with 8 items, cluster 2 with 12 items, cluster 3 with 13 items, and cluster 4 with 16 items. Based on the analysis carried out, the percentages for each cluster are:

TABLE 3. PERCENTAGE OF EACH CLUSTER

Centroids	Amount	Percentage
Cluster 0	6	11%
Cluster 1	8	*14%*
Cluster 2	12	22%
Cluster 3	13	23%
Cluster 4	16	30%
Total	55	100%

4.5 Recommendation

The recommendation of centroids and foods is a critical aspect of the K-Means Clustering Algorithm implementation on food products shown in Table 4.

TABLE 4. RECOMMENDATION OF CENTROIDS AND FOODS

No	Centroids	Foods
1.	Cluster 0	Ceker mercon, Curos, Pisang keju, Roti Bakar, Rujak cireng, Sambal.
2.	Cluster 1	*Ayam Taliwang, Paket Sei Barcode Iga, Paket Sei Barcode Lidah, Paket Sei Barcode Sapi, Sei Kemasan 250 gr, Sei Kemasan 500 gr, Sei 100 gram, Sei 500 gr Ayam,*
3.	Cluster 2	Brulee Bomb, Dimsum, Fries Platters, Indoset, Nasi Bakar, Nasi Kucing, Risoles, Sate Telur Puyuh, Sei Ayam, Sei iga, Sei Lidah, Sei Sapi
4.	Cluster 3	Ayam bakar paha, Ayam bakar sayap, Sate ati, sate, baso urat, sate usus, sate cikua, sate jando, sate kikil, sate kulit ayam, sate ampela, siomay barcode, tahu bacem, tempe bacem
5.	Cluster 4	Cilok krispi, Crab stick, Daun singkong, Dumpling ayam, dumpling keju, fish roll, kepala ayam, kuah, nasi goreng sei, otak-otak udang, sate cumi, sate sosis, sate sosis kecil, sosis merah, spaghetti brulee, tahu bakso

5. CONCLUSIONS

Based on the results of the research that has been carried out, it can be concluded as follows. Using Rapidminer software and using the k-means algorithm with sales data for 11 months with calculations carried out to produce 5 clusters. Based on the comparison results of 3 K-means algorithms with different K values, namely 3, 4, 5, the result from Davies Bouldin with a value close to 0 is the value with K 5, with the result from Davies Bouldin being -0.912.

REFERENCES

[1] J. Mikha Barus, M. Zunaidi, and F. Sonata, "Implementasi Data Mining Untuk Klasterisasi Penjualan Makanan Dan Minuman Menggunakan Metode K-Means," *Jurnal CyberTech*, vol. 3, no. 6, pp. 1195–1207, 2020 [Online]. Available: https://ojs.trigunadharma.ac.id/

[2] M. A. Jassim and S. N. Abdulwahid, "Data Mining preparation: Process, Techniques and Major Issues in Data Analysis," *IOP Conference Series: Materials Science and Engineering*, vol. 1090, no. 1, p. 012053, Mar. 2021. DOI: 10.1088/1757-899X/1090/1/012053

[3] T. Rak and R. Żyła, "Using Data Mining Techniques for Detecting Dependencies in the Outcoming Data of a Web-Based System," *Applied Sciences*, vol. 12, no. 12, p. 6115, Jun. 2022. DOI: 10.3390/app12126115

[4] F. Zuha and G. Achuthan, "Analysis of Data Mining Techniques and its Applications," *International Journal of Computer Applications*, vol. 140, no. 3, pp. 6–14, Apr. 2016. DOI: 10.5120/ijca2016909249

[5] T. Hidayat and B. E. Putro, "Analisis Karakteristik Konsumen Hotel 'X' dengan Menggunakan Metode K-Means Clustering," *Jurnal Media Teknik dan Sistem Industri*, vol. 4, no. 2, p. 53, Sep. 2020. DOI: 10.35194/jmtsi.v4i2.995

[6] N. Astuti, J. N. Utamajaya, and A. Pratama, "Penerapan Data Mining Pada Penjualan Produk Digital Konter Leppangeng Cell Menggunakan Metode K-Means Clustering," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 3, p. 754, Jun. 2022. DOI: 10.30865/jurikom.v9i3.4351

[7] R. Askia Kurniawan, M. Siddik Hasibuan, R. Syahrul Ramadhan, I. Komputer, and U. Islam Negeri Sumatera Utara, "Application of the K-Means algorithm for clustering places to eat in Batubara," *COSIE (JOURNAL OF COMPUTER SCIENCE AND INFORMATICS ENGINEERING*, vol. 01, no. 1, pp. 10–18, 2022.

[8] D. Triyansyah and D. Fitrianah, "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing," *Jurnal Telekomunikasi dan Komputer*, vol. 8, no. 3, p. 163, Oct. 2018. DOI: 10.22441/incomtech.v8i3.4174

[9] N. Normah, S. Nurajizah, and A. Salbinda, "Penerapan Data Mining Metode K-Means Clustering Untuk Analisa Penjualan Pada Toko Fashion Hijab Banten," *Jurnal Teknik Komputer*, vol. 7, no. 2, pp. 158–163, Jul. 2021. DOI: 10.31294/jtk.v7i2.10553

[10] K. Fatmawati and A. P. Windarto, "Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (DBD) Berdasarkan Provinsi," *Computer Engineering, Science and System Journal*, vol. 3, no. 2, p. 173, Aug. 2018. DOI: 10.24114/cess.v3i2.9661

[11] T. Amalina, D. Bima, A. Pramana, and B. N. Sari, "Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food," *Jurnal Ilmiah Wahana Pendidikan*, vol. 8, no. 15, pp. 574–583, 2022 [Online]. DOI: 10.5281/zenodo.7052276

[12] W. Romadhona, "Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means," *KOMPUTEK*, vol. 6, no. 2, p. 83, Oct. 2022. DOI: 10.24269/jkt.v6i2.1328

AUTHORS

Lasmi Lasmini Gumelar

Graduated from the Department of Informatics, Faculty of Engineering, Perjuangan University of Tasikmalaya, Indonesia.

Ruuhwan, M.Kom.

Lecturer in Department of Informatics, Faculty of Engineering, Perjuangan University of Tasikmalaya, Indonesia.

Missi Hikmatyar, M.Kom.

Lecturer in Department of Informatics, Faculty of Engineering, Perjuangan University of Tasikmalaya, Indonesia.