EasyChair Preprint

№ 11378

Application of the K-Means Algorithm in Determining the Best-Selling Restaurants on the Shopee Food Application in Karawang City

Agustia Hananto, April Hananto, Bayu Priyatna, Baenil Huda and Wahyu Pratama Setiawan

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

November 23, 2023

APPLICATION OF THE K-MEANS ALGORITHM IN DETERMINING THE BEST-SELLING RESTAURANTS ON THE SHOPEE FOOD APPLICATION IN KARAWANG CITY

Agustia Hananto[1] April Lia Hananto[2] Bayu Priyatna[3] Baenil Huda[4] Pratama Wahyu Setiawan[5]

[1,2,3,4] Universitas Buana Perjuangan Karawang, Karawang 41361, Indonesia
aprilia@ubpkarawang.ac.id

Abstract. Determining the best restaurant and choosing the most popular restaurant on the Shopee food application is a major problem among people who want to enjoy fast food in the city of Karawang. Shopee food is a food buying and selling application that is widely used by people to buy food online. Therefore, with this research, people can easily determine which restaurant they want to choose. One of the methods used in determining the best-selling restaurants in this study is by applying the K-Means Clustering Algorithm, by taking a data selection and processing approach. The results obtained in this study show that out of 100 restaurant data it is divided into 5 clusters, of which among the 5 clusters there is only 1 cluster that has the best-selling restaurant data, namely cluster 0 with a total of 32 restaurants.

Keywords: K-Means Clustering, Restaurant, RapidMiner.

1 Introduction

In the industrial era 4.0, the role of digital is really needed by many people, including when buying food online, the Shopee Food platform is an online food shopping application that makes it easy for consumers who are looking for various types of restaurants that provide various food menus [1]. In general, consumers do not know which restaurant is the best with a variety of food menu choices, so it becomes a reference in choosing the taste of various food menus that they want to order according to their tastes based on the most popular or best-selling choices on the application, especially when consumers have just visited a city or city. areas they have never visited [2].

With this research, we can contribute to how to determine the best-selling restaurants by applying the K-Means algorithm. The K-Means Clustering algorithm is a method where data that has the same characteristics is grouped into the same cluster and data that has different characteristics is grouped into other groups [3]. In the context of Shopee Food, especially in Kara- wang City, the K-Means algorithm can be used to cluster food purchasing patterns that are frequently purchased by consumers, so that they can determine the most popular or best-selling restaurants in Karawang City [4].

With recommendations for best-selling restaurants, Shopee Food users in Karawang City can easily determine restaurants that are popular and easily sought after by many consumers [5]. Users will get a better and faster experience ordering food and choos-

ing popular restaurants, while restaurants can increase their visibility and attractiveness. on the Shoope Food platform, especially in Karawang City [6].

2 Research Method

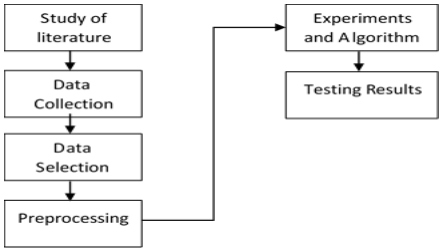


Figure 1. Research Method Flow

In figure 1 above, the researcher conducted a literary study related to the research theme, then collected data, then selected the data to be used and preprocessed the data. Once the data is ready, the process of implementing the k-means algorithm will be carried out using a rapid miner and then testing will be carried out.

2.1 Literature Study

Literature study is a series of activities related to methods of collecting library data, reading and taking notes, and managing research materials, even though it is research with literature studies, you do not have to go to the field and meet with respondents [7], [8], [9].

2.2 Data Collection

Table 1. Data Collection Table

No	Restaurant Name	Ratings	Menu Type	Quantity Sold	Type of Food	Type of Restaurant	Information
1	Mie Gacoan Karawang	4,8	4	>2000	Fried noodle	Restaurant food	Best Seller
2	M'c Donald Galuh Mas	4,8	16	>2000	Fried Chicken, Burger	Restaurant food	Best Seller
3	PHD by pass AY	4,9	12	>1000	Pizza, Pasta	Restaurant food	Best Seller
4	Janji Jiwa	4,9	7	800	Coffe, Bakery	Coffe, Baker	Best Seller
5	Lazzato Jatirasa	4,9	16	501	Fried chicken, Burger	Restaurant food	Best Seller
6	Kebuli Mamake	4,8	12	423	The middle east	Restaurant food	Best Seller
25	Ray's Dessert & Cake	4,7	7	101	Bread	Snacks	Not sold

Data collection obtained from the Shopee Food application, data is taken from several restaurants that collaborate with partners with the Shopee application. The restaurant data taken is only restaurants within the city of Karawang.

2.3 Data Selection

It is a selection process where the selected data will be converted into a data format that is appropriate to the data analysis used [9]. Once the data has been collected, you can immediately select which data is not needed and which data is needed.

2.4 Preprocessing

The thing that must be done before carrying out data clustering techniques is to prepare a dataset and understand the dataset by checking the data first [8]. Based on the data obtained, the data is still raw and cannot be directly used in the clustering process [10].

2.5 Algorithm Experimentation and Testing

The algorithm used for cluster formation is K-Means. K-Means is a non-hierachical clustering data method that can group restaurant data into several clusters based on the similarity of the data, so that restaurant data that has the same characteristics is grouped into another cluster [11].

3 Result and Discussion

3.1 Data and Information Processing

RapidMiner is software for data processing. RapidMiner extracts patterns from large datasets by combining statistical methods, artificial intelligence and databases [12]. RapidMiner makes it easy for users to calculate large amounts of data using operators [14]. This operator functions to modify data. Data is connected via nodes to the operator and then just connects it to the node to see the results [13], [15], [16].

3.2 Data Selection Stage

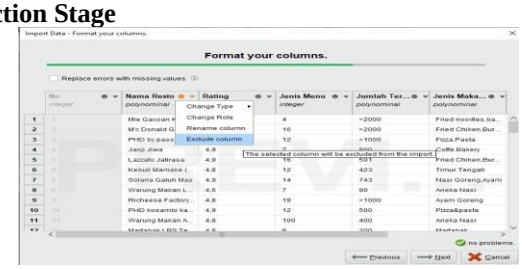


Figure 2. Data Selection Stage

Because the restaurant name has no effect on K-Means clustering, the restaurant name is not included in the next process. To remove the restaurant name from the data, you can select Exclude Column.

3.3 Nominal to Numerical Stage

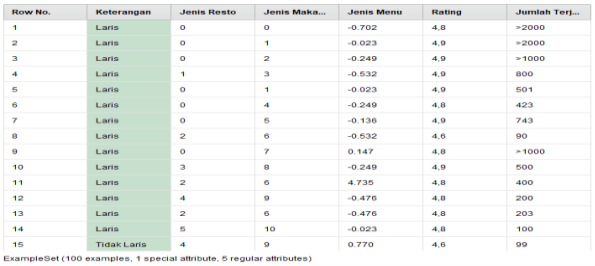
Nominal to Numerical is the stage of the process of changing data which was originally in the form of letters into numbers.

Row No.	Keterangan	Rating	Jenis Menu	Jumlah Terj...	Jenis Maka...	Jenis Resto
1	Laris	4,8	4	>2000	Fried noodle...	Restoran cep...
2	Laris	4,9	16	>2000	Fried Chiken,...	Restoran cep...
3	Laris	4,9	12	>1000	Fizza,Pasta	Restoran cep...
4	Laris	4,9	7	800	Coffe,Bakery	Coffe,Bakery
5	Laris	4,9	16	501	Fried Chiken,...	Restoran cep...
6	Laris	4,8	12	423	Timur Tengah	Restoran cep...
7	Laris	4,9	14	743	Nasi Goreng,...	Restoran cep...
8	Laris	4,6	7	90	Aneka Nasi	Restoran Pra...
9	Laris	4,8	19	>1000	Ayam Goreng	Restoran cep...
10	Laris	4,9	12	500	Pizza&pasta	Pizza&pasta
11	Laris	4,8	100	400	Aneka Nasi	Restoran Pra...
12	Laris	4,8	8	200	Martabak	Martabak
13	Laris	4,8	8	203	Aneka Nasi	Restoran Pra...
14	Laris	4,8	16	100	Sunda	Restoran Se...

ExampleSet (100 examples, 1 special attribute, 5 regular attributes)

Figure 3. Data that is still in nominal form

Looking at the data above in the Food Type and Restaurant Type columns it is still in letter/nominal form, therefore a Nominal to Numerical process is needed.



Row No.	Keterangan	Jenis Resto	Jenis Maka...	Jenis Menu	Rating	Jumlah Terj...
1	Laris	0	0	-0.702	4,8	>2000
2	Laris	0	1	-0.023	4,9	>2000
3	Laris	0	2	-0.249	4,9	>1000
4	Laris	1	3	-0.532	4,9	800
5	Laris	0	1	-0.023	4,9	501
6	Laris	0	4	-0.249	4,8	423
7	Laris	0	5	-0.136	4,9	743
8	Laris	2	6	-0.532	4,6	90
9	Laris	0	7	0.147	4,8	>1000
10	Laris	3	8	-0.249	4,9	500
11	Laris	2	6	4.735	4,8	400
12	Laris	4	9	-0.476	4,8	200
13	Laris	2	6	-0.476	4,8	203
14	Laris	5	10	-0.023	4,8	100
15	Tidak Laris	4	9	0.770	4,6	99

ExampleSet (100 examples, 1 special attribute, 5 regular attributes)

Figure 4. Data in numerical form

Judging from the data above in the Dining Type and Restaurant Type columns, they are already in numeric form and you can proceed to the clustering process.

3.4 K-Means Clustering Stage

Table 2. K-Means Clustering Results

Cluster 0	:32 restaurant
Cluster 1	:15 restaurant
Cluster 2	:19 restaurant
Cluster 3	:17 restaurant
Cluster 4	:17 restoram
Total Number of Restaurants	100 restaurants

From the results of the table above, it can be seen that 100 restaurant data are divided into 5 clusters, in cluster 0 there are 32 best-selling restaurants. Cluster 1 has 15 best-selling restaurants, in cluster 2 there are 19 best-selling restaurants, in cluster 3 there are 17 best-selling restaurants, and finally cluster 4 has 17 best-selling restaurants. In

the results of research using the K-Means clustering algorithm method, in cluster 0 there are restaurants that sell the most and with high ratings. Cluster 0 is the cluster that sells the most among the other clusters. The table below is a table of distribution of restaurants in cluster 0.

Table 3. Determining the Best Sellers

No.	Restaurant Name	Sold	Ratings	Information
1	Noodle Gacoan Karawang	>2000	4.8	Best Selling
2	M'c Donald Galuh Mas	<2000	4.9	Bestseller
3	PhD Bypass Karawang	>1000	4.9	Bestseller
4	Richeese Factory (Gramed)	1000	4.8	Bestseller
5	KFC - Cikampek Mall	<1000	4.9	Bestseller

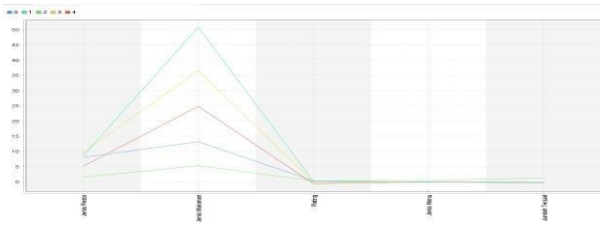


Figure 6. Relationship between Centroid Points of Each Cluster

From the graphic above, it is clear that the determinant of the best-selling restaurant lies in the rating label and the number sold, so if the number sold is small and the rating is high it cannot be declared a bestseller, but if the number sold is large but the rating is stable then the restaurant can be declared a bestseller.

4 Conclusion

From the research results it can be concluded that restaurant data analysis using the RapidMiner application can extract the necessary information and group restaurant data into 5 groups/clusters from 100 restaurant data. The results of data extraction show the sales level of the best-selling restaurants along with the relationship between restaurant ratings. From the results of the discussion it can be seen that the top 5 restaurants in cluster 0 are Noodle Gacoan Karawang, M'c Donald Galuh Mas, PhD Bypass Karawang, Richeese Factory (Gramed) and KFC - Cikampek Mall is the best-selling restaurant in Karawang City.

References

[1] AM Dewi, "Optimizing the Shopee Food E-Commerce Application as a Marketing Media for MSMEs During the Covid 19 Pandemic," J. Ris. Manaj. and Business, vol. 16, no. 2, p. 77, 2022, doi: 10.21460/jrmb.2021.162.393.

[2] I. Vania and R. Simbolon, "The Influence of the ShopeeFood Promo on Shopee Users' Purchase Interest (In the South Tangerang Area)," Fak. Econ. Univ. Adventist Indonesia., pp. 46–58, 2021.

[3] A. Nur Khormarudin, "Data Mining Techniques: K-Means Clustering Algorithm," J. Computer Science., pp. 1–12, 2016, [Online]. Available: https://ilmukomputer.org/category/datamining/.

[4] A. Kholik, "Implementation of Data Mining Menu Selection Patterns," 2020.

[5] E. - and SP Tamba, "Application of Apriori Data Mining Algorithm in Determining Raw Material Stock in Fisherman's Restaurants Using the Association Rule Method," J. Sist. Inf. and Computer Science. Prima(JUSIKOM PRIMA), vol. 5, no. 2, pp. 97–102, 2022, doi: 10.34012/jurnalsisteminformationdanilmukomputer.v5i2.2407.

[6] Y. Indriany, S. Hastuti, and D. Indrisetno PV, "ANALYSIS OF BRAND IMAGE AND CONSUMER ATTITUDE TOWARDS PURCHASING DECISIONS ON ONLINE FOOD APPLICATIONS (Case Study of Go Food, Grab Food, Shopee Food Application Customers)," J. Lentera Bisnis, vol . 11, no. 2, p. 205, 2022, doi: 10.34127/jrlab.v11i2.539.

[7] D. Judithia, "Adaptation Process of the Fakfak Student Association in Bandung City," J. Chem. Inf. Model., vol. 53, no. 9, pp. 54–69, 2019, [Online]. Available: https://elibrary.unikom.ac.id/id/eprint/1558/

[9] B. Priyatna, A. Hananto, A. Solehudin, and A. L. Hananto, "Determining the Formulation of Mosquito Repellent Production Raw Materials Using Simple Additive Weighting (SAW)," vol. 11, pp. 482–489, 2023.

[10] Y. Alpian, M. S. Sumantri, Yufiarti, S. W. Anggraeni, and A. L. Hananto, "Self-Concept, Logical Thinking Patterns, and Emotional Intelligence in Relation to Digital Literacy Ability," *Res Mil.*, vol. 12, no. 3, pp. 691–710, 2022.

[11] BRCT I et al., "Implementation of k-means clustering on rapidminer for analysis of accident-prone areas," no. April, pp. 58–62, 2017.

[12] S. García, S. Ramírez-Gallego, J. Luengo, J.M. Benítez, and F. Herrera, "Big data preprocessing: methods and prospects," *Big Data Anal.*, vol. 1, no. 1, pp. 1–22, 2016, doi: 10.1186/s41044-016-0014-0.

[13] AND Saran, "Chapter V - 1 Chapter V - 2," Econ. Islam., no. variable X, pp. 46–47, 2019, [Online]. Available: https://dspace.uii.ac.id/bitstream/handle/123456789/25104/%5BCetak National Library%5D Islamic Economy Solutions to the Challenges of the Age.pdf?sequence=1&isAllowed=y#page=71

[14] A. Deolika, K. Kusrini, and ET Luthfi, "Word Weighting Analysis in Text Mining Classification," J. Teknol. Inf., vol. 3, no. 2, p. 179, 2019, doi: 10.36294/jurti.v3i2.1077.

[15] A. Hananto, T. K. A. Rahman, G. Brotosaputro, A. Fauzi, A. L. Hananto, and B. Priyatna, "Parameters Monitoring Automation Kiln Manufacture Based Optical Character Recognition (OCR) with The Template Matching Method," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 6, pp. 621–635, 2023.

[16] M. Ghazali, N. M. Suaib, S. Sulaiman, N. Ibrahim, A. L. Hananto, and A. Beran, "Preliminary Work on Design Thinking: Addressing Challenges using Low-fidelity Prototyping with Rural Teenagers," *Proceeding - 2018 Int. Conf. ICT Rural Dev. Rural Dev. through ICT Concept, Des. Implic. IC-ICTRuDEv 2018*, no. October, pp. 158–161, 2018, doi: 10.1109/ICICTR.2018.8706867.