

Information System for Culinary Product Selection Using Clustering K-Means and Weighted Product Method

Anindya Khrisna Wardhani
Information Systems
Diponegoro University
Semarang, Indonesia
nindywardhani77@gmail.com

Catur Edi Widodo
Department of Physics
Faculty of Science and Mathematics
Diponegoro University
Semarang, Indonesia
catur.ediwidodo@gmail.com

Jatmiko Endro Suseno
Department of Physics
Faculty of Science and Mathematics
Diponegoro University
Semarang, Indonesia
jatmikoendro@gmail.com

Abstract—Typical foods are foods that have characteristics that cannot be found in other areas. The number of restaurants that sell typical food made consumers confused in choosing where to buy foods. This study aims to build information systems to provide culinary product recommendations based on the criteria desired by consumers. The clustering process used the k-means method used to classify the categories of food desired by consumers into six categories: appetizers, soups, desserts, snacks and drinks. In addition to the k-means clustering method, we also used a weighted product method that serves as a culinary selection recommendation rank. The result of the research is a system that can be used for culinary selection by using criteria of food category, number of menu, price, facility and distance. Case study was conducted in Semarang culinary tour. The result of clustering obtained from this research is there are 17 special dining house entrees, 6 specialty soup restaurants, 51 restaurants selling staple dishes, 11 selling desserts, 13 restaurants selling snacks and 6 restaurants selling specialty drinks. While the weighted product method produces *Wingkorolls* as the first restaurant recommendation for dessert menu with a score of 0.2267.

Keywords—*decision support system; data mining; k-means; weighted product; culinary*

I. INTRODUCTION

Diversity of typical food in an area is to make a lot of local and foreign tourists visiting the area to enjoy the culinary tour. The number of restaurants that sell typical food make consumers are in confusion for choosing foods. There are several factors desired by the user that this case is called as a multicriteria case. One method that can be used in the case of multicriteria is the weighted product method. The weighted product method is used for calculating the ranking of several restaurants based on the criteria desired by the user ie menu category, number of menu, average price, distance and facilities.

Besides, using weighted product method, this research uses data mining clustering technique. Clustering is a data mining technique used for grouping data objects into clusters so that the data owned by each cluster is different [7]. The k-means

clustering method is used for grouping food categories based on home data. Grouping of food categories consists of appetizers, soups, staple dishes, desserts, snacks and drinks.

II. METHODOLOGY

A. Clustering K-means

Clustering is a data mining technique used for grouping data objects into clusters so that the data held by each cluster is different [7]. Clustering analysis is useful for obtaining information from data sets in fields such as bioinformatics, pattern recognition, image processing, data mining, marketing and economics, and others [2].

The cluster technique has two methods in the grouping of hierarchical clustering and non-hierarchical clustering. hierarchical clustering is a method of grouping data that works by grouping two or more data that have similarities or similarities, then the process is continued to another object that has two closeness, this process continues until the cluster forms a tree where there is a hierarchy or a clear level between objects from the most similar to the least similar. But logically all objects will eventually form a cluster [3].

Clustering k-means is an algorithm used in data grouping. This algorithm is able to minimize the distance between the data to the cluster. Basically the use of this algorithm in the clustering process depends on the data obtained and the conclusion to be reached at the end of the process[6].

K-means algorithm basically do 2 processes. They are process of detection of cluster central location and member search process from each cluster. The clustering process begins by identifying the data to be grouped, X_{ij} ($i = 1, \dots, n; j = 1, \dots, m$) with n is the amount of data to be grouped and m is the number of variables. At the beginning of the iteration, the center of each cluster is set free (arbitrary), C_{kj} ($k = 1, \dots, k; j = 1, \dots, m$). Then, calculate the distance between each data with each cluster center. To calculate the distance of the I-data (x_i) at the center of the k-cl cluster (c_k), given the name (d_{ik}), we can use the euclidean distance formula. A data will be a member of the

k-cluster if the distance of the data to the cluster-k center is at least small when compared to the distance to the center of another cluster [3].

The basic process of k-means algorithm begins with determining the number of clusters to be established and setting the cluster center. After that, use the euclidean distance to calculate the distance of each data to the center of the cluster with the data using the formula:

$$d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \quad (1)$$

Then group the data into the cluster with the shortest distance with the formula:

$$Min \sum_k^k = d_{ik} = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2} \quad (2)$$

Then calculate the new cluster center using the formula:

$$C_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p} \quad (3)$$

where :

- x_{ij} cluster of k
- p number of cluster members of - k

Repeat steps two to four so that no more data is moved to another cluster.

B. Weighted product

Weighted Product is a method of completion by using multiplication to associate attribute values, where the value must be raised first with the attribute weights in question. This process is similar to the normalization process. Preferences for alternative S_i are given as follows:

$$S_i = \prod_{j=1}^n X_{ij}^{W_j}; \text{ dengan } i = 1, 2, \dots, m \quad (4)$$

where :

- S_i Vector value S_i
- X_{ij} Alternate value against criteria
- W_j The weighted rank is positive for the profit attribute, and is negative for the cost attribute.

where the value of $\sum W_j = 1$. W_j is a positive rank for attribute profits, and is negative for cost attributes. As for calculating the relative preference values of each alternative, the following equations are given:

$$V_i = \frac{\prod_{j=1}^n X_{ij}^{W_j}}{\prod_{j=1}^n X_{j*}^{W_j}}; \text{ dengan } i = 1, 2, \dots, m. \quad (5)$$

where :

- V_i Alternative preference value
- X_{ij} Alternative attribute value
- W_j Weight Criteria Value

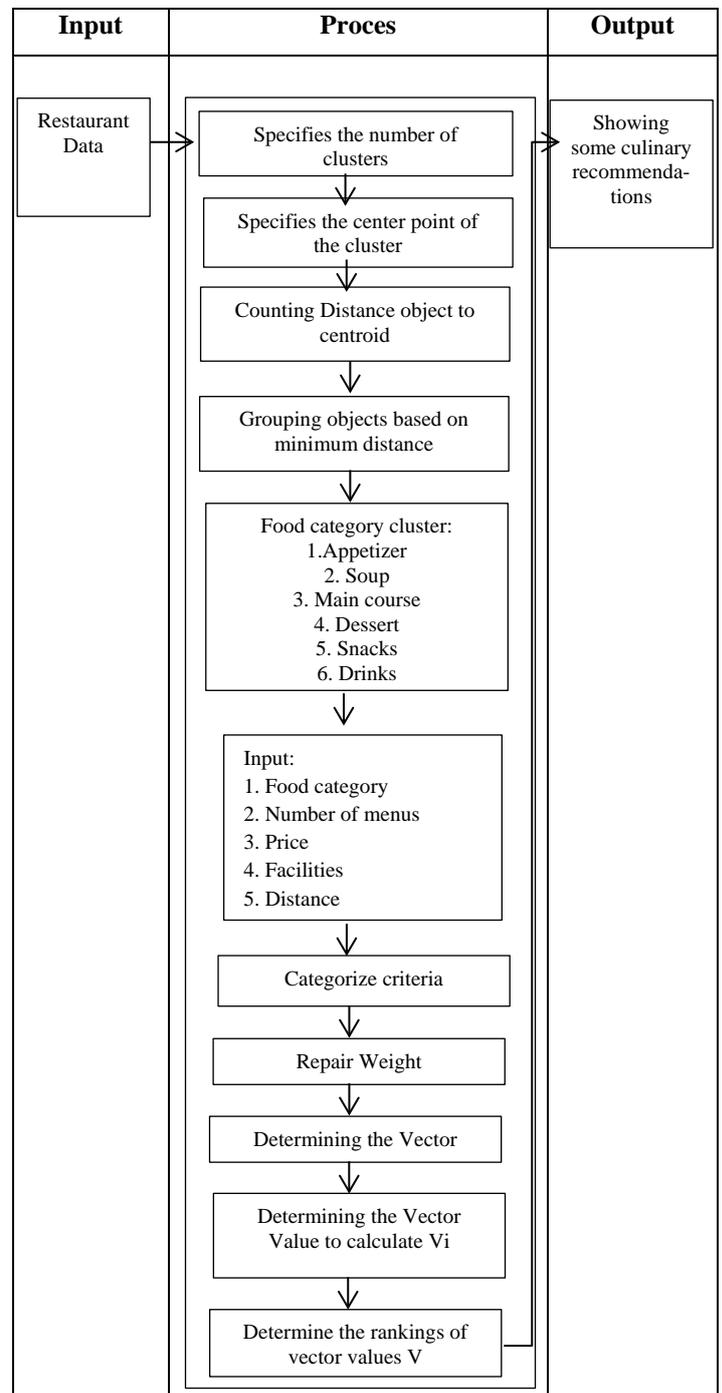


Fig 1. Information system framework

III. RESULTS

The first step is grouping restaurant menu data based on food category using k-means clustering method. In this research, use case study of culinary tour data in Semarang City. There are 105 restaurants that sell food typical of Semarang City.

TABLE 1. SAMPLE DATA

No	Restaurant	A	S	MC	D	S	Dr
1	Semarang roru cake				6		
2	Wife Cake				5		
3	Wingkorolls				9		
4	Lontong tahu gimbal Pak Polo			1			
5	Tahu Gimbal Bu Ir			1			
6	Tahu Gimbal Surya Rejeki			1			
7	Tahu Gimbal bu Galuh			2			
8	Tahu Gimbal Bu Kristu			2			
9	Tahu Gimbal Surya Rizki			1			
10	Tahu Gimbal Pak Timbul			1			
11	Tahu Gimbal Cita rasa			1			
12	Tahu Gimbal Gang Baru			5			
13	Mangut Fiza		3				
14	Excellent Bakery				1		
15	Bandeng juwana Elrina Pandanaran	4		8	1	10	
16	Toko Roti Selina				1		
17	Mie Kopyok Mbah Wok			1			
18	Mie Kopyok 88			1			
19	Mie Kopyok mas antok			1			
20	Wedang Tahu Tjoa						1

Information :

- A = Appetizer
- S = Soup
- MC = Main Course
- D = Dessert
- Dr = Drink

Iteration 1st

- Number of cluster : 6
- Determination of initial centers *cluster*

TABLE II. PRELIMINARY CLUSTER CENTER POINTS

No	Restoran	A	S	MC	D	S	Dr
1	Wife Cake				5		
2	Mangut Fiza		3				
3	Excellent Bakery				1		
4	Bandeng juwana Elrina Pandanaran	4		8	1	10	

5	Mie Kopyok Pak Dhuwur				1		
6	Wedang Tahu Tjoa						1

- The calculation of center distance *cluster*
 Calculate the distance between the data with an early center of the cluster. Then we will get the value of distance matrix as follows:
 1st data distance to the center of the cluster :
 C0=1
 C1=6,7082
 C2=5
 C3=14,3178
 C4=6,0827
 C5=6,0827
 and proceed onwards to the data to calculate the 2ndN towards the center initial cluster distance matrix to be obtained.

- Grouping the data
 Distance calculation results at the point 3 will do the comparison and the selected distance closest data with the center, cluster this distance would indicate that the data which have the shortest distance to be in one group at the center of the cluster. Based on the matrix obtained in obtained grouping as follows:
 C0: Data 1, 2, 3
 C1: Data 13
 C2: Data 14,16
 C3: Data 15
 C4: Data 4,5,6,7,8,9,10,11,12,17,18,19
 C5: Data 20

- Determination of centers cluster the new
 After getting members of each cluster and then centers cluster a new are calculated based on data member each cluster that have been center cluster a new in the matrix table as follows:

TABLE III. CENTER NEW CLUSTER

C0	0	0	0	0	0	0
C1	0	3	0	0	0	0
C2	0	0	0	0,666666667	0	0
C3	4	0	8	1	1	0
C4	0	0	1,38462	0	0	0
C5	0	0	0	0	0	0,5

Iteration then performed in the same manner until there are no changes to the data within a cluster.

Processing k-means using 105 data restaurant in Semarang City by Rapid Miner with the initial number *cluster* by 6 pieces, then the results of *cluster* existing 17 items of cluster 0, 6 items of cluster 1, 51 items of cluster 2, 11 items of cluster 3, 13 items of cluster 4 and 6 items of cluster 5 with the total amount of data is 105.

Cluster Model

```
Cluster 0: 17 items
Cluster 1: 6 items
Cluster 2: 51 items
Cluster 3: 11 items
Cluster 4: 13 items
Cluster 5: 6 items
Total number of items: 105
```

Fig. II. Results of cluster data K-means in the implementation of Rapid Miner

	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4	cluster_5
0	2.667	7.333	0	0.218	0	0
0	0	0	0	0.138	0	0
0	4.333	0.500	0	0.897	0	0
0	0.333	0	5.857	0.034	16	0
29	8	0.333	0	0.287	0	0
0	0	0	0	0.069	0	0

Fig. III. The calculation result between the distance of the cluster and the centroid

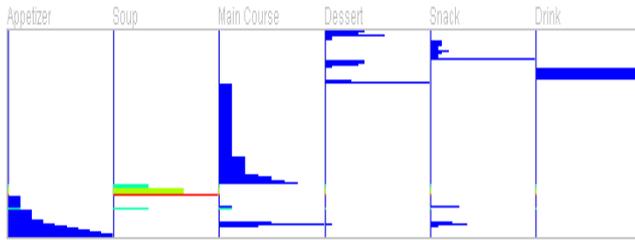


Fig. IV. Graph data of Clustering K-means

Data from the categories of food categories are used to process rankings on the weighted product method. In the process of ranking the weighted product method, the criteria used are food category, number of menu, price, facility and distance. Details of the criteria used in the process of weighted product method can be seen from the table as follows:

- Food Criteria

On the menu criteria, the data is taken from the clustering results. So the menu criteria only as a process of filter data to be processed. For example, the user wants the desserts category then, the data to be processed using the method of weighted product is data from the restaurant based on dessert cluster.

TABLE IV. FOOD CRITERIA

Category	Code
Appetizer	1
Soup	2
Main Course	3
Dessert	4
Snack	5
Drinks	6

- Criteria Number of Menu
Criteria of the number of menus is the number of menus based on the categories contained in the restaurant.
- Price Criteria
The price criteria used is the average price by category.

TABLE V. AVERAGE PRICE OF DESSERT

Category	Range	Code
Cheap	<35000	1
Medium	35.100-60.000	2
Expensive	>60000	3

4. Facility Criteria

The facility criteria are the facilities in the restaurant.

TABLE VI. FACILITY

Category	Code
No facilities	0
Toilet	1
Air conditioning(AC)	2
Wifi	3
Charging corner	4
Fan	5
Fan, toilet	6
AC,Wifi	7
AC,toilet	8
AC,wifi,charging corner	9
AC,charging corner,toilet	10
AC,wifi,toilet	11
AC,wifi,charging corner,toilet	12

Decision makers give preference weight for criteria number of menu, price, facility and distance as:

$$W = (5, 4, 2, 3).$$

Where :

- 5 very important,
- 4 important,
- 3 quite important,
- 2 not so important,
- 1 not important

TABLE VII. SAMPLE DATA

Restaurant	Number of menus	Average of price	Facility	Distance
Semarang roru cake	6	3	2	1.4
Wife Cake	5	3	2	2.5
Wingkorolls	9	2	2	2.9
Pisang Plenet Assoy	6	1	0	0.85
Pisang Plenet Pak Tuko	6	1	0	4.3
Pisang Plenet Ory	5	1	0	4.2
Nature Nest	1	3	2	0.6

Toko Roti Selina	1	1	2	1.7
Excellent Bakery	1	1	2	2.2
Es puter cong lik cabang Ahmad Dahlan	4	1	0	0.45
Es puter cong lik cabang Dargo	16	1	0	2.7

Previously done prior weighting of the previous preference weight of $W = (5, 4, 3, 3)$. W_j is W index to j . So for W_1 is 5, W_2 is 3 and so on. And $\sum W_j$ is the sum of W . So for the improvement of W_1 weight becomes:

$$W_1 = \frac{5}{5+4+2+3} = 0.36$$

TABLE VIII. RESULT OF WEIGHT

	Weight
W1	0.36
W2	0.29
W3	0.14
W4	0.21

The next stage is to determine the value of the vector S by lifting and multiply the value of each of these criteria by the weight that has been fixed before. The calculation of Vector S is as follows:

$$S_1 = (6^{0.36})(3^{-0.29})(2^{0.14})(1,4^{-0.21}) = 1.423$$

TABLE IX. RESULT OF VECTOR S

	Vector S
S1	1.423
S2	1.1798
S3	1.5895
S4	0
S5	0
S6	0
S7	0.892
S8	0.9857
S9	0.9337
S10	0
S11	0

Then after determining the value of the vector to be used is calculate the Preferences (V_i) for ranking. So Results from Calculating Preferences V_1 are as follows:

$$V_1 = \frac{1.423}{1.423+1.1798+1.5895+0+0+0+0.892+0.9857+0.9337+0+0} = 0.2032$$

TABLE X. RESULT OF PREFERENCES (V_i)

	V_i
V1	0.2032
V2	0.1684

V3	0.2267
V4	0
V5	0
V6	0
V7	0.1273
V8	0.1407
V9	0.1332
V10	0
V11	0

From the above calculation, the value of V_3 determines the greatest value so that in other words V_3 is the best alternative choice with the weighting given by the decision maker.

IV. CONCLUSIONS

Information systems for the selection of culinary products can be determined based on the clusters of typical food categories and ranking in accordance to the user desire. Clusters are formed into six clusters of appetizers, soups, staple dishes, desserts, snacks and drinks. While the method of weighted product can be used for ranking restaurant desired by user with weighting and based on criteria of food category, number of menu, average price, facility and distance. The results of the groupings obtained from this study were 17 specialty restaurants, 6 specialty restaurants, 51 restaurants selling main dishes, 11 selling desserts, 13 restaurants selling snacks and 6 restaurants selling specialty drinks. While the weighted product method produces Wingkorolls as the first restaurant recommendation for the dessert menu with a score of 0.2267. With this system, can facilitate consumers in choosing culinary products.

REFERENCES

- [1] H. Arif and I.T. Utami, "Decision Support System to Determine the Location New Cafe Suncafe as a Tourist Culinary Tour in District Pringsewu by Using Simple Additive Weighting," *IJISCS*, vol. 1, pp. 10-21, 2015.
- [2] C. Zhang, and Z. Fang, "An improved k-means clustering algorithm," *J. of Inform. and Comput. Sci.*, vol. 10, pp. 193-199, 2013.
- [3] J. O. Ong, "Implementasi algoritma k-means clustering untuk menentukan strategi marketing President University [Implementations of k-means clustering algorithm to determine marketing strategy of President University]," *JITI*, vol. 12, pp. 10 – 20, 2013.
- [4] J. Yadav and M. Sharma, "A Review of K-mean Algorithm", *Int. J. of Engineer. Tren. and Techno. (IJETT)*, vol. 4, pp. 123-129, 2013.
- [5] T. Kapila, S. Illangarathne, and H. Arumawadu, "K-means clustering for segment web search result," *Int. J. of Eng. Works*, vol. 2, pp. 79-83, 2015.
- [6] K. Kouser, and Sunita, "A comparative study of K Means Algorithm by Different Distance Measures," *Int. J. of Innov. Res. in Comp. and Comm. Eng.*, vol. 1, pp. 49-59, 2013.
- [7] R. R. Unnati and J. Chaita, "Implementing and Improvisation of K-means Clustering Algorithm," *Int. J. of Comp. Sci. and Mob. Comp.*, vol. 5, pp. 191-203, 2016.
- [8] K. Shavita and C. Chandrasekar, "Vertical handover decision schemes using SAW and WPM for network selection in Heterogeneous wireless network," *GJCST*, vol. 11, pp. 19-24, 2011.
- [9] Y. Zheng, Q. Li, and Lan, "Density k-means: a new algorithm for centers initialization for k-means," *IEEE*, pp. 958-961, 2015.