



Nutritional Analysis Of Traditional Indonesian Food Using Machine Learning

Federico Anggito Ryseno¹, Fito Mardianto², Yoga Prasetyo Wibowo³, Faishal Nugraha⁴, Eka Kusuma Pratama⁵

^{1,2,3,4,5} Universitas Bina Sarana Informatika, Depok, Indonesia
15210016@bsi.ac.id*

Abstract

This research aims to analyze the nutritional composition of traditional Indonesian food using the clustering method. Nutritional data from 1,346 types of Indonesian food, including calories, protein, fat and carbohydrates, was analyzed with the help of RapidMiner Studio. The clustering results produced three main clusters: (1) foods low in nutrients, (2) foods high in nutrients, and (3) foods with moderate nutrients. Through statistical analysis, variations were found in the number of foods with certain nutritional values in each attribute. This research provides better insight into the nutritional composition of traditional Indonesian foods that are frequently consumed by the public. It is hoped that the results of this research can help individuals choose foods that suit their body's needs, as well as contribute to encouraging a healthier lifestyle in society.

Keywords: Nutritional analysis, Clustering, Traditional Indonesian food, RapidMiner Studio

1. Introduction

Indonesia is known for its abundant culinary riches. However, with the diversity of food and drinks available, people often have difficulty determining the nutritional health of the food and drinks they consume. This concern is related to a lack of understanding regarding nutritional composition and its implications for body health.

This lack of understanding is a significant challenge, considering the importance of nutrition in maintaining a healthy body. Therefore, this research aims to contribute to expanding understanding of the nutritional composition of traditional Indonesian foods that are often consumed by Indonesian people.

Clustering is widely used in various fields such as biology, psychology, and economics. Clustering results vary due to the number of changes in cluster parameters, so the main challenge of cluster analysis is that the number of clusters or number of model parameters is rarely known, and must be determined before grouping [1]. This research discussion chooses the k-means algorithm because it is detailed, simple and generally known. In this technique, groups objects into clusters or several groups [2]. With this approach, it is hoped that we can identify patterns in the nutritional composition of food and drinks, which can then help in providing health labels.

Through this research, it is hoped that it can provide a better understanding to the public about the importance of choosing healthy food to maintain a healthy body. Apart from that, it is also hoped that the results of this research can become a basis for developing further efforts to increase nutritional awareness in the community [3], [4].

2. Method

2.1. Data Mining

A concept used to produce a rule in the discovery of knowledge is called data mining. Data mining or data mining is a method, technique, artificial intelligence and machine extracted learning so as to produce useful knowledge and information Which stored in a large database. In principle, data mining inherits many aspects and techniques fields of science. Data mining is not something new, because data mining is the root of it various fields of science [5], [6]. Data mining has a multitude of benefits for managing raw data into knowledge or a collection of information that can be used as effective decision support [7], [8].

2.2. Clustering

Clustering is a technique from one of the data mining functionalities, the Clustering algorithm is algorithm for grouping a number of data into certain data groups (Cluster) [2]. Each Cluster has Centroid is a quantity calculated from the average value of each item from a Cluster and also has medoid which is the item located in the most center [9], [10]. By using this clustering, we can classify dense areas, find overall distribution patterns, and find interesting relationships between data attributes.

The aim of data clustering is to minimize the objective function set in the clustering process, which generally tries to minimize variation within a cluster. And minimizing variations between clusters [11], [12].

2.3. K-Means Method

"The k-means method is a grouping method related to the k-medoids method which partitions a set of data into a number of clusters that have been determined right from the start." The k-means method is very simple to run and apply, quite fast, easy to use, and easy to adapt, commonly used in various small to medium applications [9], [11]. Scientifically, k-means is one of the most important methods in the use of data mining. The k-means clustering method is one of the non-hierarchical grouping methods whose aim is to group objects, starting with identifying the data to be clustered. In the first iteration, the main points of each cluster are determined freely. Then the distance between the data and each main point in the cluster is calculated. And to calculate the distance of the i th data to the center of the j th cluster which is named $(D(1_2))$ [13], [2]. K-means is an iterative clustering algorithm [7]. The K-means algorithm starts with selecting randomly K , K here is the number of clusters you want to form. Then set the K value randomly, temporarily this value becomes the center of the cluster or what is usually called the centroid, mean or "means" [14], [13].

2.4. RapidMiner

Rapid Miner is a tool used in techniques in the machine learning environment, data mining, text mining and predictive analytics [15]. RapidMiner is software for data processing. Using data mining principles and algorithms, RapidMiner extracts patterns from large data sets by combining statistical methods, artificial intelligence and databases.

RapidMiner makes it easy for users to calculate large amounts of data using operators. This operator functions to modify data. The data is connected to the nodes on the operator then we just need to connect it to the result node to see the results. The results shown by RapidMiner can also be displayed visually with graphs. Making RapidMiner one of the software choices for extracting data using data mining methods [16].

3. Results and Discussion

This research uses one data, namely nutritional data for traditional Indonesian food. The following is some of the research data used

3.1. Data Input Stage.

Row No.	calories	proteins	fat	carbohydrate	name
1	280	9.200	28.400	0	Abon
2	513	23.700	37	21.300	Abon haruwan
3	0	0	0.200	0	Agar-agar
4	45	1.100	0.400	10.800	Akar tonjong ...
5	37	4.400	0.500	3.800	Aletoge segar
6	85	0.900	6.500	7.700	Alpukat segar
7	96	3.700	0.600	19.100	Ampas kacang...
8	414	26.600	18.300	41.300	Ampas Tahu
9	75	4.100	2.100	10.700	Ampas tahu k...
10	67	5	2.100	8.100	Ampas tahu ...
11	184	18.800	14	0	Anak sapi da...
12	174	19.600	10	0	Anak sapi da...
13	190	19.100	12	0	Anak sapi da...
14	99	4.600	1	18	Andaliman s...
15	25	1.600	0.200	5.300	Andewi

ExampleSet (1,346 examples,0 special attributes,5 regular attributes)

Fig. 1: Table Nutritional Data

In this data there are 1,346 types of Indonesian food which contain nutrients containing calories, proteins, fat, carbohydrates

3.2. Clustering data using RapidMiner

Results from clustering data related to food

- a Display cluster data using the Cluster Model. From the initial data, it is divided into 3 clusters, including:
- Cluster 0: 835 items
 - Cluster 1: 110 items
 - Cluster 2: 401 items

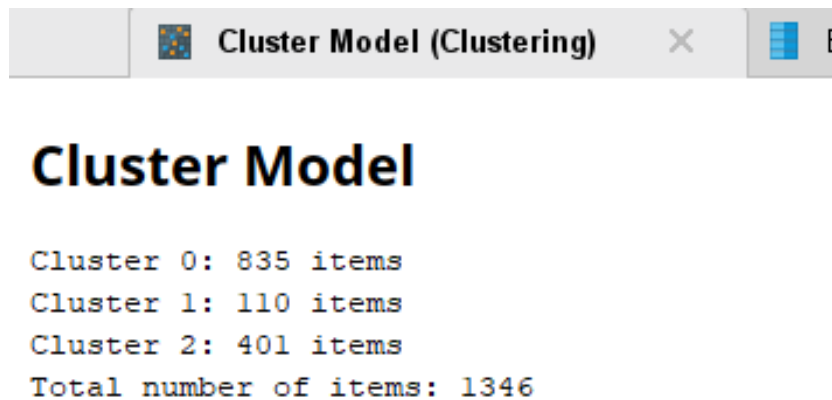


Fig. 2: shows the Cluster Model

- b Display data centroids using the Centroid Table. The following data centroids are used:

Attribute	cluster_0	cluster_1	cluster_2
calories	94.955	565.118	329.377
proteins	6.818	16.644	14.807
fat	2.553	40.833	8.940
carbohydrate	12.292	38.555	49.053
name	671.582	687.200	673.736

Fig. 3: Displays the Centroid Table

Based on Figure 3 above, it can be concluded that the centroid data used in RapidMiner Studio is as follows:

1. Cluster 0: Indonesian food which has low calories, low protein, low fat, low carbohydrate.
 2. Cluster 1: Indonesian food which has high calories, high protein, high fat, moderate carbohydrate.
 3. Cluster 2: Indonesian food which has moderate calories, moderate protein, moderate fat, high carbohydrate.
- c Displays statistics on each attribute

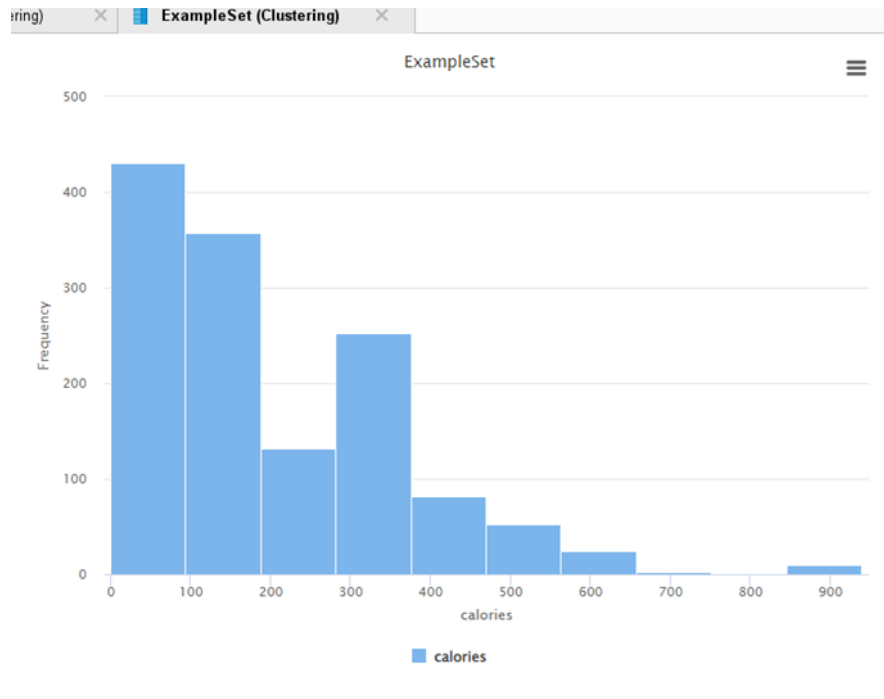


Fig. 4: Displays the Calories Attribute Statistics

Based on Figure 3 above, it can be seen that there are 431 foods that are low in calories and there are 10 foods that are high in calories.

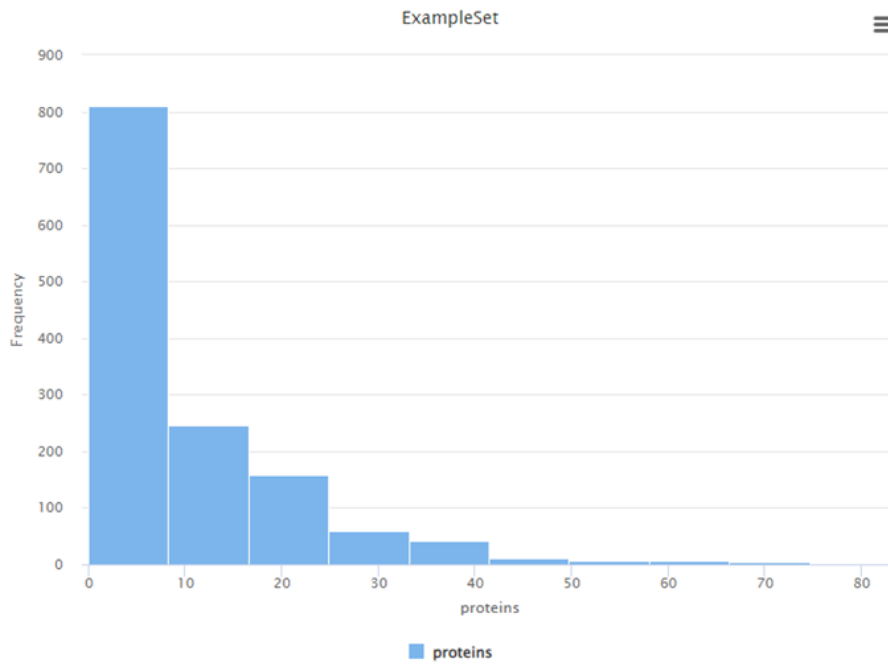


Fig. 5: Displays Protein Attribute Statistics

Based on figure 4 above, it can be seen that there are 810 foods that have low protein and there are 5 foods that have the highest protein

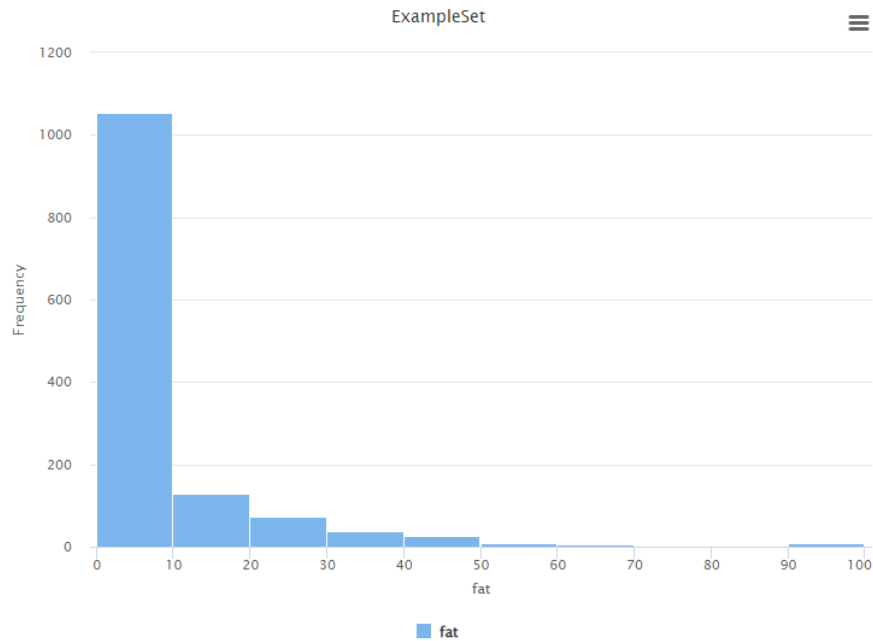


Fig. 6: Displays Fat Attribute Statistics

Based on Figure 5 above, it can be seen that there are 1,055 foods that have low fat content. So there are around 78.31% of foods that are low in fat.

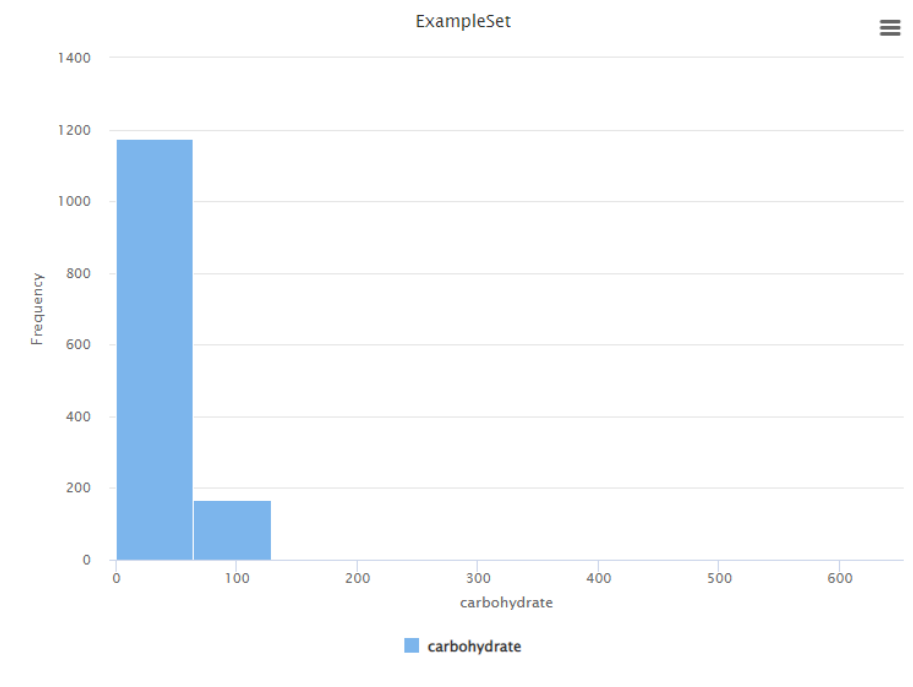


Fig. 6: Displays Carbohydrate Attribute Statistics

Based on figure 6 above, it can be seen that there are 1,175 foods that have carbohydrates and there are 169 foods that have high carbohydrates.

The final result of the clustering process with the help of Rapid Miner Studio is knowledge regarding the nutritional composition of Indonesian foods that are often consumed by the public. With the results of this research, it is hoped that the public can sort out what foods are good for our bodies to consume.

With the help of the Rapid Miner Studio software used in this research, the following knowledge was obtained:

1. There are 835 Indonesian foods that have low calories, low protein, low fat, low carbohydrate.
2. There are 110 Indonesian foods that have high calories, high protein, high fat, moderate carbohydrate
3. There are 401 Indonesian foods that have moderate calories, moderate protein, moderate fat, high carbohydrate .

4. Conclusion

In this research, Indonesian food nutrition data was used with the help of Rapid Miner Studio. The data consists of 1,346 types of Indonesian food which includes nutrients such as calories, protein, fat and carbohydrates. The results of this research obtained 3 clusters. Cluster (0) foods with low calories, protein, fat, carbohydrates. Cluster (1) eats foods with higher calories, protein, fat and carbohydrates compared to others. Cluster (2) foods with moderate calories, protein, fat, carbohydrates. Based on statistical analysis of each attribute (calories, protein, fat, carbohydrates) various information can be concluded such as the number of foods with certain nutritional values in each attribute. It is hoped that the results of this research can help the community choose food that suits their body's needs, so that it can contribute to encouraging a healthier lifestyle for the community.

References

- [1] S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.
- [2] I. Nuryani and D. Darwis, "Analisis Clustering Pada Pengguna Brand Hp Menggunakan Metode K-Means," *Pros. Semin. Nas. Ilmu Komput.*, vol. 1, no. 1, pp. 190–211, 2021.
- [3] A. F. Sari, M. Safii, D. Suhendro, and I. S. Damanik, "Implementation Data Mining of Employment Contract Extension at Indosat Using Naïve Bayes," *J. Artif. Intell. Eng. Appl.*, vol. 1, no. 1, 2021, doi: 10.59934/jaiea.v1i1.52.
- [4] Fica Oktavia Lusiana, "Analysis Of Multiple Regression Data Mining Methods On The Prediction Of Ibtidaiyah School Registration," *J. Artif. Intell. Eng. Appl.*, vol. 1, no. 3, 2022, doi: 10.59934/jaiea.v1i3.92.
- [5] J. Nasir, "Penerapan Data Mining Clustering Dalam Mengelompokkan Buku Dengan Metode K-Means," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 690–703, 2021, doi: 10.24176/simet.v11i2.5482.
- [6] S. B. Mayaza, R. Buaton, and S. Ramadani, "Application of Data Mining in Analyzing the Effect of Parents' Employment and Education Level on Student Behavior Using the A PRIORI Method (Case Study: SDN 024769 Binjai)," *J. Artif. Intell. Eng. Appl.*, vol. 3, no. 1, 2023, doi: 10.59934/jaiea.v3i1.279.
- [7] Y. R. Sari, A. Sudewa, D. A. Lestari, and T. I. Jaya, "Penerapan Algoritma K-Means Untuk Clustering Data Kemiskinan Provinsi Banten Menggunakan Rapidminer," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 5, no. 2, p. 192, 2020, doi: 10.24114/cess.v5i2.18519.
- [8] N. Edelwais, M. Simanjuntak, and M. Sihombing, "Regional Data Mining Implementation Of Contraceptive Equipment Users In The City Of Binjai By Type Using Clustering method," *J. Artif. Intell. Eng. Appl.*, vol. 3, no. 1, 2023, doi: 10.59934/jaiea.v3i1.284.
- [9] S. Hajar, A. A. Novany, A. P. Windarto, A. Wanto, and E. Irawan, "Penerapan K-Means Clustering pada ekspor minyak kelapa sawit menurut negara tujuan," *Semin. Nas. Teknol. Komput. Sains 2020*, pp. 314–318, 2020.
- [10] A. R. Arianti, N. Novriyenni, and I. Ambarita, "The Grouping Of Types Of Tax Revenue In Binjai City Uses The K-Means Clustering Method Algorithm," *J. Artif. Intell. Eng. Appl.*, vol. 3, no. 1, 2023, doi: 10.59934/jaiea.v3i1.286.
- [11] Y. D. Darmi and A. Setiawan, "Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk," *J. Media Infotama*, vol. 12, no. 2, pp. 148–157, 2017, doi: 10.37676/jmi.v12i2.418.
- [12] Erlisa Santri and D. Brahma, "Analysis Of Poverty Level Mapping In Riau Province Using The K-Means Method," *J. Artif. Intell. Eng. Appl.*, vol. 3, no. 2, 2024, doi: 10.59934/jaiea.v3i2.414.
- [13] P. Studi and S. Informasi, "Penerapan Metode K-Means Pada Pengelompokan Pengangguran Di Indonesia," vol. 6, pp. 61–74.
- [14] Z. Aras and Sarjono, "Analisis Data Mining Untuk Menentukan Kelompok Prioritas Penerima Bantuan Bedah Rumah Menggunakan Metode Clustering K-Means (Studi Kasus : Kantor Kecamatan Bahar Utara)," *J. Manaj. Sist. Inf.*, vol. 1, no. 2, pp. 159–170, 2016.
- [15] D. Novianti, "Implementasi Algoritma Naïve Bayes Pada Data Set Hepatitis Menggunakan Rapid Miner," *Paradig. - J. Komput. dan Inform.*, vol. 21, no. 1, pp. 49–54, 2019, doi: 10.31294/p.v21i1.4979.
- [16] B. Rahmat C.T.I. et al., "Implementasi k-means clustering pada rapidminer untuk analisis daerah rawan kecelakaan," *Semin. Nas. Ris. Kuantitatif Terap. 2017*, no. April, pp. 58–60, 2017.