



## Perbandingan Hasil Klasifikasi Rasa Minuman Thai Tea yang Paling Digemari Menggunakan K-means dan K-medoids

### *Comparison of Classification Results from The Most Preferred Flavor of Thai Tea Using K-means and K-medoids*

Dita Rizki Amalia\*, Riri Narasati, Ahmad Faqih  
STMIK IKMI Cirebon

Corresponding Authors: ditarizkiamalia@gmail.com\*, narasati56@gmail.com,  
faqih\_ahmad56@yahoo.com

Riwayat Artikel: Dikirim; Diterima; Diterbitkan

#### Abstrak

Data Mining saat ini semakin marak digunakan obaik oleh instansi, perusahaan, maupun organisasi. Dalam hal ini Peneliti tertarik meneliti tentang minuman yang sedang menjadi tren dalam masyarakat karena melihat minat yang besar di kalangan masyarakat., Ranah dalam penelitian ini adalah menggunakan data mining dengan k-medoids dan k-means, dimana dalam pengelompokkan kedua algoritma ini memperoleh hal yang sama namun cara yang berbeda k-menas dengan mengambil nilai rata-rata sedangkan k-medoids dengan mengambil nilai tengah. Langkah dalam data mining adalah data seleksi, data cleaning, data authentication, data integration, dan data transformation. Hasil penelitian menunjukkan kedua metode menghasilkan cluster yang sama yaitu rekomendasi rasa thai tea original dengan susu dan greentea dengan susu dengan nilai centroid 0,286 sesuai dengan target penelitian.

**Kata kunci:** Data Mining, K-Means, K-Medoids B, Clustering.

#### Abstract

Data Mining is increasingly used by agencies, companies and organizations. In this case the researcher is interested in researching about drinks that are becoming a trend in society because they see a great interest among the community. The realm in this research is to use data mining with k-medoids and k-means, where in the grouping these two algorithms get things that same but different ways k-menas by taking the average value while k-medoids by taking the middle value. The steps in data mining are data selection, data cleaning, data authentication, data integration, and data transformation. The results showed that both methods produced the same cluster, namely the recommendation of the original Thai tea flavor with milk and greentea with milk with a centroid value of 0.286 according to the research target.

**Keywords:** Data Mining, K-Means, K-Medoids B, Clustering.

#### PENDAHULUAN

Salah satu bahan minuman alami yang sangat populer dimasyarakat diantaranya teh. Teh mengandung komponen bioaktif yang disebut polifenol. Secara umum polifenol dalam tanaman terdiri atas flavonoid dan asam fenolat. Golongan terbesar dari polifenol adalah flavonoid yang merupakan yang juga sangat efektif digunakan sebagai antioksidan (Dea dkk, 2018). Dewasa ini tren minuman berbahan dasar teh, diantaranya thai tea, dimana minuman ini sudah banyak dikonsumsi oleh berbagai kalangan rasa dan juga berbagai rasa yang menarik. Melihat fenomena ini maka penulis tertarik untuk mengetahui rasa mana yang paling diminati dengan metode data mining.

“Data Mining merupakan disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data” (Han dkk, 2011). Jadi dengan teknologi data mining ini penulis dapat menemukan pola yang tepat untuk menentukan rasa thai tea



yang paling banyak dipilih oleh pembeli. Ada salah satu metode yang dapat digunakan dalam data mining yaitu clustering merupakan metode yang menggunakan teknik unsupervised learning di mana pada prosesnya tidak membutuhkan label pada dataset seperti pada data training (Tiwari dan Singh, 2011).

Clustering memiliki beberapa jenis algoritma, diantaranya adalah k-means dan k-medoids. K-means Algoritma K-Means merupakan algoritma klasterisasi yang mengelompokkan data berdasarkan titik pusat klaster (centroid) terdekat dengan data ciri khas lainnya adalah data dikelompokkan dengan memaksimalkan kemiripan data dalam satu klaster dan memperkecil kemiripan data antar klaster. Menurut penelitian (Wu et al, 2008) dan (Macqueen, 1967) pengelompokan data dengan menggunakan k-means ini lebih cepat dan sederhana. Sedangkan k-medoids adalah menggunakan medoid (median) dan tidak menentukan nilai rata - rata dari objek dalam cluster sebagai titik acuan, melainkan objek yang paling terletak dipusat sebuah cluster sehingga meminimalisir jumlah dari ketidaksamaan setiap objek dan titik acuan yang sesuai (medoid) (Sofi dkk, 2017).

Dengan adanya kelebihan dari masing-masing clustering ini peneliti dapat melakukan komparasi dari sebuah dataset yang sama. Hal ini juga sangat penting untuk menentukan kebijakan dalam sebuah perusahaan maupun organisasi. Karena dapat menjadi penentu keberlanjutan dari sebuah organisasi maupun perusahaan. Dalam penelitian akan melakukan perbandingan dalam sebuah dataset dari hasil kuisisioner pelanggan minuman thai tea.

Dalam penerapan data mining untuk penelitian ini peneliti menggunakan sebuah tools yang dapat merepresentasikan hasil dari kedua algoritma clustering yaitu rapid miner RapidMiner adalah software/perangkat lunak untuk pengolahan data, dengan menggunakan prinsip dan algoritma data mining, Rapidminer mengekstrak pola-pola dari data set yang besar dengan mengkombinasikan metode statistika, kecerdasan buatan dan database (Silalahi, 2017), dengan adanya tools ini dapat membantu peneliti dalam melihat perbandingan antara penggunaan kedua metode dalam data mining yang paling sesuai. Hasil yang diharapkan dari penelitian ini adalah dapat menunjukkan adanya rekomendasi rasa minuman thai tea yang paling sesuai untuk konsumen.

## **METODE**

### **1. Tahapan Penelitian**

Proses penelitian ini melalui empat tahapan, tahapan pertama adalah pra penelitian, dimana pada tahapan ini peneliti akan melakukan identifikasi masalah, studi pustaka, dan pengumpulan data melalui kuisisioner. Tahap kedua adalah proses persiapan data mining dimana ada lima langkah yaitu :

1. Data Selection
2. Data Cleaning
3. Data Authentication
4. Data Integration
5. Data Transformation.

Tahap yang ketiga adalah proses pengelompokan data menggunakan algoritma yang diinginkan. Dan tahapan keempat adalah knowledge presentation yaitu mempresentasikan informasi yang dibutuhkan (Heni, 2017).

#### **1. Data Seleksi**

Data Selection merupakan proses meminimalkan jumlah data yang digunakan untuk proses mining dengan tetap merepresentasikan data aslinya. Data selection dapat berupa sampling, denoising, dan feature extraction.

#### **2. Praproses (Data Cleaning)**

Data Cleaning merupakan proses untuk dapat mengatasi nilai yang hilang, noise dan data yang tidak konsisten



3. Transformasi

Data transformation dilakukan untuk mengubah bentuk dan format data. Hal ini sangat membantu memudahkan pengguna pada proses mining atau memahami hasil yang didapat.

4. Pengujian Sistem

Pada fase ini yang dilakukan adalah menerapkan algoritma atau metode pencarian pengetahuan. Pada proses ini menggunakan algoritma k-means.

5. Interpretasi dan Evaluasi

Setelah dilakukan pengelompokan via algoritma K-means maka diperoleh Knowledge Presentation yaitu proses yang dapat merepresentasikan informasi yang dibutuhkan

**2. Algoritma k-means**

Langkah-langkah dalam algoritma k-means adalah sebagai berikut:

1. Tentukan jumlah kluster (k) terlebih dahulu, tetapkan pusat kluster sembarang.
2. Hitung jarak setiap data ke pusat kluster menggunakan persamaan:

$$d_{ik} = \sqrt{\sum_{j=1}^m \{X_{ij} + C_{kj}\}^2} \quad (1)$$

3. Kelompokkan data ke dalam kluster yang dengan jarak yang paling pendek menggunakan:

$$\min \sum_{k=1}^k d_{ik} \quad (2)$$

4. Hitung pusat kluster yang baru menggunakan persamaan (Sismadi, 2013):

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

**3. Algoritma K-medoids**

Langkah-langkah K-Medoids adalah:

1. Pilih poin k sebagai inisial centroid / nilai tengah (medoids) sebanyak k cluster.
2. Cari semua poin yang paling dekat dengan medoids, dengan cara menghitung jarak vektor antar dokumen dengan menggunakan Euclidian Distance.

Rumusnya adalah sebagai berikut:

$$d(x, y) = \sqrt{\sum_{i=1}^n |x_i - y_i|^2} \quad (4)$$

Dimana:

- d(x,y) = jarak antara data ke-i dan data ke-j
- x<sub>i1</sub> = nilai atribut ke satu dari data ke-i
- y<sub>j1</sub> = nilai atribut ke satu dari data ke-j
- n = jumlah atribut yang digunakan

3. Secara acak, pilih poin yang bukan medoids.
4. Hitung total jarak antar medoid.
5. Jika TD baru < TD awal, tukar posisi medoids dengan medoids baru, jadi medoids baru.
6. Ulangi langkah 2-5 sampai medoids tidak berubah.

**HASIL DAN PEMBAHASAN**

Berikut ini adalah proses untuk kedua algoritma dengan data yang sama:

1. Data Seleksi



Data yang dipilih dalam penelitian berasal dari data pelanggan thai tea di kota Cirebon dengan menggunakan random sample dikarenakan jumlah penduduk diketahui namun gerai yang ada di kota Cirebon tidak diketahui:

Tabel 1:  
Data Kuesioner

No	Responden	SOAL																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	R1-R	3	4	5	4	3	3	4	3	3	5	4	3	3	3	4	3	4
2	R2-R	5	5	5	5	5	5	5	3	3	3	4	5	5	4	1	1	1
3	R3-R	4	3	3	3	4	4	4	4	4	5	5	5	3	4	3	4	3
4	R4-R	4	4	5	5	5	4	5	4	4	4	5	5	4	4	4	5	4
5	R5-R	5	2	2	5	3	5	4	3	4	3	4	2	2	2	3	3	4
6	R6-R	4	3	3	5	3	3	3	4	4	2	4	3	4	2	3	4	2
7	R7-R	4	3	4	4	4	4	4	4	4	3	4	4	4	4	3	4	3
8	R1-D	4	4	4	5	5	5	5	5	4	4	3	3	3	4	4	4	5
9	R2-D	3	3	4	3	5	2	4	4	3	3	3	5	4	5	4	4	4
10	R3-D	3	2	2	4	3	4	4	4	3	3	4	2	2	2	3	3	3
11	R4-D	4	2	3	5	5	4	4	3	3	3	4	4	4	3	3	3	3
12	R5-D	4	4	4	5	5	2	4	4	5	4	4	3	4	3	4	5	4
13	R6-D	5	2	4	4	2	4	4	2	5	2	4	3	3	2	5	5	2
14	R7-D	5	4	5	4	5	4	5	5	5	4	4	5	5	4	4	5	4
15	R1-L	4	4	5	4	4	4	4	4	3	4	4	3	3	4	4	4	4
16	R2-L	3	2	2	4	3	4	4	2	4	2	4	3	3	3	4	4	3
17	R3-L	4	2	4	4	4	3	3	3	4	3	4	2	2	2	3	3	3
18	R4-L	4	3	3	4	4	3	3	3	4	2	4	4	4	2	2	4	2
19	R5-L	5	2	5	5	3	5	5	5	5	3	3	5	5	1	5	5	3
20	R6-L	5	5	5	3	5	5	5	3	1	5	1	5	1	5	5	1	5
21	R7-L	4	4	4	4	1	5	4	4	4	2	4	1	1	1	3	4	3

## 2. Data Cleaning

Dalam proses ini peneliti menghilangkan data yang tidak sesuai dari item yang tidak sesuai dari pembahasan penelitian, maka diperoleh hasil sebagai berikut:

Tabel 2:  
Data Kuisioner yang Telah Disaring

Responden	Thai Tea Dengan Susu	Thai Tea Tanpa Susu	Cokelat	Kopi Dengan Susu	Kopi Tanpa Susu	Green Tea Dengan Susu	Greentea Tanpa Susu
R1-R	3	5	4	3	3	3	4
R2-R	3	3	4	5	4	1	1
R3-R	4	5	5	3	4	4	3
R4-R	4	4	5	4	4	5	4
R5-R	4	3	4	2	2	3	4
R6-R	4	2	4	4	2	4	2
R7-R	4	3	4	4	4	4	3
R1-D	4	4	3	3	4	4	5
R2-D	3	3	3	4	5	4	4
R3-D	3	3	4	2	2	3	3
R4-D	3	3	4	4	3	3	3
R5-D	5	4	4	4	3	5	4
R6-D	5	2	4	3	2	5	2
R7-D	5	4	4	5	4	5	4
R1-L	3	4	4	3	4	4	4
R2-L	4	2	4	3	3	4	3
R3-L	4	3	4	2	2	3	3
R4-L	4	2	4	4	2	4	2
R5-L	5	3	3	5	1	5	3



R6-L	1	5	1	1	5	1	5
R7-L	4	2	4	1	1	4	3

### 3. Data Transformation

Dalam proses ini peneliti melakukan pengecekan apakah kolom data sudah sesuai dengan format, jika sudah maka dilakukan ke proses selanjutnya.

### 4. Data Integration (Pengujian sistem)

Disinilah proses clustering dilakukan

Kluster = 6

Dengan pembagian 6 kluster maka diperoleh hasil berikut:

Tabel 3:

K-means dengan 6 Klaster

Cluster	Rasa	Nilai Centroid
cluster_0	Kopi Dengan Susu	0,143
cluster_1	Greentea Tanpa Susu	0,143
cluster_2	Kopi Tanpa Susu	0,143
cluster_3	Cokelat	0,143
cluster_4	Thai Tea Dengan Susu	0,286
cluster_4	Green Tea Dengan Susu	0,286
cluster_5	Thai Tea Tanpa Susu	0,143

Dari tabel 3 diperoleh data bahwa cluster dengan nilai tertinggi untuk k-means adalah cluster 4 dengan nilai 0,286 yaitu rasa thai tea dengan susu dan greentea dengan susu.

Kluster=6

Tabel 4:

K-medoids dengan 6 Klaster

Cluster	Rasa	Nilai Centroid
cluster_0	Kopi Dengan Susu	0,143
cluster_1	Greentea Tanpa Susu	0,143
cluster_2	Kopi Tanpa Susu	0,143
cluster_3	Cokelat	0,143
cluster_4	Thai Tea Dengan Susu	0,286
cluster_4	Green Tea Dengan Susu	0,286
cluster_5	Thai Tea Tanpa Susu	0,143

Dari tabel dapat terlihat bahwa rasa yang paling disukai berdasarkan clustering dengan k-medoids adalah rasa pada cluster 4 yaitu thai tea dan green tea dengan susu.

### 5. Data Transformation

#### a. K-means

Dari hasil perolehan data maka dapat disimpulkan bahwa rasa yang paling disukai dengan metode k-means adalah rasa greentea dan thai tea dengan susu, hal ini ditunjukkan dalam diagram berikut ini sesuai dengan cluster seperti dalam gambar 1.

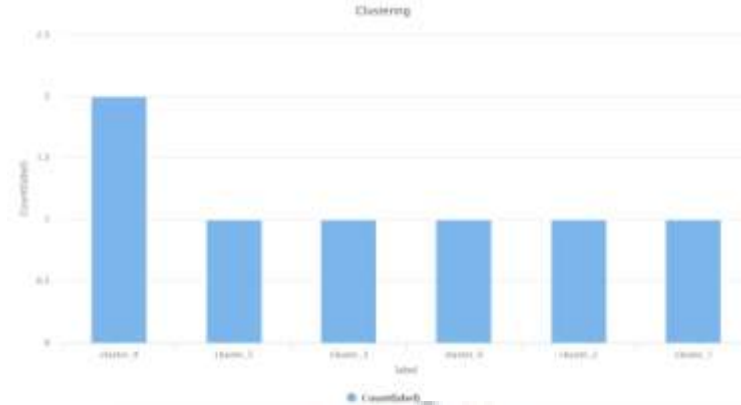
#### b. K-medoids

Dari hasil perolehan data maka dapat disimpulkan bahwa rasa yang paling disukai dengan metode k-medoids adalah rasa greentea dan thai tea dengan susu, hal ini ditunjukkan dalam diagram berikut ini sesuai dengan cluster seperti dalam gambar 2.

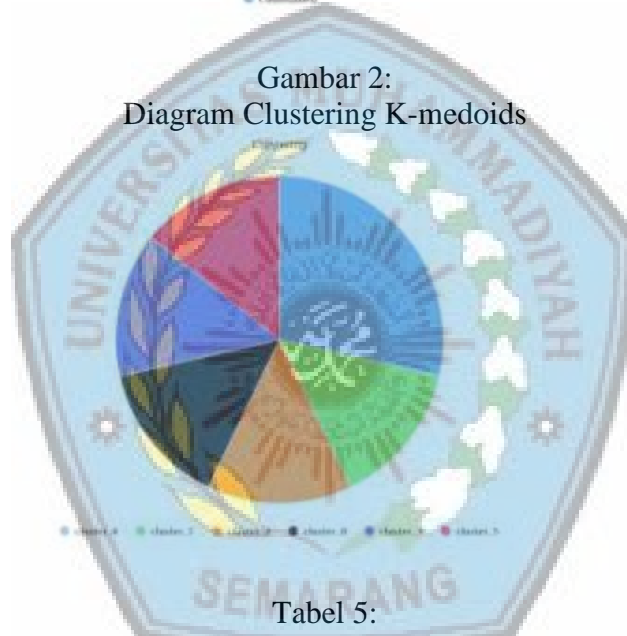
Akan tetapi walaupun dihasilkan output yang sama penulis menemukan hal yang berbeda dalam proses clusterisasi dalam data k-means dan k-medoids diantaranya sebagaimana dalam

Tabel 5. Dari tabel tersebut dihasilkan sebuah temuan, walaupun memiliki hasil yang sama, namun dalam proses peletakan cluster untuk item lain berbeda.

Gambar 1:  
Diagram Clustering K-means



Gambar 2:  
Diagram Clustering K-medoids



Tabel 5:  
Perbandingan K-medoids dan K-Means

Rasa	K-medoids	K-means
Kopi Dengan Susu	cluster_0	cluster_0
Thai Tea Tanpa Susu	cluster_1	cluster_5
Cokelat	cluster_2	cluster_3
Kopi Tanpa Susu	cluster_3	cluster_2
Thai Tea Dengan Susu	cluster_4	cluster_4
Green Tea Dengan Susu	cluster_4	cluster_4
Greentea Tanpa Susu	cluster_5	cluster_1



## KESIMPULAN

Dari hasil pembahasan diperoleh hasil cluster yang sama untuk rekomendasi rasa yaitu rasa green tea dan thai tea dengan susu, walaupun dengan nilai centroid yang berbeda antara kedua metode tersebut, hal ini dikarenakan kedua algoritma memiliki cara yang berbeda dalam memetakan ke dalam cluster masing-masing. Untuk kedepannya diharapkan penelitian ini menjadi bahan pertimbangan untuk membandingkan metode yang berbeda pada data yang sama serta dapat dijadikan acuan sebagai kebijakan baik lembaga, perusahaan, maupun organisasi.

## SARAN

Untuk penelitian ini sebaiknya bisa dikembangkan lebih lanjut ke dalam skala data yang lebih besar agar menghasilkan kebijakan yang lebih mencakup banyak pihak, serta mencakup banyak kalangan serta usia. Dalam perkembangannya semoga dapat menjadi bahan pertimbangan bagi pelaku bisnis dalam mengembangkan usahanya dalam berbagai bidang.

## UCAPAN TERIMA KASIH

1. Kemenristekdikti sebagai pihak yang memberikan kesempatan dalam melakukan penelitian ini
2. STMIK IKMI Cirebon sebagai wadah dalam memberikan izin penelitian
3. Dinas Perindustrian dan Perdagangan Kota Cirebon
4. Dinas Kependudukan dan Catatan Sipil Kota Cirebon
5. Gerai Mandja Cheese Tea

## DAFTAR PUSTAKA

- Dea Ira Lelita, Ir. Rohadi, M.P., Aldila Sagitaning Putri, S.Si, M.Si. Sifat Antioksidatif Ekstrak Teh (*Camellia Sinensis* Linn.) Jenis Teh Hijau, Teh Hitam, Teh Oolong dan Teh Putih Dengan Pengeringan Beku (Freeze Drying). 2018. Jurnal USM. Semarang. ]
- Han, J. Kamber, M & Jian, Pei. 2011. Data Mining : Concepts and techniques, Third Edition. America: Morgan Kauffman, San Francisco.
- Harival Zayuka, Surya Michrandi Nasution, ST., MT., Yudha Purwanto, ST., MT. e- Proceeding of Engineering : Vol.4, No.2 PP 2182-2190. 2017.
- Heni Sulastri, Acep Irham Gufroni. Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia. Jurnal Teknologi dan Sistem Informasi - Vol. 03 No. 02, pp 299-305, 2017.
- J. Macqueen, "Some methods for classification and analysis of multivariate observations," Proc. 5th Berkeley Symp. Math. Stat., vol. 233, no. 233, pp. 281-297, 1967.
- M. Tiwari and R. Singh, "Comparative Investigation of K-Means and K-Medoid Algorithm on Iris Data," vol. 4, no. 8, pp. 69-72, 2012.
- Silalahi Mesri, JURNAL TEKNOLOGI DAN SISTEM INFORMASI - VOL. 03 NO. 02 .PP 299-305. 2017.
- Sismadi, "Vol. IX No.1, Maret 2013 Techno Nusa Mandiri PERBANDINGAN TINGKAT AKURASI PEMETAAN ALGORITMA K-MEANS DAN FUZZY C-MEANS NASABAH ASURANSI JIWA SYARIAH," Techno Nusa Mandiri, vol. IX, no. 1, pp. 109-119, 2013.
- Sofi Defiyanti<sup>1</sup>, Mohamad Jajuli, Nurul Rohmawati. JURNAL TEKNOSI, Vol. 03, No. 01, pp211-218, April 2017
- X. Wu et al., "Top 10 algorithms in data mining," Knowl. Inf. Syst., vol. 14, no. 1, pp. 1-37, 2008.