p-ISSN: -
e-ISSN: -
NEI Volume 1 No 1 Desember 2025
NUSANTARA ENGINEERING INSIGHT
Engineering and Agroindustry Journal
Journal Homepage: nrri-journal.com/index.php/nei-j

Implementasi Metode Cluster Preferensi Pelanggan Teh Kemasan dengan K-Means

(*Implementation of the Packaged Tea Customer Preference Cluster Method Using K-Means*)

Juwita Sari[1*], M. Agusfian[2], Wahyu Adi Guna [3]

[1] Departemen Teknologi Agroindustri, Fakultas Teknologi Pertanian, IPB University, Indonesia. juwitasari02121998@gmail.com
[2] Cendekia Nusantara Institut, Nusa Tenggara Barat, Indonesia. magusfian89@gmail.com
[3] Generasi Agroindustri Lestari (GALERI) Nusa Tenggara Barat, Indonesia. wahyuadiguna1999@gmail.com

* Corresponding Author: juwitasari02121998@gmail.com

ARTICLE INFO

Article history:
Received Juni 23, 2025
Revised September 12, 2025
Accepted November 1, 2025
Available online 10 December, 2025

Kata Kunci: Cluster, preference, packaged tea, K-Means

This is an open access article under the CC BY-SA license. Copyright © 2025 by Author. Published by Nusantara Engineering Insight

A B S T R A K

Tea is a product widely consumed by the Indonesian population. Amid increasing market competition, an appropriate marketing strategy is essential. One approach to optimizing marketing efforts is to conduct market segmentation through customer preference clustering. This study utilized a packaged tea sales dataset (secondary data) obtained from Kaggle.com and processed it using the K-Means clustering method through the WEKA tool. The clustering results produced five cluster patterns, with the most dominant patterns found in two clusters: JAVANA tea and Sosro bottled tea. Based on the RFM model analysis, JAVANA tea was identified as the most preferred product, particularly among young consumers, with a time span of 15:46. Meanwhile, Sosro bottled tea ranked second, preferred by consumers across all age groups, with a time span ranging from 09:58 to 16:05.

1. PENDAHULUAN

Teh merupakan minuman yang sangat populer di Indonesia dan banyak digemari masyarakat. Cita rasa serta aromanya yang khas membuat teh menjadi pilihan minuman yang disukai banyak orang. Dari sisi kesehatan, teh diketahui memiliki berbagai manfaat, antara lain membantu memperbaiki sel tubuh yang mengalami kerusakan, menjaga kelembutan kulit, membantu program penurunan berat badan, serta berperan dalam pencegahan kanker dan penyakit jantung. Selain itu, teh juga dapat membantu menurunkan kadar kolesterol serta mendukung kelancaran peredaran darah (Soraya, 2007).

Industri minuman teh kemasan di Indonesia telah berkembang pesat dalam dekade terakhir, didorong oleh perubahan gaya hidup konsumen modern yang menginginkan produk berkualitas tinggi dengan kenyamanan konsumsi (Miao, 2025). Namun, dalam lanskap pasar yang semakin kompetitif, produsen teh kemasan menghadapi tantangan signifikan dalam memahami keragaman preferensi pelanggan yang terus berkembang.

Setiap konsumen memiliki kebutuhan, preferensi, dan perilaku pembelian yang unik, yang dipengaruhi oleh berbagai faktor seperti harga, rasa, desain kemasan, merek, dan kesadaran kesehatan. Pemahaman yang mendalam tentang segmentasi preferensi pelanggan menjadi kunci strategis bagi perusahaan untuk merancang strategi pemasaran yang efektif, mengoptimalkan pengembangan

produk, dan meningkatkan kepuasan pelanggan. Sehingga dalam persaingan pasar yang sangat ketat perusahaan dituntut untuk memiliki strategi dalam pemasaran. Salah satu upaya untuk memaksimalkan pemasaran tersebut yaitu melakukan analisis segmentasi pasar yaitu dengan mengklaster preferensi pelanggan.

Teknologi data mining dan *machine learning* menawarkan solusi inovatif untuk menganalisis pola kompleks dalam perilaku konsumen yang sebelumnya sulit diidentifikasi melalui metode konvensional (Kapti & Dwi Astuti, 2025). Di antara berbagai algoritma *clustering* yang tersedia, K-Means merupakan salah satu metode paling efektif dan efisien untuk segmentasi pelanggan dalam berbagai konteks industri (Lathifah & Azzahra, 2025). Algoritma ini mampu mengelompokkan konsumen dengan karakteristik serupa ke dalam *cluster* yang homogen secara internal namun berbeda secara eksternal, sehingga pola preferensi dapat diidentifikasi dengan lebih jelas (Hidayat *et al*., 2025).

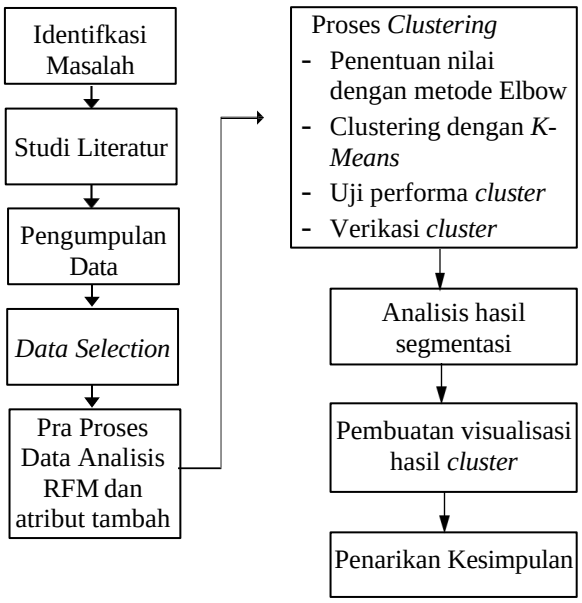
Dalam konteks data mining, *clustering* merupakan teknik yang bertujuan untuk mengelompokkan data ke dalam beberapa himpunan, di mana setiap kelompok memiliki tingkat kemiripan yang tinggi dibandingkan kelompok lainnya (Kodinariya, 2013). Jumlah *cluster* memengaruhi kualitas hasil pengelompokan: jumlah *cluster* yang terlalu sedikit dapat menyederhanakan struktur data namun berpotensi menghilangkan informasi penting, sedangkan jumlah *cluster* yang terlalu banyak dapat meningkatkan kesamaan dalam kelompok namun mengaburkan struktur data secara keseluruhan (Kuo, 2015).

K-Means merupakan salah satu metode *clustering* yang paling banyak digunakan karena kemampuannya meminimalkan kesalahan pengelompokan pada ruang Euclidean dan menghasilkan pembagian kelompok yang jelas berdasarkan kemiripan karakteristik data (Li, 2008).

Berdasarkan hal tersebut, penelitian ini bertujuan untuk menerapkan metode K-Means dalam mengelompokkan pelanggan berdasarkan nilai RFM, mengidentifikasi preferensi konsumen terhadap produk teh kemasan melalui analisis frekuensi pembelian, serta memetakan produk yang paling potensial untuk dipromosikan kepada segmen pelanggan yang tepat.

2. METODE

Penelitian ini dilakukan dengan mengolah dataset penjualan produk teh kemasan (data sekunder) yang diambil dari Kaggle.com. Selanjutnya data yang diperoleh dioleh dengan pengklasteran metode K-Means menggunakan *tools* WEKA. Adapun alur penelitian dapat dilihat pada Gambar 1.



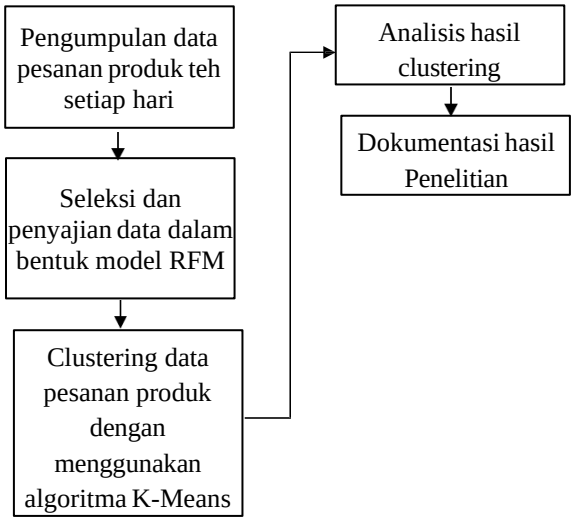
Gambar 1. Alur Penelitian

2.1. Model RFM

1. Kriteria R (recency) berarti skala kekinian sebuah produk diukur ketika produk ditransaksikan atau dipesan.
2. Kriteria F (frequency) Membutuhkan atribut yang merepresentasikan berapa kali produk dipesan. Kriteria ini ini dapat dilihat dari berapa banyak produk dengan nama yang sama muncul dalam data histori.
3. Atribut yang dibutuhkan adalah atribut jumlah total setiap produk yang dipesan selama periode analisis.
4. Kriteria M (*monetary*) atribut terakhir dari RFM adalah monetary (terlihat pada kolom kelima tabel 3). Atribut ini berhubungan dengan harga atau keuntungan yang didapat bila barang

tersebut terjual.

Algoritma K-Means adalah algoritma klasik untuk memecahkan masalah *clustering* (Zhao *et al.*, 2010). Metode K- Means merupakan metode non-hirarkis. Metode ini merupakan teknik penyekatan (partition) yang membagi atau memisahkan objek k ke daerah yang terpisah (Kodinariya *et al.*, 2013). Metode K-Means digunakan untuk mengelompokkan n vektor berdasarkan atribut partisi k, dimana k <n, tergantung pada beberapa tindakan. Pusat *cluster* adalah mean (nilai tengah) semua vector pada *cluster* tertentu. Algoritma ini dimulai dengan memilih k centroid awal secara acak (*randomly*), kemudian memberikan nilai vektor ke centroid terdekat dengan Euclidean distance dan menghitung ulang centroid baru. Proses ini berulang sampai vektor tidak lagi mengubah *cluster* antar iterasi (Birant, 2011). Alur algoritma K-Means ditunjukkan pada gambar 2 berikut:



Gambar 2. Tahapan *Clustering* K-Means

3. HASIL DAN PEMBAHASAN

Data yang digunakan untuk pemodelan *clustering* adalah data penjualan produk teh kemasan sebanyak 6.641 data dengan 5 atribut yaitu jenis produk, quantity, invoice date, unit price dan customer id. Sesuai dengan kebutuhan RFM, data dilakukan *preprocess* dengan memilih data yang diperlukan yaitu jenis produk, *quantity*, *invoice date*, unit price dan customer ID. Berikut uraian mengenai atribut yang dibutuhkan untuk model RFM :

a. Kriteria R (*recency*)

Recency berarti skala kekinian sebuah produk diukur ketika produk ditransaksikan atau dipesan. Pada penelitian ini penilaian menggunakan skala 1-4 secara berurutan. Semakin kecil nilai skala R maka produk tersebut semakin usang dari segi pemesanan. Detail penilaian Recency ditampilkan pada Gambar 3

b. Kriteria F (*frequency*)

Membutuhkan atribut yang merepresentasikan berapa kali produk dipesan. Kriteria ini dapat dilihat dari berapa kali produk dengan nama yang sama muncul dalam data histori. Atribut yang dibutuhkan adalah atribut jumlah total setiap produk yang dipesan selama periode analisis.

c. Kriteria M (*monetary*)

Atribut terakhir dari RFM adalah monetary (terlihat pada Gambar 3). Atribut ini berhubungan dengan harga atau keuntungan yang didapat bila barang tersebut terjual.

Final cluster centroids:

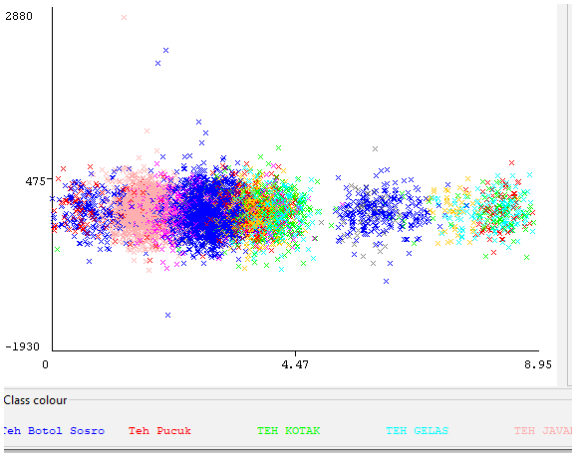
Attribute	Full Data (6641.0)	Cluster# 0 (1162.0)	1 (1601.0)
Jenis Produk	Teh Botol Sosro	Teh Botol Sosro	Teh Botol Sosro
Quantity	13.4757	12.1007	10.5078
InvoiceDate	12/6/2020 9:58	12/14/2020 12:58	1/23/2021 13:34
UnitPrice	3.0871	2.9861	3.3048
CustomerID	15386.8794	13444.4389	15277.8495

Attribute	2 (1400.0)	3 (1845.0)	4 (633.0)
Jenis Produk	Teh Botol Sosro	TEH JAVANA	Teh Botol Sosro
Quantity	11.7193	21.7024	3.4123
InvoiceDate	12/12/2020 16:05	12/2/2020 15:46	12/6/2020 9:58
UnitPrice	3.0287	1.6426	7.0617
CustomerID	17341.0164	15236.7381	15344.054

Gambar 3. Hasil Cluster Data Penjualan Teh Kemasan

Hasil *cluster* menunjukkan terdapat 5 *cluster* bahwa jumlah yang paling banyak adalah pada produk teh JAVANA (*cluster* 3) dengan total transaksi 1.6426 pada pukul 15.46 oleh customer dengan Id 15236-7381 yaitu kalangan muda. Kemudian *cluster* 1 menunjukkan terbanyak ke 2 yaitu produk teh botol sosro dengan jumlah transaksi 3.3048 pada pukul 13.34 oleh customer 15277.8435 yang merupakan konsumen anak muda dan kalangan dewasa. *Cluster* 2 tetap pada produk teh botol sosro dengan

jumlah produk yang dibeli 11.7293 dengan total transaksi 3.0287 pukul 14.05 oleh id pelanggan 17341-0164 yaitu kalangan remaja. Pada *cluster* 0 merupakan jumlah kuantitas produk terjual terbanyak ke 4 yang merupakan produk teh botol sosro juga dengan jumlah 12.1007 pada pukul 12.58 total transaksi 2.9841 oleh pelanggan id 13444 - 4385 kalangan anak muda dan remaja. Pada *cluster* 4 merupakan produk teh botol sosro juga dengan kuantitas terendah dari 5 *cluster* yaitu jumlah produk terjual 3.4123 pada pukul 9.58 dengan total transaksi 7.041



Gambar 4. Hasil Visualisasi *Clustering*

Dari hasil visualisasi menunjukkan bahwa terdapat 5 pola *cluster* yang terbentuk dan yang paling padat adalah *cluster* merah muda dan biru. Dimana *cluster* terbanyak ialah teh JAVANA dan kedua teh botol sosro.

3. KESIMPULAN

1. Beradasarkan hasil clustering terdapat 5 pola *cluster* terbentuk dan pola paling banyak terdapat pada 2 *cluster* yaitu pada produk teh JAVANA dan teh botol sosro.
2. Dari hasil *cluster* dengan menganalisis model RFM maka produk yang menjadi preferensi paling banyak adalah porduk teh JAVANA yang diminati oleh kalangan anak muda dengan rentang waktu 15.46 dan kedua ialah produk teh botol sosro yang diminati oleh semua kalangan pada rentang waktu 9.58-16.05.

SARAN

Disarankan kepada perusahaan atau retailer untuk mempromosikan teh JAVANA dan teh botol sosro di *market place* atau *platform digital* lainnya pada waktu menjelang siang hingga sore dengan kedua produk terutama produk JAVANA dikalangan anak muda.

DAFTAR PUSTAKA

Aditya, Tubagus. 2002. Optimalisasi Distribusi Teh Botol Sosro di PT Sasana Caraka Mekarjaya Unit Cakung Tugu. Skripsi. Jurusan Ilmu-ilmu Sosial Ekonomi Pertanian. Fakultas

D. Birant, Data Mining Using RFM Analysis, in Knowledge-Oriented Applications in Data Mining, Turkey: KImito Funatso, 2011, pp. 91–108

Hidayat, N., D. P. Gevano, A. Putra, and R. Ilahi. 2025. Segmentasi Pelanggan Menggunakan K-Means *Clustering* Berdasarkan Data Kepribadian dan Pola Konsumsi 6(5):3914–3924.

J. Zhao, W. Zhang, and Y. Liu 2010. Improved K-Means cluster algorithm in telecommunications enterprises customer segmentation, Proc. 2010 IEEE Int. Conf. Inf. Tehory Inf. Secur. ICITIS 2010, pp. 167–169

Kapti dan Dwi Astuti. 2025. Implementasi Algoritma K - Means Clustering untuk Segmentasi Pelanggan Implementation of K - Means Clustering Algorithm for 14:2751–2767.

Kodinariya, Trupti M. & Makwana, Prashant R.,(2013). Review on determining number of cluster in K-Means Clustering. International Journal of Advance Research in Computer Science and Management Studies, I(6), pp. 90-95

Lathifah, S. N., and Z. F. Azzahra. 2025. AI-Driven Customers Segmentation Using K-Means Clustering. G-Tech: Jurnal Teknologi Terapan 9(1):320–329.

Miao, H. 2025. Consumer Preference and Product Differentiation in the

Chinese Tea Market : A Conjoint Analysis Approach 2(3):158–161.
Pertanian. Institut Pertanian Bogor.
R. J. Kuo, S. H. Lin, and Z.-Y. Chen 2015. Integration of Particle Swarm Optimization and Immune Genetic Algorithm- Based Dynamic Clustering for Customer Clustering, Int. J. Artif. Intell. Tools, vol. 24, no. 5, p. 1550019
Soraya, N. 2007. Sehat dan Cantik Berkat Teh Hijau. Penebar Swadaya, Jakarta.
Sumarwan, Ujang. 2002. Analisis Hubungan antara Persepsi Popularitas, Persepsi Kualitas, Persepsi Harga dengan Merek Teh Celup yang Dikonsumsi. Staff Pengajar Jurusan Gizi Masyarakat dan Sumberdaya Keluarga dan Magister Management Agribusiness. Fakultas Pertanian. Institut Pertanian Bogor.
W. Li. 2008. Modified K-Means Clustering Algorithm, 2008 Congr. Image Signal Process., pp.618–621.