

## Implementasi Algoritma K-Means dalam Pengelompokan Data Harga Laptop

**Roza Marmay<sup>1\*</sup>, Ridha Luthvina<sup>2</sup>, Okti Ulandari<sup>3</sup>**

<sup>1,2</sup>Program Studi Manajemen Logistik Industri Agro, <sup>3</sup>Program Studi Teknologi Rekayasa Bioproses Energi Terbarukan Politeknik ATI Padang

\*Corresponding Author e-mail: rozamarmay.rm@gmail.ac.id

**Abstract:** This study aims to cluster laptop price data using the K-Means algorithm to obtain a more structured and interpretable price segmentation. The data used in this study consist of laptop price data that have undergone a preprocessing stage, including data cleaning, data transformation, and data normalization to ensure optimal data quality. The determination of the optimal number of clusters was conducted using the Elbow Method by analyzing the Within Cluster Sum of Squares (WCSS) values, which indicated that the optimal number of clusters is  $k = 3$ . Subsequently, the K-Means algorithm was applied to group the laptop price data into three clusters based on price characteristics. The clustering results reveal three main groups, namely low-price, mid-price, and high-price laptop clusters. This segmentation provides a clear overview of the laptop market conditions and highlights the differences in price ranges among the clusters. The results demonstrate that the K-Means algorithm is able to cluster laptop price data effectively and consistently. The resulting segmentation can be utilized as a basis for decision-making for consumers as well as business stakeholders in developing marketing strategies and determining appropriate laptop pricing.

**Keywords:** K-Means, Clustering, Data Mining, Laptop Price, Elbow Method

**Abstrak:** Penelitian ini bertujuan untuk mengelompokkan data harga laptop menggunakan algoritma *K-Means* guna memperoleh segmentasi harga yang lebih terstruktur dan mudah dipahami. Data yang digunakan merupakan data harga laptop yang telah melalui tahap *preprocessing*, meliputi pembersihan data, transformasi data, dan normalisasi untuk memastikan kualitas data yang optimal. Penentuan jumlah *cluster* dilakukan menggunakan *Elbow Method* dengan menganalisis nilai *Within Cluster Sum of Squares* (WCSS), yang menunjukkan bahwa jumlah *cluster* optimal berada pada  $k = 3$ . Selanjutnya, algoritma *K-Means* diterapkan untuk mengelompokkan data harga laptop ke dalam tiga *cluster* berdasarkan karakteristik harga. Hasil pengelompokan menunjukkan adanya tiga kelompok utama, yaitu *cluster* harga rendah, harga menengah, dan harga tinggi. Segmentasi ini memberikan gambaran yang jelas mengenai kondisi pasar laptop serta perbedaan rentang harga antar kelompok. Hasil penelitian menunjukkan bahwa algoritma *K-Means* mampu mengelompokkan data harga laptop secara efektif dan konsisten. Segmentasi yang dihasilkan dapat dimanfaatkan sebagai dasar pengambilan keputusan bagi konsumen maupun pelaku usaha dalam menentukan strategi pemasaran dan penetapan harga produk laptop.

Kata kunci: *K-Means, Clustering, Data Mining, Harga Laptop, Elbow Method*

## Pendahuluan

Permintaan akan peralatan elektronik, terutama laptop, telah meningkat karena pertumbuhan teknologi informasi yang pesat. Di berbagai industri, termasuk bisnis, pendidikan, dan pekerjaan profesional, laptop telah menjadi alat yang sangat diperlukan. Pembeli seringkali kesulitan memilih laptop yang paling sesuai dengan kebutuhan dan anggaran mereka karena banyaknya merek laptop, karakteristik, dan variasi harga yang tersedia di pasaran.

Ketersediaan data harga laptop dalam jumlah besar belum sepenuhnya dimanfaatkan secara optimal untuk menghasilkan informasi yang mendukung pengambilan keputusan. Perbedaan rentang harga yang cukup signifikan serta variasi spesifikasi produk sering kali menyulitkan konsumen maupun pelaku usaha dalam memahami pola segmentasi harga secara objektif. Analisis harga yang dilakukan secara manual cenderung bersifat subjektif dan tidak mampu menangkap pola tersembunyi dalam data berskala besar. Oleh karena itu, diperlukan pendekatan analisis data yang sistematis dan berbasis komputasi. Salah satu pendekatan yang relevan adalah clustering, yang memungkinkan pengelompokan data berdasarkan tingkat kemiripan karakteristik tanpa memerlukan label sebelumnya. Penerapan metode clustering pada data harga laptop menjadi penting untuk mengidentifikasi segmentasi harga yang jelas dan terstruktur. Penelitian ini menawarkan penerapan algoritma K-Means sebagai solusi untuk mengungkap pola harga laptop secara objektif, sehingga memberikan kontribusi dalam analisis data produk elektronik serta mendukung pengambilan keputusan berbasis data.



Prinsip algoritma K-Means yang sederhana, efektif, dan mendasar menjadikannya salah satu teknik pengelompokan yang paling populer. Perhitungan dasar yang diterapkan yaitu berdasarkan jarak terdekat ke pusat klaster (centroid). Teknik ini membagi data menjadi banyak klaster. K-Means dapat membantu mengklasifikasikan komputer ke dalam kelompok-kelompok yang berbeda, seperti harga rendah, menengah, dan tinggi. sehingga memudahkan analisis dan pengambilan keputusan.

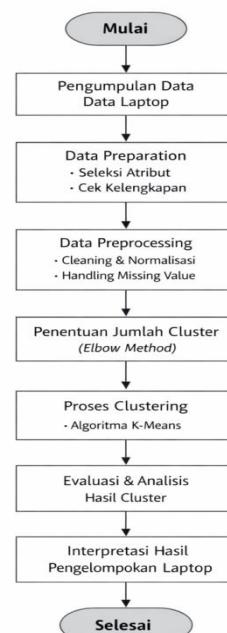
Penelitian ini bertujuan untuk mengelompokkan data harga laptop menggunakan algoritma K-Means dengan menekankan pada penyusunan segmentasi harga yang lebih aplikatif dan mudah diinterpretasikan oleh pengguna. Berbeda dengan penelitian sebelumnya yang umumnya hanya berfokus pada penerapan algoritma K-Means tanpa mengaitkan hasil clustering dengan kebutuhan pengguna, penelitian ini diawali dengan tahap pembersihan data yang lebih sistematis untuk meningkatkan kualitas data sebelum proses pengelompokan. Selain itu, penelitian ini menekankan interpretasi hasil clustering berdasarkan perspektif anggaran (budget-based segmentation), sehingga setiap cluster yang terbentuk merepresentasikan kelompok harga yang relevan dengan kemampuan finansial pengguna. Dengan pendekatan tersebut, hasil penelitian diharapkan tidak hanya menunjukkan performa algoritma K-Means, tetapi juga memberikan nilai tambah dalam mendukung pengambilan keputusan konsumen melalui pemetaan segmentasi harga laptop yang lebih kontekstual dan praktis.

## Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining, khususnya teknik *clustering* menggunakan algoritma K-Means. Tujuan penelitian adalah mengelompokkan data laptop berdasarkan harga, sehingga dapat memberikan informasi yang lebih terstruktur dalam analisis pembelian laptop.

### 1. Sumber dan Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data sekunder berupa data laptop yang mencakup nama laptop, merek, model, harga, serta spesifikasi pendukung seperti kapasitas RAM dan penyimpanan SSD. Data dikumpulkan dari sumber daring (online) melalui kaggle.com yang menyediakan informasi produk laptop. Jumlah data sebanyak 306 data, yang nantinya akan melewati beberapa preprocessing data sebelum dilakukan pengolahan seperti yang terlihat pada Gambar 1 dibawah ini.



Gambar 1. Alur Penelitian

## 2. Tahapan Penelitian

Gambar 1 merupakan alur penelitian secara umum yang telah dilakukan, dimana uraian dari alur tersebut dapat dilihat pada penjelasan berikut:

### a. Data *Preparation*

Tahap awal penelitian meliputi pengumpulan dan pemeriksaan data untuk memastikan kelengkapan dan konsistensi data. Data yang tidak lengkap atau tidak relevan diidentifikasi untuk diproses lebih lanjut.

### b. Data *Preprocessing*

Pada tahap ini dilakukan pembersihan dan pengolahan data, meliputi:

- 1) Pemilihan atribut yang relevan untuk proses clustering (Nama laptop ,Brand, Model, Harga).
- 2) Normalisasi data menggunakan metode *StandardScaler* agar setiap fitur memiliki skala yang seimbang.
- 3) Memisahkan data Nama Laptop yang isinya terdiri dari Nama Laptop, RAM dan SSD
- 4) Penanganan data kosong (missing value).
- 5) Transformasi data ke dalam bentuk numerik.

### c. Penentuan Jumlah *Cluster*

Penentuan jumlah *cluster* optimal dilakukan menggunakan *Elbow Method* dengan mengamati nilai *Within Cluster Sum of Squares* (WCSS). Titik siku (*elbow*) pada grafik digunakan sebagai dasar penentuan jumlah *cluster* terbaik.

### d. Proses *Clustering* dengan K-Means

Setelah jumlah *cluster* ditentukan, algoritma K-Means diterapkan untuk mengelompokkan data laptop berdasarkan kemiripan karakteristik. Setiap data akan ditempatkan pada *cluster* dengan jarak terdekat terhadap pusat *cluster* (*centroid*).

### e. Evaluasi dan Analisis Hasil

Hasil *clustering* dianalisis untuk mengetahui karakteristik masing-masing *cluster*, seperti kelompok laptop dengan harga rendah, menengah, dan tinggi. Analisis ini digunakan untuk memberikan interpretasi terhadap hasil pengelompokan yang diperoleh.

## 3. Tools dan Perangkat Lunak

Penelitian ini menggunakan bahasa pemrograman Python dengan bantuan pustaka:

- a. Pandas untuk pengolahan data,
- b. *Scikit-learn* untuk preprocessing dan algoritma K-Means,
- c. Matplotlib untuk visualisasi grafik *Elbow Method*.

## Hasil dan Pembahasan

Berdasarkan hasil penerapan algoritma K-Means *Clustering* terhadap data harga laptop, diperoleh pengelompokan data ke dalam tiga *cluster* utama. Penentuan jumlah *cluster* dilakukan menggunakan *Elbow Method*, yang menunjukkan titik siku (*elbow*) pada nilai  $k = 3$  seperti yang terlihat pada Gambar 2. Hal ini mengindikasikan bahwa tiga cluster merupakan jumlah yang paling optimal untuk merepresentasikan variasi data harga laptop.

### Perhitungan *Elbow Method*

*Elbow Method* digunakan untuk menentukan jumlah *cluster* ( $k$ ) yang optimal pada algoritma *K-Means* dengan mengamati nilai *Within-Cluster Sum of Squares* (WCSS) atau *inertia*. Meskipun demikian, algoritma *K-Means* memiliki beberapa keterbatasan, antara lain sensitif terhadap keberadaan *outlier* serta ketergantungan pada inisialisasi *centroid* awal yang dapat memengaruhi hasil pengelompokan. Oleh karena itu, tahap pembersihan dan normalisasi data dilakukan untuk meminimalkan pengaruh *outlier* dan meningkatkan stabilitas hasil *clustering*.

### 1. Rumus Jarak (*Euclidean Distance*)

K-Means menggunakan jarak *Euclidean* untuk menghitung kedekatan data dengan pusat *cluster*:

$$d(x_i, c_j) = \sqrt{\sum_{n=1}^m (x_{in} - c_{jn})^2} \quad \dots \dots \dots \quad (i)$$

## Keterangan :

$X_j = \text{data ke } - i$

$c_j$  = centroid cluster ke -j

$m = \text{jumlah atribut}$

## 2. Rumus WCSS (*Within-Cluster Sum of Squares*)

WCSS merupakan total jarak kuadrat antara setiap data dengan *centroid cluster*-nya:

## Keterangan :

$k$  = jumlah cluster

$C_i = cluster$  ke - j

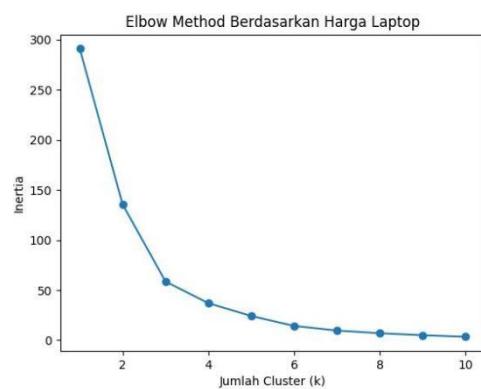
$X_i$  = data pada cluster ke - j

$c_i$  = centroid cluster ke -j

### 3. Prinsip *Elbow Method*

Nilai WCSS

- a) Tinggi saat jumlah cluster kecil
  - b) Menurun seiring bertambahnya jumlah cluster
  - c) Jumlah cluster optimal ditentukan pada titik "siku" (elbow), yaitu saat penurunan WCSS mulai melambat secara signifikan.



Gambar 2. *Elbow Method*

Seperti yang terlihat pada Gambar 2 grafik *Elbow Method* menunjukkan penurunan nilai WCSS yang signifikan terjadi saat nilai k bernilai 1 hingga 3. Setelah k=3, penurunan WCSS cenderung landai. Oleh karena itu, jumlah *cluster* optimal ditetapkan sebanyak 3 cluster

Ketiga cluster tersebut dapat diterjemahkan dengan kelompok sebagai berikut:

- ### 1. Cluster 0 (Laptop Harga Rendah)

*Cluster* ini berisi laptop dengan harga relatif rendah. Umumnya laptop pada *cluster* ini ditujukan untuk kebutuhan dasar seperti administrasi perkantoran, pembelajaran, dan penggunaan sehari-hari. Spesifikasi perangkat cenderung standar dengan kapasitas RAM dan penyimpanan yang lebih kecil.

- ## 2. Cluster 1 (Laptop Harga Menengah)

Cluster ini terdiri dari laptop dengan harga menengah yang memiliki spesifikasi lebih baik dibandingkan *cluster* harga rendah. Laptop pada kelompok ini umumnya cocok untuk

kebutuhan multitasking, pekerjaan profesional ringan, serta penggunaan aplikasi produktivitas yang lebih kompleks.

### 3. Cluster 2 (Laptop Harga Tinggi)

Cluster ini mencakup laptop dengan harga tinggi yang biasanya memiliki spesifikasi unggulan, seperti RAM besar, kapasitas SSD tinggi, serta prosesor dengan performa tinggi. Laptop dalam cluster ini ditujukan untuk kebutuhan profesional, desain grafis, pengolahan data, maupun aktivitas komputasi berat lainnya.

Setelah didapat nilai  $k=3$ , selanjutnya yang harus dilakukan adalah melakukan perhitungan K-Means berdasarkan nilai  $k$  tersebut.

#### Algoritma K-Means

Algoritma K-Means merupakan metode *clustering* yang bertujuan mengelompokkan data ke dalam  $k$  cluster berdasarkan tingkat kemiripan, yang diukur menggunakan jarak tertentu (umumnya jarak *Euclidean*).

##### 1. Penentuan Jumlah Cluster

##### 2. Inisialisasi Centroid

Centroid awal ( $c_j$ ) dipilih secara acak dari data:

$$c_j = (c_{j1}, c_{j2}, \dots, c_{jm})$$

Keterangan:

$$j = 1, 2, \dots, k$$

$m$  = jumlah atribut

##### 3. Menghitung Jarak Data ke Centroid

Jarak antara data  $x_i$  dan centroid  $c_j$  di hitung menggunakan Euclidean Distance dalam rumus persamaan ke- $i$ .

##### 1. Penentuan Cluster

Setiap data akan dimasukkan ke cluster dengan jarak terdekat

##### 5. Pembaruan Centroid

Centroid baru dihitung dari rata-rata seluruh data dalam cluster:

$$c_j^{new} = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i$$

Keterangan :

$$C_j = \text{cluster ke - } j$$

$$|C_j| = \text{jumlah data dalam cluster ke - } j$$

##### 6. Fungsi Objektif K-Means

K-Means bertujuan meminimalkan Within-Cluster Sum of Squares (WCSS) seperti persamaan- $ii$

##### 7. Kriteria Berhenti

Proses iterasi dihentikan apabila:

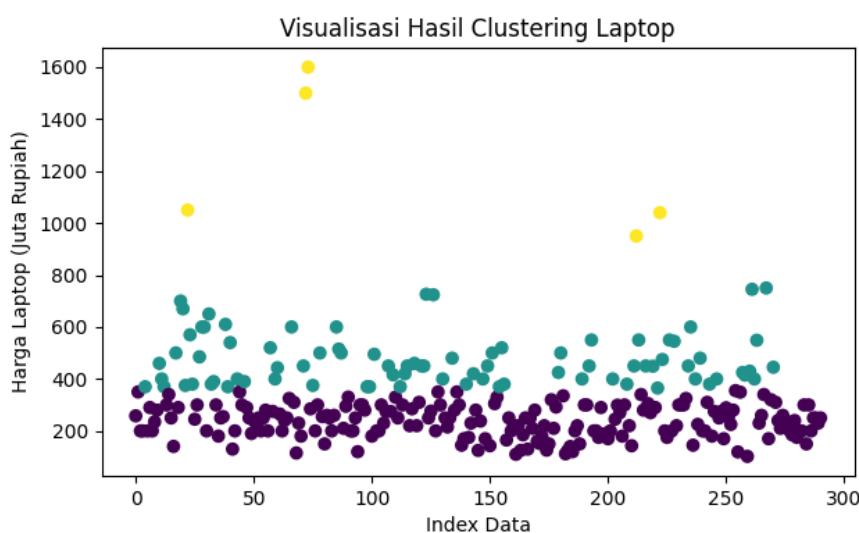
- Centroid tidak berubah, atau
- Tidak ada perpindahan data antar cluster, atau
- Iterasi maksimum tercapai

Algoritma K-Means bekerja dengan cara menghitung jarak antara data dan centroid menggunakan jarak Euclidean, kemudian mengelompokkan data ke dalam cluster terdekat. Proses ini diulang hingga tidak terjadi perubahan centroid atau jumlah iterasi maksimum tercapai.

Tabel 1. Hasil *Cluster*

Cluster	Jumlah	Rata- rata Harga(juta)	Harga(juta) Min	Harga Max (juta)	Kategori
0	203	235.83	101.999	355.0	Harga Rendah
1	83	474.36	365.0	750.0	Harga Menengah
2	5	1227.96	950.0	1599.9	Harga Tinggi

Hasil *clustering* pada tabel 1. menunjukkan adanya segmentasi harga yang jelas, yaitu kelompok laptop dengan harga rendah, menengah, dan tinggi. Setiap cluster memiliki karakteristik yang berbeda dan dapat merepresentasikan segmentasi pasar laptop secara tepat. Proses iterasi K-Means menghasilkan centroid yang stabil, menandakan bahwa proses pengelompokan telah mencapai kondisi konvergen. Penyebaran *cluster* tersebut dapat dilihat pada gambar dibawah ini.



Gambar 3. Visualisasi hasil *Cluster*

### Kesimpulan

Berdasarkan hasil penerapan algoritma K-Means pada data harga laptop, dapat disimpulkan bahwa metode ini mampu mengelompokkan data secara efektif berdasarkan tingkat kemiripan karakteristik numerik yang digunakan. Penentuan jumlah *cluster* optimal dilakukan menggunakan *Elbow Method*, yang menunjukkan titik siku pada nilai k=3, sehingga data dikelompokkan ke dalam tiga *cluster* utama.

Dengan demikian, algoritma K-Means terbukti efektif dalam membantu analisis pengelompokan data harga laptop. Penelitian ini juga menunjukkan bahwa metode *clustering* dapat menjadi pendekatan yang relevan untuk pengolahan dan analisis data produk elektronik

### Rekomendasi

Pengambilan data pada penelitian kali ini cukup sulit, karena terbatasnya data berbasis bahasa indonesia yang tersedia di platform online yang sudah terekstrak. Saran untuk penelitian selanjutnya adalah lakukan pengambilan data mandiri menggunakan tools yang memungkinkan, data review yang dibahas bisa dikembangkan ke produk lainnya dan menggunakan metode selain Kmeans.

## Daftar Pustaka

- Agung Barata, M., Ayuni, I. S., Kartini, A. Y., & Alawi, Z. (n.d.). (2024). Algoritma K-Means dalam Clustering Produk Skincare Untuk Menentukan Strategi Pemasaran. JIP (Jurnal Informatika Polinema) . ISSN: 2614-6371 E-ISSN: 2407-070X
- Ariati, I., Nugraha Norsa, R., Akhsan, L., & Heikal, J. (n.d.). (2023) Segmentasi Pelanggan Menggunakan K-Means Clustering Studi Kasus Pelanggan UHT Milk Greenfield. Jurnal Ilmiah Indonesia, 2023(7), 629–643. <https://doi.org/10.36418/cerdika.xxx>
- Bao Cho (2025) K-means clustering algorithm: a brief review. (2021). Academic Journal of Computing & Information Science, 4(5). <https://doi.org/10.25236/ajcis.2021.040506>
- BrightLocal. (2020). Local Consumer Review Survey 2020. Available Online at:<https://www.brightlocal.com/research/local-consumer-review-survey-2020>
- Darmawan, S. A. D., & Karmilasari. (2024). Penerapan Metode K-Means Clustering dan Simple Moving Average untuk Memprediksi Jenis Penyakit di Provinsi Jawa Timur. Jurnal Teknologi Informasi Dan Ilmu Komputer, 11(4), 877–886. <https://doi.org/10.25126/jtiik.1148703>
- Erwin. (2009). Analisis Market Basket dengan Algoritma Apriori dan FP-Growth. Jurnal Generic , 4.Association Rule Metode Apriori. Informasi dan Teknologi Ilmiah(INTI) , III.
- Faisal, M., Sri Utami, W., & Pratiwi, N. (2024). Bulletin of Information Technology (BIT) Pemanfaatan Data Mining Menggunakan Metode K-Means Untuk Analisa Komoditas Telur Ayam. 5(4), 247–254. <https://doi.org/10.47065/bit.v5i2.1670>
- Fatkhudin, A., Khambali, A., Artanto, F. A., Putra Zade, N. A., & Muhammadiyah Pekajangan Pekalongan, U. (2023). Implementasi Algoritma Clustering K-Means Dalam Pengelompokan Mahasiswa Studi Kasus (Prodi Manajemen Informatika). Jurnal Minfo Polgan, 12(2). <https://doi.org/10.33395/jmp.v12i2.12494>
- Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. Jurnal Nasional Teknologi Dan Sistem Informasi, 5(1), 17–24. <https://doi.org/10.25077/teknosi.v5i1.2019.17-24>
- Habibah Annisa Hasna, & Rachbini Widarto. (2025). Segmentasi Pasar Menggunakan K-Means Clustering Berdasarkan Kualitas, Harga, Fasilitas, Kenyamanan, dan Kelengkapan Produk. Musytari : Jurnal Manajemen, Akuntansi, dan Ekonomi. ISSN : 3025-9495. DOI : 10.8734/mnmae.v1i2.359
- Haque, T. U., Saber, N. N., & Shah, F. M. (2018). Sentiment analysis on large scale Amazon product reviews. 2018 IEEE International Conference on Innovative Research and Development (ICIRD). doi:10.1109/icird.2018.8376299
- Iriane, R. (2023). KLIK: Kajian Ilmiah Informatika dan Komputer Penerapan Data Mining Untuk Prediksi Penjualan Produk Pangan Hewan Menggunakan Metode K-Nearest Neighbor. Media Online, 3(5), 509–515. <https://djournals.com/klik>
- Kurniawan, B., Effendi, S., & Salim Sitompul, O. (2012). Klasifikasi Konten Berita dengan Metode Text Mining. Jurnal Dunia Teknologi Informasi , I, 14-19.
- Litbang. (2019). APJII: Jumlah Pengguna Internet di Indonesia Tembus 171 Juta Jiwa. Litbang Kemendagri. <https://litbang.kemendagri.go.id/website/apjii-jumlah-pengguna-internet-diindonesia-tembus-171-juta-jiwa/>
- Liu, Bing. (2020). Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. Cambridge University Press
- Milatina, Syukur, A., & Supriyanto, C. (2012). Pengaruh Text Preprocessing pada Clustering Dokumen Teks Berbahasa Indonesia. Jurnal Teknologi Informasi
- Oti, E. U., Olusola, M. O., Eze, F. C., & Enogwe, S. U. (2021). Comprehensive Review of K-Means Clustering Algorithms. International Journal of Advances in Scientific

Research and Engineering, 07(08), 64–69.  
<https://doi.org/10.31695/ijasre.2021.34050>

- Priyatman, H., Sajid, F., & Haldivany, D. (2019). JEPIN (Jurnal Edukasi dan Penelitian Informatika) Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa. ISSN(e): 2548-9364 / ISSN(p) : 2460-0741
- Radovanovic, M., & Ivanovic, M. (2008). Text Mining: Approaches and Application.
- Ririanti. (2014). Implementasi Algoritma FP-Growth pada Aplikasi Prediksi Persediaan Sepeda Motor(Studi Kasus PT.Pilar Deli Labumas). Pelita Informatika Budi Darma (VI).
- Siringoringo, R. (n.d.).(2019).Text Mining dan Klasterisasi Sentimen Pada Ulasan Produk Toko Online. e-ISSN : 2621-234X
- Sulistiyawati, A., & Supriyanto, E. (n.d.). Implementasi Algoritma K-means Clustering dalam Penetuan Siswa Kelas Unggulan. 15(2). 25-36. P-ISSN: 1412-9663, E-ISSN : 2656-3525
- Vanaja, S., Belwal, M. (2018). Aspect-Level Sentiment Analysis on E-Commerce Data. Proceedings of the International Conference on Inventive Research in Computing Applications (ICIRCA 2018). doi:10.1109/ICIRCA.2018.8597286