Implementasi Algortima K-Means Data Penjualan UMKM Hijab Leshamysha

Raka Panangkaran Ferdinandus, Hendro Nugroho*[1], Muhammad Rafel Daffa Eka Firmansyah[2], Gusti Eka Yuliastuti[3]
[1]Prodi Teknik Informatika,Teknik Elektro dan Teknologi Informasi, Institut Teknologi Adhi Tama Surabaya
[2]Prodi Teknik Informatika,Teknik Elektro dan Teknologi Informasi, Institut Teknologi Adhi Tama Surabaya
[3]Prodi Teknik Informatika,Teknik Elektro dan Teknologi Informasi, Institut Teknologi Adhi Tama Surabaya
*Email: dosh3ndro@itats.ac.id

Abstract. *The K-Means algorithm is used for data clustering based on similarity. The K-Means method is used for this study to identify sales products at Leshamysha Hijab UMKM. The dataset used comes from sales data for 2023, which includes 38 product items. In its implementation, two approaches were used: standard K-Means and K-Means with centroid using the average value (mean) and variance. The clustering process was tested and evaluated using two metrics, namely the Sum of Squared Error (SSE) and the Davies-Bouldin Index (DBI). The evaluation results show that the K-Means method with the initial centroid provides clustering results with an SSE value of 0.9215 and a DBI of 1.6697, indicating a fairly good level of data closeness. K-Means with the Centroid approach produces an SSE value of 41.3127 and a DBI of 2.3607, indicating a fairly high level of uniformity.*
Keywords: *K-Means, SSE, DBI, UMKM*

Abstrak. *Algoritma K-Means digunakan untuk pengelompokan data berdasarkan kemiripan. Metode K-Means digunakan untuk penelitian ini untuk mengidentifikasi produk penjualan di UMKM Hijab Leshamysha. Dataset yang digunakan berasal dari data penjualan selama tahun 2023, yang mencakup 38 item produk. Dalam implementasinya, dilakukan dua pendekatan: K-Means standar dan K-Means dengan centroid menggunakan nilai rata-rata (mean) dan varian. Proses clustering diuji dan dievaluasi menggunakan dua metrik, yaitu Sum of Squared Error (SSE) dan Davies-Bouldin Index (DBI). Hasil evaluasi menunjukkan bahwa metode K-Means dengan centroid awal memberikan hasil clustering nilai SSE 0,9215 dan DBI 1,6697 yang menunjukkan tingkat kedekatan data cukup baik. K-Means dengan pendekatan Centroid menghasilkan nilai SSE 41,3127 dan DBI 2,3607 yang menujukkan tingkat keseragaman cukup tinggi*
Kata Kunci: *K-Means, SSE, DBI, UMKM*

1. Pendahuluan

Pemanfaatan teknologi informasi dalam pengembangan Usaha Mikro, Kecil, dan Menengah atau yang disingkat UMKM banyak digunakan untuk membantu pemasaran dan meningkatkan penjualan(Nugroho et al. 2023). Dalam pemanfaatan teknologi informasi, peneliti mengembangkan implementasi informasi data penjualan pada produk UMKM Hijab Leshamysha. Tujuan penelitian ini adalah memahami pola penjualan tiap produk secara kuantitatif. Pola pengelompokan atau clustering penjualan perlu dilakukan untuk mengetahui produk mana saja yang memiliki tingkat penjualan tinggi, sedang dan rendah.

Pada penelitian sebelumnya untuk memahami clustering menggunakan metode K-Means pada data konsumsi listrik(Aisyi Maulidhia et al. 2025). Penggunaan algoritma K-Means dengan cara membagi data ke dalam sejumlah klaster berdasarkan kedekatan terhadap titik pusat (*centroid*). Adapun penelitian lain Algoritma K-Mean digunakan untuk clustering strategi promosi berdasarkan data

murid(Lin, Iswari, and Fredlina 2024). Metode K-Means clustering digunakan untuk membantu system data mining big data yang menghasilkan efisiensi analisis clustering dan frekuensi item set(Liu 2025).

Berdasarkan penelitian sebelumnya, maka tujuan penelitian ini untuk mengelompokan item produk berdasarkan harga dan jumlah data penjualan UMKM Hijab Leshamysha, dan untuk mengetahui item yang kurang laku dengan kombinasi dengan item lainnya.

2. Tinjauan Pustaka

Algoritma clustering adalah teknik *unsupervised learning* yang bertujuan untuk mengelompokan data berdasarkan kemiripan antar data. Algoritma K-Means merupakan salah satu metode clustering yang membagi data menjadi beberapa kelompok sehingga data yang karakteristik sama berada pada cluster yang sama(Riches et al. 2025), sebaliknya yang berbeda berada pada cluster yang berbeda(Azzahra and Amru Yasir 2024). Untuk mendapatkan nilai clustering dengan menggunakan algoritma K-Means dengan menggunakan algoritma sebagai berikut(Azzahra and Amru Yasir 2024).

1. Menentukan K sebagai jumlah cluster yang digunakan
2. Menghasilkan nilai acak untuk pusat cluster awal sebanyak K(Rahman et al. 2025)
3. Menggunakan rumus jarak terdekat menggunakan *Euclidean Distance*(Nahoujy 2025)

   $$d(xi - \mu j) = \sqrt{\sum (xi - \mu j)^2} \quad (1)$$

   Dimana Xi adalah data kriteria dan µj adalah centroid pada cluster ke-j
4. Mengklasifikasi setiap data terdekat dengan centroid terkecil

   $$C_k = \frac{1}{n_k} \sum d_i \quad (2)$$

   Dimana n_k adalah jumlah data dalam cluster k dan d_i adalah jumlah dari nilai jarak yang masuk dalam masing-masing cluster
5. Melakukan perulangan dari langkah 2 hingga 5 sampai anggota tiap cluster tidak ada yang berubah
6. Jika langkah terakhir telah terpenuhi, maka nilai pusat cluster (µj) pada iterasi teakhir digunakan sebagai parameter klasifikasi data.

Dari hasil Algoritma K-Means dilakukan evaluasi menggunakan metode Davies Bouldin Index (DBI) dan Sum of Sequenced Error (SSE). Tujuan pengujian DBI untuk memaksimalkan jarak antara satu cluster dengan cluster lainnya, serta mencari nilai untuk meminimalkan jarak antar data dokumen yang terdapat di dalam cluster yang sama(Liu 2025)(Aslam 2025). Untuk mencari nilai DBI didasarkan pencarian Sum of Sequence Within Cluster (SSW) yang merupakan persamaan yang didapat untuk mengetahui matrik kohesi dalam cluster ke-1 yang dirumuskan pada persamaan 3

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_{i,}, c_i) \quad (3)$$

Dalam persamaan 3 dimana m_i merupakan jumlah data dalam cluster ke-1, c_i adalah centroid cluster ke-1 dan d merupakan jarak pada setiap data ke centroid menggunakan jarak Euclidean. Setelah itu mencari nilai Sum of Square between Cluster (SSB) dengan menggunakan persamaan 4

$$SSB_{i,j} = d(C_i, C_j) \quad (4)$$

Setelah SSB sudah diperoleh, maka kemudian akan dilakukan pengukuran rasio untuk mengetahui sebuah nilai perbandingan antara cluster ke-i dan ke-j. persamaan yang digunakan untuk menghitung rasio dapat dilihat pada persamaan 5.

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (5)$$

Nilai rasio yang diperoleh tersebut akan digunakan untuk menghitung nilai DBI dengan persamaan 6.

$$DBI = \frac{1}{k} \sum_{i=1}^{k} max(R_{ij}) \quad (6)$$

Pengujian menggunakan SSE merupakan hasil penjumlahan dari seluruh jarak masing-masing data dengan titik pusat clusternya. Semakin kecil nilai SSE yang didapat, semakin seragam data yang

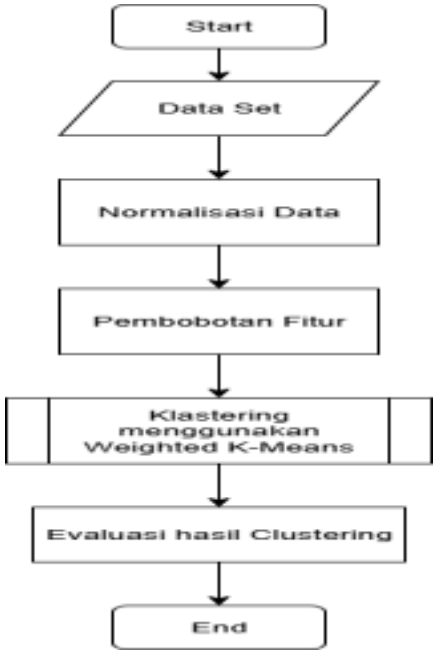
ada di dalam masing-masing cluster, dan semakin baik cluster yang dihasilkan. Untuk mendapatkan nilai SSE menggunakan persamaan 7.

$$SSE = \sum_{i=1}^{n}(x_{ij} - x_i)^2 \quad (7)$$

3. Metode Penelitian

Rancangan sistem secara keseluruhan disajikan dalam Gambar 1, yang menggambarkan alur proses secara terstruktur dan menyeluruh. *FlowChart* Alur Rangkaian dapat dilihat pada gambar 1. Flowchart menggambarkan alur proses penerapan metode Weighted K-Means dalam melakukan clustering data penjualan. Proses diawali dari tahap "Start" yang menandakan dimulainya sistem, kemudian dilanjutkan dengan pengambilan dataset sebagai bahan utama untuk analisis. Dataset yang digunakan berisi informasi seperti harga produk dan volume penjualan. Setelah itu, dilakukan proses normalisasi data menggunakan metode Min-Max untuk menyamakan skala nilai antar atribut, sehingga tidak ada atribut yang mendominasi dalam proses pengelompokan.

Tahap berikutnya adalah pembobotan fitur, yaitu proses pemberian bobot pada setiap atribut menggunakan metode entropy. Pembobotan ini bertujuan untuk memberikan nilai yang proporsional sesuai dengan tingkat variasi dan pengaruh setiap fitur terhadap hasil clustering. Selanjutnya, proses clustering dilakukan menggunakan algoritma Weighted K-Means, di mana bobot fitur yang telah dihitung sebelumnya digunakan untuk meningkatkan akurasi pengelompokan. Setelah proses clustering selesai, dilakukan evaluasi terhadap hasil clustering menggunakan metrik seperti Sum of Squared Error (SSE) dan Davies-Bouldin Index (DBI) untuk mengukur kualitas dan validitas pengelompokan yang telah dilakukan. Proses diakhiri dengan "End", menandakan bahwa sistem atau metode telah selesai dijalankan dan hasil clustering siap untuk dianalisis atau digunakan dalam pengambilan keputusan.



Gambar 1. FlowChart penelitian

Proses diawali dengan pemilihan metode K-Means yang dimodifikasi, di mana penentuan titik pusat (centroid) tidak dilakukan secara acak, melainkan dihitung berdasarkan nilai rata-rata (mean) dan variansi dari data. Setelah itu, ditentukan jumlah cluster yang akan digunakan dalam proses pengelompokan, misalnya tiga cluster untuk kategori produk laku, sedang, dan kurang laku.

Selanjutnya, dilakukan perhitungan jarak antara setiap titik data terhadap centroid menggunakan metode pengukuran seperti Euclidean Distance.

Setelah jarak dihitung, setiap data akan dikelompokkan ke dalam cluster yang memiliki jarak terdekat dari centroid. Kemudian, dilakukan pembaruan posisi centroid dengan menghitung rata-rata dari seluruh anggota data dalam masing-masing cluster. Setelah centroid diperbarui, perhitungan jarak antara titik data dan centroid kembali dilakukan untuk mengevaluasi apakah ada perubahan dalam keanggotaan cluster. Jika terjadi perubahan, proses perhitungan dan pengelompokan akan diulang hingga tidak ada lagi perubahan dalam pembentukan cluster. Jika semua data sudah stabil dan tidak ada perpindahan cluster, maka proses berhenti dan hasil clustering akhir ditetapkan. Proses kemudian diakhiri dengan pengembalian hasil pengelompokan untuk digunakan pada tahap analisis atau implementasi selanjutnya.

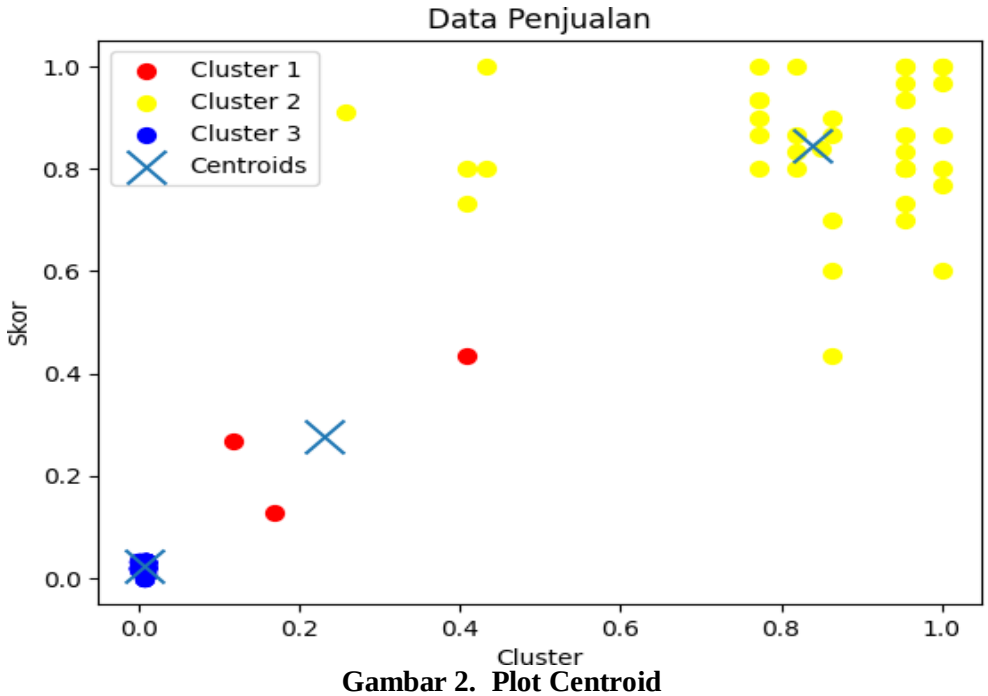
4. Hasil dan Pembahasan

Dataset yang digunakan dalam penelitian ini berasal dari data penjualan produk Leshamysha, yang telah dikumpulkan dan disusun secara sistematis. Data tersebut mencakup informasi mengenai harga per item dari masing-masing produk, serta volume penjualan. Dataset tersebut berisi informasi mengenai 38 item produk Jilbab, lengkap dengan klasifikasi penjualan untuk setiap item dalam kurun waktu tersebut. Setiap item dalam dataset memiliki sejumlah atribut yang dijadikan sebagai kriteria analisis, seperti harga satuan dan volume penjualan bulanan. Rincian atribut serta nilai-nilai yang digunakan dalam proses analisis ditampilkan pada Tabel 1.

Tabel 1. Dataset Penjualan UMKM Leshamysha

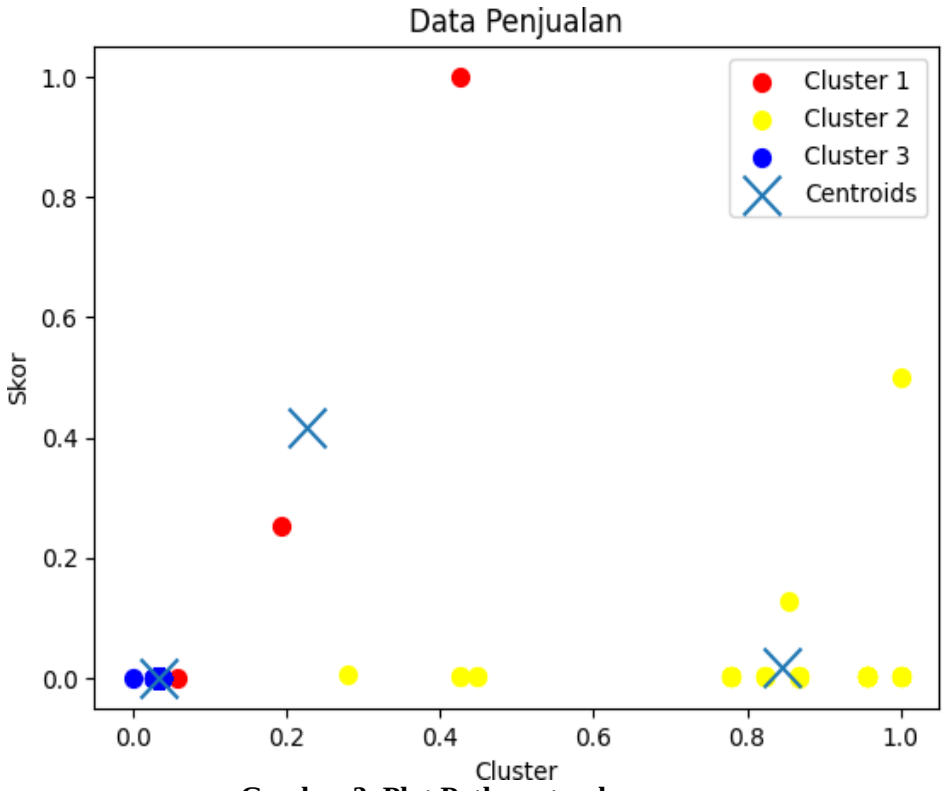
No	Item	Harga Satuan	Januari	Februari	Maret	April	Mei	Juni
1	Ceruty-Black	Rp 45.000	24	24	24	14	18	30
2	Ceruty-Navy	Rp 45.000	22	22	25	16	15	24
3	Diamond Crepe-Black	Rp 47.500	24	30	24	20	17	25
4	Diamond Crepe-Black Blue Gray	Rp 47.500	30	28	30	20	11	24
5	S4 Cotton Import Lc- Army Green	Rp 90.000	30	30	25	20	18	30
6	S4 Cotton Import Lc- Asygreen	Rp 90.000	24	24	29	16	15	25
7	S4 Ultraglossy Lc-Silver Sky	Rp 90.000	25	21	24	14	28	29
8	S4 Ultraglossy Lc-Taupe	Rp 90.000	26	29	30	16	17	24

Hasil K-Means dapat disajikan dalam bentuk grafik pada Gambar 2 Kode ini menampilkan visualisasi hasil clustering dari algoritma K-Means, Terdiri dari 3 cluster berwarna berbeda (merah, kuning, biru) Ditambahkan centroid (pusat cluster) berwarna default dengan marker “x”.



Gambar 2. Plot Centroid

Dalam uji coba metode K-Means ini, (*centroid*) dilakukan dengan menggunakan pendekatan matematis berdasarkan rata-rata (*mean*) dan varian, sesuai dengan yang dijelaskan pada persamaan 6 dan 7. Proses ini menghasilkan nilai *Sum of Squared Error* (SSE) sebesar 41.31276013, yang menunjukkan tingkat kesalahan kuadrat antar data terhadap pusat clusternya. Sementara itu, Davies-Bouldin Index (DBI) diperoleh dengan nilai 2.360747114, yang mengindikasikan kualitas pemisahan antar kluster. Hasil pembentukan *cluster* secara lengkap dapat dilihat pada Gambar 3.



Gambar 3. Plot Python standar

Program tersebut digunakan untuk menampilkan visualisasi hasil clustering K-Means dalam bentuk scatter plot dua dimensi. Data yang digunakan telah dikelompokkan ke dalam tiga cluster, yang masing-masing ditampilkan dengan warna berbeda: merah untuk Cluster 1, kuning untuk Cluster 2, dan biru untuk Cluster 3. Baris pertama hingga ketiga menggunakan plt.scatter untuk memplot data dari masing-masing cluster berdasarkan hasil label y_kmeans, dengan koordinat sumbu-x dari kolom ke-0 dan sumbu-y dari kolom ke-1 pada array X

5. Kesimpulan

K-Means Standar menghasilkan nilai SSE = 0.9215 dan DBI = 1.6697, yang menunjukkan tingkat kedekatan data dalam kluster cukup baik dan pemisahan antar cluster cukup jelas., K-Means dengan Penentuan Centroid menghasilkan nilai SSE = 41.3127 dan DBI = 2.3607, yang menunjukkan tingkat keseragaman data dalam kluster masih cukup tinggi, serta pemisahan antar cluster kurang optimal.. K-Means Standar (dengan inisialisasi acak) menghasilkan nilai SSE dan DBI lebih rendah, artinya menghasilkan klaster yang lebih padat dan terpisah jelas. Sebaliknya, metode K-Means Centroid (menggunakan mean dan varians) menghasilkan klaster dengan kualitas lebih tinggi untuk dataset ini.

Referensi

Aisyi Maulidhia, Alief Nur, Indri Ika Widyastuti, Friska Intan Sukarno, Rahmat Basya Sharys Tsany, and Thomas Brian. 2025. “Implementasi Perbandingan Algoritma K-Means Dan DB-Scan Pada Beban Listrik Rumah Tangga.” *INTEGER: Journal of Information Technology* 10(1): 85–94. doi:10.31284/j.integer.2024.v10i1.7479.

Aslam, Bilal. 2025. “Identifying Optimistic Stocks with K-Means Clustering Algorithm.” International Review of Economics and Finance 104(February). doi:10.1016/j.iref.2025.104579.

Azzahra, Lutfhia, and Amru Yasir. 2024. “Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food.” Jurnal Ilmu Komputer dan Sistem Informasi 3(1): 1–10. doi:10.70340/jirsi.v3i1.88.

Lin, Lin Andriani, Ni Made Satvika Iswari, and Ketut Queena Fredlina. 2024. “Implementasi K-Means Clustering Untuk Mengelompokkan Data Murid Sebagai Acuan Dalam Menentukan Strategi Promosi (Studi Kasus: Elite Kid Courses).” Jurnal Teknologi Informasi Dan Komunikasi 15(2): 398–412. doi:10.51903/jtikp.v15i2.888.

Liu, Lihua. 2025. “Application of K-Means Supported by Clustered Systems in Big Data Association Rule Mining.” Systems and Soft Computing 7(February). doi:10.1016/j.sasc.2025.200211.

Nahoujy, Mahdi Rahimi. 2025. “Applying a K-Means Model to TSD Data to Find Categories for the Structural Assessment of Flexible Pavements.” Transportation Engineering 20(February): 100342. doi:10.1016/j.treng.2025.100342.

Nugroho, Hendro, Suparjo Suparjo, Chrisna A.D., and Linda A. Lestari. 2023. “Implementasi Website Profil UMKM NTZ Leather Sebagai Penunjang Kegiatan Pemasaran Atau Promosi Produk Jaket Kulit Di Kecamatan Candi, Sidoarjo.” JPP IPTEK (Jurnal Pengabdian dan Penerapan IPTEK) 7(1): 19–24. doi:10.31284/j.jpp-iptek.2023.v7i1.3747.

Rahman, Syed Ahmad Fadhli Syed Abdul, Khairul Nizam Abdul Maulud, Uznir Ujang, Wan Shafrina Wan Mohd Jaafar, and Biswajeet Pradhan. 2025. “A Framework for Extracting and Characterizing Landform Structures Under Buildings Using Elevation Thresholds and K-Means Algorithm.” KSCE Journal of Civil Engineering 29(12): 100324. doi:10.1016/j.kscej.2025.100324.

Riches, Naomi O, Ramkiran Gouripeddi, Robert M Silver, and Julio C Facelli. 2025. “Air Pollution Mixtures and Stillbirth : K-Means Cluster Analysis.” Hygiene and Environmental Health Advances 16(April): 100153. doi:10.1016/j.heha.2025.10015