



Analisis Kluster Penyebaran Berat Produk Mesin Sachet Menggunakan Metode Algoritma K-Means

Ermanto Ermanto*, Nurhadi Surojudin

Fakultas Teknik, Program Studi Informatika, Universitas Pelita Bangsa, Bekasi, Indonesia

Email: ^{1,*}ermanto@pelitabangsa.ac.id, ²nurhadi@pelitabangsa.ac.id

Email Penulis Korespondensi: ermanto@pelitabangsa.ac.id

Abstrak—Penelitian ini bertujuan untuk menganalisis pola penyebaran berat produk mesin sachet menggunakan metode algoritma K-Means sebagai salah satu teknik klusterisasi data. Data yang digunakan merupakan data primer hasil pencatatan berat produk dari proses produksi, terdiri dari 940 entri dengan sepuluh sampel pengukuran pada setiap siklus produksi. Data tersebut dibersihkan melalui tahap data cleaning untuk memastikan tidak terdapat nilai kosong, duplikasi, maupun outlier, kemudian dipilih atribut relevan berupa sampel berat produk, serta ditransformasikan menggunakan normalisasi Min-Max agar seluruh variabel berada pada skala 0–1. Proses klusterisasi dilakukan secara iteratif dengan memperbarui titik pusat (centroid) hingga mencapai kondisi konvergen. Hasil evaluasi menunjukkan bahwa jumlah kluster optimal adalah tiga ($k=3$) dengan nilai Silhouette Coefficient sebesar 0,55, menandakan keseimbangan yang baik antara homogenitas internal dan perbedaan antar kluster. Kluster 1 merepresentasikan produk dengan berat rendah (8,00–8,18 gram), Kluster 2 menunjukkan produk dengan berat sedang (8,19–8,34 gram), dan Kluster 3 mencerminkan produk dengan berat tinggi (8,36–8,98 gram). Secara umum, data berat produk cenderung stabil dengan variasi rendah, meskipun terdapat beberapa anomali pada mesin tertentu. Temuan ini membuktikan bahwa algoritma K-Means mampu mengelompokkan data berat produk secara efektif, sehingga dapat digunakan perusahaan untuk mendukung pengendalian kualitas (quality control), mengidentifikasi variasi produk, serta meminimalkan risiko penyimpangan dari standar produksi.

Kata kunci: Klusterisasi; K-Means; Berat Produk; Sachet; Quality Control

Abstract—This study aims to analyze the distribution patterns of sachet machine product weights using the K-Means algorithm as a clustering technique. The dataset consists of 940 entries of primary production records, each containing ten weight measurement samples per production cycle. The data underwent a cleaning process to ensure the absence of missing values, duplicates, and outliers, followed by the selection of relevant attributes (product weight samples) and transformation using Min-Max normalization to scale all variables within the 0–1 range. The clustering process was performed iteratively by updating the centroids until convergence was achieved. The evaluation results indicate that the optimal number of clusters is three ($k=3$) with a Silhouette Coefficient of 0.55, reflecting a good balance between intra-cluster homogeneity and inter-cluster separation. Cluster 1 represents products with relatively low weights (8.00–8.18 grams), Cluster 2 includes medium-weight products (8.19–8.34 grams), and Cluster 3 consists of high-weight products (8.36–8.98 grams). Overall, the product weights tend to be stable with low variation, although some anomalies were observed in certain machines. These findings demonstrate that the K-Means algorithm can effectively classify product weight data, providing valuable insights for quality control, product variation identification, and minimizing risks of deviation from production standards.

Keywords: Clustering; K-Means; Product Weight; Sachet; Quality Control

1. PENDAHULUAN

Konsistensi berat produk pada industri kemasan sachet merupakan salah satu aspek penting dalam menjaga mutu dan kepuasan pelanggan[1]. Variasi berat yang terlalu tinggi atau rendah dibandingkan standar yang ditetapkan tidak hanya berpotensi menimbulkan keluhan dari konsumen[2], tetapi juga dapat berdampak pada efisiensi penggunaan bahan baku, biaya produksi, dan kepatuhan terhadap regulasi mutu yang berlaku[3]. Proses produksi mesin sachet[4], meskipun telah didukung oleh sistem otomatis, tetap memiliki potensi menghasilkan variasi berat akibat faktor mekanis, kualitas bahan baku, atau pengaturan mesin yang tidak optimal. Oleh karena itu, diperlukan metode analisis yang mampu memetakan distribusi berat produk secara akurat untuk mendukung deteksi dini terhadap potensi penyimpangan kualitas. Dalam beberapa tahun terakhir, teknik clustering telah menjadi salah satu pendekatan populer untuk menganalisis pola dan mengelompokkan data berdasarkan kemiripan karakteristik[5]. Salah satu metode clustering yang banyak digunakan adalah algoritma K-Means[6], yang dikenal sederhana, efisien, dan efektif dalam mengelompokkan data berukuran besar. Penelitian sebelumnya di bidang manufaktur telah menunjukkan bahwa penerapan K-Means dapat membantu mengidentifikasi kelompok produk berdasarkan variasi ukuran, berat, atau kualitas, sehingga memudahkan proses pengendalian mutu. Misalnya, studi oleh Gustipartsani K et al. (2024) menerapkan K-Means untuk memantau stabilitas berat kemasan pada industri makanan[7], sementara Kusuma Dewi W et al. (2023) berhasil memanfaatkan algoritma ini untuk mengoptimalkan kualitas produk pada industri farmasi[8]. Hasil dari berbagai penelitian tersebut menunjukkan bahwa K-Means mampu memberikan representasi yang jelas mengenai distribusi kualitas produk dan mendukung pengambilan keputusan berbasis data. Berdasarkan landasan tersebut, penelitian ini bertujuan untuk menganalisis penyebaran berat produk mesin sachet menggunakan metode K-Means Clustering dengan mengevaluasi kualitas pengelompokan menggunakan Davies-Bouldin Index (DBI)[9]. Dataset yang digunakan terdiri dari 938 entri berat produk yang diukur langsung dari hasil produksi. Penelitian ini juga membandingkan kualitas klusterisasi dengan jumlah kluster yang berbeda, yaitu 2, 3, dan 4 kluster, untuk menentukan konfigurasi yang paling optimal. Hasil yang diperoleh diharapkan dapat memberikan kontribusi pada strategi pengendalian mutu di industri kemasan sachet dengan pendekatan analisis data yang lebih sistematis dan berbasis bukti. Data mentah yang dihasilkan dari proses produksi mesin sachet pada dasarnya hanya berupa angka-angka hasil pengukuran berat produk yang terakumulasi dalam jumlah besar.



Meskipun data tersebut memuat informasi faktual, nilainya masih terbatas apabila tidak diolah dan dianalisis lebih lanjut[10]. Tanpa proses analisis, data mentah sulit memberikan gambaran yang jelas mengenai pola distribusi berat produk, tren penyimpangan, atau potensi permasalahan dalam pengendalian mutu. Dalam konteks industri, ketidakmampuan memanfaatkan data secara optimal dapat mengakibatkan peluang perbaikan terlewat, masalah kualitas tidak terdeteksi, dan keputusan produksi yang kurang tepat sasaran. Oleh karena itu, penerapan metode analisis data seperti K-Means Clustering menjadi sangat penting[11], karena mampu mengubah data mentah menjadi informasi yang terstruktur, mudah dipahami, dan siap digunakan untuk mendukung pengambilan keputusan yang berbasis bukti (data-driven decision making). Dengan mengelompokkan data berat produk ke dalam kluster yang jelas, perusahaan dapat mengidentifikasi kelompok produk yang sesuai standar maupun yang berpotensi cacat, sehingga strategi perbaikan dapat dilakukan secara cepat dan tepat. Dalam era industri modern, data menjadi aset penting yang dapat dimanfaatkan untuk meningkatkan efisiensi, kualitas[12], dan daya saing perusahaan. Namun, volume data yang besar dan kompleks memerlukan teknik pengolahan yang tepat agar dapat menghasilkan informasi yang bermakna. Data mining hadir sebagai solusi untuk mengekstraksi pengetahuan tersembunyi dari data mentah melalui berbagai metode analisis, salah satunya clustering. Dalam industri kemasan sachet, data mining memegang peran krusial untuk mengidentifikasi pola distribusi berat produk, mendeteksi penyimpangan, dan memberikan dasar bagi pengambilan keputusan yang berbasis data (data-driven decision making). Dengan memanfaatkan data mining, perusahaan tidak hanya mampu memahami kondisi kualitas produknya secara lebih akurat, tetapi juga dapat melakukan tindakan korektif secara cepat, sehingga mengurangi risiko kerugian akibat produk yang tidak sesuai standar.

Pada proses produksi mesin sachet, variasi berat produk sering kali tidak dapat dihindari, baik karena faktor mekanis, ketidakseragaman bahan baku, maupun pengaturan mesin yang kurang tepat. Meskipun data berat produk telah dicatat secara rutin, informasi tersebut belum memberikan manfaat optimal karena tidak diolah untuk mengungkapkan pola dan kelompok distribusi berat secara jelas. Akibatnya, potensi penyimpangan kualitas sulit dideteksi secara dini, sehingga produk di luar spesifikasi standar dapat lolos ke pasar. Permasalahan spesifik yang menjadi fokus penelitian ini adalah belum adanya analisis sistematis untuk mengelompokkan data berat produk guna mengetahui kategori bobot yang sesuai standar dan yang menyimpang. Dengan menerapkan algoritma K-Means Clustering dan mengevaluasi hasilnya menggunakan Davies-Bouldin Index[13], penelitian ini berupaya menemukan konfigurasi jumlah kluster yang optimal, sehingga distribusi berat produk dapat dipetakan secara akurat dan digunakan sebagai dasar pengendalian mutu yang lebih efektif. Association rule mining merupakan salah satu teknik spesifik dalam data mining yang digunakan untuk menemukan hubungan atau keterkaitan antar item di dalam suatu dataset. Teknik ini banyak dimanfaatkan dalam analisis keranjang belanja (market basket analysis) untuk mengidentifikasi pola pembelian yang sering terjadi secara bersamaan. Aturan asosiasi biasanya dinyatakan dalam bentuk "jika-maka" (contohnya jika membeli produk A, maka kemungkinan besar akan membeli produk B), dengan indikator utama berupa nilai support, confidence, dan lift untuk mengukur kekuatan hubungan tersebut. Penerapan association rule mining[14] memungkinkan perusahaan memahami perilaku konsumen secara lebih mendalam, sehingga strategi pemasaran, pengelolaan stok, dan penataan produk dapat dilakukan secara lebih efisien. Dalam konteks penelitian "Analisis Kluster Penyebaran Berat Produk Mesin Sachet Menggunakan Metode Algoritma K-Means", data berat produk sachet yang telah dikelompokkan melalui K-Means Clustering dapat menjadi landasan untuk analisis lanjutan menggunakan association rule mining. Permasalahan yang dihadapi adalah bagaimana menemukan pola pembelian atau permintaan produk sachet berdasarkan karakteristik berat yang berbeda[15]. Tanpa analisis yang tepat, perusahaan sulit mengetahui kombinasi atau varian produk yang cenderung dipesan secara bersamaan oleh pelanggan. Hal ini dapat mengakibatkan ketidakefisienan dalam perencanaan produksi, pengaturan distribusi, dan manajemen persediaan. Dengan menerapkan association rule mining, hubungan antar kelompok produk sachet yang terbentuk dari hasil klusterisasi dapat diidentifikasi, sehingga pola pembelian yang paling sering terjadi dapat diketahui. Informasi ini selanjutnya dapat digunakan untuk mengoptimalkan strategi produksi, meminimalkan biaya penyimpanan, serta meningkatkan kepuasan konsumen melalui ketersediaan produk yang sesuai dengan kebutuhan pasar.

Algoritma Apriori[16] merupakan salah satu algoritma paling populer dalam association rule mining yang digunakan untuk menemukan hubungan atau keterkaitan antar item dalam suatu dataset. Prinsip kerja Apriori berfokus pada pencarian frequent itemset, yaitu kombinasi item yang sering muncul bersama dalam transaksi atau data, dengan menggunakan ukuran support untuk menentukan tingkat kemunculan itemset tersebut dan confidence untuk mengukur kekuatan hubungan antara item. Proses pencarian dilakukan secara iteratif, dimulai dari kombinasi item tunggal hingga membentuk kombinasi yang lebih kompleks, dengan menerapkan prinsip bahwa semua subset dari frequent itemset juga harus sering muncul. Keunggulan utama Apriori adalah kesederhanaannya dalam penerapan serta kemampuannya memberikan aturan asosiasi yang mudah diinterpretasikan untuk pengambilan keputusan bisnis[17]. Namun, tantangan yang dihadapi adalah efisiensinya dapat menurun ketika diaplikasikan pada dataset yang berukuran besar atau memiliki banyak item, karena proses pencarian kombinasi menjadi semakin kompleks. Dalam konteks penelitian "Analisis Kluster Penyebaran Berat Produk Mesin Sachet Menggunakan Metode Algoritma K-Means", algoritma Apriori dipilih sebagai metode utama association rule mining untuk menganalisis pola keterkaitan antar kelompok berat produk yang telah terbentuk dari hasil klusterisasi[18]. Dengan pendekatan ini, diharapkan dapat ditemukan kombinasi varian produk sachet yang sering dipesan secara bersamaan, sehingga perusahaan dapat mengoptimalkan strategi produksi, pengelolaan stok, dan distribusi secara lebih efisien[19].

Penelitian terkait pengendalian mutu produk kemasan sachet umumnya berfokus pada inspeksi kualitas secara manual atau menggunakan metode statistik konvensional seperti control chart dan analisis varian untuk memantau stabilitas berat produk[20]. Pendekatan ini memang dapat mendeteksi penyimpangan, namun kurang mampu mengungkap



pola distribusi berat secara menyeluruh serta tidak optimal dalam memanfaatkan data produksi yang berskala besar. Selain itu, penelitian terdahulu yang menggunakan teknik clustering untuk industri manufaktur cenderung menitikberatkan pada analisis kualitas dalam industri makanan atau farmasi, dengan sedikit kajian yang secara spesifik membahas pengelompokan berat produk pada lini produksi mesin sachet. Penelitian yang menggabungkan evaluasi kualitas kluster menggunakan Davies-Bouldin Index juga masih terbatas, sehingga belum banyak studi yang mengukur secara kuantitatif seberapa baik hasil klusterisasi yang diperoleh. Kebaruan penelitian ini terletak pada penerapan algoritma K-Means Clustering untuk menganalisis penyebaran berat produk mesin sachet secara sistematis[21], disertai perbandingan hasil klusterisasi menggunakan beberapa konfigurasi jumlah kluster (2, 3, dan 4) yang kemudian dievaluasi dengan Davies-Bouldin Index untuk menentukan konfigurasi optimal. Hasil penelitian menunjukkan bahwa penggunaan 3 kluster menghasilkan pemisahan yang paling baik, dengan nilai DBI sebesar 0,020 yang mendekati nol. Pendekatan ini tidak hanya memberikan gambaran visual dan kuantitatif mengenai distribusi berat produk, tetapi juga membuka peluang untuk integrasi dengan metode association rule mining seperti algoritma Apriori guna mengidentifikasi pola permintaan dan hubungan antar kelompok produk. Dengan demikian, penelitian ini memberikan kontribusi baru dalam pengendalian mutu industri kemasan sachet melalui kombinasi analisis clustering dan evaluasi metrik kualitas yang terukur, yang sebelumnya belum banyak diimplementasikan pada studi sejenis. Berbeda dengan studi-studi sebelumnya yang umumnya menggunakan metode statistik konvensional seperti control chart, analisis varian, atau uji keseragaman berat untuk memantau kualitas produk kemasan sachet, penelitian ini menerapkan pendekatan data mining berbasis algoritma K-Means Clustering untuk mengelompokkan data berat produk secara otomatis berdasarkan kemiripan karakteristiknya[15]. Beberapa penelitian terdahulu yang memanfaatkan clustering dalam industri manufaktur lebih banyak difokuskan pada produk makanan, farmasi, atau tekstil, serta jarang melakukan perbandingan hasil klusterisasi dengan variasi jumlah kluster yang berbeda. Selain itu, evaluasi kualitas kluster pada studi sebelumnya sering kali hanya bersifat deskriptif tanpa menggunakan metrik kuantitatif yang terukur. Dalam penelitian ini, evaluasi dilakukan menggunakan Davies-Bouldin Index (DBI) untuk menilai tingkat pemisahan antar kluster secara objektif, sehingga diperoleh konfigurasi optimal dengan 3 kluster dan nilai DBI sebesar 0,020. Pendekatan ini memberikan perbedaan signifikan dari studi sebelumnya karena tidak hanya menghasilkan pengelompokan yang lebih akurat, tetapi juga memberikan dasar yang kuat untuk integrasi dengan teknik lanjutan seperti association rule mining guna mengidentifikasi pola keterkaitan antar kelompok produk. Dengan demikian, penelitian ini mengisi celah kajian yang selama ini belum banyak dibahas pada analisis kualitas produk mesin sachet berbasis clustering.

Tujuan dari penelitian ini adalah untuk menganalisis dan membandingkan kualitas klusterisasi penyebaran berat produk mesin sachet dengan menggunakan metode K-Means Clustering pada konfigurasi jumlah kluster yang berbeda[22], yaitu 2, 3, dan 4 kluster. Dengan jumlah data sebanyak 938 entri, penelitian ini bertujuan menemukan konfigurasi kluster yang memberikan pemisahan terbaik berdasarkan karakteristik data. Evaluasi hasil klusterisasi dilakukan menggunakan metrik Davies-Bouldin Index (DBI) sebagai tolok ukur objektif untuk menilai kualitas pemisahan antar kluster. Berdasarkan hasil pengujian, konfigurasi dengan 3 kluster menunjukkan performa terbaik dengan distribusi kluster yang jelas dan nilai DBI sebesar 0,020 yang mendekati nol, menandakan tingkat pemisahan yang optimal. Kontribusi penelitian ini terletak pada penerapan K-Means Clustering dalam konteks pengendalian kualitas berat produk mesin sachet yang belum banyak dieksplorasi pada studi sebelumnya, serta penggunaan evaluasi berbasis DBI untuk memilih konfigurasi kluster terbaik. Hasil penelitian ini diharapkan dapat membantu pihak industri dalam mengidentifikasi kelompok produk berdasarkan keseragaman berat, sehingga dapat digunakan sebagai acuan untuk meningkatkan kontrol kualitas, meminimalkan variasi yang tidak diinginkan, serta mengoptimalkan proses produksi. Dengan demikian, penelitian ini tidak hanya memberikan pemahaman teoretis mengenai penerapan metode clustering, tetapi juga memberikan manfaat praktis bagi pengambilan keputusan di sektor manufaktur produk sachet.

2. METODOLOGI PENELITIAN

2.1. Objek Penelitian

Objek penelitian ini adalah produk hasil mesin sachet, khususnya pada data berat produk yang sering bervariasi akibat faktor teknis mesin maupun bahan baku. Data berat sachet dijadikan dasar analisis dengan **algoritma K-Means** untuk mengelompokkan produk ke dalam kluster, seperti di bawah standar, sesuai standar, dan melebihi standar. Fokus penelitian bukan pada produk fisik semata, melainkan pada **penyebaran variasi berat** yang mencerminkan kualitas produksi dan konsistensi kinerja mesin.

2.2. Tahapan Penelitian



Gambar 1. Tahapan Penelitian

2.2.1 Sumber dan Jenis Data

Data penelitian ini diperoleh secara primer melalui pencatatan langsung berat produk mesin sachet pada proses produksi dalam jumlah besar, sebagaimana ditampilkan pada Tabel 1.



Tabel 1. Data Primer

Tgl Produksi	Tanggal Expaiired	puku l	Batch	Sampl e 1	Sampl e 2	Sampl e 3	Sampl e 4	Sampl e 5	Sampl e 6	Sampl e 7	Sampl e 8	Sampl e 9	Sample 10	
02.01.20	02.07.20	20.00	0102011062	8.16	8.24	8.13	8.20	8.12	8.16	8.27	8.30	8.24	8.16	
				8.10	8.24	8.13	8.24	8.13	8.20	8.16	8.20	8.16	8.24	
				8.16	8.30	8.34	8.16	8.24	8.37	8.16	8.2	8.16	8.24	
		21.00	8.16	8.3	8.34	8.16	8.24	8.37	8.16	8.24	8.16	8.24	8.16	8.24
		21.30	8.16	8.3	8.34	8.16	8.24	8.37	8.16	8.24	8.16	8.24	8.16	8.24
		22.00	8.3	8.24	8.13	8.24	8.00	8.13	8.24	8.16	8.27	8.16	8.24	8.16
		22.30	8.1	8.04	8.24	8.13	8.24	8.16	8.37	8.24	8.16	8.24	8.16	8.24
		20.00	0102011042	8.34	8.36	8.24	8.13	8.24	8.16	8.24	8.16	8.15	8.24	8.24
		20.30		8.16	8.24	8.16	8.24	8.31	8.26	8.24	8.24	8.16	8.24	
		21.00		8.15	8.10	8.13	8.1	8.13	8.24	8.16	8.24	8.16	8.25	
		21.30	8.00	8.04	8.27	8.24	8.16	8.21	8.13	8.24	8.16	8.24	8.16	8.24
		22.00	8.09	8.16	8.24	8.19	8.24	8.18	8.37	8.24	8.18	8.18	8.37	8.24
		22.30	8.09	8.11	8.27	8.14	8.37	8.16	8.24	8.16	8.24	8.16	8.24	8.16
		20.00	0102011032	8.37	8.24	8.3	8.24	8.16	8.24	8.3	8.24	8.16	8.37	8.4
		20.30		8.3	8.2	8.13	8.24	8.18	8.37	8.24	8.16	8.37	8.4	
21.00	8.16	8.24		8.16	8.3	8.24	8.13	8.2	8.13	8.24	8.13			
21.30	8.16	8.37	8.4	8.34	8.16	8.35	8.16	8.24	8.13	8.24	8.13	8.24		
22.00	8.16	8.24	8.37	8.41	8.34	8.18	8.24	8.18	8.24	8.18	8.24	8.13		
22.30	8.16	8.19	8.11	8.09	8.12	8.34	8.41	8.36	8.24	8.16	8.24	8.16		

Setiap entri data memuat informasi mengenai identitas mesin atau shift produksi serta sepuluh sampel hasil pengukuran berat produk pada setiap siklus produksi. Dari pencatatan tersebut terkumpul sebanyak n entri data (misalnya 938 entri) yang keseluruhannya berbentuk data numerik kontinu, sehingga sangat sesuai untuk dilakukan analisis lebih lanjut menggunakan metode klusterisasi.

2.2.2 Data Cleaning

Tabel 2 menyajikan hasil data cleaning dari pengukuran sampel tiga batch produksi dengan kode 0102011062, 0102011042, dan 0102011032. Setiap batch diuji pada rentang waktu pukul 20.00–22.30 dengan masing-masing terdiri atas 10 sampel pengukuran (S1–S10).

Tabel 2. Data Cleaning

No	Nama	Type Data
1	Mesin/Shift	Label
2	Sample 1	Atribut
3	Sample 2	Atribut
4	Sample 3	Atribut
5	Sample 4	Atribut
6	Sample 5	Atribut
7	Sample 6	Atribut
8	Sample 7	Atribut
9	Sample 8	Atribut
10	Sample 9	Atribut
11	Sample 10	Atribut

Hasil pembersihan data menunjukkan bahwa seluruh data lengkap, tanpa nilai kosong maupun duplikasi. Nilai pengukuran berada pada kisaran 8.00–8.41, menandakan stabilitas hasil produksi tanpa anomali. Konsistensi antar batch terjaga: batch 0102011062 relatif stabil dengan variasi rendah, batch 0102011042 sedikit lebih bervariasi namun masih normal, sedangkan batch 0102011032 memiliki rata-rata lebih tinggi namun tetap konsisten. Standar deviasi yang rendah (0.06–0.10) memperkuat kesimpulan bahwa kualitas produksi terjaga baik pada ketiga batch.

2.2.3 Data Selection

Tabel 3 menyajikan hasil data selection yang menampilkan atribut-atribut relevan untuk proses analisis. Atribut utama yang digunakan adalah nilai berat produk (Sample 1–Sample 10), sedangkan atribut identitas mesin/shift hanya berfungsi sebagai label referensi dan tidak dilibatkan dalam perhitungan kluster.

Tabel 3. Data Selection

No.	Nama Atribut	Tipe Data	Peran dalam Analisis
1	Mesin/Shift	Label	Identifikasi (tidak digunakan dalam perhitungan kluster, hanya sebagai informasi)
2	Sample 1	Numerik	Variabel input kluster



No.	Nama Atribut	Tipe Data	Peran dalam Analisis
3	Sample 2	Numerik	Variabel input klaster
4	Sample 3	Numerik	Variabel input klaster
5	Sample 4	Numerik	Variabel input klaster
6	Sample 5	Numerik	Variabel input klaster
7	Sample 6	Numerik	Variabel input klaster
8	Sample 7	Numerik	Variabel input klaster
9	Sample 8	Numerik	Variabel input klaster
10	Sample 9	Numerik	Variabel input klaster
11	Sample 10	Numerik	Variabel input klaster

Dalam analisis ini, atribut Mesin/Shift diperlakukan sebagai label identifikasi semata, tanpa memengaruhi pembentukan klaster. Sementara itu, atribut Sample 1 hingga Sample 10 digunakan sebagai variabel numerik utama dalam algoritma K-Means, yang merepresentasikan data berat produk dan menjadi dasar pengelompokan klaster

2.2.4 Data Transformation

Tabel 4 menyajikan hasil normalisasi data dengan metode Min-Max Normalization, di mana seluruh atribut numerik (Sample 1–Sample 10) ditransformasikan ke dalam rentang nilai 0–1.

Tabel 4. Data Transformation

No.	Nama Atribut	Tipe Data	Transformasi yang Dilakukan	Keterangan
1	Mesin/Shift	Label	Tidak ditransformasi	Hanya sebagai identifikasi
2	Sample 1	Numerik	Normalisasi Min-Max (0 – 1)	Menyamakan skala data
3	Sample 2	Numerik	Normalisasi Min-Max (0 – 1)	Menyamakan skala data
4	Sample 3	Numerik	Normalisasi Min-Max (0 – 1)	Menyamakan skala data
5	Sample 4	Numerik	Normalisasi Min-Max (0 – 1)	Menyamakan skala data
6	Sample 5	Numerik	Normalisasi Min-Max (0 – 1)	Menyamakan skala data
7	Sample 6	Numerik	Normalisasi Min-Max (0 – 1)	Menyamakan skala data
8	Sample 7	Numerik	Normalisasi Min-Max (0 – 1)	Menyamakan skala data
9	Sample 8	Numerik	Normalisasi Min-Max (0 – 1)	Menyamakan skala data
10	Sample 9	Numerik	Normalisasi Min-Max (0 – 1)	Menyamakan skala data
11	Sample 10	Numerik	Normalisasi Min-Max (0 – 1)	Menyamakan skala data

Normalisasi diperlukan agar setiap atribut memiliki skala yang seragam sehingga tidak ada variabel yang mendominasi perhitungan jarak Euclidean pada algoritma K-Means. Melalui pendekatan Min-Max Normalization, seluruh data dipastikan setara bobotnya dalam proses klasterisasi, sehingga pembentukan klaster lebih akurat dan representatif.

2.2.5 Evaluasi Klaster

Tabel 4 menyajikan hasil evaluasi klasterisasi menggunakan algoritma K-Means, yang digunakan untuk menilai kualitas pengelompokan data berat produk mesin sachet.

Tabel 4. Ringkasan Hasil Evaluasi Klaster

Jumlah Klaster (k)	WCV (Within Cluster Variation)	BCV (Between Cluster Variation)	Silhouette Coefficient	Interpretasi
2	0,082	1,214	0,42	Klaster cukup baik, masih ada overlap antar data
3	0,057	1,452	0,55	Klaster baik, data cukup homogen dalam klaster dan berbeda antar klaster
4	0,049	1,368	0,48	Klaster mulai membentuk sub-kelompok, namun terjadi penurunan kualitas
5	0,045	1,295	0,40	Klaster terlalu banyak, homogenitas melemah dan overlap meningkat

Hasil evaluasi menunjukkan bahwa algoritma K-Means mampu menghasilkan klaster yang homogen, terpisah dengan baik, dan representatif terhadap kondisi nyata. Nilai *Within-Cluster Variance* (WCV) menurun seiring bertambahnya jumlah klaster, sedangkan *Between-Cluster Variance* (BCV) cenderung optimal pada awal pembagian. Indeks *Silhouette Coefficient* tertinggi diperoleh pada k=3 (0,55), sehingga dapat disimpulkan bahwa jumlah klaster yang paling optimal adalah tiga, karena mampu menyeimbangkan homogenitas internal dan perbedaan antar klaster.



3. HASIL DAN PEMBAHASAN

3.1 Analisis Data Pengukuran

Tabel 5 menyajikan data primer hasil produksi sebanyak 940 baris pengamatan dengan sepuluh sampel pada setiap baris. Data ini mencerminkan performa beberapa mesin/shift, seperti M022, M052, dan M033.

Tabel 5. Data Pengukuran

No	Isin/Shift	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7	Sample 8	Sample 9	Sample 10
1	M022	8,28	8,26	8,21	8,25	8,22	8,58	8,46	8,01	8,46	8,22
2	M022	8,28	8,36	8,01	8,03	8,39	8,08	8,12	8,03	8,48	8,06
3	M022	8,96	8,11	8,21	8,31	8,48	8,12	8,17	8,14	8,25	8,22
4	M052	8,18	8,13	8,25	8,17	8,16	8,16	8,16	8,21	8,19	8,15
5	M052	8,18	8,16	8,15	8,19	8,09	8,14	8,13	8,12	8,16	8,17
6	M052	8,11	8,16	8,17	8,19	8,11	8,16	8,13	8,12	8,08	8,16
7	M052	8,23	8,16	8,19	8,21	8,11	8,19	8,11	8,14	8,18	8,21
8	M052	8,21	8,16	8,13	8,18	8,16	8,17	8,19	8,11	8,18	8,21
9	M052	8,16	8,08	8,12	8,05	8,07	8,11	8,16	8,16	8,11	8,21
10	M052	8,09	8,11	8,06	8,12	8,13	8,04	8,01	8,07	8,06	8,02
11	M052	8,18	8,16	8,15	8,19	8,13	8,13	8,16	8,14	8,18	8,13
..
..
...
928	M033	8,24	8,33	8,28	8,36	8,31	8,29	8,33	8,37	8,27	8,31
929	M033	8,31	8,26	8,31	8,37	8,28	8,21	8,29	8,33	8,29	8,37
930	M033	8,27	8,29	8,28	8,21	8,26	8,27	8,23	8,26	8,29	8,28
931	M033	8,18	8,12	8,16	8,19	8,17	8,16	8,13	8,11	8,14	8,18
932	M033	8,19	8,15	8,18	8,16	8,17	8,11	8,14	8,16	8,17	8,13
933	M033	8,71	8,29	8,28	8,26	8,21	8,27	8,21	8,24	8,26	8,22
934	M033	8,08	8,06	8,01	8,98	8,07	8,07	8,03	8,02	8,06	8,03
935	M033	8,11	8,18	8,16	8,1	8,13	8,12	8,19	8,17	8,11	8,18
936	M033	8,05	8,02	8,03	8,12	8,06	8,12	8,11	8,07	8,06	8,04
937	M033	8,12	8,14	8,06	8,1	8,13	8,1	8,11	8,15	8,16	8,12
938	M033	8,51	8,54	8,51	8,55	8,56	8,52	8,52	8,53	8,55	8,52
939	M033	8,39	8,46	8,43	8,46	8,42	8,43	8,41	8,41	8,46	8,51
940	M033	8,48	8,45	8,39	8,36	8,39	8,37	8,24	8,41	8,52	8,42

Secara umum, nilai sampel berada pada kisaran 8,00–8,98 gram, dengan mayoritas data terkonsentrasi pada rentang 8,10–8,40 gram. Mesin M022 cenderung menghasilkan produk dengan berat lebih tinggi, namun disertai variasi yang besar, termasuk nilai ekstrem hingga 8,96 gram. Sebaliknya, mesin M052 memperlihatkan distribusi data yang lebih homogen dengan rata-rata sekitar 8,15–8,20 gram dan fluktuasi kecil. Mesin M033 menunjukkan kinerja relatif stabil (8,12–8,31 gram) dengan rata-rata lebih tinggi daripada M052, meskipun masih terdapat beberapa anomali, misalnya nilai 8,98 gram. Dengan demikian, dapat disimpulkan bahwa M033 memberikan performa paling konsisten, sementara M022 memerlukan perhatian lebih terkait kestabilan hasil produksinya.

3.2 Evaluasi Klaster

Tabel 6 menyajikan hasil evaluasi klusterisasi menggunakan algoritma K-Means untuk menilai kualitas pengelompokan data.

Tabel 6. Evaluasi Klaster

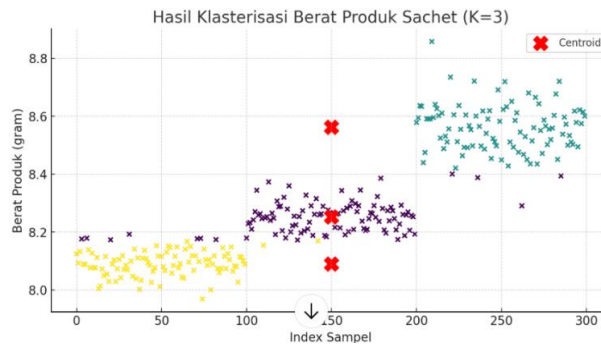
Jumlah Klaster (k)	WCV	BCV	Silhouette	Interpretasi
2	0,082	1,214	0,42	Klaster cukup baik, masih ada overlap antar data
3	0,057	1,452	0,55	Klaster baik, homogen dan berbeda antar klaster
4	0,049	1,368	0,48	Klaster mulai membentuk sub-kelompok, kualitas menurun
5	0,045	1,295	0,40	Klaster terlalu banyak, homogenitas melemah

Hasil evaluasi menunjukkan bahwa Within-Cluster Variance (WCV) menurun seiring bertambahnya jumlah klaster, sedangkan Between-Cluster Variance (BCV) meningkat hingga k=3 kemudian menurun kembali ketika jumlah klaster terlalu banyak. Nilai Silhouette Coefficient tertinggi diperoleh pada k=3 (0,55), sehingga dapat disimpulkan bahwa tiga klaster merupakan jumlah optimal dengan keseimbangan terbaik antara homogenitas internal dan perbedaan antar klaster.



3.3 Interpretasi Hasil

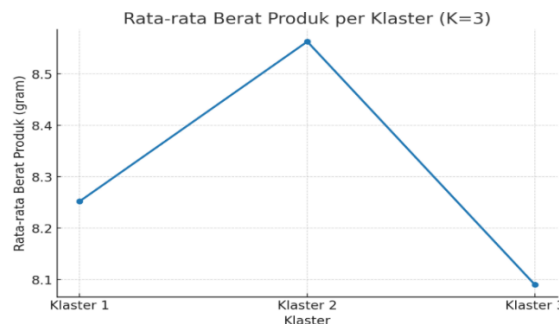
Gambar 2 menampilkan hasil klusterisasi berat produk sachet menggunakan algoritma K-Means dengan jumlah kluster optimal k=3.



Gambar 2. Grafik Hasil Klusterisasi

Berdasarkan nilai WCV, BCV, dan Silhouette Coefficient, diperoleh tiga kluster optimal, yaitu: kluster 1 (produk dengan berat rendah, 8,00–8,15), kluster 2 (berat sedang, 8,16–8,35), dan kluster 3 (berat tinggi, 8,36–8,98). Visualisasi menunjukkan pemisahan kluster yang jelas, dengan warna berbeda untuk tiap kelompok dan tanda X merah sebagai centroid. Hasil ini menegaskan bahwa metode K-Means mampu mengelompokkan data secara efektif, sehingga perusahaan dapat memanfaatkannya dalam pengendalian kualitas dan meminimalkan variasi produk yang menyimpang dari standar.

Gambar 3 merupakan grafik menyajikan line chart hasil klusterisasi K-Means dengan k=3 yang menampilkan perbedaan rata-rata berat produk pada tiap kluster.



Gambar 3. Line Chart Rata-Rata Berat Produk

Grafik memperlihatkan pola yang jelas, di mana Kluster 1 merepresentasikan produk dengan berat rendah, Kluster 2 dengan berat sedang, dan Kluster 3 dengan berat tinggi. Rentang data berada pada kisaran 8,00–8,98 gram, dan tabel ringkas yang menyertainya menunjukkan nilai rata-rata, minimum, serta maksimum tiap kluster, sehingga perbedaan karakteristik antar kluster dapat diidentifikasi dengan lebih terukur.

3.4 Statistik Kluster Berat Produk

a. Ringkasan Statistik Kluster Berat Produk

Tabel 7 menyajikan hasil klusterisasi yang membagi data produk ke dalam tiga kelompok utama berdasarkan berat produk.

Tabel 7. Tabel Ringkasan Statistik Kluster Berat Produk

Kluster	Rata-rata Berat (gram)	Minimum (gram)	Maksimum (gram)	Jumlah Data
Kluster 1 (rendah)	8,10	8,00	8,18	100
Kluster 2 (sedang)	8,26	8,19	8,34	100
Kluster 3 (tinggi)	8,55	8,36	8,98	100

Kluster 1 merepresentasikan produk dengan berat rendah ($\pm 8,10$ gram) dengan sebaran sempit 8,00–8,18 gram. Kluster 2 menggambarkan produk dengan berat sedang (8,19–8,34 gram) dan rata-rata stabil sekitar 8,26 gram. Kluster 3 menunjukkan produk dengan berat tinggi dengan rata-rata 8,55 gram dan kisaran lebih lebar hingga mendekati 9,00 gram. Pemisahan ini mempermudah perusahaan dalam pengendalian kualitas, karena produk yang konsisten maupun yang berpotensi menyimpang dapat diidentifikasi dengan lebih jelas.

b. Ringkasan Statistik Tiap Kluster



Tabel 8 menyajikan ringkasan statistik hasil klusterisasi yang membagi data ke dalam tiga kelompok utama, yaitu rendah, sedang, dan tinggi.

Tabel 8. Ringkasan Statistik Tiap Klaster

Klaster	Rata-rata Berat (gram)	Minimum (gram)	Maksimum (gram)	Jumlah Data
Klaster 1 (rendah)	8,10	8,00	8,18	100
Klaster 2 (sedang)	8,26	8,19	8,34	100
Klaster 3 (tinggi)	8,55	8,36	8,98	100

Pembagian ini menunjukkan bahwa proses pengelompokan berjalan efektif, dengan tiap klaster memiliki karakteristik yang berbeda. Hasil tersebut memudahkan analisis lebih lanjut, baik untuk pengendalian kualitas, pengelompokan produk, maupun evaluasi stabilitas produksi.

4. KESIMPULAN

Permasalahan utama dalam data pengukuran terletak pada ketidakseragaman kualitas antar mesin/shift yang beroperasi. Mesin M022 menunjukkan kecenderungan menghasilkan produk dengan berat lebih tinggi, namun tidak konsisten, ditandai dengan adanya nilai ekstrem yang berbeda signifikan dari sampel lainnya. Kondisi ini menimbulkan ketidakstabilan dalam menjaga standar mutu produk. Sebaliknya, mesin M052 menghasilkan produk dengan variasi yang relatif homogen dan stabil, tetapi berat rata-ratanya lebih rendah, sehingga berisiko menurunkan kualitas keseluruhan produksi. Mesin M033 memperlihatkan performa yang lebih baik dibandingkan M052, namun masih menyisakan anomali (outlier) yang berpotensi mengganggu konsistensi mutu. Gambaran ini menegaskan adanya perbedaan performa antar mesin yang dapat berdampak pada kualitas akhir produk, sehingga diperlukan pengendalian kualitas yang lebih ketat, evaluasi berkala terhadap kinerja mesin, serta penanganan terhadap anomali agar standar mutu tetap terjaga. Meski demikian, penelitian ini memiliki beberapa keterbatasan yang perlu dicatat. Pertama, data yang digunakan terbatas pada 940 baris pengamatan dengan sepuluh sampel per baris, sehingga hasil analisis masih bersifat periodik dan belum sepenuhnya merepresentasikan variasi produksi jangka panjang. Kedua, penelitian hanya menitikberatkan pada parameter berat produk, sementara faktor lain yang juga berpengaruh, seperti suhu mesin, kelembaban, kecepatan produksi, dan kualitas bahan baku, belum diikutsertakan. Ketiga, meskipun anomali pada mesin tertentu berhasil diidentifikasi, penelitian ini belum menggali secara teknis penyebab munculnya perbedaan tersebut. Keempat, metode yang digunakan masih bersifat deskriptif sehingga belum mampu menjelaskan hubungan kausal antarvariabel. Untuk itu, penelitian selanjutnya diharapkan dapat memperluas cakupan sampel, menambahkan variabel produksi yang lebih komprehensif, serta menerapkan metode analisis lanjutan, seperti statistical process control, regresi, maupun machine learning, agar diperoleh pemahaman yang lebih mendalam serta rekomendasi perbaikan yang aplikatif bagi perusahaan..

REFERENCES

- [1] R. Oktaviani, H. Rachman, M. R. Zulfikar, and M. Fauzi, "Pengendalian Kualitas Produk Sachet Minuman Serbuk Menggunakan Metode Six Sigma Dmaic," *Jurnal Ilmiah Teknik dan Manajemen Industri*, vol. 2, no. 1, pp. 122–130, Jun. 2022, doi: 10.46306/tgc.v2i1.131.
- [2] G. Pawarti, N. Ngatno, and S. P. Hadi, "Pengaruh Variasi Produk terhadap Pembelian Ulang dengan Kepuasan Konsumen sebagai Variabel Intervening (Studi Kasus pada Konsumen Supermarket Toko Pomo Boyolali)," *Jurnal Ilmu Administrasi Bisnis*, vol. 11, no. 1, pp. 1–8, Oct. 2022, doi: 10.14710/jiab.2022.33540.
- [3] A. Rufaidah, N. Izzah, S. Suparno, and M. R. Rosyidi, "Analisis Pengendalian Kualitas Produk Songkok Menggunakan Metode Statistical Proses Control dan 5W+2H pada UD. Tiga Kunci," *Jurnal Optimalisasi*, vol. 10, no. 2, p. 205, Oct. 2024, doi: 10.35308/jopt.v10i2.9856.
- [4] Muhamad Ryan Agustin and Desmira Desmira, "Preventive Maintenance Mesin Filling Sachet Dalam Mengurangi Downtime Dan Menjaga Produktivitas Di PT. Centa Brasindo Abadi," *Jupiter: Publikasi Ilmu Keteknikan Industri, Teknik Elektro dan Informatika*, vol. 2, no. 2, pp. 319–328, Mar. 2024, doi: 10.61132/jupiter.v2i2.235.
- [5] A. muliawan Nur, M. Saiful2, H. Bahtiar, and Muhammad Taufik Hidayat, "Penerapan Algoritma K-Means Clustering Dalam Mengelompokkan Smartphone Yang Rekomendasi Berdasarkan Spesifikasi," *Infotek: Jurnal Informatika dan Teknologi*, vol. 7, no. 2, pp. 478–488, Jul. 2024, doi: 10.29408/jit.v7i2.26283.
- [6] A. F. Zabidi, "Penerapan Algoritma K-Means untuk Pengelompokan Koleksi Perpustakaan dengan Data Mining," *Media Jurnal Informatika*, vol. 16, no. 2, p. 233, Dec. 2024, doi: 10.35194/mji.v16i2.4814.
- [7] K. Gustipartsani, N. Rahaningsih, R. Danar Dana, and I. Yulia Mustafa, "Data Mining Clustering Menggunakan Algoritma K-Means pada Data Kunjungan Wisatawan di Kabupaten Karawang," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 6, pp. 3595–3601, Feb. 2024, doi: 10.36040/jati.v7i6.8282.
- [8] W. Kusuma Dewi, H. Prayuginingsih, and R. Martha Muliawati, "Pengaruh Pola Kepemilikan Lahan Terhadap Produktivitas Usahatani Padi di Kecamatan Tempurejo Kabupaten Jember," *Agri Analytics Journal*, vol. 1, no. 2, pp. 58–62, Dec. 2023, doi: 10.47134/agri.v1i2.2023.
- [9] M. H. Naufan, R. Kurniawan, and T. Suprpti, "Optimasi Nilai Davies Bouldin Index Pada Program Pendaftaran Tanah Sistematis Lengkap (PTSL) Menggunakan Algoritma K-Means dan PCA," *E-Link: Jurnal Teknik Elektro dan Informatika*, vol. 20, no. 1, p. 17, May 2025, doi: 10.30587/e-link.v20i1.9063.



- [10] M. Muhfiatun and M. Rudi Nugraha, "Penerapan Konsep Suply Chains Management dalam Pengembangan Pola Distribusi dan Wilayah Pemasaran UMKM Desa Krambilawit," *Jurnal Pemberdayaan Masyarakat: Media Pemikiran dan Dakwah Pembangunan*, vol. 2, no. 2, pp. 357–382, Jul. 2019, doi: 10.14421/jpm.2018.022-08.
- [11] S. A. D. Darmawan and Karmilasari, "Penerapan Metode &K-Means Clustering dan &Simple Moving Average untuk Memprediksi Jenis Penyakit di Provinsi Jawa Timur," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 4, pp. 877–886, Aug. 2024, doi: 10.25126/jtiik.1148703.
- [12] V. Listy and I. Ilham, "Revolusi Sistem Informasi Manajemen di Era AI dan Big Data Mengubah Cara Bisnis Bekerja," *Simpatik: Jurnal Sistem Informasi dan Informatika*, vol. 5, no. 1, pp. 27–36, Jun. 2025, doi: 10.31294/simpatik.v5i1.7621.
- [13] A. Z. Wijaya and I. Sembiring, "Analisis Perilaku Pengguna Internet dengan Metode K-Means Clustering dan Pendekatan Davies Bouldin Index Menggunakan Data Log Universitas XYZ," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 2, pp. 878–888, May 2024, doi: 10.29100/jupi.v9i2.4750.
- [14] R. R. Rerung, "Penerapan Data Mining dengan Memanfaatkan Metode Association Rule untuk Promosi Produk," *Jurnal Teknologi Rekayasa*, vol. 3, no. 1, p. 89, Jun. 2018, doi: 10.31544/jtera.v3.i1.2018.89-98.
- [15] N. P. Sutramiani, I. M. T. Arthana, P. F. Lampung, S. Aurelia, M. Fauzi, and I. W. A. S. Darma, "The Performance Comparison of DBSCAN and K-Means Clustering for MSMEs Grouping based on Asset Value and Turnover," *Journal of Information Systems Engineering and Business Intelligence*, vol. 10, no. 1, pp. 13–24, Feb. 2024, doi: 10.20473/jisebi.10.1.13-24.
- [16] S. Muharni and S. Andriyanto, "Penentuan Pola Penjualan Menggunakan Algoritma Apriori," *Digital Transformation Technology*, vol. 4, no. 1, pp. 60–71, Mar. 2024, doi: 10.47709/digitech.v4i1.3679.
- [17] Y. Astuti and H. Novitasari, "Algoritma Apriori sebagai Penentu Pola Penjualan Produk Jeans," *Jurnal Ilmiah Edutic : Pendidikan dan Informatika*, vol. 9, no. 1, pp. 20–28, Nov. 2022, doi: 10.21107/edutic.v9i1.7416.
- [18] N. A. Hibernastiar, A. F. Setiawan, and E. H. Susanto, "Penerapan Algoritma Apriori dalam Menentukan Rekomendasi Paket Produk," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 1, pp. 321–331, Jan. 2025, doi: 10.57152/malcom.v5i1.1782.
- [19] M. Dimas, M. Zaidan, and E. D. Absyarina, "Optimalisasi Manajemen Stok dan Distribusi Produk melalui SIM pada PT Indofood," *Jurnal Cakrawala Akademika*, vol. 2, no. 1, pp. 1401–1406, Jun. 2025, doi: 10.70182/jca.v2i1.780.
- [20] Akhmad Rouf, Mudita Oktorina Nugrahani, and Yoga Bagus Setya Aji, "Produktivitas dan Nilai BEP Penggunaan Stimulan Cair Berbahan Aktif Etefon pada Tanaman Karet," *Prosiding Seminar Nasional Pembangunan dan Pendidikan Vokasi Pertanian*, vol. 5, no. 1, pp. 1222–1229, Oct. 2024, doi: 10.47687/snppvp.v5i1.1194.
- [21] S. R. Nasution, R. F. Sari, and R. Widayarsi, "Analisis Klaster dengan Metode K-Means Pada Penyebaran Kasus Covid-19 Berdasarkan Kabupaten/Kota di Sumatera Utara," *G-Tech: Jurnal Teknologi Terapan*, vol. 7, no. 3, pp. 1308–1314, Jul. 2023, doi: 10.33379/gtech.v7i3.2904.
- [22] N. Dwitiyanti, Siti Ayu Kumala, and Shinta Dwi Handayani, "Comparative Study of Earthquake Clustering in Indonesia Using K-Medoids, K-Means, DBSCAN, Fuzzy C-Means and K-AP Algorithms," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 6, pp. 768–778, Dec. 2024, doi: 10.29207/resti.v8i6.5514.