Department of Digital Business
Journal of Artificial Intelligence and Digital Business (RIGGS)
Homepage: https://journal.ilmudata.co.id/index.php/RIGGS
Vol. 4 No. 3 (2025) pp: 4221-4228
P-ISSN: 2963-9298, e-ISSN: 2963-914X

Analisis Pola Pemakaian Bahan Produksi Menggunakan Algoritma K-Means Clustering di PT. Loista Indonesia

Shifa Hairul Nisah[1], Nova Dwi Maelani[2], Niken Ayu Dara[3], Sefrika Entas[4]
[1234]Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika
shifahaerulnisah663@gmail.com, nopasmaelani03@gmail.com, nikenayudr2003@gmail.com, sefrika.sfe@bsi.ac.id

Abstrak

Industri furnitur tidak hanya dituntut menghasilkan produk berkualitas, tetapi juga harus mampu mengelola pemakaian bahan produksi secara efisien. Salah satu masalah yang kerap muncul adalah perbedaan antara rencana kebutuhan bahan (Bill of Material/BoM) dengan realisasi di lapangan, yang bisa berujung pada sisa bahan atau bahkan pemborosan. Penelitian ini mencoba menghadirkan solusi melalui penerapan algoritma K-Means Clustering sebagai bagian dari pendekatan data mining. Studi dilakukan di PT. LOISTA INDONESIA dengan memanfaatkan data historis pemakaian bahan produksi. Tahapan penelitian meliputi pra-pemrosesan data, perhitungan selisih pemakaian dan tingkat efisiensi, normalisasi, hingga pengelompokan dengan K-Means baik secara manual maupun menggunakan RapidMiner. Hasil analisis menunjukkan terbentuknya tiga pola utama: kelompok efisien sesuai rencana, kelompok dengan pemakaian lebih rendah dari rencana (potensi sisa), dan kelompok dengan pemakaian lebih tinggi dari rencana (potensi boros). Evaluasi kualitas klaster menggunakan Davies Bouldin Index (DBI) menghasilkan nilai 0,234 yang menandakan struktur pengelompokan cukup baik. Temuan ini tidak hanya membantu perusahaan memahami perilaku pemakaian bahan, tetapi juga menjadi pijakan untuk memperbaiki akurasi perencanaan, mengurangi inefisiensi, serta mendorong pengambilan keputusan yang lebih cerdas dan berbasis data.

Kata kunci: Clustering, Efisiensi Bahan, K-Means, Rapidminer

1. Latar Belakang

Industri furnitur memiliki tingkat kompleksitas tinggi dalam mengelola bahan baku. Ketepatan antara *Bill of Material* (BoM) dan realisasi pemakaian aktual merupakan faktor kunci untuk menjaga efisiensi produksi. Namun, kenyataannya masih sering ditemukan ketidaksesuaian yang berimplikasi serius pada biaya, waktu, dan kepuasan pelanggan. PT. Loista Indonesia, misalnya, mencatat rata-rata selisih pemakaian bahan mencapai 385% pada Februari–Maret 2025, yang berakibat pada penumpukan stok dan potensi pemborosan produksi. Fakta ini menunjukkan perlunya pendekatan baru dalam mengolah data produksi agar dapat menghasilkan insight yang berguna untuk evaluasi dan perbaikan perencanaan material.

Sejumlah penelitian terdahulu telah menunjukkan efektivitas algoritma K-Means dalam mengelompokkan data yang kompleks. Bagustio et al. (2024) membuktikan penerapannya dalam menganalisis data penjualan industri kecantikan[1], sementara Mustakim et al. (2025) berhasil menggunakan metode ini untuk segmentasi transaksi nasabah di Bank BJB[2]. Arifin et al. (2022) menggunakannya untuk memprediksi stok bahan baku produksi[3], sedangkan Alvianatinova et al. (2024) menerapkannya pada pengelompokan data penjualan supermarket[4]. Kusuma et al. (2023) juga menegaskan manfaat K-Means dalam mengelola persediaan barang di perusahaan retail. Bahkan Marsono et al. (2021) dan Febianda et al. (2020) mengkombinasikan K-Means dengan metode evaluasi seperti *Davies Bouldin Index* (DBI) dan *Silhouette Score* untuk mengukur kualitas klasterisasi[5][6].

Kendati demikian, mayoritas penelitian masih berfokus pada data transaksi penjualan atau perilaku konsumen, bukan pada konsumsi bahan baku dalam proses manufaktur. Celah ini penting untuk ditutup, mengingat pola pemakaian material yang tidak sesuai BoM dapat menjadi sumber inefisiensi yang nyata di lini produksi. Penelitian ini hadir dengan sudut pandang berbeda: memanfaatkan data historis penggunaan bahan produksi furnitur untuk menemukan pola efisiensi dan inefisiensi dengan algoritma K-Means. Fokus ini memberikan kontribusi unik karena menempatkan pengendalian material sebagai titik analisis utama, bukan sekadar aspek hilir seperti penjualan.

Kebaruan lain terletak pada integrasi analisis *clustering* dengan sistem ERP berbasis cloud yang sudah digunakan perusahaan. Jika penelitian sebelumnya berhenti pada penyajian laporan atau prediksi stok, maka studi ini berupaya mengolah data ERP menjadi dasar pengambilan keputusan strategis. Dengan cara ini, perusahaan tidak hanya dapat mengidentifikasi lini produksi yang boros atau efisien, tetapi juga melakukan koreksi terhadap BoM, mengurangi pemborosan, dan meningkatkan ketepatan perencanaan material. Pendekatan ini sejalan dengan tren transformasi digital industri manufaktur yang menekankan pemanfaatan data sebagai fondasi pengambilan keputusan [7].

Berdasarkan paparan tersebut, penelitian ini difokuskan untuk menjawab dua pertanyaan kunci: (1) bagaimana pola penggunaan bahan baku produksi furnitur dapat dipetakan menggunakan algoritma K-Means, dan (2) sejauh mana hasil klasterisasi dapat dimanfaatkan untuk meningkatkan akurasi BoM, efisiensi pengadaan, dan pengendalian material. Dengan menjawab pertanyaan ini, penelitian diharapkan tidak hanya memberi kontribusi akademis pada bidang *data mining*, tetapi juga manfaat praktis bagi industri furnitur dalam meningkatkan daya saing melalui efisiensi berbasis data.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis klaster berbasis algoritma K-Means. Metode ini dipilih karena mampu mengelompokkan pola penggunaan bahan produksi secara objektif berdasarkan data numerik yang tercatat dalam sistem produksi [8]. Proses penelitian disusun agar dapat direplikasi dengan mengikuti tahapan yang jelas, mulai dari pengumpulan data hingga analisis hasil akhir.

Data penelitian bersumber dari catatan produksi aktual perusahaan manufaktur furnitur pada periode Februari hingga Maret 2025. Informasi yang digunakan mencakup kode bahan baku, kebutuhan bahan berdasarkan BoM (*Bill of Material*), jumlah bahan yang dikeluarkan, serta stok yang tersisa. Dari keseluruhan catatan Surat Perintah Kerja (SPK) pada periode tersebut, peneliti mengambil sampel menggunakan teknik purposive sampling, dengan kriteria data harus lengkap, termasuk dalam kategori produk utama furnitur, serta mewakili beberapa jenis proses produksi.

Sebelum dilakukan analisis, data melalui tahap pra-pemrosesan berupa normalisasi numerik menggunakan metode min–max scaling dengan rentang 0–1. Normalisasi ini dilakukan agar skala tiap atribut numerik berada dalam rentang yang sama, sehingga algoritma K-Means dapat bekerja secara optimal. Setelah itu, dilakukan transformasi data dengan menambahkan atribut baru berupa selisih pemakaian dan tingkat efisiensi, yang penting untuk menghasilkan klaster yang lebih bermakna.

Penerapan algoritma K-Means dilakukan menggunakan perangkat lunak RapidMiner Studio, sebuah aplikasi data mining yang menyediakan antarmuka grafis dan operator analisis. Algoritma ini bekerja dengan membagi data ke dalam sejumlah klaster berdasarkan kedekatan jarak antara data dengan titik pusat (*centroid*), lalu memperbarui centroid secara berulang hingga konvergen[8].

Hasil klaster kemudian dianalisis dengan memperhatikan nilai centroid serta distribusi data di dalam tiap klaster. Atribut selisih pemakaian digunakan sebagai indikator utama, di mana nilai yang mendekati nol menunjukkan efisiensi penggunaan bahan. Analisis ini mengacu pada teori Material Usage Variance yang menegaskan bahwa selisih antara kebutuhan bahan dan realisasi pemakaian dapat mencerminkan tingkat efisiensi proses produksi [9]. Penelitian ini dilaksanakan di PT Loista Indonesia, Tangerang, Banten, pada periode April hingga Juni 2025.

3. Hasil dan Diskusi

Data yang telah dikumpulkan masuk ke tahap pra-pemrosesan untuk memastikan kebersihan dan kesiapan sebelum dianalisis lebih lanjut. Pada tahap ini dilakukan pembersihan data, pengisian nilai yang hilang, pemilihan atribut yang relevan, penghapusan atribut yang tidak diperlukan, serta normalisasi agar data lebih bermakna (Ana, 2024). Proses ini penting karena hasil analisis sangat bergantung pada kualitas data yang digunakan. Dalam penelitian ini pra-pemrosesan mencakup pula transformasi dan normalisasi data sebagai bagian dari persiapan sebelum penerapan algoritma K-Means.

Transformasi Data

Transformasi dilakukan dengan menambahkan dua atribut turunan yang bersifat kuantitatif. Atribut pertama adalah *Selisih Pemakaian*, yang dihitung dari Qty BoM dikurangi Qty Keluar, untuk menunjukkan perbedaan antara jumlah bahan yang direncanakan dengan jumlah aktual yang dipakai. Atribut kedua adalah *Persentase Efisiensi*, dihitung dengan rumus (Qty BoM / Qty Keluar) × 100, yang memberikan gambaran mengenai tingkat

DOI: https://doi.org/10.31004/riggs.v4i3.2606

efisiensi penggunaan bahan produksi. Semakin tinggi persentasenya, semakin dekat hasil penggunaan bahan dengan jumlah yang telah direncanakan. Penambahan atribut ini dimaksudkan agar informasi penting yang tersembunyi dalam data mentah dapat diekstraksi, sehingga proses pengelompokan dengan K-Means dapat didasarkan pada deviasi rencana serta efektivitas pemakaian bahan.

Normalisasi Data

Normalisasi merupakan langkah penting untuk menyeragamkan skala nilai pada atribut yang berbeda, karena K-Means sangat bergantung pada perhitungan jarak dengan metrik Euclidean. Tanpa normalisasi, atribut dengan nilai absolut besar bisa mendominasi hasil pembentukan cluster. Dalam penelitian ini digunakan metode *Min-Max Scaling* yang mengubah rentang nilai setiap atribut ke skala standar 0 hingga 1 [10]. Formula yang digunakan adalah

$$Xnorm = (X - Xmin) / (Xmax - Xmin)$$

Sebagai contoh, pada item "DEKKSON - HINGE DEKKSON ES SL 5 INCH" dengan Selisih Pemakaian sebesar 11.016 dan Efisiensi Pemakaian sebesar 69% (0,69), hasil normalisasi yang diperoleh masing-masing adalah 0.3733 dan 0.6900. Walaupun rentang asli efisiensi sudah mendekati 0–1, normalisasi tetap diterapkan untuk memastikan konsistensi data. Hasil akhir dari tahap ini adalah data terstandardisasi berupa Selisih Pemakaian dan Efisiensi Pemakaian yang menjadi input utama dalam proses clustering menggunakan K-Means.

Penerapan Algoritma K-Means

Algoritma K-Means Clustering merupakan salah satu metode pengelompokan data non-hirarki yang digunakan untuk mempartisi data ke dalam beberapa kelompok (cluster). Data dalam satu cluster memiliki kemiripan sifat, sedangkan dengan cluster lain berbeda [11]. Pengelompokan dilakukan dengan menghitung jarak antar data terhadap titik pusat kelompok (centroid).

Langkah awal penerapan algoritma ini adalah menentukan jumlah cluster (K). Dalam penelitian ini ditetapkan tiga cluster untuk mengelompokkan pola pemakaian bahan produksi, yaitu pola efisien sesuai rencana, pola kurang dari rencana dengan potensi sisa, dan pola lebih dari rencana dengan potensi boros. Selanjutnya, ditentukan centroid awal dari tiga data terpilih yang mewakili variasi nilai Selisih Pemakaian (Normalized) dan Efisiensi Pemakaian (Normalized).

Tabel 1. Centroid Awal

Cluster	Selisih Pemakaian (Normalized)	Efisiensi Pemakaian (Normalized)
C1	0.1867	1.0000
C2	0.2208	0.5000
C3	0.0000	0.0000

Langkah selanjutnya adalah menghitung jarak setiap data terhadap centroid menggunakan rumus *Euclidean Distance*, di mana jarak terdekat akan menentukan keanggotaan data dalam cluster tertentu. Untuk mempermudah ilustrasi, digunakan lima data dari Tabel III.4 sebagai contoh perhitungan. Data tersebut terdiri atas Data 1 yaitu BLOCKBOARD 18 X 1220 X 2440 MM UTY (BoM=3) dengan koordinat (0.1867, 1.0000), Data 2 yaitu CLOSET CABINET – STANDARD ROOM (BoM=12) dengan koordinat (0.3053, 0.4167), Data 3 yaitu HPL TACO TH 611 AL STEEL FOIL (BoM=4) dengan koordinat (0.2208, 0.5000), Data 4 yaitu BLOCKBOARD 18 X 1220 X 2440 MM UTY (BoM=15) dengan koordinat (0.4408, 0.0000), serta Data 5 yaitu DEKKSON – HINGE DEKKSON ES SL 6 INCH (0.0000, 0.0000). Hasil dari perhitungan jarak pada iterasi pertama ditunjukkan pada tabel berikut:

Tabel 2. Hasil Perhitungan Iterasi ke-1 (Contoh 5 Data)

Data	Koordinat (x,y)	Jarak ke C1	Jarak ke C2	Jarak ke C3	Cluster
1	(0.1867, 1.0000)	0.0000	0.5012	1.0173	C1
2	(0.3053, 0.4167)	0.6013	0.0875	0.5487	C2
3	(0.2208, 0.5000)	0.5012	0.0000	0.5437	C2
4	(0.4408, 0.0000)	1.0173	0.5487	0.4408	C3
5	(0.0000, 0.0000)	1.0173	0.5437	0.0000	C3

DOI: https://doi.org/10.31004/riggs.v4i3.2606

Setelah data dialokasikan ke cluster masing-masing, langkah berikutnya adalah menghitung centroid baru dari tiap cluster. Nilai centroid baru diperoleh dari rata-rata koordinat seluruh data dalam satu cluster.

Tabel 3. Centroid Baru Iterasi ke-2

Cluster	Selisih Pemakaian (Normalized)	Efisiensi Pemakaian (Normalized)
C1	0.1867	1.0000
C2	0.2630	0.4583
C3	0.2206	0.0000

Proses perhitungan jarak, pengelompokan data, dan pembaruan centroid terus diulang sampai posisi centroid stabil atau tidak terjadi lagi perpindahan data antar cluster. Namun, karena jumlah data dalam penelitian ini mencapai 150 item, perhitungan manual lengkap akan sangat panjang. Oleh sebab itu, perangkat lunak seperti RapidMiner digunakan untuk mempercepat analisis, sedangkan pemahaman langkah manual tetap penting agar hasil dapat diinterpretasikan secara tepat.

Penerapan dengan RapidMiner

Untuk menganalisis keseluruhan 150 item bahan produksi secara komprehensif dan efisien, sekaligus memvalidasi pemahaman dari perhitungan manual, digunakan perangkat lunak *RapidMiner*. Aplikasi ini merupakan platform analisis data yang menyediakan berbagai operator untuk mendukung tugas-tugas *data mining*, termasuk algoritma *K-Means Clustering*. Proses analisis dengan RapidMiner dilakukan secara sistematis, mulai dari pembacaan data hingga evaluasi hasil clustering.

Tahap awal dalam menentukan jumlah cluster optimal dilakukan dengan menguji nilai K secara bertahap, mulai dari K = 2 hingga K = 5, menggunakan algoritma K-Means Clustering. Untuk setiap nilai K, data yang telah dinormalisasi dikelompokkan, kemudian dihitung nilai *Davies Bouldin Index* (DBI) dengan memanfaatkan operator *Clustering Performance Evaluation* pada RapidMiner. Perlu dipahami bahwa *Davies Bouldin Index* (DBI) merupakan rasio antara jarak dalam cluster dengan jarak antar cluster, sehingga semakin rendah nilai DBI menunjukkan kualitas clustering yang semakin baik (Riady et al., 2024).

Dalam proses ini, digunakan operator *multiply* untuk menggandakan dataset agar memungkinkan dilakukannya proses clustering berulang. Selanjutnya, operator K-Means digunakan untuk mengelompokkan data ke dalam jumlah cluster yang telah ditentukan, kemudian dievaluasi menggunakan operator *Cluster Distance Performance*. Evaluasi ini bertujuan untuk menilai efektivitas hasil clustering dan menentukan jumlah cluster optimal berdasarkan nilai minimum DBI.

Berdasarkan hasil pengujian, diperoleh jumlah cluster optimal sebanyak tiga (K=3) dengan nilai DBI sebesar 0.234. Hal ini berarti data bahan produksi dikelompokkan menjadi tiga cluster, yang dipilih karena memiliki nilai DBI terkecil sehingga kualitas pengelompokan dianggap paling baik. Berikut merupakan hasil uji penentuan jumlah cluster optimal.

Tabel. 4. Nilai DBI

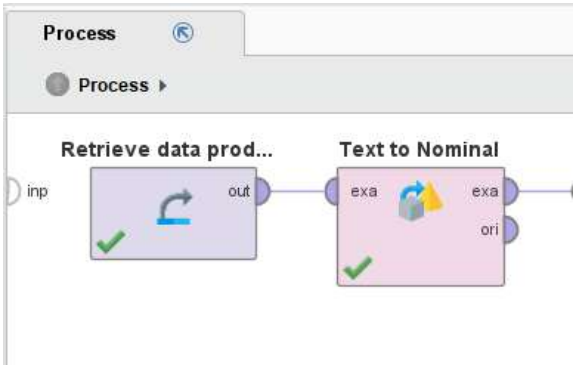
Cluster (K)	Davies Bouldin
2	0,377
3	0,234
4	0,458
5	0,356

Analisis K-Means Clustering

Tahap awal analisis K-Means Clustering adalah pra-pemrosesan data, yang sangat penting untuk menjembatani antara data mentah dengan kebutuhan teknis algoritma K-Means. Pada tahap ini digunakan beberapa operator di RapidMiner agar data dapat diproses dengan baik.

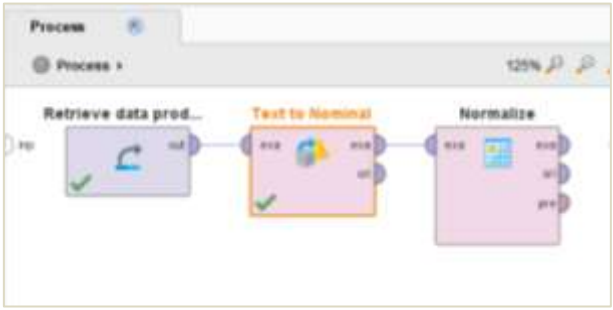
Salah satu operator yang digunakan adalah **Text To Nominal**. Fungsi utama operator ini adalah mengatasi masalah inkompatibilitas atribut *Nama Barang* yang masih berbentuk teks, karena algoritma K-Means hanya dapat melakukan perhitungan jarak pada data numerik. Operator Text to Nominal mengubah data teks bebas menjadi format nominal yang terstruktur sehingga dapat dikenali dan diolah lebih lanjut oleh RapidMiner (Gambar III.3).

DOI: https://doi.org/10.31004/riggs.v4i3.2606



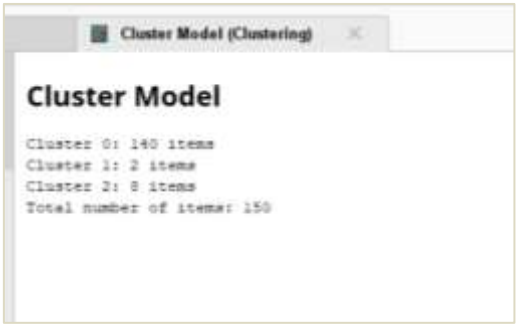
Gambar 1. Operator *Text To Nominal*

Tahap berikutnya adalah normalisasi data menggunakan operator Normalize Data. Normalisasi diperlukan agar semua data numerik berada dalam skala yang sama sehingga perhitungan jarak menjadi lebih adil. Metode normalisasi yang digunakan adalah Z-Transformation, yaitu mengurangi rata-rata data dari semua nilai kemudian membaginya dengan simpangan baku. Dengan metode ini, distribusi data akan memiliki rata-rata nol dan varians satu. Normalisasi ini sering disebut normalisasi statistik, yang bermanfaat karena tetap mempertahankan distribusi asli data dan tidak terlalu dipengaruhi oleh *outlier* [12].



Gambar 2. Operator *Normalize* Data

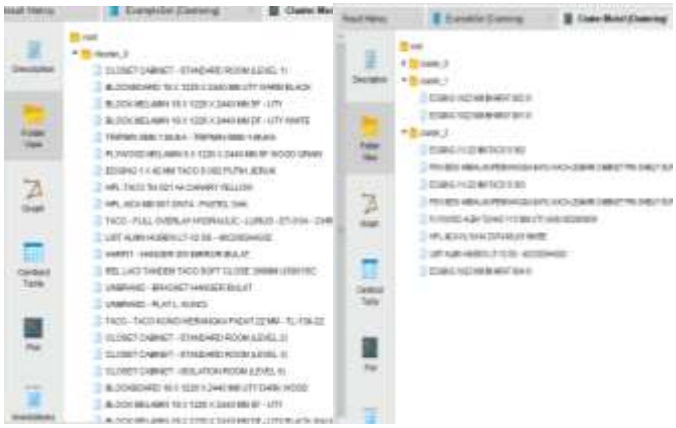
Setelah data siap, tahap inti analisis adalah penerapan algoritma **K-Means Clustering** menggunakan operator Clustering pada RapidMiner. Pada tahap ini, parameter algoritma dikonfigurasi, khususnya jumlah cluster (K) yang ditetapkan sebesar tiga sesuai hasil penentuan cluster optimal. Setelah eksekusi proses, RapidMiner menghasilkan sebuah **Cluster Model** yang merangkum hasil pengelompokan data.



Gambar 3. Cluster Model

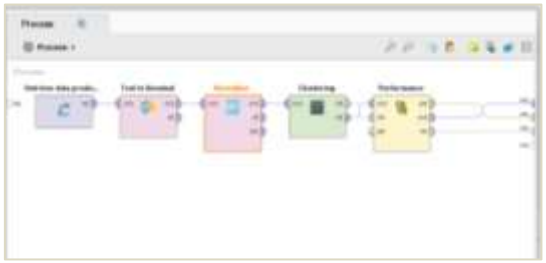
Dari hasil clustering, diperoleh bahwa dari total 150 item data, sebagian besar atau 140 item masuk ke dalam Cluster 0 (C0) yang diinterpretasikan sebagai pola pemakaian efisien atau sesuai rencana. Sementara itu, hanya 2 item masuk ke Cluster 1 (C1) yang menggambarkan pola pemakaian kurang dari rencana atau potensi sisa, dan 8 item masuk ke Cluster 2 (C2) yang menunjukkan pola pemakaian lebih dari rencana atau berpotensi boros. Ringkasan ini memperlihatkan bahwa model berhasil mengklasifikasikan seluruh data ke dalam tiga cluster yang terstruktur dengan baik.

DOI: https://doi.org/10.31004/riggs.v4i3.2606



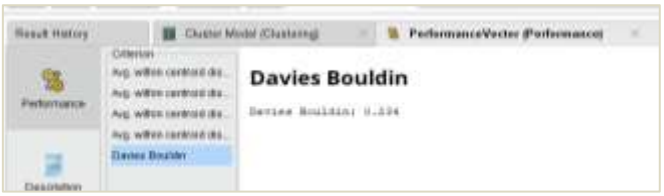
Gambar 4. Tampilan Folder View

Untuk menilai efektivitas pengelompokan, digunakan operator *Cluster Distance Performance*.



Gambar 5. Tampilan *Operator Performance*

Evaluasi dilakukan dengan mengukur nilai DBI. Semakin kecil nilai DBI, semakin baik kualitas klasterisasi karena cluster semakin berbeda satu sama lain. Dari hasil evaluasi diperoleh nilai DBI sebesar **0.234**, yang menunjukkan kualitas clustering sangat baik [13], sebagaimana ditunjukkan pada Gambar 6.



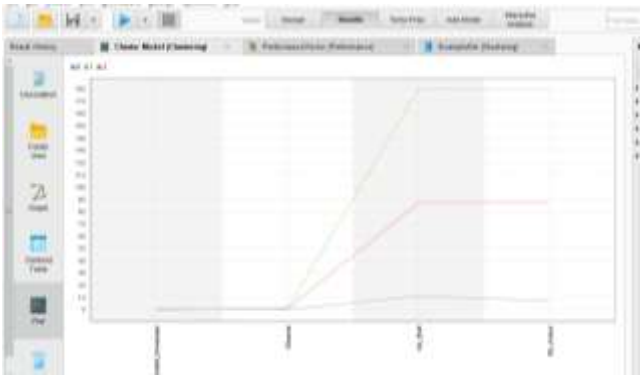
Gambar 6. Tampilan *Performance Vector*

Selain itu, hasil clustering juga ditampilkan dalam bentuk *Centroid Table*, yang menggambarkan posisi rata-rata setiap cluster pada tiap atribut. Centroid merupakan titik pusat yang mewakili karakteristik data dalam sebuah cluster [14]. Pada tabel centroid, atribut *Selisih Pemakaian* untuk Cluster 0 memiliki nilai 0.029 yang berarti pemakaian bahan hampir sesuai perencanaan, sedangkan Cluster 1 dan Cluster 2 sama-sama memiliki nilai -0.405 yang menunjukkan pemakaian kurang dari rencana dan berpotensi menimbulkan sisa bahan. Untuk atribut *Efisiensi*, Cluster 0 memiliki nilai -0.056 yang mencerminkan kondisi efisiensi relatif normal, sedangkan Cluster 1 dan Cluster 2 memiliki nilai 0.781 yang menunjukkan efisiensi tinggi. Pada atribut *Qty BoM*, Cluster 1 memiliki nilai tertinggi (180), diikuti Cluster 2 (87), dan Cluster 0 terendah (10.888). Pola yang sama juga terlihat pada atribut *Qty Keluar*, di mana Cluster 1 tertinggi (180), Cluster 2 menengah (87), dan Cluster 0 terendah (6.758). Secara umum, untuk atribut *Selisih Pemakaian* dan *Efisiensi*, ketiga cluster memiliki nilai yang cenderung mendekati nol, sehingga variansinya relatif kecil[15].

DOI: https://doi.org/10.31004/riggs.v4i3.2606

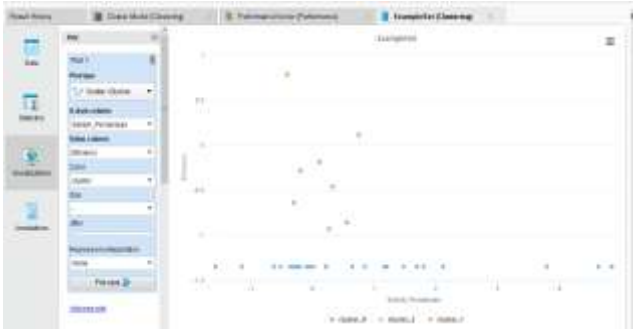
Gambar 7. *Centroid Table*

Hasil centroid ini kemudian divisualisasikan dalam bentuk Plot. Visualisasi menggunakan *scatter plot* atau *bubble plot*, dengan sumbu X menunjukkan atribut *Selisih Pemakaian* dan sumbu Y menunjukkan atribut *Efisiensi*. Warna pada titik-titik menggambarkan label cluster, yaitu biru untuk Cluster 0, oranye untuk Cluster 1, dan hijau untuk Cluster 2. Dari visualisasi terlihat bahwa Cluster 0 merupakan klaster terbesar dengan titik-titik yang tersebar luas di sekitar titik nol, menunjukkan pola pemakaian efisien sesuai rencana. Cluster 1 hanya terdiri dari satu titik data dengan selisih pemakaian -0.405 dan efisiensi 0.781, yang merepresentasikan pola pemakaian kurang dari rencana atau potensi sisa bahan. Sementara itu, Cluster 2 juga terdiri dari satu titik dengan nilai centroid identik dengan Cluster 1, yang menunjukkan pola pemakaian lebih dari rencana atau boros. Pemisahan Cluster 1 dan Cluster 2 terjadi karena algoritma memandang data tersebut sebagai titik dengan karakteristik khusus, yang mungkin dipengaruhi oleh kondisi numerik pada inisialisasi centroid atau keberadaan *outlier*.



Gambar 8. Plot

Tahap akhir adalah visualisasi hasil clustering, yang semakin memperjelas perbedaan karakteristik antar cluster. Dengan demikian, analisis K-Means Clustering menggunakan RapidMiner mampu mengidentifikasi pola penggunaan bahan produksi menjadi tiga kelompok utama, yaitu pemakaian efisien, pemakaian kurang dari rencana/potensi sisa, dan pemakaian lebih dari rencana/potensi boros.



Gambar 9. *Visualizations*

DOI: https://doi.org/10.31004/riggs.v4i3.2606

4. Kesimpulan

Berdasarkan hasil analisis yang dilakukan, diperoleh nilai Davies Bouldin Index (DBI) dengan variasi jumlah cluster. Nilai DBI terkecil berada pada jumlah cluster tiga (K=3) dengan skor 0,234, sehingga dapat disimpulkan bahwa pembentukan tiga cluster merupakan struktur yang paling optimal dibandingkan jumlah cluster lainnya. Hal ini menunjukkan bahwa pemisahan antar cluster pada jumlah tiga lebih jelas dan homogenitas dalam cluster lebih baik, sehingga model pengelompokan yang dihasilkan lebih representatif terhadap data yang dianalisis. Temuan ini dapat menjadi dasar dalam penerapan metode clustering pada kasus serupa yang memerlukan pembagian kelompok data secara optimal. Aplikasi dari hasil ini berpotensi membantu pengambilan keputusan yang lebih tepat berdasarkan pola kelompok yang terbentuk. Namun, penelitian selanjutnya dapat mempertimbangkan penggunaan indeks validasi lain atau kombinasi metode evaluasi untuk memperoleh hasil yang lebih komprehensif, sehingga kualitas pengelompokan dapat semakin terjamin.

Referensi

[1] A. P. Bagustio, A. I. Purnamasari, and I. Ali, "Analisis data penjualan menggunakan algoritma K-Means clustering pada Toko Kecantikan Putri," *PROSISKO J. Pengemb. Ris. dan Obs. Sist. Komput.*, vol. 11, no. 2, pp. 159–167, 2024, doi: 10.30656/prosisko.v11i2.7928.

[2] R. Mustakim, D. Ruhimat, S. Suhada, and R. Yulistria, "Analisis segmentasi frekuensi transaksi ATM dengan K-Means clustering pada Bank BJB Kantor Cabang Tasikmalaya," *J. Sist. Inf.*, vol. 15, no. April, pp. 60–72, 2025, [Online]. Available: https://repository.bsi.ac.id/repo/files/425537/download/JURNAL.pdf

[3] N. Arifin, R. H. Irawan, and I. N. Farida, *Algoritma K-Means untuk memprediksi stok bahan baku produksi*. Universitas Nusantara PGRI Kediri, 2022.

[4] V. Alvianatinova, I. Ali, N. Rahaningsih, and A. Bahtiar, "Penerapan Algoritma K-Means Clustering dalam Pengelompokan Data Penjualan Supermarket Berdasarkan Cabang (Branch," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1529–1535, 2024, doi: 10.36040/jati.v8i2.8993.

[5] M. Marsono, D. Saripurna, and M. Zunaidi, "Analisis data mining pada strategi penjualan produk PT Aquasolve Sanaria dengan menggunakan metode K-Means clustering." 2021. doi: 10.53513/jsk.v4i1.60.

[6] K. Febianda, D. E. Ratnawati, and B. Rahayudi, "Analisis pengelompokkan penjualan rattan furniture pada PT. Hymsa Indotraco berbasis algoritma K-Means clustering," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 6, pp. 1882–1887, 2020, [Online]. Available: http://j-ptiik.ub.ac.id

[7] I. G. I. Sudipa, "Representation knowledge data mining." 2022. [Online]. Available: https://repository.bsi.ac.id/repo/files/375053/download/Buku-Data-Mining.pdf

[8] A. R. Mulyawan, W. Gata, and S. Alfarizi, "Marketing maps pada Lembaga Amil Zakat menggunakan algoritma clustering dan association rules," *Sistemasi*, vol. 9, no. 1, pp. 36–46, 2022, doi: 10.32520/stmsi.v9i1.572.

[9] A. Haladu, "Material usage variance and decision-making: An analysis of block manufacturing industries in." Kano Metropolis, 2023.

[10] N. Nurjanah, N. Suarna, and W. Prihartono, "Implementasi K-Means clustering untuk mengelompokan tingkat pengangguran," *J. Teknol. Inf.*, vol. 8, no. 2, pp. 2462–2468, 2024.

[11] A. N. Ayu, N. B. Nugroho, and M. Syaifuddin, "Penerapan data mining untuk menentukan persediaan barang pada PT," *Deli Food menggunakan Metod. K-Means. J. Sist. Inf.*, vol. 4, no. 7, 2021, [Online]. Available: https://ojs.trigunadharma.ac.id/

[12] RapidMiner, "RapidMiner 9 operator reference manual." 2021. [Online]. Available: https://docs.rapidminer.com/latest/studio/operators/rapidminer-studio-operator-reference.pdf

[13] E. S. Ana, "Analisa tingkat penjualan makanan dan minuman pada Restoran Sumoboo dengan klasterisasi menggunakan algoritma K-Means," *Insa. J. Inf. Syst. Manag. Innov.*, vol. 4, no. 1, pp. 45–54, 2024, doi: https://repository.bsi.ac.id/repo/51352/.

[14] L. Anggilia, "Analisa clustering penyebaran penyakit tuberkulosis menggunakan algoritma K-Means di Provinsi Sumatera Selatan." 2024. [Online]. Available: https://repository.bsi.ac.id/repo/53686/

[15] I. Wijaya, "Perbandingan tools data mining untuk pengelompokan siswa putus sekolah dengan algoritma K-Means." 2024. [Online]. Available: https://repository.bsi.ac.id/repo/52950/

DOI: https://doi.org/10.31004/riggs.v4i3.2606