CLUSTERING PENGGUNAAN BAHAN BAKU MENGGUNAKAN K-MEANS DI DUA NAGA CORPORATION

Karisna Wahyuningtyas[1], Dwi Hartanti[2], Hanifah Permatasari[3]

[123]Program Studi Sistem Informasi, Ilmu Komputer, Universitas Duta Bangsa Surakarta

Jln. Bhayangkara No.55, Tipes, Kec.Serengan, Kota Surakarta, Jawa Tengah 57154
[1] 220101021@mhs.udb.ac.id, [2] dwihartanti@udb.ac.id, [3] hanifah_permatasari@udb.ac.id

Abstract

Inventory management at Dua Naga Corporation is currently performed manually, creating the potential for stockout and overstock situations due to difficulties in identifying raw material usage patterns. This study aims to automatically cluster raw materials using the K-Means clustering algorithm based on the CRISP-DM framework. The dataset consists of 62,922 raw material usage transactions representing 615 unique raw material types. The clustering process was conducted based on raw material usage characteristics to identify groups with different levels of consumption. The results indicate that the optimal number of clusters is three, representing high-, medium-, and low-usage categories. Model evaluation demonstrated excellent clustering performance, with a Silhouette Coefficient Index (SCI) of 0.928 and a Davies–Bouldin Index (DBI) of 0.735. The clustering model was integrated into a Streamlit-based web application to support real-time inventory monitoring. The implementation of this model assists the company in prioritizing inventory control, improving stock management efficiency, and supporting more accurate decision-making for raw material planning.

Keywords : *Clustering, Raw Materials, K-means, Silhouette Coefficient, CRISP-DM*

Abstrak

Manajemen persediaan di Dua Naga Corporation saat ini masih dilakukan secara manual sehingga berpotensi menimbulkan masalah *stockout* dan *overstock* akibat sulitnya mengidentifikasi pola penggunaan bahan baku. Penelitian ini bertujuan untuk mengelompokkan bahan baku secara otomatis menggunakan algoritma K-Means Clustering dengan pendekatan CRISP-DM. Data yang digunakan terdiri dari 62.922 transaksi penggunaan bahan baku yang mewakili 615 jenis bahan baku. Proses clustering dilakukan berdasarkan karakteristik penggunaan bahan baku untuk mengidentifikasi kelompok dengan tingkat penggunaan yang berbeda. Hasil penelitian menunjukkan bahwa jumlah klaster optimal adalah tiga klaster, yang merepresentasikan kategori penggunaan tinggi, sedang, dan rendah. Evaluasi model menunjukkan performa yang sangat baik dengan nilai *Silhouette Coefficient Index* (SCI) sebesar 0,928 dan *Davies-Bouldin Index* (DBI) sebesar 0,735. Hasil clustering kemudian diimplementasikan ke dalam aplikasi berbasis web menggunakan Streamlit untuk mendukung pemantauan persediaan secara *real-time*. Penerapan model ini membantu perusahaan dalam menentukan prioritas pengendalian persediaan, meningkatkan efisiensi pengelolaan stok, serta mendukung pengambilan keputusan yang lebih tepat dalam perencanaan kebutuhan bahan baku.

Kata kunci : *Clustering, Bahan Baku, K-Means, Silhouette Coefficient, CRISP-DM*

1. PENDAHULUAN

Manajemen persediaan bahan baku memegang peranan strategis dalam menjamin keberlangsungan operasional, menjaga kualitas output, serta mengoptimalkan efisiensi biaya pada sektor

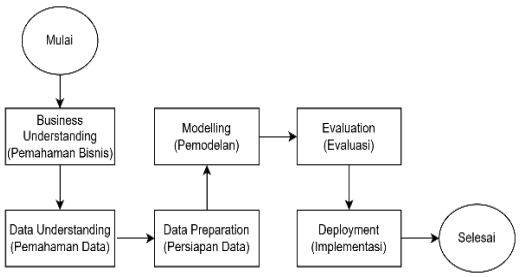
industri manufaktur [1]. Bagi sebuah holding company seperti Dua Naga Corporation yang mengelola empat unit usaha dengan karakteristik produk dan kebutuhan material yang beragam, kompleksitas pola penggunaan bahan baku menjadi tantangan tersendiri dalam menjaga keseimbangan stok. Ketidakakuratan dalam pengelolaan ini berisiko memicu berbagai masalah operasional, mulai dari kekurangan stok (*stockout*) yang menghambat produksi, penumpukan bahan (*overstock*) yang jarang digunakan, hingga meningkatnya biaya penyimpanan [2].

Saat ini, sistem pengendalian persediaan di Dua Naga Corporation masih dilakukan secara manual sehingga perusahaan kesulitan mengidentifikasi bahan baku yang sering, sedang, maupun jarang digunakan dari data historis yang ada. Berbagai penelitian terdahulu telah dilakukan untuk memberikan solusi atas permasalahan manajemen persediaan melalui pengelompokan data. Penelitian oleh Rizkyandri dkk. (2023) mengimplementasikan metode K-Means Clustering untuk menentukan persediaan barang pada toko *retail*, yang menghasilkan tiga klaster barang yaitu fast moving, medium moving, dan slow moving untuk membantu prioritas pengadaan (*restock*) [3]. Selanjutnya, Habibah (2024) menganalisis pengendalian persediaan bahan baku menggunakan metode *ABC Analysis* terhadap penjadwalan produksi, yang mengelompokkan bahan baku ke dalam kategori A, B, dan C berdasarkan tingkat kepentingan penggunaannya. Penelitian lain oleh Melisa dkk. (2025) menerapkan algoritma K-Means Clustering untuk optimalisasi persediaan berdasarkan pola penjualan, yang berhasil mengkategorikan produk menjadi sangat populer, populer, dan kurang populer guna meminimalkan risiko *overstock* maupun *stockout* [4].

Meskipun metode pengelompokan terbukti efektif, sebagian besar penelitian sebelumnya berfokus pada barang jadi (*retail*) atau menggunakan metode klasifikasi konvensional yang hanya bertumpu pada variabel tingkat kepentingan atau penjualan akhir. Berbeda dengan kajian tersebut, penelitian ini berfokus pada pola penggunaan bahan baku mentah dalam lingkup industri manufaktur yang lebih kompleks dengan mengekstraksi lima variabel utama, yaitu volume penggunaan, frekuensi transaksi, rata-rata penggunaan, maksimum penggunaan, dan standar deviasi penggunaan [5]. Untuk mendukung hal ini, penelitian ini mengembangkan sistem klastering otomatis menggunakan algoritma K-Means yang dibangun menggunakan bahasa pemrograman Python dan *framework* Streamlit. Sistem antarmuka berbasis web ini dirancang sebagai pendukung keputusan (Decision Support System) yang dapat memproses data mutasi stok secara otomatis dan memvisualisasikan hasil segmentasi, sehingga strategi pengendalian persediaan di Dua Naga Corporation dapat berjalan lebih cepat, akurat, dan efisien [6][7].

2. METODOLOGI PENELITIAN

Cross-Industry Standard Process for Data Mining (CRISP-DM) adalah sebuah kerangka kerja sistematis digunakan sebagai standar industri dalam proses pengolahan data untuk mengungkap pola tersembunyi yang relevan dengan kebutuhan bisnis. Dalam konteks analisis klastering pola penggunaan bahan baku di Dua Naga Corporation dengan penerapan algoritma K-Means, proses CRISP-DM diawali pemahaman terhadap tujuan operasional perusahaan terkait pengelolaan persediaan. Pemodelan CRISP-DM memiliki enam tahapan utama, yaitu Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, serta tahap Deployment [8]. Pada tahapan data preparation dilakukan pembersihan data mentah (*data cleaning*), agregasi data transaksi keluar (out), pembentukan lima variabel utama, dan normalisasi skala data menggunakan Standard Scaler. Pada tahapan modelling dilakukan proses pengelompokan data numerik bahan baku ke beberapa kategori tingkat penggunaan menggunakan algoritma K-Means. Berikut gambaran mengenai setiap tahap dalam metodologi secara rinci.



Gambar 1. Tahapan Penelitian

2.1. Business Understading (Pemahaman Bisnis)

Pada tahap awal berfokus pada pemahaman terhadap tujuan operasional serta identifikasi masalah utama perusahaan. Permasalahan di Dua Naga Corporation adalah tidak tersedianya informasi yang terstruktur mengenai pola penggunaan bahan baku dari waktu ke waktu. Oleh karena itu, tujuan pemodelan ini untuk menghasilkan segmentasi bahan baku yang akurat ke dalam beberapa kategori tingkat penggunaan untuk mendukung keputusan pengadaan dan manajemen persediaan yang lebih efisien

2.2. Data Understanding (Pemahaman Data)

Pada tahap ini, pengumpulan data sekunder dilakukan dengan mengekstraksi data historis mutasi persediaan bahan baku perusahaan periode Desember 2024 hingga November 2025 dalam format file *Spreadsheet* (CSV). Eksplorasi data awal dilakukan melalui Jupyter Notebook untuk memahami struktur data mentah, mengidentifikasi jumlah bahan baku, mengidentifikasikan tipe atribut data yaitu tanggal, kode bahan baku, jenis transaksi, dan kuantitas (gram), serta menganalisis nilai kosong (*missing values*) atau format yang tidak konsisten.

2.3. Data Preparation (Persiapan Data)

Tahap persiapan data merupakan langkah krusial untuk memastikan kualitas data sebelum masuk ke tahap pemodelan [10]. Data mentah melalui proses pembersihan data dengan mengambil data khusus transaksi pengeluaran barang (out), selanjutnya dilakukan agregasi berdasarkan atribut kode bahan baku untuk mentransformasi data transaksi menjadi 615 entitas unik. Tahap ini dilakukan pembentukan fitur (*feature engineering*) lima variabel utama yang menjadi kriteria pengelompokan, yaitu volume penggunaan, frekuensi transaksi, rata-rata penggunaan, maksimum penggunaan, dan standar deviasi penggunaan. Agar perhitungan jarak pada algoritma tidak bias, kelima variabel tersebut dinormalisasi menggunakan Standard Scaler [11].

2.4. Modelling (Pemodelan)

Pada tahap pemodelan diimplementasikan menggunakan algoritma K-Means Clustering dengan penghitungan metrik *Euclidean Distance* [12]. Untuk menentukan kedekatan antar titik data secara geometris

Rumus *Euclidean Distance:*

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Algoritma beroperasi secara iteratif untuk meminimalkan fungsi objektif atau jarak *intra-cluster* guna mendapatkan segmentasi data yang objektif sebagai berikut:

$$J = \sum_{i=1}^{K} \sum_{x_j \in c_i} ||x_j - \mu_i||^2$$

Keterangan:
K: Jumlah klaster
c_i: Klaster ke-i
x_j: Titik data dalam klaster
μ_i: Centroid (pusat) klaster ke-i

Penentuan jumlah klaster optimal (k) divalidasi secara matematis melalui pengujian Elbow dan Silhouette Coefficient Index (SCI). Melalui mekanisme penghitungan berulang, bahan baku dikelompokkan ke dalam tiga segmen kategori (k=3), yaitu Kelas A (penggunaan tinggi), Kelas B (penggunaan sedang), dan Kelas C (penggunaan rendah), setiap klaster merepresentasikan tingkat pola penggunaan yang tinggi. Proses ini bertujuan menghasilkan segmentasi material yang akurat guna mendukung efisiensi strategi pengendalian persediaan perusahaan.

2.5. Evaluation (Evaluasi)

Tahap evaluasi memastikan validitas dan reliabilitas hasil pengelompokan, kualitas model dievaluasi secara statistik menggunakan metrik Silhouette Coefficient Index (SCI) dan Davies-Bouldin Index (DBI). Pengukuran ini bertujuan untuk memastikan kepadatan di dalam klaster dan pemisahan antar-klaster terbentuk secara jelas, konsisten, dan tidak tumpang tindih [13].

2.6. Deployment (Implementasi)

Tahap akhir melibatkan implementasi model algoritma yang divalidasi ke dalam prototipe sistem antarmuka berbasis web memanfaatkan *framework* Streamlit dan Bahasa pemrograman Python. Sistem ini dirancang untuk memproses unggahan file data CSV secara otomatis, menjalankan klastering, dan menampilkan *output* berupa visualisasi grafik serta tabel kategori bahan baku secara *real-time* sebagai dasar pengambilan keputusan manajemen operasional [18] [19].

3. HASIL DAN PEMBAHASAN

Bagian ini menguraikan hasil eksperimen pengolahan data untuk segmentasi pola penggunaan bahan baku pada Dua Naga Corporation menggunakan kerangka kerja CRISP-DM.

3.1. Business Understading (Pemahaman Bisnis)

Tahap awal ini difokuskan untuk memahami kebutuhan perusahaan dalam memperbaiki pengelolaan stok dengan pengelompokan bahan baku yang terukur. Berdasarkan pengamatan, ditemukan bahwa pengelolaan 615 jenis bahan baku unik mengalami hambatan operasional karena pola pemakaiannya yang sering berubah-ubah tidak menentu. Hal ini menyebabkan sistem pencatatan stok secara manual sudah tidak efektif lagi untuk menjaga keseimbangan barang di gudang. Oleh karena itu, penelitian ini bertujuan untuk membangun sistem pembagian otomatis untuk mengelompokkan bahan baku ke dalam tiga tingkat prioritas yaitu Kelas A (Vital), Kelas B (Esensial), dan Kelas C (Umum).

3.2. Data Understanding (Pemahaman Data)

Tahap ke dua yaitu pengumpulan data sekunder dilakukan dengan mengekstraksi arsip mutasi persediaan perusahaan. Melalui eksplorasi pada Jupyter Notebook, teridentifikasi *dataset* mentah dengan total 68.605 baris transaksi yang mencakup atribut tanggal, kode bahan baku, jenis transaksi, dan kuantitas (gram).

```
=== Data Setelah Pembersihan ===
```

	Tanggal	Kode	Jenis	Qty
1	25-Aug-2025	CMJ0001CMJ	IN	79100.0
2	25-Aug-2025	OJR0002OJR	IN	54350.0
3	25-Aug-2025	JXE0003JXE	IN	227750.0
4	25-Aug-2025	RLT0004RLT	IN	54550.0
5	25-Aug-2025	VDE0005VDE	IN	13250.0
...	...	...	...	...
68601	23-Aug-2025	RPE1119RPE	OUT	250.0
68602	23-Aug-2025	HOU0261HOU	OUT	99360.0
68603	23-Aug-2025	CTR1289CTR	OUT	350.0
68604	23-Aug-2025	VDE0005VDE	OUT	220000.0
68605	23-Aug-2025	RLD0611RLD	OUT	1205.0

```
Total data setelah pembersihan: 68,605 baris
```

Gambar 2. Pengumpulan Data

3.3. Data Preparation (Persiapan Data)

Tahap persiapan data merupakan fase krusial untuk mentransformasi dataset mentah menjadi format yang optimal bagi algoritma K-Means Clustering. Proses ini meliputi pembersihan, agregasi data, hingga standarisasi nilai agar hasil pengelompokan akurat. Adapun langkah-langkah detailnya adalah sebagai berikut:

3.3.1 Penyaringan Transaksi (Data Filtering)

Pada tahap penyaringan data ini terdapat total keseluruhan data mentah sebanyak 68.605 baris, peneliti melakukan ekstraksi khusus pada jenis transaksi pengeluaran barang (OUT). Langkah ini bertujuan agar analisis berfokus pada pola penggunaan bahan baku di perusahaan. Setelah melalui tahap penyaringan dan pembersihan nilai kosong (*NaN*) pada kolom kuantitas, diperoleh sebanyak 62.922 baris transaksi siap diolah.

=== Penyaringan Data Transaksi Setelah Pembersihan (Transaksi OUT) ===

	Tanggal	Kode Bahan Baku	Jenis Transaksi	Kuantitas
1	26-Aug-2025	NWC0192NWC	OUT	1200.0
2	26-Aug-2025	OCE0147OCE	OUT	5100.0
3	26-Aug-2025	RCE0078RCE	OUT	900.0
4	26-Aug-2025	UCE0235UCE	OUT	1350.0
5	26-Aug-2025	CSF0861CSF	OUT	2160.0
...	...	...	...	...
62917	23-Aug-2025	RPE1119RPE	OUT	250.0
62918	23-Aug-2025	HOU0261HOU	OUT	99360.0
62919	23-Aug-2025	CTR1289CTR	OUT	350.0
62920	23-Aug-2025	VDE0005VDE	OUT	220000.0
62921	23-Aug-2025	RLD0611RLD	OUT	1205.0

Total data setelah pembersihan: 62,921 baris

Gambar 3. Penyaringan Data

3.3.2 Agregasi Data Pembentukan Fitur (Feature Engineering)

Pada tahap ini, data transaksi harian kemudian diringkas (diagregasi) berdasarkan atribut kode bahan baku untuk mengubah unit analisis dari "transaksi" menjadi "entitas bahan baku". Proses ini menghasilkan *dataset* baru dengan 615 entitas bahan baku unik yang direpresentasikan oleh lima variabel kriteria utama:

A. Volume total: akumulasi total kuantitas penggunaan bahan baku
B. Frequency count: intensitas atau jumlah frekuensi transaksi penggunaan
C. Rata_rata_penggunaan: rata rata jumlah penggunaan dalam setiap transaksi
D. Maksimum penggunaan: nilai penggunaan tertinggi yang pernah tercatat
E. Standar deviasi penggunaan: tingkat volatilitas atau keragaman pola penggunaan

```
--- Data Agregasi Selesai ---
Total 615 Bahan Baku unik sedia untuk clustering.
      KODE BB  Volume_Total  Frequency_Count  Rata_Rata_Penggunaan
0  AWO0051AWO       3917.30               39            100.443590
1  BGP0241BGP        699.75                4            174.937500
2  BQZ0159BQZ          0.00                1              0.000000
3  CCI0826CCI        154.20               64              2.409375
4  CDO1632CDO       2206.00                3            735.333333

   Maksimum_Penggunaan  Standar_Deviasi_Penggunaan
0               340.00                   76.984451
1               183.75                   12.748162
2                 0.00                    0.000000
3                68.00                    9.096898
4              2000.00                 1095.237569
```

Gambar 4. Agregasi Data

3.3.3 Normalisasi Data (Data Scaling)

Tahap akhir dalam persiapan data adalah melakukan normalisasi data untuk menyeimbangkan pengaruh lima kriteria utama karena adanya perbedaan besaran nilai yang mencolok antara variabel seperti Volume Total dan Frequency Count. Algoritma K-Means bekerja berdasarkan perhitungan jarak sehingga menerapkan metode Standard Scaler agar setiap variabel memiliki skala yang seragam dengan rata-rata nol dan standar deviasi satu. Langkah ini sangat penting untuk memastikan seluruh variabel kriteria dari 615 entitas bahan baku memberikan kontribusi yang seimbang dalam pemodelan, sehingga hasil pengelompokan bersifat objektif dan tidak didominasi variabel dengan angka yang besar saja.

```
=== Penskalaan Data (StandardScaler) ===
Data berjaya diskalakan.
   Volume_Total  Frequency_Count  Rata_Rata_Penggunaan  Maksimum_Penggunaan
0     -0.124092        -0.208547             -0.188000            -0.128461
1     -0.124406        -0.323838             -0.185757            -0.128811
2     -0.124474        -0.333720             -0.191026            -0.129223
3     -0.124459        -0.126196             -0.190953            -0.129071
4     -0.124259        -0.327132             -0.168879            -0.124737

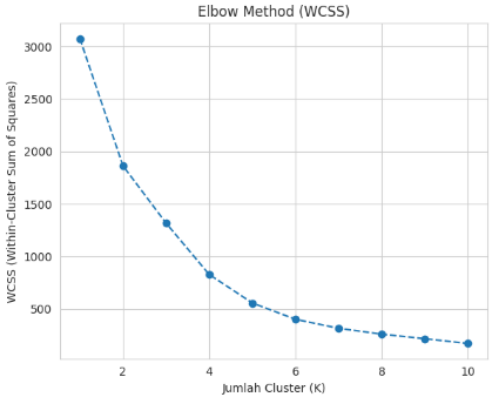
   Standar_Deviasi_Penggunaan
0                   -0.171262
1                   -0.172935
2                   -0.173266
3                   -0.173030
4                   -0.144749
```

Gambar 5. Normalisasi Data

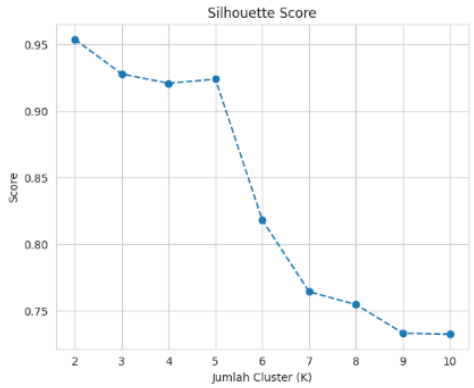
3.4. Modelling (Pemodelan)

Pada tahap modelling, mengimplementasikan algoritma K-Means Clustering untuk mengelompokkan 615 entitas bahan baku yang telah dinormalisasi. Sebelum mengelompokkan data final, dapat melakukan serangkaian pengujian untuk mencari jumlah klaster optimal (k) dengan rentang pengujian dari k=1 hingga k=10. Penentuan jumlah klaster terbaik divalidasi menggunakan perpaduan dua metode evaluasi, yaitu metode Elbow (berdasarkan nilai jarak WCSS) dan Silhouette Score.

```
=== Mencari Nilai K Optimal (1 sampai 10) ===
```



Gambar 6. Pemodelan Elbow method



Gambar 7. Pemodelan Silhouette Score

Berdasarkan analisis visual pada grafik kurva Elbow, terlihat penurunan nilai jarak antar data mulai melandai secara signifikan (membentuk pola siku) pada titik k=3. Hal ini juga didukung grafik Silhouette Score yang menunjukkan titik tersebut memiliki nilai yang paling stabil dan memadai untuk pembagian kelompok. Oleh karena itu, nilai k=3 ditetapkan sebagai parameter klaster yang paling optimal. Pemilihan parameter ini juga sangat sesuai yaitu membagi manajemen inventaris ke dalam tiga prioritas Kelas A (Vital), Kelas B (Esensial), dan Kelas C (Umum).

3.5. Evaluation (Evaluasi)

Tahap evaluasi mengukur kualitas pengelompokan algoritma K-Means melalui metrik validitas dan analisis klaster. Berdasarkan pengujian terhadap 615 entitas bahan baku, diperoleh hasil sebagai berikut:

3.5.1 Analisis Metrik Validitas Klaster

Model menghasilkan nilai Inertia (SSE) sebesar 1315.957, Silhouette Coefficient (SCI) sebesar 0.928 (menunjukkan alokasi objek yang sang at akurat), dan Davies-Bouldin Index (DBI) sebesar 0.735 yang membuktikan pemisahan antar-klaster sangat baik.

```
=== Penerapan K-Means dengan K=3 ===

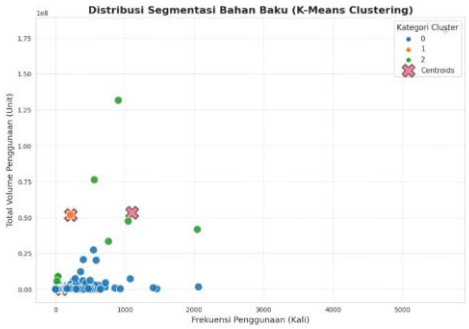
Metrik Evaluasi Clustering:
- Sum of Squared Error (SSE) / Inertia: 1315.957
- Silhouette Coefficient Index (SCI): 0.928
- Davies Bouldin Index (DBI): 0.735
```

Gambar 8. Metrik Evaluasi Clustering

3.5.2 Distribusi Klaster

Setelah validitas model teruji, dilakukan interpretasi terhadap hasil segmentasi untuk menentukan kategori inventaris menjadi tiga tingkat prioritas:

1. Kelas A (Vital - Cluster 2): Terdiri dari 10 item strategis dengan rata-rata volume penggunaan tertinggi (~53,4 juta unit).
2. Kelas B (Esensial - Cluster 1): Mencakup 1 item dengan karakteristik penggunaan tingkat menengah (~51,6 juta unit).
3. Kelas C (Umum - Cluster 0): Kelompok mayoritas berjumlah 604 item dengan volume penggunaan terendah namun stabil.



Gambar 9. Visualisasi Segmentasi Bahan Baku

3.5.3 Ekstraksi Data Bahan Baku

Tahap ini mentransformasikan hasil daftar identitas bahan baku yang siap digunakan secara praktis oleh perusahaan. Model berhasil memetakan setiap Kode BB ke dalam kategori prioritasnya secara spesifik, kemudian diekspor ke dalam format CSV. Dokumen ini menjadi luaran nyata yang memudahkan tim gudang Dua Naga Corporation dalam mengidentifikasi serta membedakan perlakuan stok antara kategori Vital, Esensial, dan Umum secara cepat dan akurat.

```
==================================================
--- KELAS A (VITAL) (Total: 10 item) ---
==================================================
   KODE BB  Volume_Total  Frequency_Count  Rata_Rata_Penggunaan  Standar_Deviasi_Penggunaan
CSF0861CSF  180474302.63             5623          32095.732283                36300.253599
OEP0145OEP  131586585.41              906         145239.056744                255321.545679
ITH8165ITH   76195090.00              560         136062.660714                208185.485641
UCE0235UCE   47417946.05             1049          45202.999094                115742.994487
INP0008INP   41686615.90             2045          20384.653252                 65005.980652
EUV0075EUV   33320402.70              765          43556.081961                 99410.867271
ORQ1473ORQ    8895238.00               36         247089.944444                208738.901801
OEP1358OEP    8512670.00               35         243219.142857                292657.089226
RIN1657RIN    5591340.00               22         254151.818182                110149.947858
PCE8146PCE     500000.00                1         500000.000000                     0.000000

[INFO] Daftar lengkap 10 item telah disimpan ke file: 'Daftar_BB_KELAS_A_VITAL.csv'

==================================================
--- KELAS B (ESENSIAL) (Total: 1 item) ---
==================================================
   KODE BB  Volume_Total  Frequency_Count  Rata_Rata_Penggunaan  Standar_Deviasi_Penggunaan
ITH0089ITH    51662354.0              216         239177.564815                774245.153334

[INFO] Daftar lengkap 1 item telah disimpan ke file: 'Daftar_BB_KELAS_B_ESENSIAL.csv'

==================================================
--- KELAS C (UMUM) (Total: 604 item) ---
==================================================
   KODE BB  Volume_Total  Frequency_Count  Rata_Rata_Penggunaan  Standar_Deviasi_Penggunaan
RIN8238RIN   27364058.00              549          49843.457195                76620.875381
RIN0009RIN   20608954.60              402          51266.056219                66574.318517
CSF0373CSF   20218255.40              587          34443.365247                26614.578466
UCE8166UCE   12185311.02              363          33568.349917                64734.634614
VDE0005VDE    7471563.00              286          26124.346154                54343.973117
PCE0180PCE    7236988.90             1079           6707.125950                19814.376570
VZR0041VZR    6133400.50              499          12291.383768                34196.677680
TRM0006TRM    5972615.00              395          15120.544304                37037.410657
DCM0548DCM    5787385.00              258          22431.724806                39472.167363
VUE0148VUE    4463399.60              721           6190.568100                15013.061622
... (dan 594 item lainnya)
```

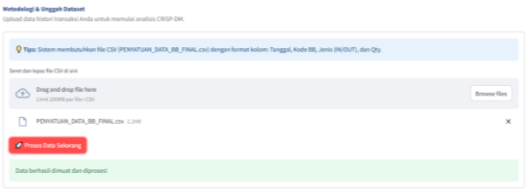
Gambar 10. Bahan Baku Per Cluster

3.6. Deployment (Implementasi)

Tahap akhir deployment mengintegrasikan model K-Means yang divalidasi ke web berbasis Streamlit. Sistem ini dirancang untuk mengotomatisasi seluruh alur analitik secara real-time, mulai dari pengunggahan data mutasi stok mentah, pemprosesan, hingga eksekusi algoritma klastering. Luaran analisis disajikan melalui visualisasi grafik dan tabel pemetaan kategori bahan baku (Vital, Esensial, dan Umum) yang dapat langsung diunduh. Implementasi ini secara praktis berfungsi sebagai sistem pendukung keputusan (Decision Support System) untuk membantu manajemen Dua Naga Corporation dalam merumuskan strategi pengendalian inventaris yang efisien dan berbasis data.

3.6.1 Halaman Upload Dataset Sistem

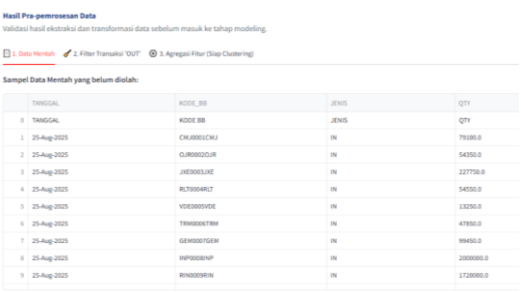
Pada tahap awal implementasi, sistem menyediakan fitur unggah dataset dalam format CSV yang berisi data histori mutasi bahan baku perusahaan. Pengguna dapat mengunggah file yang memuat atribut tanggal transaksi, kode bahan baku, jenis transaksi, dan kuantitas penggunaan. Setelah proses unggah berhasil dilakukan, sistem secara otomatis membaca dataset dan menyiapkan data untuk diproses ke tahapan analisis berikutnya. Fitur ini dirancang untuk mempermudah integrasi data operasional perusahaan ke dalam sistem clustering tanpa perlu proses input manual.



Gambar 11. Upload Dataset

3.6.2 Halaman Pra-Pemprosesan Data

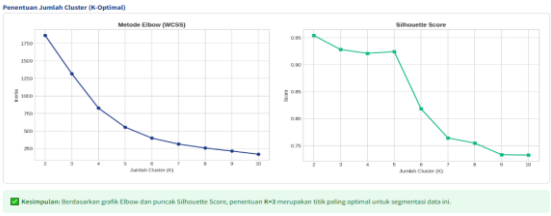
Pada tahap implementasi sistem streamlit menyediakan fitur pra-pemprosesan data secara otomatis untuk memastikan data yang digunakan proses clustering dalam kondisi optimal. Sistem melakukan validasi struktur dataset, penyaringan transaksi berdasarkan jenis pengeluaran barang (OUT), serta pembersihan data yang tidak konsisten. Selain itu, sistem menampilkan hasil agregasi data sebagai bentuk transformasi dari data transaksi mentah menjadi data analitik yang siap digunakan pada tahap pemodelan K-Means Clustering. Implementasi fitur ini bertujuan meningkatkan efisiensi pengolahan data dan meminimalkan kesalahan manual dalam proses analisis persediaan bahan baku.



Gambar 12. Fitur Pra-Pemprosesan Data

3.6.3 Penentuan Jumlah Cluster Optimal

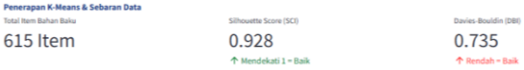
Menampilkan proses evaluasi jumlah cluster optimal menggunakan metode Elbow dan Silhouette Score. Grafik Elbow digunakan untuk melihat titik penurunan nilai WCSS yang mulai melandai, sedangkan grafik Silhouette Score digunakan untuk mengevaluasi kualitas pemisahan antar cluster. Berdasarkan hasil visualisasi kedua metode tersebut, sistem menetapkan nilai k=3 sebagai jumlah cluster paling optimal untuk segmentasi bahan baku. Implementasi visualisasi ini membantu pengguna memahami proses pemilihan parameter clustering secara lebih objektif dan terukur.



Gambar 13. Evaluasi Cluster

3.6.4 Hasil Clustering

Sistem menampilkan hasil clustering terhadap 615 entitas bahan baku yang telah diproses menggunakan algoritma K-Means. Hasil evaluasi model menunjukkan nilai Silhouette Coefficient Index (SCI) sebesar 0,928 dan Davies-Bouldin Index (DBI) sebesar 0,735, yang mengindikasikan kualitas pengelompokan sangat baik. Informasi hasil clustering ditampilkan secara otomatis melalui antarmuka web untuk mempermudah pengguna dalam melakukan analisis dan pengambilan keputusan inventaris secara real-time.



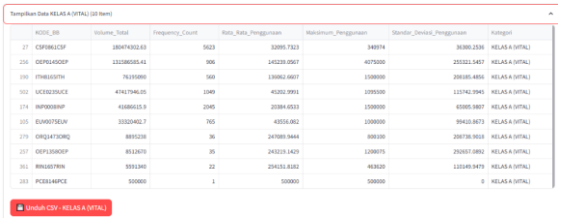
Gambar 14. Hasil Clustering

3.6.5 Rekomendasi Strategis

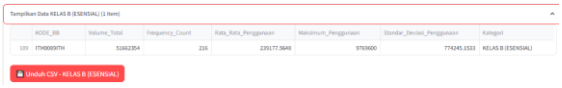
Pada tahap akhir implementasi, sistem menampilkan interpretasi hasil clustering sebagai bagian dari fungsi *Decision Support System* untuk mendukung pengelolaan persediaan bahan baku perusahaan. Halaman menyajikan pembagian bahan baku ke dalam tiga kategori prioritas, yaitu Kelas A (Vital), Kelas B (Esensial), dan Kelas C (Umum), yang disertai dengan rekomendasi strategi pengendalian stok sesuai karakteristik masing-masing klaster. Sistem merekomendasikan penerapan *tight control* dengan evaluasi stok harian dan penyediaan *safety stock* pada Kelas A, *normal control* pada Kelas B, serta *simple control* melalui mekanisme *bulk order* pada Kelas C untuk meningkatkan efisiensi biaya penyimpanan dan distribusi.



Gambar 15. Interpretasi dan Rekomendasi



Gambar 16. Daftar Bahan Baku Kelas A



Gambar 17. Daftar Bahan Baku Kelas B



Gambar 18. Daftar Bahan Baku Kelas C

Pada gambar 16, 17 dan 18 pengguna juga dapat meninjau serta mengunduh daftar entitas bahan baku secara terpisah pada setiap kategori klaster yaitu Kelas A sebanyak 10 item, Kelas B sebanyak 1 item, dan Kelas C sebanyak 604 item, sehingga proses validasi data bahan baku dan pengunduhan dalam format CSV dapat dilakukan secara lebih cepat, akurat, dan responsif secara *real-time*.

Jumlah anggota klaster yang tidak seimbang menunjukkan bahwa distribusi penggunaan bahan baku pada perusahaan bersifat heterogen. Kelas B yang hanya terdiri dari satu item mengindikasikan adanya bahan baku dengan karakteristik penggunaan yang unik dan berada di antara kelompok penggunaan tinggi (Kelas A) dan penggunaan rendah (Kelas C). Kondisi ini mencerminkan pola distribusi data yang tidak merata, sehingga jumlah anggota klaster yang berbeda merupakan hasil yang wajar. Selain itu, nilai SCI sebesar 0,928 dan DBI sebesar 0,735 menunjukkan bahwa klaster yang terbentuk memiliki kualitas pemisahan yang baik, sehingga hasil pengelompokan tetap valid untuk mendukung pengambilan keputusan pengendalian persediaan.

4. Kesimpulan dan Saran

Berdasarkan rangkaian proses penelitian yang telah dilakukan menggunakan kerangka kerja CRISP-DM, dapat ditarik beberapa kesimpulan utama sebagai berikut:

1) Segmentasi Inventaris Terukur: Penelitian ini berhasil mengolah 62.922 baris transaksi pengeluaran (OUT) barang dari 615 entitas bahan baku unik. Transformasi data dari daftar transaksi harian menjadi 5 variabel kriteria (Total Volume, Frekuensi Transaksi, dan Rata rata Pengunaan, Standar Deviasi Penggunaan) memungkinkan Perusahaan dapat melihat karakteristik setiap bahan baku secara lebih objektif dibandingkan metode manual
2) Akurasi Pengelompokan: Melalui metode Elbow dan Silhouette Score, ditetapkan bahwa jumlah klaster paling optimal adalah k=3. Model ini membagi bahan baku ke dalam tiga kategori yaitu: Cluster 2 (Kelas A-Vital) sebanyak 10 item dengan karakteristik penggunaan tinggi, Cluster 1 (Kelas B-Esensial) sebanyak 1 item dengan karakteristik penggunaan sedang, dan Cluster 0 (Kelas C-Umum) sebanyak 604 item dengan karakteristik penggunaan rendah
3) Kualitas Model Tinggi: Hasil evaluasi menunjukkan performa model yang sangat kuat, dengan nilai Silhouette Coefficient Index (SCI) sebesar 0.928 dan Davies-Bouldin Index (DBI) sebesar 0.735. Skor SCI yang mendekati angka 1 membuktikan bahwa setiap bahan baku telah dikelompokkan ke dalam klaster yang sangat tepat dengan batas pemisahan antar-kelompok yang jelas
4) Solusi Digital Praktis: Implementasi model ke dalam antarmuka berbasis web menggunakan Streamlit memberikan solusi nyata bagi manajemen operasional. Sistem ini mampu melakukan pengelompokkan otomatis secara real-time, sehingga memudahkan tim gudang dalam mengidentifikasi material prioritas tanpa harus melakukan perhitungan manual yang rumit.

Saran dari studi ini adalah sebagai berikut:

1) Mengembangkan Model Analisis Prediksi: tingkat akurasi model melalui skor Silhouette sangat memuaskan, K-Means ini sebatas melakukan pengelompokan berdasarkan data historis pengeluaran barang. Oleh sebab itu, disarankan untuk melakukan integrasi dengan metode *forecasting* seperti Long Short-Term Memory (LSTM) atau Random Forest agar sistem tidak hanya mampu mengelompokkan bahan baku, tetapi juga memprediksi volume kebutuhan bahan baku di masa depan secara lebih dinamis.
2) Penyesuaian Strategi Pengendalian Stok: terdapat perbedaan jumlah anggota klaster yang signifikan, perusahaan disarankan menerapkan strategi kontrol yang berbeda (*differentiated control*). Prioritas pengawasan ketat dan penyediaan stok pengaman (*safety stock*) difokuskan pada Kelas A guna menjamin kelancaran produksi. Sementara itu, untuk Kelas C yang berjumlah besar, disarankan melakukan evaluasi rutin untuk mencegah penumpukan barang yang tidak produktif serta mengoptimalkan efisiensi ruang penyimpanan di gudang.

5. UCAPAN TERIMA KASIH

Peneliti menyampaikan ucapan terima kasih kepada Dua Naga Corporation atas dukungan serta pemberian data primer yang mendukung kelancaran penelitian ini. Ucapan terima kasih juga disampaikan kepada Supervisor Data Analyst dan Tim Gudang atas bantuan, arahan teknis, dan wawasan operasional yang diberikan selama proses penelitian. Selain itu, peneliti mengucapkan terima kasih kepada dosen pembimbing, Ibu Dwi Hartanti, S.Kom., M.Kom dan Ibu Hanifah Permatasari S.Kom., M.Kom, atas bimbingan, masukan, serta dukungan yang diberikan selama pelaksanaan penelitian ini. Peneliti berharap hasil ini dapat memberikan manfaat serta menjadi referensi dalam pengembangan pengelolaan persediaan berbasis data pada perusahaan manufaktur.

Daftar Pustaka:

[1] V. Kumar dan A. K. Singh, “Inventory Classification and Control Using ABC Analysis and K-Means Clustering for Manufacturing Industries,” International Journal of Supply Chain Management, vol. 9, no. 3, hal. 88–96, 2021.
[2] N. Habibah, “Analisis Pengendalian Persediaan Bahan Baku Menggunakan Metode ABC Analysis terhadap Penjadwalan Produksi,” Jurnal Manajemen Industri dan Logistik, vol. 8, no. 1, hal. 45–53, 2024.
[3] M. Rizkyandri, A. Setiawan, dan D. Pratama, “Penerapan Algoritma K-Means Clustering untuk Penentuan Persediaan Barang pada Toko Retail,”

Jurnal Informatika dan Sistem Informasi, vol. 9, no. 2, hal. 112–120, 2023.

[4] S. Melisa, S. Defit, dan R. Sovia, "Penerapan Algoritma K-Means Clustering untuk Optimalisasi Persediaan Liquid Vape Berdasarkan Data Penjualan," Jurnal KomtekInfo, vol. 12, no. 1, hal. 38–45, 2025. DOI: 10.35134/komtekinfo.v12i1.620.

[5] R. Y. Pratama, A. Fauzi, dan D. Kurniawan, "Penerapan Algoritma K-Means Clustering untuk Pengelompokan Persediaan Bahan Baku pada Industri Manufaktur," Jurnal Informatika Rekayasa Elektronik (JIRE), vol. 8, no. 2, hal. 145–154, 2024. DOI: 10.36595/jire.v8i2.512.

[6] D. Ramdhan, G. Dwilestari, R. D. Dana, A. Ajiz, dan Kaslani, "Clustering Data Persediaan Barang Dengan Menggunakan Metode K-Means," MEANS (Media Informasi Analisa dan Sistem), vol. 7, no. 1, 2022. DOI: 10.54367/means.v7i1.1826.

[7] Z. Zhang dan S. Wang, "Application of CRISP-DM Model in Manufacturing Data Mining for Inventory Management," Procedia Computer Science, vol. 199, hal. 437–444, 2022. DOI: 10.1016/j.procs.2022.01.054.

[8] D. T. Larose dan C. D. Larose, Discovering Knowledge in Data: An Introduction to Data Mining, 2nd ed. Hoboken, New Jersey: John Wiley & Sons, 2014. DOI: 10.1002/9781118874059.

[9] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems, 2nd ed. California: O'Reilly Media, 2019.

[10] W. McKinney, Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, 2nd ed. Sebastopol, California: O'Reilly Media, 2018.

[11] Y. Liu dan X. Li, "Evaluation of Clustering Algorithms Based on Silhouette Coefficient and Davies-Bouldin Index in Data Mining," Journal of Computer Science and Technology, vol. 36, no. 4, hal. 785–794, 2022.

[12] N. D. Putri, A. Wibowo, dan M. F. Rifai, "Penerapan Algoritma K-Means Clustering untuk Klasifikasi Persediaan Barang pada Sistem Inventori Gudang," Jurnal Media Informatika Budidarma, vol. 6, no. 3, hal. 1234–1242, 2022. DOI: 10.30865/mib.v6i3.4125.

[13] R. A. Siregar dan D. Abdullah, "Implementasi Data Mining Menggunakan Metode K-Means untuk Pengelompokan Persediaan Barang pada Perusahaan Manufaktur," Jurnal Sistem Informasi dan Teknologi Informasi, vol. 11, no. 2, hal. 95–104, 2023. DOI: 10.30865/json.v11i2.5632.

[14] F. A. Ramadhan, B. Prasetyo, dan S. Nurhayati, "Analisis Pengendalian Persediaan Bahan Baku Menggunakan K-Means Clustering pada Industri Manufaktur," Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 7, no. 4, hal. 812–820, 2023. DOI: 10.29207/resti.v7i4.5120.

[15] A. Rahmadani dan N. Nursyahira, "Implementasi Algoritma K-Means Untuk Clustering Data Inventori," Indonesian Journal of Informatic Research and Software Engineering (IJIRSE), vol. 5, no. 1, 2025. DOI: 10.57152/ijirse.v5i1.1832.

[16] W. Windraningsih, R. R. Setiawan, dan Supriyono, "Application of K-Means Clustering for Consumable Inventory Expenditure at BKPSDM of Kudus Regency," Jurnal Teknologi Informasi dan Pendidikan, vol. 19, no. 1, 2026. DOI: 10.24036/jtip.v19i1.1106.

[17] A. N. Anggraini, A. Amali, dan M. S. Anwar, "Demand-Based Product Classification Using K-Means with Intermittency Metrics," Edumatic: Jurnal Pendidikan Informatika, vol. 10, no. 1, 2026. DOI: 10.29408/edumatic.v10i1.34435.

[18] Streamlit Inc., "Streamlit Documentation for Data Application Development," 2025. [Daring]. Tersedia pada: https://docs.streamlit.io/. [Diakses: 08-Mei-2026].

[19] Scikit-learn Developers, "Clustering Performance Evaluation Metrics Documentation," 2025. [Daring]. Tersedia pada: https://scikit-learn.org/stable/modules/clustering.html. [Diakses: 08-Mei-2026].

[20] S. Vijayarani dan M. Dhayanand, "Inventory Management and Analysis Using K-Means Clustering Algorithm Approach," International Journal of Computer Applications, vol. 146, no. 13, hal. 15–22, 2020.