

## **IMPLEMENTASI DATA MINING PADA STOK PENGGUNAAN BARANG DI GMF AEROASIA MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING**

**Tri Bayu Pamungkas<sup>1</sup>, Siti Maesaroh<sup>2</sup>, Pebri Ardiansyah<sup>3</sup>**

<sup>1</sup>Fakultas Ilmu Komputer, <sup>2</sup>Universitas Mercu Buana

<sup>3</sup>Fakultas Sains dan Teknologi, <sup>3</sup>Universitas Raharja

Email: <sup>1</sup>tribayu.pamungkas25@gmail.com, <sup>2</sup>[siti.maesaroh@mercubuana.ac.id](mailto:siti.maesaroh@mercubuana.ac.id),

<sup>3</sup>[pebri@raharja.info](mailto:pebri@raharja.info)

### *Abstract*

*The rapid development in technology has greatly benefited inventory management in the industrial sector. At PT GMF AeroAsia, a company specializing in commercial aircraft maintenance, inventory management is of utmost importance. During aircraft repairs, various materials are used and recorded in a database. To optimize the abundant data, data mining methods such as k-means clustering can be applied to group similar data based on their characteristics. In this research, the dataset used consists of stock usage data at GMF AeroAsia from 2016 to 2021. This dataset includes information on received goods, utilized goods, and expired goods, with the objective of categorizing items into high, medium, and low levels of expiration to identify potential losses. Prior to the research, the dataset was prepared by removing outliers and normalizing using min-max normalization. The results obtained from K-Means Clustering with the assistance of RapidMiner produced three clusters. Cluster 1 consisted of 13 data points with a low quantity of expired goods, cluster 2 consisted of 5 data points with a moderate quantity of expired goods, and cluster 3 consisted of 2 data points with the highest quantity of expired goods. The clustering results obtained from the k-means algorithm in this research were relatively good, as indicated by the evaluation using the Davies-Bouldin Index, which obtained a value of 0.494. This research provides valuable insights for PT GMF AeroAsia Tbk. in inventory management and preventing losses due to expired goods.*

*Keywords: Data Mining, Clustering, K-Means, RapidMiner, Davies Bouldin*

### **PENDAHULUAN**

Perkembangan teknologi telah memberikan manfaat signifikan dalam berbagai bidang, termasuk pengelolaan stok persediaan barang di industri. Stok persediaan barang merujuk pada barang-barang yang dimiliki oleh pemilik bisnis atau perusahaan untuk digunakan dalam proses produksi. Persediaan tersebut merupakan aset lancar yang sangat penting dalam menciptakan keuntungan bagi perusahaan (Ramadhan et al., 2022). Dalam melakukan perawatan pesawat, PT GMF AeroAsia Tbk. menggunakan berbagai macam material sesuai kebutuhan untuk memperbaiki bagian pesawat tertentu. Pada perbaikan terdapat bahan yang digunakan untuk memperbaiki bagian-bagian pesawat, dimana bahan-bahan ini memiliki jumlah dan jenis sangat bervariasi. Jumlah jenis materi ini tercatat dengan baik pada *database* sebagai bahan evaluasi serta untuk pengadaan selanjutnya.

Kumpulan data yang banyak bisa memberikan kita sebuah informasi jika dapat mengolahnya dengan tepat. Salah satu cara mengambil informasi dari data itu dengan *data mining*. *Data mining* adalah rangkaian proses yang digunakan untuk mengungkap nilai tambah berupa informasi baru yang sebelumnya tidak diketahui secara manual dari sekumpulan data. Proses ini melibatkan pencarian pola-pola yang ada dalam data dengan tujuan untuk mengubah data menjadi informasi yang lebih berharga. Informasi ini ditemukan melalui proses ekstraksi dan pengenalan pola-pola penting atau menarik dari data yang terdapat dalam sekumpulan data tersebut. (Rahman, 2017). *Clustering* termasuk salah satu metode dari *data mining* (Gustientiedina et al., 2019) yang memiliki kemampuan untuk mengelompokkan data ke suatu

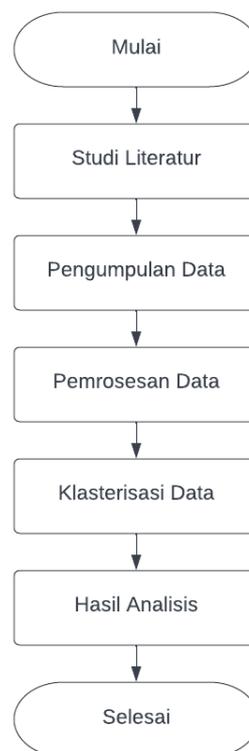
kelompok berdasarkan karakteristik yang serupa, sementara data yang memiliki karakteristik berbeda akan dikelompokkan ke kelompok lainnya. *K-Means* merupakan algoritma *clustering* sederhana yang bersifat tanpa arahan (*unsupervised*) (Ramdhan et al., 2022). Algoritma ini populer dalam pengelompokan karena kesederhanaan dan efisiensinya, serta diakui sebagai salah satu dari sepuluh algoritma *data mining* terbaik oleh IEEE.(Gustientiedina et al., 2019).

Ramdhan pada penelitiannya menggunakan algoritma *K-Means* pada dataset dengan 249 data, berhasil membagi data menjadi 2 *cluster* menggunakan bantuan alat *RapidMiner*. *Cluster 1* berisi 42 data dan *cluster 2* dengan 207 data. Dari hasil analisis yang dilakukan menunjukkan nilai akurasi sebesar 100% (Ramdhan et al., 2022). Pada penelitian lainnya, Sari (Sari Wulan et al., 2018) berhasil mengimplementasikan algoritma *K-Means Clustering* pada 34 data jumlah Persentase Balita yang Pernah Mendapat Imunisasi Campak berdasarkan provinsi. Hasil yang didapat adalah terdapat 3 *cluster* yaitu Provinsi Tinggi (C1) dengan 21 provinsi, kluster Provinsi Sedang (C2) dengan 12 provinsi, dan kluster Provinsi Rendah (C3) dengan 1 provinsi(Sari Wulan et al., 2018).

Berdasarkan uraian sebelumnya, maka penelitian ini dilakukan untuk meningkatkan efisiensi dalam pengelolaan stok, dengan memanfaatkan teknik *data mining* dan algoritma *K-Means Clustering* untuk mengelompokkan data stok persediaan ke dalam kelompok berdasarkan tingkat kedaluwarsa. Tujuan penelitian ini adalah mencegah kerugian di masa depan dengan mengoptimalkan persediaan barang. Dengan menggunakan metode ini, diharapkan dapat memprediksi dan membagi data penggunaan stok barang ke dalam beberapa kelompok, sehingga memudahkan pengambilan keputusan dalam pengadaan produk/barang dan melakukan efisiensi yang lebih baik.

### **Metode Penelitian**

Langkah-langkah penelitian yang akan dilakukan dalam penelitian ini ditunjukkan dalam Gambar 1 di bawah ini.



**Gambar 1. Tahapan Penelitian**

### *Data Mining*

*Data mining* adalah suatu proses yang melibatkan penerapan satu atau lebih teknik pembelajaran mesin (*machine learning*) untuk secara otomatis menganalisis dan mengeksplorasi pengetahuan dari data (Ramdhan et al., 2022). *Output* dari *data mining* ini digunakan untuk mendukung pengambilan keputusan bisnis dengan menganalisis data yang tersimpan dalam *database* atau kumpulan data.

Tahapan dalam proses *data mining* dapat dijelaskan secara umum sebagai berikut (Luthfi dalam (Ramdhan et al., 2022)).

- a. *Data selection*, Sebelum memulai proses penggalian informasi dalam *data mining*, langkah pertama adalah memilih (seleksi) data yang relevan dari sekumpulan data operasional.
- b. *Pre-processing/Cleaning*, Sebelum melaksanakan proses *data mining*, diperlukan langkah pembersihan (*cleaning*) pada data yang menjadi fokus penggalian data (KDD). *Data Cleaning* mencakup tindakan seperti menghapus data duplikat, memeriksa konsistensi data, dan memperbaiki kesalahan seperti kesalahan pengetikan.
- c. *Transformation Coding*, adalah proses yang dilakukan pada data yang telah dipilih untuk mempersiapkannya agar sesuai untuk proses *data mining*.
- d. *Data mining*, adalah suatu proses untuk menemukan pola atau informasi menarik dalam data yang telah dipilih menggunakan teknik atau metode khusus.
- e. *Interpretation/Evaluation*, Hasil pola informasi yang dihasilkan dari proses *data mining* perlu disajikan dalam format yang dapat dipahami oleh pihak-pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut interpretasi (Luthfi dalam (Ramdhan et al., 2022)).

*Data mining* dapat dikelompokkan berdasarkan tugas yang dapat dilakukan (Luthfi dalam (Ramdhan et al., 2022)), yaitu:

- a. Deskripsi : Mencari metode untuk mengilustrasikan pola dan tren yang terdapat dalam data.
- b. Estimasi : Estimasi serupa dengan klasifikasi, namun variabel target estimasi lebih berfokus pada nilai numerik daripada kategori.
- c. Prediksi : Mirip dengan klasifikasi, tetapi dalam prediksi, nilai hasilnya terkait dengan masa depan.
- d. Klasifikasi : Pada klasifikasi, terdapat variabel target berkategori. Sebagai ilustrasi, pendapatan dapat diklasifikasikan ke dalam tiga kategori, yakni tinggi, sedang, dan rendah.
- e. Pengklasteran : Pengklasteran melibatkan pengelompokan record, observasi, atau objek berdasarkan kemiripan mereka. Kluster merupakan kumpulan record yang memiliki kesamaan satu sama lain dan berbeda dari record-record dalam kluster lain.
- f. Asosiasi : Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul bersama dalam satu waktu.

### *Clustering*

Pengertian dari *Clustering* adalah suatu proses pengelompokan atau penggolongan objek berdasarkan informasi yang diperoleh dari data, yang bertujuan untuk memaksimalkan kesamaan antara anggota dalam satu kelompok dan meminimumkan kesamaan antara kelompok/*clusternya* (Rahman, 2017).

### *K-Means*

*K-Means* merupakan salah satu metode nonhierarki untuk pengelompokan data, yang memungkinkan data dipartisi menjadi dua kelompok atau lebih. Metode ini akan memisahkan data ke dalam kelompok-kelompok, di mana data dengan karakteristik yang serupa

dikelompokkan bersama, sedangkan data dengan karakteristik yang berbeda akan dikelompokkan kedalam kelompok lain yang serupa (Gustientiedina et al., 2019). Karena kesederhanaan dan pengelompokannya yang cepat dari kumpulan data besar dan data *outliers*, algoritma *K-Means* menjadi populer.

Berikut adalah langkah-langkah dalam algoritma *K-Means* (Rahman, 2017):

1. Menentukan jumlah *k-cluster* yang akan dibentuk.
2. Menentukan *k-centroid* (pusat *cluster*) secara acak.
3. Menghitung jarak antara setiap data dengan masing-masing *centroid* menggunakan rumus jarak *Euclidean Distance* (Nabila et al., 2021) dengan rumus (1) berikut :

$$D(x_i, \pi_i) = \sqrt{\sum_{i=1}^n (x_i - \pi_i)^2} \quad (1)$$

Dimana  $d(x_i, \mu_i)$  adalah jarak antara data ke- $i$  dengan *centroid*  $\mu$  pada *cluster*  $n$ ,  $x_i$  adalah bobot data ke- $i$  yang ingin dicari jaraknya, dan  $\mu_i$  adalah bobot data ke- $i$  pada pusat *cluster*.

4. Mengelompokkan data berdasarkan jarak terdekat antara data dan *centroid*.
5. Menentukan nilai *centroid* baru yang dilakukan dengan menghitung rata-rata dari *cluster* yang bersangkutan menggunakan rumus (2):

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

Dimana:

$n_k$  = jumlah data dalam *cluster*  $k$

$d_i$  = jumlah nilai jarak yang termasuk dalam masing-masing *cluster*

6. Mengulangi langkah 3-5 hingga tidak ada perubahan pada anggota setiap *cluster*.

### *RapidMiner*

*RapidMiner* adalah sebuah perangkat lunak dengan sifat terbuka (open source). *RapidMiner* menyediakan solusi dalam melakukan analisis *data mining*, *text mining*, dan analisis prediksi. *RapidMiner* merupakan perangkat lunak mandiri yang digunakan untuk analisis data dan juga sebagai mesin *data mining* yang dapat diintegrasikan dengan produk lainnya (Nabila et al., 2021).

### *Davies Bouldin Index*

*Davies-bouldin index* adalah salah satu metode evaluasi internal yang digunakan untuk mengukur kualitas *cluster* dalam metode *clustering data* berdasarkan kohesi dan separasi. Dalam *clustering*, kohesi mengacu pada tingkat kedekatan data dengan pusat *cluster* yang mereka miliki, sementara separasi berhubungan dengan jarak antara pusat *cluster* yang berbeda (Nabila et al., 2021).

## HASIL DAN PEMBAHASAN

Berdasarkan penjelasan dasar teori sebelumnya, bagian ini akan menjelaskan cara algoritma bekerja dan tahapan proses yang akan dilakukan untuk melakukan *clustering*.

### Metode Pengumpulan Data

Dalam penelitian ini, data diperoleh melalui pengajuan permohonan penelitian secara *online* kepada perusahaan GMF Aeroasia. Surat permohonan tersebut menjelaskan tujuan penelitian, metodologi yang akan digunakan, dan manfaat yang diharapkan dari penelitian ini. Metode pengumpulan data akan menggunakan data penggunaan stok barang di PT GMF AeroAsia dari tahun 2016 hingga 2021. Data ini mencakup informasi tentang penggunaan stok barang, terutama barang kimia habis pakai. Tahap awal dalam pengumpulan data adalah mengumpulkan data mentah yang diperlukan untuk penelitian ini. Data mentah ini kemudian akan diproses dan dibersihkan (*data cleaning*) untuk menghilangkan data yang tidak relevan atau tidak

lengkap. Setelah mendapatkan dataset yang telah dibersihkan, langkah selanjutnya adalah mengekstraksi kolom-kolom yang relevan dengan permasalahan penelitian.

**a. Persiapan Data**

Dataset yang telah diperoleh mencakup rentang waktu dari tahun 2016 hingga 2021 dan terdiri dari total 47.636 baris data. Informasi mengenai dataset yang digunakan dalam penelitian ini dapat ditemukan pada gambar 2 berikut.:

**Gambar 2. Dataset Mentah**

**b. Pre-Processing**

Tahap awal pada proses ini adalah pembersihan data (*data cleaning*), di mana variabel atau fitur yang sesuai dipilih dari dataset untuk pengolahan lebih lanjut. Pada penelitian ini, terdapat empat kolom yang dipilih, yaitu Deskripsi, Penerimaan per tahun, Penggunaan per tahun, dan Kedaluwarsa per tahun. Selanjutnya, dipilih dan diambil 30 baris data yang menjadi fokus utama untuk dianalisis penggunaannya pada setiap tahunnya. Data yang telah melalui proses pembersihan dapat ditemukan dalam Tabel 1 berikut ini:

**Tabel 1. Tuliskan deskripsi tabel**

Barang	Rata-rata Penerimaan Barang per Tahun	Rata-rata Penggunaan Barang per Tahun	Rata-rata Barang Expired per Tahun
MULTI PURPOSE SEALANT WHITE,90 ML (TRP)	719,17	660	59,17
PENETRATING OIL	1431,17	1351,83	79,33
ALEXIT FILLER HARDENER 343-15	16,67	16,5	0,17
FLUID - HYDRAULIC, PETROLEUM BASE	751,33	712,83	38,5
ALEXIT FST 343-60 FILLER WHITE 10KG	216,17	214,17	2
.....	.....	.....	.....
FUELSTAT RESINAE PLUS	330	311,83	60,67

**c. Centroid Data**

Dalam penelitian ini jumlah kluster yang ditentukan oleh peneliti adalah 3 *cluster* (K) agar sesuai dengan kelompok yang akan dihasilkan yaitu barang dengan kerugian besar, sedang, dan rendah. Setelah menentukan jumlah K, langkah selanjutnya adalah menentukan titik *centroid* awal secara acak. *Centroid* awal adalah titik pusat dari kluster pertama yang didapat secara acak. Tabel 2 di bawah ini menunjukkan *centroid* awal yang digunakan dalam penelitian ini.

**Tabel 2. Centroid Awal**

<i>Centroid</i>	Rata-rata penerimaan barang per tahun	Rata-rata penggunaan barang per tahun	Rata-rata barang <i>expired</i> per tahun
C1	0	0	0
C2	1	1	0.327
C3	0.170	0.161	1

Setelah memperoleh *centroid* awal, langkah berikutnya adalah menghitung jarak antara setiap data dengan *centroid* awal menggunakan rumus *Euclidean Distance*.

$$D(1,1) : \sqrt{(0,971 - 0)^2 + (0,915 - 0)^2 + (0,957 - 0)^2} = 1,641$$

$$D(1,2) : \sqrt{(0,971 - 1)^2 + (0,915 - 1)^2 + (0,957 - 0,327)^2} = 0,636$$

$$D(1,2) : \sqrt{(0,971 - 0,170)^2 + (0,915 - 0,161)^2 + (0,957 - 1)^2} = 1,1$$

Lakukan perhitungan seperti diatas untuk menentukan jarak terdekat data ke *centroid* awal dengan membandingkan antara 3 *cluster*. Hasil perhitungan pada keseluruhan data yang diteliti ditampilkan pada tabel 3 berikut.

**Tabel 3. Hasil Iterasi ke-1**

Data	Penerimaan	Penggunaan	<i>Expired</i>	C1	C2	C3	Jarak terdekat	<i>Cluster</i>
1	0,971	0,915	0,957	1,642	0,636	1,101	0,636	C2
2	0,000	0,000	0,000	0,000	1,452	1,027	0,000	C1
3	0,276	0,281	0,030	0,395	1,063	0,983	0,395	C1
4	0,744	0,725	0,367	1,102	0,378	1,024	0,378	C2
5	0,771	0,763	0,345	1,139	0,330	1,074	0,330	C2
6	0,509	0,514	0,095	0,730	0,729	1,029	0,729	C2
7	0,219	0,219	0,079	0,320	1,132	0,924	0,320	C1
8	0,403	0,411	0,032	0,577	0,889	1,027	0,577	C1
9	0,038	0,038	0,004	0,054	1,399	1,013	0,054	C1
10	0,036	0,035	0,026	0,057	1,397	0,991	0,057	C1
11	1,000	1,000	0,327	1,452	0,000	1,359	0,000	C2
12	0,170	0,161	1,000	1,027	1,358	0,000	0,000	C3
13	0,434	0,416	0,344	0,693	0,813	0,752	0,693	C1
14	0,137	0,124	0,189	0,265	1,237	0,812	0,265	C1
15	0,910	0,934	0,024	1,304	0,323	1,448	0,323	C2
16	0,046	0,043	0,041	0,075	1,381	0,975	0,075	C1
17	0,199	0,193	0,132	0,308	1,153	0,869	0,308	C1
18	0,183	0,184	0,057	0,266	1,186	0,944	0,266	C1
19	0,161	0,161	0,051	0,233	1,218	0,949	0,233	C1
20	0,433	0,420	0,981	1,152	1,042	0,369	0,369	C3

Langkah selanjutnya adalah menentukan nilai *centroid* yang baru berdasarkan hasil pengelompokan sebelumnya. Perhitungan nilai *centroid* baru dilakukan dengan rumus seperti berikut ini.

Menghitung rata-rata dari setiap *cluster* yang terbentuk.

$$C1 = [(0,000+0,276+0,219+0,403+0,038+0,036+0,434+0,137+0,046+0,199+0,183+0,161)/12, (0,000+0,281+0,219+0,411+0,038+0,035+0,416+0,124+0,043+0,193+0,184+0,161)/12, (0,000+0,030+0,079+0,032+0,004+0,026+0,344+0,189+0,041+0,132+0,057+0,051)/12]$$

$$C1 = [0,178; 0,176; 0,082]$$

$$C2 = [(0,971+0,744+0,771+0,509+1,000+0,910)/6, (0,915+0,725+0,763+0,514+1,000+0,934)/6, (0,957+0,367+0,345+0,095+0,327+0,024)/6]$$

$$C2 = [0,818; 0,809; 0,353]$$

$$C3 = [(0,170+ 0,433)/2, (0,161+ 0,420)/2, (1,000+0,981)/2]$$

$$C3 = [0,302; 0,291; 0,330]$$

Jadi, didapat *centroid* untuk iterasi kedua seperti berikut.

**Tabel 4. Centroid baru untuk iterasi ke-2**

<i>Centroid</i>	Rata-rata penerimaan barang per tahun	Rata-rata penggunaan barang per tahun	Rata-rata barang expired per tahun
C1	0,178	0,176	0,082
C2	0,818	0,809	0,353
C3	0,302	0,291	0,330

Setelah menemukan titik *centroid* baru, langkah berikutnya adalah mengulangi perhitungan jarak seperti yang telah dijelaskan sebelumnya untuk menempatkan data ke dalam kluster baru yang sesuai. Jika langkah-langkah tersebut diulang dengan cara yang sama dan data dalam kluster tidak berubah antara hasil sebelumnya dan hasil selanjutnya. Dalam penelitian ini, posisi kluster pada iterasi ke-3 tetap, tidak berubah dari iterasi ke-2, sebagaimana yang terlihat pada tabel 5 berikut ini dengan perhitungan iterasi ke-3.

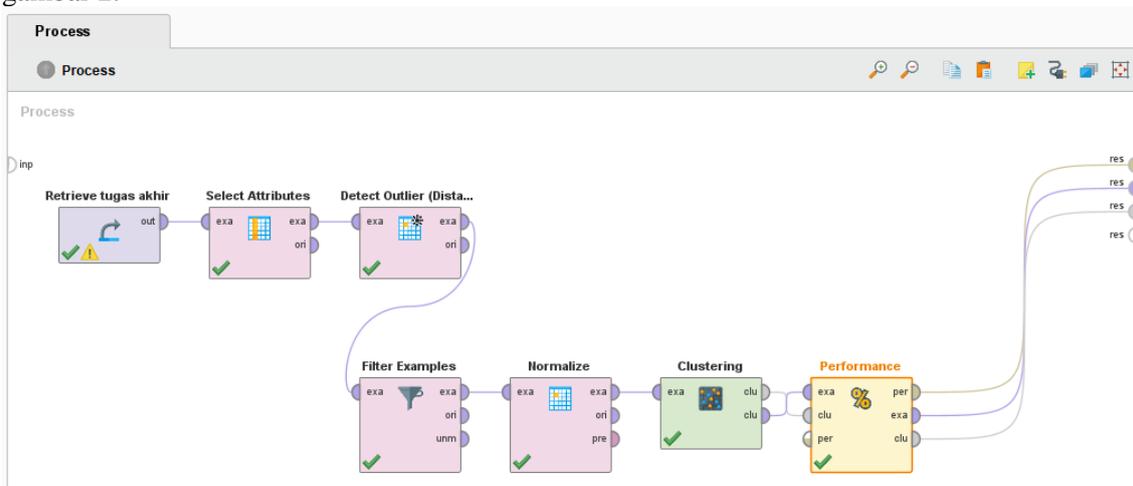
**Tabel 5. Hasil akhir pada iterasi ke3**

Data	Penerimaan	Penggunaan	Expired	C1	C2	C3	Jarak terdekat	Cluster
1	0,971	0,915	0,957	1,393	0,633	0,916	0,633	C2
2	0,000	0,000	0,000	0,263	1,203	1,075	0,263	C1
3	0,276	0,281	0,030	0,153	0,822	0,961	0,153	C1
4	0,744	0,725	0,367	0,839	0,112	0,879	0,112	C2
5	0,771	0,763	0,345	0,876	0,065	0,928	0,065	C2
6	0,509	0,514	0,095	0,474	0,498	0,946	0,474	C1
7	0,219	0,219	0,079	0,060	0,884	0,918	0,060	C1
8	0,403	0,411	0,032	0,330	0,658	0,972	0,330	C1
9	0,038	0,038	0,004	0,211	1,151	1,052	0,211	C1
10	0,036	0,035	0,026	0,207	1,147	1,032	0,207	C1
11	1,000	1,000	0,327	1,190	0,266	1,196	0,266	C2
12	0,170	0,161	1,000	0,918	1,121	0,185	0,185	C3
13	0,434	0,416	0,344	0,439	0,549	0,672	0,439	C1
14	0,137	0,124	0,189	0,125	0,979	0,835	0,125	C1
15	0,910	0,934	0,024	1,056	0,363	1,311	0,363	C2
16	0,046	0,043	0,041	0,191	1,131	1,015	0,191	C1

17	0,199	0,193	0,132	0,058	0,900	0,870	0,058	C1
18	0,183	0,184	0,057	0,027	0,938	0,947	0,027	C1
19	0,161	0,161	0,051	0,038	0,970	0,959	0,038	C1
20	0,433	0,420	0,981	0,966	0,833	0,185	0,185	C3

**d. Pembuatan Model**

Penelitian ini menggunakan *RapidMiner* sebagai alat untuk menjalankan algoritma *K-Means Clustering*. Untuk membuat model di *RapidMiner*, terdapat beberapa tahap yang perlu dilakukan, yaitu: mengimpor data, melakukan *pre-processing* data, memilih algoritma, mengkonfigurasi algoritma, melatih model, mengevaluasi model, dan menyimpan model. Pada penelitian ini, pembuatan model di *RapidMiner* memiliki desain proses yang tergambar pada gambar 2.



**Gambar 2. Desain proses pada *RapidMiner***

Algoritma “*k-Means*” menggunakan parameter  $k=3$  dengan *measure types* “*Mixed Measures*”. Lalu menambahkan operator “*Cluster Distance Performance*” untuk mengevaluasi hasil *cluster* dengan parameter utama yaitu “*Davies Bouldin*”. Setelah berhasil dijalankan, terdapat 3 tabs yang akan muncul pada bagian “*Results*”. Pemodelan berhasil mengklasterisasi dataset dengan total 20 data menjadi 3 kelompok.

### Cluster Model

```
Cluster 0: 13 items
Cluster 1: 5 items
Cluster 2: 2 items
Total number of items: 20
```

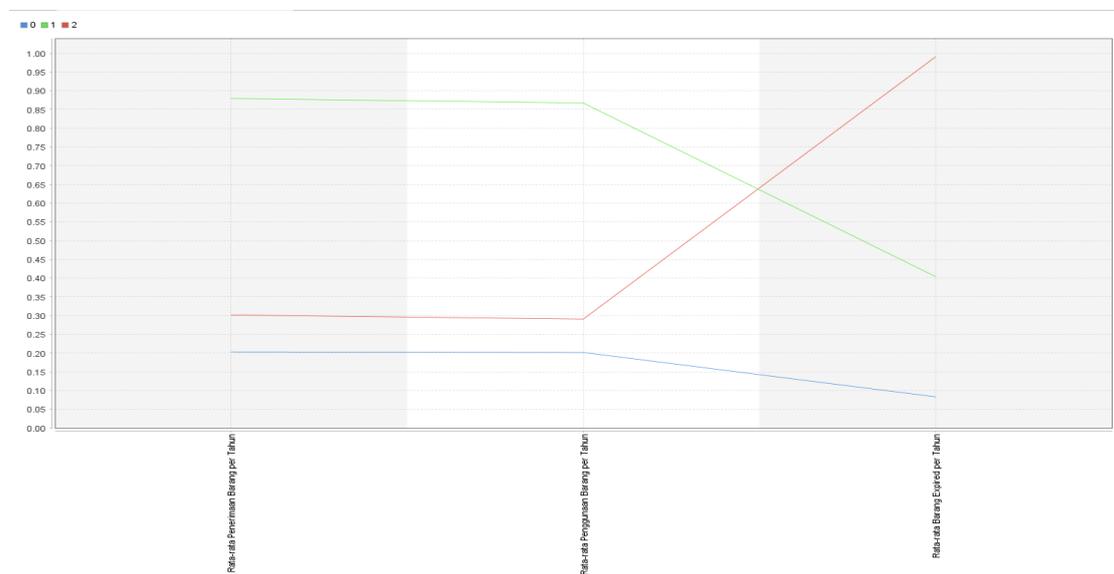
**Gambar 3. Hasil *cluster* pada *RapidMiner***

Tampilan hasil *clustering* menggunakan algoritma *K-Means* pada *RapidMiner* dapat dilihat pada gambar 4 di bawah ini.

Row No.	id	Barang	cluster	outlier	Rata-rata Penerimaan Barang ...	Rata-rata Penggunaan Barang ...	Rata-rata Barang Expired ...
1	1	MULTI PURPOSE SEALANT WHITE,90 ML (TRP)	cluster_1	false	0.971	0.915	0.957
2	2	ALEXIT FILLER HARDENER 343-15	cluster_0	false	0	0	0
3	3	ALEXIT FST 343-60 FILLER WHITE 10KG	cluster_0	false	0.276	0.281	0.030
4	4	ALEXIT INT PAINT BAC 7363 WHITE	cluster_1	false	0.744	0.725	0.367
5	5	PENETRATING OIL 1 EA = 10 OZ	cluster_1	false	0.771	0.763	0.345
6	6	LANDING GEAR FLUID (BMS3-32CT2)	cluster_0	false	0.509	0.514	0.095
7	7	SOLVENT EMULSION DEGREASER	cluster_0	false	0.219	0.219	0.079
8	8	CORROSION INHIBITING COMPOUND REMOVER	cluster_0	false	0.403	0.411	0.032
9	9	DEODORIZER	cluster_0	false	0.038	0.038	0.004
10	10	DISINFECTANT	cluster_0	false	0.036	0.035	0.026
11	11	MULTI PURPOSE SEALANT CLEAR (TRP)	cluster_1	false	1	1	0.327
12	12	AVIOX FINISH GREY	cluster_2	false	0.170	0.161	1
13	13	A/C INTERIOR SANITIZER AND DISINFECTANT	cluster_0	false	0.434	0.416	0.344
14	14	HI VIS SPRAY GEL	cluster_0	false	0.137	0.124	0.189
15	15	SEALANT: SILICONE,RED	cluster_1	false	0.910	0.934	0.024
16	16	ALEXIT HARDENER 345-55	cluster_0	false	0.046	0.043	0.041
17	17	AIRCRAFT TOILET DEODORIZE	cluster_0	false	0.199	0.193	0.132
18	18	CORROSION INHIBITIVE SEALANT	cluster_0	false	0.183	0.184	0.057
19	19	PUTTY/BODYFILLER HARDENER	cluster_0	false	0.161	0.161	0.051
20	20	FUELSTAT RESINAE PLUS	cluster_2	false	0.433	0.420	0.981

**Gambar 4. Hasil Clustering K-Means pada RapidMiner**

Clustering pada penelitian ini didasarkan pada jenis barang yang ada pada data penerimaan, penggunaan, dan expired barang yang ada di GMF AeroAsia yang terbagi menjadi 3 kluster. Kluster-kluster ini memberikan wawasan yang berharga mengenai pola dan karakteristik data barang yang dapat digunakan untuk meningkatkan manajemen persediaan dan pengawasan barang yang kedaluwarsa. Gambar 4.19. di atas menunjukkan bahwa dari 20 data barang. Terdiri dari cluster 1 (cluster\_0) dengan 13 data, cluster 2 (cluster\_1) dengan 5 data dan cluster 3 (cluster\_2) dengan 2 data.



**Gambar 5. Plot Cluster**

Attribute	cluster_0	cluster_1	cluster_2
Rata-rata Penerimaan Barang per Tahun	0.203	0.879	0.302
Rata-rata Penggunaan Barang per Tahun	0.202	0.867	0.291
Rata-rata Barang Expired per Tahun	0.083	0.404	0.991

**Gambar 6. Centroid Cluster**

Berdasarkan Gambar 6, dapat ditarik kesimpulan bahwa *RapidMiner* menggunakan *centroid* yang memiliki informasi sebagai berikut:

- Cluster 1* : Data barang yang rata-rata penerimaan dan penggunaan barangnya rendah dan rata-rata barang *expired* yang rendah.
- Cluster 2* : Data barang yang rata-rata penerimaan dan penggunaan barangnya tinggi dan rata-rata barang *expired* yang sedang.
- Cluster 3* : Data barang yang rata-rata penerimaan dan penggunaan barangnya sedang dan rata-rata barang *expired* yang tinggi.

Setelah mengimplementasikan menggunakan alat *RapidMiner*, hasil yang diperoleh sejalan dengan hasil perhitungan manual. *Cluster* yang memiliki nilai tertinggi mewakili barang-barang dengan tingkat kedaluwarsa yang tinggi. *Cluster* tengah mewakili barang-barang dengan tingkat kedaluwarsa sedang, sementara *cluster* terbawah menggambarkan barang-barang dengan tingkat kedaluwarsa yang rendah.

## Davies Bouldin

Davies Bouldin: 0.494

**Gambar 6. Hasil Performance Davies-bouldin**

Berdasarkan hasil evaluasi menggunakan *Davies-bouldin Index* dengan nilai seperti gambar diatas, dapat disimpulkan bahwa klastering yang dilakukan dengan algoritma *k-means* pada penelitian ini relatif baik.

### KESIMPULAN

Hasil analisis dan penelitian terhadap data penggunaan barang di GMF AeroAsia menggunakan algoritma *k-means* menunjukkan bahwa data berhasil dikelompokkan menjadi tiga klaster. Klaster pertama terdiri dari 13 data barang dan menggambarkan kelompok barang dengan tingkat kedaluwarsa yang rendah. Klaster kedua terdiri dari 5 data barang dan mencerminkan kelompok barang dengan tingkat kedaluwarsa yang sedang. Walaupun jumlah data dalam klaster ini lebih sedikit, keberadaannya menunjukkan perlunya perbaikan dalam pengelolaan persediaan dan pengendalian jumlah barang yang kedaluwarsa. Klaster ketiga terdiri dari 2 data barang dan mewakili kelompok barang dengan tingkat kedaluwarsa yang tinggi. Walaupun jumlah data dalam klaster ini sangat sedikit, keberadaannya mengindikasikan adanya masalah serius dalam manajemen persediaan yang harus segera ditangani. PT GMF AeroAsia perlu melakukan evaluasi dan perbaikan lebih lanjut terhadap pengawasan, pengendalian, dan perencanaan persediaan guna mengurangi jumlah barang yang kedaluwarsa.

Hasil klastering ini memberikan pemahaman yang lebih mendalam tentang pola dan karakteristik data barang di GMF AeroAsia. Informasi ini dapat digunakan oleh PT GMF AeroAsia untuk meningkatkan manajemen persediaan, memperkuat pengawasan, dan

mengurangi pemborosan sumber daya dengan fokus pada klaster yang memiliki tingkat kedaluwarsa yang lebih tinggi. Penelitian ini bertujuan untuk meningkatkan pemahaman tentang pentingnya pengelolaan barang yang kedaluwarsa dan pengendalian persediaan dalam industri penerbangan. Dengan menemukan barang-barang yang memiliki risiko tinggi untuk kedaluwarsa, perusahaan dapat mengambil tindakan proaktif untuk mengurangi jumlah barang kedaluwarsa, meningkatkan efisiensi operasional, dan mengoptimalkan penggunaan sumber daya.

Penelitian ini menggunakan jumlah sampel yang relatif kecil. Untuk mendapatkan hasil yang lebih akurat di masa depan, disarankan untuk memperluas ukuran sampel dengan menggabungkan lebih banyak data dan variabel lainnya. Dengan menggunakan lebih banyak data, dapat meningkatkan validitas kesimpulan penelitian dan memberikan pemahaman yang lebih komprehensif tentang pola dan karakteristik data barang.

#### Daftar Pustaka

- Ramadhan, D., Dwilestari, G., Dinar Dana, R., & Ajiz, A. (2022). *Clustering Data Persediaan Barang dengan Menggunakan Metode K-Means*. 7(1). <https://media.neliti.com/media/publications/506485-none-96de277a.pdf>
- Andryana, S., & Mardiani, E. (2021). Implementasi *Data Mining Untuk Menentukan Persediaan Stok Obat Di Enok Menggunakan Metode K-Means Clustering* (Vol. 8, Issue 3). <http://jurnal.uinsu.ac.id/index.php/jistech/article/view/11905/5480>
- Anggraini, N., Jasmir, Jusia, A. P., Informasi, S., Dinamika Bangsa, S., & Sudirman, J. J. J. (2019). Penerapan Metode *K-Means Clustering* untuk Menentukan Persediaan Stok Barang Pada Toko Pemsart Jambi. <http://ejournal.stikom-db.ac.id/index.php/jimti/article/view/690>
- Rahman, A. T. (2017). Coal Trade Data *Clusterung Using K-Means* (Case Study PT. Global Bangkit Utama).
- Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma *K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru*. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17–24. <https://doi.org/10.25077/teknosi.v5i1.2019.17-24>
- Mawaddah, S. (2021). IMPLEMENTASI ALGORITMA C 5.0 DALAM PREDIKSI STOK BARANG BERDASARKAN PENJUALAN BAHAN PERTANIAN CV. MITRA SEJATI. *JISTech (Journal of Islamic Science and Technology) JISTech*, 6(1), 40–51. <http://jurnal.uinsu.ac.id/index.php/jistech>
- Hendini, A. (2016). PEMODELAN UML SISTEM INFORMASI MONITORING PENJUALAN DAN STOK BARANG (STUDI KASUS: DISTRO ZHEZHA PONTIANAK): Vol. IV (Issue DESEMBER). <https://ejournal.bsi.ac.id/ejurnal/index.php/khatulistiwa/article/view/1262/1027>
- Muningsih, E., & Kiswati, D. S. (2015). Penerapan Metode *K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang*. *Jurnal Bianglala Informatika*, 3. <https://ejournal.bsi.ac.id/ejurnal/index.php/Bianglala/article/view/570>
- Salsabila, N. (2019). KLASIFIKASI BARANG MENGGUNAKAN METODE *CLUSTERING K-MEANS* DALAM PENENTUAN PREDIKSI STOK BARANG. <http://etheses.uin-malang.ac.id/16985/1/14650031.pdf>
- Nabila, Z., Rahman Isnain, A., & Abidin, Z. (2021). ANALISIS *DATA MINING UNTUK CLUSTERING* KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA *K-MEANS*. *Jurnal Teknologi Dan Sistem Informasi (JTISI)*, 2(2), 100. <http://jim.teknokrat.ac.id/index.php/sisteminformasi/article/view/868>
- Lesmana, A., & Gunawan, W. (2022). PERBANDINGAN ALGORITMA *K-MEANS* DAN *K-MEDOIDS* DALAM *PENCLUSTERAN DATA PENJUALAN PT. UNITED TEKNOLOGI INTEGRASI*. <https://repository.mercubuana.ac.id/60339/8/Cover%20-%20AGUNG%20LESMANA.pdf>

- Setiawan, F. A., Sadikin, M., & Kaburuan, E. R. (2022). Analisis Permasalahan Perangkat Pencetak Menggunakan Metode Algoritma *K-Means* dan *K-Medoids*. *Teknika*, 11(2), 77–84. <https://doi.org/10.34148/teknika.v11i2.471>
- Wulan Sari, R., Wanto, A., & Perdana Windarto, A. (2018). IMPLEMENTASI *RAPIDMINER* DENGAN METODE *K-MEANS* (STUDY KASUS: IMUNISASI CAMPAK PADA BALITA BERDASARKAN PROVINSI). <http://ejurnal.stmik-budidarma.ac.id/index.php/komik/article/view/930/805>
- Yunita Ratna Sari, S., Sudewa, A., Ayu Lestari, D., & Tri Ika, J. (2020). PENERAPAN ALGORITMA *K-MEANS* UNTUK *CLUSTERING* DATA KEMISKINAN PROVINSI BANTEN MENGGUNAKAN *RAPIDMINER*. 5(2), 2502–2714. <https://jurnal.unimed.ac.id/2012/index.php/cess/article/view/18519/pdf>
- Kasini, & Nani, H. (2022). Implementation of *data mining to determine* stock inventory at kenza grocery stores using the *k-means clustering* method. *Jurnal Mantik*, 6(3), 2685–4236. <https://www.iocscience.org/ejournal/index.php/mantik/article/view/3366/2566>
- Hardianto, R., Ramadhan, H., Putra Pane, E., & Yunefri, Y. (2022). *K-Means Clustering* in Determining the Category of Stock Items In Angkasa Mart. *Knowbase : International Journal of Knowledge in Database*, 2(1), 30. <https://doi.org/10.30983/ijokid.v2i1.5411>
- Fahikra, F., Aliska Giopani, L., Fitriah, M., Cindya Dwyne, Z., & Syahidatul Helma, S. (2022). Perbandingan Algoritma *K-Means* dan *K-Medoids* Untuk Pengelompokan Suhu di Provinsi Riau. 2(2), 128–134. <https://journal.irpi.or.id/index.php/ijirse/article/view/434/211>