

## Perbandingan Analisis *Cluster* Algoritma *K-Means* Dan *AHC* Dalam Perencanaan Persediaan Barang Pada Perusahaan Manufaktur

Ellang Putro Priambodo<sup>1\*</sup>, Arief Jananto<sup>2</sup>

Program Studi Sistem Informasi, Universitas Stikubank Semarang

Jl. Tri Lomba Juang No.1 Mugassari, Semarang, Indonesia

\*e-mail Corresponding Author: ellang65.kaell@gmail.com

### Abstract

*Inventory control is very important for PT. Multi Lestari (goods production/sales company), so as not to incur maintenance costs for goods that are produced in excess, or cause losses if customer requests cannot be met. This study tested the K-Means algorithm and the Agglomerative Hierarchical Clustering (AHC) algorithm in classifying past sales data based on sales parameters. Inventory prediction for future sales is done by interpreting the results of the cluster formed by the clustering algorithm used. The cluster center that is formed based on the average number of sold parameters is used as a reference to determine the amount of inventory for each type of item that is included in the group members of a particular cluster. The test results show that both the K-Means algorithm and the AHC algorithm can classify the goods sold based on the similarity of the average number sold. However, both produce different amounts, so further studies are needed on which algorithm produces more accurate inventory predictions, based on real sales results that have been made in the past.*

*Keywords: Data Mining; Grouping; Cluster Center; Inventory Prediction*

### Abstrak

Pengendalian persediaan barang menjadi sangat penting bagi PT. Multi Lestari (perusahaan produksi/penjualan barang), agar tidak menimbulkan biaya pemeliharaan bagi barang yang diproduksi secara berlebihan, atau menimbulkan kerugian jika permintaan pelanggan tidak dapat dipenuhi. Penelitian ini menguji algoritma *K-Means* dan algoritma *Agglomerative Hierarchical Clustering* (AHC) dalam mengelompokkan data penjualan barang masa lalu berdasarkan parameter-parameter penjualan. Prediksi persediaan untuk penjualan masa mendatang dilakukan dengan menginterpretasikan hasil *cluster* yang terbentuk oleh algoritma *clustering* yang digunakan. Pusat *cluster* yang terbentuk berdasarkan parameter *rerata jumlah terjual* digunakan sebagai acuan untuk menentukan *jumlah persediaan* barang bagi setiap jenis barang yang termasuk dalam anggota kelompok sebuah *cluster* tertentu. Hasil uji menunjukkan algoritma *K-Means* maupun algoritma *AHC* keduanya dapat mengelompokkan barang yang terjual berdasarkan kemiripan rerata jumlah terjual. Namun keduanya menghasilkan jumlah yang berbeda, sehingga perlu kajian lebih lanjut mengenai algoritma mana yang menghasilkan prediksi persediaan yang lebih akurat, dengan berpatokan pada Hasil penjualan riil yang telah dilakukan pada masa lalu.

**Kata kunci:** *Data Mining; Pengelompokan; Pusat Klaster; Prediksi Persediaan*

### 1. Pendahuluan

Pengendalian atau perencanaan persediaan merupakan suatu upaya penyediaan sumber daya atau barang-barang yang diperlukan dalam proses produksi atau dalam penjualan. Hal ini bertujuan untuk memastikan proses produksi berjalan atau penjualan menjadi lancar. Perencanaan persediaan juga bertujuan untuk mengurangi munculnya risiko yang dapat terjadi seperti kekurangan barang [1]. Sebaliknya, bila persediaan barang berlebih atau *over stock*, maka akan terjadi peningkatan pada beban biaya penyimpanan dan perawatan barang-barang di dalam gudang, sehingga mengakibatkan pemborosan [2]. Perencanaan persediaan yang baik

akan dapat meminimalisir biaya persediaan, sehingga memberikan keuntungan bagi perusahaan [3].

PT. Multi Lestari adalah perusahaan yang bergerak dibidang manufaktur yang memproduksi barang dengan standar keamanan dan mutu tinggi sesuai dengan Standar Nasional Indonesia (SNI). Ketersediaan barang dalam jumlah yang tepat pada perusahaan ini menjadi sangat penting untuk menghindari produksi barang yang kurang diminati oleh pembeli dalam jumlah yang besar. Pada kenyataannya, sering terjadi kelebihan persediaan barang sebagai akibat tidak diketahuinya barang apa saja yang paling diminati oleh pembeli, dan di sisi lain terdapat beberapa jenis barang yang justru kehabisan persediaan. Diperlukan suatu upaya untuk merencanakan persediaan dengan tepat agar tidak terjadi kelebihan persediaan atau sebaliknya kekurangan persediaan.

Perkembangan ilmu pengetahuan khususnya di dunia komputasi memungkinkan penggunaan model-model komputasi untuk mendukung perencanaan atau peramalan bisnis, termasuk merencanakan persediaan produksi dan persediaan penjualan. Dalam kasus prediksi (baik prediksi secara diskrit maupun prediksi secara linear), terdapat beberapa model komputasi yang dapat digunakan, misalnya algoritma berbasis kluster dan algoritma berbasis klasifikasi [4]. Dalam kelompok algoritma berbasis kluster, terdapat dua algoritma yang sering digunakan, yaitu algoritma *K-Means Clustering* dan algoritma *Agglomerative Hierarchical Clustering* (AHC). Kedua algoritma ini telah digunakan dalam prediksi berbagai urusan bisnis, seperti dalam hal penjualan [5 – 7], kesehatan [8 -10], pendidikan [11, 12], dan berbagai urusan bisnis lainnya [13, 14].

Penelitian kami menguji dan membandingkan kinerja algoritma *K-Means Clustering* dan algoritma AHC dalam memprediksi persediaan produksi/penjualan barang manufaktur, studi kasus pada perusahaan manufaktur PT. Multi Lestari. Hasil perbandingan kinerja kedua algoritma menjadi pertimbangan bagi manajemen perusahaan untuk memiliki algoritma terbaik untuk menjadi alat bantu dalam merencanakan dan mengendalikan persediaan barang.

## 2. Tinjauan Pustaka

Penelitian-penelitian terdahulu yang sejenis, yang telah berupaya menyelesaikan masalah perencanaan persediaan barang menggunakan algoritma komputasi:

- 1) Penelitian Rachmawati, Syafirullah, dan Faiz [15] menggunakan metode EOQ (*Economic Order Quantity*) dan ROP (*Reorder Point*). Metode EOQ bekerja dengan menentukan kuantitas pesanan persediaan dengan cara meminimalkan biaya penyimpanan dan biaya pemesanan persediaan. Model ini dapat mengatasi masalah berkaitan dengan ketidakpastian melalui persediaan pengaman (*safety stock*). Adapun ROP diterapkan untuk memberikan informasi keadaan di mana harus diadakan pesanan lagi sehingga kedatangan atau penerimaan barang yang dipesan adalah tepat pada waktunya. Penerapan model EOQ yang dikombinasikan dengan ROP tidak akan membuat *stock* barang dari *stock opname* tidak menumpuk dan bahkan bisa mengetahui jumlah pemesanan yang optimal untuk dipesan kembali, sehingga dengan menggunakan metode ini para pebisnis tidak perlu repot untuk mengetahui jumlah barang yang akan dipesan atau diorder.
- 2) Penelitian Hayami, Sunanto, dan Oktaviandi [16] menggunakan metode *single exponential smoothing* dalam memprediksi penjualan barang *furniture*. *Single exponential smoothing* bekerja dengan mempertimbangkan bobot data sebelumnya dengan memberikan bobot pada setiap data periode untuk membedakan prioritas atas suatu data, digunakan pada peramalan jangka pendek yang biasanya hanya 1 bulan ke depan yang mengasumsikan bahwa data berfluktuasi di sekitar nilai mean yang tetap tanpa trend atau pola pertumbuhan konsisten. Nilai prediksi ditentukan dari nilai alpha yang paling cocok dari perhitungan kesalahan prediksi hingga menghasilkan nilai yang paling kecil. Pada penelitian tersenut, data yang diuji merupakan data penjualan 4 bulan terakhir untuk memprediksi penjualan bulan berikutnya. Hasil prediksi penjualan menjadi dasar untuk menentukan jumlah persediaan.
- 3) Penelitian Virrayyani dan Sutikno [17] menggunakan metode *Adaptive Neuro-Fuzzy Inference System* (ANFIS). Pada penelitian tersebut, ANFIS diimplementasikan dalam sebuah aplikasi sistem prediksi penjualan barang. Prosedur prediksi menggunakan analisis runtun waktu. Aturan ANFIS menggunakan model *fuzzy* Takagi-Sugeno dan fungsi keanggotaan tipe *generalized bell* dengan 2 data masukan untuk 1 data target. Pelatihan dan pengujian ANFIS untuk semua produk menghasilkan *Mean Absolute Percentage* (MAPE) rata-rata pelatihan sebesar 8,73 % dan MAPE rata-rata pengujian sebesar 13,58%. Hasil

MAPE pengujian tersebut kurang dari batas toleransi *error*, yaitu 20 %. Batas toleransi tersebut berdasarkan penafsiran Batey dan Friedrich di mana  $MAPE < 10\%$  merupakan perkiraan yang sangat baik dan  $10\% < MAPE < 20\%$  merupakan perkiraan yang baik. Hasil prediksi penjualan menjadi dasar untuk menentukan jumlah persediaan.

- 4) Penelitian Alfani, Rozi, dan Sukmana [19] menggunakan metode *K-Nearest Neighbor* (K-NN) untuk memprediksi penjualan produk. Konsep kerja K-NN dalam memprediksi penjualan adalah dengan menganalisis penjualan produk berdasarkan status penjualan (tidak laris, laris, dan sangat laris) berbasis teknik klasifikasi berdasarkan parameter jumlah terjual. Jumlah terjual berdasarkan kelompok barang digunakan sebagai acuan untuk menentukan jumlah persediaan setiap kelompok barang.

Berbeda dengan metode-metode yang digunakan pada penelitian terdahulu yang kami tinjau, konsep perencanaan persediaan barang yang kami lakukan menggunakan teknik *llustering* dengan menguji 2 algoritma klustering (Algoritma *K-means Clustering* dan algoritma *Agglomerative Hierarchical Clustering* (AHC)). Prinsip kerja dalam penentuan persediaan barang berbasis kluster yang kami usulkan (*state of the art*) adalah: mengelompokkan berbagai jenis item barang yang diproduksi/dijual menjadi sejumlah tertentu kelompok (*cluster*) untuk mengetahui item barang apa saja yang terjual dalam jumlah *paling banyak*, terjual dengan jumlah *banyak*, terjual dengan jumlah *sedang*, terjual dengan jumlah *sedikit*, dan terjual dengan jumlah *paling sedikit*, sehingga untuk jumlah persediaan ditentukan berdasarkan kelompok item barang pada *cluster* masing-masing. Setiap kelompok menghasilkan pusat *cluster* berdasarkan parameter tertentu, salah satu diantaranya adalah pusat *cluster* berdasarkan parameter *jumlah terjual*. Pusat *cluster* untuk parameter *jumlah terjual* digunakan sebagai acuan untuk menentukan jumlah persediaan setiap kelompok barang.

### 3. Metodologi

#### 3.1 Algoritma K-Means

Algoritma *K-Means* merupakan algoritma pengelompokan iteratif yang melakukan partisi set data ke dalam sejumlah *k cluster* yang sudah ditetapkan di awal. Algoritma *K-Mean Clustering* disajikan berikut [4]:

- 1) Menentukan jumlah kluster (*k*) yang diinginkan pada dataset
- 2) Menentukan pusat kluster awal (*Centroid*) dengan mengambil nilai terkecil, rata-rata dan nilai terbesar.
- 3) Menghitung jarak terdekat disetiap data dengan *Centroid*. Menghitung jarak terdekat dengan *Centroid* digunakan rumus *Euclidean distance*. Rumus bisa dilihat dibawah ini:

$$d(x_i, \mu_j) = \sqrt{(x_i - \mu_j)^2} \dots \dots \dots (1)$$

Keterangan:

- $x_i$  : Data kriteria
- $\mu_j$  : *Centroid* pada *cluster* ke *j*

- 4) Menghitung lagi pusat *Cluster* dengan anggota *Cluster* yang sekarang. Adapun rumusnya bisa dilihat dibawah ini:

$$\mu_j(t + 1) = \frac{1}{N_{sj}} \sum_j \epsilon_{sj} x_j \dots \dots \dots (2)$$

Keterangan:

- $\mu_j(t + 1)$  : *Centroid* baru pada iterasi ke 1
- $N_{sj}$ : Banyak data pada *cluster* *sj*;

#### 3.2 Algoritma Agglomerative Hierarchical Clustering (AHC)

Pengelompokan hirarki (*Hierarchical Clustering*) adalah metode dengan cara melakukan pengelompokan objek-objek data kedalam sebuah kelompok hirarki. Metode *Agglomerative* (metode penggabungan) adalah strategi pengelompokan hirarki yang dimulai dengan setiap objek dalam satu cluster yang terpisah kemudian membentuk *cluster* yang semakin membesar. Jadi, banyaknya cluster awal adalah sama dengan banyaknya objek [4]. Algoritma AHC disajikan berikut:

- 1) Menghitung matriks jarak *Euclidean* (seperti pada Formula 1).
- 2) Menggabungkan dua cluster terdekat.  
Jika jarak objek a dengan b memiliki nilai jarak paling kecil dibandingkan jarak antar objek lainnya dalam matriks jarak *Euclidean*, maka gabungan dua *cluster* pada tahap pertama adalah  $d_{ab}$ .
- 3) Memperbarui matriks jarak sesuai dengan teknik pengelompokan *Agglomerative method*. Jika  $d_{ab}$  adalah jarak terdekat dari matriks jarak *Euclidean*, maka rumus untuk metode *agglomerative* adalah:

$$d_{(ab)c} = \min \{d_{a,c}; d_{b,c}\} \quad d_{(ab)c} = \text{average} \{d_{a,c}; d_{b,c}\} \quad d_{(ab)c} = \max \{d_{a,c}; d_{b,c}\}$$

*rumus single linkage*
*rumus average linkage*
*rumus complete linkage*

- 4) Mengulangi langkah 2 dan 3 sampai hanya tersisa satu cluster
- 5) Menyusun Dendrogram

**3.3 Tahapan Penelitian berbasis CRISP-DM**

CRISP-DM (*Cross-Industry Standard Process for Data Mining*) menyajikan prosedur atau tahapan-tahapan ril yang dilalui dalam upaya pencapaian tujuan akhir yang diharapkan dalam riset/pengkajian [19]. Beberapa tahapan utama dalam prosedur CRISP-DM yang digunakan adalah sebagai berikut:

- 1) *Business Understanding*  
Pada penelitian datamining ini bertujuan membandingkan hasil *cluster* algoritma *K-Means clustering* dan *Hierarchical clustering* pada data stok barang.
- 2) *Data Understanding*  
Data yang dipakai merupakan data stok barang pada PT Multi Lestari periode bulan Februari 2021 sampai bulan Juni 2021. potongan dataset asli PT Multi Lestari dapat dilihat pada Gambar 1.

Gambar 1. Potongan Dataset Produksi/penjualan pada PT. Multi Lestari

- 3) *Data Preparation*  
Pada tahap ini data akan di olah kembali karena pada data asli tidak semua data dan atribut akan di pakai pada penelitian. Berikut beberapa tahapan pada fase persiapan data:
  - a. Pemilihan Atribut
  - b. Pada tahap ini dilakukan seleksi pada atribut mana yang akan digunakan pada penelitian. Atribut hasil seleksi dapat dilihat pada Tabel 1.

*Tabel 1 Atribut Terseleksi*

No	Atribut
1.	Nama_barang
2.	Stok_awal
3.	Stok_masuk
4.	Stok_keluar
5.	Stok_akhir

c. Pembersihan Data

Pada tahapan ini yaitu proses pembersihan data yaitu dengan menghapus data yang tidak lengkap, kosong, ataupun hilang dan menghapus data barang lama yang sudah di pakai.

d. Transformasi Data

Pada tahap ini akan di lakukan tranformasi data pada data *Excel* untuk mempermudah proses *clustering*.

Hasil transformasi atribut terseleksi (Atribut Stok\_Awal, Barang\_Masuk, Barang\_Keluar, Stok\_Akhir) disajikan pada Tabel 2.

Tabel 2 Tranformasi Atribut Hasil seleksi

atribut	Jumlah transaksi	klasifikasi	Konversi data
Stok_awal	10.000 lebih	Tinggi	4
Barang_masuk	5000-10.000	Sedang	3
Barang_keluar	1-5000	Rendah	2
Stok_akhir	0	kosong	1

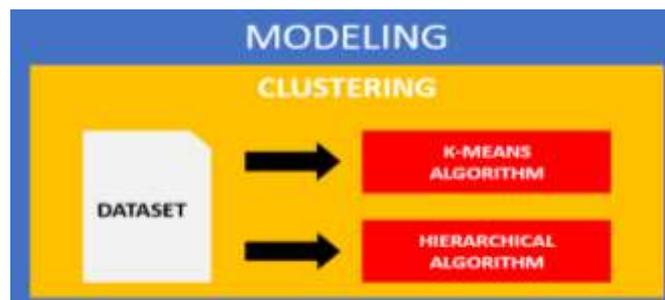
Atribut Jenis barang merupakan atribut baru yang di buat untuk mengelompokan data pada tiap jenis barang. Dengan klasifikasi menjadi 7 kelompok yang dapat dilihat pada Tabel 3.

Tabel 3 Konversi Data Jenis\_barang

Jenis barang	Konversi data
packing	1
logo	2
blister	3
etiket	4
duplex	5
masterbox	6
komponen	7

4) Fase Pemodelan (Modeling)

Pada tahap pemodelan ini akan menggunakan dua algoritma data mining yaitu algoritma *K-Means Clustering* dan *Hierarchical Clustering* dengan *tool* datamining *R studio* yang kemudian akan didapatkan hasil klister pada tiap stok barang. Model aplikasi dapat dilihat pada Gambar 2.



Gambar 2. Model Proses Sistem

5) Evaluation

Pada tahap ini melakukan evaluasi terhadap hasil dari proses data mining yang dihasilkan pada tahap pemodelan sebelumnya. Evaluasi apakah metode clustering yang menggunakan algoritma *K-means clustering* dan *Hierarchical Clustering* mendapat hasil yang sesuai dengan tujuan awal penelitian.

## 4. Hasil dan Pembahasan

### 4.1 Install package Rstudio

Beberapa *packages* yang diperlukan pada implementasi algoritma *K-Means* dan *Hierarchical Clustering* agar dapat berjalan pada Rstudio yang perlu diinstall sebagai berikut:

```
Install.package("readxl")
library(readxl)
```

*Packages* yang berfungsi untuk membaca file *excel* pada Rstudio.

```
Install.package("tidyverse")
library(tidyverse)
```

*Packages* yang berisi *package* lain seperti *reader*, *ggplot2*, *tidyr*, *dplyr* dan lainnya. beberapa fungsi seperti *reader* untuk membaca dan mengimport file, *ggplot2* untuk sebagai kebutuhan visualisasi dan *packages* lain sebagai kebutuhan proses *clustering*.

```
Install.package("cluster")
library(cluster)
```

*Packages* yang berfungsi untuk menghitung dan memproses *clustering*.

```
Install.package("factoextra")
library(factoextra)
```

*Packages* yang berfungsi untuk memvisualisasikan hasil *clustering*.

```
library(writexl)
```

*Packages* yang berfungsi untuk mengexport hasil *clustering* dalam bentuk file *excel*.

### 4.2 Data Preparation

Tahap pertama adalah mempersiapkan data yang akan diolah pada Rstudio. Berikut *Source code* pada proses data preparation:

*Source code* yang berfungsi untuk membaca dan mengimport file "februari\_fx" yang berupa file *excel* yaitu data stok barang bulan februari kemudian ditempatkan pada variabel "data\_februari":

```
data_februari <- read_excel("februari_fx.xlsx")
```

*Source code* yang berfungsi melihat variabel *data\_februari* yang telah dibuat sebelumnya. Data bulan februari berjumlah 87 *record* dengan 6 kolom. Hasil import file dapat dilihat pada Gambar 3.

```
View(data_februari)
```

	Nama_Barang	Stok_awal	Stok_masuk	Stok_keluar	Stok_akhir	Jenis_barang
1	Hose Clamp 3/4	4	1	4	4	1
2	Karet Hek	4	1	4	4	1
3	Seang Hitam	5	4	4	2	1
4	TRAJI PACONG	2	2	2	2	1
5	HOSE CLAMP MODENA 3/4	2	2	2	2	1
6	LOGO ALLISON 20	2	1	2	2	2
7	LOGO LUXURY 27	2	1	2	2	2
8	LOGO LUXURY 20	2	1	2	2	2
9	LOGO KAZUO 20	2	1	2	2	2
10	LOGO ZAMAN 20	2	2	2	2	2
11	LOGO QATAR 20	2	1	2	2	2
12	LOGO QUARDON 20	2	1	2	2	2
13	LOGO GASCOMP 20	2	4	3	4	2
14	LOGO GASCOMP 23	2	1	2	2	2
15	LOGO SUPERLOCK 20	4	1	2	4	2

Gambar 2. Potongan Dataset Data Stok Barang Bulan Februari

*Source code* yang berfungsi melihat rincian data pada variabel *data\_februari*. Tampilan rincian data dapat dilihat pada Gambar 4.

```
summary(data_februari)
```

```
> summary(data_februari)
  Nama_Barang      Stok_awal      Stok_masuk      Stok_keluar
Length:87      Min.   :2.000      Min.   :1.000      Min.   :1.000
Class :character 1st Qu.:2.000      1st Qu.:1.000      1st Qu.:2.000
Mode  :character Median :2.000      Median :1.000      Median :2.000
                Mean  :2.862      Mean  :1.494      Mean  :2.598
                3rd Qu.:4.000      3rd Qu.:1.000      3rd Qu.:4.000
                Max.  :4.000      Max.  :4.000      Max.  :4.000

  Stok_akhir      Jenis_barang
Min.   :2.000      Min.   :1.000
1st Qu.:2.000      1st Qu.:2.000
Median :2.000      Median :5.000
Mean   :2.851      Mean   :4.517
3rd Qu.:4.000      3rd Qu.:7.000
Max.   :4.000      Max.   :7.000
```

Gambar 3. Rincian Data Bulan Februari.

Source code yang berfungsi untuk menghapus jika terdapat nilai na's atau data yang kosong maupun hilang.

```
februari <- na.omit(data_februari)
```

Source code yang berfungsi untuk mengambil data numerik yang berada pada kolom kedua sampai kolom kelima.

```
februari_data.numerik <- februari[2:5]
```

Source code yang berfungsi untuk transformasi skala data\_februari. Transformasi data numerik dilakukan dengan standarisasi data. Jika rentang nilai antar variabel memiliki perbedaan skala yang cukup besar yang dapat menyebabkan bias dalam analisis cluster. Hasil tranformasi skala dapat dilihat pada Gambar 5.

```
februari_datafix <- scale(februari_data.numerik)
```

	Stok_awal	Stok_masuk	Stok_keluar	Stok_akhir
1	1.2238827	-0.4838248	1.5442117	1.2142905
2	1.2238827	-0.4838248	1.5442117	1.2142905
3	0.1483494	2.4528794	1.5442117	-0.8985750
4	-0.9271838	0.4950766	-0.6581886	-0.8985750
5	-0.9271838	0.4950766	-0.6581886	-0.8985750
6	-0.9271838	-0.4838248	-0.6581886	-0.8985750
7	-0.9271838	-0.4838248	-0.6581886	-0.8985750
8	-0.9271838	-0.4838248	-0.6581886	-0.8985750
9	-0.9271838	-0.4838248	-0.6581886	-0.8985750
10	-0.9271838	0.4950766	-0.6581886	-0.8985750
11	-0.9271838	-0.4838248	-0.6581886	-0.8985750
12	-0.9271838	-0.4838248	-0.6581886	-0.8985750
13	-0.9271838	2.4528794	0.4430116	1.2142905
14	-0.9271838	-0.4838248	-0.6581886	-0.8985750
15	1.2238827	-0.4838248	-0.6581886	1.2142905

Gambar 4. Potongan Hasil Transformasi Skala.

### 4.3 Eksekusi Menggunakan Algoritma K-means

Dalam proses clustering jumlah cluster yang dibentuk yaitu tiga cluster. Berikut Source code yang digunakan dalam proses perhitungan algoritma K-Mean pada Rstudio.

Source code yang berfungsi untuk memproses klasterisasi data\_februari dengan jumlah tiga cluster. Hasil clustering kemudian akan ditempatkan pada variabel "februari\_final".

```
februari_final <- kmeans(februari_datafix, centers = 3)
```

*Source code* yang berfungsi untuk menampilkan hasil proses clustering pada proses sebelumnya. Hasil perhitungan dapat dilihat pada Gambar 6 dan *cluster vector* dan *cluster sum of squares* dapat dilihat pada Gambar 7.

```
print(februari_final)
K-means clustering with 3 clusters of sizes 16, 46, 25

Cluster means:
  Stok_awal Stok_masuk Stok_keluar Stok_akhir
1  1.0222202 -0.4838248 -0.4517136  0.9501823
2 -0.8336592 -0.2710202 -0.6581886 -0.8756090
3  0.8797120  0.8083250  1.5001637  1.0030040
```

Gambar 5 Tampilan *Cluster Means*.

Pada Gambar 5 dapat diketahui jumlah *cluster* yang terbentuk berjumlah tiga *cluster* dengan anggota *cluster* pertama berjumlah 16, anggota *cluster* kedua berjumlah 46 dan anggota *cluster* ketiga berjumlah 25. *Cluster mean* dari setiap atribut dikelompokkan menjadi tiga *cluster*.

```
Clustering vector:
[1] 3 3 3 2 2 2 2 2 2 2 2 2 2 2 3 2 1 2 2 2 2 1 2 1 1 2 2 2 2 2 2 3 1 2 2 2
[35] 2 1 2 1 1 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
[69] 3 3 3 3 2 2 3 3 2 3 1 1 1 1 1 1 1 1 2 1

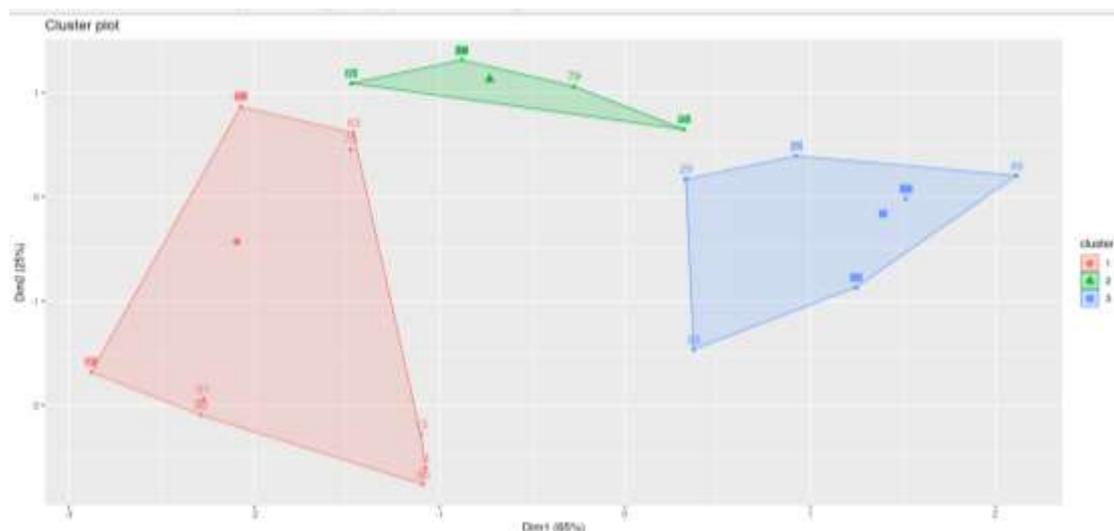
Within cluster sum of squares by cluster:
[1]  9.123596 17.157631 71.905629
(between_SS / total_SS = 71.5 %)
```

Gambar 6. Tampilan *Clustering Vector* dan *Cluster Sum of Squares*.

Pada Gambar 7 dapat diketahui bahwa pengelompokan dengan 86 *record* pada tiap *cluster* dan ditampilkan pada sebuah *clustering vector*. Nilai sum of squares pada *cluster* satu 9.123596, *cluster* kedua bernilai 17.157631 dan *cluster* ketiga bernilai 71.905629. Dengan nilai perhitungan pembagian *between sum of squares* 71.5%.

*Source code* yang berfungsi untuk visualisasi hasil analisa perhitungan *K-mean* dalam grafik. Hasil visualisasi dapat dilihat pada Gambar 8.

```
fviz_cluster(februari_final,data=februari_datafix)
```



Gambar 7. Grafik *Plot Visualisasi K-means*.

*Source code* yang berfungsi untuk menampilkan hasil *clustering* data dalam tabel. Hasil visualisasi *K-Means* dalam bentuk tabel dapat dilihat pada gambar 9.

```
februari_finalakhir=data.frame(februari,februari_final$cluster)
view(februari_finalakhir)
```

↑	Nama_Barang	Stok_awal	Stok_masuk	Stok_keluar	Stok_akhir	Jenis_barang	februari_final.cluster
1	Hose Clamp 3/4	4	1	4	4	1	3
2	Karet Hek	4	1	4	4	1	3
3	Selang Hitam	3	4	4	2	1	3
4	TALI PACKING	2	2	2	2	1	2
5	HOSE CLAMP MODENA 3/4	2	2	2	2	1	2
6	LOGO ALLISON 20	2	1	2	2	2	2
7	LOGO LUXURY 27	2	1	2	2	2	2
8	LOGO LUXURY 20	2	1	2	2	2	2
9	LOGO KAZUKI 20	2	1	2	2	2	2
10	LOGO ZAMAN 20	2	2	2	2	2	2
11	LOGO QATAR 20	2	1	2	2	2	2
12	LOGO QUARDON 20	2	1	2	2	2	2
13	LOGO GASCOMP 20	2	4	3	4	2	3
14	LOGO GASCOMP 23	2	1	2	2	2	2
15	LOGO SUPERLOCK 20	4	1	2	4	2	1

Gambar 8 Visualisasi K-Means Clustering dalam Bentuk Tabel.

Source code yang berfungsi untuk menampilkan ringkasan rata-rata dari masing-masing cluster pada tiap atribut. Hasil dapat dilihat pada Gambar 10.

```
februari[2:5] %>%
  mutate(Cluster=februari_final$cluster) %>%
  group_by(Cluster) %>%
  summarise_all("mean")
```

	Cluster	Stok_awal	Stok_masuk	Stok_keluar	Stok_akhir
	<int>	<dbl>	<dbl>	<dbl>	<dbl>
1	1	3.81	1	2.19	3.75
2	2	2.09	1.22	2	2.02
3	3	3.68	2.32	3.96	3.8

Gambar 9 Ringkasan Hasil Clustering.

Source code yang berfungsi untuk mengexport hasil akhir K-Means clustering dalam bentuk file excel.

```
write_xlsx(x = februari_finalakhir, path = "kmean_februari_final.xlsx", col_names = TRUE)
```

#### 4.4 Eksejusi Algoritma Hierarchical

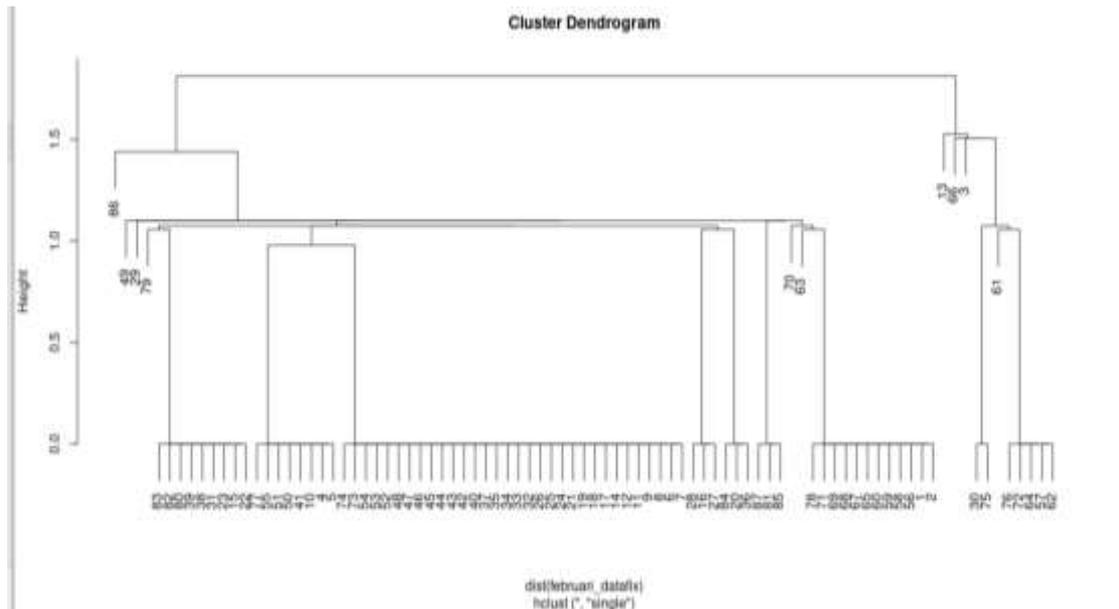
Dalam proses clustering jumlah cluster yang dibentuk yaitu tiga cluster. Berikut sourcecode yang digunakan dalam proses perhitungan algoritma Hierarchical pada Rstudio.

Sourcecode yang berfungsi untuk memproses klasterisasi dengan algoritma Hierarchical metode single link pada data\_februari. Hasil clustering kemudian akan ditempatkan pada variabel "hc\_februari".

```
hc_februari <- hclust(dist(februari_datafix), method="single")
```

Sourcecode yang berfungsi untuk menampilkan hasil grafik plot algoritma Hierarchical atau dendrogram berdasarkan atribut "No" pada data\_februari. Hasil dapat dilihat pada Gambar 11.

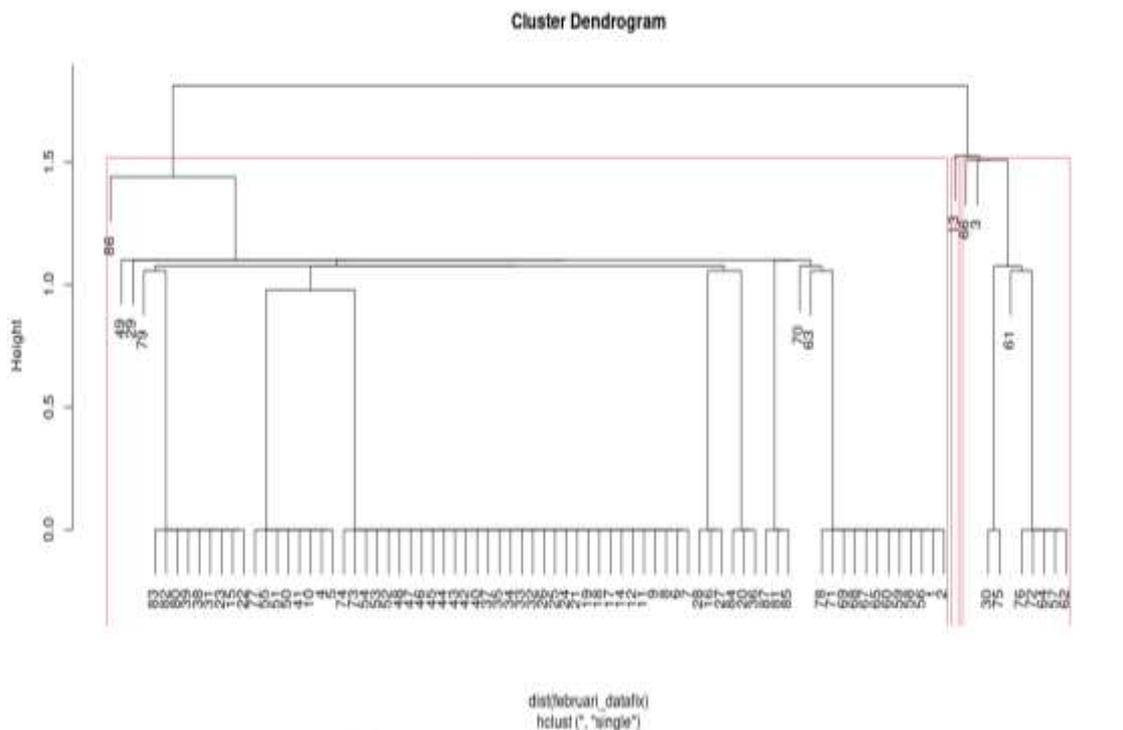
```
plot(hc_februari)
```



Gambar 10. Dendrogram *Hierarchical Clustering*.

Source code yang berfungsi untuk membentuk 3 cluster. Hasil dapat dilihat pada Gambar 12.

```
rect.hclust(hc_februari,k=3)
groups<-cutree(hc_februari,k=3)
```



Gambar 11. Dendrogram *Hierarchical Clustering* dengan Tiga Cluster.

Source code yang berfungsi untuk menampilkan hasil clustering data dalam tabel. Hasil visualisasi *Hierarchical Clustering* dalam bentuk tabel dapat dilihat pada gambar 13.

```
februari_finalakhir=data.frame(februari, final)
view(februari_finalakhir)
```

	Nama_Barang	Stok_awal	Stok_masuk	Stok_keluar	Stok_akhir	Jenis_barang	groups
1	Hose Clamp 3/4	4	1	4	4	1	1
2	Karet Hek	4	1	4	4	1	1
3	Selang Hitam	3	4	4	2	1	2
4	TALI PACKING	2	2	2	2	1	1
5	HOSE CLAMP MODENA 3/4	2	2	2	2	1	1
6	LOGO ALLISON 20	2	1	2	2	2	1
7	LOGO LUXURY 27	2	1	2	2	2	1
8	LOGO LUXURY 20	2	1	2	2	2	1
9	LOGO KAZUKI 20	2	1	2	2	2	1
10	LOGO ZAMAN 20	2	2	2	2	2	1
11	LOGO QATAR 20	2	1	2	2	2	1
12	LOGO QUARDON 20	2	1	2	2	2	1
13	LOGO GASCOMP 20	2	4	3	4	2	3
14	LOGO GASCOMP 23	2	1	2	2	2	1
15	LOGO SUPERLOCK 20	4	1	2	4	2	1

Showing 1 to 15 of 87 entries. 7 total columns

Gambar 12. Visualisasi *Hierarchical Clustering* dalam Bentuk Tabel.

Source code yang berfungsi untuk meng-export hasil akhir *Hierarchical clustering* dalam bentuk file excel.

```
write_xlsx(x = februari_finalakhir, path = "kmean_februari_final.xlsx", col_names = TRUE)
```

#### 4.5 Pembahasan

Setelah melakukan perhitungan *Algoritma K-Means* dan *Hierarchical* menggunakan data bulan Februari sampai bulan Juni dengan Rstudio diperoleh hasil akhir *cluster* seperti pada Tabel 4.

Tabel 4 Hasil *Clustering* Berdasarkan Jumlah Setiap Anggota *Cluster*.

Bulan	Cluster	K-mean	Hierarchical
Februari	1	16	76
	2	46	10
	3	25	1
Maret	1	27	86
	2	8	18
	3	72	3
April	1	27	83
	2	29	20
	3	48	1
Mei	1	24	22
	2	53	83
	3	30	2
Juni	1	25	14
	2	92	133
	3	32	2

Detail data barang yang termasuk dalam setiap *cluster* yang terbentuk disajikan pada Gambar 14.

Bulan	Cluster	K-Means									Hirarchical								
		packing	logo	blister	etiket	duplex	masterbox	komponen	jumlah	packing	logo	blister	etiket	duplex	masterbox	komponen	jumlah		
Februari	1	0	2	2	2	2	0	8	16	11	15	3	11	4	8	24	76		
	2	9	13	1	9	2	8	4	46	1	0	0	1	0	0	8	10		
	3	3	1	0	1	0	0	20	25	0	1	0	0	0	0	0	1		
Maret	1	3	1	1	1	0	0	21	27	11	15	2	11	10	9	28	86		
	2	0	1	0	3	0	0	4	8	2	0	1	1	0	0	14	18		
	3	10	15	2	8	10	9	18	72	0	1	0	0	0	0	2	3		
April	1	0	2	1	1	3	0	20	27	10	11	2	9	12	10	29	83		
	2	2	3	1	1	2	0	20	29	1	0	1	1	0	0	17	20		
	3	9	7	1	8	7	10	6	48	0	1	0	0	0	0	0	1		
Mei	1	2	1	1	0	0	0	20	24	1	0	1	0	0	0	20	22		
	2	9	13	0	10	8	10	3	53	10	15	1	12	11	10	24	83		
	3	1	2	1	2	3	0	21	30	1	1	0	0	0	0	0	2		
Juni	1	0	2	2	3	5	0	13	25	1	0	0	1	0	0	12	14		
	2	7	16	3	28	21	13	4	92	9	19	6	30	27	13	29	133		
	3	3	1	1	1	1	0	25	32	0	0	0	1	0	0	1	2		

Gambar 13. Hasil Rekapitulasi Jumlah Anggota *Clustering* Berdasarkan Jenis Barang.

Menurut Prasetyo [20] algoritma *clustering* tergolong dalam algoritma deskriptif, yang berarti bahwa hasil proses (pembentukan *cluster*) tidak akan berarti (bermakna) jika tidak dilakukan pemaknaan (interpretasi). Dalam hal ini, kami melakukan pemaknaan (interpretasi) untuk menterjemahkan bagaimana hasil klaster dapat mencerminkan perencanaan persediaan.

Berdasarkan *summary* hasil *cluster* Tabel 4, dapat diinterpretasikan hal-hal berikut:

1) Pada bulan Februari:

- Rerata penjualan barang yang tergolong dalam *Cluster* 1 sebesar 16 buah berdasarkan prediksi algoritma *K-Means*, dan 76 buah berdasarkan prediksi algoritma *Hirarchical*. Ini berarti bahwa jumlah persediaan yang harus dipersiapkan untuk barang-barang yang masuk dalam kelompok (*cluster* 1) adalah 16 buah jika berpatokan pada algoritma *K-Means* atau 76 buah jika berpatokan pada algoritma *Hirarchical*.
- Rerata penjualan barang yang tergolong dalam *Cluster* 2 sebesar 46 buah berdasarkan prediksi algoritma *K-Means*, dan 10 buah berdasarkan prediksi algoritma *Hirarchical*. Ini berarti bahwa jumlah persediaan yang harus dipersiapkan untuk barang-barang yang masuk dalam kelompok (*cluster* 2) adalah 46 buah jika berpatokan pada algoritma *K-Means* atau 10 buah jika berpatokan pada algoritma *Hirarchical*.
- Rerata penjualan barang yang tergolong dalam *Cluster* 3 sebesar 25 buah berdasarkan prediksi algoritma *K-Means*, dan 1 buah berdasarkan prediksi algoritma *Hirarchical*. Ini berarti bahwa jumlah persediaan yang harus dipersiapkan untuk barang-barang yang masuk dalam kelompok (*cluster* 3) adalah 25 buah jika berpatokan pada algoritma *K-Means* atau 1 buah jika berpatokan pada algoritma *Hirarchical*.

2) Pada bulan Maret:

- Rerata penjualan barang yang tergolong dalam *Cluster* 1 sebesar 27 buah berdasarkan prediksi algoritma *K-Means*, dan 86 buah berdasarkan prediksi algoritma *Hirarchical*. Ini berarti bahwa jumlah persediaan yang harus dipersiapkan untuk barang-barang yang masuk dalam kelompok (*cluster* 1) adalah 27 buah jika berpatokan pada algoritma *K-Means* atau 86 buah jika berpatokan pada algoritma *Hirarchical*.
- Rerata penjualan barang yang tergolong dalam *Cluster* 2 sebesar 8 buah berdasarkan prediksi algoritma *K-Means*, dan 18 buah berdasarkan prediksi algoritma *Hirarchical*. Ini berarti bahwa jumlah persediaan yang harus dipersiapkan untuk barang-barang yang masuk dalam kelompok (*cluster* 2) adalah 8 buah jika berpatokan pada algoritma *K-Means* atau 18 buah jika berpatokan pada algoritma *Hirarchical*.
- Rerata penjualan barang yang tergolong dalam *Cluster* 3 sebesar 72 buah berdasarkan prediksi algoritma *K-Means*, dan 3 buah berdasarkan prediksi algoritma *Hirarchical*. Ini berarti bahwa jumlah persediaan yang harus dipersiapkan untuk barang-barang yang masuk dalam kelompok (*cluster* 3) adalah 72 buah jika berpatokan pada algoritma *K-Means* atau 3 buah jika berpatokan pada algoritma *Hirarchical*.

Formulasi hal serupa berlaku bagi data pada bulan April, Mei, dan Juni.

Berdasarkan hasil interpretasi, algoritma *K-Means* dan algoritma *Hirarchical* telah menghasilkan rekomendasi jumlah persediaan barang untuk proses produksi/penjualan di masa mendatang, dengan berpatokan pada rerata nilai setiap pusat *cluster* yang terbentuk. Hanya saja, jumlah persediaan barang yang direkomendasikan untuk setiap bulannya (selama 5 bulan data

yang diuji) mengalami perbedaan, sehingga muncul beberapa pertanyaan yang masih perlu dikaji lebih lanjut, yaitu: pada percobaan bulan apa/berapa menghasilkan prediksi yang lebih mencerminkan keadaan sebenarnya?

Selanjutnya, hasil yang diperoleh oleh algoritma *K-Means* berbeda dengan hasil yang diperoleh oleh algoritma *Hierarchical*. Pertanyaan yang muncul adalah: algoritma mana yang menghasilkan prediksi yang lebih akurat? Pertanyaan ini membuka peluang tindak lanjut di masa mendatang, berupa penelitian lanjutan untuk menguji akurasi masing-masing algoritma, agar manajemen perusahaan dapat menentukan pilihan penggunaan algoritma yang lebih akurat dalam merencanakan persediaan barang.

## 5. Simpulan

Baik algoritma *K-Means* maupun algoritma *Hierarchical* keduanya dapat mengelompokkan barang-barang yang terjual berdasarkan kemiripan rerata jumlah yang terjual pada setiap bulannya. Rerata jumlah yang terjual untuk setiap kelompok barang menjadi acuan dalam mempersiapkan persediaan untuk penjualan di masa mendatang. Ke 2 algoritma ini menghasilkan jumlah barang yang berbeda pada setiap kelompok yang tercipta, sehingga sebelum digunakan sebagai acuan dalam perencanaan persediaan, perlu kajian lebih lanjut di masa mendatang mengenai algoritma mana yang menghasilkan proses prediksi persediaan yang lebih akurat, dengan berpatokan pada Hasil penjualan ril yang telag dilakukan di masa lalu.

## Daftar Referensi

- [1] A.R. Khorida, "Analisis Perencanaan Dan Pengendalian Persediaan Barang Jadi Pada PT. Pardic Jaya Chemicals". *Balance Vocation Accounting Journal*, vol. 1, no. 1, pp. 1-14, 2017.
- [2] E. Khikmawati, M. Anggraini, and K. Anwar, "Analisis Perencanaan Biaya Persediaan Produk Semen Melalui Pendekatan Perencanaan Kebutuhan Bahan Baku (Material Requirement Planning)". *Jurnal Rekayasa, Teknologi, dan Sains*, vol. 1, no. 1, pp. 28-35, 2017.
- [3] H. Chrisna, "Analisis manajemen persediaan dalam memaksimalkan pengendalian internal persediaan pada pabrik sepatu ferradini Medan". *Jurnal Akuntansi Bisnis dan Publik*, vol. 8, no. 2, pp. 82-92, 2018.
- [4] E. Prasetyo, *Data Mining*, Yogyakarta: Penerbit ANDI, 2014.
- [5] S. Butsianto and N.T. Mayangwulan, "Penerapan Data Mining Untuk Prediksi Penjualan Mobil Menggunakan Metode K-Means Clustering". *J. Nas. Komputasi dan Teknol. Inf*, vol. 3, no. 3, pp. 187-201, 2020.
- [6] D. Anggarwati, O. Nurdian, I. Ali, and D.A. Kurnia, "Penerapan Algoritma K-Means Dalam Prediksi Penjualan Karoseri". *Jurnal Data Science & Informatika*, vol. 1, no. 2, pp. 58-62, 2021.
- [7] D. Swanjaya, "Integrasi Metode Agglomerative Hierarchical Clustering dan Backpropagation Pada Model Peramalan Penjualan". In *Prosiding SEMNAS INOTEK (Seminar Nasional Inovasi Teknologi)*, vol. 5, no. 3, pp. 132-141, 2021.
- [8] M. W. Goni, D. Gustian, and F. Sembiring, "Implementasi K-Means Dalam Pengelompokan Penyebaran COVID-19 di Jawa Barat". *Progresif: Jurnal Ilmiah Komputer*, vol. 17, no. 2, pp. 107-118, 2021.
- [9] M.F.I. Al-Rizkil. Widaningrum, and G.A. Buntoro, "Prediksi Penyebaran Penyakit TBC dengan Metode K-Means Clustering Menggunakan Aplikasi Rapidminer". *JTERA (Jurnal Teknol. Rekayasa)*, vol. 5, no. 1, pp. 1-10, 2020.
- [10] A. Jananto, "Penerapan Algoritma K-Means Clustering Untuk Perencanaan Kebutuhan Obat di Klinik Citra Medika". *Progresif: Jurnal Ilmiah Komputer*, vol. 18, no. 1, pp. 69-76, 2022.
- [11] R. Helilintar and I.N. Farida, "Penerapan Algoritma K-Means Clustering Untuk Prediksi Prestasi Nilai Akademik Mahasiswa". *Jurnal sains dan Informatika*, vol. 4, no. 2, pp. 80-87, 2018.
- [12] H. Priyatman, F. Sajid, and D. Haldivany, "Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa". *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, vol. 5, no. 1, pp. 62-66, 2019.
- [13] W. Purba, W. Siawin, M.N. Nababan, N.P. Dharshinni, and S. Aisyah, "Implementasi Data Mining Untuk Pengelompokan Dan Prediksi Karyawan Yang Berpotensi Phk Dengan Algoritma K-Means Clustering". *Jurnal Sistem Informasi Dan Ilmu Komputer Prima (Jusikom Prima)*, vol. 2, no. 2, pp. 85-90, 2019.

- 
- [14] R.M. Sari and V. Tasril, "Prediksi Jumlah APBD Kota Payakumbuh dengan metode K-Means". *Jurnal Ipteks Terapan*, vol. 14, no. 1, pp. 45-50, 2020.
- [15] S.A. Rachmawati, L. Syafirullah, and M.N. Faiz, "Perancangan Sistem Pengendalian Persediaan Barang Menggunakan Metode EOQ dan ROP Berbasis Web". In *Prosiding Seminar Nasional Terapan Riset Inovatif (SENTRINOV)*. *Bol.* 6, no. 1, pp. 778-786), 2020.
- [16] N. L. W. S. R. Ginantra and I. B. G. & Anandita, "Penerapan Metode Single Exponential Smoothing Dalam Peramalan Penjualan Barang". *J-SAKTI (Jurnal Sains Komputer dan Informatika)*, vol. 3, no. 2, pp. 433-441, 2019.
- [17] A. Virrayyani and S. Sutikno, "Prediksi Penjualan Barang Menggunakan Metode Adaptive Neuro-Fuzzy Inference System (ANFIS)". *Khazanah Informatika: Jurnal Ilmu Komputer dan Informatika*, vol. 2, no. 2, pp. 57-63, 2016.
- [18] F. Rozi and F. Sukmana, "Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor". *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 6, no. 1, pp. 155-160, 2021.
- [19] I. Budiman, T. Prahasto, and Y. Christyono, "Data Clustering Menggunakan Metodologi CRISP-DM Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma," *J. Sist. Inf. Bisnis*, vol. 1, no. 3, pp. 15–16, 2014
- [20] E. Prasetyo, *Data Mining: Konsep dan Aplikasi Menggunakan MATLAB*, Yogyakarta: Penerbit ANDI, 2012.