

# Penerapan Data Mining Untuk Memprediksi Penjualan Bahan Bakar Minyak Menggunakan Algoritma K-Means Clustering (Studi Kasus: PT. Anugerah Alam Semesta)

Rahayul Kahfi, Muhammad Syahrizal

Program Studi Teknik Informatika, STMIk Budi Darma, Medan, Indonesia

Email: rahayul.kahfi90@gmail.com

**Abstrak**—Bisnis selalu dilalui dengan bersaing semua perusahaan selalu memikirkan cara untuk cara membuat perusahaan agar selalu maju dan perusahaan dapat terus bertahan dalam mengembangkan bisnis jika tidak perusahaan akan mengalami kemunduran. Perusahaan ingin mengetahui jumlah penjualan pada PT. Anugerah Alam Semesta Raya untuk memperoleh analisa data penjualan yang paling banyak diminati terhadap penjualan jenis-jenis bahan bakar. Salah satu faktor dilakukannya pengclusteran pada penjualan bahan bakar minyak adalah untuk memudahkan dalam penjualan bahan bakar minyak dengan cara adanya pengkelompokkan. Salah satu masalah dalam persediaan adalah kesulitan dalam menentukan besarnya jumlah persediaan yang harus disediakan dalam memenuhi jumlah permintaan. Sering terjadi suatu perusahaan mempunyai jumlah persediaan terlalu sedikit dibanding dengan permintaan konsumen. Melakukan tinjau langsung ke perusahaan tersebut mengenai proses penjualan bahan bakar pada perusahaan tersebut dengan adanya pengolahan dengan cara mengclusterkan data-data penjualan sebelumnya untuk menghasilkan sebuah prediksi mengenai penjualan selanjutnya agar dapat lebih baik. Penjualan yang tidak stabil akan mengakibatkan prekenomian sebuah perusahaan dapat melemah dengan kata lain perusahaan harus mendapatkan suatu teknik dalam hal mengelolah penjualan bahan bakar minyak kedepannya, sehingga dapat menghasilkan sebuah prediksi dengan akurasi yang baik dalam mengambil sebuah kesimpulan.

**Kata Kunci:** Data Mining, Penjualan, Bahan Bakar Minyak, K-Means Clustering.

**Abstract**—Business is always traversed by competing all companies always thinking of ways to make the company always go forward and the company can continue to survive in developing the business if the company will not decline. The company wants to know the number of sales at PT. Anugerah Alam Semesta Raya to obtain the most popular sales data analysis for the sale of fuel types. One of the factors in clustering the sale of fuel oil is to facilitate the sale of fuel oil by means of grouping. One problem in inventory is the difficulty in determining the large amount of inventory that must be provided in meeting the demand. It often happens that a company has too little inventory compared to consumer demand. Reviewing the company directly regarding the process of selling fuel at the company by processing it by clustering the previous sales data to produce a prediction about further sales in order to be better. Unstable sales will result in a company's prekenomian can be weakened with the word another company must obtain a technique in terms of managing sales of fuel oil in the future, so that it can produce a prediction with good accuracy in drawing a conclusion.

**Keywords:** Data Mining, Sales, Fuels, K-Means Clustering.

## 1. PENDAHULUAN

Data Mining adalah Serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data dengan melakukan penggalian pola-pola dari data dengan tujuan untuk memanipulasi data menjadi informasi yang lebih berharga yang diperoleh dengan cara mengekstraksi. [1]

PT. Anugerah Alam Semesta Raya merupakan usaha dalam penjualan bahan bakar minyak ke mitra-mitra yang sudah terjalin kerjasama dengan PT. Anugerah Alam Semesta Raya. Salah satu mitra yang dibahas pada kasus ini adalah PT. Anugerah Alam Semesta Raya. Perusahaan ingin mengetahui jumlah penjualan pada PT. Anugerah Alam Semesta Raya untuk memperoleh analisa data penjualan yang paling banyak diminati terhadap penjualan jenis-jenis bahan bakar.

Salah satu faktor dilakukannya pengclusteran pada penjualan bahan bakar minyak adalah untuk memudahkan dalam penjualan bahan bakar minyak dengan cara adanya pengkelompokkan. Salah satu masalah dalam persediaan adalah kesulitan dalam menentukan besarnya jumlah persediaan yang harus disediakan dalam memenuhi jumlah permintaan. Sering terjadi suatu perusahaan mempunyai jumlah persediaan terlalu sedikit dibanding dengan permintaan konsumen. Keadaan ini dapat menyebabkan perusahaan mengeluarkan biaya yang lebih besar lagi untuk memenuhi jumlah permintaan. Selain itu, hal ini dapat menyebabkan hilangnya kepercayaan konsumen terhadap perusahaan dan bisa saja konsumen akan beralih ke produk sejenis dari perusahaan lain, sehingga dapat mengurangi kesempatan perusahaan untuk memperoleh laba. Penjualan yang tidak stabil akan mengakibatkan prekenomian sebuah perusahaan dapat melemah dengan kata lain perusahaan harus mendapatkan suatu teknik dalam hal mengelolah penjualan bahan bakar minyak kedepannya, sehingga dapat menghasilkan sebuah prediksi dengan akurasi yang baik dalam mengambil sebuah kesimpulan.

K-Means clustering adalah metode pengelompokan data yang banyak menjadi data yang sederhana dengan adanya pengelompokan. K-Means adalah sebuah proses untuk mengelompokan data ke dalam beberapa cluster atau kelompok sehingga data dalam satu cluster memiliki tingkat kemiripan yang maksimum dan data antar cluster memiliki kemiripan yang minimum, clustering ini tentu sangat berguna dan bisa menemukan grup atau kelompok yang tidak dikenal dalam data. [2]

Pada penelitian terdahulu yang dilakukan oleh Asroni yang diterbitkan pada Jurnal Ilmiah semesta teknika, Vol 18, maret 2015 dengan judul “Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang”. Menjelaskan bahwa variabel-variabel yang akan dijadikan acuan adalah nilai matakuliah algoritma dan pemrograman, fisika dasar, kalkulus dan index prestasi kumulatif (IPK). Dari permasalahan diperlukan sebuah pengelompokan terhadap kriteria-kriteria yang ada. Kriteria tersebut di proses dengan menggunakan metode pengelompokan k-means. [3]

Peneliti selanjutnya yang dilakukan oleh Elly muningsih, yang diterbitkan pada jurnal bianglala informatika, Vol 3 No.1 Maret 2015, ISSN : 2338-9761) dengan judul “Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang”. Manajemen stok yang dilakukan secara tidak akurat dan asal-asalan akan menyebabkan biaya simpan yang tinggi dan tidak ekonomis, karena bisa terjadi kekosongan atau kelebihan produk tertentu. Hal ini tentu akan sangat merugikan semua pelaku usaha seperti halnya *online shop*. [4]

## 2. METODOLOGI PENELITIAN

### 2.1 Data Mining

Data mining, sering juga disebut Sebagai *Knowledge Discovery In Database (KDD)*. *Knowledge Discovery* dan *Data Mining (KDD)* adalah sebuah daerah interdisipliner yang berfokus pada metodologi untuk mengekstrak pengetahuan yang berguna dari data. Pertumbuhan yang berkelanjutan cepat data online karena Internet dan meluasnya penggunaan database telah menciptakan kebutuhan yang besar untuk KDD metodologi. Tantangan untuk mengekstrak pengetahuan dari data menarik berdasarkan penelitian dalam statistik, database, pengenalan pola, pembelajaran mesin, visualisasi data, optimasi, dan komputasi kinerja tinggi, untuk memberikan solusi penemuan kecerdasan dan web bisnis lanjutan. Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam *database*, data warehouse, atau penyimpanan informasi lainnya. Data mining berkaitan dengan bidang ilmu-ilmu lain, seperti *database system*, data *warehousing*, *statistik*, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi. Selain itu, data mining didukung oleh ilmu lain seperti *neural network*, pengenalan pola, *spatial data analysis*, *image database*, *signal processing*. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi.[6]

### 2.2 Penjualan

Aktivitas penjualan merupakan pendapatan utama perusahaan karena jika aktivitas penjualan produk maupun jasa tidak dikelola dengan baik maka secara langsung dapat merugikan perusahaan. Hal ini dapat disebabkan karena sasaran penjualan yang diharapkan tidak tercapai dan pendapatan pun akan berkurang. Untuk lebih jelasnya dapat dilihat dari pengertian penjualan itu sendiri.

Produk adalah barang atau jasa yang dapat diperjual belikan. Dalam marketing, produk adalah apapun yang bisa ditawarkan ke sebuah pasar dan bisa memuaskan sebuah keinginan atau kebutuhan. Dalam tingkat pengecer, produk sering disebut sebagai *merchandise*. Dalam manufaktur, produk dibeli dalam bentuk barang mentah dan dijual sebagai barang jadi. Produk yang berupa barang mentah seperti metal atau hasil pertanian sering pula disebut sebagai komoditas.

### 2.3 Algoritma K-Means Clustering

K-Means merupakan algoritma clustering yang berulang-ulang. Algoritma K-Means dimulai dengan pemilihan secara acak K, K disini merupakan banyaknya Cluster yang ingin dibentuk dengan melakukan proses iterasi. Iterasi ini akan terus dilakukan sampai tercipta kondisi konvergen dimana tidak ada perubahan posisi pada data lagi. Secara lebih detail, algoritma *K-means Clustering* adalah sebagai berikut:

1. Definisikan jumlah K *cluster*.
2. Inisialisasi nilai pusat *cluster (centroid)* sebagai nilai acuan awal. *Centroid* ini dapat diperoleh secara acak.
3. Untuk setiap komponen atau nilai data, hitung dan tandai jarak (*distance*) ke *centroid* awal kemudian masukkan data tersebut ke *centroid* yang paling dekat jaraknya dengan menggunakan rumus *Euclidian distance*.
4. Hitung dan ubah kembali *centroid* tiap *cluster* sebagai rata-rata dari seluruh anggota kelompok (*group cluster* tersebut).
5. Cek semua data kembali dan taruh setiap data yang terdekat dengan *centroid* baru. Jika anggota tiap *cluster* tidak berubah (konvergen), maka langkah berhenti dan jika masih berubah, kembali ke langkah 2.

Rumus *Euclidian Distance*:

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

Dengan ketentuan adalah

D = Jarak X=Data Y= Centroid i= Banyak Data

### 3. HASIL DAN PEMBAHASAN

Untuk meningkatkan penjualan bahan bakar minyak secara tidak langsung perusahaan sudah memikirkan bagaimana prosedur persediaan sampai dengan pemasaran. Namun beberapa kendala muncul saat bahan bakar minyak sudah dipasarkan tidak memiliki suatu kepastian seberapa banyak yang sudah terjual, apakah dinyatakan penjualan tertinggi, sedang ataupun terendah. Untuk memberikan suatu gambaran penulis menggali informasi dari data – data yang sudah menumpuk dengan menerapkan suatu algoritma yang dapat melakukan klasifikasi penjualan dinyatakan masuk kategori tertentu.

Sebagai data awal dalam melakukan pengujian Algoritma K-means terhadap pengelempokan penjualan Bahan bakar minyak maka disajikan data pada tabel 1, Data yang digunakan dipilih secara acak sebanyak 10 data yang dinyatakan sebagai sampel. Adapun data penjualan bahan bakar minyak dinyatakan dalam satuan per ton untuk memudahkan berikut ini adalah sajian data penjualan.

**Tabel 1.** Jumlah Penjualan Bahan Bakar Minyak

Minggu Ke	Persediaan	Terjual
1	196493	190520
2	979271	892329
3	455595	451329
4	380899	361766
5	29978	19670
6	360226	318743
7	256277	14484
8	723529	71913
9	81380	74079
10	80769	57890

#### 3.1 Penerapan Algoritma K-Means

Berdasarkan pada uraian tersebut, maka algoritma K-Means *Clustering* akan melakukan proses iterasi atau pengulangan langkah-langkah berikut sampai terjadi kestabilan (konvergen) untuk memprediksi penjualan bahan bakar minyak pada PT. Anugerah Alam Semesta Raya berikut ini data pada proses cluster untuk penjualan bbm adalah sebagai berikut. Berdasarkan data di atas diketahui berdasarkan produk penjualan bahan bakar minyak adalah sebagai berikut. Sebagai bagian dari penggunaan Algoritma K-Means maka beberapa persiapan didefinisikan yaitu :

Jumlah cluster terdiri dari 3 yaitu dengan Kategori Output adalah ( Penjualan Meningkat Pada Cluster 1, Penjualan Sedang Pada Cluster 2 dan Penjualan Menurun Pada Cluster 3) jumlah data 10 dan jumlah atribut 2. Setelah diketahui nilai k dan pusat *cluster* awal selanjutnya mengukur jarak antara pusat *cluster* menggunakan *euclidian distance*, kemudian akan didapatkan matriks jarak yaitu C1, C2 dan C3 sebagai berikut:

Penentuan pusat awal cluster

Di ambil data ke-2 sebagai pusat cluster ke 1 979271 892329

Di ambil data ke-3 sebagai pusat cluster ke 2 723529 71913

Di ambil data ke-5 sebagai pusat cluster ke 3 29978 190520

Penentuan jarak cluster dilakukan untuk mengetahui jarak tiap clusternya dengan cara sebagai berikut. Setelah pusat cluster diketahui selanjutnya menghitung nilai tersebut dengan memasukkan nilai cluster kedalam rumus di atas untuk diproses kemudian.

**Iterasi Pertama** : Adapun untuk perhitungan cluster 1 adalah sebagai berikut :

$$D(x_1,y_1)=\sqrt{(196493 - 979271)^2 + (190520 - 892329)^2}$$

$$D(x_1,y_1)=1051322$$

$$D(x_2,y_2)=\sqrt{(196493 - 979271)^2 + (190520 - 892329)^2}$$

$$D(x_2,y_2)=0$$

$$D(x_3,y_3)=\sqrt{(196493 - 979271)^2 + (190520 - 892329)^2}$$

$$D(x_3,y_3)=684629.5$$

$$D(x_4, y_4) = \sqrt{(196493 - 979271)^2 + (190520 - 892329)^2}$$

$$D(x_4, y_4) = 799716$$

$$D(x_5, y_5) = \sqrt{(196493 - 979271)^2 + (190520 - 892329)^2}$$

$$D(x_5, y_5) = 1289454$$

$$D(x_6, y_6) = \sqrt{(196493 - 979271)^2 + (190520 - 892329)^2}$$

$$D(x_6, y_6) = 843929.9$$

$$D(x_7, y_7) = \sqrt{(196493 - 979271)^2 + (190520 - 892329)^2}$$

$$D(x_7, y_7) = 1137248$$

$$D(x_8, y_8) = \sqrt{(196493 - 979271)^2 + (190520 - 892329)^2}$$

$$D(x_8, y_8) = 859352.3$$

$$D(x_9, y_9) = \sqrt{(196493 - 979271)^2 + (190520 - 892329)^2}$$

$$D(x_9, y_9) = 1214801$$

$$D(x_{10}, y_{10}) = \sqrt{(196493 - 979271)^2 + (190520 - 892329)^2}$$

$$D(x_{10}, y_{10}) = 1226211$$

Adapun untuk perhitungan cluster 2 adalah sebagai berikut :

$$D(x_1, y_1) = \sqrt{(196493 - 723529)^2 + (190520 - 71913)^2}$$

$$D(x_1, y_1) = 540217$$

$$D(x_2, y_2) = \sqrt{(196493 - 723529)^2 + (190520 - 71913)^2}$$

$$D(x_2, y_2) = 859352$$

$$D(x_3, y_3) = \sqrt{(196493 - 723529)^2 + (190520 - 71913)^2}$$

$$D(x_3, y_3) = 464484$$

$$D(x_4, y_4) = \sqrt{(196493 - 723529)^2 + (190520 - 71913)^2}$$

$$D(x_4, y_4) = 448787$$

$$D(x_5, y_5) = \sqrt{(196493 - 723529)^2 + (190520 - 71913)^2}$$

$$D(x_5, y_5) = 695516$$

$$D(x_6, y_6) = \sqrt{(196493 - 723529)^2 + (190520 - 71913)^2}$$

$$D(x_6, y_6) = 439220$$

$$D(x_7, y_7) = \sqrt{(196493 - 723529)^2 + (190520 - 71913)^2}$$

$$D(x_7, y_7) = 470768$$

$$D(x_8, y_8) = \sqrt{(196493 - 723529)^2 + (190520 - 71913)^2}$$

$$D(x_8, y_8) = 0$$

$$D(x_9, y_9) = \sqrt{(196493 - 723529)^2 + (190520 - 71913)^2}$$

$$D(x_9, y_9) = 642153$$

$$D(x_{10}, y_{10}) = \sqrt{(196493 - 723529)^2 + (190520 - 71913)^2}$$

$$D(x_{10}, y_{10}) = 642913$$

Adapun untuk perhitungan cluster 3 adalah sebagai berikut :

$$D(x_1, y_1) = \sqrt{(196493 - 29978)^2 + (190520 - 19670)^2}$$

$$D(x_1, y_1) = 238572.8$$

$$D(x_2, y_2) = \sqrt{(196493 - 29978)^2 + (190520 - 19670)^2}$$

$$D(x_2, y_2) = 1289454$$

$$D(x_3, y_3) = \sqrt{(196493 - 29978)^2 + (190520 - 19670)^2}$$

$$D(x_3, y_3) = 606201$$

$$D(x_4, y_4) = \sqrt{(196493 - 29978)^2 + (190520 - 19670)^2}$$

$$D(x_4, y_4) = 490076.8$$

$$D(x_5, y_5) = \sqrt{(196493 - 29978)^2 + (190520 - 19670)^2}$$

$$D(x_5, y_5) = 0$$

$$D(x_6, y_6) = \sqrt{(196493 - 29978)^2 + (190520 - 19670)^2}$$

$$D(x_6, y_6) = 445543$$

$$D(x_7, y_7) = \sqrt{(196493 - 29978)^2 + (190520 - 19670)^2}$$

$$D(x_7, y_7) = 226358$$

$$D(x_8, y_8) = \sqrt{(196493 - 29978)^2 + (190520 - 19670)^2}$$

$$D(x_8, y_8) = 695516$$

$$D(x_9, y_9) = \sqrt{(196493 - 29978)^2 + (190520 - 19670)^2}$$

$$D(x_9, y_9) = 74849.9$$

$$D(x_{10}, y_{10}) = \sqrt{(196493 - 29978)^2 + (190520 - 19670)^2}$$

$$D(x_{10}, y_{10}) = 63564$$

Iterasi pertama didapat berdasarkan perhitungan data di atas menghasilkan tabel berikut ini :

**Tabel 2.** Hasil Perhitungan Cluster

Persediaan/barel	Terjual	c1	c2	c3	Jarak Terpendek
196493	190520	1051322	540217	238573	C2
979271	892329	0	859352	1289454	C1
455595	451329	684630	464484	606201	C2
380899	361766	799716	448787	490077	C2
29978	19670	1289454	695516	0	C3
360226	318743	843930	439220	445543	C2
256277	14484	1137248	470768	226358	C3
723529	71913	859352	0	695516	C2
81380	74079	1214801	642153	74849.9	C3
80769	57890	1226211	642913	63564.9	C3

**Pengelompokan Data :** Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat *cluster*, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat *cluster* terdekat. Dengan cara membandingkan hasil *cluster* dan diambil yang paling kecil.

**Tabel 3.** Hasil Pengelompokan

Persediaan	Terjual	c1	c2	c3	Jarak
196493	190520	1051322	540217	238573	238573
979271	892329	0	859352	1289454	0
455595	451329	684630	464484	606201	464484
380899	361766	799716	448787	490077	448787
29978	19670	1289454	695516	0	0
360226	318743	843930	439220	445543	445543
256277	14484	1137248	470768	226358	226358
723529	71913	859352	0	695516	0
81380	74079	1214801	642153	74849.9	74849.9
80769	57890	1226211	642913	63564.9	63564.9

Berdasarkan tabel 3. diatas maka dibentuk cluster baru dengan mengambil nilai terkecil. Untuk memperjelas maka ditampilkan dengan tabel 4. berikut ini.

**Tabel 4.** Hasil Cluster pertama

Persediaan	Terjual	c1	c2	c3
196493	190520			*
979271	892329	*		
455595	451329		*	
380899	361766		*	
29978	19670			*
360226	318743		*	
256277	14484			*
723529	71913		*	
81380	74079		*	
80769	57890			*

**Iterasi Kedua :** dilakukan setelah hasil iterasi pertama sudah selesai adapun perhitungannya adalah sebagai berikut :

**Tabel 5.** Hasil Pengelompokan Iterasi I

c1	c2	c3
		*
*		
	*	
	*	
		*
	*	
		*

c1	c2	c3
	*	
	*	
		*

Untuk mengasilkan proses iterasi baru maka dilanjutkan dengan menentukan cluster baru pada tabel 4.6 berikut ini:

**Tabel 6.** Hasil Iterasi

Persediaan	Terjual	c1	c2	c3
196493	190520	979271	400325.8	140879.3
979271	892329	892329	255566	70641
455595	451329			
380899	361766			
29978	19670			
360226	318743			
256277	14484			
723529	71913			
81380	74079			
80769	57890			

Data pada tabel 6. akan dijadikan sebagai acuan dalam menentukan iterasi ke dua berikut adalah uraian untuk pusat cluster baru

Di ambil data ke-2 sebagai pusat cluster ke 1	979271	892329
Di ambil data ke-3 sebagai pusat cluster ke 2	400325.8	255566
Di ambil data ke-5 sebagai pusat cluster ke 3	140879	70641

**Tabel 7.** Hasil Perhitungan Cluster

Persediaan	Terjual	c1	c2	c3	Jarak
196493	190520	1051322	213960	132151	132151
979271	892329	0	860607	1173913	0
455595	451329	684630	203415	493933	203415
380899	361766	799716	107962	377311	107962
29978	19670	1289454	439095	122054	122054
360226	318743	843930	74828.7	331161	331161
256277	14484	1137248	280839	128336	128336
723529	71913	859352	371737	582651	371737
81380	74079	1214801	366966	59598.5	59598.5
80769	57890	1226211	375756	61447.8	61447.8

Setelah tabel 7. selesai maka dilanjutkan dengan menentukan cluster dengan memilih aktifitas terkecil. Hasil ditampilkan pada tabel 8. berikut ini.

**Tabel 8.** Hasil Cluster ke 2

Persediaan	Terjual	c1	c2	c3
196493	190520			*
979271	892329	*		
455595	451329		*	
380899	361766		*	
29978	19670			*
360226	318743			*
256277	14484			*
723529	71913		*	
81380	74079		*	
80769	57890		*	

Iterasi dilakukan sampai dengan menghasilkan pengelompokan data yang sama dengan iterasi sebelumnya, dalam uji coba digunakan sampai dengan iterasi 2 dengan menghasilkan kesimpulan dalam data berikut ini.

**Tabel 9.** Hasil Akhir Iterasi

Minggu	Persediaan	Terjual	Naik	Sedang	Turun
1	196493	190520			*
2	979271	892329	*		

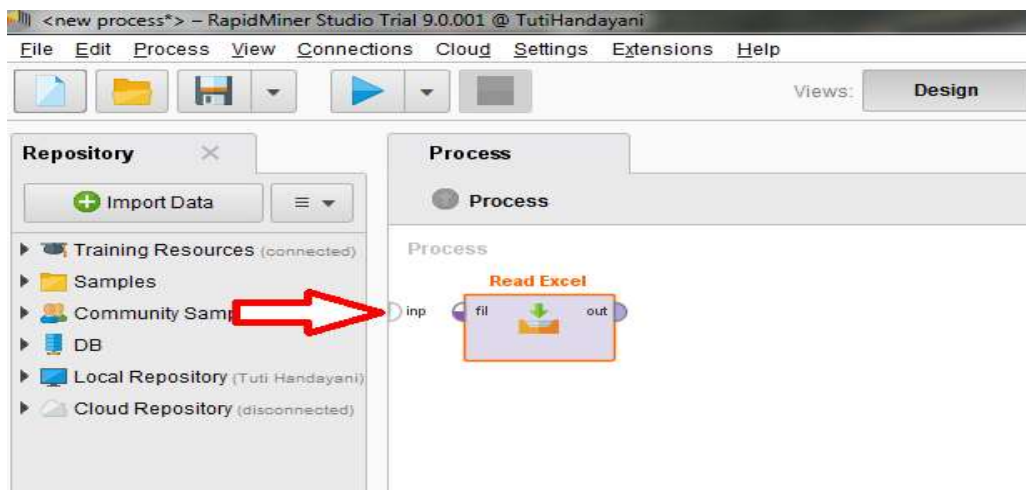
Minggu	Persediaan	Terjual	Naik	Sedang	Turun
3	455595	451329		*	
4	380899	361766		*	
5	29978	19670			*
6	360226	318743			*
7	256277	14484			*
8	723529	71913		*	
9	81380	74079		*	
10	80769	57890		*	

Sehingga dalam proses pengelompokan ini ditemukan bahwa

1. Penjualan dengan kategori tertinggi pada minggu ke 2
2. Penjualan dengan kategori Sedang yaitu pada minggu ke 3,4,8,9 dan 10
3. Penjualan terendah yaitu minggu 1, 5,6 dan 7

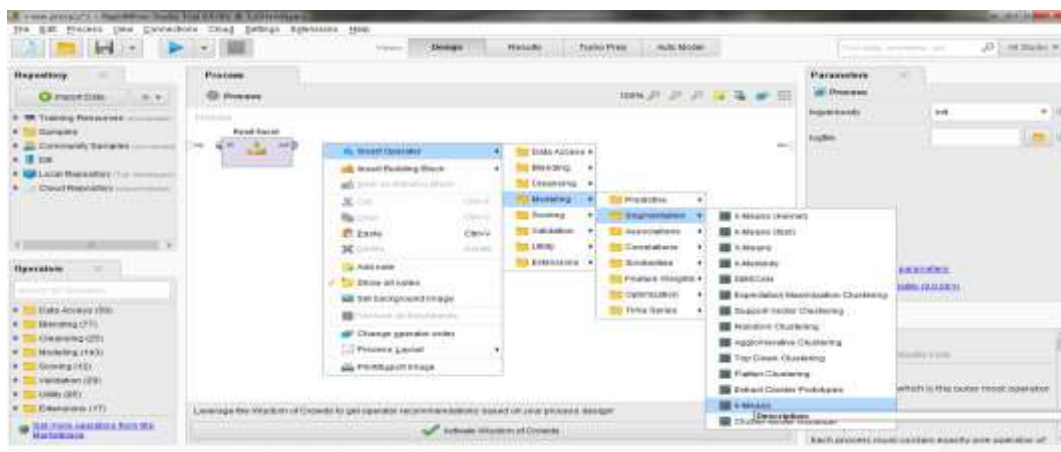
### 3.2 Pengujian

Pada implementasi dan pengujian disini kita menggunakan sebuah software RapidMiner Studio versi 9.0. Setelah selesai import data s, maka pada Main Process terlihat sebuah operator baru Read Excel yang sudah berisi file Penjualan.xlsx yang di import langsung dari file excel. Data tersebut siap dilakukan pengujian.



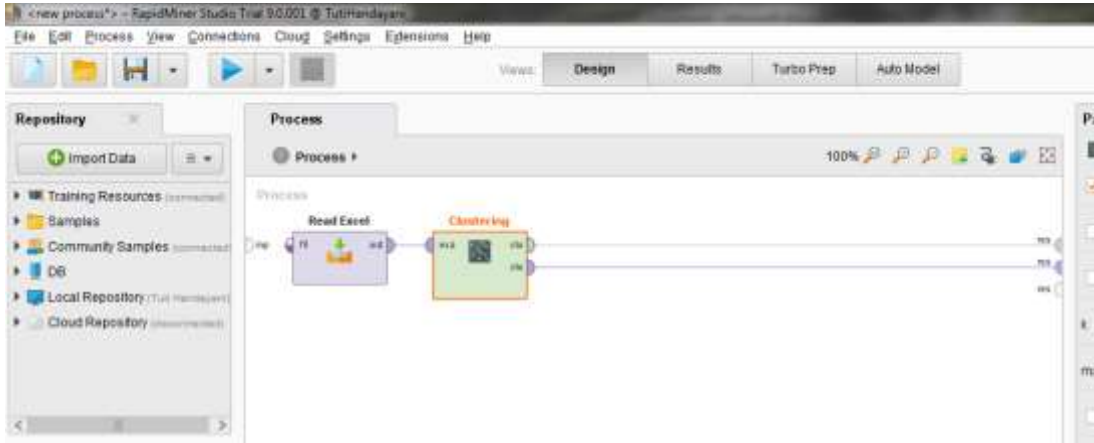
Gambar 1. Tampilan Operator Data Setelah Import Data

Pada gambar 1. menampilkan data yang sudah di import, sehingga tombol pada operator read excel tidak ada lagi tanda seru warna kuning (menandakan operator telah berisi data dan siap diolah). Langkah selanjutnya adalah menambahkan operator K-Means dengan cara klik kanan → Insert Operator → Modeling → Segmentation → K-Means.



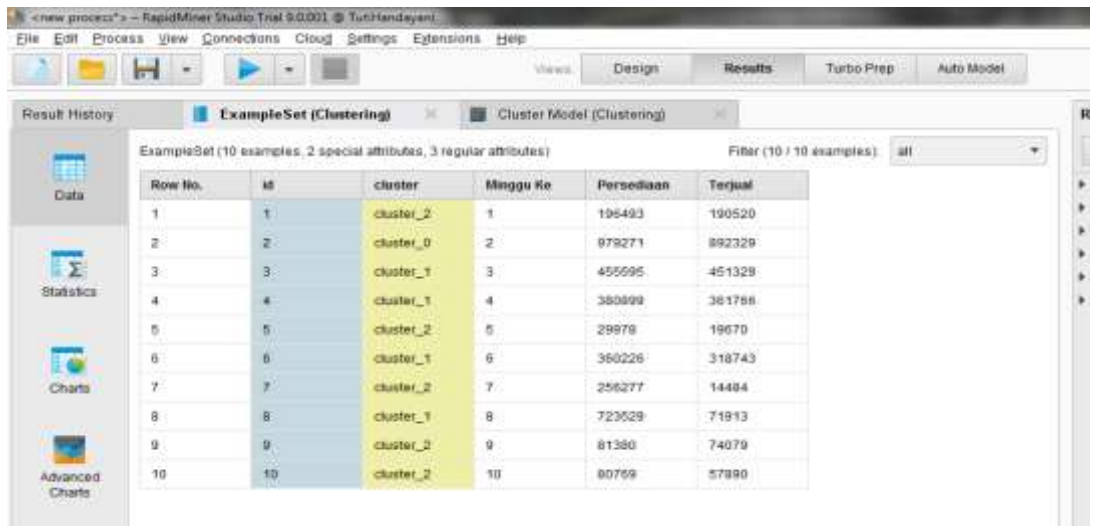
Gambar 2. Menambahkan Operator K-Means

Setelah menambahkan operator K-Means pada lembaran kerja hubungkan operator tersebut ke data file Penjualan yang sudah diimport, apabila sudah terhubung klik tombol play ▶ Adapun hasilnya sebagai berikut:



Gambar 3. Tampilan Proses K-Means

Setelah tombol play diklik akan diperoleh hasil seperti gambar 4. di bawah ini

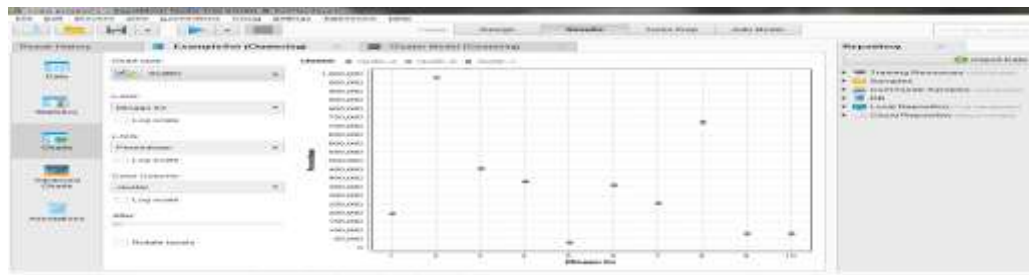


Gambar 4. Tampilan Running Data Proses K-Means

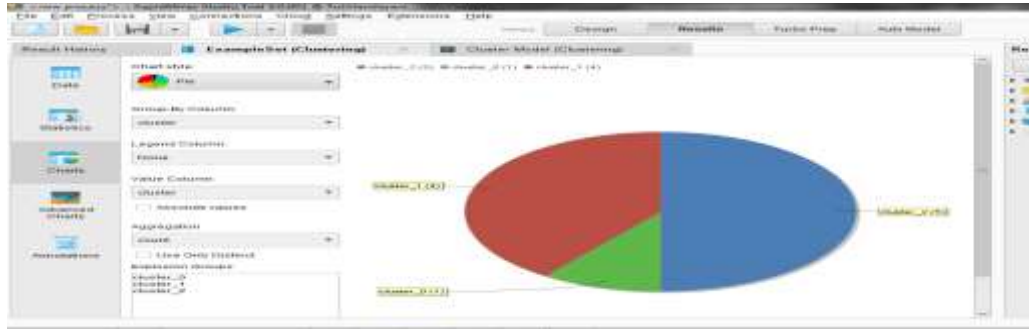
Dari hasil pengujian dapat disajikan dalam beberapa bentuk, adapun bentuknya sebagai berikut :

1. Chart

Chart merupakan tampilan grafik hasil pengelompokan data penjualan dengan jumlah cluster sebanyak 3. Adapun hasilnya dapat dilihat pada gambar 5. di bawah ini



Gambar 5. Tampilan Cluster Dalam Bentuk Chart



Gambar 7. Tampilan Cluster Dalam Bentuk Pie

2. Cluster Model

Pada cluster model dapat dilihat beberapa tampilan hasil cluster, yaitu Text View yang merupakan tampilan hasil pengelompokan berdasarkan urutan cluster dan jumlah data dalam masing-masing cluster. Adapun hasilnya dapat dilihat pada gambar 8. di bawah ini



Gambar 8. Tampilan Cluster Dalam Text View

Folder View merupakan tampilan data bagian-bagian cluster secara keseluruhan, dimana masing-masing anggota cluster ditampilkan berdasarkan minggu. Adapun hasilnya dapat dilihat pada gambar 9. di bawah ini



Gambar 9. Tampilan Cluster Dalam Folder View

### 4. KESIMPULAN

Dari hasil analisa dan pengujian pengelompokan data penjualan dengan algoritma k-means clustering, dapat diambil beberapa kesimpulan, yaitu :

1. Algoritma K-Means Clustering dapat diimplementasikan dengan menggunakan Database penjualan bahan bakar minyak karena dapat dijadikan sebagai informasi yang sangat berharga dalam pengambilan keputusan untuk mempersiapkan stok persediaan yang diperlukan kemudian.
2. Aplikasi rapid miner dengan versi 9.0 sangat membantu dalam proses pengelompokan data penjualan sangat efisien dan dapat mempercepat proses hasil penjualan bahan bakar minyak di PT.Anugerah Alam Semesta Raya.

## REFERENCES

- [1] Johan Oscar Ong, *Jurnal Ilmiah Teknik Industri*, vol. 12, p. 1, juni 2013.
- [2] Sri Mulyati, *Prosiding Seminar Ilmiah Nasional Teknologi Komputer*, vol. 1, no. Padang, p. 2, oktober 2015.
- [3] Asroni, *Jurnal Ilmiah Semesta Teknika*, vol. 18, Magelang, p. 3, maret 2015.
- [4] Elly muningsih, *jurnal bianglala informatika*, vol. 3, Yogyakarta, p. 4, maret 2015.
- [5] Guntur Setiawan, *implementasi dalam birokrasi pembangunan*, bandung, p. 5, 2002.
- [6] Jiawei dan Kamber, Micheline han, *data mining*, p. 6, 2006.
- [7] Budi Santosa, *data mining teknik pemanfaatan data untuk keperluan bisnis*, Yogyakarta, p. 7, 2007.
- [8] Luthfiq Emha Kusri, *algoritma data mining*, Yogyakarta, p. 8, 2009.
- [9] Dominikus Juju Feri Sulianta, *Data mining meramalkan bisnis perusahaan*, jakarta, p. 9, 2010.
- [10] Juan santana finn lee, *data mining:meramalkan bisnis perusahaan*, no. jakarta, p. 11, 2010.
- [11] Abu Bakar dan Wibowo Arif, *akuntansi keuangan dasar2*, jakarta, p. 12, 2008.
- [12] Oliviera, *pendekatan partisi dan clustering*, p. 13, 2007.
- [13] Bian S.,et.al Everitt, *Cluster Analisis*, p. 14, 2011.
- [14] Brian S.,et.al Everitt, *Cluster Analisis 5th Edition*, p. 15, 2011.
- [15] Everitt et al, *agglomerative clustering*, p. 16, 2011.
- [16] Adi Nugroho, *analisa dan perancangan*, no. bandung, p. 17.
- [17] Haviluddin, *Unified modelling language*, p. 18, 2011.
- [18] D. P. Utomo and B. Purba, "Penerapan Datamining pada Data Gempa Bumi Terhadap Potensi Tsunami di Indonesia," *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, vol. 1, pp. 846-853, 2019.
- [19] R. Amelia and D. P. Utomo, "ANALISA POLA PEMESANAN PRODUK MODERN TRADE INDEPENDENT DENGAN MENEREPAKAN ALGORITMA FP. GROWTH (STUDI KASUS: PT. ADAM DANI LESTARI)," *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, vol. 3, no. 1, pp. 416-423, 2019.
- [20] E. Bulolo, *Data Mining Untuk Perguruan Tinggi*. 2020.