
Implementasi Data Mining Untuk Pengelompokan Data Penjualan Berdasarkan Pola Pembelian Menggunakan Algoritma K-Means Clustering Pada Toko Syihan

Nurin Fadhilah Adani*, Ahmad Fitri Boy, S.Kom., M.Kom **, Rendy Syahputra, S.Kom., M.Kom ***

* Program Studi Sistem Informasi, STMIK Triguna Dharma

** Program Studi Sistem Informasi, STMIK Triguna Dharma

*** Program Studi Sistem Informasi, STMIK Triguna Dharma

Article Info

Article history:

Received Jun 12th, 201x

Revised Aug 20th, 201x

Accepted Aug 26th, 201x

Keyword:

Data Mining,
Penjualan Sembako,
K-Means

ABSTRAK

Toko Syihan merupakan sebuah toko yang bergerak dalam bidang penjualan sembako. Pada Toko Syihan terdapat beberapa permasalahan yang kerap muncul. Dalam melakukan penjualan sering pelanggan mengalami kekecewaan karena seringnya stok yang seharusnya ingin dibeli oleh pelanggan kosong. Karena kekecewaan tersebut jumlah pelanggan menurun. Adapun proses-proses yang dilakukan masih manual sehingga masih sering terjadi kesalahan dalam pencatatan data-data yang ada sehingga menyebabkan kurangnya efisiensi waktu yang digunakan. Ketersediaan data yang ada tidak digunakan dengan baik sehingga data penjualan tidak dimanfaatkan dengan sebaik mungkin yang dapat merancang strategi bisnis dalam meningkatkan produk. Untuk menyelesaikan masalah tersebut dapat menggunakan aplikasi data mining, yaitu dengan memanfaatkan data yang ada untuk menggali informasi baru. Salah satu teknik yang ada pada data mining adalah clustering. Clustering dipilih karena dapat mengelompokkan data-data sesuai dengan karakteristik yang diinginkan, dalam penelitian ini berarti mengelompokkan data penjualan sembako di Toko Syihan. Adapun algoritma clustering yang digunakan adalah K-Means Clustering diintegrasikan pada aplikasi pemrograman berbasis dekstop. Dengan menerapkan aplikasi data mining ini maka hasil yang diperoleh yaitu sistem ini berhasil mengelompokkan data penjualan sembako menjadi 3 cluster (kelompok). Cluster 1 terdiri dari 12 data, cluster 2 terdiri dari 10 data dan cluster 3 yang merupakan cluster terkecil terdiri dari 3 data.

Copyright © 2019 STMIK Triguna Dharma.

All rights reserved.

Corresponding Author

Nama : Nurin Fadhilah Adani

Program Studi : Sistem Informasi STMIK Triguna Dharma

Email : nurinfhdlh@gmail.com

1. PENDAHULUAN

Dalam dunia bisnis sekarang ini menuntut berbagai kalangan untuk mengembangkan bisnis mereka agar bertahan dalam dunia persaingan. Untuk mencapai hal yang dibutuhkan itu para pembisnis harus meningkatkan produk, menambahkan jenis produk baru dan pengurangan biaya operasional perusahaan. Sistem informasi sangat

dibutuhkan untuk pihak manajemen sebagai tolak ukur dalam memutuskan keputusan yang dapat memberikan dampak keberlangsungan perusahaan. Maka dari itu penting bagi pelaku usaha untuk menganalisa ketersediaan stok. Salah satu yang dilakukan ialah dengan menganalisa data transaksi yang ada pada sebuah aplikasi data mining.

Toko Syihan merupakan sebuah toko yang bergerak dalam bidang penjualan sembako. Toko ini setiap harinya memenuhi kebutuhan pelanggan sehingga dapat ditentukan strategi pola penjualan. Untuk melakukan itu toko membutuhkan sumber informasi yang cukup banyak untuk dapat dianalisis lebih lanjut.

Pada Toko Syihan terdapat beberapa permasalahan yang kerap muncul. Dalam melakukan penjualan sering pelanggan mengalami kekecewaan karena seringnya stok yang seharusnya ingin dibeli oleh pelanggan kosong. Karena kekecewaan tersebut jumlah pelanggan menurun. Adapun proses-proses yang dilakukan masih manual sehingga masih sering terjadi kesalahan dalam pencatatan data-data yang ada sehingga menyebabkan kurangnya efisiensi waktu yang digunakan. Ketersediaan data yang ada tidak digunakan dengan baik sehingga data penjualan tidak dimanfaatkan dengan sebaik mungkin yang dapat merancang strategi bisnis dalam meningkatkan produk.

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang terkait dari berbagai database besar [1]. Untuk memenuhi kebutuhan manajemen diatas, banyak cara yang dapat ditempuh. Salah satunya dengan pemanfaatan data perusahaan. Data mining merupakan teknik yang menggabungkan teknik analisis data dan menemukan pola-pola yang penting pada data [2]. Data mining merupakan proses informasi dengan mencari pola dan hubungan yang tersembunyi pada tumpukan data. Data mining disebut juga sebagai *Knowledge In Database (KDD)* yaitu kegiatan pengumpulan pemakaian data lampau untuk menemukan pola atau hubungan terhadap data yang ukurannya besar. Banyak algoritma yang dapat digunakan untuk menemukan pola pembelian konsumen, salah satunya K-Means Clustering.

K-Means merupakan salah satu metode dalam data mining yang digunakan untuk mempartisi data yang ada kedalam beberapa cluster sehingga data yang memiliki karakteristik yang sama akan dikelompokkan kedalam satu cluster dan data dengan karakteristik yang berbeda akan dikelompokkan kedalam cluster lain. Selain itu juga pada penelitian lain seperti [3], [4] dijelaskan bahwasanya metode K-Means dapat diterapkan untuk eksplorasi data pelanggan untuk kontekstual marketing voice over long term evolution, analisa dan penerapan data mining untuk menentukan kubikasi air terjual berdasarkan pengelompokan pelanggan. Dari data yang dianalisa dengan Algoritma K-Means Clustering akan menghasilkan satu tujuan yaitu mendapatkan kelompok data yang akan dipromosikan. Dari latar belakang diatas maka disusunlah penelitian ini dengan judul “Implementasi Data Mining Untuk Pengelompokan Data Penjualan Berdasarkan Pola Pembelian Menggunakan Algoritma K-Means Clustering Pada Toko Syihan”

2. METODE PENELITIAN

2.1 Data Mining

Data adalah kumpulan informasi yang digunakan dalam proses pengambilan kesimpulan maupun pengambilan keputusan. Data merupakan bentuk jamak dari bentuk tunggal *datum* atau data-item. Data adalah kenyataan yang menggambarkan suatu kejadian dan kesatuan nyata [4].

Data mining merupakan suatu proses untuk mendapatkan informasi baru dari kumpulan data dengan menggunakan algoritma dan teknik yang melibatkan bidang ilmu statistik, mesin pembelajaran dan sistem manajemen *database*. Data mining digunakan untuk ekstraksi informasi penting yang tersembunyi dari *dataset* yang besar. Dengan adanya Data mining maka akan didapatkan suatu permata berupa pengetahuan di dalam kumpulan data yang banyak jumlahnya.

Banyak istilah yang digunakan untuk menunjukkan proses Data mining. Untuk dapat memberikan pemahaman tentang Data mining, berikut ini adalah beberapa fakta yang terjadi seperti banyaknya organisasi, baik dari dunia bisnis maupun pemerintah berurusan dengan sejumlah informasi dan juga pengelolaan basis data informasi tersebut, dan bukan tidak mungkin termasuk didalamnya kebutuhan akan pembangunan *Data Warehouse* dalam skala besar. Seringkali data yang tersimpan tidak dapat secara langsung dianalisa dengan metode statistik standar, hal ini disebabkan karena adanya beberapa *record* yang hilang ataupun karena datanya dalam dimensi ukuran kualitatif bukan kuantitatif.

Data mining merupakan sebuah proses ekstraksi data menjadi informasi atau pengetahuan yang baru. Data yang diolah sebelumnya merupakan data yang bersifat implisit dan biasanya dianggap tidak berguna serta data dalam jumlah besar [5].

Terdapat empat tugas utama data mining yaitu sebagai berikut :

1. *Predictive Modelling*

Predictive modelling digunakan untuk membangun sebuah model untuk target *variable* sebagai fungsi dari *explanatory variable*. *Explanatory variable* dalam hal ini merupakan semua atribut yang digunakan untuk melakukan prediksi, sedangkan variabel target merupakan atribut yang akan diprediksi nilainya. *Predictive modelling* dibagi menjadi dua tipe yaitu : *Classification* digunakan untuk memprediksi nilai dari target *variable* yang *discrete* (diskrit) dan *regression* digunakan untuk memprediksi nilai dari target *variable* yang *continue* (berkelanjutan).

2. Association Analysis

Association analysis digunakan untuk menemukan aturan asosiasi yang memperlihatkan kondisi-kondisi nilai atribut yang sering muncul secara bersamaan dalam sebuah himpunan data.

3. Cluster Analysis

Tidak seperti klasifikasi yang menganalisa kelas data obyek yang mengandung label. *Clustering* menganalisa objek data tanpa memeriksa kelas label yang diketahui. Label-label kelas dilibatkan di dalam data *training*, karena belum diketahui sebelumnya. *Clustering* merupakan proses pengelompokan sekumpulan objek yang sangat mirip.

4. Anomaly Detection

Anomaly detection merupakan metode pendeteksian suatu data dimana tujuannya adalah menemukan objek yang berbeda dari sebagian besar objek lain. *Anomaly* dapat dideteksi dengan menggunakan uji statistik yang menerapkan model distribusi atau probabilitas untuk data.

Data mining merupakan ilmu baru yang berakar dari berbagai bidang ilmu seperti kecerdasan buatan (*Artificial Intelligence*), mesin *learning*, statistik dan *database*. Oleh karena itu dalam menggali informasi, data mining menggunakan beberapa teknik, antara lain [6] :

1. Association Discovery

Association Discovery adalah teknik mempelajari sekumpulan data dan untuk menunjukkan hubungan antara kemunculan atribut-atribut dalam data. Teknik ini mencoba untuk menyiapkan nilai-nilai yang muncul pada saat bersamaan dalam setiap barisnya dan menampilkan hasil keluaran yang disimpulkan dalam sebuah *rule*.

2. Clustering

Clustering adalah proses pengumpulan data yang serumpun dari sebuah *dataset* yang lebih besar. Teknik ini menyingkapkan sejumlah kelompok-kelompok yang digunakan sebagai masukan datanya. Dengan *Clustering* kelompok minoritas yang tersebar dikelompokkan dalam sebuah kelompok besar yang memiliki kemiripan entitas. *Clustering* dapat juga digunakan untuk mendeteksi secara otomatis *cluster* dari *record-record* yang berdekatan dengan pengertian tertentu di dalam keseluruhan variabel-variabel.

3. Sequential Discovery

Sequential Discovery adalah teknik mencari pola-pola diantara peristiwa-peristiwa yang muncul dalam periode waktu. Metode ini dapat digunakan untuk mencari pola komoditas yang terjadi berulang kali. Teknik ini terkonsentrasi pada kebiasaan yang sama yang sering muncul di kemudian hari. Menurut laporan tersebut diungkapkan bahwa algoritma yang dipakai memiliki kriteria sebagai berikut :

- One Scan* : algoritma ini membutuhkan paling sedikit satu kali pelarikan basis data untuk membuat *cluster* yang diinginkan.
- Anytime Algorithm* : algoritma ini selalu dapat menyediakan jawaban yang terbaik setiap saat selama komputasi dijalankan.
- Interruptable and Incremental* : algoritma ini dapat ditunda, dapat dihentikan dan dijalankan lagi, dan hasil sementara yang muncul dapat disimpan untuk kelanjutan komputasi data baru lebih lanjut.
- Limited RAM Requirement* : algoritma ini mampu bekerja dalam kondisi memori yang terbatas.
- Forward-only Cursor* : algoritma ini mampu untuk beroperasi pada basis data dengan kursor (proses) bergerak maju hingga akhir data.

4. Classification

Classification adalah proses pengumpulan data bersama-sama yang didasarkan atas sekumpulan kesamaan yang awalnya telah ditentukan oleh seorang analis sebelum analisa dimulai. Teknik ini memeriksa data yang telah diklasifikasikan dan dikumpulkan dalam grup bersama-sama sesuai dengan aturan keanggotaannya. Aturan keanggotaan bisa mempunyai komponen waktu, komponen geografis, komponen kuantitatif. Proses klasifikasi ini dapat dibagi menjadi tiga *fase* :

- Learning* : algoritma yang mencari sejumlah *record* dari *training set* dan menciptakan sebuah deskripsi tentang model klasifikasi. Model ini dibuat kecil dan persis pada saat yang sama.
- Model yang sudah dibuat harus diuji dengan serangkaian uji coba dalam sebuah basis data. Sebagai *training set*, *record-record* uji coba harus merupakan kumpulan yang sudah lebih dahulu diklasifikasikan. *Record* untuk uji coba harus berbeda dengan *training set*. *Model classification* akan selalu bekerja dengan sempurna dalam *training set* yang digunakan.
- Classification*, model *iterasi* yang digunakan untuk membuat klasifikasi pada sisa *record-record* dalam basis data yang ada.

5. Neural Network

Neural Network merupakan sebuah metode khusus untuk pengendalian identifikasi pola yang digunakan pada *trend* perkiraan berdasarkan kebiasaan yang telah diketahui sebelumnya. Suatu *trend* atau kecenderungan dapat diidentifikasi sebagai pergerakan yang ditunjukkan berdasarkan pada kebiasaan yang lama. Inti dari pemrosesan didasarkan pada pengolahan data yang meniru dari fungsi sistem saraf tubuh manusia. Pengetahuan dapat dipelajari dari suatu kumpulan data yang terpisah dan kompleks. Ada tiga lapisan (*layers*) pada network, lapisan kiri yang menerima masukan (*input*), lapisan tersembunyi (*hidden layer neurons*) yang menjalankan fungsi kerja pemrosesan datanya, dan lapisan kanan yang menampilkan analisa hasil keluaran. Ketika *network* dilatih (*trained*) dengan informasi yang dimasukkan melalui masukan (*input*) akan diproses oleh lapisan tengah

(*neurons*) yang menjadikannya cerdas dalam sejumlah elemen data yang ada dan mengeluarkan hasil pada lapisan keluarnya.

Data mining serangkaian proses seperti yang dijelaskan sebagai berikut:

1. Pembersihan data (untuk membuang data yang tidak konsisten dan *noise*).
2. Integrasi data (penggabungan data dari beberapa sumber)
3. Transformasi data (data diubah menjadi bentuk yang sesuai untuk di-*mining*)
4. Aplikasi teknik data mining, proses ekstraksi pola dari data yang ada
5. Evaluasi pola yang ditemukan (proses interpretasi pola menjadi pengetahuan yang dapat digunakan untuk mendukung pengambilan keputusan).
6. Presentasi pengetahuan (dengan teknik visualisasi).

2.2 K-Means Clustering

K-Means merupakan salah satu metode pengelompokan data nonhierarki (sekatan) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data kedalam kelompok yang lain. Adapun tujuan pengelompokan data ini adalah meminimalkan variasi didalam suatu kelompok dan memaksimalkan variasi antar kelompok [11].

Pada dasarnya *clustering* merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain. *Clustering* merupakan salah satu metode *data mining* yang bersifat tanpa arahan (*unsupervised*), maksudnya metode ini diterapkan tanpa adanya latihan (*taining*) dan tanpa ada guru (*teacher*) serta tidak memerlukan target output. Dalam data mining ada dua jenis metode clustering yang digunakan dalam pengelompokan data, yaitu *hierarchical clustering* dan *non-hierarchical clustering*.

Hierarchical clustering adalah suatu metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang memiliki kedekatan kedua. Demikian seterusnya sehingga cluster akan membentuk semacam pohon dimana ada hierarki (tingkatan) yang jelas antar objek, dari yang paling mirip sampai yang paling tidak mirip. Secara logika semua objek pada akhirnya hanya akan membentuk sebuah *cluster*. Dendrogram biasanya digunakan untuk membantu memperjelas proses hierarki tersebut.

Berbeda dengan metode *hierarchical clustering*, metode *non-hierarchical clustering* justru dimulai dengan menentukan terlebih dahulu jumlah cluster yang diinginkan (dua cluster, tiga cluster, atau lain sebagainya). Setelah jumlah cluster diketahui, baru proses cluster dilakukan tanpa mengikuti proses hierarki. Metode ini biasa disebut dengan *K-Means Clustering*.

Adapun langkah-langkah melakukan clustering dengan metode *K-Means* adalah sebagai berikut [12]:

1. Menentukan banyaknya cluster (*k*) untuk jumlah cluster dari dataset yang ada.
2. Menentukan *k* sebagai Centroid, biasanya dilakukan secara acak (random).
3. Hitung jarak data dengan centroid menggunakan rumus jarak menggunakan rumus Euclidean

$$D(i, j) = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

$D(i, j)$ = Jarak data ke *i* ke pusat cluster *j*

X_{ki} = Data ke *i* pada atribut data ke *k*

X_{kj} = Titik pusat ke *j* pada atribut ke *k*

4. Kelompokkan data berdasarkan kedekatan dengan centroid kemudian perbaharui nilai centroid baru dengan lokasi dari pusat cluster menggunakan persamaan.

$$\mu_j(t+1) = \frac{1}{N_{sj}} \sum_{j \in s_j} x_j$$

Dimana $\mu_j(t+1)$ merupakan centroid baru pada iterasi ke (t+1) dan N_{sj} merupakan banyaknya data pada cluster *sj*.

5. Lakukan langkah 2 sampai 4 sampai anggota tiap cluster tidak ada yang berubah.

2.3 Unified Modeling Language (UML)

UML (*Unified Modelling Language*) diagram memiliki tujuan utama untuk membantu tim pengembangan proyek berkomunikasi, mengeksplorasi potensi desain, dan memvalidasi desain arsitektur perangkat lunak atau pembuat program. Komponen atau notasi UML diturunkan dari 3 (tiga) notasi yang telah ada sebelumnya yaitu Grady Booch, OOD (*Object-Oriented Design*), Jim Rumbaugh, OMT (*Object Modelling Technique*), dan Ivar Jacobson OOSE (*Object-Oriented Software Engineering*).

UML (*Unified Modelling Language*) adalah suatu alat untuk memvisualisasikan dan mendokumentasikan hasil analisa dan desain yang berisi sintak dalam memodelkan sistem secara visual. Juga merupakan satu kumpulan konvensi pemodelan yang digunakan untuk menentukan atau menggambarkan sebuah sistem *software* yang terkait dengan objek.

3. ANALISIS DAN HASIL

3.1 Analisis

Dalam menentukan pengelompokan data penjualan sembako digunakan beberapa jenis data diantaranya yaitu data variabel, data primer dari perusahaan dan data hasil inialisasi. Dalam aplikasi data mining menentukan pengelompokan data penjualan sembako, maka harus ditetapkan variabel-variabel yang digunakan sebagai acuan untuk penilaian dalam proses pengujian. Variabel-variabel tersebut dapat dilihat pada tabel di bawah ini :

Tabel 3.2 Variabel Yang Digunakan

No	Kode	Variabel	Keterangan
1	V1	Stok	Stok barang sembako
2	V2	Jumlah terjual	Jumlah terjual barang sembako
3	V3	Harga satuan	Harga satuan dari barang sembako

Agar data pada tabel 3.1 di atas dapat diolah dengan menggunakan metode k-means clustering, maka data harga harus disederhanakan agar hasil pengelompokan lebih optimal, dengan cara harga barang dibagi 1000, sehingga menghasilkan tabel hasil penyederhanaan seperti dijelaskan pada tabel di bawah berikut ini.

Tabel 3.3 Data Penjualan Sembako

Kode	Jenis Barang	Stok	Jlh Terjual	Harga Satuan
01	Beras 5 kg	25	16	52
02	Minyak goreng 1 liter	30	18	12
03	Detergen bubuk 1 kg	26	20	14,5
04	Pasta gigi	28	22	12
05	Sabun mandi	29	17	3,5
06	Susu kental manis	35	21	11
07	Kecap manis	21	16	7,5
08	Saus	22	15	6,5
09	Beras 10 kg	32	20	105
10	Gula 1kg	45	33	11
11	Minyak goreng 2 liter	22	15	27,5
12	Gas elpiji 3 kg	38	30	19
13	Mie Instan	20	18	3
14	Beras 30 kg	15	10	295
15	Pewangi lantai	20	11	7,5
16	Tissu wajah	19	15	10
17	Cairan cuci piring	34	24	6,5
18	Telur papan	23	18	34
19	Tepung beras	32	15	12
20	Gula Merah	17	15	21
21	Air mineral sedang	28	22	3
22	Air mineral besar	33	23	5
23	Roti kemasan	34	21	6
24	Detergen bubuk 500 gr	22	15	6
25	Garam	25	11	5,5

Algoritma k-means clustering dalam menentukan pengelompokan data penjualan sembako di Toko Syihan dapat dijabarkan sebagai berikut.

1. Iterasi 1

Untuk menentukan pusat (centroid) awal pada iterasi ke-1 ini ditentukan dengan acak (random) dari data yang sudah ada. Pada kasus ini pusat centroid awal adalah data ke-6, data ke-13 dan data ke-18:

Tabel 3.4 Titik Pusat (Centroid) Awal Cluster

Centroid	Kode	Alamat	V1	V2	V3
Centroid 1	06	Susu kental manis	35	21	11
Centroid 2	13	Mie Instan	20	18	3
Centroid 3	18	Telur papan	23	18	34

Perhitungan jarak dari data ke-1 terhadap titik pusat cluster adalah sebagai berikut di bawah ini:

$$D(1,1) = \sqrt{(25 - 35)^2 + (16 - 21)^2 + (52 - 11)^2} \\ = 42,50$$

$$D(1,2) = \sqrt{(25 - 20)^2 + (16 - 18)^2 + (52 - 3)^2}$$

$$= 49,30$$

$$D(1,3) = \sqrt{(25 - 23)^2 + (16 - 18)^2 + (52 - 34)^2}$$

$$= 18,22$$

Dan seterusnya dilakukan perhitungan jarak untuk data ke-2 sampai data ke-25 dengan rumus seperti di atas. Sehingga akan didapatkan hasil perhitungan jarak setiap data terhadap pusat cluster baru sebagai berikut dijelaskan pada tabel di bawah ini:

Tabel 3.5 Hasil Perhitungan Distance Score Iterasi 1

Kode	Jenis Barang	C1	C2	C3	Cluster
01	Beras 5 kg	42,50	49,30	18,22	3
02	Minyak goreng 1 liter	5,92	13,45	23,09	1
03	Detergen bubuk 1 kg	9,71	13,12	19,83	1
04	Pasta gigi	7,14	12,69	22,91	1
05	Sabun mandi	10,40	9,07	31,10	2
06	Susu kental manis	0,00	17,26	26,12	1
07	Kecap manis	15,27	5,02	26,65	2
08	Saus	15,01	5,02	27,68	2
09	Beras 10 kg	94,05	102,72	71,60	3
10	Gula 1kg	15,62	30,23	35,19	1
11	Minyak goreng 2 liter	21,85	24,76	7,23	3
12	Gas elpiji 3 kg	12,41	26,91	24,37	1
13	Mie Instan	17,26	0,00	31,14	2
14	Beras 30 kg	284,92	292,15	261,25	3
15	Pewangi lantai	18,36	8,32	27,57	2
16	Tissu wajah	17,12	7,68	24,52	2
17	Cairan cuci piring	5,50	15,63	30,22	1
18	Telur papan	26,12	31,14	0,00	3
19	Tepung beras	6,78	15,30	23,96	1
20	Gula Merah	21,45	18,49	14,63	3
21	Air mineral sedang	10,68	8,94	31,65	2
22	Air mineral besar	6,63	14,07	31,08	1
23	Roti kemasan	5,10	14,63	30,23	1
24	Detergen bubuk 500 gr	15,17	4,69	28,18	2
25	Garam	15,17	8,96	29,42	2

Setelah diketahui anggota tiap-tiap cluster kemudian pusat cluster baru dihitung berdasarkan data anggota tiap-tiap cluster sesuai dengan rumus pusat anggota cluster. Perhitungan pusat cluster baru dicari dengan menghitung nilai rata-rata (*average*).

Selanjutnya hitung data pada cluster sampai berdasarkan jumlah anggota yang dimiliki, sehingga didapat hasil pembentukan pusat centroid baru dari nilai *average* pada setiap cluster sebagai berikut:

Tabel 3.9 Hasil Pembentukan Pusat Cluster Baru

Centroid	V1	V2	V3
Centroid 1	33,5	22,7	10,9
Centroid 2	22,889	15,556	5,833
Centroid 2	22,333	15,667	89,083

2. Iterasi Ke-2

Setelah dilakukan perhitungan pusat centroid maka selanjutnya menugaskan kembali setiap objek/data dengan pusat cluster baru. Dengan cara melakukan perhitungan jarak pada iterasi ke-2 sebagai berikut :

a. Perhitungan jarak dari data ke-1 terhadap pusat cluster adalah:

$$D(1,1) = \sqrt{(25 - 33,5)^2 + (16 - 22,7)^2 + (52 - 10,9)^2}$$

$$= 42,501$$

$$D(1,2) = \sqrt{(25 - 22,889)^2 + (16 - 15,556)^2 + (52 - 5,833)^2}$$

$$= 46,217$$

$$D(1,3) = \sqrt{(25 - 22,333)^2 + (16 - 15,667)^2 + (52 - 89,083)^2}$$

$$= 37,180$$

Dan seterusnya dilakukan perhitungan jarak untuk data ke-2 sampai data ke-25 dengan rumus seperti di atas. Sehingga akan didapatkan hasil perhitungan jarak setiap data terhadap pusat cluster baru sebagai berikut :

Tabel 3.10 Hasil Perhitungan Distance Score pada Iterasi 2

Kode	Jenis Barang	C1	C2	C2	Cluster
01	Beras 5 kg	42,501	46,217	37,180	3
02	Minyak goreng 1 liter	5,962	9,725	77,498	1
03	Detergen bubuk 1 kg	8,746	10,225	74,799	1
04	Pasta gigi	5,652	10,280	77,550	1
05	Sabun mandi	10,368	6,699	85,853	2
06	Susu kental manis	2,269	14,248	79,283	1
07	Kecap manis	14,584	2,558	81,595	2
08	Saus	14,522	1,243	82,586	2
09	Beras 10 kg	94,151	99,684	19,120	3
10	Gula 1kg	15,439	28,634	83,134	1
11	Minyak goreng 2 liter	21,612	21,692	61,588	1
12	Gas elpiji 3 kg	11,796	24,705	73,229	1
13	Mie Instan	16,332	4,727	86,146	2
14	Beras 30 kg	284,985	289,328	206,125	3
15	Pewangi lantai	18,185	5,646	81,750	2
16	Tisu wajah	16,442	5,727	79,156	2
17	Cairan cuci piring	4,615	13,971	83,818	1
18	Telur papan	25,806	28,273	55,136	1
19	Tepung beras	7,921	11,016	77,690	1
20	Gula Merah	20,822	16,280	68,295	2
21	Air mineral sedang	9,651	8,699	86,501	2
22	Air mineral besar	5,929	12,583	85,074	1
23	Roti kemasan	5,211	12,374	84,068	1
24	Detergen bubuk 500 gr	14,682	1,062	83,086	2
25	Garam	15,437	5,032	83,756	2

Dari hasil perhitungan metode k-means di atas diperoleh 3 cluster dengan penjelasan sebagai berikut:

1. Cluster 1 memiliki 12 anggota cluster dengan karakteristik jenis sembako yang jumlah terjualnya cukup banyak perharinya dan harga yang relatif tidak terlalu mahal.
2. Cluster 2 merupakan cluster dengan 10 anggota cluster. Karakteristik jenis sembako pada cluster 2 ini merupakan barang-barang dengan harga yang murah dan memiliki stok yang banyak
3. Cluster 3 merupakan cluster dengan jumlah anggota terkecil yaitu 3 anggota cluster. Karakteristik jenis sembako pada cluster ini merupakan jenis sembako utama dan memiliki harga yang cukup mahal.

3.2 Hasil

Implementasi sistem menjelaskan mengenai hasil sistem yang telah dibangun. Terdiri dari beberapa form input dan beberapa laporan. Berikut di bawah ini dijelaskan lebih detail.

1. Form Data Penjualan

Tampilan ini berisikan tentang data penjualan yang berfungsi sebagai media dalam memasukan data penjualan baru dan juga mengedit serta menghapus data penjualan. Tampilan *form* dirancang agar mudah untuk digunakan oleh *user*. Adapun tampilan *form* sebagai berikut:

Kode Barang	Jenis Barang	Stok	Jln Terjual	Harga Satuan	Satuan
001	Beras 5 kg	25	16	52000	Kilogram
002	Minyak goreng 1 liter	30	18	12000	Liter
003	Detergen bubuk 1 kg	26	20	14500	Kilogram
004	Pasta gigi	28	22	12000	Pcs
005	Sabun mandi	29	17	3500	Pcs
006	Susu kental manis	35	21	11000	Pcs
007	Kecap manis	21	16	7500	Pcs
008	Saus	22	15	6500	Pcs
009	Beras 10 kg	19	20	100000	Kilogram

Gambar Form Data Penjualan

2. Form Proses Titik Cluster

Tampilan Form Proses Titik Cluster ini berfungsi untuk menampilkan hasil inialisasi dan memilih 3 data sebagai titik pusat cluster. Adapun cara penggunaannya dengan terlebih dahulu memilih (*double click*) pada *listview* pertama maka titik pusat cluster akan tampil pada *listview* kedua. Tampilan form dapat dilihat pada gambar berikut.

Kode Barang	JenisBarang	stok	jlnTerjual	harga
010	Gula 1kg	45	33	11
011	Minyak goreng 2 liter	22	15	28
012	Gas elpiji 3 kg	38	30	19
013	Mie Instan	20	18	3
014	Beras 30 kg	15	10	295
015	Pewangi lantai	20	11	8
016	Tissu wajah	19	15	10
017	Cairan cuci piring	34	24	6
018	Telur papan	23	18	34

Kode Barang	JenisBarang	stok	jlnTerjual	harga
006	Susu kental manis	35	21	11
013	Mie Instan	20	18	3
018	Telur papan	23	18	34

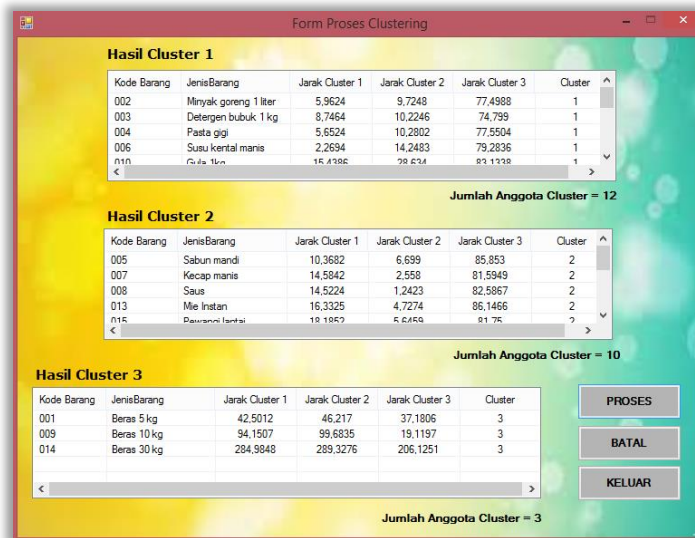
Gambar Tampilan Form Proses Titik Cluster

Adapun fungsi-fungsi dari tombol yang terdapat dalam form yaitu :

- Simpan : Menyimpan data Proses Titik Cluster baru
- Batal : Membatalkan proses data dan membersihkan form

3. Form Proses Clustering

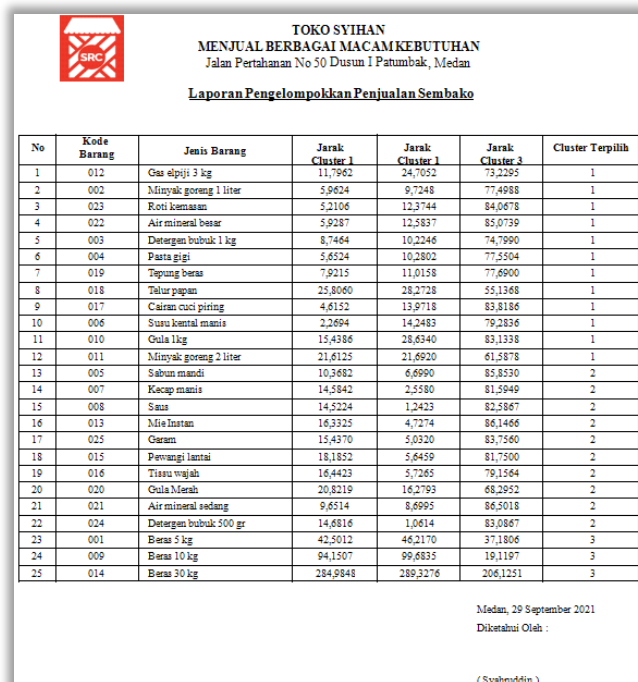
Tampilan Form Proses Clustering ini berfungsi untuk melakukan proses perhitungan jarak data ke titik cluster menggunakan metode *K-Means Clustering* dan menampilkan hasil perhitungan. Adapun hasil perhitungannya tampil dalam bentuk listview. Klik tombol Proses untuk memulai perhitungan dengan metode *K-Means Clustering*. Tampilan form sebagai berikut :



Gambar Tampilan Hasil Form Proses Clustering

4. Laporan

Pengujian program yang dilakukan menghasilkan output berupa laporan hasil analisa. Form Laporan ini berfungsi untuk melihat hasil perhitungan Cluster dengan metode K-Means Clustering beserta informasi lain mengenai Titik Cluster tersebut. Adapun hasil Cluster akan tampil pada kolom Cluster. Tampilan preview dapat dilihat dibawah ini:



Gambar Laporan Hasil Clustering

4. KESIMPULAN

Berdasarkan pembahasan dan evaluasi dari bab terdahulu, maka dapat ditarik beberapa kesimpulan sebagai berikut :

1. Dalam menganalisa pola penjualan sembako di Toko Syihan dilakukan dengan menganalisa data mentah penjualan sembako dari Toko Syihan kemudian data tersebut diinisialisasikan kedalam bentuk angka lalu dilakukan proses perhitungan distance score sehingga terbentuk hasil cluster (hasil pengelompokkan).
2. Perancangan aplikasi data mining yang mengadopsi algoritma K-Means Clustering diperlukan sebuah bahasa pemrograman Visual Basic .Net 2010 serta sebuah aplikasi database Microsoft Access. Dan dalam pemodelan sistem menggunakan pemodelan Unified Modeling Language (UML).
3. Implementasi sistem dilakukan dengan mengintegrasikan algoritma K-Means Clustering kedalam baris kode program. Sehingga menghasilkan laporan hasil pengelompokkan yang dapat membantu pemilik dalam melihat pengelompokan penjualan sembako.

REFERENSI

- [1] C. A. Sugianto, A. H. Rahayu, and A. Gusman, "Algoritma K-Means Untuk Pengelompokkan Penyakit Pasien Pada Puskesmas Cigugur Tengah," no. August, 2020, doi: 10.47292/joint.v2i2.30
- [2] Wirdah Choiriah, "Penggunaan Algoritma Apriori Data Mining Untuk Mengetahui Tingkat Kesetiaan Konsumen (*Brand Loyalty*) Terhadap Merek Kendaraan Bermotor (Studi Kasus Dealer Honda Rumbai)," no. Februari, 2016.
- [3] Baginda Harahap, "Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung)," 2019.
- [4] E. Iswandy, "Analisa Dan Perancangan Sistem Informasi Penagihan Purchasing Order Customer Studi Kasus Pada Cv. Vertical Cipta Relasi Padang Dengan Metode Centralized Data Processing," *J. TEKNOIF*, vol. 4, no. Oktober, p. 14, 2016.
- [5] R. Yanto and R. Khoiriah, "Implementasi Data Mining dengan Metode Algoritma Apriori dalam Menentukan Pola Pembelian Obat," pp. 102–113.
- [6] A. Setiawan, et al, "Rancang Bangun Edugame The World of Word Berbasis unity 3D Dengan Implementasi Speech Recognition" *Jurnal Teknik Informatika*, 2017.
- [7] A. Rifa'i, G. Guntoro Setiaji, and V. Vydia, "PENGUNAAN METODE K-MEANS PADA ANALISA DAN KLASIFIKASI CAPRES 2019 DI TWITTER," *Pengemb. Rekayasa dan Teknol.*, vol. 15, no. 1, pp. 43–47, 2019, [Online]. Available: <http://journals.usm.ac.id/index.php/jprt/index>.
- [8] N. Dwitri, J. A. Tampubolon, S. Prayoga, and P. P. P. A. N. W. F. I. R. H. Zer, "PENERAPAN ALGORITMA K-MEANS DALAM MENENTUKAN TINGKAT PENYEBARAN PANDEMI COVID-19 DI INDONESIA," vol. 4, no. 1, pp. 128–132, 2020.
- [9] F. Indriyani and E. Irfiani, "Clustering Data Penjualan Pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means," 2019.
- [10] Indra Griha and George Pri Hartawan, "Perancangan Aplikasi Koperasi Simpan Pinjam berbasis Web", *Jurnal Ilmiah Ilmu Ekonomi*, 2017.
- [11] E. Affandi and T. Syahputra STMIK Triguna Dharma, "J-SISKO TECH Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD Pemodelan Uml Manajemen Sistem Inventory," vol. 1, no. 2, pp. 14–25, 2018.

BIOGRAFI PENULIS

	<p>Nama : Nurin Fadhilah Adani Nirm : 2017021133 Program Studi : Sistem Informasi STMIK Triguna Dharma Deskripsi : Mahasiswa Stambuk 2017 pada Program Studi Sistem Informasi yang memiliki minat dan fokus dalam Desain</p>
	<p>Nama : Ahmad Fitri Boy, S.Kom., M.Kom NIDN : 0104058001 Program Studi : Sistem Informasi STMIK Triguna Dharma Deskripsi : Dosen Tetap STMIK Triguna Dharma yang aktif mengajar Prestasi : Pemenang hibah PDP sebanyak 1 kali pada tahun yang sama yaitu 2015 dan dosen terfavorit 2011</p>
	<p>Nama : Rendy Syahputra, S.Kom., M.Kom NIDN : 0125059201 Program Studi : Sistem Informasi STMIK Triguna Dharma Deskripsi : Dosen Tetap STMIK Triguna Dharma yang aktif mengajar dan fokus pada bidang ilmu komputer dengan bidang keilmuan perakitan komputer dan jaringan</p>