

PENERAPAN METODE CLUSTERING K-MEANS DALAM PENGELOMPOKAN PENJUALAN PRODUK

Yulia Darmi¹, Agus Setiawan²

^{1,2}Program Studi Teknik Informatika Universitas Muhammadiyah Bengkulu
Jl. Bali, Kel. Kampung Bali, Kec. Teluk Segara, Kota Bengkulu, Bengkulu 38119

¹yuliadarmi@yahoo.com

²agus_setiawan@yahoo.com

ABSTRACT

In the minimarket, the product is a staple for sale. Products in the minimarket are among products that sell well with unsold products. Given this problem, it is necessary to create a system that can classify sellable products and non-salable products, conducted in MM.TIKA Minimarket Bengkulu and implemented in June to July 2015. K-Means algorithm is not affected to the order of objects used, this is proven when the author tries to randomly determine the starting point of the cluster center of one of the objects at the beginning of the calculation. The number of cluster members generated amounts to the same when using other objects as the starting point of the cluster center. However, this only affects the number of iterations performed. The purpose to make the application and analyze the sales of products on supermarket MM.TIKA Bengkulu City using K-Means method. With this system can provide the ease of ease to analyze the grouping of product sales at supermarket MM.TIKA Bengkulu City, determine and classify sales of products that sell and sell less. K-Means Cluster Analysis algorithm basically can be applied to problems in understanding consumer behavior, identifying new product opportunities in the market and K-Means algorithm can also be used to summarize objects of large quantities making it easier to describe the properties or characteristics of each- Each group. The methods used in data collection are observation and interview. With this application, minimarket owners can see the results of grouping products that sell and not sell. So, if there are products that do not sell, minimarket owners can find other alternatives for unsold products can be sold.

Keywords: K-Means Clustering, Data Mining, Product

INTISARI

Pada minimarket, produk merupakan bahan pokok yang dijual belikan. Produk di minimarket terdapat antara produk yang laku dengan produk yang tidak laku. Dengan adanya masalah ini, perlu untuk menciptakan sistem yang dapat mengelompokkan produk laku dan produk yang tidak laku, yang dilakukan di Minimarket MM.TIKA Bengkulu dan dilaksanakan pada bulan Juni sampai Juli 2015. Algoritma K-Means tidak terpengaruh terhadap urutan objek yang digunakan, hal ini dibuktikan ketika penulis mencoba menentukan secara acak titik awal pusat cluster dari salah satu objek pada permulaan perhitungan. Jumlah keanggotaan cluster yang dihasilkan berjumlah sama ketika menggunakan objek yang lain sebagai titik awal pusat cluster tersebut. Namun, hal ini hanya berpengaruh pada jumlah iterasi yang dilakukan. Tujuan untuk membuat aplikasi dan menganalisis penjualan produk pada swalayan MM.TIKA Kota Bengkulu menggunakan metode K-Means. Dengan adanya sistem ini dapat memberikan manfaat kemudahan untuk menganalisis pengelompokkan penjualan produk pada swalayan MM.TIKA Kota Bengkulu, menentukan dan mengklasifikasikan penjualan produk yang laku dan kurang laku. Algoritma K-Means Cluster Analysis pada dasarnya dapat diterapkan pada permasalahan dalam memahami perilaku konsumen, mengidentifikasi peluang produk baru dipasaran dan algoritma K-Means ini juga dapat digunakan untuk meringkas objek dari jumlah besar sehingga lebih memudahkan untuk mendiskripsikan sifat- sifat atau karakteristik dari masing-masing kelompok. Metode yang digunakan pengumpulan data adalah observasi dan wawancara. Dengan adanya aplikasi ini, pemilik minimarket dapat melihat hasil pengelompokkan produk yang laku dan tidak laku. Maka, bila terdapat produk yang tidak laku, pemilik minimarket dapat mencari alternative lain agar produk yang tidak laku dapat menjadi laku.

Kata Kunci : *K-Means, Clustering, Data Mining, Produk*

I. PENDAHULUAN

Data mining merupakan suatu langkah dalam *knowlegde discovery in databases* (KDD) yang memiliki teknik menganalisa data untuk digali informasi tersembunyi dalam jumlah besar dan kompleks, sehingga menghasilkan output berupa

karakteristik atau pola dari data tersebut. Salah satu teknik analisa *Data Mining* adalah analisis kelompok (*cluster analysis*) yang lebih dikenal dengan *Clustering*. *Clustering* merupakan metode analisis data yang tujuannya mengelompokkan data dengan karakteristik yang sama ke suatu wilayah yang sama. Salah satu pendekatan yang digunakan dalam

mengembangkan metode *clustering* yaitu metode *K-Means*, dimana metode ini merupakan salah satu metode pengelompokan data *nonhierarki* (sekatan) yang berusaha mempartisi data ke dalam bentuk dua atau lebih kelompok (*cluster*) yang berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama.

Data Mining sudah banyak diterapkan dalam berbagai bidang, salah satunya metode *clustering* khususnya metode *K-Means*, antara lain oleh Andika (2008) menjelaskan tentang algoritma *K-Means clustering* dapat diimplementasikan dengan model perangkat lunak untuk verifikasi citra sidik jari poin *minutiae*. Suprihatin (2011) menyatakan bahwa teknik *clustering K-Means* digunakan untuk menentukan nilai huruf ujian akhir pada Universitas Ahmad Dahlan (UAD).

Swalayan MM.TIKA merupakan salah satu bidang usaha yang bergerak di bidang penjualan kebutuhan sehari-hari. Di mana swalayan ini setiap harinya harus memenuhi kebutuhan konsumen. Untuk dapat melakukan hal tersebut, maka membutuhkan sumber informasi yang cukup banyak untuk dapat dianalisis lebih lanjut. Pada Swalayan MM.TIKA, terdapat beberapa permasalahan yang kerap muncul mengenai penjualan.

Berdasarkan hal tersebut maka diharapkan bisa mempermudah pihak Swalayan MM. TIKA dalam menemukan informasi dari tumpukan-tumpukan data yang berguna bagi Swalayan MM. TIKA. Tidak hanya itu penulis juga mengharapkan bisa memberikan rangsangan kepada pengelola data supaya lebih aktif dalam menggali informasi yang dibutuhkan oleh Swalayan MM.TIKA, sehingga bisa menjawab kebutuhan dari Swalayan. Penulis berharap dapat membantu memberikan informasi mengenai hasil *Clustering* data penjualan.

Dengan demikian, dibutuhkan analisis yang terkomputerisasi menggunakan *software* data mining yang menunjang arus data dan informasi sesuai dengan kebutuhan dari proses-proses tersebut.

II. TINJAUAN PUSTAKA

A) Algoritma klasifikasi K-Means

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. Kemudian tetapkan nilai-nilai K secara random, untuk sementara nilai tersebut menjadi pusat dari cluster atau biasa disebut dengan centroid, mean atau “means”. Hitung jarak setiap data yang ada terhadap masing-masing centroid menggunakan rumus Euclidian hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Klasifikasikan setiap data

berdasarkan kedekatannya dengan centroid. Lakukan langkah tersebut hingga nilai centroid tidak berubah (stabil) (Snati, 2012 :54).

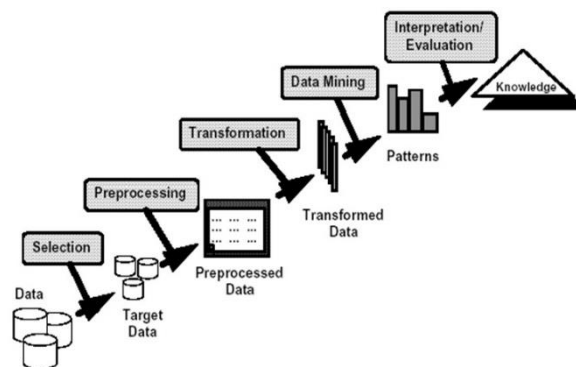
B) Data Mining

Menurut Widodo (2013:1) Data mining adalah analisa terhadap data untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut. Sedangkan menurut Sulianta (2010:17) Data mining adalah metoda yang digunakan untuk mengekstraksi informasi prediktif tersembunyi pada database, ini adalah teknologi yang sangat potensial bagi perusahaan yang sangat potensial bagi perusahaan dalam memberdayakan data *warehouse*.

Secara garis besar, data mining dapat dikelompokkan menjadi 2 kategori utama, yaitu:

Deskriptive mining, yaitu proses untuk menemukan karakteristik penting dari data dalam satu basis data. Teknik data mining yang termasuk *descriptive mining* adalah *clustering*, *asosiation*, dan *sequential mining*.

Predictive, yaitu proses untuk menemukan pola dari data dengan menggunakan beberapa variable lain di masa depan. Salah satu teknik yang terdapat dalam *predictive mining* adalah klasifikasi. Secara sederhana data mining biasa dikatakan sebagai proses penyaring atau “menambang” pengetahuan dari sejumlah data yang besar. Istilah lain untuk data miing adalah *Knowlegde Discovery in Database (KDD)*. Walaupun data mining sendiri adalah bagian dari tahapan proses KDD seperti yang terlihat pada Gambar 1.



Gambar 1. KDD

C) Clustering

Menurut Widodo (2013:9) *Clustering* atau klasifikasi adalah metode yang digunakan untuk membagi rangkaian data menjadi beberapa group berdasarkan kesamaan-kesamaan yang telah ditentukan sebelumnya. *Cluster* adalah sekelompok atau sekumpulan objek-objek data yang similar satu sama lain dalam *cluster* yang sama dan dissimilar

terhadap objek-objek yang berbeda cluster. Objek akan dikelompokkan ke dalam satu atau lebih cluster sehingga objek-objek yang berada dalam satu cluster akan mempunyai kesamaan yang tinggi antara satu dengan yang lainnya. Objek-objek dikelompokkan berdasarkan prinsip memaksimalkan kesamaan objek pada cluster yang sama dan memaksimalkan ketidaksamaan pada cluster yang berbeda. Kesamaan objek biasanya diperoleh dari nilai-nilai atribut yang menjelaskan objek data, sehingga objek-objek data biasanya dipresentasikan sebagai sebuah titik dalam ruang multidimensi.

Dengan menggunakan clustering ini, kita dapat mengkalsifikasikan daerah yang padat, menemukan pola-pola distribusi secara keseluruhan, dan menemukan keterkaitan yang menarik antara atribut data. Dalam data mining, usaha difokuskan pada metode-metode penemuan untuk cluster pada basis data berukuran besar secara efektif dan efisien. Beberapa kebutuhan clustering dalam data mining meliputi skalabilitas, kemampuan untuk menangani tipe atribut yang berbeda mampu menangani dimensionalitas yang tinggi, menangani data yang mempunyai noise, dan dapat diterjemakan dengan mudah.

Adapun tujuan dari data *clustering* ini adalah untuk meminimalisasikan objektif function yang di-set dalam proses *clustering*, yang pada umumnya berusaha meminimalisasikan variasi dalam suatu *cluster*. Dan meminimalisasikan variasi antar cluster.

Secara garis besar, terdapat beberapa metode klasifikasi data. Pemilihan metode *clustering* tergantung pada tipe data dan tujuan *clustering* itu sendiri.

D) Metode K-means

K-Means merupakan salah satu metode data clustering non hierarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbedadikelompokkan ke dalam kelompok yang lainnya.

K-Means adalah metode clustering berbasis jarak yang membagi data ke dalam sejumlah cluster dan algoritma ini hanya bekerja pada atribut numeric. Algoritma *K-Means* termasuk partitioning clustering yang memisahkan data ke k daerah bagian yang terpisah. Algoritma *K-Means* sangat terkenal karena kemudahan dan kemampuannya untuk mengcluster data yang besar dan data outlier dengan sangat cepat. Dalam algoritma *K-Means*, setiap data harus termasuk ke cluster tertentu dan bisa dimungkinkan bagi setiap data yang termasuk cluster tertentu pada suatu tahapan proses, pada tahapan berikutnya berpindah kecluster lainnya.

Pada dasarnya penggunaan algoritma dalam melakukan proses clustering tergantung dari data yang ada dan konklusi yang ingin dicapai. Untuk itu digunakan algoritma *K-Means* yang didalamnya membuat aturan sebagai berikut:

- 1) Jumlah Cluster perlu diinputkan
- 2) Hanya memiliki atribut bertipe numeric.

Algoritma *K-Means* merupakan metode non-hierarki yang pada awalnya mengambil sebagian banyaknya komponen populasi untuk dijadikan pusat cluster awal. Pada tahap ini pusat cluster dipilih secara acak dari sekumpulan populasi data. Berikutnya *K-Means* menguji masing-masing komponen di dalam populasi data dan menandai komponen tersebut ke salah satu pusat cluster yang telah didefinisikan tergantung dari jarak minimum antar komponen dengan tiap-tiap cluster. Posisi pusat cluster akan dihitung kembali sampai semua komponen data digolongkan ke dalam tiap-tiap pusat cluster dan terakhir akan terbentuk posisi pusat cluster yang baru.

Algoritma *K-Means* pada dasarnya melakukan dua proses, yakni proses pendeteksian lokasi pusat tiap cluster dan proses pencarian anggota dari tiap-tiap cluster. Cara kerja algoritma *K-Means*:

- 1) Tentukan k sebagai jumlah cluster yang ingin dibentuk.
- 2) Bangkitkan k centroid (titik pusat cluster) awal secara random.
- 3) Hitung jarak setiap data ke masing-masing centroid.
- 4) Setiap data memilih centroid yang terdekat.
- 5) Tentukan posisi centroid yang baru dengan cara menghitung nilai rata-rata dari data-data yang terletak pada centroid yang sama.
- 6) Kembali ke langkah-3 jika posisi centroid baru dengan centroid yang lama tidak sama.

E) Produk

Produk menurut Kotler dan Armstrong (1996:274) adalah : “*A product as anything that can be offered to a market for attention, acquisition, use or consumption and that might satisfy a want or need*”. Artinya produk adalah segala sesuatu yang ditawarkan ke pasar untuk mendapatkan perhatian, dibeli, dipergunakan dan yang dapat memuaskan keinginan atau kebutuhan konsumen. Menurut Stanton, (1996:222), “*A product is asset of tangible and intangible attributes, including packaging, color, price quality and brand plus the services and reputation of the seller*”. Artinya suatu produk adalah kumpulan dari atribut-atribut yang nyata maupun tidak nyata, termasuk di dalamnya kemasan, warna, harga, kualitas dan merk ditambah dengan jasa dan reputasi penjualannya. Menurut Tjiptono (1999:95) secara konseptual produk adalah pemahaman subyektif dari produsen atas “sesuatu” yang bisa

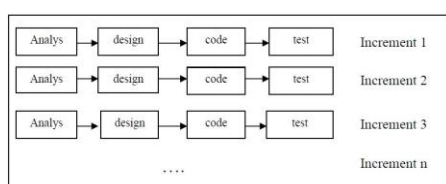
ditawarkan sebagai usaha untuk mencapai tujuan organisasi melalui pemenuhan kebutuhan dan keinginan konsumen, sesuai dengan kompetensi dan kapasitas organisasi serta daya beli. Lima Tingkatan Produk Menurut Kotler (2003:408) ada lima tingkatan produk, yaitu core benefit, basic product, expected product, augmented product dan potential product. Penjelasan tentang kelima tingkatan produk adalah :

- 1) *Core benefit (namely the fundamental service of benefit that costumer really buying)* yaitu manfaat dasar dari suatu produk yang ditawarkan kepada konsumen.
- 2) *Basic product (namely a basic version of the product)* yaitu bentuk dasar dari suatu produk yang dapat dirasakan oleh panca indra.
- 3) *Expected product (namely a set of attributes and conditions that the buyers normally expect and agree to when they purchase this product)* yaitu serangkaian atribut-atribut produk dan kondisi-kondisi yang diharapkan oleh pembeli pada saat membeli suatu produk.
- 4) *Augmented product (namely that one includes additional service and benefit that distinguish the company's offer from competitor's offer)* yaitu sesuatu yang membedakan antara produk yang ditawarkan oleh badan usaha dengan produk yang ditawarkan oleh pesaing.
- 5) *Potential product (namely all of the argumentations and transformations that this product that ultimately undergo in the future)* yaitu semua argumentasi dan perubahan bentuk yang dialami oleh suatu produk dimasa datang. Klasifikasi Produk Banyak klasifikasi suatu produk yang dikemukakan ahli pemasaran, diantaranya pendapat yang dikemukakan oleh Kotler.

III. METODOLOGI PENELITIAN

A) Model Pengembangan Sistem

Dalam penelitian ini model pengembangan sistem yang digunakan yaitu model *incremental* yang merupakan model pengembangan sistem pada *software engineering* berdasarkan *requirement software* yang dipecah menjadi beberapa fungsi atau bagian sehingga model pengembangannya secara bertahap. Model pengembangan sistem dalam penelitian ini ditunjukkan pada Gambar 2.



Gambar 2. Desain Pemodelan *Incremental*

B) Proses metode clustering k-means

Pertama akan kita bahas dulu metode *cluster* secara statistik untuk non hirachical method yaitu: *K-Means Clustering*.

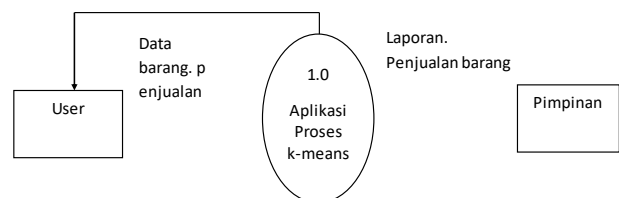
Algoritma:

- 1) Partisi item menjadi K initial cluster
- 2) Lakukan proses perhitungan dari daftar item, tandai item untuk kelompok yang mana berdasarkan pusat (mean) yang terdekat (dengan menggunakan *distance* dapat digunakan *Euclidean distance*). Hitung kembali pusat centroid untuk item baru yang diterima pada cluster tersebut dari cluster yang kehilangan item.
- 3) Ulangi step 2 hingga tidak ada lagi tempat yang akan ditandai sebagai cluster baru.

Pusat awal *cluster* atau *centroid* didapatkan secara random, untuk penentuan awal *cluster* di asumsikan: Pusat Cluster 1: (5. 30) dan Pusat Cluster 2: (25. 6).

C) DFD (Data Flow Diagram)

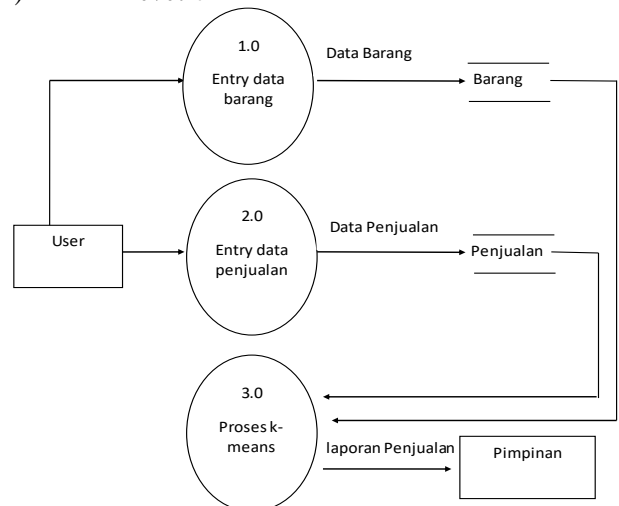
1) Diagram Konteks



Gambar 3. Diagram Konteks

Gambar 3 menunjukan bahwa *user* dan proses clustering k-means saling berhubungan, terlihat *user* mengirimkan data user ke sistem proses *clustering k-means* dan proses *clustering k-means* membalas dengan mengirimkan data barang penjualan dan data barang penjualan mengirimkan ke aplikasi, dari aplikasi membuat laporan penjualan laporan barang kemudian ke pimpinan.

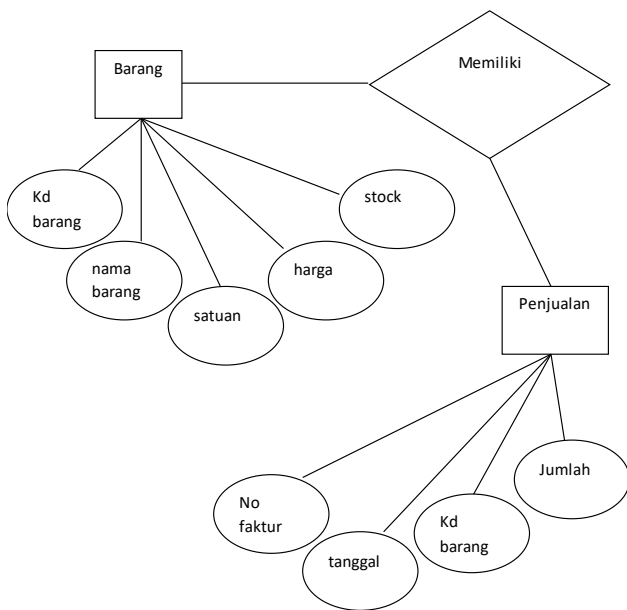
2) DFD Level 0



Gambar 4. DFD Level 0

Berdasarkan Gambar 4 dapat dijelaskan bahwa terdapat tiga buah alur proses yaitu antara data barang, data penjualan dan proses metode k-means. Hubungan pertama data barang menjelaskan proses penginputan data barang dan data penjualan barang, aplikasi menghasilkan sebuah laporan penjualan barang dan mengirimkan laporan data penjualan barang ke pimpinan. Sedangkan hubungan yang kedua yaitu menjelaskan hubungan antara user dan data barang, user menginputkan data penjualan menghasilkan poses clustering k-means di aplikasi.

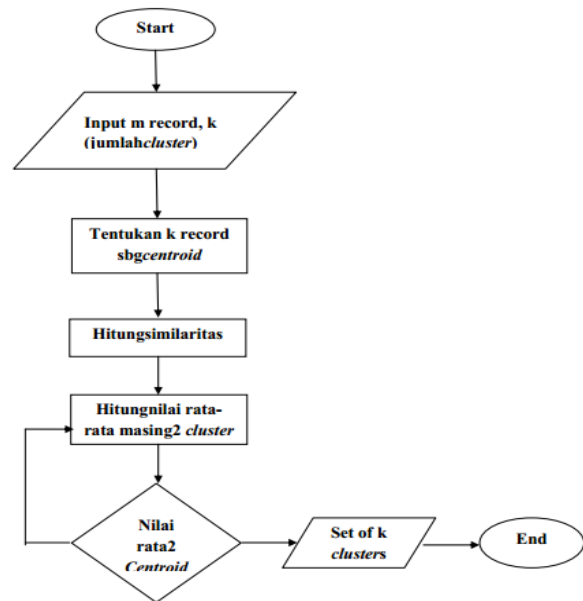
3) *Pemodelan Analisis ERD*



Gambar 5. ERD

E) *Rancangan Menu*

Pada tahap ini akan dilakukan proses utama yaitu segmentasi atau pengelompokkan data Penjualan barang yang diakses dari database, yaitu sebuah metode clustering algoritma K-Means. Berikut ini merupakan diagram flowchart dari algoritma K-Means dengan asumsi bahwa parameter input adalah jumlah dataset sebanyak n data dan jumlah inisialisasi centroid $K=3$ sesuai dengan penelitian. Dari banyak data penjualan yang diperoleh, diambil 14 jenis barang untuk dijadikan sampel untuk penerapan algoritma k-means dalam penjurusan mahasiswa. Percobaan dilakukan dengan menggunakan parameter-parameter berikut.



Gambar 6. Flowchart K-Means

D) *Perancangan File Tabel*

4) *Tabel Data Barang*

Nama File : tbbarang
Primary Key : kd_barang

Tabel 1. Tabel Data Barang

Nama Field	Type	Size	Keterangan
Kd_barang*	Varchar	30	Kode barang
nmbarang	Varchar	8	Nama barang
satuan	Varchar	20	Satuan
harga	Varchar	12,2	Harga
stock	Int	4	Stock

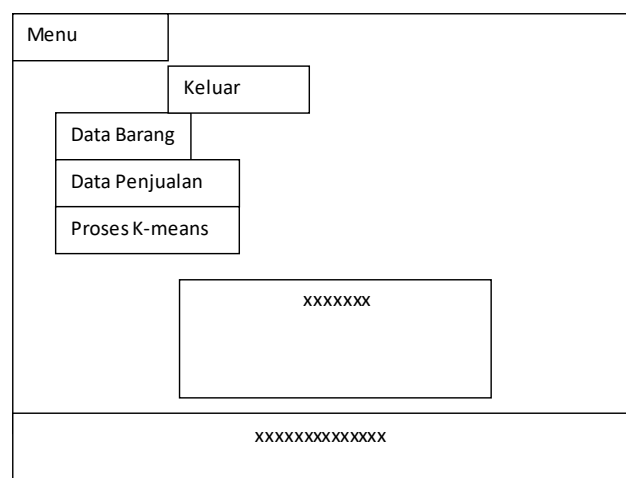
5) *Tabel Data Penjualan*

Nama File : tbpenjualan
Primary Key : no_fak

Tabel 2. Tabel Data tbpenjualan

Nama Field	Type	Size	Keterangan
no_fak*	Varchar	10	No_faktur
tgl	Date	30	Tanggal
kd_barang	Varchar	20	Kode Barang
jml	Intr	5	Jumlah

6) *Rancangan Menu Pembuka*



Gambar 7. Rancangan Menu Pembuka

7) Rancangan Menu Data Barang

xxxxxxxxxxxxxx					
Kode Barang	xxxxxxxxxxxxxx				
Nama Barang	xxxxxxxxxxxxxx				
Satuan	xxxxxxxxxxxxxx				
Harga	xxxxxxxxxxxxxx				
Stock	xxxxxxxxxxxxxx				
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 25%;">Simpan</td> <td style="width: 25%;">Edit</td> <td style="width: 25%;">Hapus</td> <td style="width: 25%;">Keluar</td> </tr> </table>		Simpan	Edit	Hapus	Keluar
Simpan	Edit	Hapus	Keluar		

Gambar 8. Rancangan Menu Data Barang

8) Rancangan Menu Data Penjualan

xxxxxxxxxxxxxx					
No Fak	xxxxxxxxxxxxxx				
Tanggal	Xx/xx/xxxxx				
Kode Barang	xxxxxxxxxxxxxx				
Nama Barang	xxxxxxxxxxxxxx				
Jumlah	xxxxxxxxxxxxxx				
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 25%;">Simpan</td> <td style="width: 25%;">Edit</td> <td style="width: 25%;">Hapus</td> <td style="width: 25%;">Keluar</td> </tr> </table>		Simpan	Edit	Hapus	Keluar
Simpan	Edit	Hapus	Keluar		

Gambar 9. Rancangan Menu Data Penjualan

9) Rancangan Menu Proses K-Means

xxxxxxxxxxxxxx										
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 25%;">xxx</td> <td style="width: 25%;">xxx</td> <td style="width: 25%;">xxx</td> <td style="width: 25%;">xxx</td> </tr> <tr> <td> </td> <td>xxx</td> <td>xxx</td> <td> </td> </tr> </table>	xxx	xxx	xxx	xxx		xxx	xxx		<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="height: 100px; vertical-align: top;">xxxxxxxxxxxxxx</td> </tr> </table>	xxxxxxxxxxxxxx
xxx	xxx	xxx	xxx							
	xxx	xxx								
xxxxxxxxxxxxxx										
<table border="1" style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%;">Proses</td> <td style="width: 50%;">Keluar</td> </tr> </table>		Proses	Keluar							
Proses	Keluar									

Gambar 10. Rancangan Menu Proses K-Means

F) Perancangan Pengujian

Dalam tahapan ini user melakukan pengujian dengan metode pengujian yang dipakai adalah *black box testing*. *black box testing* atau test fungsional adalah pengujian program yang dilakukan oleh pengembang (Programmer) dengan memberikan input tertentu dan melihat hasil yang didapatkan dari input tersebut. dengan kata lain, *black box testing* terfokus pada fungsionalitas sistem. dalam melaksanakan *blackbox testing*, kami menggunakan beberapa kriteria yang akan diujikan. kriteria-kriteria tersebut antara lain :

- 1) Kemampuan Interface sistem untuk menjalankan fungsinya.
- 2) Kemampuan sistem untuk menjalankan fungsi interface.
- 3) Kemampuan sistem untuk menangani input-input form yang berada di luar boundary sistem.
- 4) Kemampuan sistem untuk menangani masalah keamanan.

IV. PEMBAHASAN

A) Analisa Sistem Aktual

Dari hasil penelitian yang dilakukan di Minimarket MM.TIKA Bengkulu, didapatkan sistem yang sudah berjalan dan digunakan saat ini masih manual. Disamping itu, pemilik Minimarket MM.TIKA Bengkulu mempunyai kesulitan dalam mengklasifikasian produk yang laku dengan yang tidak laku. Hal tersebut tentunya akan sulit karena banyaknya produk yang masuk dan keluar setiap harinya.

B) Analisa Sistem Baru

Dari kekurangan sistem yang sedang berjalan tersebut maka dibuatlah suatu aplikasi pengelompokkan penjualan produk dengan menerapkan metode k-means dengan menggunakan bahas pemograman visal basic 6.0. Aplikasi ini nantinya dapat mengklasifikasikan penjualan produk yang laku dan kurang laku. Sehingga pemesanan barang yang kurang laku dapat dikurangi.

C) Perhitungan jarak pusat cluster

Untuk mengukur jarak antara data dengan pusat cluster digunakan Euclidian distance, kemudian akan didapatkan matrik jarak sebagai berikut :

Rumus euclidian distance :

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

X = Pusat cluster

Y = data

Dari 5 data yang dijadikan sampel telah dipilih pusa tawal cluster yaitu C1(5 , 30), dan C2 (25 , 6). Lalu dilakukan penghitungan jarak dari sisa sampel data dengan pusat cluster yang dimisalkan dengan M(a,b), dimana a merupakan total penjualan,dan b harga barang yang diperkecil menjadi angka puluhan juta agar cara penghitungan lebih mudah.

- M1= (14500,5)
- M2= (14.500,4)
- M3= (14.500,2)
- M4= (14.500,3)
- M5= (14.500,1)
- M6= (4.700,40)
- M7=(12.500,10)
- M8= (7.200,4)
- M9= (13.200,19)

$$M10 = (14.500, 10)$$

$$M11 = (10.750, 3)$$

$$M12 = (5.900, 6)$$

$$M13 = (9.500, 10)$$

$$M14 = (9.600, 7)$$

Hitung Euclidean distance dari semua data ketiap titik pusat pertama :

$$D11 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 5)^2 + (5 - 30)^2} = 14494.16$$

$$D12 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 5)^2 + (5 - 30)^2} = 17.37261$$

$$D13 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 5)^2 + (5 - 30)^2} = 19.13407$$

$$D14 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 5)^2 + (5 - 30)^2} = 20.03117$$

$$D15 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 5)^2 + (5 - 30)^2} = 20.96404$$

$$D16 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 5)^2 + (5 - 30)^2} = 11.2443$$

$$D17 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 5)^2 + (5 - 30)^2} = 15.12786$$

$$D18 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 5)^2 + (5 - 30)^2} = 7.349633$$

$$D19 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 5)^2 + (5 - 30)^2} = 12.53122$$

$$D20 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 5)^2 + (5 - 30)^2} = 16.79792$$

$$D21 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 5)^2 + (5 - 30)^2} = 13.06766$$

Dengan cara yang sama hitung jarak tiap titik ketitik pusat ke-2 dan kita akan mendapatkan:

$$D21 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 25)^2 + (5 - 30)^2} = 14498.08$$

$$D22 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 25)^2 + (5 - 30)^2} = 12.70692$$

$$D23 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 25)^2 + (5 - 30)^2} = 13.13431$$

$$D24 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 25)^2 + (5 - 30)^2} = 12.88363$$

$$D25 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 25)^2 + (5 - 30)^2} = 13.45482$$

$$D26 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 25)^2 + (5 - 30)^2} = 34.35196$$

$$D27 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 25)^2 + (5 - 30)^2} = 11.40157$$

$$D28 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 25)^2 + (5 - 30)^2} = 5.570015$$

$$D29 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 25)^2 + (5 - 30)^2} = 17.39554$$

$$D210 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 25)^2 + (5 - 30)^2} = 13.27904$$

$$D211 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 25)^2 + (5 - 30)^2} = 9.25571$$

Dari hasil penghitungan Euclidean distance, kita dapat membandingkan :

Tabel 3. Hasil Iterasi 1

	C1	C2
M1	14494.16	14498.08
M2	17.37261	12.70692
M3	19.13407	13.13431
M4	20.03117	12.88363
M5	20.96404	13.45482
M6	11.2443	34.35196
M7	15.12786	11.40157
M8	7.349633	5.570015
M9	12.53122	17.39554
M10	16.79792	13.27904
M11	13.06766	9.25571

{M9,M10,M11,M12,M13,M14} :anggota C1

{M1,M2,M3,M4,M5,M6,M7,M8} :anggota C2

Selanjutnya lakukan cek untuk setiap item untuk ditandai kembali. Perhitungan kwadrat jarak (*squared distances*) diberikan pada Tabel 4.

Tabel 4. Perhitungan Kwadrat Jarak

Cluster	Koordinate dari Centroid			
	A	B	C	D
A	0	0	1	9
(BCD)	2	4	5	5

Kita lihat setiap item yang baru telah ditandai untuk cluster berdasarkan centroid(pusat) terdekat maka proses telah dihentikan. Sehingga dengan K=2 cluster maka terbentuk cluster sebagai berikut : A dan (BCD).

Permasalahan yang dikaji dalam tulisan ini adalah bagaimana penggunaan metode K-Means Cluster Analysis dalam pengklasifikasian karakteristik suatu objek, tujuan yang ingin penulis capai adalah mengkaji metode K-Means Cluster Analysis dalam pengklasifikasian karakteristik berdasarkan set variable yang dibentuk. Metode ukuran jarak yang digunakan dalam menghitung jarak objek terhadap centroid yaitu persamaan jarak Euclidian. Pada algoritma K-Means Cluster Analysis terdapat beberapa langkah yang harus dilakukan yaitu sebagai berikut:

- 1) Tentukan jumlah cluster.
- 2) Alokasikan objek ke dalam cluster secara random.
- 3) Hitung centroid sampel yang ada di masing-masing cluster.
- 4) Alokasikan masing-masing objek ke centroid terdekat.
- 5) Kembali ke langkah 3 apabila masih ada objek yang berpindah cluster atau masih ada perubahan nilai centroid, ada yang di atas nilai threshold yang ditentukan atau apabila perubahan nilai pada objective function yang digunakan di atas nilai threshold yang ditentukan.

Sebagai bahan bacaan dalam tulisan ini penulis juga membandingkan beberapa karya ilmiah yang mengupas materi yang sama tentang Analisis Cluster, Rismawan dan Kusumadewi dalam penelitiannya mencoba membangun suatu system untuk mengelompokkan objek yang ada berdasarkan status gizi dan ukuran rangka dari objek yang diambil dengan memasukkan parameter kondisi fisik dari objek orang tersebut. Pengelompokan objek dilakukan dengan menggunakan metode K-Means Cluster yaitu mengelompokkan n buah objek ke dalam k kelas berdasarkan jaraknya dengan pusat kelas. Hasil dari penelitian ini terhadap 20 objek sampel diperoleh 3 kelompok mahasiswa berdasarkan nilai BMI (Body Mass Index) dan ukuran rangka, yaitu BMI normal dengan kerangka besar, BMI obesitas sedang dengan kerangka sedang, dan BMI obesitas berat dengan kerangka kecil.

Jika diberikan sekumpulan objek () maka algoritma K-Means Cluster Analysis akan mempartisi X dalam k buah cluster, setiap cluster memiliki centroid dari objek – objek dalam cluster tersebut. Pada tahap awal algoritma K-Means Cluster Analysis dipilih secara acak k buah objek sebagai centroid, kemudian jarak antara objek dengan centroid dihitung dengan menggunakan jarak euclidian, objek ditempatkan dalam cluster yang terdekat dihitung dari titik tengah cluster. Centroid baru ditetapkan jika semua objek sudah ditempatkan dalam cluster terdekat. Proses penentuan centroid dan penempatan objek dalam cluster diulangi sampai nilai centroid konvergen (centroid dari semua cluster tidak berubah lagi). Secara umum metode K-Means Cluster Analysis menggunakan algoritma sebagai berikut [6]:

- 1) Tentukan k sebagai jumlah cluster yang di bentuk. Untuk menentukan banyaknya cluster k dilakukan dengan beberapa pertimbangan seperti pertimbangan teoritis dan konseptual yang mungkin diusulkan untuk menentukan berapa banyak cluster.
- 2) Bangkitkan k Centroid (titik pusat cluster) awal secara random. Penentuan centroid awal dilakukan secara random/acak dari objek-objek yang tersedia sebanyak k cluster, kemudian untuk menghitung centroid cluster ke-I berikutnya.
- 3) Hitung jarak setiap objek ke masing-masing centroid dari masing -masing cluster. Untuk menghitung jarak antara objek dengan centroid penulis menggunakan Euclidian Distance.
- 4) Alokasikan masing - masing objek ke dalam centroid yang paling terdekat. Untuk melakukan pengalokasian objek kedalam masing - masing cluster pada saat iterasi secara umum dapat dilakukan dengan dua cara yaitu dengan hard k-means, dimana secara tegas setiap objek dinyatakan sebagai anggota cluster dengan mengukur jarak kedekatan sifatnya terhadap titik

pusat cluster tersebut, cara lain dapat dilakukan dengan fuzzy C-Means.

- 5) Lakukan iterasi, kemudian tentukan posisi centroid baru dengan menggunakan persamaan.
- 6) Ulangi langkah 3 jika posisi centroid baru tidak sama. Pengecekan konvergensi dilakukan dengan membandingkan matriks group assignment pada iterasi sebelumnya dengan matrik group assignment pada iterasi yang sedang berjalan. Jika hasilnya sama maka algoritma k-means cluster analysis sudah konvergen, tetapi jika berbeda maka belum konvergen sehingga perlu dilakukan literasi berikutnya.

D) Hasil Perancangan Sistem

1) Menu Utama

Adapun awal dari tampilan aplikasi ini adalah file menu utama, maka pertama kali pada layar akan terlihat tampilan beberapa pilihan menu seperti terlihat pada gambar 4.1 berikut. Dalam tampilan ini terdapat menu data barang, data penjualan, dan proses metode k-means.



Gambar 11. Menu Utama

2) Tampilan Menu Barang

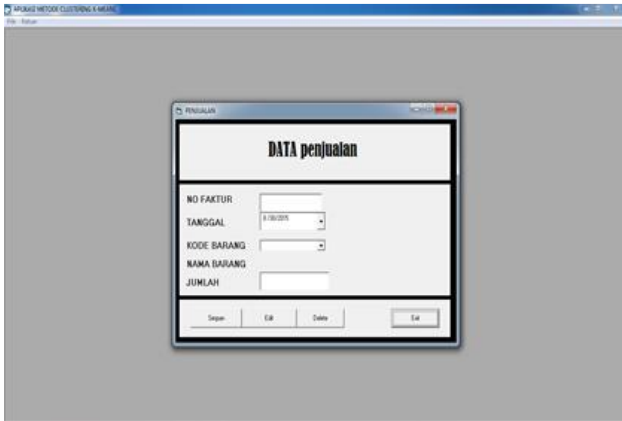
Form menu utama merupakan antarmuka yang memberikan arahan pada pemakai untuk penggunaan fasilitas yang tersedia pada system ini. Tampilan form menu data barang akan terdapat menu entri data seperti kode barang, nama barang, satuan, harga serta stock. Setelah itu pilih tombol simpan agar data disimpan di database .



Gambar 12. Tampilan Menu Barang

3) *Tampilan Menu Data Penjualan*

Menu data penjualan ini adalah tampilan penginputan data untuk penjualan yang terdiri dari no faktur, tanggal, kode barang, nama barang dan jumlah. Dan selain itu juga terdapat tombol simpan untuk menyimpan data, edit untuk merubah data, hapus untuk menghapus data dan exit untuk keluar dari menu data penjualan.



Gambar 13. Menu Data Penjualan

4) *Tampilan Menu Proses K-means*

Menu data kriteria ini menampilkan dari kolom harga (x), jumlah penjualan produk (y), D11 adalah hasil perhitungan Euclidean distance dari semua data ketiap titik pusat pertama, D12 hasil perhitungan Euclidean distance dari semua data ketiap titik ke-2, C1 adalah Pusat Cluster 1: (5. 30), C2 adalah Pusat Cluster 2: (25. 6), tombol proses untuk memproses hasil dari clustering k-means. Rumus euclidian distance : $d = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$

X = Pusat cluster Y = data, Lalu dilakukan penghitungan jarak dari sisa sampel data dengan pusat cluster yang dimisalkan dengan M(a,b), dimana a merupakan total penjualan, dan b harga barang yang diperkecil menjadi angka puluhan juta agar cara penghitungan lebih mudah.

- M1= (14500,5)
- M2= (14.500,4)
- M3= (14.500,2)
- M4= (14.500,3)
- M5= (14.500,1)
- M6= (4.700,40)
- M7=(12.500,10)
- M8= (7.200,4)
- M9= (13.200,19)
- M10= (14.500,10)

Hitung Euclidean distance dari semua data ketiap titik pusat pertama :

$$D11 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 5)^2 + (5 - 30)^2} = 14494.16$$

$$D12 = \sqrt{(M_{1x} - C_{2x})^2 + (M_{1y} - C_{2y})^2} = \sqrt{(14500 - 25)^2 + (5 - 6)^2} = 17.37261$$

$$D13 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 5)^2 + (5 - 30)^2} = 19.13407$$

$$D14 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 5)^2 + (5 - 30)^2} = 20.03117$$

Dengan cara yang sama hitung jarak tiap titik ketitik pusat ke-2 dan kita akan mendapatkan:

$$D21 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 25)^2 + (5 - 30)^2} = 14498.08$$

$$D22 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 25)^2 + (5 - 30)^2} = 12.70692$$

$$D23 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 25)^2 + (5 - 30)^2} = 13.13431$$

$$D24 = \sqrt{(M_{1x} - C_{1x})^2 + (M_{1y} - C_{1y})^2} = \sqrt{(14500 - 25)^2 + (5 - 30)^2} = 12.88363$$

Dari hasil penghitungan Euclidean distance, kita dapat membandingkan :

Tabel 5. Hasil Iterasi 1

	C1	C2
M1	14494.16	14498.08
M2	17.37261	12.70692
M3	19.13407	13.13431
M4	20.03117	12.88363
M5	20.96404	13.45482
M6	11.2443	34.35196
M7	15.12786	11.40157
M8	7.349633	5.570015

- {M9,M10,M11,M12,M13,M14} : anggota C1
- {M1,M2,M3,M4,M5,M6,M7,M8} : anggota C2



Gambar 14. Menu Data Proses K-means

E) *Pengujian Sistem*

Pengujian *blackbox* (*blackbox testing*) adalah salah satu metode pengujian perangkat lunak yang berfokus pada sisi fungsionalitas, khususnya pada input dan output aplikasi (apakah sudah sesuai dengan apa yang diharapkan atau belum). Tahap pengujian atau testing merupakan salah satu tahap yang harus ada dalam sebuah siklus pengembangan perangkat lunak (selain tahap perancangan atau desain).

Tabel 6. Pengujian Black-Box pada Validasi Login Admin Aplikasi Sistem Pakar

No.	Skenario pengujian	Test case	Hasil yang diharapkan	Hasil pengujian	Kesimpulan
1.	Mengosongkan semua isian data barang, lalu langsung mengklik tombol 'Simpan'.	Kode barang: - Nama barang: - Satuan: - Harga : - Stock :	Sistem akan menyimpan data yang diinput	Sesuai harapan	Valid
2.	Hanya mengisi data Kode barang, Nama barang, Satuan, Harga, Stock, lalu langsung mengklik tombol 'Simpan'.	Kode barang: - Nama barang: - Satuan: - Harga : - Stock	Sistem akan menyimpan data yang diinput	Sesuai harapan	Valid
3.	Kode barang, Nama barang, Satuan, Harga, Stock, lalu langsung mengklik tombol 'Masuk'.	Kode barang : kata, Nama barang : kata, Satuan : Nomor, Harga : Nomor, Stock	Sistem akan menyimpan data yang diinput	Sesuai harapan	Valid
5.	Menginputkan sintaks <i>SQL Injection</i> pada data barang, lalu mengklik tombol 'simpan'.	Nama barang : kata, Satuan : Nomor, Harga : Nomor, Stock	Sistem akan menyimpan data yang diinput	Sesuai harapan	Valid

Keterangan: - = kosong (tidak diisi).

V. PENUTUP

A) Kesimpulan

Berdasarkan hasil dan pembahasan, maka kesimpulan dari penelitian ini antara lain :

- 1) Pengklasteran dengan menggunakan metode K-Means didapatkan untuk produk yang laku dengan yang tidak laku.
- 2) Semakin banyak data penjualan barang yang diinput, maka semakin banyak clustering centroid yang terbentuk di proses metode k-means.
- 3) Produk yang laku terdiri dari makanan, minuman dan produk yang tidak laku terdiri dari kebanyakan dari kosmetik

B) Saran

Adapun saran untuk penelitian selanjutnya yaitu :

- 1) Ada beberapa metode klastering yang lainnya yang masih perlu dianalisis untuk dijadikan sebagai referensi untuk mengembangkan metode k-means
- 2) Menggunakan metode atau algoritma klastering yang lainnya untuk pengelompokkan produk

VI. DAFTAR PUSTAKA

Agusta, Y. 2007. K-means-Penerapan, Permasalahan dan Metode Terkait. Jurnal Sistem dan Informatika Vol.3 (Februari 2007): 47-60.

Ems, TIM. 2012. Web Progreming for Beginners. Jakarta: PT Elex Media Komputindo.

Feri Sulianta, Dominikus juju, (2010), Data Mining Meramalkan Bisnis. Perusahaan, Jakarta : Elex Media Komputindo

Hidayat, T., dan Istiadah, N. 2011 . Panduan Lengkap Menguasai SPSS 19 untuk Mengolah Data Statistik Penelitian. Jakarta: Media Kita.

Jurnal by Ahmad Afif. Clustering Data Penjualan dan Persediaan Barang pada PT. Sayap Mas Utama dengan Metode K-Means. STMIK GI MDP

Jurnal by Johan Oscar Ong. Implementasi Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing President University. Program Studi Teknik Industri, President University, Jln. Jln. Ki Hajar Dewantara, Kota Jababeka, Cikarang Baru, Bekasi

Kurniawan, Rulianto.2010. PHP dan MySQL. Palembang. Maxikom

Pahevi, Said Mirza. 2013. Pembangunan Basis Data. Jakarta: PT. Elex Media Komputindo

Santosa, B. 2007. Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis. Yogyakarta: Graha Ilmu

Widodo.2004. Psikologi Belajar. Jakarta: RinekaCipta.