

KLASTERISASI PENJUALAN ALAT-ALAT BANGUNAN MENGUNAKAN METODE K-MEANS (STUDI KASUS DI TOKO ADI BANGUNAN)

M.Hasyim Siregar,S.Kom.,M.Kom

¹Program Studi Teknik Informatika, Fakultas Teknik, Universitas Islam Kuantan Singingi
Jl. Gatot Subroto Km. 7 Jake, Teluk Kuantan
E-mail : hasyimsiregar92@gmail.com

Abstract

In the world of business competition today, we are required to continually develop business to always survive in the competition. To achieve this there are a few things that can be done is to improve the quality of the product, adding the type of product and operational cost reduction company with how to use data analysis of the company. Data mining is a technology that automate the process to find interesting patterns and sensitive from the large data sets. This allows human understanding about finding patterns and scalability techniques. The store Adi Bangunan is a shop which is engaged in the sale of building materials and household who have such a system on supermarket namely buyers took own goods that will be purchased. Sales data, purchase goods or reimbursed some unexpected is not well ordered, so that the data is only function as archive for the store and cannot be used for the development of marketing strategy. In this research, data mining applied using the model of the process of K-Means that provides a standard process for the use of data mining in various areas used in the classification of because the results of this method can be easily understood and interpreted.

Keywords : *Data mining, K-Means, Clustering*

Abstrak

Dalam persaingan dunia bisnis saat ini, kita dituntut untuk senantiasa mengembangkan bisnis agar selalu bertahan dalam persaingan. Untuk mencapai hal tersebut, ada beberapa hal yang bisa dilakukan yaitu dengan meningkatkan kualitas produk, penambahan jenis produk, dan pengurangan biaya operasional perusahaan dengan cara menggunakan analisis data perusahaan. Data mining adalah sebuah teknologi yang mengotomatisasi proses untuk menemukan pola menarik dan sensitif dari kumpulan-kumpulan data yang besar. Ini memungkinkan pemahaman manusia tentang menemukan pola dan skalabilitas teknik. Toko Adi Bangunan merupakan sebuah toko yang bergerak dalam bidang penjualan bahan-bahan bangunan dan peralatan rumah tangga yang memiliki sistem seperti pada swalayan yaitu pembeli mengambil sendiri barang yang akan dibeli. Data-data penjualan, pembelian barang maupun pengeluaran tidak terduga tidak tersusun dengan baik, sehingga data tersebut hanya berfungsi sebagai arsip bagi toko dan tidak dapat dimanfaatkan untuk pengembangan strategi pemasaran. Pada penelitian ini, data mining diterapkan menggunakan model proses K-Means yang menyediakan proses standar penggunaan data mining pada berbagai bidang digunakan dalam klasifikasi karena hasil metode ini mudah dipahami dan diinterpretasikan.

Kata Kunci : *Data mining, K-Means, Clustering, Toko bangunan, penjualan*

1. PENDAHULUAN

Kemajuan teknologi informasi sudah semakin berkembang pesat disegala bidang kehidupan. Banyak sekali data yang dapat dihasilkan oleh teknologi informasi yang canggih, mulai dari bidang industri, ekonomi, ilmu dan teknologi maupun berbagai aspek kehidupan lainnya. Dalam persaingan dunia bisnis saat ini, kita dituntut untuk senantiasa mengembangkan bisnis agar selalu bertahan dalam persaingan. Untuk mencapai hal tersebut, ada beberapa hal yang bisa dilakukan yaitu dengan meningkatkan kualitas produk, penambahan jenis produk, dan pengurangan biaya operasional perusahaan dengan cara menggunakan analisis data perusahaan. Namun hal ini tidak dimanfaatkan oleh salah satu toko yang berada di Teluk Kuantan yaitu Toko Adi Bangunan. Toko Adi Bangunan merupakan sebuah toko yang bergerak dalam bidang penjualan bahan-bahan bangunan dan peralatan rumah tangga yang memiliki sistem seperti pada swalayan yaitu pembeli mengambil sendiri barang yang akan dibeli. Data-data penjualan, pembelian barang maupun pengeluaran tidak terduga tidak tersusun dengan baik, sehingga data tersebut hanya berfungsi sebagai arsip bagi toko dan tidak dapat dimanfaatkan untuk pengembangan strategi pemasaran. Data-data yang telah tersedia dapat dijadikan sebagai sistem pengambilan keputusan untuk solusi bisnis serta dukungan infrastruktur di bidang teknologi yang merupakan penyebab munculnya suatu teknologi data mining. Data mining berguna untuk memberikan solusi kepada para pengambil keputusan dalam bisnis guna meningkatkan bisnis perusahaan. Informasi transaksi yang besar membutuhkan sistem secara otomatis mencari hubungan antara item dalam data di Database [1]. Dapat di ambil contoh dari data kasir di toko adi bangunan. Database penjualan menyimpan jumlah record transaksi penjualan yang besar. Setiap record memberikan daftar item barang yang dibeli oleh pelanggan dalam setiap transaksi penjualan. Pimpinan akan tertarik untuk mengetahui jika ada beberapa kelompok item barang secara konsisten dibeli secara bersama. Pimpinan dapat menggunakan data tersebut dalam pengaturan layout toko untuk meletakkan item barang secara optimal dengan keterkaitan satu dengan lainnya, dapat pula digunakan dalam promosi, atau dalam design katalog dan untuk mengidentifikasi. Salah satu cara yang dapat diterapkan adalah dengan menerapkan penggunaan *Data Mining*, Karena di dalam data mining sendiri terdapat cara dan teknik dalam pemenuhan kebutuhan salah satunya adalah kebutuhan informasi yang luas, dan dari informasi yang kita dapat bila kita gunakan sebagai suatu keputusan atau menentukan sebuah kualitas dalam menentukan suatu keputusan [2]. Klasifikasi adalah proses yang digunakan untuk mengkategorikan sekelompok objek menjadi kelompok tertentu. [3]

Menerapkan algoritma pengelompokan k-means tradisional dan jarak pengukuran Euclidean dari kesamaan dipilih untuk digunakan dalam analisis nilai [4]. Algoritma K-means bekerja dengan baik untuk dokumen tertentu [5]. Pada penelitian ini analisa data mining dilakukan dengan menggunakan metode *K-Means clustering*. Dengan menggunakan metode ini, data-data yang telah didapatkan dapat dikelompokkan kedalam beberapa *cluster* berdasarkan kemiripan dari data-data tersebut, sehingga data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster* dan yang memiliki karakteristik yang berbeda dikelompokkan dalam *cluster* yang lain yang memiliki karakteristik yang sama [6]. Teknik preprocessing data sering diterapkan pada kumpulan data agar lebih bersih, konsisten dan bebas dari kebisingan. Normalisasi digunakan untuk menghilangkan data yang berlebihan

dan memastikan bahwa cluster kualitas baik dihasilkan yang dapat meningkatkan efisiensi algoritma clustering [7]. Dengan adanya pengelompokan-pengelompokan data seperti ini, diharapkan pimpinan dapat melakukan pemasaran dengan strategi yang tepat untuk melayani kebutuhan konsumen.

2. TINJAUAN PUSTAKA

Dalam beberapa tahun terakhir banyak penelitian dan perkembangan tentang *Data Mining*. Nama *data mining* sebenarnya mulai dikenal sejak tahun 1990, ketika pekerjaan pemanfaatan data menjadi sesuatu yang penting dalam berbagai bidang, mulai dari bidang akademik, bisnis, hingga medis (Prasetyo, 2014). Data Mining dapat diterapkan pada berbagai bidang yang mempunyai sejumlah data, tetapi karena wilayah penelitian dengan sejarah yang belum lama, dan belum melewati masa ‘remaja’, maka data mining masih diperdebatkan posisi bidang pengetahuan yang memilikinya. Maka Daryl Pregibon menyatakan bahwa “data mining adalah campuran dari statistik, kecerdasan buatan, dan riset basis data” yang masih berkembang. Menurut (Gokila *et al*), *Data mining* adalah sebuah teknologi yang mengotomatisasi proses untuk menemukan pola yang menarik dan sensitif dari kumpulan-kumpulan data yang besar. Ini memungkinkan pemahaman manusia untuk menemukan pola dan skalabilitas teknik. Teknik data mining digunakan untuk melakukan mining deskriptif (menggambarkan pengelompokan sifat umum) atau mining prediktif (mencoba memprediksi berdasarkan kesimpulan klasifikasi data) pada data volume yang besar. . *Data mining* menurut (Han dan Kamber, 2006). Dapat di definisikan sebagai proses mengekstrak atau menambang pengetahuan yang dibutuhkan dari sejumlah data besar. Keluaran dari data mining bisa dipakai untuk memperbaiki pengambilan keputusan dimasa depan (Siska, 2016). Menurut sumber (Prasetyo, 2014). Data mining atau KDD bertujuan untuk memanfaatkan data dalam basis data dengan mengolahnya sehingga menghasilkan informasi baru yang berguna.

2.1 Tahapan Proses *K-means*

Algoritma *k-means* adalah metode *clustering* yang paling sederhana dan umum. Hal ini dikarenakan *K-means* mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu komputasi yang relatif cepat dan efisien. Dalam penelitian ini dilakukan pengelompokan berdasarkan karakter barang yang ada pada data penjualan toko Adi Bangunan yang digunakan yaitu :

1. Laris (C1)
2. Tidak Laris (C2)

Langkah selanjutnya adalah menentukan jumlah *cluster*, adapun *cluster* yang diinginkan terdiri atas 2 *cluster*. Kemudian diberikan nilai *centroid* yang diambil secara acak dari tabel diatas, dengan menggunakan formula : Dimana j adalah nomor index kriteria atau $\{ j \mid 1 \leq j \leq 3, j \in A \}$ dan $j = 1..3$. Langkah berikutnya adalah menghitung jarak menggunakan rumus *Euclidean* index random centroidnya adalah sebagai berikut yang mana untuk C1 diambil dari data yang mempunyai nilai tertinggi dan c2 diambil data yang mempunyai nilai terendah.

Menghitung jarak setiap data yang ada terhadap nilai centroid, ditampilkan untuk data iterasi pertama dengan rumus berikut:

$$D_{L_i}(x_2, x_1) = \|x_2 - x_1\| \sum_{j=1}^p |x_{2j} - x_{1j}|$$

$$D_{L_2}(x_2, x_1) = \|x_2 - x_1\| = \sqrt{\sum_{j=1}^p (x_{2j} - x_{1j})^2}$$

3. METODE PENELITIAN

Dalam metode penelitian ada beberapa urutan kerja yang harus diikuti, urutan kerja ini merupakan langkah-langkah yang harus dilalui dan dikerjakan sesuai dengan pokok permasalahan agar tidak menyimpang dari batasan masalah yang ditentukan dan penelitian ini berjalan dengan baik. Langkah-langkah yang harus dilakukan mulai dari masalah yang ditemukan sampai mendapatkan solusi dari masalah tersebut, sehingga hasil penelitian ini dapat diterapkan dengan baik. Adapun kerangka kerja yang digunakan di dalam pelaksanaan penelitian ini seperti gambar berikut :



Gambar 3.1 Kerangka Kerja Penelitian

3.1 KDD (*Knowledge Discovery in Database*)

Data mining juga disebut dengan KDD, yaitu kegiatan yang meliputi pengumpulan, pemakaian data untuk menemukan keteraturan, pola dan hubungan data set data yang besar.

3.2 Cleaning Data

Pada penelitian ini jumlah data yang diperoleh sebanyak 10 buah, dengan dilakukannya pembersihan (*cleaning*) tidak ditemukan data yang memiliki nilai noise, sehingga data dapat dipergunakan seluruhnya.

4. HASIL DAN PEMBAHASAN

4.1 Data Awal

Dataset yang digunakan dalam penelitian ini adalah hasil dari *database* pada data penjualan 1 Juli 2017 sampai dengan 9 Agustus 2017 yang berjumlah 10 buah data dengan instrumen berdasarkan stok awal, penjualan dan stok akhir. Data yang digunakan pada penelitian ini dapat dilihat pada tabel 4.1 di bawah ini :

Tabel 4.1 Data Awal

No	Nama Barang	Stok Awal	Total Penjualan	Stok Akhir
1	Cat Semprot Bwin Silver Mett	6	3	3
2	Cat Semprot Bwin Black Dof	10	10	0
3	Cat Air Vitara Pail	3	3	0
4	Cat Minyak Avian 1kg	104	76	28
5	Cat Noodroof Abu-Abu 5kg	4	1	3
6	Trico Politur 1kg 801 Vermis	4	1	3
7	Tinner Botol	12	6	6
8	Tiner Cobra 5l	4	4	0
9	Tiner Labalaba 5l	4	3	1
10	Besi Cor 10 Ck	100	137	13

Sumber (Toko Adi Bangunan, 2017)

4.2 Algoritma K-Means

1. Eksperimen Algoritma K-means

Tabel 4.2 Centroi Awal

Centroi Awal			
C1	104	76	28
C2	3	3	0

Pada proses ini untuk menentukan centroid awal C1 diambil dari data ke-4 dan C2 diambil dari data ke-3 berdasarkan dari jumlah tertinggi dan jumlah terendah.

Menghitung jarak setiap data yang ada terhadap nilai centroid, ditampilkan untuk data iterasi pertama berikut:

$$C_{1,2}=(3;3;0) \quad X_{1}=(3;3;0) = 0,00 \quad D_{11} = \sqrt{(104 - 104)^2 + (76 - 76)^2 + (28 - 28)^2} \quad X_{1}=(104;76;28) \\ C_{1,1}=(104;76;28)$$

$$D_{1,2} = \sqrt{(3 - 3)^2 + (3 - 3)^2 + (0 - 0)^2}$$

$$= 0$$

Begitu seterusnya sampai data 10 hingga diperoleh data sebagai berikut:

Tabel 4.3 Jarak Euclidean

Data ke	Stok Awal	Total Penjualan	Stok Akhir	C1	C2	Nilai Minimum
1	6	3	3	747,20	12,00	12,00
2	10	10	0	898,86	9,90	9,90
3	3	3	0	908,62	0,00	0,00
4	104	76	28	0,00	908,62	0,00
5	4	1	3	750,00	11,24	11,24
6	4	1	3	750,00	11,24	11,24
7	12	6	6	599,60	45,49	45,49
8	4	4	0	907,22	1,41	1,41
9	4	3	1	852,81	2,00	2,00
10	100	137	13	286,13	334,42	286,13

Langkah berikutnya adalah menentukan letak cluster dengan cara membandingkan antara dua cluster, nilai minimum merupakan nilai yang menjadi nilai pilihan. Jika ditemukan nilai yang paling kecil (minimum) maka dapat dimasukkan kedalam cluster tersebut. Untuk lebih jelas lihat table 4.4 dibawah ini:

Tabel 4.4 Penentuan Cluster Baru

Data ke	Stok Awal	Total Penjualan	Stok Akhir	C1	C2
1	6	3	3		1
2	10	10	0		1
3	3	3	0		1
4	104	76	28	1	
5	4	1	3		1
6	4	1	3		1
7	12	6	6		1
8	4	4	0		1
9	4	3	1		1
10	100	137	13	1	

Kemudian menentukan nilai centroid baru, nilai ini ditentukan oleh data yang masuk kedalam cluster, berdasarkan table diatas (data 1-10) diperoleh nilai sebagai berikut :

- Cluster 1 terdapat 2 data
- Cluster 2 terdapat 8 data

Untuk menentukan nilai centroid baru (contoh pada Cluster 1 terdapat 2 data dapat dilakukan dengan cara berikut:

$$ck = \frac{\text{jumlah dari nilai yang masuk kedalam cluster}}{\text{Jumlah data yang masuk}}$$

Untuk centroid pertama (C1) :

$$CK_1 = \frac{104+100}{2} = 102$$

$$CK_2 = \frac{76+137}{2} = 106,5$$

$$CK_3 = \frac{28+13}{2} = 20,5$$

Untuk centroid kedua (C2) terdapat 8 data:

$$CK_1 = \frac{6+10+3+4+4+12+4+4}{8} = 5,875$$

$$CK_2 = \frac{3+10+3+1+1+6+4+3}{8} = 3,875$$

$$CK_3 = \frac{3+0+0+3+3+6+0+1}{8} = 2$$

Secara keseluruhan diperoleh nilai centroid:

Tabel 4.5 Cluster Baru

C1	102	106,5	20,5
C2	5,875	3,875	2

Untuk mencari nilai centroid selanjutnya, ulangi langkah diatas.setelah nilai centroid baru ditemukan maka ulangi langkah perhitungan jarak yaitu pada langkah sebelumnya hingga memasukan data kedalam cluster.

Data akhir adalah jika langkah-langkah diatas diulang dengan langkah-langkah yang sama hingga menghasilkan data pada suatu cluster tepat sama dari data sebelumnya dengan data pada langkah sesudahnya ,atau dengan kata lain data tidak berubah lagi posisinya pada cluster.posisi pada tiap cluster terakhir dapat dilihat pada table berikut:

Table 4.6 Nilai Cluster Interasi 2

Data ke	Stok Awal	Total Penjualan	Stok Akhir	C1	C2	Nilai Minimum
1	6	3	3	447,417	1,884	1,884
2	10	10	0	553,578	11,385	11,385
3	3	3	0	563,474	7,005	7,005
4	104	76	28	86,816	797,781	86,816
5	4	1	3	450,244	4,432	4,432
6	4	1	3	450,244	4,432	4,432
7	12	6	6	345,158	22,483	22,483
8	4	4	0	562,061	5,879	5,879
9	4	3	1	522,785	3,069	3,069
10	100	137	13	86,816	284,039	86,816

Langkah berikutnya adalah menentukan letak cluster dengan cara membandingkan antara dua cluster, nilai minimum merupakan nilai yang menjadi nilai pilihan. Jika ditemukan nilai yang paling kecil (minimum) maka dapat dimasukkan kedalam cluster tersebut. Untuk lebih jelas lihat table 4.7 dibawah ini:

Tabel 4.7 Penentuan Cluster Baru Interasi 2

Data ke	Stok Awal	Total Penjualan	Stok Akhir	C1	C2
1	6	3	3		1
2	10	10	0		1
3	3	3	0		1
4	104	76	28	1	
5	4	1	3		1
6	4	1	3		1
7	12	6	6		1
8	4	4	0		1
9	4	3	1		1
10	100	137	13	1	

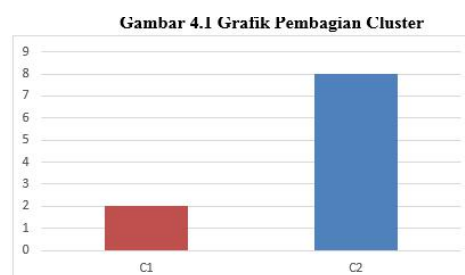
Hasil

Berdasarkan eksperimen perhitungan algoritma k-means hasil yang diperoleh sebagai berikut :

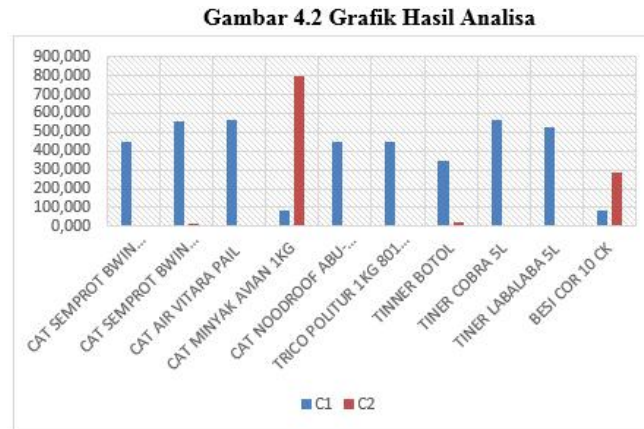
Cluster 1 = 2 Anggota

Cluster 2 = 8 Anggota

Berdasarkan hasil cluster diatas, maka dilakukan analisa dalam pengelompokan barang. Analisa dilakukan dengan menghitung rata-tata dari setiap anggota cluster berdasarkan kategori yang telah ditentukan, berikut hasil analisa dapat digambarkan dalam bentuk grafik sebagai berikut :



Berdasarkan hasil cluster diatas, maka dilakukan analisa dalam pengelompokan barang. Analisa dilakukan dengan menghitung rata-tata dari setiap anggota cluster berdasarkan kategori yang telah ditentukan, berikut hasil analisa dapat digambarkan dalam bentuk grafik sebagai berikut :



Dari dua cluster yang terbentuk jika dilihat dari nilai rata-rata centroid masing-masing cluster maka metode K-Means dapat memberi gambaran bahwa cluster 1 adalah cluster yang penjualan kurang laris pada toko adi bangunan. Sedangkan cluster 2 adalah cluster yang dikategorikan barang yang laris di pasaran pada kelompok ini hanya dua barang yang menunjukkan tingkat penjualan yang tinggi.

Kesimpulan

Dalam penelitian ini ini dapat disimpulkan beberapa hal yang diperlukan untuk menganalisis penjualan yang terjadi di Toko Adi Bangunan pada 1 Juli 2017 sampai dengan 9 Agustus 2017. Data tersebut diolah menggunakan *k-mean clustering* untuk menentukan kelompok barang yang laris dan yang kurang laris dalam penelitian tersebut, peneliti menarik kesimpulan sebagai berikut:

1. Penerapan metode clustering dapat menentukan pembelian stok barang yang diperlukan dengan cepat, dari penelitian yang dilakukan diketahui kelompok barang yang laris sebanyak 10 item sehingga prioritas pembelian stok barang diarahkan pada 10 item tersebut.
2. Penerapan metode k-mean dengan menggunakan bantuan software rapid miner menghasilkan item barang yang laris dan yang kurang laris sehingga pihak manajemen dapat menentukan prioritas pembelian stok barang pada barang barang yang laris.

6.2 Saran

Setelah penulis menyelesaikan tahap akhir dari penelitian ini, penulis menyadari masih banyak kekurangan-kekurangan yang mesti diperbaiki dan dipenuhi, antara lain :

1. Penelitian ini masih jauh dari sempurna dengan hanya menggunakan data penjualan selama 40 hari yaitu pada tanggal 1 Juli 2017 sampai dengan 9 Agustus 2017.
2. Penulis mengharapkan agar hasil dari penelitian ini dapat bermanfaat bagi penulis dan khususnya Toko Adi Bangunan dalam mengambil keputusan penyediaan stok barang.
- 3.

DAFTAR PUSTAKA

- [1] E.T. Luthfi, “Penerapan Data Mining Algoritma Asosiasi Untuk meningkatkan penjualan”, JURNAL DASI, ISSN: 1411-3201, Vol. 10, No. 1, Maret 2009.
- [2] E. Ikhwan, D. Nofriansyah dan Sriani, “Penerapan Data Mining dengan Algoritma Fp Growth untuk Mendukung Strategi Promosi Pendidikan (Studi Kasus Kampus STMIK Triguna Dharma)” , Jurnal Ilmiah Saintikom, ISSN : 1978-6603, *Vol.14, No. 3, September 2015*.
- [3] Oyelade, Oladipupo dan Obagbuwa, “Application of k-Means Clustering algorithm for prediction of Students’ Academic Performance” , ISSN 1947-5500, (*IJCSIS International Journal of Computer Science and Information Security, Vol. 7, _o. 1, 2010*).
- [4] S. Khatri dan K. Garg, “Document Clustering Using Improved K-Means Algorithm”, ISSN 2091-2730, International Journal of Engineering Research and General Science Volume 4, Issue 3, May-June, 2016.
- [5] Jayant Tikmani, Sudhanshu Tiwari, Sujata Khedkar, “An Approach to Customer Classification using k-means” , ISSN : 2320-9801, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 3, Issue 11, November 2015
- [6] J.O. Ong, “Implementasi Algoritma *K-Means Clustering* Untuk Menentukan Strategi Marketing President University” ISSN 1412-6869, Jurnal Ilmiah Teknik Industri, Vol. 12, No. 1, Juni 2013.
- [7] Deepali Virmani ,Shweta Taneja ,Geetika Malhotra, “Normalization based K means Clustering Algorithm” , ISSN: 2349-6495, *International Journal of Advanced Engineering Research and Science (IJAERS), [Vol-2, Issue-2, Feb.- 2015]*.