PENERAPAN METODE K-*MEANS CLUSTERING* PADA PERUSAHAAN

Asrul Sani

Program Studi Teknik Informatika
STMIK Widuri
Jalan Palmerah Barat No.353, Jakarta Selatan – 12210
e-mail: **asrul_sani09@yahoo.com**

Abstrak

Tujuan dari penelitian ini adalah untuk mengetahui seberapa besar hasil dari pengelompokan barang berpengaruh terhadap kebutuhan dari konsumen. Pengelola perusahaan harus memperhatikan aspek dari jumlah jenis barang maupun artikel dari barang tersebut. Semakin lengkap koleksi/artikel dan variasi produk yang ada maka kebutuhan dari konsumen akan menjadi terpenuhi. Koleksi barang yang tersedia akan dibagi menjadi beberapa cluster sehingga dapat diketahui produk/artikel mana yang paling diminati sehingga produk tersebut arus banyak. Metode yang digunakan pada penelitian ini adalah menggunakan teknik Clustering K-Means yang merupakan salah satu metode yang populer dan banyak digunakan. Metode Clustering K-Means akan mencari partisi yang maksimum dari data dengan meminimalkan kriteria jumlah kesalahan kuadrat dengan prosedur iterasi yang optimal. Variabel yang digunakan pada penelitian ini adalah kode produk, jumlah transaksi, harga produk. Hasil akhir penelitian didapatkan pengelompokan jenis artikel dengan jumlah stok terbanyak dan sedikit.

Kata kunci: *Clustering, K-Means, Data Mining.*

1. Pendahuluan

Teknologi *data mining* pada sebuah perusahaan pada dasarnya agar dapat membantu mempercepat proses pengambilan keputusan secara tepat, memungkinkan perusahaan untuk mengelola informasi yang terkandung di dalam data transaksi menjadi sebuah pengetahuan (*knowledge*) yang baru. Melalui pengetahuan yang didapat, perusahaan dapat meningkatkan pendapatannya dan mengurangi biaya, dan pada akhirnya dimasa yang akan datang perusahaan dapat lebih kompetitif [1] [2] [3]. Peningkatan *revenue* perusahaan merupakan dampak yang paling bisa dirasakan. Ketika sebuah strategi penjualan dijalankan, fokus utama perusahaan tidak lagi kepada bagaimana mendapatkan pelanggan baru yang potensial (*prospecting customer*), tetapi bagaimana menjual lebih banyak produk kepada pelanggan yang sudah ada (*existing customer*). Dalam hal ini dengan mengelompokkan jenis produk yang sudah tersedia menjadi beberpa kelompok sesuai dengan kategori dari produk tersebut [2].

Dengan adanya informasi yang relevan, pengelompokan jenis produk dapat digunakan untuk meningkatkan *performance* dari penjualan sehingga dapat menentukan langkah untuk peningkatan stok pada produk secara tepat. Salah satu cara yaitu dengan memanfaatkan teknologi *data mining* untuk pengelompokan produk tersebut yaitu dengan teknik *clustering K-Means* yaitu metode yang sering digunakan untuk proses pengelompokan. *K-Means Clustering* merupakan salah satu teknik data mining yang memberikan deskripsi cluster pada sebuah produk. Sehingga hasil simulasi dapat memberikan pengetahuan jenis produk atau artikel yang akan menjadi stok terbanyak sehingga mampu meningkatkan *revenue* perusahaan [2]; [4].

Persediaan barang atau stok adalah salah satu bagian yang paling aktif dalam suatu operasi perusahaan atau organisasi yang secara terus menerus diproduksi, diubah kemudian dijual kembali. Persediaan didefinisikan sebagai simpanan material yang berupa bahan mentah, barang dalam proses dan barang jadi. Sedangkan pengendalian persediaan merupakan aktivitas mempertahankan jumlah persediaan pada tingkat yang dikehendaki, sehingga persediaan ini berfungsi untuk mempermudah jalannya operasi perusahaan yang dilakukan secara berturut-turut untuk proses bisnis [2]; [5]. Persediaan adalah sebagai suatu aktivitas yang meliputi barang pemilik organisasi dengan maksud untuk dijual dalam suatu waktu atau periode usaha tertentu atau persediaan barang barang yang masih dalam pengerjaan proses produksi ataupun persediaan bahan baku yang menunggu penggunanya dalam proses produksi [3]; [6].

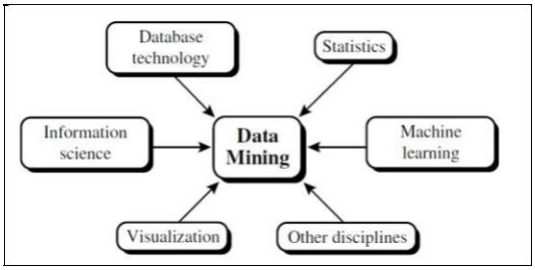
Data adalah sesuatu yang belum mempunyai arti bagi penerimanya dan masih memerlukan adanya suatu pengolahan. Data bisa merujuk pada suatu keadaan, gambar, suara, huruf, angka, matematika, bahasa ataupun simbol-simbol lainnya yang bisa kita gunakan sebagai bahan untuk melihat lingkungan, obyek, kejadian ataupun suatu konsep [7]; [8]; [9]. Informasi merupakan hasil pengolahan dari sebuah model, formasi, organisasi, ataupun suatu perubahan bentuk dari data yang memiliki nilai tertentu, dan bisa digunakan untuk menambah pengetahuan bagi yang menerimanya [4].

Data Mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data Mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengindentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [10]; [11]. Menurut Gartner Group *data mining* adalah suatu proses menemukan hubungan yang berarti, pola dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika [3]. Definisi *data mining* berdasarkan Han Jiawei dan M. Kamber (2006) adalah proses mengekstraksi pola-pola yang menarik (implicit, tidak di ketahui sebelumnya dan berpotensi untuk dimanfaatkan) dari data yang berukuran besar [1].

Data mining merupakan pertemuan beberapa disiplin ilmu seperti *database*, statistik, *machine learning*, visualisasi, dan ilmu informasi. Selain itu pendekatan *data mining* yang dapat diterapkan pada bidang lain seperti *neural network*, *fuzzy*, *inductive logic programming*, dan komputasi kinerja tinggi [1]; [9].

data, observasi atau kasus berdasarkan kemiripan objek yang diteliti. Sebuah *cluster* adalah suatu kumpulan data yang mirip dengan lainnya atau ketidakmiripan data pada kelompok lain [3]. *Clustering* didefinisikan dengan membagi objek data dalam bentuk, entitas, contoh, ketaatan, unit ke dalam beberapa jumlah kelompok (grup, bagian atau kategori) [1].

Proses *clustering* bertujuan untuk meminimalkan terjadinya *objective function* yang diset dalam proses *clustering* yang pada umumnya digunakan untuk meminimalisasikan variasi dalam suatu *cluster* dan memaksimalkan variasi antar *cluster* atau dengan kata lain data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster* yang sama dan data yang memiliki karakteristik berbeda dikelompokkan ke dalam kelompok lain [12].



Gambar 1 *Data Mining* sebagai Pertemuan Beberapa Disiplin Ilmu [1]

2. Metode Penelitian

Metode penelitian menggunakan pendekatan algoritma K-Means. Algoritma *K-Means* merupakan salah satu algoritma dalam fungsi c*lustering* atau pengelompokan.

Gambar 2 Flowchart *K-Menas Clustering*

Proses *clustering* dengan algoritma K-Means adalah sebagai berikut:

1. Tentukan banyaknya *cluster yang diinginkan*
2. Alokasikan data sesuai dengan jumlah *cluster* yang telah ditentukan
3. Tentukan nilai *centroid* pada tiap-tiap *cluster*
4. Hitung jarak terdekat dengan menggunakan rumus Euclidean
5. Tampilkan hasil berdasarkan jarak terendah dari hasil perhitungan step 4
6. Jika belum didapatkan hasil yang sesuai, iterasi kembali dilanjutkan dengan menggunakan step 3. Iterasi akan dihentikan jika hasil clustering sudah sama dengan iterasi sebelumnya [12] [11].

Untuk menentukan nilai centroid tentukan bedasarkan nilai range yang berada pada sumber data yang ada dengan melakukan pemilihan sesuai dengan nilai centroid yang dipilih.

Untuk menentukan jarak digunakan rumus Euclidean sebagai berikut:

$$dist = \sqrt{\sum_{k=1}^{n}(p_k - q_k)^2}$$

Keterangan:

dist	: Jarak Obyek
p_k	: Koordinat dari Obyek p
q_k	: Koordinat dari Obyek q
k	: Urutan dari koordinat

Persediaan barang yang tersedia merupakan kumpulan dari beberapa koleksi / artikel, dimana artikel artikel tersebut bertambah sesuai dengan kebutuhan dari konsumen. Sedangkan artikel artikel yang nilai jualnya sudah berkurang kemungkinan besar tidak akan diproduksi kembali. Pokok permasalahannya adalah dari beberapa koleksi yang laku di pasaran, sangat banyak jenis koleksi koleksi dengan banyak kategori yang jadi produk laku. Dengan menggunakan K-Means Clustering ini diharapkan akan membantu pengelola perusahaan dalam memproduksi ulang koleksi koleksi yang tepat sesuai dengan tujuan perusahaan yaitu meningkat-kan nilai tambah terhadap perusahaan.

Data yang diambil dan diolah berdasarkan nama artikel, penjualan perbulan, dan harga jual. Jenis data yang digunakan adalah data kuantitatif. Data kuantitatif adalah jenis data yang dapat dihitung, berupa angka atau nominal. Data historis transaksi penjualan adalah jenis data kuantitatif karena berupa angka atau nominal dan dapat

dihitung. Lebih spesifik lagi, data yang digunakan berupa data matriks, yaitu jenis data yang memiliki objek dan atribut.

Sumber Data yang digunakan dalam penelitian adalah data primer dan data sekunder. Sumber data primer merupakan sumber data yang diperoleh secara langsung dari sumber asli dan tidak melalui media perantara. Data historis transaksi penjualan yang digunakan diperoleh secara langsung dari objek penelitian melalui wawancara dan dokumentasi. Sedangkan data sekunder merupakan sumber data penelitian yang diperoleh secara tidak langsung melalui media perantara diperoleh dan dicatat oleh pihak lain. Data sekunder pada umumnya berupa bukti catatan atau laporan historis yang dipublikasikan. Data sekunder yang di maksud dalam penelitian ini adalah sumber data yang digunakan untuk menunjang kelengkapan teori data primer.

Populasi penelitian ini adalah data historis transaksi penjualan bulan Januari s/d Desember 2017 dengan data master kode koleksi/artikel yang mengandung kategori Jas dan Celananya saja yang memiliki 37 jenis produk dan berjumlah 468 transaksi. Sampel adalah sebagian jumlah obyek yang diteliti. Menurut Gay & Diehl (1992), "Semakin besar sampelnya maka kecenderungan lebih representatif dan hasilnya lebih digenalisir, maka ukuran sampel dapat
diterima tergantung pada jenis dari penelitiannya".

Berdasarkan pengertian tersebut, maka semakin besar sampel semakin representative dan hasilnya lebih digeneralisir. Besarnya sampel pada penelitian ini berdasarkan pada rumus Slovin yaitu:

$$= \frac{468}{. ^2+1} = \frac{468}{468 . 0.05^2 + 1} = 216$$

Keterangan:
= sampel N =
populasi;
e = nilai presisi 95% atau sig. = 0,05 (*margin of error* yang di tetapkan 5%)

1. Hasil dan Pembahasan

Dari hasil perhitungan sampel yang didapatkan adalah 216 transaksi, maka kemudian dari hasil tersebut dapat isusun tabel atribut dan nilai yang digunakan pada penelitian ini.

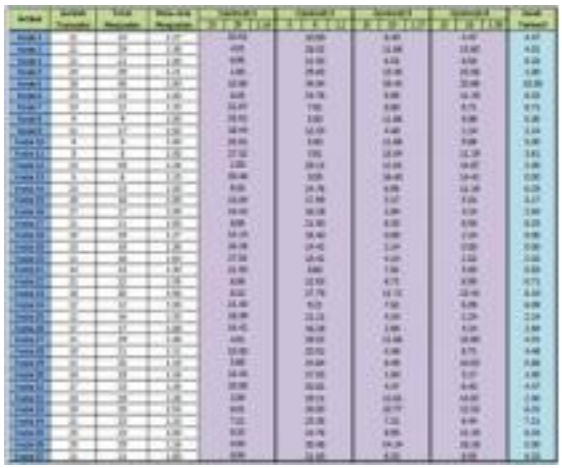
Tabel 1 Atribut dan Nilai

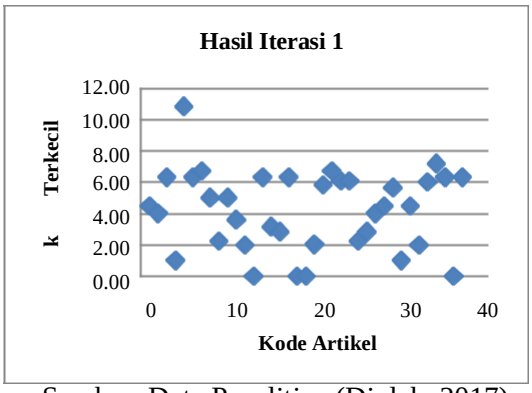
Artikel	Jumlah Transaksi	Total Penjualan	Rata-rata Penjualan
Kode 1	11	14	1.27
Kode 2	21	29	1.38
Kode 3	21	21	1.00
Kode 4	24	29	1.21
Kode 5	19	38	2.00
Kode 6	23	23	1.00
Kode 7	10	12	1.20
Kode 8	9	9	1.00
Kode 9	11	17	1.55
Kode 10	9	9	1.00
Kode 11	8	8	1.00
Kode 12	23	29	1.26
Kode 13	5	6	1.20
Kode 14	23	23	1.00
Kode 15	18	18	1.00
Kode 16	17	17	1.00
Kode 17	21	21	1.00
Kode 18	15	19	1.27
Kode 19	13	18	1.38
Kode 20	11	18	1.64
Kode 21	10	15	1.50
Kode 22	21	22	1.05
Kode 23	19	30	1.58
Kode 24	12	12	1.00
Kode 25	12	16	1.33
Kode 26	17	17	1.00
Kode 27	21	29	1.38
Kode 28	19	21	1.11
Kode 29	21	26	1.19
Kode 30	16	19	1.19
Kode 31	17	23	1.35
Kode 32	23	29	1.26
Kode 33	19	29	1.53
Kode 34	21	23	1.10
Kode 35	23	23	1.00
Kode 36	25	29	1.16
Kode 37	21	21	1.00

Sumber: Data Penelitian (Diolah, 2017)

Clustering data dikelompokkan menjadi 4 kelompok berdasarkan data yang tersedia. Masing masing kelompok ditentukan nilai centroid yang didapatkan dari nilai minimum, nilai maximum dan nilai tengah dari data yang tersedia. Proses cluster dilakukan dengan cara memperhitungkan jarak terdekat dengan menggunakan rumus Euclidean. Dan hasilnya seperti pada tabel di bawah ini.

Tabel 2 Perhitungan Cluster Iterasi 1





Sumber: Data Penelitian (Diolah, 2017)

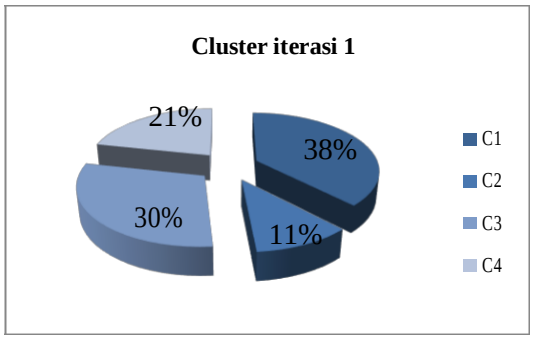
Tabel 3 Hasil Cluster Iterasi 1

Artikel	C1	C2	C3	C4
Kode 1				1
Kode 2	1			
Kode 3			1	
Kode 4	1			
Kode 5	1			
Kode 6	1			
Kode 7				1
Kode 8		1		
Kode 9				1
Kode 10		1		
Kode 11		1		
Kode 12	1			
Kode 13		1		
Kode 14	1			
Kode 15			1	
Kode 16			1	
Kode 17			1	
Kode 18			1	
Kode 19				1
Kode 20				1
Kode 21				1
Kode 22			1	
Kode 23	1			
Kode 24				1
Kode 25				1
Kode 26			1	
Kode 27	1			
Kode 28			1	
Kode 29	1			
Kode 30			1	
Kode 31			1	
Kode 32	1			
Kode 33	1			
Kode 34	1			
Kode 35	1			
Kode 36	1			
Kode 37			1	

Sumber: Data Penelitian (Diolah, 2017)

Proses *K-Means Clustering* pada iterasi 1 sudah didapatkan sesuai dengan tabel 3. Data yang sudah dikelompokkan pada iterasi 1 adalah hasil dari proses clustering dimana hasil iterasi 1 dapat juga dilihat pada gambar di bawah ini:

Gambar 3 Hasil Iterasi 1



Gambar 4 Cluster Iterasi 1

Proses K-Means clustering akan terus melakukan iterasi hingga data hasil clustering akan sama dengan hasil iterasi sebelumnya. Proses ini dilakukan berulang ulang dengan menggunakan metode penentuan centroid hingga hasil clustering pada iterasi tersebut akan menghasilkan clustering yang persis sama dengan hasil clustering pada iterasi sebelumnya.

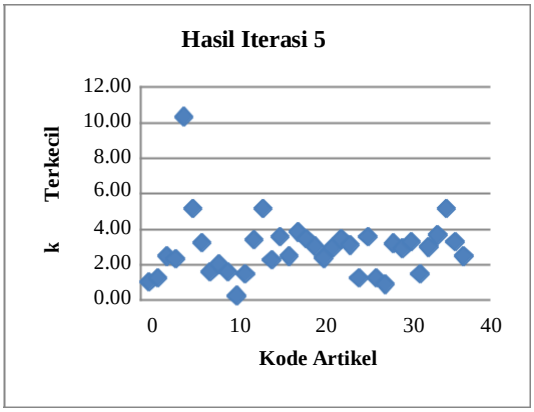
Proses selanjutnya adalah clustering pada iterasi ke 2, yang hasil cluster belum sama dengan iterasi 1. Selanjutnya dilakukan iterasi ke 3 yang hasilnya masih ada perbedaan pada iterasi ke 2 dan begitu seterusnya hingga didapatkan hasil yang sesuai.

Pada iterasi ke 5 akhirnya didapatkan clustering yang sama dan sesuai dengan hasil pada iterasi ke 4. Berikut ditampilkan data hasil clustering dan grafik hasil clustering pada iterasi ke 5 di bawah ini:

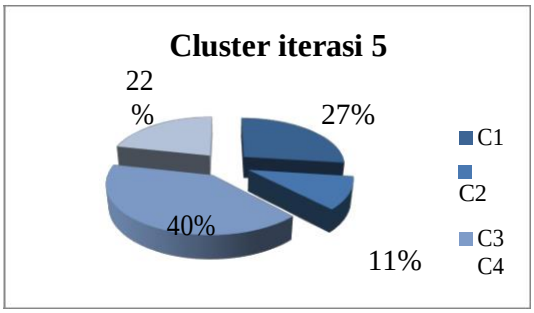
Tabel 4 Hasil Cluster Iterasi 5

Artikel	C1	C2	C3	C4
Kode 1				1
Kode 2	1			
Kode 3			1	
Kode 4	1			
Kode 5	1			
Kode 6			1	
Kode 7				1
Kode 8		1		
Kode 9				1
Kode 10		1		
Kode 11		1		
Kode 12	1			
Kode 13		1		
Kode 14			1	
Kode 15			1	
Kode 16			1	
Kode 17			1	
Kode 18			1	
Kode 19				1
Kode 20				1
Kode 21				1
Kode 22			1	
Kode 23	1			
Kode 24				1
Kode 25				1
Kode 26			1	
Kode 27	1			
Kode 28			1	
Kode 29	1			
Kode 30			1	
Kode 31			1	
Kode 32	1			
Kode 33	1			
Kode 34			1	
Kode 35			1	
Kode 36	1			
Kode 37			1	

Sumber: Data Penelitian (Diolah, 2017)



Gambar 5 Hasil Iterasi 5



Gambar 6 Cluster Iterasi 5

Hasil iterasi ke 5 menunjukkan hasil yang sama dengan iterasi ke 4. Hasil ini merupakan

hasil akhir dari proses clustering. Dari 37 data artikel yang ditampilkan, ada 10 artikel pada cluster 1 (C1) dengan transaksi tertinggi yang terdiri dari kode 2, kode 4, kode 5, kode 12, kode 23, kode 27, kode 29, kode 32, kode 33 dan kode
36. Pada cluster 2 (C2), ada 4 artikel dengan transaksi terendah yang terdiri dari kode 8, kode 10, kode 11, dan kode 13. Padacluster 3 (C3) didapatkan ada 15 artikel dengan transaksi rata rata 1 yaitu kode 3, kode 6, kode kode 15, kode 16, kode 17, kode 18, kode 22, kode 26, kode 28, kode 30, kode 31, kode 34, kode 35 dan kode 37. Untuk cluster 4 (C4) ada 8 artikel dengan transaksi rata rata 2 terdiri dari kode 1, kode 7, kode 9, kode 19, kode 20, kode 21, kode 24, dan kode 25

1. **Simpulan dan Rekomendasi**

Untuk melakukan pengelompokan atas transaksi yang terjadi pada perusahaan dapat dilakukan dengan menerapkan K-Means Clustering. Data data yang diperoleh, diproses dengan menggunakan software Microsoft Excel ataupun dengan menggunakan software lainnya seperti SPSS ataupun Rapidminer. Nilai centroid ditentukan dengan membagi menjadi 4 kelompok dimana C1 merupakan level transaksi tertinggi, C2 merupakan level transaksi terendah, C3 dan C4 merupakan level transaksi rata rata 1 dan rata rata 2. Dari hasil clustering didapatkan untuk nilai C1 berjumlah 10 kode artikel, nilai untuk C2 berjumlah 4 kode artikel, nilai untuk C3 berjumlah 15 artikel dan untuk C4 berjumlah 8 artikel.

Daftar Pustaka:

1 Jie Han and Micheline Kamber, *Data Mining: Concepts and Techniques*, 2nd ed., Asma Stephan, Ed. San Fransisco: Morgan Kaufmann Publishers, June 2006.

2 Budi Santosa, *DATA MINING: Teknik Pemanfaatan Data Untuk Keperluan Bisnis (Teori dan Aplikasi)*. Yogyakarta: PT. Graha Ilmu, 2007.

3 Daniel T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*. New Jersey: John Wiley & Sons, Inc., 2005.

4 Carlo Vercellis, *Multivariate Data Analysis*. New Jersey: A John Wiley and Sons, Ltd., 2009.

Andy Wijaya, Muhammad Arifin, and Tony Soebijono, "Sistem Informasi Perencanaan," *Jurnal Sistem Informasi (JSIKA)*, vol. 2, no. 1, pp. 14-20, May 2013.