

IMPLEMENTASI DATA MINING PEMILIHAN PELANGGAN POTENSIAL MENGGUNAKAN ALGORITMA K-MEANS

Afrisawati (12110955)

Mahasiswa Program Studi Teknik Informatika STMIK Budi Darma Medan
Jl. Sisingamangaraja No. 338 Sp. Limun Medan
www.stmik-budidarma.ac.id //Email: afrisawati.budidarma@gmail.com

ABSTRAK

CV Cahaya Sejati adalah salah satu perusahaan yang bergerak di bidang penjualan kosmetik. Salah satu hal yang penting dalam pengelolaan pelanggan adalah bagaimana suatu perusahaan dapat mempertahankan pelanggan yang dimilikinya. Usaha untuk mempertahankan pelanggan ini menjadi hal yang penting bagi CV Cahaya Sejati mengingat semakin banyaknya usaha retail yang menjual produk yang sama. Untuk mencegah terjadinya perpindahan pelanggan ini, maka perlu diketahui kelompok pelanggan yang potensial, sehingga perusahaan bisa melindungi pelanggan potensial tersebut dengan cara memberikan pelayanan prima dan memberikan hadiah.

Penerapan data mining dapat membantu untuk menganalisa data yang diperoleh dari transaksi. Pengelompokan tersebut dilakukan dengan menggunakan metode clustering. Dalam mengelompokan data tersebut digunakan tools Tanagra untuk mendapatkan data pelanggan potensial. Algoritma yang digunakan adalah algoritma K-Means, informasi yang ditampilkan berupa nilai sentroid dari tiap-tiap cluster dan kelompok-kelompok pelanggan potensial.

Kata Kunci: data mining, metode clustering, Algoritma K-Means, data transaksi pelanggan.

1.Pendahuluan

1.1 Latar Belakang Masalah

Di tengah persaingan bisnis yang sangat ketat dewasa ini, perusahaan-perusahaan retail mulai mengalihkan perhatian mereka dari sekedar mengembangkan produk dan layanan yang unggul ke arah penciptaan pengalaman personal pelanggan. Hal ini dilakukan dengan suatu kesadaran yang utuh bahwa hubungan antara perusahaan dengan pelanggan sangat penting untuk menunjang perkembangan dan kelangsungan perusahaan. Pihak manajemen perusahaan retail harus mampu untuk mengenali pelanggan potensialnya dan mempercayainya dengan meningkatkan pemahaman perusahaan akan kebutuhan mereka sebagai individu sehingga dapat mempertahankan loyalitasnya terhadap perusahaan.

Salah satu hal yang penting dalam pengelolaan pelanggan adalah bagaimana suatu perusahaan dapat mempertahankan pelanggan yang dimilikinya. Usaha untuk mempertahankan pelanggan ini menjadi hal yang penting bagi CV Cahaya Sejati mengingat semakin banyaknya usaha retail yang menjual produk yang sama. Untuk mencegah terjadinya perpindahan pelanggan ini, maka perlu diketahui kelompok pelanggan yang potensial, sehingga perusahaan bisa melindungi pelanggan potensial tersebut dengan cara memberikan pelayanan prima dan memberikan hadiah. Hadiah yang diberikan perusahaan kepada pelanggan berupa cindara mata dan dilakukan secara kontinu di setiap tahunnya. CV Cahaya Sejati adalah salah satu perusahaan retail yang menyadari akan pentingnya hubungan antara pelanggan yang loyal dengan keberhasilan bisnis perusahaan.

1.2 Perumusan Masalah

Berdasarkan latar belakang di atas, dapat dirumuskan beberapa masalah, yaitu:

1. Bagaimana mengelompokkan data yang tersedia dengan algoritma K-Means?
2. Bagaimana mendapatkan data pelanggan yang potensial dengan menggunakan algoritma K-Means?
3. Bagaimana mendapatkan pelanggan potensial dengan mengaplikasikan tools tanagra?

1.3 Batasan Masalah

Pada penyusunan skripsi ini permasalahannya dibatasi pada:

1. Data yang diolah merupakan data sampel konsumen yang bertransaksi dari bulan September 2012-April 2013.
2. Menggunakan algoritma K-Means dalam pengelompokannya.
3. Data yang ditampilkan hanya berupa data-data pelanggan yang potensial.
4. Data yang diolah dalam data mining merupakan data yang disimpan dalam bentuk excel.
5. Pada penulisan ini tidak membahas tentang sistem pendukung keputusan.
6. Menggunakan aplikasi tanagra untuk menerapkan teknik data mining.

1.4 Tujuan dan Manfaat Penelitian

Adapun beberapa tujuan yang diharapkan pada pembuatan skripsi ini adalah:

1. Mengelompokkan data pelanggan yang bertransaksi dengan algoritma *K-Means*.
2. Mendapatkan data pelanggan yang potensial dengan menggunakan algoritma *K-Means*.
3. Mengaplikasikan tanagra untuk mendapatkan pelanggan potensial dengan algoritma *K-Means*. Manfaat diadakannya penelitian ini adalah :

 1. Mempermudah menganalisis data yang besar.
 2. Menggunakan aplikasi tanagra untuk mendapatkan pelanggan yang potensial.
 3. Diharapkan dapat membantu pihak-pihak yang berkepentingan untuk pemilihan pelanggan potensial, sehingga tepat sasaran dalam memberikan hadiah pelanggan potensial dalam rangka mempertahankan pasar ditengah persaingan yang semakin pesat.
 4. Untuk mendapatkan laporan pelanggan potensial secara cepat dan akurat.

2. Landasan Teori

2.1. Pengertian Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. Menurut Turban, dkk dalam bukunya Kusrini *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar (Kusrini, "Algoritma *Data Mining*, 2009:4").

2.2. Konsep pengelompokan *K-Means*

K-Means merupakan salah satu metode pengelompokan data *nonhierarki* (sekat) yang berusaha mempartisi data yang ada ke dalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama dan data yang berkarakteristik berbeda dikelompokkan kedalam kelompok yang lain. Adapun tujuan pengelompokan data ini adalah untuk meminimalkan fungsi objektif yang diatur dalam proses pengelompokan, yang pada umumnya berusaha meminimalkan variasi di dalam suatu kelompok dan memaksimalkan variasi antar kelompok. (sumber: Eko Prasetyo, "Data Mining: Konsep dan Aplikasi menggunakan MATLAB, 2012:178").

$$d_{\text{Euclidean}}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots \dots \dots (1)$$

Fungsi objektif yang digunakan untuk *K-Means* ditentukan berdasarkan jarak dan nilai keanggotaan data dalam kelompok. Fungsi objektif yang digunakan adalah sebagai berikut:

$$J = \sum_{i=1}^N \sum_{l=1}^K aicD(x_i, c_l)^2 \dots \dots \dots (2)$$

2.3 Pelanggan Potensial

Pelanggan potensial adalah pelanggan yang loyal terhadap perusahaan (Zulkarnain, "Ilmu Menjual, 2012:135"). Pelanggan yang potensial umumnya akan melanjutkan pembelian produk atau jasa tersebut walaupun dihadapkan pada banyak alternatif produk atau jasa yang lebih unggul dipandang dari berbagai sudut atributnya. Menurut Kotler dalam bukunya Zulkarnain, konsumen yang loyal tidak diukur dari berapa banyak dia membeli, tapi dari berapa sering dia melakukan pembelian ulang, termasuk merekomendasikan orang untuk membeli (Zulkarnain, "Ilmu Menjual, 2012:135").

3. Analisa Dan Pembahasan

Menggunakan algoritma *K-Means* dalam menganalisis datanya. Data yang akan digunakan secara acak sebanyak 30 data transaksi. Selanjutnya akan digunakan algoritma *K-Means* untuk mengelompokkan data yang ada. Data yang ada akan dikelompokkan menjadi 3 kelompok. Adapun langkah dari pengelompokan data adalah sebagai berikut:

1. $K=3$
2. Tentukan pusat cluster, $K_1=(2,1)$; $K_2=(4,3)$; $K_3=(3,2)$.
3. Hitung jarak setiap data yang ada terhadap setiap pusat cluster. data pelanggan pertama dengan pusat cluster pertama adalah:

$$d_{11} = \sqrt{(4-2)^2 + (2,255-1)^2} = 2.36$$

Jarak data pelanggan pertama dengan pusat cluster kedua:

$$d_{12} = \sqrt{(4-4)^2 + (2,255-3)^2} = 0.745$$

Jarak data pelanggan pertama dengan pusat cluster ketiga:

$$d_{13} = \sqrt{(4-3)^2 + (2,255-2)^2} = 1.03$$

Dari hasil perhitungan data pelanggan ke-1 jarak terdekat dari pusat kluster terdapat pada kelompok 2, sehingga data pelanggan ke-1 merupakan anggota dari kelompok 2. Hasil perhitungan selengkapnya pada tabel 1

Tabel 1 : Hasil perhitungan jarak setiap data pada iterasi 1

Plgn Ke	Nm Plgn	Jumlah transaksi	Total belanja	K1	K2	K3
1	A	4	2.255	2.36	0.75	1.03
2	B	2	1.788	0.79	2.34	1.02
3	C	2	1.255	0.25	2.65	1.25
4	D	4	3.322	3.06	0.32	1.66
5	E	4	3.012	2.83	0.01	1.42
6	F	3	1.261	1.03	2.00	0.74
7	G	3	1.200	1.02	2.06	0.8
8	H	2	1.190	0.19	2.69	1.29

9	I	4	1.503	2.06	1.49	1.11
10	J	2	1.512	0.51	2.49	1.11
11	K	2	1.196	0.19	2.69	1.28
12	L	2	1.309	0.31	2.62	1.22
13	M	2	1.200	0.2	2.69	1.28
14	N	2	1.107	0.11	2.75	1.34
15	O	2	1.477	0.48	2.51	1.14
16	P	2	1.235	0.24	2.67	1.26
17	Q	2	1.420	0.42	2.55	1.16
18	R	2	1.023	0.02	2.81	1.39
19	S	2	2.205	1.21	2.15	1.02
20	T	2	1.185	0.18	2.70	1.29
21	U	2	1.452	0.45	2.53	1.76
22	V	2	2.070	1.07	2.21	2.29
23	W	4	4.239	3.81	1.24	4.36
24	X	3	3.092	2.32	1.00	3.09
25	Y	3	2.662	1.94	1.06	2.66
26	Z	2	1.742	0.74	2.36	2.01
27	A1	2	1.455	0.46	2.53	1.77
28	B1	2	1.124	0.12	2.74	1.5
29	C1	3	1.280	1.04	1.98	1.28
30	D1	2	1.310	0.31	2.62	1.65

Suatu data akan menjadi anggota dari suatu kelompok yang memiliki jarak terkecil dari pusat kelompoknya. Misalkan untuk data pertama, jarak terkecil diperoleh pada kelompok 2, sehingga data pertama akan menjadi anggota dari kelompok 2. Demikian juga untuk data kedua, jarak terkecil ada pada kelompok 1, maka data tersebut akan masuk pada kelompok 1. Posisi *cluster* selengkapnya dapat dilihat pada tabel 2

Tabel 2 : Posisi *cluster* pada iterasi 1

Pln Ke	Nm Pln	Jumlah transaksi	Total belanja	K1	K2	K3
1	A	4	2.255		*	
2	B	2	1.788	*		
3	C	2	1.255	*		
4	D	4	3.322		*	
5	E	4	3.012		*	
6	F	3	1.261			*
7	G	3	1.200			*
8	H	2		*		

9	I	4	1.503	*		
10	J	2	1.512	*		
11	K	2	1.196	*		
12	L	2	1.309	*		
13	M	2	1.200	*		
14	N	2	1.107	*		
15	O	2	1.477	*		
16	P	2	1.235	*		
17	Q	2	1.420	*		
18	R	2	1.023	*		
19	S	2	2.205			*
20	T	2	1.185	*		
21	U	2	1.452	*		
22	V	2	2.070	*		
23	W	4	4.239			*
24	X	3	3.092			*
25	Y	3	2.662			*
26	Z	2	1.742	*		
27	A1	2	1.455	*		
28	B1	2	1.124	*		
29	C1	3	1.280	*		
30	D1	2	1.310	*		

Catatan: tanda (*) menyatakan keanggotaan data terhadap suatu kelompok.

4. Hitung kembali pusat kelompok dengan keanggotaan kelompok yang sekarang. Pusat kelompok adalah rata-rata dari semua data/objek dalam kelompok tertentu. Jika dikehendaki bisa juga memakai median dari kelompok tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai.

Hitung pusat kelompok baru. Untuk kelompok 1, ada 20 data yaitu data ke-2, 3, 8, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 26, 27, 28, 29 dan data ke-30 sehingga:

$$K_{11} = (2+2)/20 = 2.05$$

$$K_{12} = (1.788 + 1.255 + 1.190 + 1.512 + 1.196 + 1.309 + 1.200 + 1.107 + 1.477 + 1.235 + 1.420 + 1.023 + 1.185 + 1.452 + 2.070 + 1.742 + 1.455 + 1.124 + 1.280 + 1.310)/20 = 1.367$$

Untuk kelompok 2, ada 6 data yaitu data ke-1, 4, 5, 23, 24 dan data ke-25 sehingga:

$$K_{21} = (4+4+4+4+3+3)/6 = 3.667$$

$$K_{22} = (2.255+3.322+3.012+4.239+3.092+2.662)/6 = 3.097$$

Untuk kelompok 3, ada 4 data yaitu data ke-6, 7, 9 dan data ke-19 sehingga:

$$K_{31} = (3+3+4+2)/4 = 3$$

$$K_{32} = (1.261+1.200+1.503+2.205)/4 = 1.542$$

Maka terbentuk pusat cluster baru yaitu, K1=(2.05;1.367), K2=(3.667 ;3.097), K3=(3; 1.542)

5. Ulangi langkah 3 hingga posisi data sudah tidak mengalami perubahan.

Tabel 3 : Posisi cluster pada iterasi ke-2

Pign Ke	Nm Pign	Jumlah transaksi	Total belanja	K1	K2	K3
1	A	4	2.255		*	
2	B	2	1.788	*		
3	C	2	1.255			*
4	D	4	3.322		*	
5	E	4	3.012		*	
6	F	3	1.261			*
7	G	3	1.200			*
8	H	2	1.190			*
9	I	4	1.503			*
10	J	2	1.512			*
11	K	2	1.196			*
12	L	2	1.309			*
13	M	2	1.200			*
14	N	2	1.107			*
15	O	2	1.477			*
16	P	2	1.235			*
17	Q	2	1.420			*
18	R	2	1.023	*		
19	S	2	2.205			*
20	T	2	1.185			*
21	U	2	1.452			*
22	V	2	2.070			*
23	W	4	4.239		*	
24	X	3	3.092			*
25	Y	3	2.662			*
26	Z	2	1.742			*
27	A1	2	1.455			*
28	B1	2	1.124			*
29	C1	3	1.280			*
30	D1	2	1.310		*	

Tabel 4 : Posisi cluster pada iterasi ke-3

Pign Ke	Nm Pign	Jumlah transaksi	Total belanja	K1	K2	K3
1	A	4	2.255		*	
2	B	2	1.788	*		
3	C	2	1.255	*		
4	D	4	3.322		*	
5	E	4	3.012		*	
6	F	3	1.261		*	
7	G	3	1.200		*	
8	H	2	1.190	*		
9	I	4	1.503		*	
10	J	2	1.512	*		
11	K	2	1.196	*		
12	L	2	1.309	*		
13	M	2	1.200	*		
14	N	2	1.107		*	
15	O	2	1.477		*	
16	P	2	1.235		*	
17	Q	2	1.420		*	
18	R	2	1.023	*		
19	S	2	2.205		*	
20	T	2	1.185		*	
21	U	2	1.452		*	
22	V	2	2.070		*	
23	W	4	4.239			*
24	X	3	3.092			*
25	Y	3	2.662			*
26	Z	2	1.742		*	
27	A1	2	1.455	*		
28	B1	2	1.124	*		
29	C1	3	1.280			*
30	D1	2	1.310	*		

Tabel 5 : Posisi cluster pada iterasi ke-4

Pln Ke	Nm Pln	Jumlah transaksi	Total belanja	K1	K2	K3
1	A	4	2.255		*	
2	B	2	1.788	*		
3	C	2	1.255	*		
4	D	4	3.322			*
5	E	4	3.012			*
6	F	3	1.261			*
7	G	3	1.200			*
8	H	2	1.190	*		
9	I	4	1.503			*
10	J	2	1.512	*		
11	K	2	1.196	*		
12	L	2	1.309	*		
13	M	2	1.200	*		
14	N	2	1.107	*		
15	O	2	1.477	*		
16	P	2	1.235	*		
17	Q	2	1.420	*		
18	R	2	1.023	*		
19	S	2	2.205	*		
20	T	2	1.185	*		
21	U	2	1.452	*		
22	V	2	2.070	*		
23	W	4	4.239		*	
24	X	3	3.092		*	
25	Y	3	2.662		*	
26	Z	2	1.742	*		
27	A1	2	1.455	*		
28	B1	2	1.124	*		
29	C1	3	1.280			*
30	D1	2	1.310	*		

Tabel 6 : Posisi cluster pada iterasi ke-5

Pln Ke	Nm Pln	Jumlah transaksi	Total belanja	K1	K2	K3
1	A	4	2.255		*	

2	B	2	1.788	*		
3	C	2	1.255	*		
4	D	4	3.322		*	
5	E	4	3.012		*	
6	F	3	1.261			*
7	G	3	1.200			*
8	H	2	1.190	*		
9	I	4	1.503			*
10	J	2	1.512	*		
11	K	2	1.196	*		
12	L	2	1.309	*		
13	M	2	1.200	*		
14	N	2	1.107	*		
15	O	2	1.477	*		
16	P	2	1.235	*		
17	Q	2	1.420	*		
18	R	2	1.023	*		
19	S	2	2.205	*		
20	T	2	1.185	*		
21	U	2	1.452	*		
22	V	2	2.070	*		
23	W	4	4.239		*	
24	X	3	3.092		*	
25	Y	3	2.662		*	
26	Z	2	1.742	*		
27	A1	2	1.455	*		
28	B1	2	1.124	*		
29	C1	3	1.280			*
30	D1	2	1.310	*		

Karena pada iterasi ke-4 dan ke-5 posisi cluster tidak berubah, maka iterasi dihentikan dan hasil akhir yang diperoleh adalah 3 cluster:

1. Cluster pertama memiliki pusat (2; 1.412) Adapun pelanggan yang masuk pada cluster pertama yaitu: (B, C, H, J, K, L, M, N, O, P, Q, R, S, T, U, V, Z, A1, B1, D1).
2. Cluster kedua memiliki pusat (3.667; 3.097). Adapun pelanggan yang masuk pada cluster kedua yaitu: (A, D, E, W, X, Y).

3. *Cluster* ketiga memiliki pusat (3.25;1.311). Adapun pelanggan yang masuk pada *cluster* ketiga yaitu: (F, G, I, C1).

Dari kesimpulan di atas dapat diketahui bahwa pelanggan yang paling potensial berada pada *cluster* kedua yaitu: (A, D, E, W, X, Y)

4. Kesimpulan dan Saran

4.1. Kesimpulan

Dari penelitian yang telah dilakukan, maka dapat ditarik kesimpulan sebagai berikut :

1. Mengelompokkan data dengan algoritma K-Means dilakukan dengan cara menentukan jumlah *cluster*, hitung jarak terdekat dengan pusat *cluster*. Data dengan jarak terdekat menyatakan anggota dari *cluster* tersebut, dilakukan perhitungan kembali sampai data tidak berpindah pada *cluster* lain, untuk meminimalkan fungsi objektif.
2. Data pelanggan yang potensial didapatkan setelah perhitungan algoritma K-Means selesai, data dengan pusat *centroid* terbesarlah yang termasuk ke dalam pelanggan yang paling potensial.
3. Data pelanggan potensial didapatkan dengan cara menggunakan *tools Tanagra*, data yang diproses meliputi data nama pelanggan, data jumlah transaksi dan data total belanja. Data pelanggan potensial, didapatkan setelah menginput data pelanggan yang telah disimpan dalam *excel*, kemuadian data tersebut *diimport* ke dalam *Tanagra*.

4.2 Saran

Berdasarkan hasil penulisan yang telah dilakukan adapun saran dari penulis adalah sebagai berikut:

1. Bagi perusahaan, agar sistem ini dapat diterapkan supaya pencarian data pelanggan yang potensial lebih cepat dan akurat.
2. Hasil dari proses *data mining* ini dapat digunakan sebagai pertimbangan dalam mengambil keputusan lebih lanjut tentang penetapan pelanggan potensial.

DAFTAR PUSTAKA

- [1] Dewanti Retno, 2008. *Kewirausahaan*, Mitra Wacana Media. Jakarta
- [2] Kursini & Emha taufiq luthfi, 2009. *Algoritma Data Mining*, CV Andi Offset.Yogyakarta.
- [3] Prasetyo Eko, 2012. *Data Mining: Konsep dan Aplikasi menggunakan MATLAB*, CV Andi Offset Yogyakarta.
- [4] Santosa Budi, 2007. *Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Graha Ilmu. Yogyakarta.
- [5] Sulianta Feri & Dominikus Juju, 2010. *Data Mining: Meramalkan Bisnis Perusahaan*, PT Elex Media Komputindo. Jakarta.
- [6] Susanto Sani & Dedy Suryadi, 2010. *Pengantar Data Mining : Menggali Pengetahuan dari Bongkahan Data*, CV Andi Offset.Yogyakarta.
- [6] Zulkarnain, 2012. *Ilmu Menjual*, Graha Ilmu. Yogyakarta.