

PENGELOMPOKAN NASABAH BANK MENGGUNAKAN ALGORITMA K-MEANS UNTUK MEMBERIKAN PENAWARAN YANG TEPAT

Ismail Setiawan¹⁾

¹⁾Magister Teknik Informatika STMIK AMIKOM Yogyakarta
Jl Ring road Utara, Condongcatur, Sleman, Yogyakarta 55281
Email : ismail.s@students.amkom.ac.id

Abstrak

Penentuan nasabah bank yang layak untuk diberikan iklan penawaran menjadi salah satu pertimbangan dalam menghemat anggaran promosi. Bila satu biaya promosi menghabiskan Rp 1000 dan bank memiliki 100.000.000 nasabah sementara hanya 10% yang tertarik untuk membeli, maka bank bisa dikatan melakukan pemborosan biaya penawaran sebesar Rp 9.000.000.000.

Algoritma K-means dapat mengelompokkan data menjadi kelas-kelas tertentu sesuai dengan keinginan pengguna. Algoritma K-means melakukan pengklusteran dengan melakukan beberapa iterasi. Iterasi dianggap cukup atau berhenti saat nilai perubahan fungsi objektif berada dibawah nilai ambang batas.

Dari hasil pengklusteran data sebanyak 30 nasabah ke dalam 3 kelas didapat kelas 1 di isi oleh 6 nasabah. Nasabah ini lah yang nantinya akan dilirik oleh pihak bank untuk diberikan iklan penawaran.

Kata Kunci—penawaran, k-means, kelas

1. Pendahuluan

1.1 Latar Belakang Masalah

Salah satu operasi penting dalam data mining adalah melakukan klusterisasi terhadap data set berukuran besar[1]. Analisis kelompok (cluster analysis) adalah pekerjaan mengelompokkan data (objek) yang didasarkan hanya pada informasi yang ditemukan dalam data yang menggambarkan objek tersebut dan hubungan diantaranya[2]. Tujuannya adalah agar objek yang bergabung dalam sebuah kelompok merupakan objek-objek yang mirip (atau berhubungan) satu sama lain dan berbeda (atau tidak berhubungan) dengan objek dalam kelompok yang lain [3]. Aplikasi yang menggunakan data mining bermaksud menyelesaikan masalah dengan membangun model berdasarkan data yang sudah digali untuk diterapkan terhadap data yang lain[4].

Penelitian dibidang klusterisasi yang banyak dikembangkan saat ini berfokus pada tipe data numerik [1]. Ada banyak tipe data yang dapat diaplikasikan dengan data mining. Diantara tipe data tersebut adalah 1) Nominal, 2) Ordinal, 3) Interval dan 4) Rasio [3]. Contoh data nominal adalah kodepos dan nomor induk mahasiswa. Sedangkan tipe ordinal contoh datanya

adalah suhu (dingin, sedang dan panas). Untuk tipe data interval contoh datanya adalah tanggal 1,2,3 dan seterusnya. Sedangkan contoh Tipe data rasio adalah panjang, umur dan tinggi. Beberapa algoritma klasifikasi yang telah dikembangkan dan masuk dalam top ten algoritma data mining adalah k-means [5].

Salah satu kegiatan bank selain melakukan simpan pinjam adalah penawaran terhadap produk-produk tertentu. Produk tersebut bisa merupakan layanan tambahan dari bank itu sendiri atau prosedur kerjasama dengan perusahaan lain. Salah satu layanan tambahan dari bank adalah pemberian kartu kredit yang dapat dipakai diberbagai negara. Sedangkan produk kerjasama dengan perusahaan tertentu adalah pemberian diskon bila membeli barang dalam jumlah tertentu dan ditoko tertentu. Bank skala nasional tentu memiliki jumlah nasabah yang tidak sedikit. Bila biaya per penawaran Rp 1000 sementara jumlah nasabah ada 100.000.000 maka total biaya untuk satu periode penawaran adalah Rp 100.000.000.000 (seratus milyar rupiah). Jika dari satu periode penawaran tersebut hanya 10% nasabah yang benar-benar tertarik untuk membeli produk yang ditawarkan maka bank telah mengeluarkan dana sebesar Rp 9.000.000.000 (sembilan milyar rupiah) untuk melakukan penawaran yang tidak direspons nasabah. Artinya biaya tersebut terbuang dengan percuma.

1.2 Rumusan Masalah

Bagaimana prediksi untuk memilih nasabah yang berpotensi untuk membeli penawaran yang diberikan bila indikator yang digunakan adalah jumlah rumah dan mobil yang dimiliki.

1.3 Tujuan

Tujuan dari penelitian ini adalah untuk mengetahui berapa jumlah nasabah kelompok 1 yang akan diberikan penawaran berdasarkan hasil pengklusteran yang dilakukan menggunakan algoritma k-means.

1.4 Tinjauan Pustaka

1.4.1 Data

Ada beberapa macam untuk merepresentasikan data[3]. Misalnya atribut yang digunakan untuk menggambarkan jenis objek (bisa merupakan kuantitatif atau kualitatif) set data dapat mempunyai karakteristik yang berbed, misalnya set

data yang menggunakan nilai deret waktu (time series) atau sebuah nilai angka, bahkan berupa objek dengan hubungan khusus di dalamnya.

1.4.2 Data mining

Data mining merupakan cabang ilmu yang termasuk masih baru tetapi telah menghasilkan keuntungan yang cukup besar saat ini [4]. Salah satu teknik yang dibuat dalam data mining adalah bagaimana menelusuri data yang ada untuk membangun sebuah model, kemudian menggunakan model tersebut agar dapat mengenali pola data yang lain yang tidak berada dalam basis data yang tersimpan [3]. Secara skematis, gorunescu (2011) [4] membagi langkah pelaksanaan data mining dalam tiga aktivitas yaitu :

- a. Eksplorasi data, terdiri dari aktivitas pembersihan data, transformasi data, pengurangan dimensi, pemilihan cirri dan lain-lain.
- b. Membuat model dan pengujian validitas model, merupakan pemilihan terhadap model-model yang sudah dikembangkan yang cocok dengan kasus yang dihadapi. Dengan kata lain, dilakukan pemilihan model secara kompetitif.
- c. Penerapan model dengan data baru untuk menghasilkan perkiraan dari kasus yang ada. Tahap ini merupakan tahap yang menentukan apakah model yang telah dibangun dapat menjawab permasalahan yang dihadapi.

1.4.3 K-means

K-means merupakan salah satu metode clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster. Metode ini mempartisi data ke dalam cluster sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda di kelompokkan ke dalam cluster yang lain. Dalam statistik dan mesin pembelajaran, pengelompokan k-means merupakan metode analisis kelompok yang mengarah pada pemartisian N objek pengamatan kedalam K kelompok (cluster) di mana setiap objek pengamatan dimiliki oleh sebuah kelompok dengan mean (rata-rata) terdekat [3].

1.4.4 Nasabah

Nasabah adalah orang yang biasa berhubungan dengan atau menjadi pelanggan bank dalam hal keuangan[6]. Perusahaan yang bergerak dibidang keuangan salah satu contohnya adalah bank.

1.5 Metode Penelitian

1.5.1 Pengumpulan data

Data yang digunakan pada penelitian ini adalah data nasabah pada PT BANK ABC. Data yang dicatat adalah jumlah kepemilikan rumah dan mobil. Data tersebut adalah data yang sengaja diciptakan untuk kepentingan penelitian.

1.5.2 Proses data mining

Dalam proses mining data digunakan algoritma k-means dengan aturan-aturan yang ditetapkan [3]. Aturan tersebut adalah proses iterasi akan dihentikan jika nilai perubahan fungsi objektif tidak lebih besar dari nilai ambang batas yang ditentukan. Pada proses ini ditentukan nilai sentroid awal karena k-means merupakan algoritma unsupervised learning artinya sentroid di tentukan sendiri oleh pengguna.

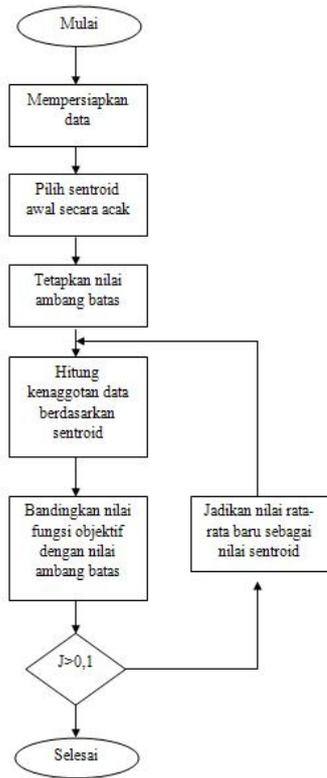
1.5.3 Proses DSS

Proses selanjutnya data baru yang didapat dari proses mining akan digunakan atau diolah untuk menghasilkan sebuah informasi pada proses pembuatan DSS. Proses DSS dihitung dengan iterasi. Iterasi yang dilakukan tergantung dari nilai yang dihasilkan apakah lebih besar atau lebih kecil dari yang telah ditetapkan. Bila lebih besar iterasi di hentikan. Bila lebih kecil iterasi terus dilakukan dengan batasan tertentu yang ditentukan oleh pengguna.

1.5.4 Informasi

Hasil akhir dari proses adalah informasi berupa nasabah yang terkelompokkan kedalam kelas yang telah ditentukan untuk kemudian dijadikan tujuan oleh pihak bank dalam pemberian penawaran. Informasi dapat berupa diagram batang atau diagram pie. Tergantung dari pihak yang membutuhkan.

Secara keseluruhan metode yang penulis lakukan dapat dilihat pada gambar 1.



Gambar 1. Alur Penelitian

2. Hasil Dan Pembahasan

Hasil dari penelitian ini menunjukkan bahwa pengklasteran nasabah dalam 3 kelas dilakukan dengan 1 kali iterasi. Anggota pada masing – masing kelas adalah 1) kelompok 1 diisi sebanyak 6 nasabah, 2) kelompok 2 diisi sebanyak 11 nasabah dan 3) kelompok 3 diisi sebanyak 11 nasabah. Ditentukan kelompok 1 (satu) adalah nasabah yang diprioritaskan untuk diberikan penawaran.

Berikut ini adalah data nasabah dengan jumlah rumah dan mobil yang dimilikinya. Nilai ambang batas yang ditentukan adalah 0,1.

Tabel 1. Nasabah PT BANK ABC

Nasabah	Jumlah mobil	Jumlah rumah
1	1	3
2	3	3
3	4	3
4	5	3
5	1	2
6	4	2
7	1	1
8	2	1
9	3	2
10	2	3

11	3	4
12	4	1
13	2	2
14	1	3
15	2	2
16	5	3
17	6	2
18	3	3
19	4	4
20	2	2
21	1	1
22	2	2
23	3	3
24	4	1
25	2	2
26	2	3
27	1	2
28	2	1
29	3	2
30	4	3

Untuk penentuan sentroid awal dilakukan secara acak oleh pengguna. Pada penelitian ini penulis memilih nilai acak sentroid pada masing- masing kelas sebagai berikut :

Tabel 1. Sentroid awal

Kelas	Sentroid 1	Sentroid 2
1	2	1
2	2	3
3	1	4

Masukan nilai sentroid tersebut dan hitung jarak antar data untuk mendapatkan posisi masing-masing data pada kelas berapa data tersebut berada. Bandingkan setiap data dengan sentroid yang dihitung dan ambil nilai terkecil, pada sentroid berapa nilai terkecil tersebut berada dan jadikan data tersebut menjadi anggota sentroid yang terdekat. Rumus untuk menghitung jarak antar data dengan sentroid digunakan *eucludian distance* dengan formula sebagai berikut

$$D(x_2, x_1) = |x_2 - x_1|_2 = \sqrt{\sum_{j=1}^p |x_2 - x_1|^2} \dots \dots \dots (1)$$

Tabel 2. Hasil penghitungan awal

Nasabah	Jumlah Rumah	Jumlah Mobil	sentroid 1		sentroid 2		sentroid 3		sentroid
			2	1	2	3	1	4	
1	1	3	2,24	1,00	1,00				3
2	3	3	2,24	1,00	2,24				2
3	4	3	2,83	2,00	3,16				2
4	5	3	3,61	3,00	4,12				2
5	1	2	1,41	1,41	2,00				3
6	4	2	2,24	2,24	3,61				3
7	1	1	1,00	2,24	3,00				1
8	2	1	0,00	2,00	3,16				1
9	3	2	1,41	1,41	2,83				3
10	2	3	2,00	0,00	1,41				2
11	3	4	3,16	1,41	2,00				2
12	4	1	2,00	2,83	4,24				1
13	2	2	1,00	1,00	2,24				3
14	1	3	2,24	1,00	1,00				3
15	2	2	1,00	1,00	2,24				3
16	5	3	3,61	3,00	4,12				2
17	6	2	4,12	4,12	5,39				3
18	3	3	2,24	1,00	2,24				2
19	4	4	3,61	2,24	3,00				2
20	2	2	1,00	1,00	2,24				3
21	1	1	1,00	2,24	3,00				1
22	2	2	1,00	1,00	2,24				3
23	3	3	2,24	1,00	2,24				2
24	4	1	2,00	2,83	4,24				1
25	2	2	1,00	1,00	2,24				3
26	2	3	2,00	0,00	1,41				2
27	1	2	1,41	1,41	2,00				3
28	2	1	0,00	2,00	3,16				1
29	3	2	1,41	1,41	2,83				3
30	4	3	2,83	2,00	3,16				2

Jarak nasabah 1 dengan sentroid masing-masing adalah.

$$D_{i1} = \sqrt{(1 - 2)^2 + (3 - 1)^2}$$

$$= \sqrt{1 + 2} = 2,24$$

$$D_{i2} = \sqrt{(1 - 2)^2 + (3 - 3)^2}$$

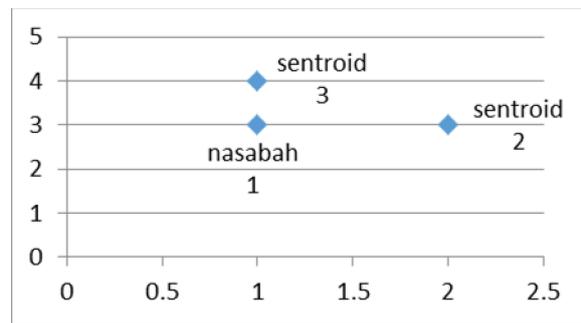
$$= \sqrt{1 + 0} = 1$$

$$D_{i3} = \sqrt{(1 - 1)^2 + (3 - 4)^2}$$

$$= \sqrt{0 + 1} = 1$$

Berdasarkan hasil perhitungan tersebut nasabah 1 memiliki nilai yang sama antara sentroid 2 dan 3.

Sistem mengelompokkan kedalam sentroid 3. Hal ini dibuktikan oleh gambar 2.



Gambar 2. Jarak nasabah 1 dengan sentroid 2 dan 3

Kemudian hitung nilai fungsi objektif berdasarkan jarak dan nilai keanggotan data dalam kelompok. Berikut rumus untuk menentukan fungsi objektif

$$J = \sum_{i=1}^N \sum_{c=1}^K \alpha_{ic} D(x_i, C_c)^2 \dots \dots \dots (2)$$

Nilai Jawal adalah 0, kemudian berdasarkan rumus diatas dengan menggunakan data pada tabel 2. didapat nilai J baru adalah 54,47. Selesih antara Jawal dan Jbaru yaitu 54,47 dan masih diatas dari nilai ambang batas yang ditentukan. Maka lakukan kembali iterasi dengan nilai sentroid baru berdasarkan rata-rata dari jumlah kelompok. Sentroid baru berdasarkan tabel 2 adalah sebagai berikut

Tabel 3. Sentroid untuk kelas 1

Nasabah	Jumlah Rumah	Jumlah Mobil	Sentroid
7	1	1	1
8	2	1	1
12	4	1	1
21	1	1	1
24	4	1	1
28	2	1	1
M	6		
Total	14	6	
Rata rata	2,33	1	

M = jumlah data

Rata -rata jumlah rumah adalah 2,33. Sedangkan rata rata jumlah mobil adalah 1.

Tabel 4. Sentroid untuk kelas 2

Nasabah	Jumlah rumah	Jumlah mobil	Sentroid
2	3	3	2
3	4	3	2
4	5	3	2
10	2	3	2

11	3	4	2
16	5	3	2
18	3	3	2
19	4	4	2
23	3	3	2
26	2	3	2
30	4	3	2
M	11		
Total	49	35	
Rata rata	3,45	3,18	

Rata rata jumlah rumah adalah 3,45. Sedangkan rata-rata jumlah mobil adalah 3,18

Tabel 5. Sentroid untuk kelas 3

Nasabah	Jumlah Rumah	Jumlah Mobil	Sentroid
1	1	3	3
5	1	2	3
6	4	2	3
9	3	2	3
13	2	2	3
14	1	3	3
15	2	2	3
17	6	2	3
20	2	2	3
22	2	2	3
25	2	2	3
27	1	2	3
29	3	2	3
M	13		
Total	30	28	
Rata rata	2,30	2,15	

Rata untuk jumlah rumah adalah 2,30. Sedangkan rata-rata jumlah mobil adalah 2,15.

Sehingga didapatkan sentroid baru untuk masing-masing kelas seperti dijelaskan pada tabel 6.

Tabel 6. Sentroid baru

Kelas	Sentroid 1	Sentroid 2
1	2,33	1
2	3,45	3,18
3	2,30	2,15

Masukan kembali nilai sentroid pada tabel 6 pada data. Berikut hasil perhitungan iterasi pertama.

Tabel 7. Iterasi 1

Nasabah	Jumlah Rumah	Jumlah Mobil	sentroid 1		sentroid 2		sentroid 3		sentroid
			2	1	2	3	1	4	
1	1	3	2,40		2,46		1,56		3
2	3	3	2,11		0,49		1,09		2
3	4	3	2,60		0,57		1,89		2
4	5	3	3,33		1,56		2,82		2
5	1	2	1,67		2,72		1,32		3
6	4	2	1,94		1,30		1,70		2
7	1	1	1,33		3,28		1,74		1
8	2	1	0,33		2,62		1,19		1
9	3	2	1,20		1,27		0,71		3
10	2	3	2,03		1,47		0,90		3
11	3	4	3,07		0,94		1,97		2
12	4	1	1,67		2,25		2,05		1
13	2	2	1,05		1,87		0,34		3
14	1	3	2,40		2,46		1,56		3
15	2	2	1,05		1,87		0,34		3
16	5	3	3,33		1,56		2,82		2
17	6	2	3,80		2,81		3,70		2
18	3	3	2,11		0,49		1,09		2
19	4	4	3,43		0,98		2,50		2
20	2	2	1,05		1,87		0,34		3
21	1	1	1,33		3,28		1,74		1
22	2	2	1,05		1,87		0,34		3
23	3	3	2,11		0,49		1,09		2
24	4	1	1,67		2,25		2,05		1
25	2	2	1,05		1,87		0,34		3
26	2	3	2,03		1,47		0,90		3
27	1	2	1,67		2,72		1,32		3
28	2	1	0,33		2,62		1,19		1
29	3	2	1,20		1,27		0,71		3
30	4	3	2,60		0,57		1,89		2

Didapat nilai Jbaru adalah 29,11 sedangkan Jlama adalah 54,47. Berdasarkan hasil perhitungan Jbaru – Jawal sama dengan -25,36 (absolute 25,36), nilai tersebut lebih kecil dari nilai ambang batas yang ditentukan yaitu 0,1 maka iterasi dihentikan.

3. Kesimpulan dan saran
- 3.1 Kesimpulan
 - a. Rata-rata jumlah rumah dan mobil pada kelompok 1 berturut-turut adalah 2.33 dan 1.
 - b. Pada kelompok 2 Rata – rata jumlah rumah dan mobil adalah 3.45 dan 3.48.
 - c. Kelompok 3 memiliki rata-rata jumlah rumah dan mobil 2.30 dan 2.15.
 - d. Bila penawaran berupa pembelian barang mewah seperti mobil maka kelompok 2 adalah sasaran yang tepat.
 - e. Penggunaan Algoritma k-means pada kasus pengelompokan nasabah berjalan dengan cepat dibuktikan dengan hanya melakukan 1 iterasi.
- 3.2 Saran
 - a. Untuk peneliti selanjutnya diharapkan menggunakan data yang real dari bank
 - b. Jumlah data yang akan diteliti sebaiknya lebih dari 100 buah.
 - c. Peneliti selanjutnya bisa membandingkan hasil dari data yang dirubah ke dimensi lebih kecil dari dimensi yang dibahas pada penelitian ini.

Daftar Pustaka

- [1]Karlita, T., Helen, A., Arifin ,F., 2006, *Pengembangan Algoritma K-Modes pada Penentuan Titik Pusat Awal untuk Mengelompokkan Penyakit pada Kacang Kedelai, IES 2006 – Politeknik Elektronika Negeri Surabaya - ITS*
- [2]Tan, P. et al., 2006, *Intoduction to data mining*. Boston, Pearson Education.
- [3]Prasetyo, E., 2012, *Data Mining Konsep dan aplikasi menggunakan matlab*, Andi, Yogyakarta
- [4]Widodo, P.P., Handayanto, R.T., Herlawati, 2013, *penerapan data mining dengan matlab, Rekayasa sains*, Bandung
- [5]Wu, X., Kumar, V., 2009, *The Top Ten Algorithms in Data Mining*, CRC Press, London
- [6] <http://kbbi.web.id/nasabah> (diakses pada tanggal 27 november 2014)

Biodata Penulis

Ismail Setiawan, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informasi STMIK DUTA BANGSA Surakarta, lulus tahun 2010..Saat ini menjadi Dosen di AMIK Harapan Bangsa Surakarta dan UNS Surakarta.