

CLUSTERING PEMILIHAN PELANGGAN POTENSIAL MENGUNAKAN ALGORITMA *K-MEANS* DI UD.MAKMUR

¹Dermawan Adi Wijaya

*Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember
Email : dermawanady@gmail.com*

Abstrak - UD.Makmur merupakan industri manufaktur yang bergerak pada pengiriman serta penjualan kayu. 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 UD.Makmur 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. Implementasi dari clustering pelanggan potensial dapat membantu untuk menganalisa data yang diperoleh dari transaksi. Pengelompokan tersebut dilakukan dengan menggunakan metode clustering. Algoritma yang digunakan adalah algoritma *K-Means*, informasi yang ditampilkan berupa nilai sentroid dari tiap-tiap cluster dan kelompok-kelompok pelanggan potensial.

Kata Kunci: *Metode Clustering, Algoritma K-Means, Pelanggan Potensial.*

1. Pendahuluan

1.1 Latar Belakang Masalah

Ditengah persaingan bisnis yang sangat ketat dewasa ini, perusahaan-perusahaan manufaktur 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 manufaktur harus mampu untuk mengenali pelanggan potensialnya dan mempercayainya dengan meningkatkan pemahaman perusahaan akan kebutuhan mereka sebagai individu sehingga dapat mempertahankan loyalitasnya terhadap perusahaan.

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

1. Bagaimana mengelompokkan data yang tersedia dengan algoritma *K-Means*?
2. Bagaimana mengelompokkan pelanggan yang potensial dengan menggunakan algoritma *K-Means*?

1.3 Tujuan 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*.

1.4 Batasan Masalah

Pada penyusunan skripsi ini permasalahannya dibatasi pada:

1. Data yang di gunakan merupakan data transaksi pada bulan September 2014 - Maret 2015.

1.5 Manfaat Penelitian

Mempermudah menganalisis data yang besar.

2. Tinjauan Pustaka

Bab ini membahas teori penunjang yang berhubungan dengan penerapan metode *K-Means clustering* pada sistem pengelompokan pelanggan potensial

2.1 UD.Makmur.

Di Kecamatan Sumber Jambe Kabupaten Jember, banyak berdiri industri pengolahan kayu rakyat yang bahan bakunya berasal dari desa-desa di kecamatan sekitarnya. Melihat hal tersebut banyak sekali potensi dari hasil perhutani yang dapat di dimanfaatkan. Salah satu industri yang memanfaatkan peluang tersebut adalah UD.Makmur . UD.Makmur merupakan industri manufaktur yang bergerak pada pengiriman serta penjualan kayu. Pada umumnya industri tersebut berbentuk industri kecil yang manajemen perusahaannya sederhana. Pada awalnya perusahaan yang berdiri pada tahun 2015 ini hanya memiliki 5 pegawai tetap yang di pekerjakan. Namun setelah usaha tersebut kini telah sedikit berkembang, setidanya UD.Makmur telah memiliki 20 Pegawai tetap serta 12 pekerja dalam pengiriman barang. Usaha yang di lakukan UD.Makmur sangat

tergantung sekali kepada konsumen atau pelanggan yang memesan barang, sehingga kepuasan pelanggan cukup berpengaruh pada usaha yang sedang di kembangkan. Pada saat ini pelanggan yang berpotensi dalam mengembangkan usaha manufaktur tersebut.

Clustering adalah suatu metode pengelompokan berdasarkan ukuran kedekatan(kemiripan).Clustering berbeda dengan group, kalau group berarti kelompok yang sama kondisinya kalau tidak ya pasti bukan kelompoknya.Tetapi kalau cluster tidak harus sama akan tetapi pengelompokannya berdasarkan pada kedekatan dari suatu karakteristik sample yang ada, salah satunya dengan menggunakan rumus jarak *euclidean*.Aplikasinya cluster ini sangat banyak, karena hampir dalam mengidentifikasi permasalahan atau pengambilan keputusan selalu tidak sama persis akan tetapi cenderung memiliki kemiripan saja. (Edi satriyanto, M,Si)

K-Means merupakan salah satu metode pengelompokan data *nonhierarki* (sekatan) 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. (sumber: Eko Prasetyo,“*Data Mining: Konsep dan Aplikasi menggunakan MATLAB*)

Metode ini mempartisi ke dalam cluster / kelompok sehingga data yang memiliki karakteristik yang sama (High intra class similarity) dikelompokkan ke dalam satu cluster yang sama dan yang memiliki karakteristik yang berbeda (Law inter class similarity) dikelompokkan pada kelompok yang lain [3]. Proses klustering dimulai dengan mengidentifikasi data yang akan dikluster, X_{ij} ($i=1, \dots, n; j=1, \dots, m$) dengan n adalah jumlah data yang akan dikluster dan m adalah jumlah variabel. Pada awal iterasi, pusat setiap cluster ditetapkan secara bebas (sembarang), C_{kj} ($k=1, \dots, k; j=1, \dots, m$). Kemudian dihitung jarak antara setiap data dengan setiap pusat cluster. Untuk melakukan penghitungan jarak data ke- i (x_i) pada pusat cluster ke- k (c_k), diberi nama (d_{ik}), dapat digunakan formula Euclidean [2] seperti pada persamaan (1), yaitu:

$$d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \dots\dots\dots(1)$$

Suatu data akan menjadi anggota dari cluster ke- k apabila jarak data tersebut ke pusat cluster ke- k bernilai paling kecil jika dibandingkan dengan jarak ke pusat cluster lainnya. Hal ini dapat dihitung dengan menggunakan persamaan (2)

$$\text{Min} \sum_{k=1}^k d_{ik} = \sqrt{\sum_{j=1}^m (x_{ij} - c_{kj})^2} \dots\dots\dots(2)$$

Selanjutnya dikelompokkan data yang menjadi anggota-anggota pada setiap cluster.Nilai pusat cluster yang baru dapat dihitung dengan cara mencari nilai rata-

rata dari data-data yang menjadi anggota pada cluster tersebut, dengan menggunakan rumus pada persamaan 3:

$$c_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p}; \dots\dots\dots(3)$$

Dimana $x_{ij} \in$ cluster ke – k $p =$ banyaknya anggota cluster ke k

Algoritma dasar dalam k-means adalah

1. Tentukan jumlah cluster (k), tetapkan pusat cluster sembarang.
2. Hitung jarak setiap data ke pusat cluster menggunakan persamaan (2.1).
3. Kelompokkan data ke dalam cluster yang dengan jarak yang paling pendek menggunakan persamaan (2.2).
4. Hitung pusat cluster yang baru menggunakan persamaan (2.3)

Ulangi langkah 2 sampai dengan 4 hingga sudah tidak ada lagi data yang berpindah ke cluster yang lain.

2.6 Pelanggan Potensial

Pelanggan potensial adalah pelanggan yang loyal terhadap perusahaan . 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. Metode Penelitian

3.1.1 Pengumpulan Data

Pada penelitian ini data untuk tahap pengujian nantinya yang digunakan adalah data dari pelanggan yang melakukan transaksi pada UD.Makmur. Sedangkan untuk data-data awal yang digunakan untuk membuat aturan-aturannya didapatkan dari buku refrensi.

Jenis data (data pengujian dan data awal) yang digunakan pada penelitian ini dari cara memperolehnya ialah data primer dimana data berasal dari sumber asli atau pertama. Data tersebut harus dicari melalui narasumber yang dijadikan objek penelitian atau orang yang dijadikan sebagai sarana mendapatkan informasi maupun data. Data dipaparkan dalam bentuk angka-angka yang pasti yang disebut data kuantitatif. Sifat dari data pengujian tersebut ialah diskrit, dimana nilainya merupakan bilangan asli. Sedangkan data awal dipaparkan dalam suatu interval tertentu (kontinyu) yang nantinya barulah di ambil kesimpulan kembali untuk mengubah data tersebut kedalam bentuk bilangan asli yang dilakukan oleh pakar mekanik tersebut. Menurut waktu pengumpulannya, data tersebut merupakan data *time series*, dari waktu ke waktu atau periode secara *historis*.

Format dari data yang didapatkan ialah data *numeric*. Jadi data-data yang dimiliki ialah dalam bentuk angka. Jumlah data pengujian yang akan digunakan ialah sejumlah data yang didapatkan selama kurang lebih 6 bulan yang merupakan data transaksi yang di lakukan oleh UD.Makmur

3.1.2 Pra Proses

Pada tahap ini, data yang telah di pra proses nantinya pusat cluster dipilih secara acak dari sekumpulan populasi data .Mengelompokkan data dengan algoritma K-Means dilakukan dengan cara menentukan jumlah *cluster*, hitung jarak terdekat dengan pusat *cluster*. Pada tahap pra proses ini data yang sudah terkumpul nantinya akan di pilih beberapa data saja yang sesuai dengan paramater yang di tentukan. Selanjutnya dengan menetapkan parameter yang berupa jumlah transaksi dan total belanja, data yang termasuk dalam jumlah transaksi dan total belanja nantinya akan di proses oleh metode K-Means. Data yang ada akan dikelompokkan menjadi 3 kelompok. Adapun langkah dari pengelompokkan data adalah sebagai berikut

1. $K=3$
2. Tentukan pusat cluster, $K_1=(2,1)$; $K_2=(4,3)$; $K_3=(3,2)$.

Data yang akan di masukan dalam Proses K-Mean

3.1.3 K-Means

Tahap selanjutnya 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 pusat cluster. Posisi pusat cluster ke dalam tiap-tiap pusat cluster dan terakhir akan terbentuk pusat cluster baru. Langkah Setelah Praproses sebagai berikut

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:

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

4. Hitung kembali pusat kelompok dengan keanggotaan kelompok yang sekarang. Pusat kelompok adalah rata-rata dari semua data/obyek 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,

$$K_1 = (2.05; 1.367), K_2 = (3.667; 3.097), K_3 = (3; 1.542)$$

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

4. Implementasi Sistem

Setelah melalui tahap analisis dan perancangan, tahap selanjutnya untuk mengembangkan suatu perangkat lunak adalah tahap implementasi dan pengujian sistem. Untuk mengetahui apakah implementasi perangkat lunak tersebut berhasil atau tidak, diperlukan pengujian. Berikut ini hasil implementasi dan pengujian dari aplikasi yang telah dibangun.

Berdasarkan hasil analisis dan perancangan sistem yang telah dilakukan, maka dilakukan implementasi "*Clustering Pelanggan Potensial Dengan Menggunakan Metode K-Means Di UD.MAKMUR*" ini di buat menggunakan bahasa pemrograman java.

4.1.1 Lingkungan Implementasi

Berikut ini beberapa perangkat keras dan perangkat lunak komputer yang dibutuhkan untuk mengimplementasikan program aplikasi "*Clustering Pelanggan Potensial Dengan Menggunakan Metode K-Means Di UD.MAKMUR*".

1. Hardware dan Software Untuk Pengujian Sistem

a. Hardware

PC dengan Processor Intel Core 2 duo 2,3Ghz, Memori 2 GB, Kartu Grafik 1 GB

b. Software

- i. Sistem Operasi Windows 7 Ultimate 32 bit
- ii. NetBeans 7.2
- iii. Microsoft Excel 2007

2. Hardware dan Software Minimal Untuk Menjalankan Program

a. Hardware

PC dengan Processor IV 1,6Ghz, Memori 512MB, KartuGrafik 512 MB

b. Software

- i. Sistem Operasi Windows XP SP3
- ii. NetBeans 6.9
- iii. Microsoft Excel 2003

4.1.2 Tampilan Implementasi Program

Adapun hasil implementasi form yang telah dirancang pada sistem sebagai berikut:

1. Tampilan Menu Utama Program



Gambar 4.1 Antarmuka Menu Utama Aplikasi

Antarmuka ini merupakan form awal yang akan ditampilkan pada saat aplikasi dijalankan, pengguna harus mengklik tombol masuk jika ingin menjalankan proses cluster.

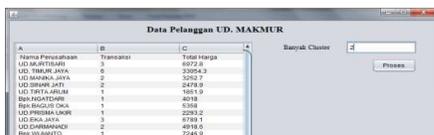
2. Antarmuka Pembacaan Data Pelanggan UD.Makmur



Gambar 4.2 Antarmuka Pembacaan Data Pelanggan

Antarmuka ini adalah antarmuka yang akan menampilkan data lengkap pelanggan UD.Makmur dalam melakukan transaksi selama 6 bulan, adapun data yang akan ditampilkan adalah nama perusahaan yang menjadi pelanggan UD.makmur, Jumlah Transaksi dan Total Harga Pembelaanjaan.

3. Antarmuka Proses Cluster K-Means



Gambar 4.3 Antarmuka Proses Cluster K means

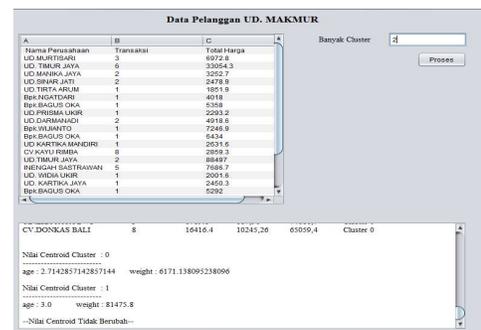
Antarmuka ini adalah antarmuka yang akan menampilkan data yang nantinya akan di melakukan cluster sesuai nilai yang di inputkan user .Adapun data yang akan ditampilkan adalah penghitungan jarak setiap data dengan masing-masing centroid, Untuk mendapatkan jarak dari masing-masing cluster pelanggan C1,C2 sampai Cn digunakan rumus euclidean distance sebagai berikut :

$$D_{L_2}(x_2, x_1) = \|x_2 - x_1\|_2 = \sqrt{\sum_{j=1}^p (x_{2j} - x_{1j})^2}$$

4. Antarmuka Inputan Centroid



Antarmuka ini merupakan antarmuka yang menampilkan inputan dari user untuk centroid awal dari masing –masing cluster. Centroid yang di masukan sebagai centroid awal merupakan data yang di ambil dari tabel pelanggan yang pemilihannya di lakukan secara acak.



Antarmuka ini adalah antarmuka yang akan menampilkan data iterasi-3, adapun data yang akan ditampilkan adalah nama perusahaan ,jumlah transaksi ,total harga,nilai cluster C1,C2....Cn dan nilai centroid akhir yang tidak berubah dari iterasi sebelumnya. Penghitungan jarak setiap data dengan masing-masing centroid, Untuk mendapatkan jarak dari masing- masing Nilai centroid dengan jumlah transaksi dan total harga digunakan rumus euclidean distance juga.

5.1 Kesimpulan

Berdasarkan data data yang diperoleh dari perusahaan UD.Makmur serta analisa dengan metode K-Means dapat diambil kesimpulan sebagai berikut :

1. Pengujian data pelanggan yang tidak berimbang lebih mudah di kelompokkan setelah dilakukan proses K-Means.
2. Proses iterasi dalam menentukan masing-masing cluster akan terus berlanjut sampai nilai dari centroid akhir dan nilai centroid sebelumnya tidak berubah.
3. Hasil yang di peroleh merupakan kelompok pelanggan yang berpotensi . Artinya pelanggan – pelanggan yang termasuk dalam data merupakan acuan untuk mempermudah pengambilan keputusan oleh perusahaan UD.Makmur

5.2 Saran

Hasil penelitian ini mungkin jauh dari kesempurnaan untuk perbaikan pengembangan yang lebih lanjut, hal-hal yang disaran penulis yaitu penambahan data set dan tambahan rule yang lebih banyak supaya dapat mengetahui tingkat keakuratan sistem lebih baik.

DAFTAR PUSTAKA

Dewanti Retno, 2008. *Kewirausahaan*, Mitra Wacana Media. Jakarta

Kursini & Emha taufiq luthfi, 2009. *Algoritma Data Maining*, CV Andi Offset.Yogyakarta.Offset. Yogyakarta.

Prasetyo Eko, 2012. *Data Mining: Konsep dan Aplikasi menggunakan MATLAB*, CV Andi Offset. Yogyakarta.

Prasetyo Eko, 2012. *Data Mining: Konsep dan Aplikasi menggunakan MATLAB*, CV Andi

A.S,Rosa dan Shalahuddin, M, 2011, Modul

Pembelajaran: Rekayasa Perangkat Lunak, Modula, Bandung

Santosa Budi, 2007. *Data Mining : Teknik*

Pemanfaatan Data untuk Keperluan Bisnis,

Graha Ilmu. Yogyakarta.

Satriyanto ,E, Clustering, Diakses 23 Maret 2015 dari

<http://www.id.scribd.com/doc/45119955/Clustering>

Sulianta Feri & Dominikus Juju, 2010. *Data Mining: Meramalkan Bisnis Perusahaan*, PT

Elex Media Komputindo. Jakarta.

Susanto Sani & Dedy Suryadi, 2010. *Pengantar Data Mining : Menggali Pengetahuan dari Bongkahan Data*, CV Andi Offset.Yogyakarta.

Sutojo,T dan Mulyanto, E., 2011, *Kecerdasan Buatan*, Andi Offset, Yogyakarta

Zulkarnain, 2012. *Ilmu Menjual*, Graha Ilmu. Yogyakarta.

