

SISTEM PENDUKUNG KEPUTUSAN PENGKATEGORIAN NASABAH MENGUNAKAN CLUSTERING DENGAN ALGORITMA K-MEANS

Elvina Sandra

AMIK Depati Parbo Kerinci
Jl. Pancasila No. 29, Sawahan – Sungai Penuh

E-Mail : elvina.sandra@gmail.com

ABSTRACT

Clustering adalah sebuah metode yang ada didalam sistem penunjang keputusan yang dapat digunakan untuk membedakan dan mengelompokan data dalam scala besar dimana dengan metode ini kita bisa menentukan karakter dari masing-masing kumpulan data. Mempelajari sifat dari nasabah adalah salah satu pokok utama yang harus dilakukan oleh seorang analis kredit agar tidak terjadinya permasalahan kredit macet dalam sebuah lembaga keuangan dengan menerapkan algoritma K_Means ini diharapkan kita bisa mempelajari bagaimana karakter – karakter nasabah pada sebuah lembaga perbankan agar kita bisa memprediksi kelancaran operasional lembaga perbankan tersebut dalam memberikan pelayanan serta menjaga kestabilan perusahaannya.

Keyword clustering, datamining, k_means

PENGANTAR

Pengambilan keputusan adalah sebuah pekerjaan yang sangat sulit dilakukan karena harus mempertimbangkan banyak hal apalagi jika keputusan tersebut memiliki pengaruh besar terhadap masa depan orang banyak ataupun masa depan kelancaran sebuah perusahaan. Disebuah lembaga keuangan nasabah memegang peranan terbesar dalam menentukan maju tidaknya sebuah lembaga keuangan, maka agar memudahkan sipembuat keputusan dalam memprioritaskan pelayanan maka kita harus mempelajari karakteristik nasabah yang dimiliki. *Clustering* adalah metode yang cocok digunakan untuk mengelompokan data nasabah tersebut karena *clustering* memiliki sifat *unsupervised learning* karena metode ini tidak memerlukan proses pelatihan data seperti klasifikasi.

Permasalahan

Dalam penelitian ini penulis menggunakan algoritma *K_Means* dalam proses *clustering* data, dimana parameter input yang digunakan adalah Usia

nasabah, tipe bidang usaha nasabah, jenis jaminan nasabah, dan track record pinjaman.

Objek penelitian kali ini adalah sebuah perusahaan yang bergerak dalam bidang keuangan dan penelitian lebih difokuskan kepada pengkajian karakteristik nasabah yang dimiliki oleh perusahaan ini dengan tujuan output yang dihasilkan nanti bisa dipergunakan sebagai acuan dalam proses pengambilan keputusan oleh para pembuat keputusan di perusahaan tersebut.

Tumpukan data nasabah di perusahaan ini sebenarnya mengandung ribuan informasi dan ilmu pengetahuan yang berharga namun diabaikan begitu saja karena sejauh ini hanya digunakan pada saat realisasi pinjaman dan baru dibuka kembali apabila terjadi penunggakan pada tagihan nasabah yang bersangkutan. Penulis akan berupaya memanfaatkan data-data tersebut sehingga bisa menghasilkan sesuatu yang berharga yang nantinya dapat dimanfaatkan dalam proses realisasi pinjaman atau pencairan dana nasabah.

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD (*Knowledge Discovery in Database*) secara keseluruhan. Data mining mampu menganalisa data yang besar menjadi informasi berupa pola yang mempunyai arti bagi pendukung keputusan. Hasil dari aplikasi data mining tersebut dievaluasi untuk menemukan suatu informasi atau pengetahuan baru yang menarik dan bernilai bagi perusahaan.

Salah satu metode yang bisa digunakan dalam Data Mining adalah metode *Clustering* yaitu sebuah metode dalam Data Mining yang bersifat tanpa arahan (*unsupervised*). Ada dua jenis data *Clustering* yang sering digunakan dalam proses pengelompokan data yaitu *hierarchical* (hirarki) data clustering dan *non hierarchical* (non hirarki) data clustering. *K-Means* adalah merupakan salah satu algoritma atau metode data clustering non hirarki yang berusaha mempartisi data yang ada kedalam bentuk satu atau lebih cluster atau kelompok. Metode ini mempartisi data kedalam cluster atau kelompok sehingga data yang memiliki

karakteristik yang sama dikelompokkan kedalam satu kelompok yang sama dan data dengan karakteristik yang berbeda di kelompokkan kedalam kelompok yang lain.

K-Means merupakan algoritma *clustering* yang berulang-ulang. Algoritma *K-Means* dimulai dengan pemilihan secara acak K , K disini merupakan banyaknya *cluster* yang ingin dibentuk. Kemudian tetapkan nilai-nilai K secara random, untuk sementara nilai tersebut menjadi pusat dari *cluster* atau sering disebut dengan centroid, Mean atau ‘Means’. Hitung jarak setiap data yang ada terhadap masing-masing centroid menggunakan rumus Euclidian hingga ditemukan jarak yang paling dekat dari setiap data dengan centroid. Klasifikasikan setiap data berdasarkan kedekatannya dengan centroid.lakukan langkah tersebut hingga nilai centroid tidak berubah (stabil) oleh Tedy Rismawan dan Sri Kusuma Dewi (2008)

Adapun tujuan dari pengelompokkan data ini adalah untuk meminimalisasikan fungsi obyektif yang diset dalam proses pengelompokkan, yang pada umumnya berusaha meminimalisasikan variasi didalam suatu kelompok dan memaksimalkan variasi antar kelompok.

Konsep dasar dari algoritma *K-Means* adalah pencarian pusat *cluster* ditetapkan berdasarkan jarak setiap data ke pusat *cluster*. Proses *clustering* dimulai dengan mengidentifikasi data yang akan di *cluster*, x_{ij} ($I=1, \dots, n; j=1, \dots, m$) dengan n adalah jumlah data yang akan di-*cluster* dan m adalah jumlah variabel.(Irwan Budiman, 2012)

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 perhitungan jarak data ke- I , (x_i) pada pusat *cluster* ke- k (C_k), diberi nama (d_{ik}), dapat digunakan formula *Euclidian* yaitu ;

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

Suatu data akan menjadi anggota dari *cluster* ke- j apabila jarak data tersebut ke pusat *cluster* ke- j bernilai paling kecil jika dibandingkan dengan jarak kepusat

cluster lainnya. Selanjutnya kelompokkan data-data yang menjadi 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 rumus ;

$$C_{kj} = \frac{\sum_{h=1}^p Y_{hj}}{p}; y_{hj} = x_{ij} \in \text{cluster ke}_k$$

Setelah itu hitunglah besaran rasio antara *Between cluster variation (bcv)* dan *Within cluster variation (wcv)* dengan formula sebagai berikut :

$$\text{rasio} = \frac{bcv}{wcv}$$

Data *clustering* menggunakan metode *K-Means* ini secara umum dilakukan dengan algoritma dasar sebagai berikut :

1. Tentukan jumlah *cluster*
2. Alokasikan data ke dalam *cluster* secara random
3. Hitung *centroid* rata-rata dari data yang ada dimasing-masing *cluster*
4. Alokasikan masing-masing data ke *centroid* / rata-rata terdekat
5. Kembali ke *step* 3, apabila masih ada data yang berpindah *cluster* atau apabila perubahan nilai *centroid*, ada yang diatas nilai *threshold* yang ditentukan atau apabila perubahan nilai pada *objective function* yang digunakan diatas nilai *threshold* yang ditentukan.

Dibanyak lembaga keuangan penilaian kelayakan kredit dilakukan menggunakan metode tradisional dan model pengambilan keputusan individu (*The Satisficing Models*). Dalam penilaian rencana pembiayaan nasabah pada objek penelitian ini memiliki beberapa kebijakan yang harus dipenuhi oleh calon nasabahnya untuk selanjutnya kebijakan-kebijakan tersebut dituangkan dalam bentuk scoring.

Dari sekian banyak parameter yang tertuang dalam scoring data nasabah yang ada maka disini penulis memutuskan untuk mengambil beberapa variabel yang dapat mewakili data nasabah secara keseluruhan untuk dilakukan proses *clustering*. Ada 4 (empat) variabel yang akan penulis gunakan dimana variabel tersebut diantaranya adalah sebagai berikut :

1. Usia Nasabah

Usia calon nasabah perlu mendapatkan perhatian saat memberikan pembiayaan kredit karena usia akan mempengaruhi bagaimana kinerja dari calon nasabah dalam melakukan proses pengembalian pinjaman. Sebab usia dapat menggambarkan masa produktif dari calon nasabah.

2. Tipe Bidang Usaha

Tipe bidang usaha adalah jenis usaha yang dilakukan oleh calon nasabah dan Parameter yang satu ini sangat memegang peranan dalam meninjau kemampuan pengembalian pinjaman calon nasabah karena kita bisa melihat dan mempertimbangan seberapa besar kemampuan calon nasabah kita dalam proses pengembalian pinjaman

3. Jenis Jaminan

Jenis jaminan adalah jaminan yang diberikan calon nasabah baik jaminan yang bergerak (seperti kendaraan bermotor, barang dagangan, mesin-mesin, deposito, piutang, perhiasan, dan lain-lain) ataupun jaminan yang tidak bergerak (seperti tanah, bangunan, dan kios)

4. Track Record Pinjaman

Track record pinjaman adalah ada tidaknya hubungan calon nasabah kita dengan perusahaan keuangan lain dari sini kita bisa lihat apakah calon nasabah kita memiliki pinjaman lain di perusahaan keuangan lainnya, dan dari sini kita juga bisa melihat apakah calon nasabah memiliki masalah dengan perusahaan keuangan lainnya.

Proses pengolahan data disini menggunakan 125 sampel data nasabah yang telah dicairkan pinjamannya dan dianggap layak.

Seperti yang telah penulis jelaskan diatas berikut ini adalah langkah pengelompokkan data dengan menggunakan algoritma *K_Means* :

a. Iterasi Pertama

1. Menentukan jumlah cluster atau nilai K dalam hal ini penulis mengambil nilai $K=2$
2. Menentukan nilai pusat cluster, pada iterasi pertama bisa ditentukan secara acak ($c_1=1.5;1.5;3;1$ dan $c_2= 2;3;2.5;1.5$) dan kemudian lanjutkan dengan

menghitung jarak setiap data dengan pusat cluster hal ini dapat dilakukan dengan menggunakan rumus *Euclidian*.

Misalkan untuk menghitung jarak data nasabah pertama dengan pusat cluster pertama dapat dilakukan seperti berikut ini :

$$d_{1.1} = \sqrt{(1.5 - 1.5)^2 + (3 - 1.5)^2 + (3 - 3)^2 + (1 - 1)^2} = 1.5000$$

Jarak data nasabah kedua dengan pusat cluster kedua adalah sebagai berikut :

$$d_{1.2} = \sqrt{(1.5 - 2)^2 + (3 - 3)^2 + (3 - 2.5)^2 + (1 - 1.5)^2} = 0.8660$$

Lakukanlah proses tersebut untuk seluruh sampel data

3. Suatu data akan menjadi anggota dari suatu cluster yang memiliki jarak terkecil dari pusat clusternya. Misalkan pada data pertama, jarak terkecil diperoleh pada cluster ke 2 (hal ini dapat dilihat pada kolom nilai Min) sehingga data pertama akan menjadi anggota dari cluster kedua (dapat dilihat pada kolom keterangan) pada tabel 4.4 diatas. Kemudian hitung rasio antara besaran *Between Cluster Variation* dengan *Within cluster variation* lalu bandingkan rasio tersebut dengan rasio sebelumnya (bila sudah ada). Jika rasio tersebut membesar, lanjutkan kelangkah keempat. Jika tidak hentikan prosesnya.

Dalam hal ini perbandingan antara *Between Cluster Variation* dengan *Within Cluster Variation* adalah sebagai berikut:

$$BCV = \sqrt{((1.5 - 2)^2 + (1.5 - 3)^2 + (3 - 2.5)^2 + (1 - 1.5)^2)}$$

$$BCV = 1.7321WCV$$

$$\begin{aligned}
&= ((0.8660)^2 + (0.7071)^2 + (1.5000)^2 + (0.7071)^2 \\
&+ (1.1180)^2 + (0.0000)^2 + (1.1180)^2 + (1.8028)^2 \\
&+ (0.5000)^2 + (0.7071)^2 + (1.1180)^2 + (2.0616)^2 \\
&+ (0.5000)^2 + (0.8660)^2 + (0.7071)^2 + (1.6583)^2 \\
&+ (0.7071)^2 + (0.7071)^2 + (0.8660)^2 + (1.0000)^2 \\
&+ (0.8660)^2 + (0.8660)^2 + (0.7071)^2 + (0.7071)^2 \\
&+ (0.5000)^2 + (2.1213)^2 + (0.7071)^2 + (0.7071)^2 \\
&+ (1.0000)^2 + (0.8660)^2 + (0.7071)^2 + (2.1213)^2 \\
&+ (0.5000)^2 + (0.5000)^2 + (0.7071)^2 + (0.7071)^2 \\
&+ (0.7071)^2 + (0.5000)^2 + (0.7071)^2 + (0.7071)^2 \\
&+ (1.6583)^2 + (0.8660)^2 + (0.7071)^2 + (0.0000)^2 \\
&+ (1.8028)^2 + (0.5000)^2 + (0.5000)^2 + (0.7071)^2 \\
&+ (0.8660)^2 + (0.5000)^2 + (0.0000)^2 + (0.7071)^2 \\
&+ (0.8660)^2 + (0.8660)^2 + (0.8660)^2 + (0.7071)^2 \\
&+ (1.8707)^2 + (0.7071)^2 + (0.8660)^2 + (0.8660)^2 \\
&+ (0.7071)^2 + (1.1180)^2 + (1.1180)^2 + (0.5000)^2 \\
&+ (0.7071)^2 + (1.1180)^2 + (0.5000)^2 + (0.7071)^2 \\
&+ (0.7071)^2 + (0.7071)^2 + (0.7071)^2 + (1.1180)^2 \\
&+ (1.8708)^2 + (0.8660)^2 + (0.7071)^2 + (0.5000)^2 \\
&+ (0.7071)^2 + (0.8660)^2 + (1.3229)^2 + (0.7071)^2 \\
&+ (0.7071)^2 + (0.7071)^2 + (1.0000)^2 + (0.7071)^2 \\
&+ (0.8660)^2 + (0.8660)^2 + (0.7071)^2 + (0.7071)^2 \\
&+ (0.5000)^2 + (0.8660)^2 + (0.7071)^2 + (0.7071)^2 \\
&+ (0.8660)^2 + (0.5000)^2 + (0.5000)^2 + (0.8660)^2 \\
&+ (0.7071)^2 + (1.8028)^2 + (0.8660)^2 + (0.8660)^2 \\
&+ (0.5000)^2 + (1.8708)^2 + (0.7071)^2 + (0.7071)^2 \\
&+ (0.8660)^2 + (1.1180)^2 + (0.7071)^2 + (0.8660)^2 \\
&+ (1.8708)^2 + (0.8660)^2 + (0.7071)^2 + (0.8660)^2 \\
&+ (0.8660)^2 + (0.0000)^2 + (0.8660)^2 + (1.0000)^2
\end{aligned}$$

$$+ (0.7071)^2 + (0.8660)^2 + (0.8660)^2 + (0.7071)^2 \\ + (0.0000)^2 + (1.8708)^2$$

$$\frac{BCV}{WCV} = \frac{1.7321}{114.2500} \\ = 0.0152$$

4. Menghitung pusat cluster baru, dapat dilakukan dengan menjumlahkan semua data pada variabel yang ada pada masing-masing nasabah dan dibagi dengan jumlah keseluruhan data pada masing-masing cluster. Misalnya untuk cluster pertama ada 44 data nasabah sehingga:

$$P1(TUsia) = 1.5+1.5+1.5+1.5+2+2+1.5+1.5+1.5+2+1.5+1.5+1.5+1.5+1.5 \\ +1.5+2+1.5+2+1.5+2+1.5+1.5+1.5+2+2+1.5+1+1.5+1.5+ \\ 1.5+1.5+1+1.5+1.5+1.5+2+2+2+2+2+2+2)/44 = 1.6477$$

Dengan melakukan cara yang sama yaitu dengan menjumlahkan semua data yang ada pada variabel yang sesuai dengan cluster pertama dan dibagi dengan 44 maka akan didapatkan nilai –nilai berikut ini

Nilai P1(TTBU) sebagai berikut

$$P1(TTBU) = 2+1+1.5+2+1.5+1+2+1+2+2+1+2+2+1+1+2+2+1+1+2+2+2+2 \\ +2+1+1+2+2+2+1+2+2+2+1+2+2+2+2+2+1+2+2+1+1+2)/4 \\ 4=1.6364$$

Nilai P1(TJJ) sebagai berikut

$$P1(TJJ) = (3+3+3+3+1.5+3+3+1+3+3+1+3+3+1.5+3+3+3+3+ \\ 3+3 \\ +3+3+3+3)/44=2.8409$$

Nilai P1(TTRP) sebagai berikut

$$P1(TTRP) = (1.5+1.5+1+1.5+1.5+1+1.5+1.5+1.5+1+1.5+1+1.5+1.5 \\ +1.5+1+1.5+1.5+1.5+1.5+1.5+1+1.5+1.5+ \\ 1.5+1.5+1.5+1+1.5+1.5+1.5+1.5+1.5+1+1.5+1.5+1+ \\ 1.5+1.5+1+1.5+1.5+1)/44=1.3750$$

Sementara untuk pusat cluster baru pada cluster kedua karena jumlah datanya ada sebanyak 81 orang maka semua data pada masing-masing variabel yang terdapat dicluster kedua dibagi dengan 81.

Nilai P2(TU_{usia}) sebagai berikut :

$$\begin{aligned}
 P1(TU_{usia}) &= (1.5+2+1+1.5+2+2+2+1+2+1.5+1.5+2+2+1.5+1.5+2 \\
 &\quad +2+2+2+1.5+2+2+1.5+1.5+2+2+2+1.5+2+2+2+1.5+ \\
 &\quad 1.5+1.5+1.5+2+2+1.5+1.5+2+1.5+1.5+1.5+1.5+1.5+ \\
 &\quad 2+2+2+1.5+2+1.5+1.5+1.5+1.5+1.5+2+1.5+1.5+2+2 \\
 &\quad +1.5+2+1.5+1.5+1+1.5+1.5+1.5+1.5+2+1.5+ \\
 &\quad 1+2+1.5)/81=1.6914
 \end{aligned}$$

Nilai P2(TT_{BU}) sebagai berikut :

$$\begin{aligned}
 (TT_{BU}) &= (3+3+3+3+2+3+3+2+3+3+3+2.5+3+3+2.5+2.5+3+3+3+2.5 \\
 &\quad +3+3+3+3+3+2.5+3+3+2+3+3+3+3+3+3+2+2.5+3 \\
 &\quad +3+3+3+3+3+3+2+3+3+2.5+3+3+2.5+3+3+3+2.5 \\
 &\quad +2.5+3+2.5+2.5+3+3+3+3+3+2+2.5+2.5+2+3+2.5+2.5+2 \\
 &\quad +3+2.5+3+3+3+3+2)/81=2.7901
 \end{aligned}$$

Nilai P2(TJ_J) sebagai berikut :

$$\begin{aligned}
 (TT_{BU}) &= (3+1+3+1.5+1+3+1.5+1+2.5+2+3+3+1.5+3+2.5+3+ \\
 &\quad 3+3+1.5+2.5+3+2.5+3+3+3+3+2.5+2.5+3+3+2.5+3+3+ \\
 &\quad 3+1+3+3+1.5+1.5+3+1.5+2.5+3+1.5+1+3+2.5+3+3+2+ \\
 &\quad 1.5+1.5+3+3+3+3+3+2.5+2+3+2.5+3+3+1+3+3+1+3+ \\
 &\quad 3+1+3+3+2.5+2+2.5+2.5+1)/81=2.4259
 \end{aligned}$$

Nilai P2(TTR_P) sebagai berikut :

$$\begin{aligned}
 (TT_{BU}) &= (1+1.5+1.5+1.5+1.5+1.5+1+1.5+1+1+1.5+1+1.5+1+ \\
 &\quad 1+1.5+1+1.5+1.5+1+1+1.5+1+1.5+1.5+1+1.5+1+ \\
 &\quad 1.5+1.5+1.5+1.5+1.5+1.5+1+1+1.5+1.5+1+1+1.5+ \\
 &\quad 1.5+1.5+1.5+1+1.5+1.5+1.5+1+1+1.5+1+1.5+1+1.5+ \\
 &\quad 1.5+1+1.5+1.5+1.5+1.5+1+1.5+1.5+1+1+1.5+1.5+1+ \\
 &\quad 1.5+1+1.5+1.5+1.5+1.5+1.5+1+1.5+1.5+1.5)/81=1.3210
 \end{aligned}$$

5. Ulangi langkah 2 pada iterasi berikutnya hingga posisi data sudah tidak mengalami perubahan, atau jika nilai perbandingan antara *between cluster*

variation dengan *withen cluster variation* nya sudah sama dengan hasil pada iterasi sebelumnya atau semakin mengecil.

Dari hasil percobaan dengan $K=2$ maka dengan deretan data yang ada maka proses pengclusterannya tercapai pada iterasi ke-4, berikut ini adalah hasil dari proses klusterisasinya.

b. Iterasi Kedua

Dari proses iterasi pertama sebelumnya kita telah mendapatkan nilai untuk pusat cluster berikutnya yaitu untuk ($c_1=1.6477;1.6364;2.8409;1.3750$ dan $c_2=1.6914;2.7901;2.4259;1.3210$), dengan nilai c_1 dan c_2 tersebut kita dapat mencari jarak antara data dengan pusat cluster berikutnya, misalkan kita akan mencari jarak antara data pertama dengan pusat cluster dapat dilakukan dengan cara berikut ini:

$d_{2.1}$

$$= \sqrt{(1.5 - 1.6477)^2 + (3 - 1.6364)^2 + (3 - 2.8409)^2 + (1 - 1.3750)^2}$$

$$= 1.4308$$

$d_{2.2}$

$$= \sqrt{(1.5 - 1.6914)^2 + (3 - 2.7901)^2 + (3 - 2.4259)^2 + (1 - 1.3210)^2}$$

$$= 0.7164$$

Dari proses pengolahan data tersebut didapatkan nilai BCV dan WCV sebagai berikut:

BCV

$$= \sqrt{\frac{((1.6875 - 1.6556)^2 + (2.8000 - 1.6444)^2 + (2.4438 - 2.8000)^2 + (1.3778 - 1.3188)^2)}{4}}$$

$$= 1.2111$$

Dengan cara yang sama dengan yang kita lakukan di iterasi pertama maka akan kita dapatkan nilai wcv sebagai berikut

$$wcv = 90.9612$$

$$\frac{BCV}{WCV} = \frac{1.2111}{90.9612} = 0.0133$$

Jika dibandingkan nilai perbandingan BCV dan WCV pada iterasi pertama dan iterasi kedua pada iterasi ketiga ini nilainya masih belum memiliki kesamaan

maka kita harus melanjutkan proses pada itersi keempat dan untuk itu kita harus menentukan pusat cluster berikutnya dengan cara yang sama dengan yang kita lakukan di *iterasi kedua* maka kita dapatkan nilai untuk:

$$\begin{array}{ll} P1(Tusia) = 1.6477 & P2(Tusia) = 1.6914 \\ P1(TTBU) = 1.6364 & P2(TTBU) = 2.7901 \\ P1(TJJ) = 2.8409 & P2(TJJ) = 2.4259 \\ P1(TTRP) = 1.3750 & P2(TTRP) = 1.3210 \end{array}$$

Dengan data ini kita dapat melanjutkan keiterasi berikutnya dimana nilai diatas dapat kita jadikan sebagai pusat cluster baru pada iterasi keempat.

c. Iterasi Keempat

Dari proses iterasi ketiga sebelumnya kita telah mendapatkan nilai untuk pusat cluster berikutnya yaitu untuk ($c1=1.6477;1.6364;2.8409;1.3750$ dan $c2=1.6914;2.7901;2.4259; 1.3210$), dengan nilai $c1$ dan $c2$ tersebut kita dapat mencari jarak antara data dengan pusat cluster berikutnya, misalkan kita akan mencari jarak antara data pertama dengan pusat cluster dapat dilakukan dengan cara berikut ini:

$d_{3.1}$

$$= \sqrt{(1.5 - 1.6477)^2 + (3 - 1.6364)^2 + (3 - 2.8409)^2 + (1 - 1.3750)^2}$$

$$= 1.4308$$

$d_{3.2}$

$$= \sqrt{(1.5 - 1.6914)^2 + (3 - 2.7901)^2 + (3 - 2.4259)^2 + (1 - 1.3210)^2}$$

$$= 0.7164$$

Dari proses pengolahan data tersebut didapatkan nilai BCV dan WCV sebagai berikut:

BCV

$$= \sqrt{\frac{((1.6914 - 1.6477)^2 + (2.7901 - 1.6364)^2 + (2.4259 - 2.8409)^2 + (1.3210 - 1.3750)^2)}{4}}$$

$$= 1.2281$$

Dengan cara yang sama dengan yang kita lakukan di iterasi pertama maka akan kita dapatkan nilai wcv sebagai berikut

$$wcv = 90.8464$$

$$\frac{BCV}{WCV} = \frac{1.2281}{90.8464} = 0.0135$$

Jika dibandingkan nilai perbandingan *BCV* dan *WCV* pada iterasi pertama, iterasi kedua dan iterasi ketiga maka pada iterasi keempat ini nilainya sudah memiliki kesamaan nilai dengan proses perbandingan *BCV* dan *WCV* pada iterasi kedua, maka kita sudah dapat menghentikan proses iterasi data.

Dari hasil proses iterasi diatas didapatkan beberapa kesimpulan tentang data nasabah PNM sebagai berikut:

1. Terdapat 45 orang nasabah yang berada pada kluster pertama dan 80 orang nasabah yang tergolong kedalam kluster kedua
2. Dari proses pengklusteran yang dilakukan maka dapat disimpulkan beberapa ciri-ciri masing-masing kluster sebagai berikut:
 - a. Kluster pertama memiliki titik pusat (1.6556;1.6444; 2.8000;1.3778) yang dapat diartikan nasabahnya lebih dominan memiliki usia berkisar antara (21-45) tahun, dengan tipe bidang usaha dominan pada pertanian dan perternakan, dan jenis jaminan terbanyak adalah tanah dan bangunan, serta dengan track record pinjaman lebih sedikit yang memiliki pinjaman dilembaga keuangan lainnya.
 - b. Kluster kedua memiliki titik pusat (1.6875;2.8000;2.4438; 1.3188) ini dapat diartikan bahwa nasabahnya berusia antara (21-45) tahun, dengan tipe bidang usaha terbanyak adalah perdagangan, dan jenis jaminan yang jadi anggunan adalah kombinasi antara tanah dan mobil atau tanah dan bangunan, dan track record pinjamannya lebih dominan pada nasabah yang tidak memiliki pinjaman dilembaga keuangan lainnya.

Dengan adanya kesimpulan diatas maka hal tersebut bisa dijadikan acuan oleh manager kredit ketika akan melakukan realisasi terhadap pinjaman nasabah untuk kedepannya dan memungkinkan bisa meminimalkan terjadinya penunggakan nasabah dalam proses pengembalian pinjaman.

DAFTAR PUSTAKA

- Divisi PPL PNM . (2012). "*Accelerated Training Program*". Jakarta : Permodalan Nasional Madani
- Jong Jek Siang. (2009). "*Jaringan Syaraf Tiruan & Pemrogramannya*". Yogyakarta : Andi Offset
- Kusrini dan Luthfi Taufiq Emha. (2009). "*Algoritma Data Mining*". Yogyakarta : Andi Offset.
- Paulraj Ponniah. (2010). "*Data Warehousing : Fundamentals for IT Professional*". Canada : A John Wiley Sons. Inc. Publication.
- S. Sumanthi and S.N. Sivanandam. (2006). "*Introduction to Data Mining and its Applications*". New York : Springer.
- Sani Susanto dan Dedy Suryadi. (2010). "*Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data*". Yogyakarta : Andi Offse