ARTIKEL

ANALISIS DATA MINING UNTUK PENENTUAN RESIKO KREDIT KEPEMILIKAN KENDARAAN BERMOTOR MENGGUNAKAN ALGORITMA K- NEAREST NEIGHBOR

Oleh:

CANDRA EKO SAPUTRO

14.1.03.02.0064

Dibimbing oleh :

1. **Patmi Kasih, M.Kom**
2. **Risa Helilintar, M.Kom**

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS NUSANTARA PGRI KEDIRI

TAHUN 2018

Artikel Skripsi
Universitas Nusantara PGRI Kediri

SURAT PERNYATAAN
ARTIKEL SKRIPSI TAHUN 2018

Yang bertanda tangandibawahini:

Nama Lengkap	: Candra Eko Saputro
NPM	: 14.1.03.02.0064
Telepun/HP	: 081553515240
Alamat Surel (Email)	: candrasaputro95@gmail.com
Judul Artikel	: Analisis Data Mining Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor Menggunakan Algoritma K-Nesrest Neighbor
Fakultas – Program Studi	: Teknik – Teknik Informatika
Nama Perguruan Tinggi	: Universitas Nusantara PGRI Kediri
Alamat PerguruanTinggi	: Jl. KH. Ahmad Dahlan No.76 Kediri

Dengan ini menyatakan bahwa:

a. artikel yang saya tulis merupakan karya saya pribadi (bersama tim penulis) dan bebas plagiarisme.
b. artikel telah diteliti dan disetujui untuk diterbitkan oleh Dosen Pembimbing I dan II.

Demikian surat pernyataan ini saya buat dengan sesungguhnya. Apabila dikemudian hari ditemukan ketidaksesuaian data dengan pernyataan ini dan atau ada tuntutan dari pihak lain, saya bersedia bertanggung jawab dan diproses sesuai dengan ketentuan yang berlaku.

Mengetahui		Kediri,8 Agustus 2018
Pembimbing I	Pembimbing II	Penulis,
Patmi Kasih, M.Kom. NIDN. 0701107802	Risa Helilintar, M.Kom. NIDN. 0721058902	Candra Eko Saputro. 14.1.03.02.0064

Candra Eko Saputro | 14.1.03.02.0064 simki.unpkediri.ac.id
Teknik – Teknik Informatika || 1||

ANALISIS DATA MINING UNTUK PENENTUAN RESIKO KREDIT KEPEMILIKAN KENDARAAN BERMOTOR MENGGUNAKAN ALGORITMA K- NEAREST NEIGHBOR

Candra Eko Saputro
14.1.03.02.0064
Fakultas Teknik – Teknik Informatika
candrasaputro95@gmail.com
Patmi Kasih, M.Kom. dan Risa Helilintar, M.Kom.
UNIVERSITAS NUSANTARA PGRI KEDIRI

ABSTRAK

Perusahaan *leasing* PT Adira Dinamika Multi Finance sulit untuk menentukan calon konsumen *good cust* dan *bad cust* dengan sistem yang masih manual. Perlu penerapan sistem data mining untuk membantu menentukan calon konsumen *good cust* dan *bad cust*. Tujuan dari sistem ini adalah untuk merancang dan membangun sistem yang dapat menyimpan data konsumen kredit dan mengolahnya untuk menghasilkan kelas konsumen *good cust* dan *bad cust*.

Metode yang digunakan adalah *K-Nearest Neighbor*. *K-Nearest Neighbor* digunakan untuk mengklasifikasikan calon konsumen kredit berdasarkan jarak kedekatan atribut. Untuk menghitung jarak kedekatan antara data baru dan data pelatihan dapat menggunakan beberapa macam perhitungan jarak seperti *euclidean*, *manhattan*, dan *mahalanobois*. Dalam penelitian ini perhitungan yang digunakan adalah perhitungan jarak *euclidean*.

Tahapan proses penggunaan sistem, langkah pertama admin melakukan proses *input* data pemohon kredit yang digunakan sebagai *testing*, kemudian dihitung proses *K-Nearest Neighbor* yaitu mencari jarak terdekat *euclidean* dengan data pelanggan kredit sebagai *training*. Setelah proses hitung KNN didapat hasil *klasifikasi* yaitu calon konsumen *good cust* atau *bad cust*.

Hasil sistem yang dibuat antara lain : sistem yang dibuat mengacu pada permasalahan yang ada, dimana sistem mampu menyimpan data calon konsumen kredit dan sistem dapat mengolah data tersebut. Penerapan metode *K-Nearest Neighbor* dapat mengklasifikasikan calon konsumen kredit *good cust* dan *bad cust* berdasarkan jarak terdekatan atribut, sehingga dari hasil metode *K-Nearest Neighbor* dapat disimpulkan hasil *klasifikasi* calon konsumen kredit *good cust* dan *bad cust* sesuai kriteria yang telah ditentukan. Hasil *klasifikasi* calon konsumen kredit dapat digunakan sebagai rekomendasi penerimaan konsumen kredit oleh *manager* PT Adira Dinamika Multi Finance.

KATA KUNCI : Metode K-Nearest Neighbor, data mining, penentuan resiko kredit.

I. LATAR BELAKANG

Menurut Rivai (2006), kedit didefiniskan sebagai berikut.

> Kredit adalah penyerahan barang, jasa, atau uang dari satu pihak (kreditor/pemberi pinjaman) atas dasar kepercayaan kepada pihak lain (nasabah atau pengutang/ *borrower*) dengan janji membayar dari penerima kredit kepada pemberi kredit pada tanggal yang telah disepakati kedua belah pihak.

Agar kredit yang diberikan mencapai sasaran dan aman, maka analisis kredit perlu dilakukan. Menurut Rivai (2006), analisa kredit didefinisikan sebagai berikut.

Analisis kredit adalah kajian yang dilakukan untuk mengetahui kelayakan dari suatu permasalahan kredit. Melalui hasil analisis kreditnya, dapat diketahui apakah usaha nasabah layak (feasible), marketable (hasil usaha dapat dipasarkan), profitable (menguntungkan), serta dapat dilunasi tepat waktu. Untuk mewujudkan hal tersebut, perlu dilakukan persiapan kredit, yaitu dengan mengumpulkan informasi dan data untuk bahan analisis. Kualitas hasil analisis tergantung pada kualitas SDM, data yang diperoleh, dan teknik analisis.

Perusahaan kredit atau *leasing* memberikan jasa membantu masyarakat untuk mendapatkan kendaraan transportasi seperti sepeda motor, namun terdapat kendala-kendala yang diterima oleh perusahaaan seperti terjadinya macet dalam pembayaran kredit yang mengakibatkan kerugian yang berdampak bagi perusahaan *leasing* pada PT. Adira Dinamika Multi Finance Kediri yang meninmbulkan kerugian dan resiko. 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.

Banyaknya kiteria yang digunakan dalam menerima calon konsumen baru diperusahaan PT. Adira Dinamika Multi Finance Kediri seperti pekerjaan, jenis pekerjaan, penghasilan per-bulan, pendapatan lainnya, status, jumlah tanggungan, kepemilikan rumah, kondisi rumah, pendidikan, DP awal yang mungkin akan diterapkan. Sehingga membuat perusahaan harus membuat keputusan yang tepat agar tidak terjadinya kesalahan dalam memilih calon konsumen dengan memanfaatkan kriteria tadi. Dengan menggunakan kriteria yang sesuai dan tepat dapat menghasilkan *klasifkasi* calon konsumen kredit yang tepat, maka dibutuhkan rekomendasi yang tepat. Perekomendasi yang tepat membutuhkan data yang banyak. Algoritma *K-Nearest Neighbor* merupakan salah satu *algoritma* yang dapat menganalisa dengan jumlah data yang banyak atau bisa disebut juga data mining. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan. Dalam penelitian ini akan menggunakan algoritma *K-Nearest Neighbor* untuk menentukan resiko kredit.

Penelitian *klasifikasi* penentuan calon konsumen di perusahaan kredit kendaraan bermotor ini bertujuan untuk memberikan solusi yang tepat untuk menentukan calon konsumen *bad cust* atau *good cust* menggunakan metode data mining *k-Nearest Neighbor*. Memanfaatkan data mining atau data terdahulu kemudian diolah dengan data baru kemudian menghasilkan data target. Dengan adanya *klasifikasi* calon konsumen, diharapkan mampu membantu pemilik perusahaan dalam menentukan calon konsumen *Bad Cust* atau *Good Cust* dengan memanfaatkan hasil klasifikasi calon konsumen.

II. METODE

A. Data Mining

Menurut Pramudiono (2006), data mining di definisikan sebagai berikut.

> Data mining adalah analisa otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya.

B. K-Nearest Neighbor

Menurut Vercellis (2009), *K-Nearest Neighbor* di definisikan sebagai berikut.

> *k-nearest neighbor* (kNN) termasuk kelompok *instance-based learning*. Algoritma ini juga merupakan salah satu teknik *lazy learning*. kNN dilakukan dengan mencari kelompok k objek dalam data training yang paling dekat (mirip) dengan objek pada data baru atau data testing.

K-Nearest Neighbor adalah sebuah metode untuk melakukan *klasifikasi* terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. *K-Nearest Neighbor* merupakan metode *klasifikasi instance-based*, memilih satu objek latih yang memiliki sifat ketetanggaan (*neighborhood*) yang paling dekat.

Kedekatan dapat dianggap sebagai *invers* jarak, alias berbanding terbalik dengan jarak. Semakin kecil jarak antara dua *instance*, semakin besar "kedekatan" antara dua *instance* tersebut. Dengan demikian, *k nearest neighbours* dari sebuah *instance* x didefinisikan sebagai k *instance* yang memiliki jarak terkecil (kedekatan terbesar, *nearest*) dengan x.

Ada banyak cara untuk mengukur jarak kedekatan antara data baru dengan data lama (data *training*), diantaranya

1. Distance =

$$\sqrt{(x1-x2)^2+(y1-y2)^2}$$

2. Euclidean =

$$\sqrt{\sum_{i=1}^{k}(xi-yi)^2}$$

3. Manhattan =

$$\sum_{i=1}^{k}|xi-yi|$$

4. Minkowski =

$$\left(\sum_{i=1}^{k}(|xi-yi|)^q\right)^{1/q}$$

Menurut Larose (2005) menyatakan bahwa.

> Untuk mengukur jarak dari atribut yang mempunyai nilai besar, seperti atribut pendapatan, maka dilakukan normalisasi. Normalisasi bisa dilakukan dengan *min-max normalization* atau *Z-score standardization*. Jika data training terdiri dari atribut campuran antara numerik dan kategori, lebih baik gunakan *min-max normalization*.

Tahapan langkah algoritma K-NN.

a) Menentukan parameter k (jumlah tetangga paling dekat).
b) Menghitung kuadrat jarak *euclidean* objek terhadap data training yang diberikan.
c) Mengurutkan hasil hitung jarak *euclidean* secara *ascending* (berurutan dari nilai rendah ke tinggi)
d) Mengumpulkan kategori Y (*Klasifikasi nearest neighbor* berdasarkan nilai k).
e) Dengan menggunakan kategori *nearest neighbor* yang paling mayoritas maka dapat dipredisikan kategori objek.

C. Cunfison matrix

Confusion matrix adalah suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Rumus ini melakukan perhitungan dengan 4 keluaran yaitu: *recall, precision, acuraccy, dan eror rate.*

1. *Recall* adalah proporsi kasus positif yang diidentifikasi dengan benar. Rumus dari *recall* $= \frac{d}{(c+d)}$
2. *Precision* adalah proporsi kasus dengan hasil positif yang benar. Rumus dari *precision* $= \frac{d}{(b+d)}$
3. *Accuracy* adalah perbandingan kasus yang diidentifikasi benar dengan jumlah semua kasus. Rumus dari *accuracy* $= \frac{(a+c)}{(a+b+c+d)}$
4. *Eror Rate* adalah kasus yang diidentifikasi salah dengan sejumlah semua kasus. Rumus dari *eror rate* $= \frac{(b+c)}{(a+b+c+d)}$

Keterangan :

a jika hasil prediksi negatif dan data sebenarnya negatif.

b jika hasil prediksi positif sedangkan nilai sebenarnya negatif.

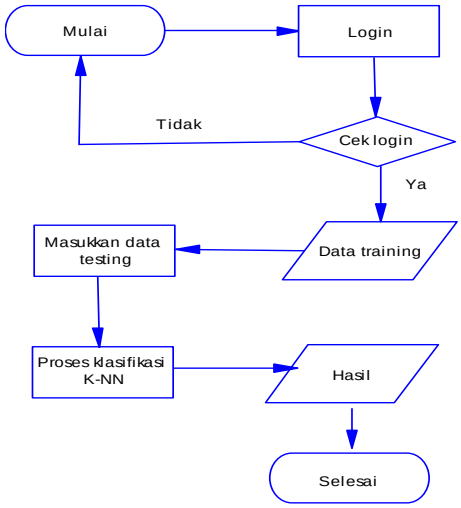
c jika hasil prediksi negatif sedangkan nilai sebenarnya positif.

d jika hasil prediksi positif dan nilai sebenarnya positif.

III. HASIL DAN KESIMPULAN

A. Implementasi Sistem

1. *Flowchart system*



Gambar 1.1 : *Flowchart Diagram*

Adapun keterangan alur dari gambar 1.1 *Flowchart Diagram* adalah sebagai berikut.

a) User sebelum melakukan proses input dan perhitungan K-NN harus login berdasarkan username dan password yang sudah ditentukan.

b) Muat data training berisi data pelanggan lama.

c) Masukkan data testing atau data pemohon kredit, kemudian data akan tersimpan secara otomatis ke dalam *database*.

d) Tahap selanjutnya proses perhitungan dengan metode *K-Nearest Neighbor*, dengan mencari jarak tetangga terdekat dari data pelanggan lamadan data pemohon kredit.

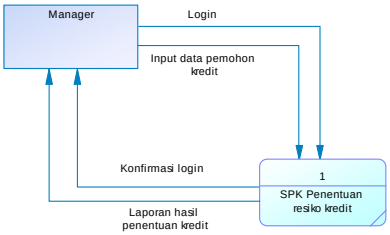
e) Hasil akan menampilkan hasil berupa klasifikasi kelas *Good Cust* atau *Bad Cust*.

f) Selesai.

2. *Context Diagram*

Diagram konteks atau disebut juga model sistem fundamental mempresentasikan seluruh elemen sistem sebagai sebuah bubble tunggal dengan data input dan output yang ditunjukkan oleh anak panah yang masuk dan keluar secara bertahap.

Berdasarkan hasil analisis, maka diagram konteks dari penerapan penentuan resiko kredit perusahaan sepeda motor, dapat dilihat pada gambar 1.2 *Context Diagram* dibawah ini :

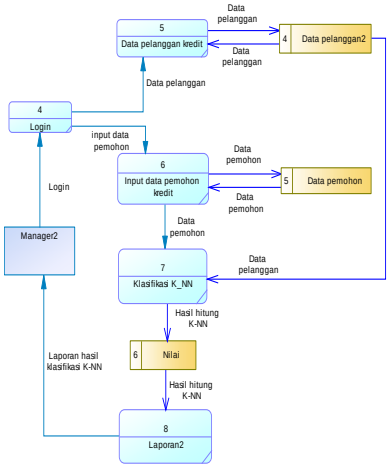


Gambar 1.2 : *Context Diagram*

Pada gambar 1.2 aktivitas pada aplikasi yaitu admin melakukan login dan input data pemohon kredit dan melakukan proses hitung KNN.

3. *DFD*

Data Flow Diagram merupakan pengembangan dari *DFD Level* 0 yang terdiri dari 5 proses yaitu *login*, muat data *training*, pengolahan data *testing*, proses hitung KNN dan hasi hitung KNN.



Gambar 1.3 : DFD Level 1

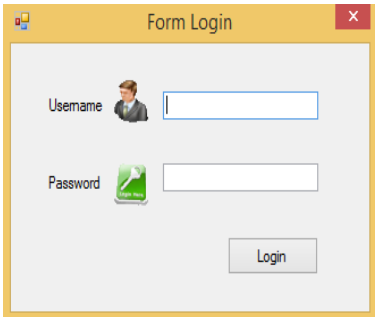
Pada gambar 1.3 DFD level 1 *Data Flow Diagram* merupakan pemecahan dari proses *context diagram* yang bersifat global. Sehingga dalam diagram iniakan lebih detai gambaran proses-proses yang akan dibuat. Pada DFD ini menganalisa lebih detail mulai dari personal yang terlibat, aliran, dan penyimpanan data pada basis data. Model ini menggambarkan sistem sebagai jaringan kerja antara fungsi yang berhubungan satu sama lain dengan aliran dan penyimpanan data pada basis data. Sebagai perangkat analisis model ini hanya dapat dimodelkan dari satu sudut pandang, yaitu sudut pandang fungsi.

B. Tampilan Program

Pada aplikasi penentuan resiko kredit kendaraan bermotor tampilan program dibuat dengan desain yang sederhana dengan tujuan untuk mempermudah penggunanya. Berikut tampilan program yang dibuat.

1. Tampilan *Login*

Sebelum mengakses aplikasi, seorang user harus melalui proses login dimana halaman login adalah tampilan yang pertama muncul ketika aplikasi dijalankan.



Gambar 1.4 : Tampilan Login

Dari gambar 1.4 tampilan *Login* seorang user harus melalui proses login dimana pada proses ini, user akan memasukkan data login berupa username dan password.

2. Tampilan Menu Utama

Menu utama adalah jembatan untuk mengakses halaman dengan fungsi yang sudah di kembangkan pada aplikasi ini, berikut menu yang di tampilakn dalam menu utama.

Gambar 1.5 Tampilan Menu Utama

Dari gambar tampilan pada halaman beranda terdapat 3 buah *button*, fungsi dari masing – masing elemen tersebut adalah sebagai berikut.

a) *Button Training*

Button ini merupakan sebuah *button* untuk mengakses halaman *training.*

b) *Button* Proses KNN

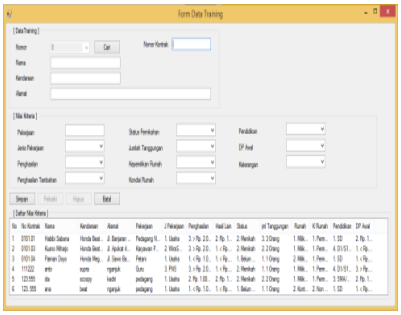
Button ini merupakan sebuah *button* untuk mengakses halaman *testing.*

c) *Button Keluar*

Button ini berfungsi untuk menutup program.

3. Tampilan *Training*

Training adalah *form* yang dapat diakses dari *form* beranda. *Form* ini digunakan sebagai akses untuk memproses data *training*. Tampilan *training* dapat dilihat pada gambar 1.6 Tampilan *Training* berikut.



Gambar 1.6 Tampilan *Training*

Dari gambar tampilan pada halaman *training* terdapat 5 buah *button*, 1 buah *list view*,16 buah *label*, 11 buah *combo box*, 3 buah *group box*, dan 5 buah *text box*. Fungsi dari masing – masing elemen tersebut adalah sebagai berikut.

a) *Button* Simpan

Button yang digunakan untuk memasukkan data pelanggan baru kemudian disimpan pada database.

b) *Button* Perbaiki

Button yang digunakan untuk melakukan perbaikan data pada data pelanggan lama.

c) *Button* Hapus

Button yang digunakan untuk menghapus data pelanggan lama.

d) *Button* Batal

Button yang digunakan untuk batal simpan data pelanggan lama.

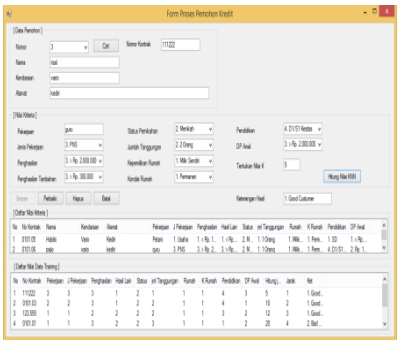
e) *Button* Cari

Button yang digunakan untuk mencari data pelanggan lama.

4. Tampilan Proses KNN

Proses KNN adalah *form* yang dapat diakses dari *form* beranda. *Form* ini digunakan sebagai input data pemohon kredit danproses perhitungan klasifikasi KNN. Tampilan

proses K-NNdapat dilihat pada gambar 1.7 Tampilan Proses KNN berikut.



Gambar 1.7 Gambar

Tampilan Proses KNN

Dari gambar tampilan pada halaman proses KNN terdapat 6 buah *button*, 2 buah *list view*, 16 buah *label*, 10 buah *combo box*, 4 buah *group box*, dan 6 buah *text box*.Fungsi dari masing – masing elemen tersebut adalah sebagai berikut.

a) *Button* Simpan

Button yang digunakan untuk memasukkan data pemohon kredit kemudian disimpan pada database.

b) *Button* Perbaiki

Button yang digunakan untuk melakukan perbaikan pada data pemohon kredit.

c) *Button* Hapus

Button yang digunakan untuk menghapus data pemohon kredit.

d) *Button* Batal

Button yang digunakan untuk batal simpan data pemohon kredit.

e) *Button* Cari

Button yang digunakan untuk mencari data pemohon kredit.

f) *Button Hitung KNN*

Button yang digunakan untuk proses hitung klasifikasi K-NN data pemohon kredit.

C. Skenario Uji Coba

Tabel 5.9 Akurasi KNN adalah tabel *confusion matrix* yang dihasilkan dengan menggunakan algoritma kNN. Perhitungan kedekatan kasus lama pada data training dengan kasus baru pada data testing, diketahui dari 30 data lapangan, 8 diklasifikasikan *bad cust* , 2 data diprediksi *bad cust* tetapi

ternyata *good cust*, 17 data kelas *good cust* diprediksi sesuai, dan 3 data diprediksi *good cust* ternyata *bad cust*. Tingkat akurasi penerapan algoritma kNN ini sebesar 83%.

Tabel 1.1 Alurasi KNN

	true bad	true good	class precission	Akurasi	eror rate
pred. Bad	8	2	0,8	83,333	0,16667
pred. Good	3	17	0,85		
class recall	0,72727	0,89474			

Dari hasil yang terdapat pada tabel 1.1 maka dapat disimpulkan bahwa hasil dari 30 data pelanggan dari PT. Adira Dinamika Multi Finance yang sudah selesai di proses klasifikasi adalah 25 data benar dan 5 data salah dalam perhitungan prediksi kolektibilitas sistem. Dari perhitungan diatas dinyatakan bahwa keberhasilan *K-Nearest Neighbor* dalam penelitian ini untuk melakukan klasifikasi data sudah valid karena semua data uji yang di tes sebagai data testing nilainya paling dekat dengan data training dengan tingkat akurasi 83.33%.

IV. PENUTUP

Berdasarkan hasil perancangan, implementasi, dan uji coba sistem yang telah dilakukan, dapat diambil beberapa kesimpulan dari hasil penelitian dan dapat disampaikan beberapa saran.

1. Kesimpulan

Beberapa kesimpulan yang dapat diambil dari penelitian ini adalah sebagai berikut.

a. Berdasarkan hasil pengujian, sistem bantu yang menggunakan metode *klasifikasi K-Nearest Neighbor* dapat digunakan sebagai salah satu alat sistem untuk membantu menentukan calon konsumen kredit *good cust* atau *bad cust*.

b. Dengan adanya sistem pendukung keputusan untuk menentukan kelayakan pemberian kredit sepeda motor bisa membantu dalam memberikan rekomendasi dan pertimbangan dalam pengambilan keputusan kredit berdasarkan kriteria yang telah ditentukan.

c. Data pelanggan kredit dan pemohon kredit berhasil

diimplementasikan kedalam metode *K-Nearest Neighbor* dengan mengubah data kriteria menjadi nilai bobot.

2. Saran

Dari proses penelitian yang dilakukan, ada beberapa saran yang bisa diterapkan guna membangun sistem lebih lanjut yaitu sebagai berikut.

a. Sistem masih terfokus pada implementasi metode perhitungan *K-Nearest Neihbor,* untuk penelitian selanjutnya desain interface yang perlu dikembangkan lagi karena masih terlalu sederhana.

2. Untuk memberikan hasil pertimbangan yang lebih baik dalam menganalisa penentuan resikokredit, hendaknya untuk pengembangan sistem selanjutnya agar dapat menambahkan kriteria pendukung lainnya agar mendapatkan hasil yang lebih akurat dalam pengambilan keputusan calon konsumen kredit

.

V. DAFTAR PUSTAKA

Larose, Daniel T. 2005. Discovering Knowledge in Data: An Introducing to Data Mining. John Willey & Sons, Inc.

Pramudiono, I. 2006. Apa itu Data Mining? Dalam http://datamining.japati.net/ egi-bin/indodm.cgi?bacaarsip&1155 527614&artikel". Diakses tanggal 16 Januari 2007.

Rivai, Veithzal.,Veithzal, Andria Permata. 2006. Credit Management Handbook. Jakarta : Raja GrafindoPersada.

Vercellis, Carlo. 2009. Business Intelligent: Data Mining and Optimization for Decision Making. Southern Gate, Chichester, West Sussex : John Willey & Sons, Ltd.

.