

ANALISA DAN PERANCANGAN APLIKASI DATA MINING PENENTUAN RESIKO KREDIT KEPEMILIKAN KENDARAAN BERMOTOR MENGUNAKAN ALGORITMA K-NEAREST NEIGHBOR

Fahrozi Zulfami

Program Sarjana Teknik Informatika, Universitas Putra Indonesia “YPTK”, Padang
Email: fahrozi.zulfahmi93@gmail.com

ABSTRAK

Adira Finance yang berada di kota Payakumbuh merupakan salah satu cabang dari perusahaan leasing adira finance Tbk yang merupakan perusahaan yang bergerak dalam membantu kredit suatu barang. Setiap harinya atau minggunya selalu ada calon nasabah yang ingin melakukan kredit kendaraan bermotor dengan menggunakan jasa perusahaan ini. Dengan banyaknya resiko kredit dalam bentuk macetnya atau ketidak lancarnya pembayaran kredit membuat perusahaan ini harus pandai dalam memilih konsumen yang akan dibantunya agar tidak membuat perusahaan mengalami kerugian. Banyaknya variabel yang digunakan dalam menerima calon konsumen baru diperusahaan ini seperti pekerjaan, pendidikan terakhir, usia, status pernikahan, penghasilan perbulannya dan pembayaran awal yang mungkin akan dilakukannya. Sehingga membuat perusahaan harus membuat keputusan yang tepat agar tidak terjadinya kesalahan dalam memilih calon konsumen dengan memanfaatkan variabel tadi, dengan menggunakan variabel yang sesuai dan tepat dapat menghasilkan kriteria yang tepat maka dibutuhkan rekomendasi yang tepat. Rekomendasi 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. Diperlukan penelitian untuk membuat suatu terobosan baru dengan memanfaatkan algoritma k-nearest neighbor dalam bentuk aplikasi data mining berbasis PHP, sehingga membantu dalam pengambilan keputusan dan penyeleksian calon konsumen.

Kata Kunci : resiko kredit, algoritma k-nearest neighbor, data mining, PHP.

1. PENDAHULUAN

Kredit merupakan suatu fasilitas keuangan yang memungkinkan seseorang atau badan usaha untuk meminjam uang untuk membeli produk dan membayarkannya kembali dalam jangka waktu yang ditentukan. Saat ini sistem kredit banyak digunakan untuk mendapatkan suatu barang seperti kendaraan transportasi.

Perusahaan kredit atau *leasing* memberikan jasa membantu masyarakat untuk mendapatkan kendaraan transportasi seperti sepeda motor, namun terdapat kendala-kendala yang diterima oleh perusahaan seperti terjadinya macet dalam pembayaran kredit yang mengakibatkan kerugian yang berdampak bagi perusahaan *leasing* yang menimbulkan 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. 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.

1.1 Rekayasa Perangkat Lunak

Rekayasa Perangkat Lunak (*software engineering*) merupakan pembangunan dengan menggunakan prinsip atau konsep rekayasa dengan tujuan menghasilkan perangkat lunak yang bernilai ekonomi yang dipercaya dan bekerja secara efisien menggunakan mesin (Rosa A. S dan M. Salahuddin 2013: 4). Proses perangkat lunak (*software process*) adalah sekumpulan aktifitas yang memiliki tujuan untuk mengembangkan atau mengubah perangkat lunak. Proses tersebut terdiri dari (Rosa, Shalahuddin : 2013):

a. Pengumpulan Spesifikasi (*Specification*)

Mengetahui apa saja yang harus dapat dikerjakan sistem perangkat lunak dan batasan pengembangan perangkat lunak.

- b. Pengembangan (*Development*)
Pengembangan perangkat lunak untuk menghasilkan sistem perangkat lunak.
- c. Validasi (*validation*)
Memeriksa apakah perangkat lunak sudah memenuhi keinginan pelanggan (*customer*).
- d. Evolusi (*Evolution*)
Mengubah perangkat lunak untuk memenuhi perubahan kebutuhan pelanggan (*customer*).

1.2 UML

Menurut Rosa dan Shalahudin (2013) Unified Modeling Language adalah sebuah standarisasi bahasa pemodelan untuk pembangunan perangkat lunak yang dibangun dengan menggunakan teknik pemrograman berorientasi objek.

Diagram UML memiliki beberapa pembagian kategori, berikut penjelasan singkat kategori tersebut adalah (Rosa dan Shalahuddin. 2013) :

- a. Use Case Diagram
Use case diagram merupakan pemodelan untuk kelakuan (*behavior*) sistem informasi yang akan dibuat. *Use case* mendeskripsikan sebuah interaksi antara satu atau lebih pengguna (aktor) dengan sistem informasi yang dibuat (Rosa dan Shalahuddin: 2013).
- b. Class Diagram
Menurut Rosa dan Shalahuddin (2013) *Class diagram* menggambarkan struktur sistem dari segi pendefinisian kelas-kelas yang akan dibuat untuk membangun sistem. Kelas memiliki apa yang disebut atribut dan metode atau operasi.
- c. Sequence Diagram
Sequence diagram menggambarkan kelakuan objek pada *use case* dengan mendeskripsikan waktu hidup objek dan *message* yang dikirimkan dan diterima antar objek (Rosa dan Shalahuddin. 2013).
- d. Collaboration Diagram
Collaboration diagram adalah bentuk lain dari *sequence diagram*. Bila *sequence diagram* diorganisir menurut waktu maka *collaboration diagram* diorganisir menurut ruang atau *space*. Berikut beberapa symbol yang sering digunakan.
- e. Activity Diagram
Menurut Rosa dan Shalahuddin (2013) *Activity diagram* menggambarkan *workflow* (aliran kerja) atau aktifitas dari sebuah sistem atau proses bisnis.
- f. Statechart Diagram
Statechart diagram atau diagram mesin yang digunakan untuk menggambarkan perubahan status atau transisi status dari sebuah mesin atau objek atau sistem (Rosa dan Shalahuddin:2013).
- g. Deployment Diagram
Deployment diagram menunjukkan konfigurasi komponen dalam proses *eksekusi aplikasi*.

1.3 Data Mining

Data Mining adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis (Fajar. 2013:3).

1.4 K-Nearest Neighbor

K-Nearest Neighbor 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* (Wu. 2009).

Algoritma k-NN bekerja berdasarkan jarak terpendek dari *query instance* ke *training data* untuk menentukan k-NN-nya. Salah satu cara untuk menghitung jarak dekat atau jauhnya tetangga menggunakan metode *euclidian distance*.

$$d_y = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2}$$

Keterangan :

X_{ik} = nilai X pada *training data*

X_{jk} = nilai X pada *testing data*

n = batas jumlah banyaknya data

1.5 PHP

PHP adalah pemrograman interpreter yaitu proses penerjemahan baris kode sumber menjadi kode mesin yang dimengerti komputer secara langsung pada saat baris kode berjalan (Alexander, 2013).

1.6 Database

Sistem basis data adalah sistem terkomputerisasi yang tujuan utamanya adalah memelihara data yang sudah diolah atau informasi dan membuat informasi tersedia saat dibutuhkan (Rosa dan Shalahahudin. 2013).

Wahana (2011) mengungkapkan : “ada beberapa istilah yang berada dalam *database*”, yaitu :

a. *Tabel*

Tabel merupakan kumpulan data yang disusun dalam bentuk baris (*record*) dan kolom (*field*) dimana tiap-tiap kolom mempunyai nama spesifik serta unik.

b. *Field*

Merupakan kolom pada tabel yang menggambarkan suatu atribut dari *record* yang menunjukkan suatu jenis data seperti kode, nama, alamat, dan lain sebagainya.

c. *Record*

Record merupakan kumpulan dari field yang menggambarkan tentang suatu unit data tertentu seperti data barang.

d. *Key*

Key merupakan *field* yang menjadi kunci pada operasi *tabel* seperti kode barang, kode pelanggan, kode transaksi, dan lain-lain.

DBMS merupakan hal yang wajar apabila satu *database* memiliki banyak *tabel* dimana antara satu *tabel* dengan *tabel* lain saling terhubung. “Ada tiga jenis hubungan antar *tabel* pada DBMS” (Wahana, 2011), yaitu :

a. *One To One*

One to one merupakan hubungan antara satu tabel dengan tabel lain dimana hanya ada satu baris yang dapat dihubungkan dengan tabel lain.

b. *One To Many*

One to many merupakan hubungan antara satu tabel dengan beberapa baris pada tabel yang lain.

c. *Many To Many*

Many to many merupakan hubungan dimana antara beberapa baris suatu tabel berhubungan dengan beberapa baris pada tabel lain.

1.7 MySQL

MySQL adalah sistem manajemen *database* relasi atau RDBMS (Relational Database Management System) yaitu aplikasi sistem yang menjalankan fungsi pengolahan data.

1.8 Kredit

Kata kredit berasal dari kata *credere* yang artinya adalah kepercayaan, maksudnya apabila seseorang memperoleh kredit, berarti mereka memperoleh kepercayaan. Sementara itu, bagi si pemberi kredit artinya memberikan kepercayaan kepada seseorang bahwa uang yang dipinjamkan pasti kembali (Sasuldyn. 2013).

Kredit bermasalah (*non performing loan*) adalah suatu keadaan dimana nasabah sudah tidak membayar sebagian atau seluruh kewajibannya kepada pihak yang telah diperjanjikannya (Glently. 2013).

2. METODE PENELITIAN

Dalam penelitian ini akan dicari nilai *ecludiean* dari hubungan data konsumen dan akan dihitung jarak setiap data konsumen sesuai dengan jarak antara data konsumen yang telah ditetapkan, dan nilai k sebagai batas dari perhitungan kelompok klasifikasi data tersebut.

Dalam setiap data ditentukan terlebih dahulu bobotnya dikarenakan data berupa variabel huruf bukan angka. Berikut pembobotan data setiap atribut :

Tabel 1. Penentuan Bobot

Atribut	Nilai Atribut	Bobot
Status Pernikahan	Menikah	0.5
	Belum	0

	Menikah	
	Duda/janda	1
Jumlah Tanggungan	0	0
	1	0.25
	2-3	0.5
	>3	1
Usia	<25	1
	25-50	0.5
	>50	0
Pendidikan	Tidak Sekolah	1
	SD	0.8
	SMP	0.6
	SMA	0.4
	DIII	0.2
	S1-S3	0
Kepemilikan Rumah	Milik Sendiri	0
	Orang Tua	0.25
	KPR	0.5
	Kontrak	1
Kondisi Rumah	Permanen	0
	Non Permanen	1
Jenis Pekerjaan	PNS	0
	Guru	0.4
	Wiraswasta	0.8
	Karyawan	1
Penghasilan	<2.000.000	1
	2 Jt-3.5 Jt	0.5
	>3.500.000	0
Pembayaran Pertama	>30%	0
	10%-30%	0.5
	<10%	1

Dari Tabel 1 dapat dilihat bahwa setiap atribut memiliki nilai atribut masing-masing dengan bobot yang telah ditentukan.

Pada perhitungannya menggunakan data baru yang ditentukan sendiri oleh penginputnya berikut salah satu contoh perhitungan yang ada pada penelitian ini.

Tabel 2. Contoh Perbandingan Data Perhitungan

Atribut	Data Training	Data Testing
X1	Belum Menikah	Belum Menikah
X2	0	0
X3	SMA	Diploma
X4	25-50	25-30
X5	Orang Tua	Orang Tua
X6	Permanen	Permanen
X7	Wiraswasta	PNS
X8	2jt – 3.5 Jt	>3.5 Jt
X9	10% - 30%	10% - 30%

$$\begin{aligned}
 &= \sqrt{(0 - 0)^2 + (0 - 0)^2 + (0.4 - 0.2)^2 + (0.5 - 0.5)^2 + (0.25 - 0.25)^2 + (0 - 0)^2 + (0.8 - 0)^2 + (0.5 - 0)^2 + (0.5 - 0.5)^2} \\
 &= \sqrt{(0 + 0 + 0.04 + 0 + 0 + 0 + 0.64 + 0.25 + 0)} \\
 &= \sqrt{(0.93)} = 0.964365076099
 \end{aligned}$$

Lakukan perhitungan yang sama kepada setiap data training yang ada dengan data testing yang sama untuk melihat perhitungannya secara manual.

Setelah melakukan setiap perhitungan didapat semua query distance kemudian seperti contoh Tabel 3 dbawah ini :

Tabel 3. Contoh hasil Query Distance setiap Data

N	Query Distance
1	1.361065759
2	0.964365076099
3	1.13578166916
4	1.68597153001
5	1.45773797371
6	1.18953772534
7	1.41862609591
8	1.71172427686
9	1.53378616502

Dari Query yang di dapatkan kemudian diurutkan sesuai dengannilai terkecil dan lihat hasil keputusan dengan meihat keputusan dominan yang terdapat pada data sebelumnya. Sebagai contoh terdapat pada Tabel 4 :

Tabel 4. Klasifikasi data testing

N	Query Distance	K yang digunakan (k=7)	Y
2	0.89	Ya	Bad Cust
6	1.107	Ya	Good Cust
3	1.14	Ya	Good Cust
7	1.38	Ya	Bad Cust
1	1.45	Ya	Good Cust
9	1.53	Ya	Bad Cust
5	1.639	Ya	Good Cust
4	1.643	-	Bad Cust
8	1.67	-	Bad Cust

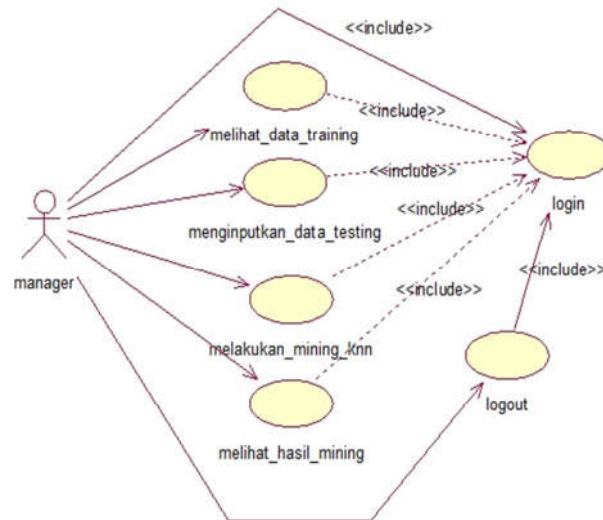
Pada Tabel 4 dapat dilihat terdapat:

Good Cust : 4

Bad Cust : 3

Hal ini dikarenakan K yang digunakan adalah 7 dan data training yang digunakan dalam pehitungan hanya 9.

Pada *use case* hanya terdapat 1 aktor yang menggunakan aplikasi ini yang akan melihat hasil proses perhitungannya



Gambar 1. Use Case Diagram

Dalam aplikasi yang dibuat nanti tidak dapat melakukan proses penginputan data training dikarenakan data yang digunakan telah ada pada perusahaannya.

3. HASIL DAN PEMBAHASAN

Aplikasi ini memiliki 3 menu utama yaitu data training, KNN, dan logout. Karena sebelum menggunakan aplikasi ini terlebih dahulu login dalam penggunaannya.



Gambar 2. Tampilan Awal

Pada menu data training menampilkan data training yang digunakan.

No.	Usia	Status	Tanggapan	Pendidikan	Pekerjaan	Kepemilikan Rumah	Kondisi Rumah	Penghasilan	Pembayaran Awal	Ramek
1	30	Manakah	Lebih 3 Anak	SI-SD	PKW	Milik Sendiri	Perumahan	Kecil Darat 2 Ji	<10%	Good Chat
2	25-30	Tidak Manakah	0	SMA	Wirawasta	Orang Tua	Perumahan	2 - 3,5 Ji	10%-30%	Bad Chat
3	25-30	Tidak Manakah	0	SI-SD	Karyawan	Orang Tua	Perumahan	Kecil Darat 2 Ji	<10%	Good Chat
4	25-30	Manakah	Lebih 3 Anak	Tidak Sekolah	Wirawasta	Milik Sendiri	Perumahan	2 - 3,5 Ji	10%-30%	Bad Chat
5	25-30	Tidak Manakah	1 Anak	Dipl	PKW	Milik Sendiri	Non-Perumahan	Kecil Darat 2 Ji	10%-30%	Good Chat
6	25-30	Tidak Manakah	1 Anak	SI-SD	PKW	Milik Sendiri	Perumahan	Kecil Darat 2 Ji	>30%	Good Chat
7	25-30	Manakah	2-3 Anak	SI-SD	Orang	Milik Sendiri	Non-Perumahan	Kecil Darat 2 Ji	<10%	Bad Chat
8	Kecil 25 The	Tidak Manakah	0	SMA	Wirawasta	Orang Tua	Non-Perumahan	Beas Darat 2 Ji	10%-30%	Bad Chat
9	25-30	Manakah	Lebih 3 Anak	SMA	Karyawan	Milik Sendiri	Perumahan	Kecil Darat 2 Ji	10%-30%	Bad Chat

Gambar 3. Tampilan Data Training

Menu KNN akan menyuguhkan penginputan dari data testing.



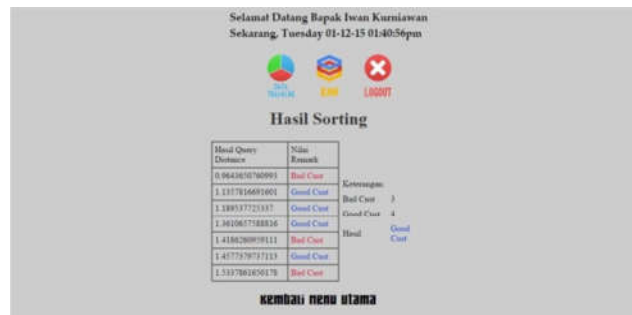
Gambar 4. Tampilan KNN

Setelah data testing di input kemudian di proses dan di sorting dan dilihat.



Gambar 5. Tampilan Proses

Setelah proses dapat dilihat hasil keputusannya sesuai dengan Gambar 7.



Gambar 6. Tampilan Hasil Keputusan

4. KESIMPULAN

Setelah melakukan perancangan sistem dan penerapan aplikasi ini dapat diambil beberapa kesimpulan, diantaranya:

1. Dengan menggunakan aplikasi ini berhasil membantu Perusahaan dalam melakukan proses mining terhadap data sebelumnya dan mengambil keputusan dengan tepat dalam menentukan konsumen baru dari data-data konsumen sebelumnya
2. Dengan menggunakan aplikasi ini dapat melihat pembayaran awal dan penghasilan dari pihak konsumen mempengaruhi dalam proses pembayaran kredit.
3. Algoritma K-Nearest Neighbor dapat digunakan dalam melihat resiko kredit yang ada namun dalam proses penginputan data testingnya harus berhati-hati dan sesuai dengan datanya.

5. DAFTAR PUSTAKA

- A.S Rosa & M.Shalahudin. 2013. *Rekayasa Perangkat Lunak Terstruktur Dan Berorientasi Objek*. Bandung : Modula.
- Fakhrurriqfi, Muhammad. dan Retantyo W. 2013. *Perbandingan Algoritma Nearest Neighbor, C4.5 dan LVQ untuk Klasifikasi Kemampuan Mahasiswa*. Jurnal Ilmu Komputer dan Elektronika FMIPA UGM Yogyakarta.
- Hermawati, Fajar Astti. 2013. *Data Mining*. Yogyakarta : Andi.
- Kusrini & Emha Taufiq Luthfi. 2009. *Algoritma Data Mining*. Yogyakarta : Andi Publisher.
- Krisandi, Nobertus. dkk. 2013. *Algoritma K-Nearest Neighbor Dalam Klasifikasi Data Hasil Produksi Kelapa Sawit Pada PT. Minamas Kecamatan Parindu*. Buletin Jurnal Ilmia Math, Stat, dan Terapan UNTAN.
- Leidiyana, Henny. 2013. *Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor*. Jurnal Pasca Sarjana Magister Ilmu Komputer STMIK Nusa Mandiri.
- Pato, Saduldyn. 2013. *Analisa Pemberian Kredit Mikro Pada Bank Syariah Mandiri Cabang Manado*. Jurnal Fakultas Ekonomi dan Bisnis FEKON UNSRAT Manado.
- Sibero, Alexander F. K. 2014. *Web Programming Power Pack*. Jakarta : Mediakom.
- Wahana. 2011. *Mastering CMS Programing With PHP & MySQL*. Yogyakarta : Andi Offset.