

IMPLEMENTASI ALGORITMA DATA MINING NAIVE BAYES PADA KOPERASI

Emerensye S. Y. Pandie

Jurusan Ilmu Komputer, Fakultas Sains dan Teknik, Universitas Nusa Cendana

INTISARI

Salah satu faktor kegagalan dalam bidang usaha perkreditan yaitu kurang akuratnya penilaian terhadap kemampuan debitur, sehingga mengakibatkan kesalahan dalam keputusan pemberian kredit yang berujung pada kemacetan kredit. Teknik *data mining* dapat digunakan untuk menilai kemampuan nasabah berdasarkan data-data masa lalu. Data debitur yang telah melalui tahapan data mining selanjutnya diproses menggunakan algoritma *data mining Naive Bayes*. *Naive Bayes* merupakan teknik prediksi berbasis probabilistik sederhana berbasis pada penerapan aturan bayes. Implementasi menggunakan Weka 3.8 dengan jumlah data 3018 *record* menghasilkan tingkat kebenaran sebesar 94 %.

Kata kunci: *credit, data mining, naive bayes, cross validation*

ABSTRACT

One of the factors of failure in the field of credit business is the lack of accurate assessment of the ability of the debtor, thus resulting in errors in credit decisions that culminate in credit congestion. Data mining techniques can be used to assess customer ability based on past data. Debtor data that has been through the stages of data mining is then processed using Naive Bayes data mining algorithm. Naive Bayes is a simple probabilistic based prediction technique based on the application of bayes rules. Implementation using Weka 3.8 with a total of 3018 records yields a truth level of 94%.

Keywords: *credit, data mining, naive bayes, cross validation*

I. PENDAHULUAN

Sebagai salah satu usaha bank yakni memberikan kredit, bank akan selalu memperhatikan resiko kredit yang terjadi dalam hal pemberian kredit, seperti nasabah yang tidak mampu membayar pinjaman yang telah diterima. Untuk mengurangi resiko yang terjadi pihak bank menganggap perlu mengambil suatu kebijakan yaitu dengan memperkuat pengawasan bank terhadap nasabah untuk mengatasi resiko tersebut.

Didalam kegiatan perkreditan sering terjadi masalah kredit macet yang disebabkan oleh gagalnya pengembalian sebagian pinjaman yang diberikan kepada para peminjam. Masalah ini dapat diatasi, salah satunya dengan mengidentifikasi dan memprediksi nasabah dengan baik sebelum memberikan pinjaman dengan cara memperhatikan data historis pinjaman. Analisa kredit dapat menggunakan salah satu jenis algoritma *data mining* yaitu *Support Vector Machine* [6].

Teknologi *data mining* memiliki peranan penting dalam dunia bisnis yang semakin kompetitif. *Data mining* mampu memanfaatkan aset penting perusahaan yaitu data bisnis yang jumlahnya sangat besar sehingga dapat menghasilkan informasi yang belum pernah terpelajari sebelumnya. Bentuk penerapan *data mining* antara lain dengan membuat aplikasi yang memanfaatkan teknologi *data mining* menggunakan algoritma C4.5 untuk klasifikasi profil kredit nasabah perbankan.

Dalam mengantisipasi kredit macet karena meningkatnya jumlah pengajuan kredit pada PT. Telkom Kandatel Surabaya Timur maka dilakukan analisa dan pola kredit dari angsuran pinjaman menggunakan metode *k-Nearest Neighbour* (k-NN). Output dari aplikasi digunakan untuk melakukan *reminding call*, *rescheduling* dan *reconditioning* kepada nasabah [9].

Industri kredit berkaitan dengan banyak masalah yang menarik bagi komunitas komputasi. Permasalahan kredit dapat dianalisa menggunakan dua teknik analisa yaitu *Neural Networks* (NNS) dan *Genetic Algorithms* (GAs) [1].

Koperasi merupakan salah satu lembaga keuangan yang melakukan transaksi kredit. Dalam rangka mengembangkan usahanya dalam bentuk pinjaman anggota, koperasi memiliki prinsip kehati-hatian sebagaimana yang diterapkan pada perbankan dengan melakukan analisa pemberian kredit kepada calon anggotanya. Peran koperasi dalam perekonomian dapat dilihat dari data Dinas Koperasi UKM Provinsi Jawa Tengah yang dimuat dalam harian Suara Merdeka edisi 30 juni 2011, pada tahun 2011 terdapat 5.809 (20%) koperasi tidak aktif yang masuk dalam daftar yang akan dihapus dari *database*. Koperasi di Jawa Tengah dengan aset mencapai Rp 19,3 trilyun/tahun dan omset Rp 25,9 trilyun menjadi penyangga kekuatan ekonomi sebesar 58%.

Berdasarkan berbagai penelitian yang telah dilakukan oleh beberapa peneliti sebelumnya mengenai analisa kredit menggunakan *data mining*, maka yang menjadi permasalahan adalah bagaimana menemukan pola kredit dari *database* koperasi untuk menganalisa pengajuan kredit kembali nasabah lama dengan mempertimbangkan *track record* dari pinjaman sebelumnya menggunakan algoritma k-NN.

II. MATERI DAN METODE

2.1 Bahan Penelitian

Penelitian ini menggunakan sumber data dari koperasi simpan pinjam sumber rejeki. Dari keseluruhan data diambil data nasabah yang melakukan pinjaman lebih dari satu kali sebagai bahan penelitian.

2.2 Proses *Cleaning Data*

Dari keseluruhan data yang ada, setelah dilakukan pertimbangan menggunakan dasar tingkat keterikatan antar hubungan data, maka dipilih tabel anggota, tabel aplikasi, tabel angsuran, tabel pinjam, tabel jaminan dan tabel plafon sebagai data yang akan di *cleaning* sebagai sumber data penelitian. Proses *cleaning* data diperlukan untuk memilih atribut mana saja dari keseluruhan data yang akan digunakan dalam membentuk tabel pola kredit. Pengkategorian kredit digunakan sebagai atribut target dari proses klasifikasi kredit. Berdasarkan SK Direksi Bank Indonesia NO.31/147/KEP/DIR tanggal 12 Nopember 1998, kategori kredit dan rentang keterlambatan didefinisikan sebagai Lancar, jika total jumlah keterlambatan angsuran sama dengan 0, dalam perhatian khusus, 1 s.d 90 hari, kurang lancar, 91 s.d. 180 hari, diragukan, 182 s.d 270 hari dan macet, diatas 271 hari keterlambatan.

Pada praktek di koperasi, hanya dikenal 3 kategori kredit, lancar, tersendat dan macet. Dengan toleransi sampai dengan 3 hari keterlambatan maka total keterlambatan 0 sampai dengan 36 hari masih dikategorikan lancar. Angka keterlambatan sampai dengan 36 hari berdasarkan kategori Bank Indonesia ada pada kategori dalam perhatian khusus, oleh karena itu kategori 0–90 hari dikonversi menjadi kategori lancar. Sementara angka keterlambatan lebih 180 hari oleh BI dikategorikan diragukan dan macet, oleh koperasi dengan pertimbangan permodalan rentang tersebut dikategorikan macet, maka sisa rentang waktu diantaranya dikategorikan sebagai tersendat.

Untuk mengkategorikan kredit dengan jangka waktu selain 12 bulan, dilakukan dengan cara menghitung persentase jangka waktu keterlambatan terhadap rentang waktu kredit. Dari hasil perhitungan persentase tersebut, digunakan untuk menghitung rentang waktu kategori kredit pada jangka waktu selain 12 bulan.

Proses *cleaning* juga diterapkan pada atribut yang memiliki sebaran data yang beragam seperti alamat, no telepon, dan nama juga terhadap atribut yang sebagian besar datanya tidak diisi tidak diikutsertakan dalam data pola. Demikian juga dengan data yang sudah terwakili oleh atribut tertentu bisa dihapus seperti atribut merk, jenis barang jaminan, tahun barang jaminan sudah terwakili oleh atribut nilai jaminan sehingga atribut-atribut tersebut bisa langsung dihapus.

Hasil analisa atribut dalam proses *data mining* dari database koperasi menghasilkan 8 atribut yang akan digunakan sebagai data pola / rumusan dalam prediksi kategori kredit dan jumlah data 3018 *record*. 8 atribut terpilih beserta tipe datanya adalah 1. Plafon, numerik. 2. Niljamin, numerik, 3. Miliksd, biner, 4. Stnikah, biner, 5. Jangka, numerik, 6. Telat, numerik, 7.bunga, biner, 8.keterangan, numerik. Data pola kredit dapat dilihat pada gambar 1.

No.	1: telat	2: niljamin	3: miliksd	4: bunga	5: status	6: plafonusul	7: jangk.
	Numeric	Numeric	Nominal	Nominal	Numeric	Numeric	Numeric
2859	93.0	3750000.0	FALSE	turun	1.0	2000000.0	6
2860	1.0	3750000.0	FALSE	turun	1.0	2000000.0	6
2861	2.0	4250000.0	TRUE	flat	1.0	1000000.0	10
2862	124.0	1250000.0	FALSE	flat	1.0	2000000.0	10
2863	4.0	5000000.0	FALSE	flat	1.0	1500000.0	10
2864	3.0	3000000.0	TRUE	turun	1.0	1500000.0	6
2865	31.0	2500000.0	FALSE	turun	1.0	2000000.0	6
2866	93.0	3000000.0	FALSE	flat	1.0	3000000.0	12
2867	9.0	5000000.0	FALSE	turun	1.0	4500000.0	6
2868	93.0	2000000.0	FALSE	flat	1.0	2000000.0	12
2869	62.0	1750000.0	FALSE	flat	1.0	1250000.0	12
2870	62.0	4250000.0	FALSE	flat	1.0	4000000.0	18
2871	2.0	3500000.0	FALSE	flat	1.0	3000000.0	12
2872	2.0	3500000.0	FALSE	flat	1.0	3000000.0	12
2873	11.0	2000000.0	FALSE	turun	1.0	2000000.0	6
2874	8.0	1500000.0	FALSE	turun	1.0	1500000.0	6
2875	4.0	2000000.0	TRUE	flat	1.0	2000000.0	10
2876	62.0	3250000.0	TRUE	flat	1.0	2000000.0	12
2877	5.0	4000000.0	FALSE	turun	1.0	2500000.0	6
2878	0.0	3500000.0	FALSE	flat	1.0	1500000.0	10
2879	93.0	2000000.0	FALSE	flat	1.0	2000000.0	12
2880	93.0	3250000.0	FALSE	turun	1.0	3000000.0	6
2881	124.0	2250000.0	FALSE	turun	1.0	2000000.0	6

Gambar 1. Data Pola Kredit

2.3 Konversi Nilai

Penelitian ini menggunakan Algoritma *Naive Bayes*. Tipe data nominal sebelumnya harus dirubah dahulu ke dalam nilai diskrit. Atribut nominal bernilai biner yaitu miliksd, bunga dan status dirubah kedalam pilihan biner 0 atau 1, seperti terlihat pada gambar 2.

Atribut	Instance	Nilai diskrit
Kategori	lancar	1
	tersendat	2
	macet	3
status	menikah	1
	tidak menikah	0
bunga	flat	1
	menurun	0
miliksd	TRUE	1
	FALSE	0

Gambar 2. Konversi Nilai

2.4 Algoritma Naive Bayes

Naive bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi keanggotaan suatu *class*. *Naive bayes* didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*. *Naive bayes* memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar.

Prediksi Bayes didasarkan pada formula teorema Bayes dengan formula umum sebagai berikut:

Teorema Bayes memiliki bentuk umum sbb :

$$P(H|X) = \frac{P(X|H) P(H)}{P(X)}$$

Keterangan :

X : data dengan *class* yang belum diketahui

H : hipotesis data X merupakan suatu *class* spesifik

P(H|X) : probabilitas hipotesis H berdasar kondisi X
(*posteriori probability*)

P(H) : probabilitas hipotesis H (*prior probability*)

P(X|H) : probabilitas X berdasar kondisi hipotesis H

P(X) : probabilitas dari X

Persamaan (1)

2.5 Standarisasi Data

Kendala utama dalam menghitung jarak secara langsung dari data adalah dalam kasus di mana data memiliki rentang nilai pengukuran yang jauh berbeda sehingga penghitungan jarak menjadi tidak efektif. Misalnya, jika atribut pendapatan adalah pendapatan tahunan dalam rupiah, dan atribut umur adalah umur seseorang dalam tahun, maka pendapatan akan memiliki pengaruh yang jauh lebih tinggi dalam Setelah proses standarisasi maka bisa dilanjutkan dengan proses perhitungan Naive Bayes.

Data pola kredit distandarisasi dalam rentang nilai [0..1] menggunakan persamaan (2) ukuran jarak dibanding umur, sehingga pengukuran menjadi tidak relevan lagi. Salah satu solusinya adalah digunakan standarisasi nilai dengan rentang [0..1] menggunakan rumusan standarisasi *min-max* seperti terlihat pada persamaan (2).

$$X_s = \frac{X - Min}{Max - Min} \dots \dots \dots (2)$$

Nilai X_s didefinisikan sebagai nilai setelah di standarisasi dimana X merupakan nilai *instance* atribut dan Min merupakan nilai minimum dalam atribut sebaliknya Max merupakan nilai maksimum dalam atribut. Data pola kredit terstandarisasi dapat dilihat pada gambar 3.

PLAFON	JAMINAN	MILIK	NIKAH	JANGKA	TELAT	BUNGA	keterangan
0,12	0,27272727	1	0	0,33333333	0,35361217	1	2
0,08	0,27272727	0	0	0,33333333	0,23574144	1	1
0,14	0,18181818	1	0	0,33333333	0,35361217	1	2
0,14	0,27272727	0	0	0	0,11787072	0	1
0,24	0,45454545	1	0	0,33333333	0,35361217	1	2
0,06	0,15454545	0	0	0,33333333	0,47148289	1	2
0,08	0,09090909	0	0	0,33333333	0,35361217	1	2
0,12	0,18181818	0	0	0,33333333	0,35361217	1	2
0,12	0,27272727	0	0	0,33333333	0,34980989	1	2
0,12	0,22727273	0	0	0,33333333	0,23574144	1	1
0,16	0,31818182	0	0	0,33333333	0,35361217	1	2
0,12	0,27272727	0	0	0,33333333	0,47148289	1	2
0,16	0,31818182	0	0	0	0,23574144	0	2
0,14	0,31818182	0	0	0,33333333	0,35361217	1	2
0,18	0,45454545	0	0	0,33333333	0,35361217	1	2
0,12	0,27272727	0	0	0,33333333	0,35361217	1	2
0,08	0,27272727	1	0	0,33333333	0,35361217	1	2
0,12	0,32727273	1	0	0	0,03041825	0	1

Gambar 3. Data Pola Kredit Terstandarisasi

2.6 K-fold cross validation

K-fold cross validation diawali dengan membagi data sejumlah *k-fold* yang diinginkan. Dalam proses *cross validation* data akan dibagi dalam *k* buah partisi dengan ukuran yang sama $D_1, D_2, D_3, \dots, D_n$ selanjutnya proses testing dan *training* dilakukan sebanyak *k* kali. Dalam iterasi ke-*i* partisi D_i akan menjadi data *testing* dan sisanya akan menjadi data *training*. Persamaan (1) diterapkan terhadap setiap *record* dalam data testing terhadap seluruh *record* pada data *training*. Pada akhir perhitungan dihitung persentase tingkat kebenaran klasifikasi sehingga diperoleh gambaran validitas data training sebagai data pola.

III. HASIL DAN PEMBAHASAN

3.1 Hasil

Hasil 10-fold *cross validation* untuk uji validitas menggunakan algoritma *Naive Bayes* menunjukkan hasil seperti terlihat pada gambar 3.

Correctly Classified Instances	2955	94.7723 %
Incorrectly Classified Instances	163	5.2277 %
Kappa statistic	0.8996	
Mean absolute error	0.0752	
Root mean squared error	0.1912	
Relative absolute error	21.5592 %	
Root relative squared error	45.7873 %	
Total Number of Instances	3118	

Atribut yang digunakan sebanyak 8 atribut yaitu total telat, nilai jaminan, status kepemilikan, jenis bunga, status perkawinan, status pernikahan, jumlah plafon diusulkan, jangka waktu kredit dengan jumlah *record* sebanyak 3018 *record*. Persentase tingkat kebenaran klasifikasi dengan algoritma *Naive Bayes* menghasilkan nilai lebih dari 94%.

IV. PENUTUP

4.1 Kesimpulan

Berdasarkan hasil pengujian maka dapat disimpulkan:

1. Algoritma *Naive Bayes* dapat digunakan untuk klasifikasi kategori kredit nasabah yang hendak mengajukan pinjaman koperasi karena memiliki tingkat kebenaran klasifikasi

sebesar 94,77%.

2. Jumlah data pola sebanyak 3018 cukup untuk dijadikan data pola karena memiliki kombinasi data kategori Tersendat sebanyak 1058, Kategori Lancar sebanyak 1865 dan sisanya berupa data Kategori Macet sebanyak 95.

4.2 Saran

Diperlukan adanya penelitian lanjutan menggunakan algoritma yang berbeda untuk membandingkan kebenaran tingkat klasifikasi data pola.

DAFTAR PUSTAKA

- [1] Chen, M.C., Huang, S.H., 2003. *Credit scoring and rejected instances reassigning through evolutionary computation techniques*. *Jurnal Expert System with Application* 24 (4), 433-441
- [2] Ye, Ruisong dan Gou, Weichuang., 2013, *A Chaos-based Image Encryption Scheme Using Mulltimodal Skew Tent Map*, <http://www.cisjournal.org/journalofcomputing/archive/vol4no10/vol4no10>, 10 Oktober 2013, diakses tanggal 23 Mei 2014.
- [2] Han, J.K., 2001. *Data mining: Concept and Technique*. San Fransisco: Morgan Kaufmann Publisher
- [3] Kohavi, R., 1995. *A Study of Cross Validation and Bootstrap for Accuracy Estimation and Model Selection*, *Lecture Note in Computer Science* 6440, 114-124
- [4] Moertini, S.V., 2003. *Towards the use of C4.5 Algorithm for classifying Banking Dataset*. *Jurnal Integral* 8 (2), 105-116
- [5] Mohd, S.S., Rayner, A., 2010. *Advanced Data mining and Applications 6th International Conference*, ADMA 2010, Chongqing, China, November 19-21, 2010, Proceedings, Part I
- [6] Oktrivianto, R., 2008. *Aplikasi Data mining untuk Memprediksi Kelas Resiko Pemberian Kredit menggunakan Support Vector Machine (SVM)*. Tesis. Surabaya: Institut Teknologi Surabaya.
- [7] Sucahyo, Y.G., 2003. *Data mining: Menggali Informasi yang Terpendam*. *Website: <http://www.ilmukomputer.com>*. Diakses tanggal 12 Januari 2011.
- [8] Sujarwanto., 2011, 5809 Koperasi di Jawa Tengah Tidak Aktif. *Harian Suara Merdeka* tanggal 30 Juni 2011
- [9] Sunaryanto, D., 2009. *Penerapan K-Nearest Neighbour untuk Prediksi Resiko Kredit di PT. Telkom Kandatel Surabaya Timur*. Tesis. Surabaya: Sekolah Tinggi Ilmu Komputer (STIKOM).