

PENERAPAN NAÏVE BAYES UNTUK PREDIKSI KELAYAKAN KREDIT

Dedy Ahmad Kurniawan¹⁾, Danny Kriestanto²⁾

Teknik Informatika, STMIK AKAKOM

e-mail: dedyahmad.akakom10@gmail.com¹⁾, danny@akakom.ac.id²⁾

ABSTRAK

Dalam dunia perbankan, pemberian kredit kepada nasabah adalah kegiatan rutinyang mempunyai resiko tinggi. Dalam pelaksanaannya, kredit yang bermasalah (kredit macet) sering terjadi akibat analisis kredit yang tidak hati-hati atau kurang cermat dalam proses pemberian kredit, maupun dari karakter nasabah yang tidak baik. Untuk mencegah terjadinya kredit macet, diperlukan adanya peramalan akurat yang salah satunya menggunakan teknologi di bidang data mining.

Dengan menggunakan teknologi di bidang data mining yang mengoptimasi proses pencarian informasi prediksi dalam basis data yang besar, serta menemukan pola-pola yang tidak diketahui sebelumnya. Naïve Bayes memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya dengan mempelajari korelasi hipotesis yang merupakan label kelas yang menjadi target pemetaan dalam klasifikasi dan evidence yang merupakan fitur-fitur yang menjadi masukan dalam model klasifikasi.

Pengolahan data berbasis data mining tersebut diharapkan dapat digunakan sebagai alat bantu dalam memprediksikan kelayakan kredit yang diperkirakan layak atau tidaknya pemohon atau nasabah untuk diberikan kredit.

Kata kunci: Data Mining, Naïve Bayes, Prediksi kelayakan kredit

ABSTRACT

In banking business, lending a money is a routine that consists of high risks. By giving a lend to a person, a bank or an organization that have authority to do that can lead to a bad credit because of a bad or recklessness in data analysis. In order to prevent these things to occur, a good forecasting tool is needed.

The using of data mining techniques in searching process of a huge amount of data to extract information can be done quickly in order to find a good prediction of money lending, and also to find unusual patterns that cannot be seen before. Naïve Bayes can predict probability based on previous experiences (data) by studying hypothesis as the target of the classification and evidence as input features in classification model.

The implemented data mining application can be used as a tool to predict the creditworthiness of a prospective borrower.

Keywords: Data Mining, Naïve Bayes, Creditworthiness prediction

I. PENDAHULUAN

Dalam dunia perbankan pemberian kredit kepada nasabah adalah kegiatan rutin yang mempunyai resiko tinggi. Dalam pelaksanaannya, kredit yang bermasalah (kredit macet) sering terjadi akibat analisis kredit yang tidak hati-hati atau kurang cermat dalam proses pemberian kredit, maupun dari karakter nasabah yang tidak baik. Untuk mencegah terjadinya kredit macet, seorang analisis kredit perbankan harus mampu mengambil keputusan yang tepat untuk menerima ataupun menolak pengajuan kredit. Untuk mengetahui kelayakan kredit di masa mendatang, diperlukan adanya peramalan yang akurat yang salah satunya menggunakan teknologi di bidang data mining.

Dengan menggunakan teknologi di bidang data mining yang mengoptimasi proses pencarian informasi prediksi dalam basis data yang besar, serta menemukan pola-pola yang tidak diketahui sebelumnya. Maka identifikasi pola data dari sistem penentuan pemberian kredit dapat dilakukan dengan pendekatan probabilitas bersyarat. Berdasarkan kemampuan belajar yang dimilikinya, maka data mining dapat dioptimasi untuk menemukan pola dimasa lalu dan mencari suatu fungsi yang akan menghubungkan pola data masa lalu dengan keluaran yang diinginkan. Naïve Bayes memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Dengan mempelajari korelasi hipotesis yang merupakan label kelas yang menjadi target pemetaan dalam klasifikasi, dan evidence yang merupakan fitur-fitur yang menjadi masukan dalam model klasifikasi.

Data mining adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (machine learning) untuk menganalisa dan mengekstraksi pengetahuan (knowledge) secara otomatis [1].

Menurut [2], data mining adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan pola atau hubungan dalam set data yang berukuran besar.

Dengan adanya aplikasi pengolahan data berbasis data mining tersebut diharapkan dapat digunakan sebagai alternatif dan alat bantu dalam memprediksikan resiko kelayakan kredit yang diperkirakan layak atau tidaknya pemohon atau nasabah untuk diberikan kredit, yang dalam penelitian ini digunakan BMT Beringharjo, Yogyakarta sebagai studi kasusnya.

Adapun perbedaan tinjauan pustaka yang dijadikan sebagai acuan dalam penelitian ini adalah seperti yang tampak pada Tabel I.

TABEL I
PERBANDINGAN TINJAUAN PUSTAKA

Referensi	Objek	Masukan	Metode	Keluaran
[3]	Klasifikasi premi asuransi	Data Nasabah	Naïve Bayes	Lancar, kurang lancar, atau tidak lancar
[4]	Koperasi Kara Audita Perwakilan BPKP Yogyakarta	Data Nasabah	Naïve Bayes	Diterima, dipertimbangkan, atau ditolak
[5]	BMT Beringharjo, Yogyakarta	Data historis nasabah yang telah diakui kelayakannya	Back Propagation	Diterima atau ditolak
Yang diajukan	BMT Beringharjo, Yogyakarta	Data Nasabah	Naïve Bayes	Layak atau Tidak Layak

II. METODE

A. Metode Penelitian

Adapun metode penelitian yang diterapkan secara berurutan adalah sebagai berikut:

1) Pengumpulan Data

Data penelitian diperoleh dengan melakukan survei dan juga meminta langsung kepada pihak pengelola BMT Beringharjo Yogyakarta.

2) Analisis Kasus

Analisis dilakukan dengan membandingkan hasil pertimbangan pihak BMT Beringharjo seperti ditunjukkan pada Tabel II, kemudian diterapkan dalam aplikasi dengan menggunakan Naïve Bayes.

3) Perancangan Aplikasi

Adapun perancangan aplikasi ini akan dibuat berbasis Java dengan bantuan perangkat lunak Netbeans. Oleh karena itu, desain perancangan dari aplikasi ini akan menggunakan UML.

4) Pengujian Aplikasi

Pengujian dilakukan dengan membandingkan hasil perhitungan manual dengan hasil perhitungan aplikasi dan dilihat persentase tingkat kesalahannya.

B. Kebutuhan Input

Data masukan yang dibutuhkan dalam sistem ini adalah data pengajuan kredit yaitu jaminan, total penghasilan, pinjaman lain, total pengeluaran, nominal kredit, status rumah, dan tempo pengembalian. Data pengajuan kredit ini nantinya akan dijadikan data untuk proses *training*, dimana data tersebut sudah diklasifikasikan sebelumnya sesuai dengan kelayakan kreditnya yaitu layak maupun tidak layak. Adapun sampel data *training* yang digunakan dapat dilihat pada Tabel II.

C. Kebutuhan Output

Kebutuhan keluaran atau *output* yang dihasilkan dari sistem ini adalah hasil prediksi data melalui proses *training* dan *testing* yang layak maupun tidak layak. Proses *testing* data dilakukan setelah adanya hasil dari proses *training* data untuk menentukan nilai *posterior probability* terhadap data baru yang diujikan.

Nilai tersebut merupakan nilai dari probabilitas akhir dari data yang diujikan sesuai dengan kelasnya yang masuk kedalam kelas layak maupun tidak layak. Dari data tersebut nantinya akan dibandingkan hasil probabilitas akhir antara kelas layak dan tidak layak. Nilai probabilitas akhir yang lebih besar akan dijadikan sebagai nilai akhir sekaligus sebagai keputusan apakah pengajuan kredit layak ataupun tidak layak.

D. Perancangan Sistem

Pemodelan yang digunakan untuk membangun sistem ini adalah pemodelan dengan menggunakan UML (*Unified Modelling Language*) yang meliputi antara lain *usecase diagram*, *activity diagram*, dan *class diagram*.

1) Usecase Diagram

Usecase diagram yang digunakan dalam aplikasi ini dapat dilihat pada Gambar 1.

2) Activity Diagram Training

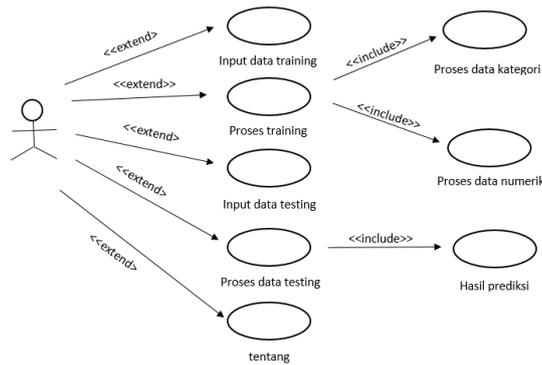
Activity diagram training ini merupakan gambaran dari sistem melakukan proses *training* terhadap data. Adapun *Activity diagram training* dapat dilihat pada Gambar 2.

3) Activity Diagram Testing

Activity diagram testing ini merupakan gambaran dari sistem melakukan proses *testing* terhadap data. Adapun perancangan *Activity diagram testing* dapat dilihat pada Gambar 3.

TABEL II
SAMPEL DATA TRAINING

No	Jaminan	Total Penghasilan	Pinjaman Lain	Total Pengeluaran	Nominal Kredit	Status Rumah	Tempo Pengembalian	Kelayakan Kredit
1	Spd mtr	5.000.000	500.000	1.500.000	3.000.000	Kontrak	3	Layak
2	Tab	-	-	1.070.000	2.500.000	Mp	3	Layak
3	Tab	-	-	215.000	700.000	Mp	3	Layak
4	Mobil	-	-	3.000.000	10.000.000	Mp	24	Layak
5	Spd mtr	-	-	1.530.000	5.000.000	Kontrak	12	Layak
6	Shm	-	-	7.107.500	140.000.000	Kontrak	24	Tidak layak
7	Spd mtr	-	-	500.000	10.000.000	Kontrak	12	Tidak layak
8	Mobil	500.000	500.000	2.500.000	25.000.000	Mp	6	Tidak layak
9	Shm	200.000	200.000	750.000	20.000.000	Mp	6	Tidak layak
10	Spd mtr	1.500.000	1.500.000	1.000.000	45.000.000	Mp	6	Tidak layak



Gambar 1. Use Case Diagram

4) Activity Diagram Naïve Bayes

Gambar 4 merupakan gambaran dari keseluruhan proses yang dilakukan oleh metode Naïve Bayes. Proses ini meliputi dari awal perhitungan yang dilakukan ketika proses *training* data hingga proses prediksi data yang baru (data *testing*). Jika data berupa data kategori, maka dapat langsung dihitung probabilitas awalnya. Jika data berupa numerik maka akan dihitung terlebih dahulu *mean*, *standar deviasi* dan *variansi* dari tiap parameter data numerik. Setelah semua data dihitung nilai *mean*, *standar deviasi* dan *variansi* menggunakan rumus diatas, maka selanjutnya akan dihitung nilai probabilitas awal menggunakan rumus distribusi Gaussian. Dari data hasil perhitungan data yang berupa kategori dan *numeric* maka akan dijumlahkan berdasarkan kelas kelayakannya (layak atau tidak layak) untuk mendapatkan probailitas akhir dari setiap kelas. Kemudian dari nilai probabilitas akhir tiap kelas tersebut akan dibandingkan dari keduanya. Nilai probabilitas akhir yang lebih besar akan menjadi hasil akhir dari prediksi. *Activity diagram* yang diimplementasikan dalam penelitian ini dapat dilihat pada Gambar 4.

5) Class Diagram

Implementasi aplikasi ini menggunakan *class diagram* seperti tampak pada Gambar 5.

III. HASIL

Masukan data *training* dilakukan di borang data *training*, dan hasilnya tampak seperti pada Gambar 6. Borang data *training* difungsikan untuk memasukkan data-data yang telah ada sebelumnya. Hal penting yang dimasukkan pada borang ini adalah jaminan, penghasilan, nominal kredit yang diinginkan, status kepemilikan rumah, dan kelayakan. Tampilan yang sama juga dilakukan terhadap borang data *testing*, seperti yang tampak pada Gambar 7, dengan tambahan nilai probabilitas, dan hasil prediksi berupa penentuan layak atau tidak-nya kredit yang diajukan untuk diterima.

Proses pengujian pertama kali dilakukan terhadap 40 data dan diperoleh hasil 37 data yang diprediksi benar (sesuai dengan metode manual) dan 3 data yang diprediksi salah. Kemudian pada pengujian terhadap 200 data, dengan 160 data *testing* dan 40 data *training* diperoleh tingkat akurasi aplikasi dapat mencapai 92.5% dengan nilai error sebesar 7,5%.

IV. PEMBAHASAN

Berdasarkan hasil penelitian, ada perbedaan yang cukup signifikan terjadi ketika jumlah data yang diujikan cukup banyak. Dari hasil uji 40 data ditambah hingga menjadi 200 data peningkatan akurasi bertambah tidak signifikan. Namun, apabila dibandingkan dengan hasil [4] dari sisi jumlah kriteria yang digunakan, tingkat akurasi dari aplikasi yang dibangun meningkat dengan signifikan dari 70% menjadi 92,5% [4]. Adapun data yang

digunakan oleh [4] adalah sebanyak 250 data yang dibagi atas 170 data *training* dan 80 data *testing*. Dari kedua hasil ini dapat disimpulkan bahwa tingkat akurasi aplikasi akan meningkat seiring dengan semakin banyaknya kriteria yang digunakan.

V. SIMPULAN DAN SARAN

Berdasarkan hasil pengujian yang telah dilakukan, maka dapat disimpulkan sebagai berikut:

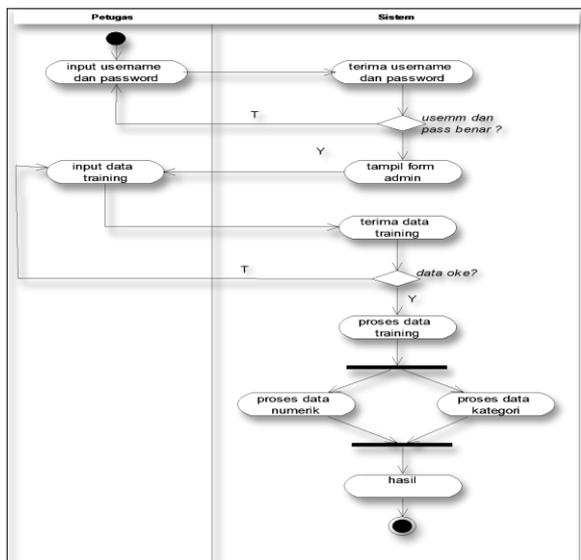
- 1) Penerapan algoritma *data mining* menggunakan Naïve Bayes dapat dilakukan untuk memprediksi kelayakan kredit pada BMT Beringharjo Yogyakarta.
- 2) Akurasi data yang diperoleh dalam penelitian ini adalah sebesar 92,5% dengan error sebesar 7,5% dari 160 data yang digunakan untuk *training* dan 40 data untuk *testing*.
- 3) Semakin banyak kriteria yang digunakan dalam proses prediksi maka akurasi data akan semakin tinggi.

Adapun saran untuk penelitian selanjutnya adalah sebagai berikut:

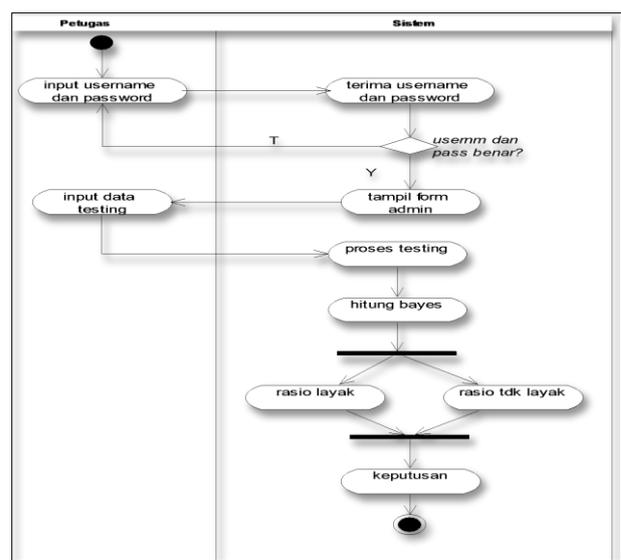
- 1) Penelitian selanjutnya sebaiknya tetap menggunakan data yang sudah ada kelayakannya untuk kemudahan proses *training* data.
- 2) Diperlukan penelitian lebih lanjut mengenai hubungan kriteria dengan akurasi prediksi kelayakan kredit.
- 3) Data dan kriteria yang digunakan perlu ditambah untuk hasil akurasi prediksi yang lebih baik.

REFERENSI

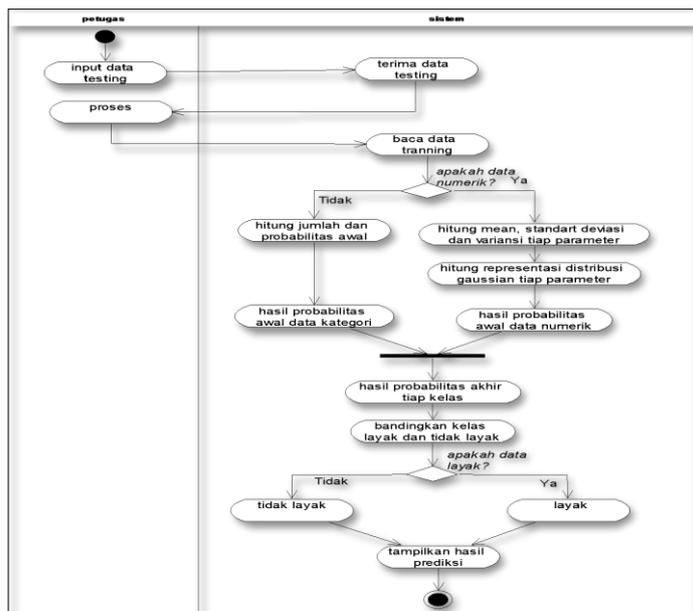
[1] Fajar Astuti Hermawati, *Data mining*, Andi, Yogyakarta, 2013
 [2] Budi Santosa, *Data mining: teknik pemanfaatan data untuk keperluan bisnis*, Graha Ilmu, Yogyakarta, 2007
 [3] Bustami, “Penerapan algoritma naïve bayes untuk mengklasifikasi data nasabah asuransi”, Universitas Malikussaleh, Aceh, 2013
 [4] Putri Wulandari, “Sistem penentuan kredit koperasi dengan menggunakan pendekatan metode naïve bayes pada studi kasus koperasi karya audita perwakilan BPKP Yogyakarta”, STMIK AKAKOM, Yogyakarta, 2014
 [5] Arif Rahman Bintoro, “Implementasi jaringan syaraf tiruan dengan metode backpropagation untuk prediksi kelayakan kredit di BMT Beringharjo, STMIK AKAKOM, Yogyakarta, 2014



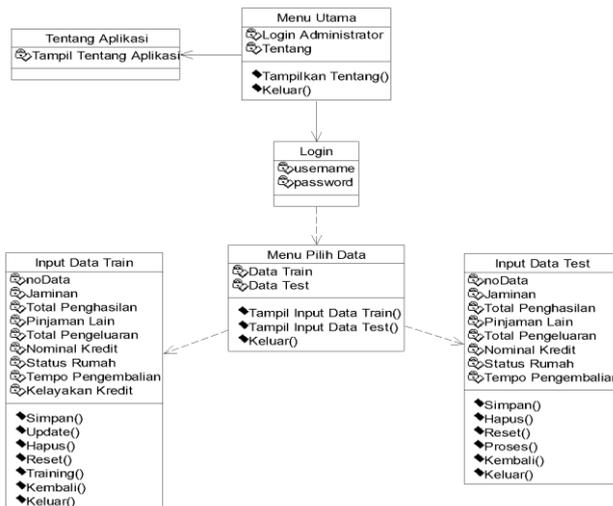
Gambar 2. Activity Diagram Training



Gambar 3. Activity Diagram Testing



Gambar 4. Activity Diagram Naive Bayes



Gambar 5. Diagram Class

No Data	Jaminan	Total Penghasilan	Pinjaman Lain	Total Pengeluaran	Desar Kredit	Status Rumah	Tempo Pengemb.	Kelayakan Kredit
1	Tabungan	2300000	300000	700000	1000000	Pribadi	3	Layak
2	Tabungan	2500000	500000	500000	500000	Pribadi	3	Layak
3	Motor	5000000	500000	1500000	3000000	Kontrak	3	Layak
4	Motor	4500000	500000	1000000	5000000	Pribadi	3	Layak
5	Tabungan	1500000	300000	750000	900000	Pribadi	3	Layak
6	Tabungan	1400000	300000	500000	900000	Kontrak	3	Layak
7	Tabungan	1500000	0	1070000	2500000	Pribadi	3	Layak
8	Tabungan	2000000	500000	1000000	1000000	Kontrak	3	Layak

Gambar 6. Contoh Data Training

No Data	Jaminan	Total Penghas..	Pinjaman Lain	Total Pengelu...	Besar Kredit	Status Rumah	Tempo Penge...	Prediksi NBC
1	Tabungan	2300000	300000	700000	1000000	Pribadi	3	TidakLayak
2	Motor	5000000	10000000	4500000	45000000	Pribadi	24	TidakLayak
3	Motor	2500000	15000000	1000000	10000000	Kontrak	12	TidakLayak
4	Mobil	2500000	5000000	2000000	7500000	Pribadi	24	TidakLayak
5	Saham	7000000	6000000	4000000	15000000	Kontrak	24	TidakLayak
6	Motor	1750000	5000000	1000000	5000000	Pribadi	6	TidakLayak
7	Motor	3000000	5000000	2000000	15000000	Pribadi	12	TidakLayak
8	Motor	4750000	4500000	3000000	25000000	Pribadi	6	TidakLayak
9	Motor	5000000	5000000	3000000	40000000	Pribadi	3	TidakLayak
10	Saham	2000000	6000000	1000000	17000000	Pribadi	12	TidakLayak

Probabilitas Layak	4.301239432536737E-20
Probabilitas Tidak Layak	3.0561685853859835E-18
Hasil Prediksi	TidakLayak

Gambar 7. Contoh Data Testing dan Nilai Probabilitas Kelayakan Kredit