

ANALISA KLASIFIKASI KARTU KREDIT MENGGUNAKAN ALGORITMA NAIVE BAYES

Reny Wijayanti¹, Sulastr²

^{1,2}Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Stikubank
e-mail: ¹renywijayanti6522@gmail.com, ²sulastr@edu.unisbank.ac.id

ABSTRAK

Pemberian kredit bank terhadap nasabah adalah kegiatan rutin yang mempunyai resiko tinggi. Begitupun kredit yang bermasalah atau kredit macet sering terjadi akibat analisis kredit 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 dalam pengajuan kredit. Diperlukan adanya prediksi yang akurat dimasa mendatang, salah satunya menggunakan teknik data mining.

Salah satu teknik data mining sendiri yaitu metode klasifikasi untuk memprediksi probabilitas keanggotaan suatu class dengan menggunakan algoritma Naive Bayes, sehingga tingkat akurasi dapat diketahui melalui proses Confusion Matrix dalam prediksi. Algoritma ini mampu menghasilkan tingkat akurasi yang baik.

Dari hasil analisa terhadap 23 variabel terdapat 3 variabel yang begitu berpengaruh kredit lancar dan macet nasabah yaitu status pembayaran pada bulan September, Agustus, dan Mei 2005. Tingkat akurasi data 3 variabel sebanyak 80.59% dengan data train sebanyak 18001 data.

Kata Kunci : Data Mining, Model Klasifikasi, Naive Bayes, Kartu Kredit.

1. PENDAHULUAN

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

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *Machine Learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakit dari berbagai database besar (Turban, dkk., 2005).

Dalam kaitannya data mining memiliki beberapa teknik salah satunya adalah teknik klasifikasi. Klasifikasi adalah pemrosesan untuk menemukan sebuah model atau fungsi yang menjelaskan dan mencirikan konsep atau kelas data, untuk kepentingan tertentu. Klasifikasi memiliki beberapa Algoritma klasifikasi data mining salah satunya yaitu *Naive Bayes*. Pada penerapan beberapa kasus teknik klasifikasi, Algoritma ini mampu menghasilkan akurasi dan performansi yang baik. *Bayesian Classification* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class (Kusrini, 2009).

Pada penelitian sebelumnya BMT Bringham Yogyakarta metode Naive Bayes digunakan sebagai alat bantu dalam memprediksi kelayakan kredit dengan merancang sebuah aplikasi menggunakan software UML (Dedy, 2016). Metode Naive Bayes dapat digunakan untuk menentukan kelayakan kredit terhadap peningkatan nilai NPL dari UKM dengan menggunakan sampel data penelitian sekunder terhadap record data baru yang belum memiliki kelas, dengan melakukan perhitungan probabilitas nilai kelas, probabilitas level golongan nilai kelas, jumlah tanggungan nilai kelas, level pinjaman dan jangka waktu, menurut (Yuli, 2015). Untuk menentukan kelayakan kredit Algoritma C4.5 dapat digunakan untuk memprediksi calon nasabah Bagian Dana Bank XY yang mengalami masalah, menurut (Angga, 2012).

2 METODE PENELITIAN

Metode yang digunakan untuk menangani permasalahan yang ada bertujuan untuk memperlihatkan bagaimana sebuah model data mining dapat digunakan untuk membantu mengetahui pola nasabah dalam mengambil kredit di salah satu Negara yaitu Taiwan pada bulan April hingga September 2005 berdasarkan atribut-atribut dari data nasabah yang mempunyai kredit lancar ataupun macet. Data ini diambil dari *UCI Machine Learning Repository*.

Dari berbagai model klasifikasi yang ada digunakan Algoritma *Naive Bayes* yang digunakan untuk mendapatkan informasi berdasarkan atribut yang merupakan sistem pengukuran statistik. Rumus untuk menghitung probabilitas :

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

Parameter 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 berdasarkan kondisi pada hipotesis H.

P(X) Probabilitas dari X.

Setelah mendapat informasi dari semua atribut yang dihitung, atribut yang paling mempengaruhi akan menjadi faktor paling penting dalam pemberian kartu kredit terhadap nasabah. Berikut adalah atribut data nya :

Tabel 1. Variabel Data

ATR	Keterangan	
X1	LIMIT_BAL	Jumlah kredit yang diberikan dalam dolar NT (termasuk kredit individu dan keluarga / pelengkap).
X2	SEX	Jenis kelamin (1 = laki-laki, 2 = perempuan)
X3	PENDIDIKAN	(1 = sekolah pascasarjana, 2 = universitas, 3 = sekolah tinggi, 4 = lainnya, 5 = tidak diketahui, 6 = tidak diketahui)
X4	MARRIAGE	Status perkawinan (1 = menikah, 2 =tunggal, 3=lainnya)
X5	UMUR	Usia di tahun
X6	PAY_0	Status pembayaran pada bulan September 2005 (-1 = bayar sepatutnya, 1 = penundaan pembayaran untuk satu bulan, 2 = penundaan pembayaran selama dua bulan, ... 8 = penundaan pembayaran selama delapan bulan, 9 = penundaan pembayaran selama sembilan bulan dan atas)
X7	PAY_2	Status pelunasan pada bulan Agustus 2005 (skala sama seperti di atas)
X8	PAY_3	Status pelunasan pada bulan Juli 2005 (skala sama seperti di atas)
X9	PAY_4	Status pelunasan pada bulan Juni 2005 (skala sama seperti di atas)

X10	PAY_5	Status pembayaran kembali pada bulan Mei 2005 (skala sama seperti di atas)
X11	PAY_6	Status pelunasan pada bulan April 2005 (skala sama seperti di atas)
X12	BILL_AMT1	Jumlah pernyataan tagihan pada bulan September 2005 (dolar NT)
X13	BILL_AMT2	Jumlah pernyataan tagihan pada bulan Agustus 2005 (dolar NT)
X14	BILL_AMT3	Jumlah pernyataan tagihan pada bulan Juli 2005 (dolar NT)
X15	BILL_AMT4	Jumlah pernyataan tagihan pada bulan Juni 2005 (dolar NT)
X16	BILL_AMT5	Jumlah pernyataan tagihan pada bulan Mei 2005 (dolar NT)
X17	BILL_AMT6	Jumlah pernyataan tagihan pada bulan April 2005 (dolar NT)
X18	PAY_AMT1	Jumlah pembayaran sebelumnya pada bulan September 2005 (dolar NT)
X19	PAY_AMT2	Jumlah pembayaran sebelumnya pada bulan Agustus 2005 (dolar NT)
X20	PAY_AMT3	Jumlah pembayaran sebelumnya pada bulan Juli 2005 (dolar NT)
X21	PAY_AMT4	Jumlah pembayaran sebelumnya pada bulan Juni 2005 (dolar NT)
X22	PAY_AMT5	Jumlah pembayaran sebelumnya pada bulan Mei 2005 (dolar NT)
X23	PAY_AMT6	Jumlah pembayaran sebelumnya pada bulan April 2005 (dolar NT)
Y	Default.payment.	Pembayaran default (1 = ya, 0 = tidak)

Sebuah proses data mining ini melalui tahap dalam analisisnya menurut *knowledge discovery in databases* (KDD) sebagai berikut :

2.1 Persiapan Data

Tahapan awal yang dilakukan dalam penelitian sebelum diproses ke dalam data mining :

Gambar 1. Persiapan Data

2.2. Transformasi Data

Menormalisasikan data dan membersihkan data yang tidak digunakan. Berikut adalah gambar hasil transformasi :

Gambar 2. Hasil Proses Transformasi

2.3. Pembagian Data

Proses data mining ini dilakukan untuk membagi data training dan data testing. Pembagian data ini adalah 60% untuk data train dan 40% data test. Sehingga menghasilkan 18001 data train dan 11999 data test dengan 24 variabel.

3. HASIL DAN PEMBAHASAN

3.1. Pemodelan Data

Proses data mining ini untuk mengimplimentasikan data train kedalam model data yang akan digunakan dengan algoritma *Naive Bayes*. Variabel yang digunakan untuk pemodelan tersebut yaitu X6, X7, dan X10. Berikut adalah hasil model naive bayes

```

> model_naive
Naive Bayes Classifier for Discrete Predictors
call:
naiveBayes.default(x = X, y = Y, laplace = laplace)
A-priori probabilities:
Y
  No      Yes
0.7787901 0.2212099
Conditional probabilities:
  X6
Y   [,1] [,2]
No -0.2194219 0.950157
Yes 0.6702662 1.399040

  X7
Y   [,1] [,2]
No -0.3012340 1.036806
Yes 0.4527875 1.513452

  X10
Y   [,1] [,2]
No -0.3985306 0.9720891
Yes 0.1863385 1.5163628

```

Gambar 3. Hasil Pemodelan Naive Bayes

Pemodelan ini menghasilkan 10 rule dengan data sebanyak 18001. Berikut adalah hasil dari rule tersebut :

```

> print(credit)
conditional inference tree with 10 terminal nodes
Response: Y
Inputs: X6, X7, X10
Number of observations: 18001
1) X6 <= 1; criterion = 1, statistic = 1913.801
2) X7 <= 1; criterion = 1, statistic = 197.6
3) X7 <= -1; criterion = 1, statistic = 47.543
4) X6 <= 0; criterion = 1, statistic = 76.029
5)* weights = 4785
4) X6 > 0
6) X10 <= -2; criterion = 0.991, statistic = 8.865
7)* weights = 514
6) X10 > -2
8)* weights = 574
3) X7 > -1
9) X10 <= 0; criterion = 1, statistic = 98.233
10)* weights = 8415
9) X10 > 0
11) X10 <= 2; criterion = 0.967, statistic = 6.471
12)* weights = 439
11) X10 > 2
13)* weights = 22
2) X7 > 1
14) X10 <= 0; criterion = 1, statistic = 27.722
15)* weights = 965
14) X10 > 0
16)* weights = 420
1) X6 > 1
17) X10 <= 0; criterion = 1, statistic = 29.443
18)* weights = 1067
17) X10 > 0
19)* weights = 800

```

Gambar 4. Hasil Rule 3 variabel

Sehingga menghasilkan perhitungan Confusion Matrix sebagai berikut :

```

> confusionMatrix(data=pred_naive, reference = Test$Y)
Confusion Matrix and Statistics

          Reference
Prediction No  Yes
No      8645 1629
Yes     700  1025

      Accuracy : 0.8059
      95% CI   : (0.7987, 0.8129)
      No Information Rate : 0.7788
      P-Value [Acc > NIR] : 2.142e-13

      Kappa : 0.3559
      Mcnemar's Test P-Value : < 2.2e-16

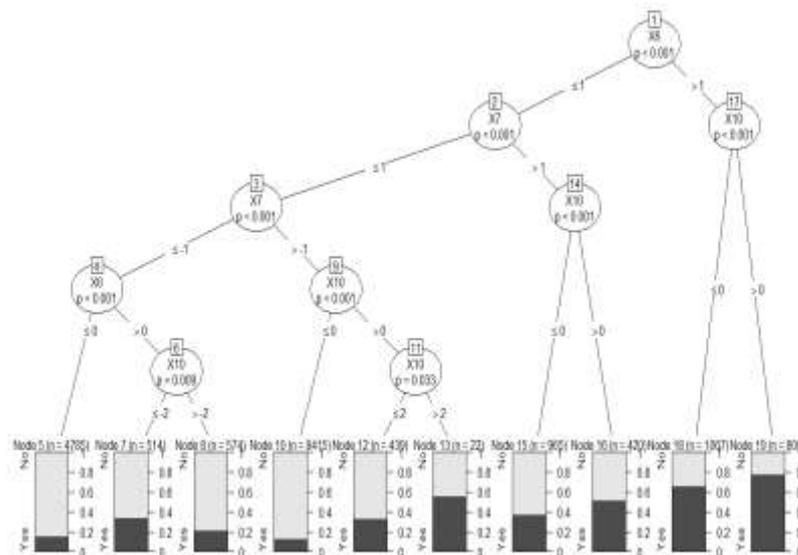
      Sensitivity : 0.9251
      Specificity : 0.3862
      Pos Pred Value : 0.8414
      Neg Pred Value : 0.5942
      Prevalence : 0.7788
      Detection Rate : 0.7205
      Detection Prevalence : 0.8562
      Balanced Accuracy : 0.6557

      'Positive' Class : No

```

Gambar 5. Hasil evaluasi

Dari pemodelan diatas maka dapat dibuat sebuah visualisasi menggunakan pohon keputusan sebagai berikut :



Gambar 6. Visualisasi 3 Variabel

4. KESIMPULAN

Dari hasil pengujian, dapat disimpulkan beberapa hal sebagai berikut :

1. Berdasarkan analisa pada penelitian ini, ketika menggunakan 23 variabel didapatkan akurasi klasifikasi untuk 18001 data yaitu 69.11%. akurasi ini masih minim sehingga penulis mencoba mengkomputasikan menggunakan 11 variabel dan mendapatkan tingkat akurasi data sebanyak 79.7%. Untuk mendapatkan keakurasian data yang lebih, peneliti mencoba mengkomputasikan menggunakan 6 variabel dan mendapat akurasi sebanyak 80.47%. Sedangkan jika menggunakan 4 variabel akurasi data sebanyak 80.52%.
2. Dan terakhir menggunakan 3 variabel memiliki akurasi data sebanyak 80.59%. Akurasi tersebut lebih baik jika dibandingkan klasifikasi terhadap semua variabel. Dimana variabel tersebut adalah X6, X7, X10 yaitu status pembayaran selama bulan September, Agustus, dan Mei 2005 dan terdapat 10 rule Hal ini disebabkan karena semua variabel tidak digunakan, sehingga mempengaruhi proses klasifikasi data.

DAFTAR PUSTAKA

- [1]. Ahmad Kurniawan, D dan Kriestanto, D. 2016. Penerapan Naive Bayes untuk Memprediksi Kelayakan Kredit. *Jurnal Informatika dan komputer*, Vol.1, No. 1, hal.19-23.
- [2] Asdi, Yudiantri (2017) Pengenalan Software R. Website: https://nanopdf.com/download/pengenalan-software-r_pdf, diakses 7 Mei 2018.
- [3] F.H, Fandi dan Hansun. 2017. Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE. *Jurnal Teknik Informatika dan Sistem Informasi*, Vol.3, No. 2.
- [4] Ginanjar Maburur, A dan Lubis, R. 2012. Penerapan Data Mining Untuk Memprediksi Kriteria Nasabah Kredit *Jurnal Komputer dan Informatika*, Vol.1, hal 53-57.
- [5] Hermawati, F. A. (2013). *Data Mining*. Yogyakarta: Penerbit Andi.
- [6] Indrajani. (2011). *Perancangan Basis Data* dalam All in 1, PT. Elex Media Komputindo, Jakarta.
- [7] Kusriani (2009), *Algoritma Data Mining*, edisi 1, Andi Offset, Yogyakarta.
- [8] Murdianingsih, Y. (2015). Klasifikasi Nasabah Baik Dan Bermasalah Menggunakan Metode Naive Bayes. *Seminar Nasional Informatika*, hal 349-356.
- [9] Retnosari, P. (2013) Implementasi Data Mining untuk menemukan hubungan antara Kota Kelahiran Mahasiswa dengan Tingkat Kelulusan Mahasiswa pada Fakultas Teknologi Informasi Unisbank, *Skripsi*, Program Studi Sistem Informasi FTI Unisbank, Semarang.
- [10] Turban, E., dkk, 2005, *Decicion Support Systems and Intelligent Systems*, Andi Offset.
- [11] Wibowo, Ari. (2011). *Prediksi Nasabah Potensial Menggunakan Metode Klasifikasi Pohon Biner*: Universitas Politeknik Negri Batam.