

Komparasi Algoritma Klasifikasi Data Mining untuk Evaluasi Pemberian Kredit

Siti Masripah^{1,*}

¹ Program Studi Manajemen Informatika; AMIK BSI Jakarta Jl.Kramat Raya No.18 Jakarta Pusat 021-3100413 ; e-mail: siti.stm@bsi.ac.id

* Korespondensi: e-mail: siti.stm@bsi.ac.id

Diterima: 27 April 2016; Review: 20 Mei 2016; Disetujui: 28 Mei 2016

Cara sitasi : Masripah S. 2016. Komparasi Algoritma Klasifikasi Data Mining untuk Evaluasi Pemberian Kredit. Bina Insani ICT Journal. 3 (1): 187 – 193.

Abstrak: Kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan pinjam meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam melunasi hutangnya setelah jangka waktu tertentu dengan pemberian bunga, pada koperasi permasalahan kredit merupakan permasalahan manajemen, dimana jika banyak nasabah yang menunggak dalam pembayaran maka akan mengganggu system keuangan yang ada, untuk itulah penelitian ini menerapkan proses analisa kredit nasabah terlebih dahulu sebelum diambil sebuah keputusan pemberian kredit, analisa keputusan memberikan kredit menggunakan algoritma klasifikasi C4.5 dan Naïve Bayes dimana kedua algoritma tersebut dilakukan penilaian, mana algoritma yang paling akurat dalam menganalisa kemampuan nasabah dalam membayar kredit, analisa berdasarkan data history. Hasil yang didapatkan dari perbandingan kedua algoritma tersebut, bahwa tingkat akurasi yang lebih baik adalah menganalisa menggunakan algoritma klasifikasi C4.5 yaitu 88.90 % sedangkan untuk tingkat akurasi menggunakan algoritma klasifikasi Naïve Bayes yaitu 80.00%.

Kata kunci: c4.5, klasifikasi, naïve bayes

Abstract: *Credit is the provision of cash or the equivalent, based on the agreement between bank lending and other parties who require the borrower to pay off debts after a certain period of time with interest, of the cooperative credit problems are management problems, whereby if a lot of customers in arrears payment it will disrupt the existing financial system, for why this study apply customer credit analysis process before taken a decision to grant credit, provide credit decision analysis using the algorithm C4.5 and Naïve Bayes classification where the algorithms are done votes, where the algorithm the most accurate in analyzing the ability of customers to pay for the credit, an analysis based on the data history. Results obtained from the comparison of the two algorithms are, that level of accuracy is even better is to analyze menggunakan C4.5 classification algorithm that is 88.90% while the level of accuracy using Naïve Bayes classification algorithm is 80.00%.*

Keywords: *clasification, c4.5, naïve bayes*

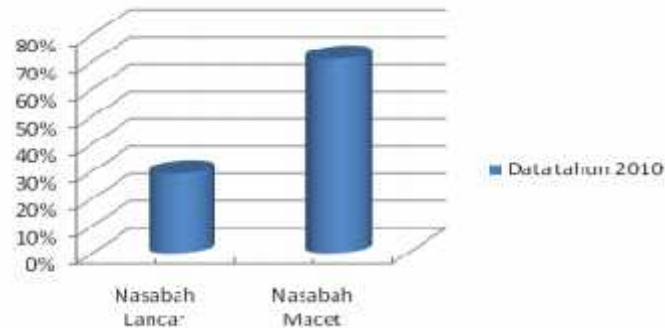
1. Pendahuluan

Dalam UU perbankan telah dijelaskan bahwa kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan pinjam meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam melunasi hutangnya setelah jangka waktu tertentu dengan pemberian bunga(UU No.10, 1998) . Berdasarkan UU tersebut sebuah kredit yang dilakukan harus berdasarkan persetujuan pinjam meminjam, dimana akan ada sebuah analisa yang dilakukan untuk menentukan sebuah langkah dalam pengambilan keputusan.

Sebuah keputusan yang diambil pasti memiliki resiko, salah satu resiko kredit adalah ketidak sanggupan nasabah dalam melakukan pembayaran angsuran. Resiko kredit adalah ketidak pastian atau fluktuasi laba dalam kegiatan kredit (Yu, Chen, Koronios, Zhu, & Guo, 2007).

Pihak pemberi kredit haruslah melakukan sebuah analisa untuk mengurangi resiko kredit, karena analisa kredit merupakan hal yang penting dalam manajemen resiko keuangan(Lai, Yu, Zhou, & Wang, 2006).

Dibawah ini merupakan data nasabah tahun 2010 yang memiliki permasalahan pada angsuran kredit, sehingga menyebabkan banyak nasabah yang tidak lancar dalam pembayaran kredit.



Gambar 1. Grafik Status Kredit nasabah

Berdasarkan data pada Gambar 1, bahwa tingkat status nasabah macet lebih besar dengan perbandingan 71% dan 29% (Masripah, 2014), hal tersebut akan menjadi sebuah permasalahan besar dalam manajemen keuangan perusahaan. Dalam penelitian ini digunakan 13 variabel yaitu nama nasabah, jenis kelamin, umur, jumlah pinjaman, jangka waktu, jumlah angsuran perbulan, type pinjaman, bi sector ekonimo, col, saldo nominative, plafon teoritis, tunggakan pokok dan tunggakan bunga.

Ada beberapa peneliti yang melakukan penelitian terkait dengan bidang data mining diantaranya Khafiizh Hastuti yang dalam penelitiannya membahas tentang komparasi algoritma klasifikasi untuk menentukan prediksi mahasiswa non aktif. Dimana penelitian tersebut untuk mengetahui tingkat akurasi dari masing-masing algoritma klasifikasi. Windy Julianto dan kawan-kawan melakukan penelitian tentang Algoritma C4.5 untuk Penilaian Kinerja Karyawan mengatakan bahwa melakukan penilaian kinerja karyawan menggunakan algoritma C4.5 dapat membantu proses monitoring karyawan untuk penentuan kenaikan gaji atau promosi jabatan. Penelitian Sri Kusumadewi dalam penelitiannya tentang klasifikasi status gizi menggunakan Nive Bayes Classification mengatakan untuk menilai status gizi seseorang. Algoritma Naïve Bayes dapat memecahkan masalah dengan cukup baik, hasil penelitian menunjukkan kinerja system sebesar 93,2%.

2. Metode Penelitian

Metode penelitian yang digunakan yaitu metode Experiment, yaitu suatu metode penyelidikan hubungan kasual menggunakan tes dikendalikan oleh peneliti(Dowson, 2009). Penelitian dilakukan menggunakan data history.

Metode experiment terdiri dari : (1) Pengumpulan Data, (2) Pengolahan Data, (3) Model yang diusulkan, (4) Pengujian Model, (5) Evaluasi dan validasi Model(Swastina, 2013).

- a. Pengumpulan Data, yaitu menggunakan data primary dan sekunder. Untuk data primer adalah data hasil uji menggunakan algoritma-algoritma yang digunakan, sedangkan data sekunder adalah data yang diperoleh dari dokumentasi atau buku, jurnal, literature yang berkaitan dengan penelitian.
- b. Pengolahan Data, data didapat dari koperasi yaitu terdiri dari 13 variabel.
- c. Model yang diusulkan, model menggunakan teknik data mining dengan model algoritma C4.5 dan Algoritma Naïve Bayes. Tahapan algoritma C4.5 :
 - 1) Menyiapkan data training
 - 2) Menentukan akar dari pohon
 - 3) Hitung nilai Gain:

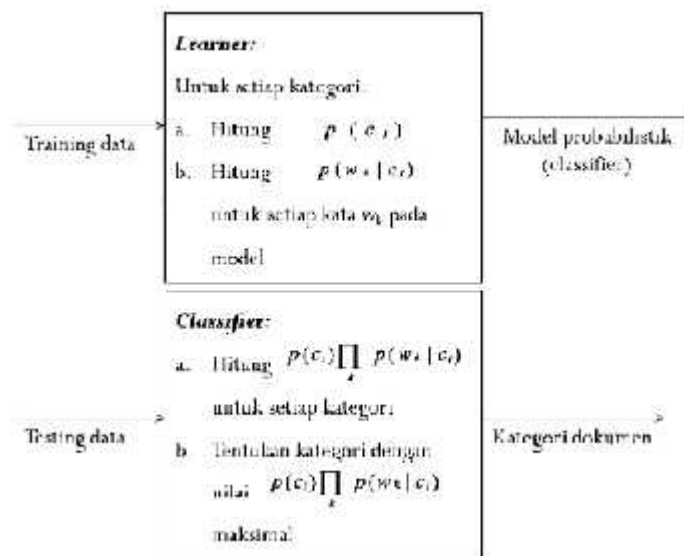
$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

- 4) Ulangi langkah ke-2 hingga semua tupel terpartisi

$$Gain(S, A) = S - \sum_{i=1}^n \frac{|S_i|}{|S|} * S_i$$

- 5) Proses partisi pohon keputusan akan berhenti saat semua tupel dalam node N mendapat kelas yang sama dan atau tidak ada atribut didalam tupel yang dipartisi lagi atau tidak ada tupel dicabang yang kosong

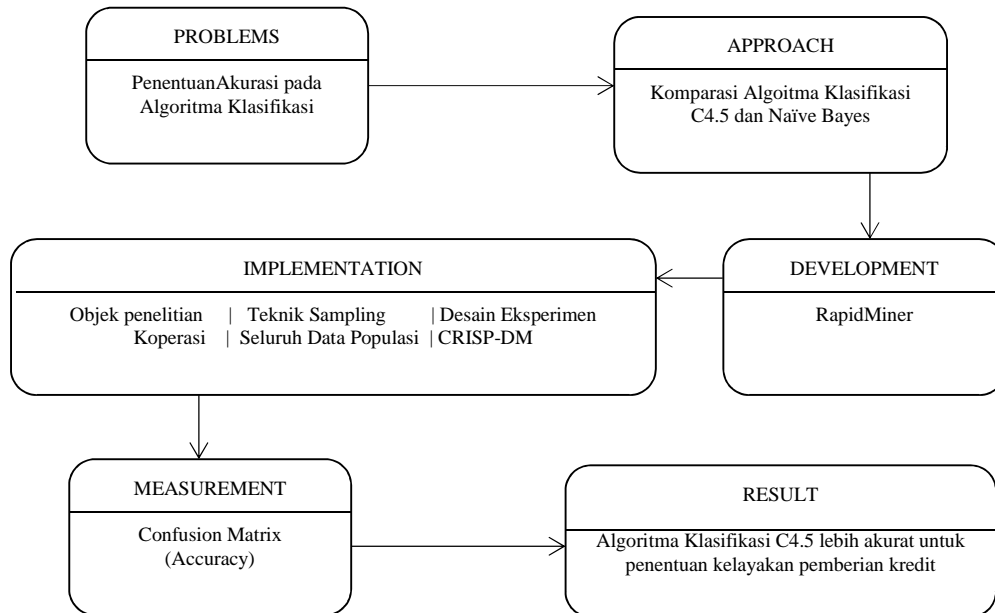
Untuk algoritma Naïve Bayes merupakan salah satu metode machine learning yang menggunakan perhitungan probabilitas, berikut gambaran klasifikasi algoritma naïve Bayes :



Gambar 2. Tahapan Proses Klasifikasi dengan Naïve Bayes

- d. Pengujian Model, data pengujian terbagi menjadi dua yaitu data training dan data testing
- e. Evaluasi dan Validasi Hasil, Evaluasi dilakukan dengan menganalisa hasil klasifikasi menggunakan pengukuran confusion matrix. Hasil dari algoritma C4.5 dengan algoritma Naïve Bayes akan dibandingkan hasilnya.

Kerangka pemikiran ini dapat dilihat pada gambar 3 berikut ini:



Gambar 3. Kerangka Pemikiran

Penjelasan untuk kerangka pemikiran pada gambar 3 yaitu:

- a. Problems
Problems atau permasalahan yang ada pada penelitian ini adalah penentuan akurasi pada algoritma klasifikasi
- b. Approach
Approach atau metode penelitian yang dipakai pada kasus ini adalah dengan komparasi algoritma C4.5 dan Algoritma Naïve Bayes
- c. Development
Development yang digunakan dalam membantu proses penelitian ini dengan tools Rapid Miner
- d. Implementation
Penelitian ini menggunakan data Koperasi, menggunakan pengambilan data seluruh data populasi, dan urutan desain eksperimen menggunakan Cripss-DM (Cross Standard Industry Process for Data Mining)
- e. Measurement
Measurement atau pengukuran yang dilakukan adalah dengan Confusion Matrix (Accuracy)
- f. Result
Result atau hasil penelitian yang didapat adalah Algoritma C4,5 lebih akurat untuk penentuan kelayakan pemberian kredit.

Data Mining

Data mining adalah sebuah proses, sehingga dalam melakukan prosesnya harus sesuai prosedur yaitu proses CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yaitu sebagai keseluruhan proses, preprocessing data, pembentukan model, model evaluasi dan akhirnya penyebaran model (Larose, 2005).

Dalam metode penelitian eksperimen digunakan model proses CRISP-DM yang memiliki enam tahapan (Larose, 2005):

- a. Business Understanding
- b. Data Understanding
- c. Data Preparation
- d. Modelling
- e. Evaluation
- f. Deployment

Dalam data mining, terdapat fungsi-fungsi yang dapat digunakan, ada enam fungsi yaitu (Susanto, 2010):

- a. Fungsi deskripsi (*description*)
Adalah cara yang digunakan untuk menggambarkan sekumpulan data secara ringkas, banyak cara yang digunakan dalam memberikan gambaran secara ringkas bagi sekumpulan data yang besar jumlahnya dan banyak macamnya yaitu deskripsi Grafis, Deskripsi Lokasi, dan Deskripsi Keragaman
- b. Fungsi estimasi (*estimation*)
Fungsi estimasi adalah fungsi untuk memperkirakan suatu hal yang sudah ada datanya. Fungsi estimasi terdiri dari dua cara yaitu Estimasi Titik dan Estimasi Selang Kepercayaan.
- c. Fungsi prediksi (*prediction*)
Fungsi prediksi adalah memperkirakan hasil dari hal yang belum diketahui, untuk mendapatkan hal baru yang akan muncul selanjutnya. Cara memprediksi dalam fungsi ini adalah Regresi Linier.
- d. Fungsi klasifikasi (*classification*)
Fungsi klasifikasi atau menggolongkan suatu data. Cara yang digunakan terdiri dari algoritma Mean Vector, algoritma K-nearest Neighbor, algoritma ID3, algoritma C4.5, dan algoritma C5.0
- e. Fungsi pengelompokan (*cluster*)
Fungsi pengelompokan, data yang dikelompokkan disebut objek atau catatan yang memiliki kemiripan atribut kemudian dikelompokkan pada kelompok yang berbeda. Algoritma yang digunakan adalah algoritma Hirarchical Clustering, algoritma Partitional Clustering, algoritma Single Linkage, algoritma Complete Linkage, algoritma Average Linkage, algoritma K-Means dan lain-lain.
- f. Fungsi asosiasi (*association*)
Fungsi asosiasi adalah untuk menemukan aturan asosiasi (*association rule*) yang mampu mengidentifikasi item-item yang menjadi objek. Algoritma yang digunakan adalah algoritma Generalized Association Rules, Quantitative Association Rule, asynchronous Parallel Mining

3. Hasil dan Analisis

Pengujian melakukan data nasabah tahun 2010, dimana terdapat 13 variabel yang ada yaitu nama nasabah, jenis kelamin, umur, jumlah pinjaman, jangka waktu, jumlah angsuran perbulan, type pinjaman, bi sector ekonimo, col, saldo nominative, plafon teoritis, tunggakan pokok dan tunggakan bunga.

	True NEGAT	True POSITIF	class precision
pred NEGAT	71	31	94,27%
pred POSITIF	80	373	82,57%
class recall	98,44%	82,44%	

Gambar 4. Hasil Performance Vector Algoritma C4.5

Pada gambar 4 hasil pengujian C4.5 menghasilkan nilai akurasi 88,90%. Nilai Class Precision 82,57% dan Class Recall 92,44%.

	true MACET	true LANCAR	class precision
pred MACET	420	10	97,96%
pred LANCAR	110	400	78,43%
class recall	81,38%	97,56%	

Gambar 5. Hasil Performance Vector Algoritma Naïve Bayes

Sedangkan pada gambar 5, hasil pengujian menggunakan algoritma Naïve Bayes menghasilkan nilai akurasi 88,00%. Nilai Class Precision 78,43% dan Class Recall 97,56%.

Jika digambarkan kedalam sebuah table, maka hasil eksperimen untuk perbandingan validasi hasil prediksi kelayakan pemberian kredit adalah sebagai berikut :

Tabel 1. Perbandingan Validasi Hasil Prediksi Kelayakan Pemberian Kredit

	C4.5 (%)	Naïve Bayes (%)
Akurasi	88,90	88,00
Presisi	82,57	78,43
Recall	92,44	97,56

Berdasarkan tabel 1, maka nilai akurasi yang didapat untuk algoritma C4.5 sebesar 88,90% sedangkan untuk akurasi pada algoritma Naïve Bayes sebesar 88,00%.

4. Kesimpulan

Dari hasil pengujian dua buah algoritma yaitu algoritma C4.5 dan Algoritma Naïve Bayes didapat sebuah kesimpulan yaitu Algoritma C4.5 memprediksi lebih akurat dari pada algoritma Naïve Bayes untuk penentuan kelayakan kredit atau evaluasi pemberian kredit kepada nasabah dengan tingkat keakuratan sebesar 82,57% dan akurasi sebesar 88,90%.

Dengan adanya penerapan algoritma C4.5 dalam memprediksi kelayakan pemberian kredit atau evaluasi pemberian kredit dapat membantu lembaga keuangan guna menentukan calon penerima kredit, dan dapat digunakan untuk bahan evaluasi keuangan yang terdapat pada perusahaan.

Referensi

- Dawson CW. 2009. Project in Computing and Information Systems a Student's Guide. England: Addison Wesley
- Hastuti K. 2012. Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Mahasiswa Non Aktif. Seminar Nasional Teknologi Informasi & Komunikasi Terapan (Semantik 2012). Semarang.
- Lai KK, Yu L, Zhou L, Wang S. 2006. Credit Risk Evaluation With Least Square Support Vector Machine.
- Larose DT. 2005. Discovering Knowledge In Data. Canada: Wiley-Interscience
- Swastina L. 2013. Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa. Jurnal GEMA AKTUALITA, Vol. 2 No. 1, Juni 2013.

- Masripah S. 2014. Application Determination of Credit Feasibility in Sharia Cooperative with C4.5 Algorithm. Seminar Internasional ISIT 2014 Proceedings Bulan Oktober 2014 ISBN 978-602-99213-8-0.
- Kusumadewi S. 2009. Klasifikasi Status Gizi Menggunakan Naive Bayesian Classification. CommIT, Vol. 3 No. 1 Mei 2009, hlm. 6 – 11
- Susanto S, Suryadi D. 2010. Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data. Yogyakarta : C.V Andi Offset
- Undang-undang Republik Indonesia Nomor 10 tahun 1998 tentang Perubahan atas Undang-undang nomor 7 tahun 1992 tentang Perbankan
- Yu L, Chen G, Koronios A, Zhu S, Guo X. 2007. Application and Comparison of Classification Techniques in Controlling Credit Risk. *World Scientific*, 111
- Julianto WA, Yunitarini R, Sophan MK. 2014. Algoritma C4.5 Untuk Penilaian Kinerja Karyawan. Universitas Trunojoyo Madura. SCAN VOL. IX NOMOR 2 JUNI 2014 ISSN : 1978-0087