

MODEL POHON KEPUTUSAN UNTUK KLASIFIKASI PERSETUJUAN KREDIT MENGUNAKAN ALGORITMA C4.5

Ivandari
Program Studi Teknik Informatika
STMIK Widya Pratama

Jl. Patriot 25 Pekalongan
Telp (0285) 427816
Email: Ivdarialkaromi@gmail.com

ABSTRAK

KREDIT MERUPAKAN SALAH SATU AKTIVITAS EKONOMI YANG BANYAK DILAKUKAN. DENGAN MELAKUKAN PEMBELIAN SECARA KREDIT MASYARAKAT SETIDAKNYA DAPAT TERBANTU DALAM MEMBELI BARANG DENGAN HARGA YANG KURANG TERJANGKAU JIKA DILAKUKAN DENGAN CARA PEMBAYARAN KONTAN. PERSETUJUAN KREDIT MERUPAKAN SALAH SATU HAL YANG PENTING DAN DAPAT MEMPENGARUHI PERKEMBANGAN BANK DAN PIHAK PEMBERI PINJAMAN. TEKNOLOGI KOMPUTER DAPAT MEMBANTU DALAM PELAKSANAAN KLASIFIKASI PERSETUJUAN KREDIT. SALAH SATU MODEL KLASIFIKASI TERBAIK YANG BANYAK DILAKUKAN DAN TERBUKTI BAIK ADALAH C4.5. HASIL DARI ALGORITMA C4.5 MERUPAKAN SEBUAH MODEL POHON KEPUTUSAN YANG DAPAT DENGAN MUDAH DIPAHAMI DENGAN BAHASA ALAMI MANUSIA. DALAM PENELITIAN INI AKAN DIBUAT SEBUAH MODEL POHON KEPUTUSAN DENGAN MENGGUNAKAN ALGORITMA C4.5 UNTUK KLASIFIKASI PERSETUJUAN KREDIT.

Kata kunci: Algoritma C4.5, Klasifikasi Persetujuan Kredit, Model Pohon Keputusan

1 Pendahuluan

1.1 Latar Belakang

Kredit merupakan aktivitas ekonomi yang banyak dilakukan pada masa sekarang ini. Banyaknya nasabah yang mengalami kesulitan dalam pembayaran angsuran kredit menyebabkan pihak pemilik dana harus lebih selektif dalam memberikan persetujuan kepada nasabahnya.

Aplikasi bantu untuk menyelesaikan masalah ini banyak dibuat dengan menggunakan berbagai metode. Tahun 2013 SPK mengenai persetujuan kredit pernah dibuat dengan studi kasus di Bank Muamalat cabang Yogyakarta [1]. Sebelumnya pada 2009 juga pernah digunakan aplikasi sejenis yang diimplementasikan pada PT.BPR Mranggen [2]. Dalam hal ini banyak juga digunakan metode Data Mining [3] untuk menentukan kelayakan nasabah dalam persetujuan kredit [4]. Dengan menggunakan metode Bayes [5], AHP [6], *decision tree* [7] serta SAW [8] banyak dibuat SPK untuk persetujuan kredit kendaraan.



Metode data mining terbukti banyak memberikan kontribusi terhadap klasifikasi terutama jika diterapkan dalam perekonomian modern. Salah satu metode yang terbukti baik dan menghasilkan model yang mudah dipahami dengan bahasa alami manusia adalah model pohon keputusan atau *decission tree* [9]. Dalam penelitian ini akan dibuat sebuah model pohon keputusan dengan menggunakan algoritma C4.5 untuk klasifikasi persetujuan kredit.

1.2 Landasan Teori

1.2.1 Data Mining

Data mining adalah proses mendapatkan sebuah informasi berharga dari kumpulan data yang sebelumnya bisa jadi kurang berguna [10]. Teknik data mining umumnya menggunakan pengenalan pola dari data yang digunakan [11]. Bidang ilmu ini merupakan sebuah proses untuk mendapatkan pengetahuan baru atau pola dari kumpulan data [10].

Dengan menggunakan proses data mining, data yang sebelumnya bisa jadi tidak berguna akan dapat lebih bermanfaat dengan cara pengenalan pola seperti halnya statistik dan matematika. Dalam data mining sendiri berdasarkan metode pembelajarannya dapat digolongkan menjadi dua bagian, yaitu *supervised learning* dan *unsupervised learning* [12]. *Supervised learning* merupakan pembelajaran dengan menggunakan label atau ada atribut tujuan. Dalam hal ini terdapat berbagai metode untuk masing masing kegunaan. Salah satunya adalah klasifikasi, estimasi, prediksi, serta asosiasi.

1.2.2 Klasifikasi

Seperti yang telah dijelaskan sebelumnya bahwa klasifikasi merupakan salah satu metode *supervised learning*. Karena di dalam proses klasifikasi terdapat proses pembelajaran dengan memanfaatkan data lampau. Proses pembelajaran ini berguna untuk mengenali pola data yang nantinya dapat diterapkan kepada data baru yang belum diketahui kelompoknya. Teknik klasifikasi sudah banyak diaplikasikan dalam dunia nyata seperti halnya dalam dunia medis [13], pendidikan [14] [15] [16] [17] [18], teknik bangunan [19], jaringan komputer [20], serta banyak digunakan dalam bidang lain.

1.2.3 Algoritma C4.5

C4.5 Merupakan pengembangan dari algoritma ID3 [11] yang dikembangkan oleh Quinlan [21]. Algoritma C4.5 banyak digunakan peneliti untuk melakukan tugas klasifikasi. Output dari algoritma C4.5 adalah sebuah pohon keputusan atau sering dikenal dengan *decission tree*. Dalam beberapa penelitian algoritma C4.5 ini menjadi pilihan terbaik dibandingkan dengan beberapa algoritma klasifikasi lain [14][20]

Decision tree sendiri merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal [22]. Dalam *decission tree* ini data yang berupa fakta dirubah menjadi sebuah pohon keputusan yang berisi aturan dan tentunya dapat lebih mudah dipahami

dengan bahasa alami. Model pohon keputusan banyak digunakan pada kasus data dengan output yang bernilai diskrit [12]. Walaupun tidak menutup kemungkinan dapat juga digunakan untuk kasus data dengan atribut numeric.

Setiap *node* dalam *decision tree* dapat merepresentasikan sebuah atribut. Sedangkan cabang dari *node* merupakan nilai dari atribut tersebut, serta daun merepresentasikan kelas. *Node* paling atas pada *decision tree* disebut sebagai *root node*. *Root node* ini tidak memiliki *input* serta bisa saja tidak memiliki *output* dan bahkan dapat memiliki *output* lebih dari satu. *Internal root* merupakan *node* percabangan yang hanya memiliki satu *input* dan memiliki minimal dua *output*. *Leaf node* atau *terminal node* merupakan *node* akhir yang hanya memiliki satu *input* serta tidak memiliki *output*.

Langkah untuk membuat sebuah *decision tree* dari algoritma C4.5 adalah sebagai berikut [21]:

1. Mempersiapkan data training, *data training* yaitu data yang diambil dari data histori yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menentukan akar pohon. Akar pohon ditentukan dengan cara menghitung *GainRatio* tertinggi dari masing-masing atribut. Sebelum menghitung *GainRatio*, terlebih dahulu menghitung *Total Entropy* sebelum dicari masing-masing *Entropy class*, adapun rumus mencari *Entropy* sebagai berikut:

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

Keterangan:

- S = Himpunan kasus
- n = jumlah partisi S
- p_i = proporsi dari S_i terhadap S

Dimana $\log_2 p_i$ dapat dihitung dengan cara:

$$\log(X) = \frac{\ln(X)}{\ln(2)}$$

3. Menghitung nilai *GainRatio* sebagai akar pohon, tetapi sebelumnya menghitung *Gain* dan *SplitEntropy (SplitInfo)*, rumus untuk menghitung *Gain* seperti dibawah ini:

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

Rumus untuk menghitung *SplitEntropy*, seperti di bawah ini:

$$SplitEntropy_A(S) = - \sum_{i=1}^n \frac{|S_i|}{|S|} * \log_2 \left(\frac{|S_i|}{|S|} \right)$$

Rumus untuk menghitung *GainRatio*, dibawah ini:

$$\text{GainRatio}(A) = \frac{\text{Gain}(A)}{\text{SplitEntropy}(A)}$$

Keterangan:

- S = Himpunan Kasus
- A = Atribut
- n = jumlah partisi atribut A
- |S_i| = jumlah kasus pada partisi ke-i
- |S| = jumlah kasus dalam S

4. Ulangi langkah ke-2 dan ke-3 hingga semua tupel terpartisi
5. Proses partisi pohon keputusan akan berhenti disaat:
 - a. Semua tupel dalam node N mendapatkan kelas yang sama
 - b. Tidak ada atribut didalam tupel yang dipartisi lagi
 - c. Tidak ada tupel didalam cabang yang kosong

1.2.4 Cross Validation

Cross validation adalah pembuktian dari sebuah metode atau pembuktian dari perhitungan performa suatu algoritma. Dalam proses pengujian sebuah metode data mining banyak digunakan *cross validation*. *Cross validation* merupakan pembuktian dengan membagi data sebagian sebagai *data training* dan sebagian yang lain sebagai *data testing* dengan komposisi tertentu. Banyak peneliti menggunakan pembagian sebanyak 10. Artinya dalam penelitian tersebut dataset dibagi menjadi 10 bagian dengan 1 bagian dijadikan sebagai *data testing* dan 9 bagian yang lain menjadi *data training*. Proses seperti itu berlanjut sampai dengan keseluruhan dataset memiliki kesempatan untuk menjadi *data testing*. Artinya dalam hal ini percobaan dilakukan sebanyak 10 kali. Validasi yang seperti ini disebut juga dengan *10folds cross validation* [23].

1.2.5 Confusion Matrix

Confusion Matrix adalah sebuah matrix yang merepresentasikan hasil dari sebuah klasifikasi. Dalam *confussion matrix* terdapat nilai dari hasil klasifikasi dengan label klasifikasi sebenarnya. Nilai inilah yang nantinya dapat dijadikan sebagai acuan untuk menghitung tingkat akurasi sebuah metode algoritma. Hasil akurasi algoritma ini diwujudkan dalam sebuah tabel [24]. Untuk mengetahui tingkat akurasi sebuah algoritma dapat digunakan rumus sederhana sebagai berikut:

$$\text{Accuracy} = \frac{\text{klasifikasi yg sesuai}}{\text{total record yang ada}}$$

Klasifikasi yang sesuai adalah jumlah *record* dimana dalam klasifikasi digolongkan dalam kelas yang sesuai dengan label aslinya. Sedangkan pembagiannya adalah jumlah *record* dari keseluruhan data.

2 Metode Penelitian

Dalam penelitian ini metode yang digunakan adalah eksperimen dengan beberapa tahapan seperti berikut:

2.1 Pengumpulan Data

Tahapan penelitian yang pertama dalam penelitian ini adalah pengumpulan data. Dalam tahap ini digunakan data *public* yaitu data *credit approval* dengan rincian dataset sebagai berikut. Dataset ini memiliki 766 *record*. Dengan golongan yaitu 210 *record* tergolong dalam klasifikasi kredit lancar dan sisanya yaitu 556 *record* tergolong dalam klasifikasi kredit macet. Dalam dataset ini terdapat 16 atribut dengan satu diantaranya adalah atribut *id* serta satu lagi merupakan atribut label yaitu status kredit.

2.2 Desain Eksperimen dan Pengujian Algoritma

Penelitian ini dilakukan dengan bantuan aplikasi rapid miner. Artinya dalam pemodelan pohon keputusan menggunakan algoritma C4.5 ini tidak sepenuhnya dilakukan dengan *coding*.

2.2.1 Perangkat Keras yang Digunakan

Dalam menunjang penelitian ini digunakan perangkat keras dengan spesifikasi yang sesuai agar kinerja dalam proses perhitungan tidak terlalu lama. Adapun perangkat keras yang digunakan adalah sebagai berikut: Processor intel core i3 3220 dengan frekuensi 3,3GHz. Sedangkan RAM yang digunakan sebesar 4GB dengan VGA intel HD Series.

2.2.2 Perangkat Lunak yang Digunakan

Dalam penelitian ini digunakan perangkat lunak untuk melakukan komputasi algoritma. Adapun perangkat lunak yang digunakan adalah sebagai berikut: Sistem operasi windows 7 dengan *software* bantu yaitu rapid miner. Rapid miner merupakan salah satu aplikasi yang banyak digunakan dalam perhitungan algoritma data mining. Aplikasi ini mudah digunakan untuk pemula sekalipun karena hanya tinggal *drag and drop* algoritma dan data. Aplikasi ini dipilih karena termasuk dalam golongan aplikasi gratis.

2.2.3 Tahap Eksperimen

Dalam tahapan ini akan dilakukan perhitungan untuk mengetahui nilai gain yang tertinggi dalam dataset. Kemudian nilai gain ini akan dijadikan sebagai atribut akar dalam pemodelan pohon keputusan klasifikasi persetujuan kredit. Selain dari itu dalam

tahapan ini juga akan dilakukan perhitungan tingkat akurasi dari model pohon keputusan terhadap klasifikasi persetujuan kredit tersebut.

2.2.3.1 Validasi

Validasi dalam sebuah penelitian adalah proses yang harus dilakukan. Dalam proses validasi umumnya digunakan beberapa cara agar mendapatkan hasil yang paling maksimal. Penelitian ini tergolong dalam penelitian data mining. Dalam penelitian data mining pada umumnya dilakukan proses validasi untuk membagi dataset menjadi beberapa bagian yang nantinya digunakan sebagai data *training* dan sebagian lainnya digunakan sebagai data *testing*. Dalam penelitian klasifikasi banyak peneliti menggunakan *cross validation* untuk tahap validasi. Selain *cross validation* sebenarnya ada *split validation* untuk proses ini. Dalam penelitian ini akan digunakan *10folds cross validation* untuk tahap validasinya.

2.2.3.2 Pengukuran akurasi algoritma

Dalam sebuah penelitian data mining hasil akhir yang diharapkan adalah tingkat akurasi yang baik. Untuk mengetahui kemampuan algoritma serta tingkat akurasi sebuah metode digunakan metode pengukuran tersendiri [26]. Banyaknya metode pengukuran yang dapat digunakan membuat penelitian harus jeli dan tepat dalam pemilihan metode tersebut.

Dalam penelitian ini digunakan *confussion matrix* untuk menghitung tingkat akurasi algoritma. *Confussion matrix* yang digunakan dalam penelitian ini memanfaatkan aplikasi rapid miner. Dikarenakan dalam validasi penelitian ini telah digunakan *10folds cross validation* artinya pengukuran akurasi algoritma dalam penelitian ini dilakukan sebanyak 10 kali. Sedangkan hasil tingkat akurasi didapatkan dengan melakukan perhitungan rata-rata dari keseluruhan percobaan.

2.3 Evaluasi Hasil

Setelah penelitian berhasil dilakukan terhadap data persetujuan kredit dengan menggunakan *rapid miner*. Maka langkah berikutnya adalah mencatat hasil dari akurasi algoritma C4.5 tersebut untuk data persetujuan kredit. Berikutnya dibangun sebuah pohon keputusan yang mewakili algoritma C4.5 tersebut dalam pemodelan aturan aturan yang ada terhadap atribut terkait. Pohon keputusan tersebut yang nantinya akan digunakan sebagai aturan dalam persetujuan kredit untuk nasabah baru

3 Hasil dan Pembahasan

3.1 Perolehan nilai gain algoritma C4.5

Dalam tahapan ini dilakukan perhitungan terhadap semua atribut untuk mengetahui nilai gain terbesar. Nilai gain terbesar nantinya akan digunakan sebagai atribut akar dalam model pohon keputusan yang ada. Sebelumnya juga dihitung nilai entropy total guna perhitungan nilai gain. adapun nilai information gain yang terbentuk dalam

penelitian ini adalah sebagai berikut sesuai dengan tabel 1 berikut. Setelah nilai *information gain* untuk setiap atribut diketahui maka nilai *split information* digunakan untuk membagi nilai tersebut agar didapati nilai *gain ratio*. Tabel 2 merupakan nilai *gain ratio* untuk setiap atribut.

Tabel 1. Nilai *information gain* semua atribut

Atribut	Nilai <i>information gain</i>
type pinjam an	0.0
bi golongan debitur	0.002368891105240975
bi sektor ekonomi	0.0043600944984576095
jenis pinjam an	0.0047529360139100295
jenis kelamin	0.007051442147994892
umur	0.023118544205183927
jkw	0.02675422888515543
saldo nominatif	0.05787698281584561
angsuran per_bulan	0.15053811852727494
jumlah pinjaman	0.21870668937148233
col	0.23436926074644862
bi gol penjamin	0.24535443936313503
tunggakan bunga	0.29654338474401754
tunggakan pokok	1.0

Tabel 2. Nilai *gain ratio* semua atribut

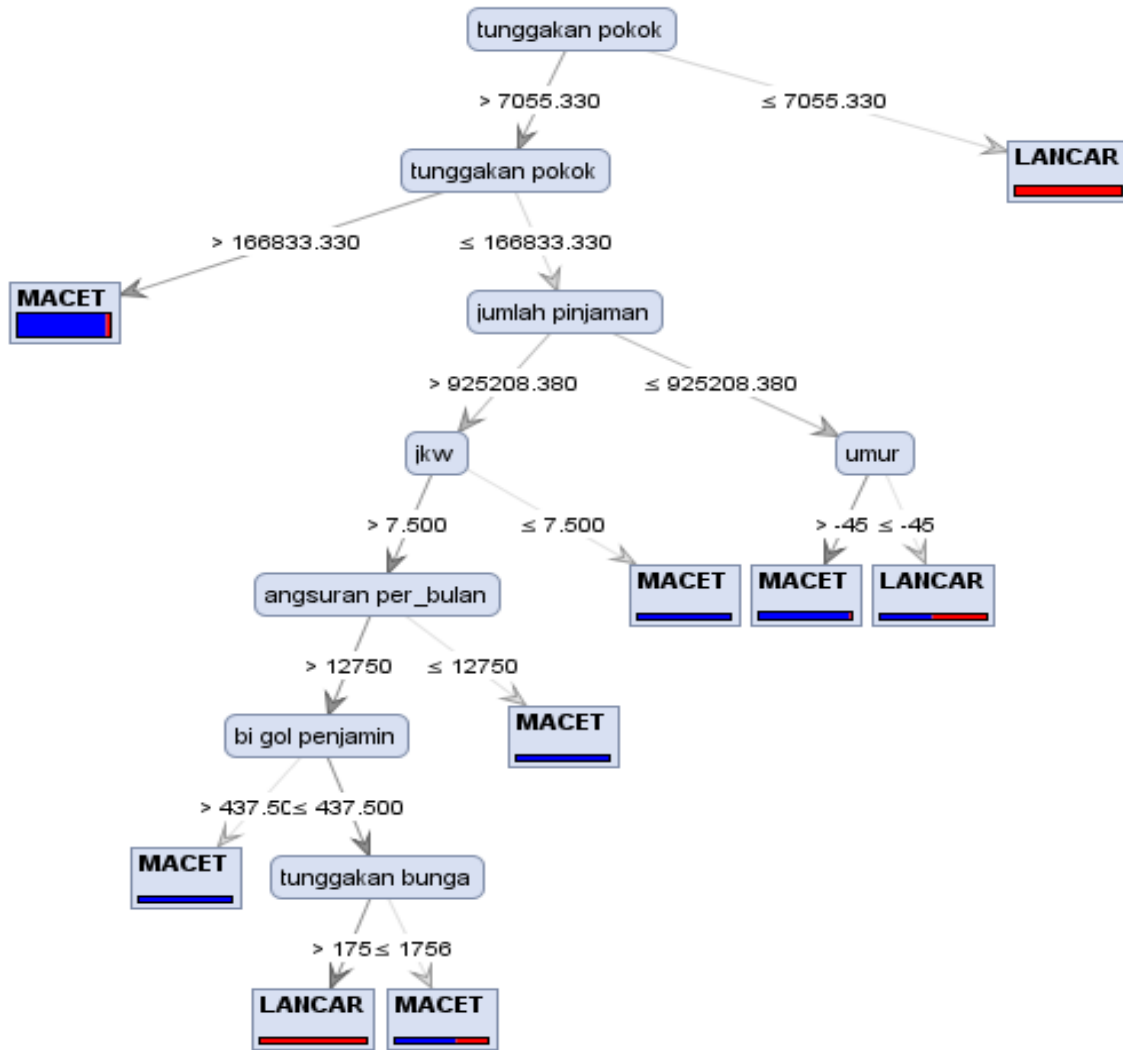
Atribut	Nilai <i>gain ratio</i>
type pinjam an	0.0
jenis pinjam an	0.0
bi golongan debitur	0.0
jenis kelamin	0.005121734949285033

angsuran per_bulan	0.037410550386613825
saldo nominatif	0.059173930645393766
umur	0.07839174374125772
jkw	0.07839174374125772
bi sektor ekonomi	0.08332029052166456
jumlah pinjaman	0.1800768079024675
bi gol penjamin	0.1906877240907161
col	0.2192153592231016
tunggakan bunga	0.22407132463764518
tunggakan pokok	1.0

3.2 Hasil Model Pohon Keputusan

Setelah semua nilai *gain ratio* diketahui maka selanjutnya dibangun sebuah model pohon keputusan untuk memudahkan dalam menentukan klasifikasi. Pohon keputusan ini merupakan sebuah model yang populer dan banyak digunakan karena mudah dipahami dalam bahasa alami manusia.

Gambar 1. Model Pohon Keputusan Persetujuan Kredit



Dari model pohon keputusan sesuai gambar 1 dapat disimpulkan bahwa node akar adalah tunggakan pokok. Jika tunggakan pokok diatas 7055.330 maka klasifikasi persetujuan kredit tergolong dalam klasifikasi kredit lancar. Kemudian jika tunggakan pokok kurang dari 166833.330 dengan jumlah pinjaman kurang dari atau sama dengan 925208.380 dengan umur lebih dari 45 maka klasifikasinya tergolong dalam kredit macet. Selebihnya dari pohon keputusan tersebut dapat diikuti karena model yang tercipta sangat mudah dipahami dengan bahasa alami manusia.

3.3 Akurasi Algoritma C4.5 untuk Persetujuan Kredit

Untuk mengukur tingkat akurasi klasifikasi digunakan *cross validation* untuk proses validasinya. Kemudian untuk pengukurannya digunakan *confussion matrix*. Dalam penelitian ini digunakan *10folds cross validation* yaitu membagi dataset menjadi 10 bagian untuk nantinya 1 bagian dijadikan sebagai data *testing* dan 9 bagian lainnya dijadikan sebagai data *training*.

Untuk pengukuran *confussion matrix* digunakan untuk mengetahui klasifikasi yang sesuai dengan aslinya dan klasifikasi yang tidak sesuai dengan klasifikasi sebenarnya. Tingkat akurasi didapatkan dari jumlah *record* dengan hasil klasifikasi yang sesuai dengan sebenarnya dibagi dengan jumlah keseluruhan *record* data. Tingkat akurasi ini nantinya dihitung untuk 10 kali percobaan sesuai dengan yang digunakan dalam proses validasi. Sehingga nantinya didapatkan nilai akurasi yaitu rata-rata dari keseluruhan tingkat akurasi yang ada.

Tabel 3 merupakan tabel *confussion matrix* dari klasifikasi persetujuan kredit yang terbentuk dari perhitungan setelah dilakukan validasi. Nilai akurasi didapatkan dari perhitungan $(540+184)$ dibagi dengan $(540+26+16+184)$. Hasilnya adalah $724/766 = 94,516\%$

Tabel 1. *Confussion Matrix*

	Benar MACET	Benar LANCAR	presisi
Prediksi MACET	540	26	95,41%
Prediksi LANCAR	16	184	92,00%
Kelas recall	97,12%	87,62%	

4 Kesimpulan dan Saran

4.1 Kesimpulan

Dari hasil penelitian yang telah dilakukan dengan melakukan klasifikasi persetujuan kredit menggunakan algoritma C4.5 maka dapat diambil kesimpulan sebagai berikut:

1. Algoritma C4.5 dapat memodelkan pohon keputusan untuk klasifikasi persetujuan kredit. Model pohon keputusan yang terbentuk dapat dengan mudah merepresentasikan aturan dari klasifikasi persetujuan kredit.
2. Tingkat akurasi klasifikasi persetujuan kredit dengan menggunakan algoritma C4.5 sebesar 94,516% dan tergolong dalam *best classification*

4.2 Saran

Dalam penelitian ini masih terdapat beberapa kekurangan yang dapat digunakan sebagai alat untuk evaluasi diri dan untuk perbaikan penelitian berikutnya. Untuk penelitian berikutnya dapat dibuat sebuah aplikasi untuk menentukan kelayakan persetujuan kredit untuk membantu pihak bank dalam memilih calon nasabah

5 Referensi

- [1] B. Dwi Cahyani, "Sistem Pendukung Keputusan Persetujuan Permohonan Kredit Pinjaman pada Bank Muamalat Indonesia Cabang Yogyakarta," *Sekol. Tinggi Manaj. Inform. dan Komput. AMIKOM Yogyakarta*, 2013.
- [2] Y. Suhari, M. Sukur, and S. Eniyati, "SISTEM PENDUKUNG KEPUTUSAN PEMBERIAN KREDIT PADA PT . BPR ARTAMANUNGGAL ABADI MRANGGEN," *Din. Inform. Fak. Teknol. Inf. Univ. Stikubank Semarang*, vol. 1, no. 1, 2009.
- [3] I. Melissa and R. S. Oetama, "Analisis Data Pembayaran Kredit Nasabah Bank Menggunakan Metode Data Mining," vol. IV, no. 1, pp. 18–27, 2013.
- [4] A. Ginanjar Mabrur and R. Lubis, "Penerapan Data Mining untuk Memprediksi Kriteria Nasabah Kredit," *J. Komput. dan Inform. (KOMPUTA)*, vol. 1, 2012.
- [5] A. Zahid, "Sistem Pendukung Keputusan Persetujuan Penerimaan Pinjaman di PD.BPR BKK Lasem dengan Menggunakan Metode Bayes," *Sekol. Tinggi Manaj. Inform. dan Komput. AMIKOM Yogyakarta*, 2013.
- [6] F. Azwany, "Sistem Pendukung Keputusan Pemberian Kredit Usaha Rakyat pada Bank Syariah Mandiri Cabang Medan menggunakan Metode AHP," *Progr. Stud. Ilmu Komput. Dep. Ilmu Komput. Fak. Mat. dan Ilmu Pengetah. Alam Univ. Sumatera Utara Medan*, 2010.
- [7] Y. Y. W, F. R. Pratikto, and A. S. Vivianne, "EVALUASI PEMOHON KREDIT MOBIL DI PT ' X ' DENGAN MENGGUNAKAN TEKNIK DATA MINING DECISION TREE," *Simp. Nas. RAPI VIII*, pp. 42–49, 2009.
- [8] A. Wahyu Oktaputra and E. Noersasongko, "Sistem Pendukung Keputusan Kelayakan Pemberian Kredit Motor Menggunakan Metode Simple Additive Weighting pada Perusahaan Leasing HD Finance," *Progr. Stud. Sist. Inf. - S1, Fak. Ilmu Komput. Univ. Dian Nuswantoro, Semarang*, pp. 1–9, 2014.
- [9] X. Wu, V. Kumar, J. Ross Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, *Top 10 algorithms in data mining*, vol. 14, no. 1. 2007, pp. 1–37.
- [10] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques 3rd Edition*. Elsevier, 2011.
- [11] D. T. Larose, *Discovering Knowledge in Data: an Introduction to Data Mining*. John Wiley & Sons, 2005.

- [12] B. Santosa, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Edisi Pert. Yogyakarta: Graha Ilmu, 2007.
- [13] A. Christobel and D. . Sivaprakasam, “An Empirical Comparison of Data Mining Classification Methods,” vol. 3, no. 2, pp. 24–28, 2011.
- [14] A. H. M. Ragab, A. Y. Noaman, A. S. Al-Ghamdi, and A. I. Madbouly, “A Comparative Analysis of Classification Algorithms for Students College Enrollment Approval Using Data Mining,” *Proc. 2014 Work. Interact. Des. Educ. Environ. - IDEE '14*, pp. 106–113, 2014.
- [15] D. Sugianti, “Algoritma Bayesian Classification untuk Memprediksi Heregistrasi Mahasiswa Baru di STMIK Widya Pratama,” no. 2, pp. 1–5, 2012.
- [16] K. Hastuti, “Analisis komparasi algoritma klasifikasi data mining untuk prediksi mahasiswa non aktif,” vol. 2012, no. Semantik, pp. 241–249, 2012.
- [17] T. H. Pudjianto, F. Renaldi, and A. Teogunadi, “Penerapan data mining untuk menganalisa kemungkinan pengunduran diri calon mahasiswa baru,” 2011.
- [18] Kusriani, S. Hartati, R. Wardoyo, and A. Harjoko, “Perbandingan metode nearest neighbor dan algoritma c4.5 untuk menganalisis kemungkinan pengunduran diri calon mahasiswa di stmik amikom yogyakarta,” vol. 10, no. 1, 2009.
- [19] A. Ashari, I. Paryudi, and A. M. Tjoa, “Performance Comparison between Naïve Bayes , Decision Tree and k-Nearest Neighbor in Searching Alternative Design in an Energy Simulation Tool,” vol. 4, no. 11, pp. 33–39, 2013.
- [20] D. Widiastuti, “Analisa Perbandingan Algoritma SVM, Naïve Bayes, dan Decision Tree dalam Mengklasifikasikan Serangan (Attack) pada Sistem Pendeteksi Intrusi,” *Jur. Sist. Inf. Univ. Gunadarma*, pp. 1–8, 2007.
- [21] J. Han and M. Kamber, “Data Mining: Concepts and Techniques Second Edition,” vol. 40, no. 6, p. 9823, Mar. 2006.
- [22] Kusriani and L. E. Taufiq, *Algoritma Data Mining*. Yogyakarta: Andi Offset, 2009.
- [23] Ian H Witten. Eibe Frank. Mark A Hall, *Data Mining 3rd*. 2011.
- [24] F. Gorunescu, *Data Mining: Concepts; Models and Techniques*. Springer, 2011.
- [25] F. Gorunescu, *Data Mining: Concept, Models and Techniques*, Vol 12. Berlin: Heidelberg: Springer Berlin Heidelberg, 2011.

- [26] D. R. Amancio, C. H. Comin, D. Casanova, G. Travieso, O. M. Bruno, F. a. Rodrigues, and L. D. F. Costa, "A systematic comparison of supervised classifiers," Oct. 2013.