

Implementasi Algoritma C4.5 Klasifikasi Nasabah Potensial ADIRA Dinamika Multi Finance Pringsewu

Hidayatus Sholihah¹⁾, Fiqih Satria²⁾, Muhammad Muslihudin³⁾

STMIK Pringsewu

Jl. Wisma Rini No. 09 Pringsewu, (0729) 22240

e-mail: dayhidas25@gmail.com¹, fiqih.satria@gmail.com², muslihudinstmikpsw@gmail.com³

Abstrak

Salah satu keberhasilan perbankan adalah mengklasifikasi nasabahnya, karena nasabah merupakan aset terpenting dalam suatu perusahaan. Adira dinamika multi finance adalah perusahaan yang bergerak di bidang pembiayaan dana secara kredit, memberikan penawaran kembali terhadap nasabah potensial yang sebelumnya telah melakukan pembiayaan kredit merupakan suatu langkah untuk tetap mendapatkan nasabah kembali. Mengetahui nasabah yang berpotensi merupakan hal yang harus diketahui seorang analis kredit. Analisis kredit dengan teknik data mining perlu dilakukan untuk mempersingkat waktu analisis pemberian kredit. Penelitian ini membentuk model decision tree C4.5 untuk klasifikasi nasabah kredit berdasarkan data yang sudah ada di Adira Finance. Preprocessing data terdiri dari data cleaning, data integration, data selection, dan data transformasi dilakukan untuk meningkatkan kualitas model klasifikasi. Proses pembentukan model decision tree C4.5 dilakukan menggunakan bantuan software weka.

Kata kunci: Data mining, Kredit, Adira Finance, Decision tree C4.5

1. Pendahuluan

Dengan pertumbuhan ekonomi yang secara pesat, kebutuhan akan informasi yang akurat sangat dibutuhkan dalam kehidupan sehari-hari. Teknologi informasi menjadi kebutuhan sekunder bagi para penggunanya, sehingga informasi akan menjadi suatu elemen penting dalam perkembangan masyarakat untuk saat ini dan waktu yang akan mendatang. Namun kebutuhan akan informasi yang tinggi kadang tidak diimbangi dengan penyajian informasi yang memadai, seringkali informasi harus di gali ulang dari data yang jumlahnya sangat besar. Penggunaan teknik data mining diharapkan dapat memberikan pengetahuan-pengetahuan yang tersembunyi, sehingga dapat menjadi informasi yang berharga.

Berdasarkan penelitian^[1] Dengan beberapa algoritma klasifikasi yang diuji terhadap data training diberikan model klasifikasi algoritma C4.5 yang mempunyai nilai akurasi yang tertinggi. Setelah diimplementasikan kepada data tes, didapatkan keputusan nasabah yang ditolak dan diterima kreditnya. Penerapan Algoritma C4.5 juga pernah diimplementasikan pada Nasabah Bank di Tasikmalaya dengan hasil klasifikasi kelayakan pemberian kredit nasabah dengan jumlah 8 atribut menghasilkan akurasi 87.36% merupakan tingkat akurasi yang baik dibandingkan dengan hasil penelitian klasifikasi kelayakan pemberian kredit nasabah yang menggunakan 4 atribut menghasilkan akurasi 79.50 %^[2]. Penelitian^[3] dengan metode keputusan decision tree menggunakan algoritma C4.5 pada calon nasabah Bank diharapkan menjadikan proses penggalian informasi lebih cepat dan optimal dengan kapasitas data yang lebih besar, sehingga kesalahan yang ditimbulkan dalam pengambilan keputusan lebih diminimalkan.

Dari beberapa penelitian di atas memungkinkan Algoritma C4.5 di terapkan pada klasifikasi nasabah potensial pada Adira dinamika multi finance yang bergerak di bidang pembiayaan dana secara kredit, memberikan penawaran kembali terhadap nasabah potensial yang sebelumnya telah melakukan pembiayaan kredit sehingga nasabah tetap mendapatkan pinjaman sesuai dengan kriteria yang ditentukan.

2. Metode Penelitian

2.1 Metode Pengumpulan Data

Dalam penelitian ini penulis metode pengumpulan data sebagai berikut :

1. Wawancara

Pada tahap ini penulis mewawancarai beberapa pihak yang terkait dalam hal ini berupa pegawai Adira Finance. Wawancara mengenai data kredit nasabah seperti pendapatan, pinjaman, angsuran, waktu dan kategori.

2. Dokumentasi

Melalui metode ini, penulis mengumpulkan dokumen, arsip dan data-data nasabah Adira Finance sebagai bahan penelitian.

3. Studi pustaka

Dalam hal ini penulis mengumpulkan dan mempelajari buku-buku maupun literatur-literatur yang berhubungan dengan masalah yang diteliti. Dari studi pustaka ini didapatkan banyak referensi yang akan mendukung penelitian ini, sekaligus untuk memperkuat pengetahuan dasar dan teori yang digunakan pada penelitian ini.

2.2 Algoritma C4.5

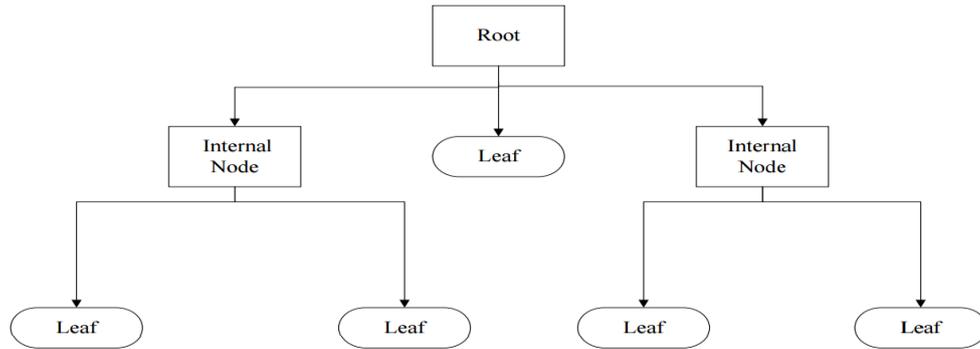
Algoritma C4.5 merupakan kelompok algoritma decision tree. Algoritma ini mempunyai input berupa training samples dan samples. Training samples berupa data contoh yang akan digunakan untuk membangun sebuah tree yang telah diuji kebenarannya. Sedangkan samples merupakan field-field data yang nantinya akan kita gunakan sebagai parameter dalam melakukan klasifikasi data. Algoritma C4.5 adalah algoritma hasil pengembangan dari algoritma ID3. Perbaikan algoritma C4.5 dari algoritma ID3 dilakukan dalam hal (Santosa, ID3):

1. Bisa mengatasi missing value
2. Bisa mengatasi data kontinyu
3. Pruning
4. Adanya aturan

Secara garis besar langkah-langkah yang dilakukan oleh algoritma C4.5 dalam membentuk pohon keputusan adalah sebagai berikut:

1. Pada saat awal pembentukan pohon akan dimulai dengan membuat suatu node yang melambangkan training sample.
2. Jika sampel-sampel tersebut memiliki kelas yang sama, maka node tersebut dijadikan leaf node dengan label kelas tersebut.
3. Jika sampel-sampel tersebut tidak memiliki kelas yang sama maka algoritma ini akan mencari gain ratio tertinggi dari atribut yang tersedia, sebagai cara untuk memilih atribut yang paling berpengaruh pada training sample yang tersedia. Nantinya atribut ini akan dijadikan atribut “penguji” atau “keputusan” pada node tersebut. Hal yang perlu diperhatikan adalah ketika atribut tersebut bernilai kontinu, maka atribut tersebut harus di diskritkan terlebih dahulu.
4. Cabang untuk setiap node akan dibentuk berdasarkan nilai-nilai yang diketahui dari atribut pengujian.
5. Algoritma ini akan terus melakukan proses yang sama rekursif untuk membentuk suatu pohon keputusan untuk setiap sample di setiap bagiannya.
6. Proses rekursif ini akan berhenti, ketika salah satu dari kondisi di bawah telah terpenuhi. Kondisi-kondisinya adalah:
 - a. Semua sample yang diberikan pada node adalah berasal dari satu kelas yang sama.
 - b. Tidak ada atribut lainnya yang dapat digunakan untuk mempartisi sample lebih lanjut.
 - c. Tidak ada sample yang memenuhi $test-attribute = .$ Dalam hal ini, sebuah daun dibuat dan dilabeli dengan kelas yang memiliki sample terbanyak (majority voting).

2.3 Model Pohon Keputusan



Gambar 1. Pohon Keputusan

Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Decision tree merupakan metode klasifikasi yang paling populer digunakan. Selain karena pembangunannya relatif cepat, hasil dari model yang dibangun mudah untuk dipahami. Pada decision tree terdapat 3 jenis node, yaitu:

1. Root Node, merupakan node paling atas, pada node ini tidak ada input dan bisa tidak mempunyai output atau mempunyai output lebih dari satu.
2. Internal Node, merupakan node percabangan, pada node ini hanya terdapat satu input dan mempunyai output minimal dua.
3. Leaf node atau terminal node, merupakan node akhir, pada node ini hanya terdapat satu input dan tidak mempunyai output. Contoh dari pohon keputusan dapat dilihat di Gambar 1.

3. Hasil dan Pembahasan

3.1 Data Penelitian

Data penelitian ini menggunakan data nasabah kredit Adira Finance. Setelah melakukan serangkaian tahapan data mining Penelitian ini menggunakan 100 dataset dengan 6 atribut *predictor* Atribut *predictor* yang digunakan dalam data kredit adalah Umur, Jenis Kelamin, Status Perkawinan, Pekerjaan, Pendapatan dan Tagihan Listrik.

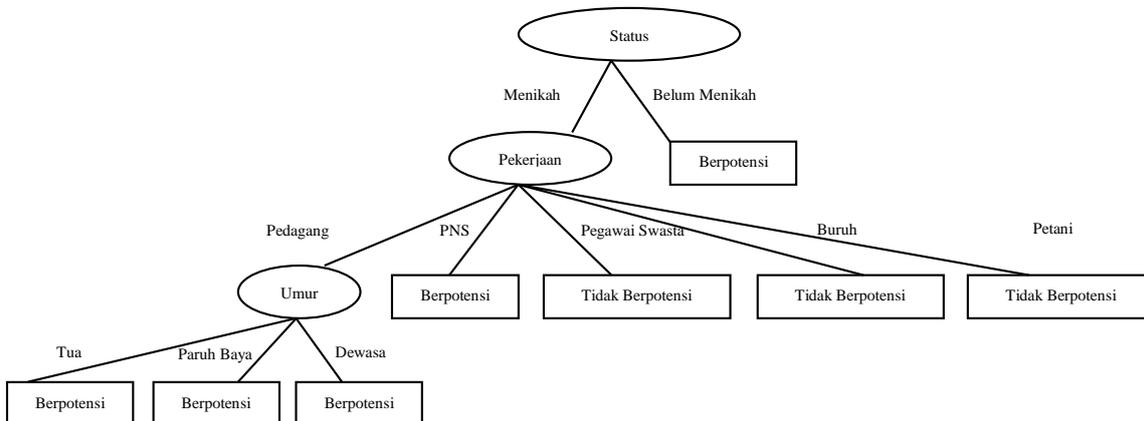
nama	Alamat	Umur	Jenis Kelamin	Status Perkawinan	Pekerjaan	Penghasilan	Tagihan Listrik	Klasifikasi
paino	Fajar Agung	Tua	Laki-laki	menikah	pedagang	Besar	Sedang	Berpotensi
mistam	Bumiayu	Paruh Baya	Laki-laki	menikah	PNS	Besar	Sedang	Berpotensi
joko prase	Pringsewu Timur	Dewasa	Laki-laki	Belum Menikah	Pegawai swasta	Besar	Sedang	Tidak berpotensi
karsono	pringsewu Selatan	Paruh Baya	Laki-laki	menikah	Pegawai swasta	Besar	Sedang	Tidak berpotensi
teguh arsi	pajarisuk	Dewasa	Laki-laki	Belum Menikah	Pegawai swasta	Besar	Sedang	Berpotensi
suyanto	Pringsewu Timur	Dewasa	Laki-laki	menikah	Pegawai swasta	Besar	Sedang	Tidak berpotensi
shinta me	Bumiarum	Dewasa	perempuan	menikah	Pegawai swasta	Besar	Sedang	Tidak berpotensi
malik dart	Rejosari	Paruh Baya	Laki-laki	menikah	PNS	Besar	Sedang	Tidak berpotensi
koko supa	Sidoarjo	Dewasa	Laki-laki	Belum Menikah	Pegawai swasta	Besar	Sedang	Berpotensi
martini	Margakaya	Dewasa	perempuan	menikah	pedagang	Besar	Sedang	Tidak berpotensi
ngadiso	pajarisuk	Paruh Baya	Laki-laki	menikah	PNS	Besar	Sedang	Berpotensi
badriyah	Pringsewu Barat	Paruh Baya	perempuan	menikah	pedagang	Besar	Sedang	Berpotensi
poniyah	Pringsewu Utara	Paruh Baya	perempuan	menikah	Buruh	Besar	Sedang	Tidak berpotensi
agus riyad	Podomoro	Dewasa	Laki-laki	Belum Menikah	Pegawai swasta	Besar	Sedang	Berpotensi
nasrudin	Margakaya	Dewasa	Laki-laki	Belum Menikah	pedagang	Besar	Sedang	Berpotensi
siswantor	podosari	Dewasa	Laki-laki	Belum Menikah	PNS	Besar	Sedang	Berpotensi
tuti suda	Pringsewu Utara	Dewasa	perempuan	Belum Menikah	PNS	Besar	Sedang	Berpotensi
Tarsani	Pringsewu Barat	Paruh Baya	perempuan	menikah	Buruh	Besar	Sedang	Tidak berpotensi
gatot muk	Pringsewu Timur	Paruh Baya	Laki-laki	menikah	Pegawai swasta	Besar	Sedang	Tidak berpotensi
tukiran	podosari	Tua	Laki-laki	menikah	Buruh	Besar	Sedang	Tidak berpotensi
murdimar	pajarisuk	Tua	Laki-laki	menikah	Pegawai swasta	Besar	Sedang	Tidak berpotensi
dwi cayok	Podomoro	Dewasa	Laki-laki	Belum Menikah	Pegawai swasta	Besar	Kecil	Berpotensi

Gambar 2. Attribute Data Kredit Nasabah

3.2 Decision Tree Menggunakan WEKA

Untuk mencari knowledge dengan algoritma C4.5 maka dilakukan langkah-langkah sebagai berikut :

1. Pada jendela utama WEKA klik tombol *explorer*. Lalu muncul jendela WEKA Explorer, kemudian pada *tab Preprocess* klik tombol *open file* untuk membuka data.
2. Selanjutnya keluar kotak dialog WEKA dan klik tombol *All* untuk mengeksekusi data.
3. Selanjutnya pada kotak dialog WEKA, aktifkan *tab classify*. Lalu klik tombol *Choose* pada *tab Classify* untuk mencari algoritma yang digunakan.
4. Pada menu *classify* pilih *Root Tree* dan klasifikasi dengan pohon keputusan, yaitu metode klasifikasi dengan struktur pohon yang merepresentasikan kriteria pembagian kelas dan kelas-kelas dengan menggunakan metode algoritma J48.
5. Setelah *classifiers*, pilihan-pilihan pengujian, dan *class* telah ditentukan, proses pembelajaran dapat dimulai dengan mengklik tombol *Start*, *user* dapat menghentikan proses ini sewaktu-waktu dengan tombol *Stop*. Saat *training* selesai, area *output classifiers* di sebelah kanan menampilkan *teks* yang menggambarkan hasil *training* dan pengujian. Sebuah *entry* baru saja muncul di kotak *Result list*.
6. Untuk membuat pohon keputusan maka klik kanan *trees-J.48* pada menu *result list* kemudian pilih *Visualize tree*.
7. Hasil *Visualize Tree* adalah hasil akhir dari pohon keputusan.



Gambar 3. Pohon keputusan

Di mana *rule* yang dihasilkan pada gambar 3 adalah :

1. IF status perkawinan me nikah pekerjaan pedagang umur Tua, THEN Berpotensi
2. IF status perkawinan menikah pekerjaan pedagang umur Paruh Baya, THEN Berpotensi
3. IF status perkawinan me nikah pekerjaan pedagang umur Dewasa, THEN Tidak Berpotensi
4. IF Menikah pekerjaan PNS, THEN Berpotensi
5. IF Menikah pekerjaan Pegawai Swasta, THEN Tidak Berpotensi
6. IF Menikah pekerjaan Buruh, THEN Tidak Berpotensi
7. IF Menikah pekerjaan Petani, THEN Tidak Berpotensi
8. IF Status Perkawinan belum menikah, THEN Berpotensi

3.3 Pembahasan Hasil Pengujian

Dari pohon keputusan dan *rule* yang terbentuk di mana status berpotensi dan tidak berpotensi dilihat dari atribut yang ada pada data awal yaitu data nasabah Adira Finance dan hasilnya atribut Status Perkawinan yang menjadi akar, pekerjaan dan umur menjadi dahan atau node . rincian jumlah nasabah yang berpotensi dan tidak berpotensi terlihat pada tabel berikut :

Tabel 1. Rincian Jumlah Berdasarkan Rule

No	Rule	Jumlah	
		Berpotensi	Tidak Berpotensi
1	Menikah, Pedagang, Tua	6	
2	Menikah, Pedagang, Paruh Baya	12	
3	Menikah, Pedagang, Dewasa		5
4	Menikah, PNS	6	
5	Menikah, Pegawai Swasta		18
6	Menikah, Petani		2
7	Menikah, Buruh		20
8	Belum Menikah	31	
Jumlah Nasabah		55	45

4. Simpulan

4.1 Kesimpulan

1. Analisis klasifikasi pada data kredit Adira Finance berhasil dilakukan dengan tahapan pengumpulan data, *preprocessing* yang terdiri dari *data cleaning*, *data integration*, *data selection*, dan *data transformation*, proses klasifikasi menggunakan metode *decision tree* C4.5.
2. Metode pohon keputusan (*decision tree*) dengan *software* Weka dapat mengidentifikasi kelayakan kredit dengan baik

4.2 Saran

1. Penelitian selanjutnya dapat menambahkan variabel-variabel yang memiliki hubungan dengan masalah pengajuan perpanjangan kredit agar nantinya tingkat pemberian keputusan yang lebih baik dapat tercapai.
2. Penelitian ini tidak mengukur tingkat akurasi hasil algoritma C4.5 karena peneliti tidak membandingkan *software* data mining, melainkan hanya mengimplementasikan *software* Weka untuk mengklasifikasikan data nasabah Adira Finance yang berpotensi dan tidak berpotensi, sehingga peneliti selanjutnya dapat melakukan perbandingan dengan *software* lain dan menghitung tingkat akurasinya.

Daftar Pustaka

- [1] H. Marcos and I. Hidayah, "Implementasi Data Mining Untuk Klasifikasi Nasabah Kredit Bank ' X ' Menggunakan Classification Rule," *Semin. Nas. Teknol. Inf. dan Multimed.* 2014, pp. 1–7, 2014.
- [2] C. R. H. Susanto, Neneng Sri Uryani, "Penerapan algoritma c4.5 untuk menganalisis kelayakan pemberian kredit nasabah," *STMIK TasikMalaya*, vol. 4, no. 1, pp. 1–7, 2013.
- [3] L. N. Rani, "Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit," *J. KomTekInfo Fak. Ilmu Komput.*, vol. 2, no. 2, pp. 33–38, 2015.
- [4] E. Turban, J. E. Aronson, and T.-P. Liang, "Decision Support Systems and Intelligent Systems," *Decis. Support Syst. Intell. Syst.*, vol. 7, p. 867, 2007.
- [5] Y. Zhang, S. Fong, J. Fiaidhi, and S. Mohammed, "Real-time clinical decision support system with data stream mining.," *J. Biomed. Biotechnol.*, vol. 2012, p. 580186, 2012.
- [6] E. Turban, R. Sharda, and D. Delen, *Decision Support and Business Intelligence Systems. Chapter 6 Artificial Neural Networks for Data Mining*, vol. 8th. 2007.
- [7] M. Muslihudin, "Analisis Prediksi Mahasiswa Tidak Tepat Waktu Menyelesaikan Studi Dengan Menggunakan Metode Algoritma C 4.5 (Studi Kasus : STMIK Pringsewu)," *Tesis IBI Darmajaya*, pp. 5–29, 2015.
- [8] Y. Mardi, "Klasifikasi Menggunakan Algoritma C4.5," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.
- [9] A. Srivastava, E.-H. S. H. E.-H. S. Han, V. Singh, and V. Kumar, "Parallel formulations of decision-tree classification algorithms," *Proceedings. 1998 Int. Conf. Parallel Process. (Cat. No.98EX205)*, vol. 24, pp. 1–24, 1998.
- [10] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques Third Edition*. 2012.