

# Prediksi Tingkat Risiko Kredit dengan Data Mining Menggunakan Algoritma *Decision Tree C.45*

Nurdiana Handayani<sup>1,\*</sup>, Herry Wahyono<sup>2</sup>, Joko Trianto<sup>3</sup>, Dwi Sidik Permana<sup>4</sup>

<sup>1</sup> Teknik Informatika, Universitas Muhammadiyah Tangerang, Tangerang, Indonesia

<sup>2</sup> Teknik Informatika, Universitas Krisnadwipayana, Jakarta Timur, Indonesia

<sup>3</sup> Teknik Informatika, Sekolah Tinggi Teknologi Informasi NIIT, Jakarta Selatan, Indonesia

<sup>4</sup> Teknik Informatika, Institut Bisnis & Informatika Kosgoro 1957, Jakarta Selatan, Indonesia

Email: <sup>1,\*</sup>nurdiana.handayani@ft-umt.ac.id, <sup>2</sup>wahyonos2000@unkris.ac.id, <sup>3</sup>joko.trianto@i-tech.ac.id, <sup>4</sup>dwi770@gmail.com

Email Penulis Korespondensi: nurdiana.handayani@ft-umt.ac.id

Submitted 25-09-2021; Accepted 05-12-2021; Published 30-12-2021

## Abstrak

Perusahaan pembiayaan dalam memberikan kredit melakukan analisis data terlebih dahulu untuk mengurangi risiko kredit. Ketika nasabah tidak membayar kredit secara lancar maka akan merugikan perusahaan. Untuk itu, analisis kredit merupakan faktor penting untuk meminimalkan risiko keuangan. Maka, dibutuhkan analisa prediksi tingkat risiko kredit berdasarkan dari data atau berkas dari nasabah. Penelitian ini bertujuan untuk melakukan prediksi tingkat risiko kredit dengan data mining menggunakan algoritma pohon keputusan C.45. Prediksi tingkat risiko yang digunakan terdapat dua kelas, yaitu lancar dan tidak lancar. Algoritma pohon keputusan C.45 memiliki fungsi untuk menemukan pengetahuan atau pola-pola kesamaan karakteristik dalam suatu kelompok atau kelas tertentu. Pada penelitian ini algoritma C.45 diimplementasikan dan dianalisa menggunakan aplikasi WEKA. Dari hasil evaluasi menggunakan *confusion matrix* didapatkan akurasi yang dihasilkan untuk 1.153 data training dengan 91 data testing dan enam atribut yang digunakan menghasilkan akurasi sebesar 79%.

**Kata Kunci:** C.45; Confusion Matrix; Data Mining; Pohon Keputusan; Risiko Kredit

## Abstract

Finance companies in providing credit conduct data analysis first to reduce credit risk. When customers do not pay credit smoothly, it will harm the company. For this reason, credit analysis is an important factor to minimize financial risk. So, it takes a predictive analysis of the level of credit risk based on data or files from customers. This study aims to predict the level of credit risk with data mining using the C.45 decision tree algorithm. There are two classes of risk level predictions, namely current and non-current. The C.45 decision tree algorithm has a function to find knowledge or patterns of characteristic similarity in a particular group or class. In this study, the C.45 algorithm was implemented and analyzed using the WEKA application. From the results of the evaluation using the confusion matrix, the accuracy generated for 1,153 training data with 91 testing data and the six attributes used produces an accuracy of 79%.

**Keywords:** C.45; Confusion Matrix; Data Mining; Decision Tree; Credit Risk

## 1. PENDAHULUAN

Adanya lembaga pembiayaan ini, masyarakat bisa memanfaatkan lembaga tersebut untuk mendapatkan pinjaman dana yang dapat digunakan untuk bantuan pembiayaan konsumen dalam memenuhi kebutuhan sehari-hari atau untuk mengembangkan sebuah usaha. *Customer finance* merupakan salah satu alternatif untuk penunjang bisnis yang dapat dimanfaatkan sebagai sumber pendanaan [1]. Melalui lembaga pembiayaan, masyarakat bisa memanfaatkan lembaga tersebut untuk memperoleh pinjaman dana yang dapat digunakan untuk bantuan pembiayaan konsumen dalam memenuhi kebutuhan sehari-hari atau untuk mengembangkan sebuah usaha. PT. WOM Finance, merupakan perusahaan yang memiliki salah satu jasanya terkait pembiayaan dana. Seperti perusahaan-perusahaan *customer finance* yang lain, PT. WOM Finance melaksanakan prinsip-prinsip yang berlaku dalam pelaksanaan sistem kredit. Salah satu produk yang memiliki banyak konsumen pada PT. WOM Finance adalah pinjaman dana dengan jaminan BPKP. PT. WOM Finance memiliki standarisasi dalam penerimaan dan penolakan calon nasabah yang akan melakukan kredit. Dalam memberikan kredit, perusahaan ini melaksanakannya melalui analisis data dari calon konsumen yang dilakukan oleh penyelia. Analisis kredit merupakan faktor penting untuk meminimalkan risiko keuangan, sehingga perlu melalui analisis. Ini dapat membantu manajerial untuk menentukan keputusan, karena dalam memutuskan informasi dapat diambil berdasarkan data [2]. Maka dari itu, apabila akan melakukan pengajuan pinjaman dana ke PT. WOM Finance terdapat syarat-syarat yang harus dipenuhi oleh calon konsumen. Jika terdapat syarat yang tidak terpenuhi maka perusahaan memiliki hak untuk menolak calon konsumen tersebut.

Akan tetapi, bisnis pembiayaan konsumen ini tidak dapat dipisahkan dari risiko kredit. Risiko kredit (*credit risk*) merupakan risiko kerugian yang berhubungan dengan kemungkinan kegagalan *counterparty* memenuhi kewajibannya atau risiko bahwa debitur tidak membayar kembali utangnya. Prediksi terhadap risiko kredit ini sangat penting untuk dilakukan oleh perusahaan. Karena apabila nasabah melakukan pembayaran kredit tidak lancar akan merugikan perusahaan. Kelancaran atau ketidak lancaran nasabah dalam membayar angsuran dapat diprediksi sebelumnya berdasarkan data menggunakan data mining. Data mining atau penambangan data merupakan metode yang bermanfaat untuk memperoleh informasi berharga dari sejumlah data yang dilakukan dengan menggunakan pengetahuan seperti statistik, matematika dan pengenalan pola [3]. Penambangan data melibatkan data besar untuk ekstrak dan identifikasi untuk ditemukan informasi yang berguna bagi perusahaan [4]. Penambangan data dapat digunakan untuk mengklasifikasikan, memprediksi, memperkirakan untuk mendapatkan informasi yang bermanfaat [5][6]. Prediksi merupakan sebuah fungsi yang dapat menemukan pola tertentu dari suatu data [7]. Pola-pola tersebut dapat diketahui dari berbagai variabel yang ada pada data. Ketika sudah

menemukan pola, maka pola yang didapat tersebut bisa digunakan untuk memprediksi variabel lain yang belum diketahui nilai ataupun jenisnya [8]. Salah satu metode prediksi yang dapat digunakan untuk penambahan data adalah dengan *decision tree* menggunakan algoritma C.45.

Algoritma C4.5 merupakan algoritma yang dapat digunakan dalam penyelesaian klasifikasi atau pengelompokan dan memiliki sifat yang prediktif [8]. Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan (*decision tree*). Pohon keputusan adalah model prediksi melalui struktur pohon atau hirarki dengan melakukan perubahan pada data kedalam pohon keputusan dan aturan-aturan keputusan [9]. Pohon keputusan bertujuan agar dapat mengeksplorasi data, menggali informasi hubungan yang tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target [10]. Pohon keputusan digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang sebelumnya kompleks kemudian diubah menjadi lebih sederhana. Penerapan pohon keputusan bermanfaat dalam membantu kinerja proses *data mining* karena memiliki kemampuan dalam membantu pengambilan keputusan yang kompleks untuk dapat disederhanakan sehingga dapat menghasilkan model yang mempermudah dalam interpretasi solusi dari permasalahan yang ada [5]. Algoritma C4.5 merupakan pengembangan dari algoritma *decision tree* lainnya yaitu ID3 yang dapat mengatasi permasalahan *missing value*, *continue data* dan *pruning* [11].

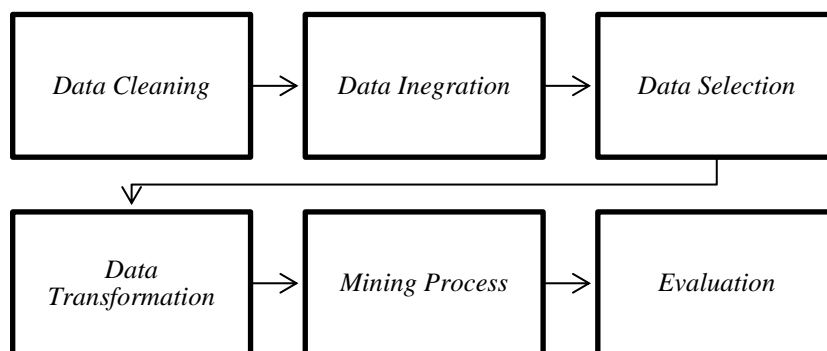
Di beberapa penelitian algoritma C4.5 memiliki akurasi yang baik. Beberapa penelitian terkait, diantaranya penelitian mengenai implementasi C4.5 yang digunakan untuk memprediksi kepuasan penumpang BRT Trans Semarang [12]. Dari hasil pengujian menunjukkan bahwa C4.5 cocok digunakan pada kasus ini, karena hasil akurasi mencapai 95%. Penelitian lainnya, Penelitian lainnya yaitu tentang klasifikasi tingkat kelulusan mahasiswa dengan pohon keputusan C.45 [13]. Berdasarkan hasil evaluasi dengan membandingkan keputusan asli dengan keputusan yang didapatkan dari metode yang diusulkan memperoleh nilai akurasi sebesar 94%. Selanjutnya, penelitian mengenai penerapan C4.5 yang diimplementasikan pada rekomendasi untuk menentukan penerimaan mitra penjualan [14]. Penelitian ini menunjukkan bahwa algoritma C4.5 mampu melakukan klasifikasi dan dapat merekomendasikan dengan hasil akurasi sebesar 92,26%.

Penelitian ini bertujuan untuk melakukan prediksi tingkat risiko kredit dengan menggunakan algoritma pohon keputusan C.45. Algoritma pohon keputusan C.45 memiliki fungsi untuk menemukan pengetahuan atau pola-pola kesamaan karakteristik dalam suatu kelompok atau kelas tertentu. Prediksi tingkat risiko yang digunakan terdapat dua kelas, yaitu lancar dan tidak lancar. Data yang digunakan untuk prediksi yaitu data yang diambil dari berkas nasabah. Pada penelitian ini algoritma C.45 diimplementasikan dan dianalisa menggunakan aplikasi WEKA.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Untuk menerapkan pendekatan data mining yang digunakan dalam penyelesaian untuk memprediksi tingkat risiko kredit dibutuhkan perencanaan yang matang melalui tahapan-tahapan penelitian yang tersusun secara terstruktur dan sesuai dengan kebutuhan. Gambar 1 berikut ini adalah tahapan penelitian yang dilakukan.



Gambar 1. Tahapan Penelitian

- 1) *Data Cleaning* (Pembersihan Data)  
*Data cleaning* merupakan sebuah tahapan yang digunakan untuk menghapus *missing values* atau *noise* pada data [3]. Apabila terdapat data yang kosong pada salah satu *row* yang ada pada file, maka *row* tersebut akan dilakukan penghapusan.
- 2) *Data Inegration* (Integrasi Data)  
 Integrasi data merupakan penggabungan data dari berbagai sumber yang disatukan ke dalam tempat penyimpanan baru [15]. Dalam hal ini, data di dapat melalui dua sumber yaitu dari file *excel* dan sistem perusahaan.
- 3) *Data Selection* (Seleksi Data)  
 Tahap seleksi data merupakan tahap pemilihan data operasional, karena tidak semua data akan digunakan [16]. Maka dari itu data yang diambil hanya data yang sesuai untuk dianalisis. Padah tahap ini tidak semua variabel akan digunakan, hanya ada beberapa variabel yang diperlukan untuk proses klasifikasi.
- 4) *Data Transformation* (Data Transformasi)

Data transformasi merupakan proses mengubah data menjadi format yang sesuai, sehingga dapat digunakan dalam pemrosesan *data mining*.

5) *Mining Process* (Proses Mining)

Tahapan ini merupakan tahapan utama, dimana algoritma *data mining* diimplementasikan untuk mendapatkan pengetahuan yang bermanfaat yang tersembunyi dari data yang ada. Algoritma C4.5 teknik *decision tree* yang membangun beberapa kaidah dan sebuah pohon keputusan agar akurasi dapat ditingkatkan. C4.5 menggunakan parameter *gain ratio* yang digunakan untuk mendapatkan variabel yang akan digunakan agar didapatkan bentuk cabang pada pohon keputusan. Menghitung *gain ratio* dapat menggunakan persamaan berikut ini.

$$Entropy(S) = \sum_{i=0}^n - pi * \log_2 pi \quad (1)$$

Dimana  $S$  adalah himpunan kasus,  $n$  adalah jumlah partisi dari  $S$ , dan  $pi$  adalah probabilitas kasus dalam partisi ke- $i$ . Sedangkan *gain* adalah ukuran efektifitas suatu atribut dalam mengklasifikasikan data yang digunakan agar dapat menentukan urutan atribut [17]. Untuk menghitung *gain* dapat melalui persamaan berikut ini:

$$Gain(S, A) = Entropy(S) - \sum_{i=0}^n \frac{|Si|}{|S|} * Entropy(Si) \quad (2)$$

6) *Evaluation* (Evaluasi)

Evaluasi bertujuan untuk mengetahui kinerja dari model yang dikembangkan dan memastikan model yang dikembangkan dapat berjalan dengan baik [18]. Untuk evaluasi pada penelitian ini menggunakan *confusion matrix*, di mana akan dihitung nilai *precision*, *recall*, dan *accuracy*. *Confusion matrix* terdiri dari *true positive*, *false positive*, *true negative* dan *false negative* untuk menghitung presisi, recall dan akurasi [14]. *Precision* adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Sedangkan *recall* adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Sedangkan *Accuracy* didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual.

### 3. HASIL DAN PEMBAHASAN

Berdasarkan analisis data, variabel yang digunakan untuk mengklasifikasikan dan memprediksi tingkat risiko kredit adalah: jenis kelamin, status perkawinan, tenor, pekerjaan, pinjaman, dan penghasilan. Masing-masing atribut dikelompokkan berdasarkan: atribut jenis kelamin terdiri dari laki-laki dan perempuan, atribut status perkawinan terdiri dari menikah dan tidak menikah, atribut tenor terdiri dari A untuk 6-10 bulan dan B untuk 12-18 bulan, atribut pekerjaan terdiri dari wiraswasta, pegawai swasta dan PNS, atribut pinjaman terdiri dari sedikit untuk 2-3,5 juta, banyak untuk 3,5-4,5 juta dan sangat banyak untuk >4,5 juta, atribut penghasilan terdiri dari rendah untuk <2 juta, sedang untuk 2-3,5 juta dan tinggi untuk >3,5 juta

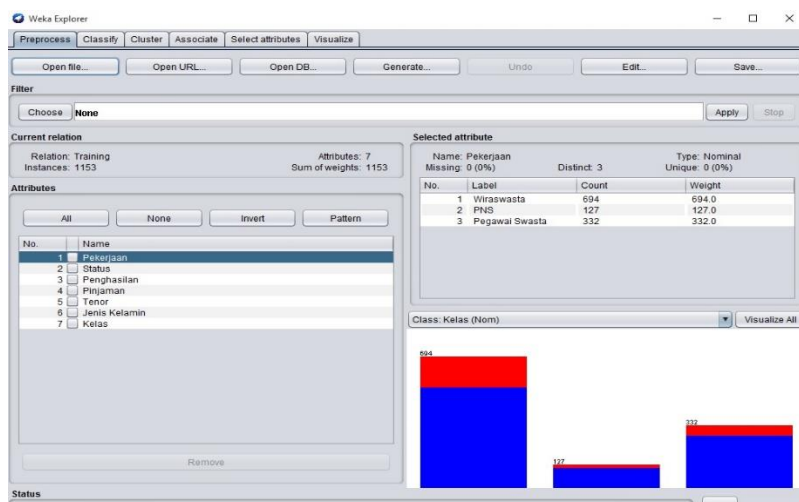
Data set yang digunakan adalah data konsumen pada PT. WOM Finance dari tahun 2018, 2019 dan 2020. Data yang digunakan sebanyak 1.153 data. Untuk data testing menggunakan sebanyak 91 data. Untuk implementasi pada pohon keputusan C4.5 dimulai dari menentukan nilai *gain* dari setiap atribut. Hasil perhitungan tersebut digunakan untuk menentukan akar pertama, yang didapatkan dari nilai *gain* yang tertinggi. Berikut hasil dari perhitungan *gain* dari variabel yang telah ditentukan. Hasil perhitungan *gain* dapat dilihat pada Tabel 1 berikut ini.

**Tabel 1.** Hasil Perhitungan *Gain* Untuk Setiap Variabel

| No | Variabel          | GainRatio   |
|----|-------------------|-------------|
| 1. | Jenis Kelamin     | 0.052111490 |
| 2. | Status Perkawinan | 0.107217712 |
| 3. | Tenor             | 0.052111490 |
| 4. | Pekerjaan         | 0.387957684 |
| 5. | Pinjaman          | 0.378289254 |
| 6. | Penghasilan       | 0.429410337 |

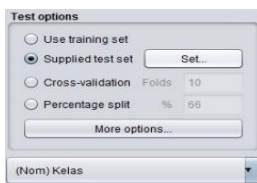
Berdasarkan tabel 1, bahwa variabel yang memiliki nilai *gain* paling besar terdapat pada variabel pinjaman, maka percabangan pertama untuk proses pembuatan *tree* terjadi pada variabel penghasilan. Penghasilan “Tinggi” memiliki keputusan “Lancar”, pinjaman “Sedang” memiliki keputusan “Tidak Lancar”, dan penghasilan “Rendah” memiliki keputusan “Tidak Lancar”. Proses untuk membentuk cabang kedua dan seterusnya menerapkan proses yang sama, yakni didasarkan pada variabel dengan nilai *gain* paling besar. Pada pembuatan cabang yang kedua, hasil *gain* dari variabel pinjaman tidak dihitung lagi.

Implementasi algoritma C4.5 menggunakan alat bantu Weka versi 3.8.2. Sebelum melakukan proses pengolahan data menggunakan Weka, data hasil *preprocessing* di *export* terlebih dahulu kedalam format yang dapat diproses oleh Weka yaitu CSV (*Comma Separated Value*). Selanjutnya, dilakukan proses *mining* dengan menggunakan *software* Weka. Gambar 2 berikut ini adalah tampilan Weka *Explorer* dimana dilakukan *setting* untuk data yang akan dilakukan klasifikasi dengan C4.5.



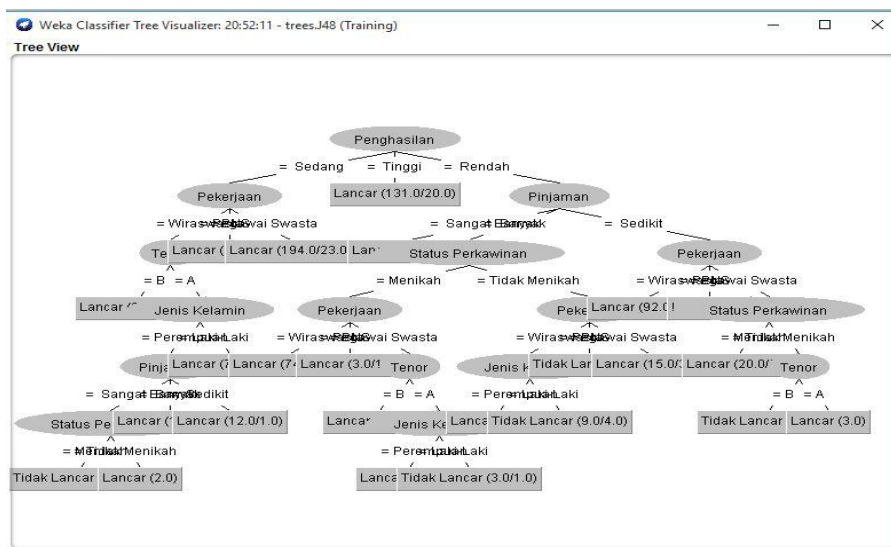
Gambar 2. Setting Data Mining Dengan Weka

Selanjutnya proses berpindah ke panel *Classify*. Sebelum proses dimulai, terlebih dahulu menentukan algoritma yang dipakai. Dalam panel *classify* terdapat beberapa algoritma yang dapat dipilih dan memiliki fungsi serta cara kerja yang berbeda. Pada penelitian ini penulis memilih untuk menggunakan algoritma C4.5. Dalam Weka algoritma C4.5 mempunyai nama J48. Sebelum proses dimulai terdapat parameter yang harus diatur, yaitu parameter *unpruned*. Parameter *unpruned* diubah dari *false* menjadi *true*. Setelah parameter algoritma C4.5 diatur maka langkah selanjutnya adalah menentukan pilihan *test* yang akan digunakan. Terdapat beberapa teknik *test* yang dapat dipilih, pada penelitian ini menggunakan pilihan teknik *Supplied test set*. Dimana menggunakan data *testing* sebagai data yang akan diprediksi berdasarkan data *training*, seperti pada Gambar 3 berikut ini.



Gambar 3. Test Options

Sebelum memilih data testing yang akan diproses, data testing tersebut harus memiliki atribut yang sama dengan data training. Mulai dari banyaknya atribut dan tipe atribut. Untuk data testing yang akan diprediksi kelasnya, isi kelas yang akan di prediksi dengan tanda tanya "?". Setelah mengatur semua parameter yang ada pada algoritma C4.5 maka setelah itu dilakukan pemilihan *test* yang akan diterapkan. Terdapat beberapa teknik *test* yang dapat dipilih, pada penelitian ini menggunakan pilihan teknik *Supplied test set*. Dimana menggunakan data *testing* sebagai data yang akan diprediksi berdasarkan data *training*. Bentuk dari *decision tree* ditampilkan pada bagian *Classifier model*. Visualisasi *decision tree* yang dihasilkan terlihat pada gambar 4 berikut ini.



Gambar 4. Hasil Decision Tree Yang Terbentuk

Pada gambar 4, terlihat percabangan pertama dari *tree* menggunakan variabel penghasilan. Pada *tree level* kedua terdapat dua cabang, masing-masing berdasarkan variabel pekerjaan dan pinjaman. Pada *tree level* ketiga terdapat tiga cabang, masing-masing berdasarkan variabel tenor, status perkawinan dan pekerjaan. Pada *tree level* keempat terdapat tiga cabang, masing-masing berdasarkan variabel jenis kelamin, pekerjaan dan status perkawinan. Pada *tree level* kelima terdapat dua cabang, masing-masing berdasarkan variabel pinjaman dan tenor. *Tree* tersebut memiliki ukuran 40 dan jumlah *leaf* sebanyak 24. Ini artinya terdapat 24 *rule* yang terbentuk, aturan tersebut diantaranya sebagai berikut:

1. Jika Penghasilan = Tinggi, maka konsumen = Lancar.
2. Jika Penghasilan = Sedang dan Pekerjaan = PNS, maka konsumen = Lancar.
3. Jika Penghasilan = Sedang dan Pekerjaan = Pegawai Swasta, maka konsumen = Lancar.
4. Jika Penghasilan = Rendah dan Pinjaman = Sangat Banyak, maka konsumen = Lancar.
5. Jika Penghasilan = Rendah dan Pinjaman = Sedikit dan Pekerjaan = Wiraswasta, maka konsumen = Lancar.
6. Jika Penghasilan = Rendah dan Pinjaman = Sedikit dan Pekerjaan = PNS, maka konsumen = Lancar.
7. Jika Penghasilan = Sedang dan Pekerjaan = Wiraswasta dan Tenor = B, maka konsumen = Lancar.
8. Jika Penghasilan = Sedang dan Pekerjaan = Wiraswasta dan Tenor = A dan Jenis Kelamin = Laki-Laki, maka konsumen = Lancar.
9. Jika Penghasilan = Rendah dan Pinjaman = Banyak dan Status = Menikah dan Pekerjaan = Wiraswasta, maka konsumen = Lancar.
10. Jika Penghasilan = Rendah dan Pinjaman = Banyak dan Status = Menikah dan Pekerjaan = PNS, maka konsumen = Lancar.
11. Jika Penghasilan = Rendah dan Pinjaman = Banyak dan Status = Tidak Menikah dan Pekerjaan = PNS, maka konsumen = Tidak Lancar.
12. Jika Penghasilan = Rendah dan Pinjaman = Banyak dan Status = Tidak Menikah dan Pekerjaan = Wiraswasta, maka konsumen = Lancar.
13. Jika Penghasilan = Rendah dan Pinjaman = Sedikit dan Pekerjaan = Pegawai Swasta dan Status = Menikah, maka konsumen = Lancar.
14. Jika Penghasilan = Sedang dan Pekerjaan = Wiraswasta dan Tenor = A dan Jenis Kelamin = Perempuan dan Pinjaman = Sedikit, maka konsumen = Lancar.
15. Jika Penghasilan = Sedang dan Pekerjaan = Wiraswasta dan Tenor = A dan Jenis Kelamin = Perempuan dan Pinjaman = Banyak, maka konsumen = Lancar.
16. Jika Penghasilan = Rendah dan Pinjaman = Banyak dan Status = Menikah dan Pekerjaan = Pegawai Swasta dan Tenor = B, maka konsumen = Lancar.
17. Jika Penghasilan = Rendah dan Pinjaman = Banyak dan Status = Tidak Menikah dan Pekerjaan = Wiraswasta dan Jenis Kelamin = Perempuan, maka konsumen = Lancar.
18. Jika Penghasilan = Rendah dan Pinjaman = Banyak dan Status = Tidak Menikah dan Pekerjaan = Wiraswasta dan Jenis Kelamin = Laki-Laki, maka konsumen = Tidak Lancar.
19. Jika Penghasilan = Rendah dan Pinjaman = Sedikit dan Pekerjaan = Pegawai Swasta dan Status = Tidak Menikah dan Tenor = B, maka konsumen = Tidak Lancar.
20. Jika Penghasilan = Rendah dan Pinjaman = Sedikit dan Pekerjaan = Pegawai Swasta dan Status = Tidak Menikah dan Tenor = A, maka konsumen = Lancar.
21. Jika Penghasilan = Sedang dan Pekerjaan = Wiraswasta dan Tenor = A dan Jenis Kelamin = Perempuan dan Pinjaman = Sangat Banyak dan Status = , maka konsumen = Lancar.
22. Jika Penghasilan = Sedang dan Pekerjaan = Wiraswasta dan Tenor = A dan Jenis Kelamin = Perempuan dan Pinjaman = Sangat Banyak dan Status = Tidak Menikah, maka konsumen = Lancar.
23. Jika Penghasilan = Rendah dan Pinjaman = Banyak dan Status = Menikah dan Pekerjaan = Pegawai Swasta dan Tenor = A dan Jenis Kelamin = Perempuan, maka konsumen = Lancar.
24. Jika Penghasilan = Rendah dan Pinjaman = Banyak dan Status = Menikah dan Pekerjaan = Pegawai Swasta dan Tenor = A dan Jenis Kelamin = Laki-Laki, maka konsumen = Tidak Lancar.

Untuk melihat prediksi yang dihasilkan oleh Weka, klik kanan *Result list*. Selanjutnya pilih menu *Visualize classifier errors* kemudian *save*. Berdasarkan prediksi yang dihasilkan oleh Weka, bahwa dari 91 data testing terdapat 2 konsumen yang diprediksi tidak lancar dan 89 konsumen diprediksi lancar. Hasil visualisasi ditampilkan pada Gambar 5 berikut ini.



ARFF-Viewer - E:\1SKRIPSI\Result.arff

File Edit View

Result.arff

Relation: Testing\_predicted

| No. | 1: Pekerjaan<br>Nominal | 2: Status Perkawinan<br>Nominal | 3: Penghasilan<br>Nominal | 4: Pinjaman<br>Nominal | 5: Tenor<br>Nominal | 6: Jenis Kelamin<br>Nominal | 7: prediction margin<br>Numeric | 8: predicted Kelas<br>Nominal | 9: Kelas<br>Nominal |
|-----|-------------------------|---------------------------------|---------------------------|------------------------|---------------------|-----------------------------|---------------------------------|-------------------------------|---------------------|
| 1   | Wiraswasta              | Tidak Menikah                   | Rendah                    | Banyak                 | B                   | Laki-Laki                   | -0.111111                       | Tidak Lancar                  |                     |
| 2   | Pegawai Swasta          | Menikah                         | Rendah                    | Sedikit                | B                   | Laki-Laki                   | 0.5                             | Lancar                        |                     |
| 3   | Pegawai Swasta          | Tidak Menikah                   | Sedang                    | Banyak                 | B                   | Perempuan                   | 0.762887                        | Lancar                        |                     |
| 4   | Pegawai Swasta          | Menikah                         | Rendah                    | Banyak                 | B                   | Perempuan                   | 0.428571                        | Lancar                        |                     |
| 5   | Wiraswasta              | Menikah                         | Rendah                    | Banyak                 | B                   | Perempuan                   | 0.486486                        | Lancar                        |                     |
| 6   | Wiraswasta              | Menikah                         | Rendah                    | Banyak                 | B                   | Perempuan                   | 0.486486                        | Lancar                        |                     |
| 7   | Pegawai Swasta          | Menikah                         | Sedang                    | Banyak                 | B                   | Laki-Laki                   | 0.762887                        | Lancar                        |                     |
| 8   | Wiraswasta              | Tidak Menikah                   | Sedang                    | Banyak                 | B                   | Laki-Laki                   | 0.513699                        | Lancar                        |                     |
| 9   | Wiraswasta              | Menikah                         | Rendah                    | Banyak                 | B                   | Perempuan                   | 0.486486                        | Lancar                        |                     |
| 10  | Wiraswasta              | Menikah                         | Rendah                    | Banyak                 | B                   | Laki-Laki                   | 0.486486                        | Lancar                        |                     |
| 11  | Pegawai Swasta          | Menikah                         | Sedang                    | Banyak                 | B                   | Perempuan                   | 0.762887                        | Lancar                        |                     |
| 12  | Wiraswasta              | Menikah                         | Rendah                    | Banyak                 | B                   | Perempuan                   | 0.486486                        | Lancar                        |                     |
| 13  | Wiraswasta              | Tidak Menikah                   | Rendah                    | Sedikit                | B                   | Laki-Laki                   | 0.565217                        | Lancar                        |                     |
| 14  | Wiraswasta              | Menikah                         | Sedang                    | Banyak                 | B                   | Laki-Laki                   | 0.513699                        | Lancar                        |                     |
| 15  | Wiraswasta              | Menikah                         | Sedang                    | Banyak                 | B                   | Perempuan                   | 0.513699                        | Lancar                        |                     |
| 16  | Wiraswasta              | Menikah                         | Rendah                    | Sedikit                | B                   | Laki-Laki                   | 0.565217                        | Lancar                        |                     |
| 17  | Wiraswasta              | Menikah                         | Sedang                    | Sangat Banyak          | B                   | Perempuan                   | 0.513699                        | Lancar                        |                     |
| 18  | Wiraswasta              | Menikah                         | Rendah                    | Banyak                 | B                   | Laki-Laki                   | 0.486486                        | Lancar                        |                     |
| 19  | Wiraswasta              | Menikah                         | Sedang                    | Sedikit                | B                   | Laki-Laki                   | 0.513699                        | Lancar                        |                     |
| 20  | Wiraswasta              | Tidak Menikah                   | Sedang                    | Banyak                 | B                   | Laki-Laki                   | 0.513699                        | Lancar                        |                     |
| 21  | PNS                     | Menikah                         | Sedang                    | Sangat Banyak          | B                   | Laki-Laki                   | 0.682927                        | Lancar                        |                     |
| 22  | Wiraswasta              | Menikah                         | Sedang                    | Banyak                 | B                   | Perempuan                   | 0.513699                        | Lancar                        |                     |
| 23  | Pegawai Swasta          | Menikah                         | Sedang                    | Sedikit                | B                   | Laki-Laki                   | 0.762887                        | Lancar                        |                     |
| 24  | Wiraswasta              | Menikah                         | Rendah                    | Sangat Banyak          | B                   | Perempuan                   | 0.463415                        | Lancar                        |                     |
| 25  | Pegawai Swasta          | Menikah                         | Sedang                    | Banyak                 | B                   | Laki-Laki                   | 0.762887                        | Lancar                        |                     |

Gambar 5. Hasil Prediksi

Selanjutnya model perlu dievaluasi guna mengetahui tingkat akurasi dari model yang dibangun. Untuk menghitung akurasi data maka digunakan *confusion matrix*. *Confusion matrix* pada dasarnya informasi akurasi yang dihasilkan didapatkan dari perbandingan hasil klasifikasi yang dihasilkan oleh sistem dengan hasil yang seharusnya. Pada tabel 2 berikut ini merupakan tabel hasil evaluasi yang telah dilakukan.

Tabel 2. Hasil Evaluasi

| Evaluasi                   | Persentase |
|----------------------------|------------|
| <i>Accuracy</i>            | 79%        |
| <i>Error Rate</i>          | 20%        |
| <i>False Positive Rate</i> | 27%        |
| <i>Recall</i>              | 100%       |
| <i>Precision</i>           | 79%        |

Dari hasil evaluasi pada Tabel 2, diperoleh informasi sebagai berikut:

1. Perhitungan dengan metode C4.5 menggunakan 1153 data training dan 91 data testing. Dengan membandingkan 2 atribut yaitu 5 atribut dan 6 atribut.
2. Hasil evaluasi akurasi data pada penelitian ini menggunakan *confusion matrix* dengan 6 atribut menghasilkan akurasi sebesar 74% sedangkan dengan 5 atribut menghasilkan akurasi yang lebih baik yaitu sebesar 79%.
3. Pemilihan atribut sangat mempengaruhi dalam pengolahan algoritma C4.5 karena keputusan bergantung pada atribut yang dipilih.

Berdasarkan hasil evaluasi hasil akurasi dipengaruhi oleh beberapa faktor, diantaranya jumlah dataset, jumlah data training dan jumlah atribut.

## 4. KESIMPULAN

Pada penelitian ini melakukan prediksi tingkat risiko kredit menggunakan algoritma *decision tree* C.45. Algoritma C.45 memiliki fungsi untuk menemukan pengetahuan atau pola-pola kesamaan karakteristik dalam suatu kelompok atau kelas tertentu. Prediksi tingkat risiko yang digunakan terdapat dua kelas, yaitu lancar dan tidak lancar. Data yang digunakan untuk prediksi yaitu data yang diambil dari berkas nasabah. Dari hasil evaluasi menggunakan *confusion matrix* didapatkan akurasi yang dihasilkan untuk 1.153 data training dengan 91 data testing dan enam atribut yang digunakan menghasilkan akurasi sebesar 79%. Akurasi dapat dipengaruhi oleh beberapa faktor, diantaranya: jumlah data training, data testing dan atribut yang digunakan. Maka, Untuk penelitian selanjutnya dapat menggunakan variasi data training, data testing dan atribut sehingga didapatkan model dengan akurasi yang terbaik.

## REFERENCES

- [1] M. Hasan, “Prediksi Tingkat Kelancaran Pembayaran Kredit Bank Menggunakan Algoritma Naïve Bayes Berbasis Forward Selection,” *Ilk. J. Ilm.*, vol. 9, no. 3, pp. 317–324, 2017.
- [2] R. I. Borman and H. Fauzi, “Penerapan Metode Perbandingan Eksponensial (MPE) Dalam Sistem Pendukung Keputusan Penerimaan Beasiswa Siswa Berprestasi Pada SMK XYZ,” *CESS J. Comput. Eng. Syst. Sci.*, vol. 3, no. 1, pp. 17–22, 2018.
- [3] R. I. Borman and M. Wati, “Penerapan Data Mining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandarlampung Dengan Algoritma Naïve Bayes,” *J. Ilm. Fak. Ilmu Komput.*, vol. 9, no. 1, pp. 25–34, 2020.
- [4] E. Darmawan, “C4.5 Algorithm Application for Prediction of Self Candidate New Students in Higher Education,” *J. Online Inform.*, vol. 3, no. 1, p. 22, 2018.
- [5] Y. Mardi, “Data Mining : Klasifikasi Menggunakan Algoritma C4.5,” *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017.
- [6] P. Prasetyawan, I. Ahmad, R. I. Borman, A. Ardiansyah, Y. A. Pahlevi, and D. E. Kurniawan, “Classification of the Period Undergraduate Study Using Back-propagation Neural Network,” in *Proceedings of the 2018 International Conference on Applied Engineering, ICAE 2018*, 2018.
- [7] M. Akbar and Y. Rahmanto, “Desain Data Warehouse Penjualan Menggunakan Nine Step Methodology Untuk Business Intelegency Pada PT Bangun Mitra Makmur,” *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 1, no. 2, pp. 137–146, 2021.
- [8] A. Fauzi, N. M. Saraswati, and R. C. S. Hariyono, “Penerapan Algoritma K-Modes dan C4.5 Untuk Prediksi Pemilihan Jurusan di Universitas Peradaban Pada Siswa SMA (Studi Kasus: SMA Islam Ta’allumul Huda Bumiayu),” *IJIR*, vol. 1, no. 2, pp. 57–64, 2020.
- [9] A. Z. Pratama, L. Kurniawati, S. Larbona, and T. Haryanti, “Algoritma C4.5 Untuk Klasifikasi Nasabah Dalam Memprediksi Kredit Macet,” *Inf. Syst. Educ. Provesionals*, vol. 3, no. 2, pp. 121–130, 2019.
- [10] E. Elisa, “Analisa dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Mengidentifikasi Faktor-Faktor Penyebab Kecelakaan Kerja Kontruksi PT.Arupadhatu Adisesanti,” *J. Online Inform.*, vol. 2, no. 1, p. 36, 2017.
- [11] N. Azwanti, “Algoritma C4.5 Untuk Memprediksi Mahasiswa Yang Mengulang Mata Kuliah (Studi Kasus Di Amik Labuhan Batu),” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 1, pp. 11–22, 2018.
- [12] A. S. Febriarini and E. Z. Astuti, “Penerapan Algoritma C4.5 untuk Prediksi Kepuasan Penumpang Bus Rapid Transit (BRT) Trans Semarang,” *Eksplora Inform.*, vol. 8, no. 2, pp. 95–103, 2019.
- [13] T. Taufiq and Y. Yudihartanti, “Penerapan Algoritma C4.5 Klasifikasi Tingkat Kelulusan Mahasiswa,” in *Seminar Nasional Ilmu Komputer (SOLITER)*, 2019, vol. 2, pp. 153–162.
- [14] M. F. Arifin and D. Fitriannah, “Penerapan Algoritma Klasifikasi C4.5 Dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus: PT Atria Artha Persada,” *InComTech*, vol. 8, no. 2, pp. 87–102, 2018.
- [15] K. Latifah, S. Wibowo, and N. Q. Nada, “Analisis dan Penerapan Algoritma C4.5 Dalam Data Mining Untuk Menunjang Strategi Promosi Prodi Informatika UPGRIS,” *J. Tek. Inform.*, vol. 11, no. 2, pp. 3–8, 2018.
- [16] R. I. Borman, B. Priopradono, and A. R. Syah, “Klasifikasi Objek Kode Tangan pada Pengenalan Isyarat Alphabet Bahasa Isyarat Indonesia (Bisindo),” in *Seminar Nasional Informatika dan Aplikasinya (SNIA)*, 2017, no. September, pp. 1–4.
- [17] Y. Handrianto and M. Farhan, “C.45 Algorithm for Classification of Causes of Landslides,” *J. Publ. Informatics Eng. Res.*, vol. 4, no. 1, pp. 120–127, 2019.
- [18] A. Sucipto, Y. Fernando, R. I. Borman, and N. Mahmuda, “Penerapan Metode Certainty Factor Pada Diagnosa Penyakit Saraf Tulang Belakang,” *J. Ilm. FIFO*, vol. 10, no. 2, p. 18, 2019.