

# Analisis Penentuan Resiko Kredit Menggunakan Algoritma C.5.0

Karina Auliasari  
Program Studi Teknik Informatika  
Institut Teknologi Nasional Malang  
Malang, Indonesia  
karina.auliasari@lecturer.itn.ac.id

Mariza Kertaningtyas  
Jurusan Teknik Industri  
Institut Teknologi Nasional Malang  
Malang, Indonesia  
mariza\_kertaningtyas@lecturer.itn.ac.id

Diah Wilis L. Basuki  
Jurusan Teknik Industri  
Institut Teknologi Nasional Malang  
Malang, Indonesia  
wilis.basuki@lecturer.itn.ac.id

**Abstract**— Risk rating is a way of labeling credit risk on data, this label reflects the level of risk and the ability of a customer or company to have financial obligations. Bank sometimes take a long time to labeling new data, update label for decision making in obligations. This research is modeling the credit risk rating so that it can be used as a knowledge base for decision making in loan application process. Before doing modeling with C.5.0 algorithm, dataset is need to prepare and cleansing first, furthermore the modeling result is analyzed and evaluated. From the evaluation of the C.5.0 modeling was formed for the classification or prediction of credit risk, it still has an error rate of 24%, so its quite feasible to use. From the cofussion matrix can be concluded that if the intersection of columns and rows falls on the same label, then the classification of the model is correct, but if the labels are different then the classification is wrong. From the cofussion matrix also it can be seen if the numbers with the diagonal to the bottom right show the data with the correct prediction from the model because they intersect on the same label.

**Keywords**—Algorithm; Bank; C.5.0; Credit; Risk Rating

**Abstrak**— Penelitian ini melakukan pemodelan untuk memprediksi nilai risk rating atau pemeringkatan resiko kredit sehingga dapat dijadikan basis pengambilan keputusan terhadap proses pengajuan pinjaman baru. Diharapkan dengan adanya model prediksi lembaga pinjaman lebih cepat dan secara tepat dalam menyetujui atau menolak pengajuan pinjaman dan meminimalisir tingkat kesalahan dalam proses pengambilan keputusan. Pemodelan keputusan menggunakan model klasifikasi pohon keputusan C5 sebagai model akurasi, berbeda dengan penelitian-penelitian pada umumnya yang menggunakan algoritma C.4.5. Tahapan yang dilakukan terdiri dari 3 tahap yaitu persiapan data, pemodelan menggunakan algoritma C.5.0 dan analisis hasil pemodelan. Dari evaluasi hasil algoritma C.5.0 pemodelan yang dibentuk untuk klasifikasi atau prediksi resiko kredit tetap memiliki tingkat kesalahan sebesar 24%, sehingga model yang dihasilkan cukup layak digunakan. Dari hasil cofussion matrix dapat disimpulkan jika perpotongan kolom dan baris jatuh pada label yang sama maka klasifikasi dari model tersebut benar, namun jika labelnya berbeda maka klasifikasinya salah. Selain itu dari tampilan cofussion matrix bisa dilihat jika angka dengan diagonal ke kanan bawah menunjukkan data dengan prediksi yang benar dari model karena berpotongan pada label yang sama.

**Keywords**—Algoritma; Bank; C.5.0; Kredit; Risk Rating

## PENDAHULUAN

Bisnis inti perbankan dan koperasi simpan pinjam adalah produk investasi dan kredit pinjaman. Manajemen resiko kredit memiliki peran penting bagi perbankan ataupun koperasi simpan pinjam dalam memberikan kredit pinjaman. Baik bank maupun koperasi memiliki analisis kredit untuk memutuskan apakah pinjaman akan diberikan kepada nasabah ataupun permintaan kredit akan ditolak. Tentunya dalam mengambil keputusan analisis mengandalkan data historis keuangan nasabah yang seringkali disebut sebagai kredit portofolio nasabah. Jika analisis memahami dengan benar data tersebut akan berubah menjadi informasi yang membantu perbankan dan koperasi simpan pinjam menghindari kerugian pemberian kredit dan pencegahan adanya nasabah yang gagal bayar kredit.

*Risk rating* atau pemeringkatan resiko merupakan salah satu cara pelabelan resiko kredit pada data, dimana label ini mencerminkan tingkat resiko dan kemampuan seorang nasabah atau perusahaan memenuhi kewajiban tanggungan keuangan. Lembaga perbankan ataupun koperasi terkadang membutuhkan waktu lama untuk melakukan pelabelan data baru, memperbarui label pada data lama dan menjadikannya dasar pengambilan keputusan.

Penilaian label risk rating menggunakan teknik data mining banyak diminati pada beberapa penelitian sepuluh tahun terakhir. Seperti penelitian yang dilakukan oleh Desai et al. di tahun 1990 an dalam mengklasifikasi nasabah untuk pinjaman internasional dan memprediksi variabel kredit berdasarkan informasi kredit yang dimiliki[1]. Selanjutnya Yobas et al. menggunakan metode fuzzy dan decision tree mengelompokkan nasabah menjadi dua kategori yaitu pembayar baik dan pembayar miskin dengan mempertimbangkan parameter historis kartu kredit dan status pekerjaan nasabah[2]. Dari beberapa penelitian tersebut banyak yang mengaplikasikan metode decision tree yaitu algoritma C4.5, diantaranya penelitian oleh Rani di tahun 2016 menggunakan algoritma C4.5 untuk mengklasifikasikan nasabah ke dalam dua kategori yaitu nasabah berpotensi dan tidak berpotensi. Penelitian yang dilakukan Rani menggunakan empat parameter yaitu kolektibilitas, nilai anggunan, kelengkapan surat dan angsuran[3]. Selanjutnya di tahun 2018 Nur et al. menggunakan tujuh parameter yaitu jenis kelamin, total pinjaman, jumlah tanggungan, status pernikahan, pendapatan utama, total biaya hidup dan status pinjaman nasabah. Nur et al. menggunakan algoritma C.4.5 untuk mengklasifikasikan nasabah ke dalam dua kelas yaitu kredit macet dan lancar[4].

Masih di tahun yang sama Mandala dan Putri juga menggunakan algoritma C.4.5 untuk meklasifikasikan nasabah ke dalam dua kelas yaitu nasabah dengan jumlah pemberian kredit besar dan kecil. Mandala dan Putri menggunakan lima parameter yaitu penghasilan, pekerjaan, jaminan, saldo rekening tabungan dan kartu keluarga[5]. Di tahun 2019 penelitian yang dilakukan oleh Khasanah menggunakan algoritma C.4.5 untuk mengklasifikasikan nasabah ke dalam dua kelas yaitu layak dan tidak layak menerima kredit, pada penelitian ini menggunakan 9 parameter yaitu jenis kelamin, jarak, domisili, pekerjaan jumlah penghasilan, jumlah pinjaman, jangka waktu, lama waktu pinjam, jumlah tanggungan dan status pernikahan[6]. Berikutnya di tahun 2020 serupa dengan penelitian yang dilakukan Khasanah, penelitian Ulfa juga mengklasifikasikan nasabah ke dalam dua kelas layak dan tidak layak menggunakan algoritma C.4.5 namun dalam penelitiannya Ulfa hanya menggunakan 5 parameter yaitu jenis kelamin, jenis usaha, pencairan pinjaman, tanggal pinjaman, dan periode pinjaman[7]. Di tahun yang sama Natalia mengklasifikasikan nasabah ke dalam dua kelas yaitu nasabah dengan kredit macet dan lancar menggunakan algoritma C.4.5. Natalia menggunakan 7 parameter yaitu jenis kelamin, tanggungan, penghasilan, pinjaman, anggungan, tempo dan angsuran[8].

Berbeda dengan penelitian-penelitian sebelumnya pada penelitian ini melakukan pemodelan untuk memprediksi nilai risk rating atau pemeringkatan resiko kredit sehingga dapat dijadikan basis pengambilan keputusan terhadap proses pengajuan pinjaman baru. Diharapkan dengan adanya model prediksi lembaga pinjaman lebih cepat dan secara tepat dalam menyetujui atau menolak pengajuan pinjaman dan meminimalisir tingkat kesalahan dalam proses pengambilan keputusan. Pemodelan keputusan menggunakan model klasifikasi pohon keputusan C5 sebagai model akurasi.

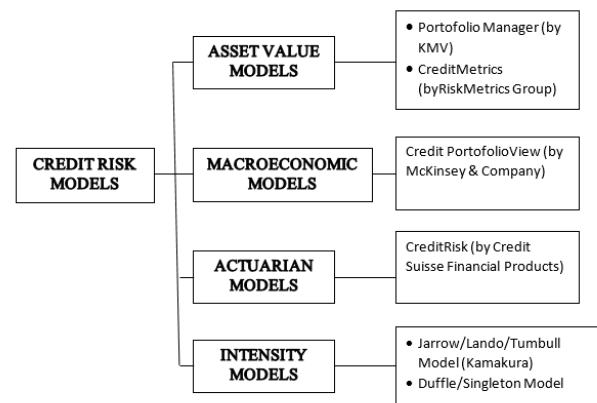
PENGUKURAN RESIKO KREDIT

Untuk menerapkan prinsip perkreditan yang sehat, lembaga perbankan maupun koperasi diminta untuk dapat mengukur resiko kredit dengan standar tertentu. Hal ini sesuai dengan kebijakan manajemen resiko di Indonesia yang diatur dalam peraturan otoritas jasa keuangan No. 18/PJOK/03/2016 tentang penerapan manajemen resiko bagi bank[9]. Dalam mengukur risiko kredit dikembangkan model dalam mengukur risiko kredit seperti pada Gambar 1. Menurut kredit resiko model pada Gambar 1 Bank harus melakukan verifikasi dengan menggunakan dua metode. Pertama, perbandingan, bank harus membandingkan hasil dengan hasil alternatif atau non-bank, dan juga harus menggunakan alat validasi kuantitatif lainnya. Analisis ini harus didasarkan pada data dan informasi yang sesuai dengan portofolio, diperbarui secara berkala dan mencakup periode pengamatan yang sesuai. Penilaian internal Bank sendiri atas kinerja model harus didasarkan pada riwayat data panjang yang mencakup serangkaian kondisi ekonomi dan idealnya satu atau lebih siklus ekonomi[10]. Oleh karena itu, penting untuk menerapkan teknologi informasi dalam menganalisa dan memberikan rekomendasi kepada pimpinan terhadap calon nasabah mengenai kredit[11].

ALGORITMA C.5.0

Algoritma C.5.0 merupakan merupakan penyempurnaan dari algoritma ID3 dan C4.5. Dalam proses pembentukan pohon keputusan nilai informasi gain tertinggi akan terpilih sebagai root bagi node selanjutnya. Algoritma ini dimulai

dengan semua data yang dijadikan akar dari pohon keputusan sedangkan atribut yang dipilih akan menjadi pembagi bagi sampel tersebut[13].



Gambar 1. Credit risk model[10]

Untuk menentukan akar dari pohon didasarkan pada nilai information gain tertinggi dari atribut-atribut yang ada. Sebelum mendapatkan nilai information gain, terlebih dahulu mencari nilai entropy. Entropy merupakan suatu parameter untuk mengukur tingkat keberagaman dari kumpulan data dan digunakan untuk menentukan seberapa informatif sebuah masukan atribut untuk menghasilkan sebuah atribut. Rumus dari entropy ditunjukkan pada persamaan 1 sebagai berikut[14]:

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

Keterangan:

- S adalah himpunan kasus
- n adalah jumlah nilai yang ada pada atribut target (jumlah kelas klasifikasi)
- pi adalah jumlah proporsi sampel (peluang) untuk kelas i

Selanjutnya menghitung information gain sebagai salah satu attribute selection measure yang digunakan untuk memilih test attribute tiap node pada tree. Untuk menghitung nilai information gain digunakan rumus seperti yang ditunjukkan dalam persamaan 2 berikut ini[14]:

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} * Entropy(S_i) \tag{2}$$

Keterangan:

- S adalah himpunan kasus
- A adalah atribut penentu
- n jumlah partisi atribut A
- |Si| adalah jumlah kasus untuk nilai i
- |S| adalah jumlah seluruh kasus dalam S
- Entropy (Si) adalah entropy untuk kasus yang memiliki nilai i

Langkah penentuan akar dengan menghitung nilai entropy hingga semua baris data terpartisi akan terus diulang. Partisi pohon keputusan akan berhenti pada tiga kondisi terpenuhi yaitu kondisi semua record dalam simpul N mendapat kelas yang sama, tidak ada atribut dalam record



proses pemodelan algoritma C5.0 digunakanlah bahasa pemrograman R pada penelitian ini, dimana berdasarkan peringkat IEEE bahasa pemrograman R menempati peringkat kelima terpopuler setelah C++ untuk analisis data [15]. Bahasa pemrograman R juga didukung oleh banyak paket library analisis data yang tersedia pada repository CRAN (*Comprehensive R Archive Network*) yang dibuat dan dikembangkan oleh komunitas yang sangat aktif [16].

Dari hasil pemodelan teks pada Gambar 4 bisa disimpulkan alur pengambilan keputusan untuk menentukan peringkat resiko seperti yang ditunjukkan pada Gambar 5. Alur pengambilan keputusan memiliki empat alur. Alur pertama jika jumlah tanggungan lebih dari 4 orang maka perlu dilihat dua kondisi berikutnya, yaitu kondisi pertama jika lama pinjaman sampai dengan 24 bulan, maka klasifikasi peringkat resiko kreditnya bernilai 4, kondisi kedua jika lama pinjaman lebih dari 24 bulan maka klasifikasi peringkat resiko kreditnya bernilai 5. Dari alur pertama dan kondisi pertama mengklasifikasikan 105 data namun 30 diantara 105 data ini salah klasifikasi, sedangkan untuk alur pertama dengan kondisi kedua mengklasifikasikan 120 data dengan 51 diantara 120 data salah klasifikasi.

Class specified by attribute 'Peringkat Resiko'

Read 800 cases (3 attributes)

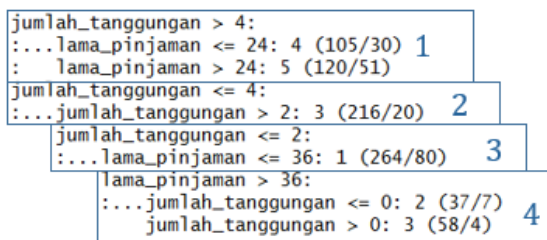
Decision tree:

```

jumlah_tanggungan > 4:
... lama_pinjaman <= 24: 4 (105/30)
: lama_pinjaman > 24: 5 (120/51)
jumlah_tanggungan <= 4:
... jumlah_tanggungan > 2: 3 (216/20)
: jumlah_tanggungan <= 2:
: ... lama_pinjaman <= 36: 1 (264/80)
: lama_pinjaman > 36:
: ... jumlah_tanggungan <= 0: 2 (37/7)
: jumlah_tanggungan > 0: 3 (58/4)
    
```

Gambar 4. Hasil pemodelan teks algoritma C5.0

Decision tree:

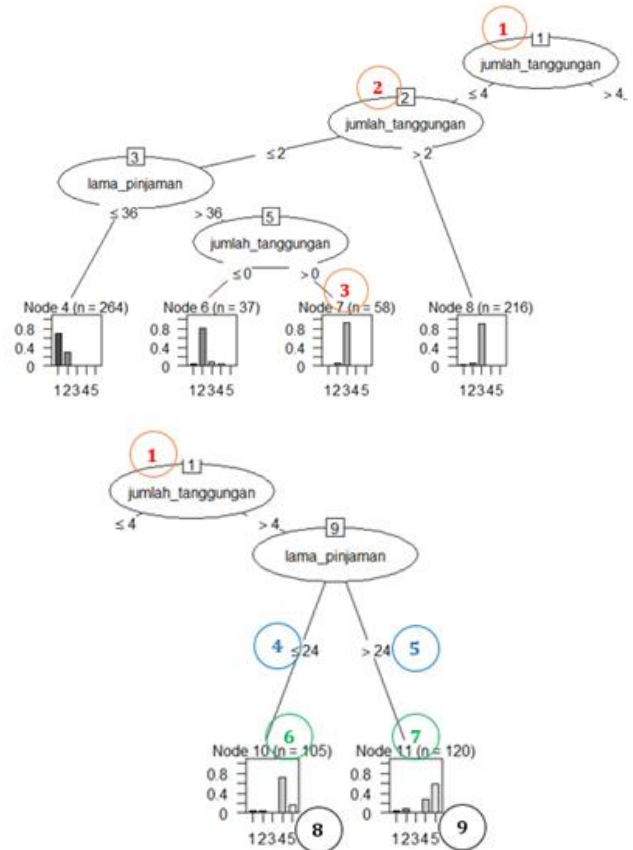


Gambar 5. Alur pengambilan keputusan

Selanjutnya pada alur kedua jika jumlah tanggungan sampai dengan 4 orang akan ada dua kondisi yaitu kondisi pertama jika jumlah tanggungan lebih dari 2 orang maka klasifikasi peringkat resiko kreditnya bernilai 3 (untuk kondisi ini data yang terklasifikasi 216 data dengan 20 data salah klasifikasi). Kondisi kedua jika jumlah tanggungan sampai dengan 2 orang maka masuk ke alur ketiga.

Pada alur ketiga sesuai kondisi sebelumnya yaitu jumlah tanggungan sampai dengan 2 orang akan dilihat terlebih dahulu nilai lama pinjaman. Jika lama pinjamannya sampai dengan 36 bulan dan jumlah tanggungan 2 orang maka klasifikasi peringkat resiko kreditnya bernilai 1 (untuk kondisi ini data yang terklasifikasi 264 data dengan 80 data salah klasifikasi).

Selanjutnya masuk pada alur keempat yaitu jika lama pinjaman lebih dari 36 bulan akan dilihat dua kondisi berikutnya yaitu jumlah tanggungan. Kondisi pertama jika lama pinjaman lebih dari 36 bulan dan jumlah tanggungan tidak ada (sama dengan 0) maka klasifikasi peringkat resiko kreditnya bernilai 2 (untuk kondisi ini data yang terklasifikasi 37 data dengan 7 data salah klasifikasi). Kondisi kedua jika pinjaman lebih dari 36 bulan dan jumlah tanggungan 1 orang (lebih besar dari 0) maka klasifikasi peringkat resiko kreditnya bernilai 3 (untuk kondisi ini data yang terklasifikasi 58 data dengan 4 data salah klasifikasi).



Gambar 6. Hasil pemodelan diagram pohon algoritma C5.0

Selain pemodelan teks algoritma C5.0 juga menghasilkan pemodelan dengan bentuk diagram pohon seperti yang ditunjukkan Gambar 6. Pada Gambar 6 lingkaran merah nomor 1,2 dan 3 menunjukkan node dan penomoran nodenya. Lingkaran merah nomor 1 adalah node nomor 1 dimana merupakan root node dengan variable penentu jumlah tanggungan, lingkaran merah nomor 2 adalah node nomor 2 dengan variable penentu jumlah tanggungan dan untuk lingkaran merah nomor 3 merupakan node nomor 7 sebagai leaf node untuk klasifikasi peringkat resiko.

Lingkaran biru menunjukkan kondisi split ke beberapa node, seperti pada Gambar 6 lingkaran nomor 4 menunjukkan split dimana lama pinjaman kurang atau sama dengan 24 bulan, untuk lingkaran biru nomor 5 menunjukkan kondisi split dimana lama pinjaman lebih dari 24 bulan. Kemudian untuk lingkaran hijau menunjukkan jumlah data yang sudah terklasifikasi, seperti yang ditunjukkan pada Gambar 6, lingkaran nomor 6 menunjukkan hasil klasifikasi sejumlah 105 data dan lingkaran nomor 7 menunjukkan hasil klasifikasi sejumlah

120 data. Bagian lingkaran terakhir yaitu lingkaran hitam menunjukkan hasil klasifikasi dan distribusinya (dengan rentang rasio antara angka 0 dan 1) seperti pada Gambar 6, lingkaran hitam nomor 8 menunjukkan peringkat resiko kredit pada node nomor 10 mayoritas bernilai 4, sehingga pada node nomor 10 model klasifikasinya bernilai 4. Lingkaran hitam nomor 9 menunjukkan peringkat resiko pada node nomor 11 mayoritas bernilai 5, oleh karena itu model node nomor 11 klasifikasinya bernilai 5.

Pada proses pembentukan pohon keputusan tidak dilakukan *tree pruning* karena dalam penelitian ini variabel yang digunakan untuk menghasilkan *decision tree* hanya dua variabel yaitu jumlah tanggungan dan lama pinjaman. Ukuran pohon keputusan yang dihasilkan sederhana sehingga tidak perlu melakukan pemangkasan pohon (*tree pruning*).

**B. Evaluasi pemodelan**

Dari pemodelan teks yang dihasilkan selanjutnya dilakukan evaluasi model klasifikasi atau prediksi data algoritma C.5.0 seperti yang ditunjukkan pada Gambar 7. Pada bagian a ada beberapa informasi yaitu ada 800 baris data sebagai training data yang diproses. Size yang bernilai 6 yaitu jumlah leaf node (node ujung) dari model yang dihasilkan algoritma C.5.0. Errors bernilai 192 baris data yang salah diklasifikasi dengan rasio prosentase kesalahan dari seluruh populasi data sebesar 24%.

Evaluation on training data (800 cases):

Decision Tree					
Size	Errors	a			
6	192(24.0%)	<<			

(a)	(b)	(c)	(d)	(e)	<-classified as
184	2	5	6	6	(a): class 1
80	30	19	6	11	(b): class 2
	3	250			(c): class 3
	2		75	34	(d): class 4
			18	69	(e): class 5

Gambar 7. Hasil evaluasi model algoritma C5.0

Dari dua input variable yang digunakan oleh model algoritma C.5.0 prosentase tingkat kepentingan variabel jumlah tanggungan 100% dan variabel lama pinjaman 73% seperti yang ditunjukkan pada Gambar 8. Dari besaran prosentase tersebut menunjukkan bahwa variabel jumlah tanggungan yang menempati root node pada model yang dihasilkan.

**Attribute usage:**

100.00% jumlah\_tanggungan  
73.00% lama\_pinjaman

Gambar 8. Prosentase tingkat kepentingan *input variable*

Bagian b pada tampilan Gambar 9 merupakan confusion matrix dengan ukuran 5x5. Bagian b ini merupakan tampilan yang menunjukkan hasil dari klasifikasi yang dilakukan model dibandingkan dengan data klasifikasi sebenarnya, sehingga bisa ditunjukkan seberapa akurat model melakukan klasifikasi atau prediksi. Confusion matrix pada Gambar 6 terdiri dari jumlah kolom dan baris yang sama. Header dari baris dan kolom merupakan representasi dari nilai class variable (peringkat resiko), dimana class variable ada 5 maka matrix berukuran 5x5. Header pada kolom menunjukkan nilai class peringkat resiko yang diprediksi atau

diklasifikasikan oleh model dengan menggunakan label (a), (b), (c), (d) dan (e). Header pada baris menunjukkan peringkat resiko pada data sebenarnya, yang masih dipresentasikan dengan (a), (b), (c), (d) dan (e). Terlihat jika (a) merupakan representasi peringkat resiko dengan nilai 1, (b) merupakan representasi peringkat resiko dengan nilai 2 dan seterusnya. Tiap perpotongan antara kolom dan baris merupakan informasi hasil prediksi dari class ada di nilai pada kolom dibandingkan data aktual classnya berada pada nilai di baris.

(a)	(b)	(c)	(d)	(e)	<-classified as
184	2	5	6	6	(a): class 1
80	30	19	6	11	(b): class 2
	3	250			(c): class 3
	2		75	34	(d): class 4
			18	69	(e): class 5

Gambar 9. Nilai baris pertama *confusion matrix*

Untuk memberikan gambaran dari *confusion matrix* ini diambil contoh penjelasan dari baris pertama dari matrix ini seperti yang ditunjukkan pada Gambar 9 bagian kotak merah. Pada baris pertama angka 184 pada kolom pertama dan baris pertama menunjukkan jumlah data yang benar klasifikasinya atau prediksinya dimana klasifikasi model terhadap data mendapatkan peringkat resiko bernilai 1 dan pada data aktual juga peringkat resikonya bernilai 1. Angka 2 pada pada kolom kedua baris pertama menunjukkan jumlah data yang salah prediksi, dimana klasifikasi model terhadap data mendapatkan peringkat resiko bernilai 2 dan pada data aktual nilai peringkat resikonya bernilai 1. Angka 5 pada kolom ketiga baris pertama menunjukkan jumlah data yang salah prediksi, dimana klasifikasi model terhadap data mendapatkan peringkat resiko bernilai 3 dan pada data aktual peringkat resikonya bernilai 1. Angka 6 pada kolom keempat dan baris pertama menunjukkan jumlah data yang salah prediksi, dimana klasifikasi model terhadap data mendapatkan peringkat resiko bernilai 4 dan pada data aktual nilai peringkat resikonya bernilai 1. Terakhir angka 6 pada kolom kelima dan baris pertama menunjukkan jumlah data yang salah prediksi dimana klasifikasi model terhadap data mendapatkan peringkat resiko bernilai 5 dan pada data aktual nilai peringkat resikonya bernilai 1. Dari tampilan *confusion matrix* dengan contoh baris pertama pada Gambar 9 menunjukkan jika perpotongan kolom dan baris jatuh pada label yang sama maka klasifikasi dari model tersebut benar, namun jika labelnya berbeda maka klasifikasinya salah.

(a)	(b)	(c)	(d)	(e)	<-classified as
184	2	5	6	6	(a): class 1
80	30	19	6	11	(b): class 2
	3	250			(c): class 3
	2		75	34	(d): class 4
			18	69	(e): class 5

Gambar 10. Data dengan prediksi benar dan salah

Dari tampilan *confusion matrix* bisa dilihat jika angka dengan diagonal ke kanan bawah menunjukkan data dengan prediksi yang benar dari model karena berpotongan pada label yang sama seperti yang ditunjukkan pada Gambar 10 pada bagian kotak-kotak berwarna biru. Untuk bagian kotak-kotak berwarna merah merupakan data dengan prediksi yang salah. Jika keseluruhan angka pada kota-kotak

biru dijumlahkan bernilai 608 data. Untuk angka pada kotak-kotak merah jika seluruhnya dijumlahkan bernilai 192 data. Maka 608 data prediksi benar dijumlahkan dengan 192 data prediksi salah jumlahnya 800 data sesuai jumlah *data training*.

#### KESIMPULAN

Pemodelan algoritma C.5.0 menggunakan dua input variable yaitu lama pinjaman dan jumlah tanggungan menghasilkan empat alur pengambilan keputusan untuk menentukan peringkat resiko pengambilan kredit. Dari dua input variable yang digunakan oleh model algoritma C.5.0 prosentase tingkat kepentingan variabel jumlah tanggungan 100% dan variabel lama pinjaman 73%. Dari besaran prosentase tersebut menunjukkan bahwa variabel jumlah tanggungan yang menempati root node pada model yang dihasilkan. Pada penelitian selanjutnya memungkinkan untuk menambah input variable seperti status kpr.

Dari evaluasi hasil algoritma C.5.0 pemodelan yang dibentuk untuk klasifikasi atau prediksi resiko kredit tetap memiliki tingkat kesalahan sebesar 24%, sehingga model yang dihasilkan cukup layak digunakan. Dari hasil *confusion matrix* dapat disimpulkan jika perpotongan kolom dan baris jatuh pada label yang sama maka klasifikasi dari model tersebut benar, namun jika labelnya berbeda maka klasifikasinya salah. Selain itu dari tampilan *confusion matrix* bisa dilihat jika angka dengan diagonal ke kanan bawah menunjukkan data dengan prediksi yang benar dari model karena berpotongan pada label yang sama. Pada penelitian berikutnya memungkinkan untuk melakukan evaluasi model terhadap data testing.

#### PENGHARGAAN

Ucapan terima kasih kami sampaikan kepada Institut Teknologi Nasional Malang yang telah memberikan dana pada pelaksanaan tahun 2020. Terima kasih juga kami sampaikan kepada program studi Teknik Informatika S1 ITN Malang dan Kepala Laboratorium Database Teknik Informatika S1 ITN Malang yang telah memfasilitasi selama pelaksanaan.

#### REFERENSI

- [1] V. Desai, J. Crook dan G. Overstreet, "A comparison of neural networks and linear scoring models in the credit union environment", *European Journal of Operations Research*, 95(1), hal.24-37, 1996.
- [2] M.B. Yobas, J.N. Crook dan P. Ross, "Credit Scoring Using Evolutionary Techniques", *IMA Journal of Mathematics Applied in Business & Industry*, Vol. 11, hal. 111-125, 2000.
- [3] L.N. Rani, "Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit", *Jurnal INOVTEK POLBENG-SERI INFORMATIKA*, Vol. 1 No. 2, hal. 126-132, November 2016.
- [4] I.T.A. Nur, N. Y. Setiawan dan F.A. Bachtiar, "Prediksi Kredit Macet Berdasarkan Preferensi Nasabah Menggunakan Metode Klasifikasi C4.5 pada Koperasi Simpan Pinjam Mitra Raya Wates", *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol. 2, No. 12, hal. 6118-6127 Desember 2018.
- [5] E.P.W. Mandala dan D.E. Putri, "Prediksi Jumlah Pemberian Kredit Kepada Nasabah Di Bank Perkreditan Rakyat Dengan Algoritma C.4.5", *UPI YPTK Jurnal KomTekInfo* Vol. 5, No. 1, hal 70-80, Juni 2018.
- [6] S.N. Khasanah, "Komparasi Algoritma C4.5 Dan Naive Bayes Untuk Menganalisa Kelayakan Pengajuan Kredit", *Jurnal Manuk Penusa* Vol. 3 No. 3, hlm. 94-98, Desember 2019.
- [7] R. T. Ulfa, "Aplikasi Berbasis Web Untuk Klasifikasi Kelayakan Pemberian Kredit pada Koperasi Bumdes Mitra Baru Menggunakan

- Algoritma C.4.5", Tugas Akhir Program Studi Sistem Informasi UIN SUSKA Riau, 2020.
- [8] P. Natalia, "Sistem Pendukung Keputusan Untuk Kelayakan Pemberian Kredit Menggunakan Algoritma C.4.5 pada KUD Sawitra" Tugas Akhir Program Studi Sistem Informasi UIN SUSKA Riau, 2020.
- [9] Otoritas Jasa Keuangan, "Penerapan Manajemen Risiko Bagi Bank Umum", POJK No. 18/POJK/03/2016, 2016.
- [10] M.A. Canbolat dan A. Gumrah, "Analysis of Credit Risk Measurement Models in the Evaluation of Credit Demands", *Universal Journal of Accounting and Finance* 3(1), hal. 16-20, 2015.
- [11] M.K. Anam, A.R. Putra, S. Fadli, M.B. Firdaus, F. Suandi dan Latifah, "Audit Teknologi Informasi Pada Sistem Perkreditan Online Terpadu Bank XYZ Cabang Perawang Menggunakan ITIL V3", *MISI (Jurnal Manajemen Informatika & Sistem Informasi)* Volume 3, No 2, hal. 90-99, Juni 2020.
- [12] A. Izyuddin dan S. Wibisono, "Aplikasi Prediksi Penjualan AC Menggunakan Decision Tree Dengan Algoritma C4.5", *MISI (Jurnal Manajemen Informatika & Sistem Informasi)* Volume 3, No 2, hal. 146-156, Juni 2020.
- [13] P.W. Kastawan, D.M. Wiharta dan I.M. Sudarma, "Implementasi Algoritma C5.0 pada Penilaian Kinerja Pegawai Negeri Sipil", *Majalah Ilmiah Teknologi Elektro*, Vol. 17, No. 3, September-Desember 2018, hal. 371-376, 2018.
- [14] A.C. Rumahorbo dan K.A. Sekarwati, "Penerapan Data Mining dengan Menggunakan Algoritma C4.5 pada Klasifikasi Fasilitas Kesehatan Provinsi di Indonesia", *Jurnal Ilmiah KOMPUTASI*, Volume 19 No: 1, hal. 27-38, Maret 2020.
- [15] B. Stephen Cass. The 2016 top programming languages, <http://spectrum.ieee.org/computing/software/the-2016-top-programming-languages>, 2016.
- [16] Tim Smith, Kevin Ushey. Arrgh: a newcomer's (angry) guide to R, <http://arrgh.tim-smith.us/>, 2016.
- [17] A. Sucipto, "Prediksi Kredit Macet Melalui Perilaku Nasabah Pada Koperasi Simpan Pinjam Dengan Menggunakan Metode Algoritma Klasifikasi C4.5", *Jurnal DISPROTEK*, Volume 6 No: 1, hal. 75-87, Januari 2015.
- [18] N. Iriadi dan N. Nuraeni, "Kajian Penerapan Metode Klasifikasi Data Mining Algoritma C4.5 Untuk Prediksi Kelayakan Kredit Pada Bank Mayapada Jakarta", *Jurnal Teknik Komputer AMIK BSI*, Volume II No: 1, hal. 132-137, Februari 2016.
- [19] B. Hermanto, S.N. Azhari dan F.P. Putra, " Analisis Kinerja Decision Tree C4.5 dalam Prediksi Potensi Pelunasan Kredit Calon Debitur", *Jurnal INOVTEK POLBENG - SERI INFORMATIKA*, Volume. 2, No: 2, hal. 189-197, November 2017.
- [20] M.H. Rifqo dan T. Arzi, "Implementasi Algoritma C4.5 Untuk Menentukan Calon Debitur Dengan Mengukur Tingkat Risiko Kredit Pada Bank BRI Cabang Curup", *Pseudocode*, Volume III Nomor 2, hal. 83-90, September 2016.
- [21] S. Saprudin, "Penerapan Particle Swarm Optimization (Pso) Untuk Klasifikasi Dan Analisis Kredit Dengan Menggunakan Algoritma C4.5", *Jurnal Informatika Universitas Pamulang* Volume 2, No. 4, hal. 214-219, Desember 2017.
- [22] S. Masripah, "Evaluasi Penentuan Kelayakan Pemberian Kredit Koperasi Syariah Menggunakan Algoritma Klasifikasi C4.5", *Jurnal Pilar Nusa Mandiri* Vol. XI, No.1 hal. 1-10, Maret 2015.