

Implementasi C4.5 Dan Algoritma K-Nearest Neighbor Untuk Prediksi Kelayakan Pemberian Kredit Menggunakan RapidMiner Studio

Edi Junaedi
Universitas Buana Perjuangan
Karawang, Indonesia
if17.edijunaedi@mhs.ubpkarawang.ac.id

Amril Mutoi Siregar
Universitas Buana Perjuangan
Karawang, Indonesia
amrilmutoi@ubpkarawang.ac.id

Euis Nurlaelasari
Universitas Buana Perjuangan
Karawang, Indonesia
euis.nurlaelasari@ubpkarawang.ac.id

Abstract — Perusahaan *multifinance* hadir untuk mengakomodir keanekaragaman kebutuhan masyarakat mulai dari pembiayaan modal kerja, pembiayaan investasi, hingga pembiayaan multiguna sesuai persetujuan otoritas jasa keuangan (OJK). Selama memiliki persyaratan dan jaminan tertentu, pengajuan kredit dapat dengan mudah dilakukan. Namun, pemberian kredit memiliki resiko yang dapat merugikan perusahaan, sehingga perusahaan harus lebih selektif dalam memberikan pemberian kredit. Untuk meminimalisasi resiko kerugian, dilakukan suatu perhitungan yang dapat memprediksi untuk kelayakan pemberian kredit dengan memperhatikan data-data dari calon nasabah. Salah satu metode yang dapat digunakan untuk prediksi kelayakan pemberian kredit adalah algoritma *K-Nearest Neighbor* dan algoritma C4.5. Pada penelitian yang telah dilakukan, algoritma *K-Nearest Neighbor* mendapatkan akurasi sebesar 80% dan algoritma C4.5 mendapatkan akurasi sebesar 87% sehingga metode ini dapat digunakan untuk membantu perusahaan dalam meminimalisasi kerugian yang disebabkan karena kredit macet.

Kata kunci — C4.5, K-Nearest Neighbor, Kredit, Prediksi

I. PENDAHULUAN

Hadirnya perusahaan *multifinance* menyediakan pelayanan untuk mengakomodir kebutuhan masyarakat mulai dari pembiayaan jual beli, pembiayaan multiguna, pembiayaan investasi, pembiayaan Syariah dan pembiayaan lainnya sesuai dengan peraturan otoritas jasa keuangan (OJK) [1]. Permasalahan yang sering dihadapi perusahaan saat memberikan pembiayaan atau kredit pada nasabah adalah lainnya nasabah dalam melakukan pembayaran yang dipengaruhi oleh faktor kesengajaan atau kondisi diluar kemampuan nasabah sehingga berdampak pada kesehatan keuangan perusahaan [2][3].

Diperlukan peran teknologi yang dapat menggali informasi tersembunyi dari data calon nasabah yang mengajukan kredit sehingga dapat memprediksi kemampuan nasabah dalam melakukan pembayaran kredit. Salah satu bagian teknologi yang dapat membantu adalah *data mining* karena *data mining* dapat dijadikan acuan dalam pengambilan keputusan, *clustering*, prediksi, denganteknik-teknik dan algoritma yang dimiliki seperti *Association Rule*, algoritma C4.5, algoritma *K-Means*, dan algoritma *K-Nearest Neighbor* [4][5][6]. Sebelumnya, data mining berhasil mendapatkan akurasi kebenaran sebesar 82% untuk analisa resiko kredit [7] dan terbukti dapat digunakan untuk memprediksi mahasiswa yang lulus tepat waktu [1].

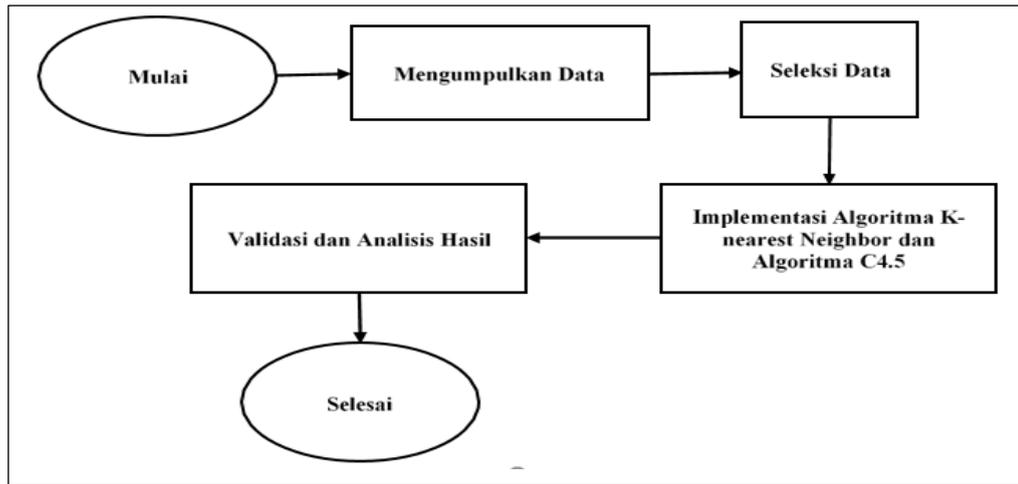
Teknik data mining yang biasa digunakan untuk prediksi adalah algoritman *K-nearest Neighbor* yang melakukan prediksi terhadap objek berdasarkan data *training* yang jaraknya paling dekat dengan objek tersebut[8] dan algoritma C4.5 yang merupakan pengembangan dari algoritma ID3 yang tidak memerlukan transformasi nilai setiap atribut pada dataset. Algoritma C4.5 dan algoritma *K-nearest Neighbor* pernah digunakan untuk memprediksi mahasiswa yang lulus tepat pada waktunya dengan data premier yang didapatkan dari pusat data Biro Administrasi Akademik dan Kemahasiswaan (BAAK) STMIK Widuri dan memperoleh hasil akurasi sebesar 90% dan algoritma C4.5 mendapatkan hasil akurasi sebesar 98.04% [9]. Selain itu, algoritma *K-nearest Neighbor* berhasil mendapatkan nilai akurasi 95% untuk klasifikasi pencemaran udara di kota Jakarta [10]. Pada penelitian yang dilakukan untuk klasifikasi penderita penyakit diabetes menggunakan algoritma C4.5 dengan menggunakan 16 variabel dataset gejala, dan 1 variabel kelas penentu klasifikasi menghasilkan akurasi yang cukup besar yaitu 97,12 % *Precision* sebesar 93,02% %, dan *Recall* sebesar 100,00% [11].

Berdasarkan data dan penelitian yang pernah dilakukan sebelumnya, maka penelitian ini akan menggunakan Algoritma C4.5 dan algoritma *K-nearest Neighbor* untuk prediksi kelayakan pemberian kredit.

II. DATA DAN METODE

A. Alur Penelitian

Berdasarkan kebutuhan perusahaan untuk memprediksi kelayakan pemberian kredit, dilakukan komputasi menggunakan data mining dan diharapkan dapat meminimalisasi resiko kredit macet yang dapat merugikan perusahaan. Maka, *flowchart* penelitian digambarkan sebagai berikut.



Gambar 1 Alur Penelitian

B. Algoritma *K-nearest Neighbor*

Algoritma *K-nearest Neighbor* merupakan jenis algoritma *supervised* yang paling tua dan populer, prinsip kerja algoritma *K-nearestNeighbor* melakukan prediksi terhadap objek berdasarkan data *training* yang jaraknya paling dekat dengan objek tersebut. Dekat atau jauhnya jarak dapat dihitung menggunakan rumus *euclidean distance* berikut.

$$distance = \sqrt{\sum_{i=1}^n (X_{i\text{training}} - X_{\text{testing}})^2} \tag{1}$$

Keterangan :

Dimana,

$(X_{i\text{training}})$: data *training* ke- *i*.

(X_{testing}) : data testing

i : record (baris) ke – *i* dari tabel

N : jumlah data *training*

Adapun tahapan untuk melakukan perhitungan dengan algoritma *K-nearest Neighbor* yaitu.

1. Tentukan parameter *K*
2. Lakukan pelatihan data untuk mendapatkan jarak antara data yang akan diprediksi.
3. Mengurutkan jarak hasil pelatihan dari yang terkecil hingga terbesar.
4. Cari jarak terdekat dengan nilai *K*.
5. Bandingkan kelas-kelas yang sesuai.
6. Tentukan banyaknya kelas tetangga terdekat dan masukan sebagai kelas yang akan diprediksi [8][12].

C. Algoritma C4.5

Algoritma C4.5 merupakan metode pohon keputusan yang dikembangkan dari algoritma ID3, algoritma C4.5 bekerja dengan membuat pohon keputusan berdasarkan data *training* yang telah disediakan [13][14]. Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 yaitu.

1. Menyiapkan data *training*. Data ini diambil dari data yang ada sebelumnya dan sudah dikelompokkan dalam kelas tertentu.
2. Setelah itu tentukan akar dari pohon. Pilih akar dari atribut, caranya adalah dengan menghitung nilai *gain* dari semua atribut, yang menjadi akar pertama adalah nilai *gain* yang paling tertinggi. Sebelum menentukan nilai *gain*, terlebih dahuluhitung nilai *entropy*, untuk menentukan nilai *entropy* yaitu dengan menggunakan rumus seperti pada persamaan berikut:

$$Entropy(S) = \sum_{i=1}^n -p_i \times \log_2 p_i \tag{2}$$

Keterangan :

- S = Himpunan Kasus
- n = Jumlah partisi S
- pi = Proporsi dari Si terhadap S.

3. Setelah itu tentukan nilai *gain* dengan rumus seperti pada persamaan berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i) \tag{3}$$

Keterangan :

- S = Kumpulan Kasus,
- A = Atribut,
- n = Jumlah anggota atribut A,
- |Si| = Jumlah kasus pada anggota ke i,
- |S| = Jumlah kasus dalam S

4. Setelah itu ulangi langkah ke-2 sampai semua *record* terpartisi secara sempurna.
5. Proses partisi pohon keputusan akan berhenti saat :
 - a. Data pada simpul N memiliki hasil yang sama.
 - b. Atribut hasil tidak ada yang dipecah lagi.
 - c. Semua *record* pada cabang tidak ada yang kosong [14].

D. RapidMiner

RapidMiner merupakan salah satu solusi data mining yang paling banyak digunakan diseluruh dunia. RapidMiner tidak hanya digunakan oleh peneliti di bidang *computer science* dan matematika saja, tetapi banyak juga digunakan untuk bisnis dan industri. Didalam dunia *computer science* dan matematika, rapid miner digunakan untuk pengembangan metode *data minuing*, *machine learning* dan metode statistik. RapidMiner dapat digunakan sebagai alat, karena menyediakan berbagai macam alat berbagai metodedari evaluasi statistik sederhana seperti analisis korelasi hingga regresi, klasifikasi dan prosedur pengelompokan serta pengurangandimensi dan optimasi parameter [15].

E. Persiapan Data

Hal pertama yang dilakukan yaitu mempersiapkan datasets yang diperoleh dari salah satu perusahaan multifinance cabang Karawang, yaitu data kredit motor. Datasets yang diperoleh berjumlah 3.863 data transaksi kredit motor, tetapi hanya di pakai sebanyak 50 datasets. Dari 50 datasets tersebut terdapat 24 anggota dengan keterangan “Lancar” dan 26 anggota dengan keterangan “Macet”. Setelah memperoleh datasets kemudian dilakukan perhitungan menggunakan algoritma K-Nearst Neighbor (KNN) danalgoritma C4.5.

Tabel 1 Contoh Data

<i>Nama Cust</i>	<i>Alamat</i>	<i>Pekerjaan</i>	<i>Penghasilan</i>	<i>Obj</i>
ADE RIDWAN	BAKAN LIO RT 029 RW 012. Kel Karyasari Kec Rengasdengklok Kabupaten Karawang	JASA	3890030	NMC
ARIP ROHMAN	DUSUN BAYUR I RT 13 RW 04 Kel Payungsari Kec Pedes Kab Karawang	BURUH/PRT	1322353	MPF
ONIN	KP SAYURAN RT 05 RW 03 Kel kertasari Kab. Bekasi	PEDAGANG	10505428	NMC
SAAN	DUSUN KRAJAN RT 02 RW 01 Kel Tambaksumur, Tirtajaya, Karawang	PETANI	7815815	NMC

F. Seleksi Data

Proses seleksi data yaitu membersihkan data dari atribut – atribut yang tidak diperlukan, data yang tidak diperlukan yaitu atribut kode kantor, nomor kontrak dan jumlah angsuran. Setelah tahap penyeleksian dilakukan, maka akan menghasilkan dataset yang akan segera diolah.

Hasil seleksi data adalah data nama customer, alamat, pekerjaan, penghasilan dan object/jenis kredit. Contoh dari hasil seleksi data bisa dilihat pada tabel 2.

Tabel 2 Dataset

No.	Nama	Penghasilan	Pengajuan	Tenor	ST	Keterangan
1	Saepudin	15 juta – 20 juta	10 juta – 15 juta	12	KPR	Lancar
2	Paisal	1 juta – 5 juta	1 juta – 5 juta	12	ROT	Lancar
3	Haris	5 juta – 10 juta	5 juta – 10 juta	6	KPR	Macet
4	Samsudin	15 juta – 20 juta	10 juta – 15 juta	12	KPR	Lancar
5	Yanto	15 juta – 20 juta	10 juta – 15 juta	12	KPR	Lancar
6	Tatang	15 juta – 20 juta	10 juta – 15 juta	12	KPR	Macet
7	Carta	5 juta – 10 juta	5 juta – 10 juta	12	KPR	Macet
.						
.						
.						
50	Much. N	10 juta – 15 juta	15 juta – 20 juta	12	KOS	Macet

ST = Status Tinggal
KPR = Kredit Pemilik Rumah
ROT = Rumah Orang Tua
KOS = Rumah Kos Kosan

Dari empat belas atribut yang terdapat pada data asli, hanya lima atribut yang dipilih karena lebih bisa dijadikan acuan dalam melakukan prediksi. Atribut yang digunakan kemudian dikonversi menjadi sebuah nilai yang dapat diolah menggunakan algoritma KNN, konversi yang dilakukan dengan mengelompokkan data dengan kriteria 1="buruk", 2="Sedang", 3="Cukup baik", 4="Sangat Baik" berdasarkan wawancara yang dilakukan dengan supervisor perusahaan *mulifinance* cabang Karawang.

1. Penghasilan

Atribut penghasilan merupakan penghasilan bulanan dari para nasabah. Pada variabel penghasilan dilakukan konversi ke sebuah nilai, ditulis dengan tabel berikut.

Tabel 3 Konversi Nilai Atribut Penghasilan

Pengajuan	Nilai
1 juta – 5 juta	1
5 juta – 10 juta	2
10 juta – 15 juta	3
15 juta – 20 Juta	4

Pada tabel 3. penghasilan 1 juta - 5 juta, 5 juta – 10 juta, 10 juta – 15 juta, dan 15 juta – 20 juta, di konversikan kedalam nilai satu, dua, tiga, dan empat.

2. Pengajuan

Atribut pengajuan merupakan jumlah nominal pinjaman yang dilakukan oleh nasabah. Pada variabel pengajuan jugadilakukan proses konversi ke sebuah nilai, disajikan pada tabel berikut.

Tabel 4 Konversi Nilai Atribut Pengajuan

Pengajuan	Nilai
1 juta – 5 juta	4
5 juta – 10 juta	3
10 juta – 15 juta	2
15 juta – 20 juta	1

variabel 1 juta – 5 juta di konversikan ke nilai empat, 5 juta – 10 juta ke nilai tiga, 10 juta – 15 juta ke nilai dua, dan 15 juta - 20 juta ke nilai satu.

3. Tenor

Atribut tenor merupakan jangka waktu dari anggota kredit. Dan pada variabel dari tenor di konversikan ke dalam nilai

Tabel 5 Konversi Nilai Atribut Tenor

Tenor	Nilai
6	1
12	2
18	3
24	4

variabel 6 dikonversikan kedalam nilai satu, variabel 12 ke nilai dua, variabel 18 ke nilai tiga, dan variabel 24 ke nilai empat.

4. Status Tinggal (ST)

Atribut status tinggal merupakan status tempat tinggal dari pada anggota kredit. Pada variabel status tinggal juga dikonversikan kedalam nilai.

Tabel 6 Konversi Nilai Atribut Penghasilan

Status Tinggal	Nilai
KPR	1
KOS	2
ROT	3
MS	4

dijelaskan bahwa variabel KPR di konversi ke nilai satu, variabel KOS ke nilai dua, variabel ROT ke nilai tiga, dan variabel MS ke nilai empat.

III. HASIL DAN PEMBAHASAN

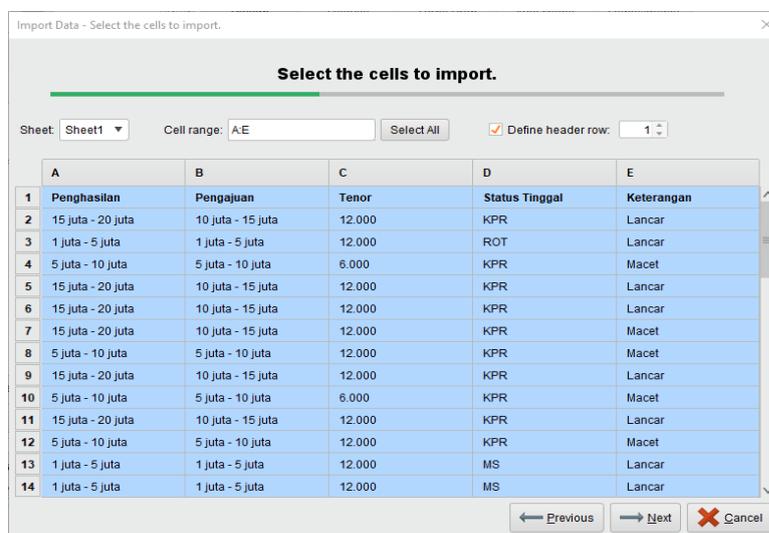
Sebelum Anda mulai memformat naskah Anda, pertama-tama tulis dan simpan konten sebagai file teks terpisah. Lengkapi semua konten dan pengubahan organisasi sebelum memformat. Harap perhatikan bagian A-D di bawah ini untuk informasi lebih lanjut tentang pengoreksian ulang, pengejaan dan tata bahasa.

Sebaiknya pisahkan dokumen teks dan grafik (tabel dan gambar) Anda hingga teks diformat dan ditata pada naskah ini. Jangankan tab, jangan menambahkan jenis penomoran halaman apa pun di naskah. Jangan beri nomor kepala teks – editor penerbit akan melakukan itu untuk Anda.

A. Perhitungan Algoritma *K-Nearest Neighbor*

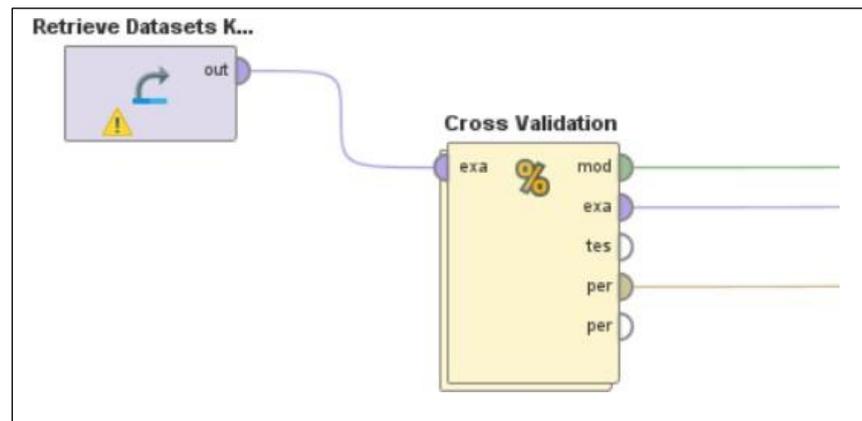
Adapun langkah untuk melakukan perhitungan algoritma *K-Nearest Neighbor* menggunakan rapidminer studio adalah:

1. Memasukan *datasets* yang sudah disiapkan, kemudian klik *import data* pada *Rapidminer*.



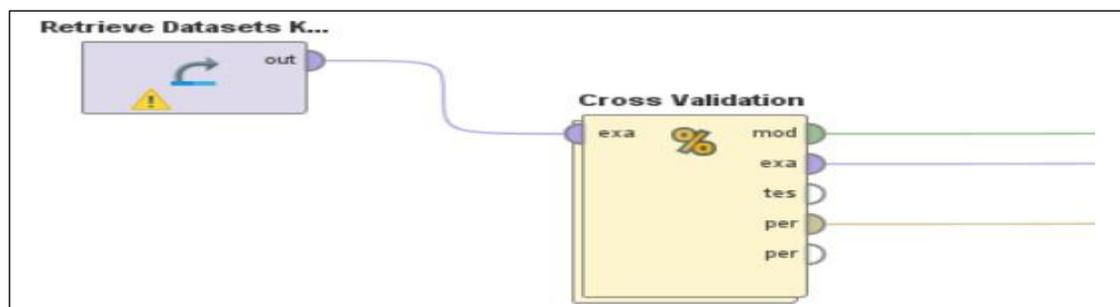
Gambar 2 Import Dataset

2. Drag operator yang diperlukan untuk mengolah *datasets*, yaitu *datasets* yang sudah di *upload* dan operator *cross validation*.



Gambar 3 Memasukan Operator *Cross validation*

3. Klik operator *cross validation*, kemudian drag operator *KNN*, *apply model*, dan *performance classification*.



Gambar 4 Pengaturan Metode dan performa

4. Kemudian hubungkan semua operator dan klik *run* untuk melihat hasilnya.

accuracy: 80.00% +/- 13.33% (micro average: 80.00%)

	true Lancar	true Macet	class prec
pred. Lancar	20	6	76.92%
pred. Macet	4	20	83.33%
class recall	83.33%	76.92%	

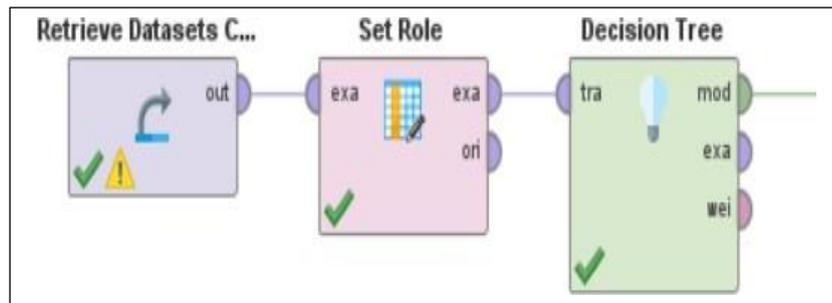
Gambar 5 Hasil Perhitungan *K-Neares Neighbor*

Pada Gambar 5 menunjukkan hasil prediksi kredit macet yang dilakukan menggunakan algoritma *K-Neares Neighbor* mendapatkan hasil akurasi sebesar 80% dengan prediksi lancar mendapatkan nilai akurasi 76.92% dan *class recal* untuk data yang terbaca lancar sebesar 83.33%

B. Perhitungan Algoritma C4.5

Untuk menghitung prediksi kelayakan pemberian kredit dengan algoritma C4.5 menggunakan rapidminer studio, dapat dilakukan dengan tahapan sebagai berikut.

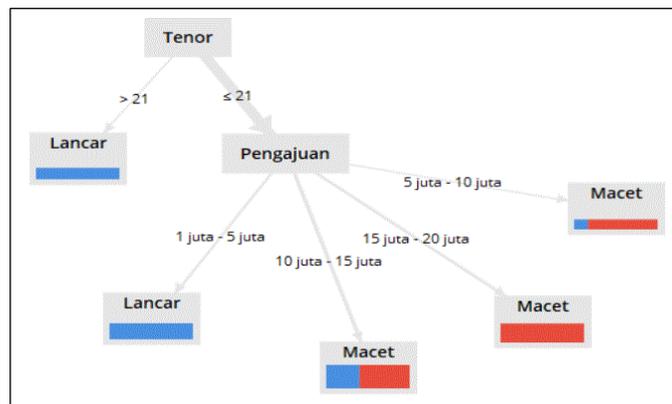
1. Memasukan *datasets* yang sudah disiapkan, kemudian klik *import* data pada *Rapidminer*. Proses import dataset C4.5 samadengan import dataset *K-Neares Neighbor* pada gambar 2.
2. Drag operator yang diperlukan untuk mengolah *datasets*, yaitu *datasets* yang sudah di *upload* dan operator *set role*, dan *decision tree*.



Gambar 6 Proses Pembuatan Pohon Keputusan

Operator *set role* berfungsi untuk menentukan label dari *datasets* yang sudah di *upload*, dan operator *decision tree* merupakan algoritma yang digunakan.

- Selanjutnya hubungkan semua operator, kemudian klik *run*, untuk melihat hasil dari pohon keputusan.



Gambar 7 Pohon Keputusan C4.5

- Untuk melihat hasil akurasi, hanya menambahkan operator *apply model*, dan *performance* saja, lalu hubungkan, dan klik *run* untuk melihat hasil akurasi.

accuracy: 86.00%			
	true Lancar	true Macet	class precision
pred. Lancar	17	0	100.00%
pred. Macet	7	26	78.79%
class recall	70.83%	100.00%	

Gambar 8 Hasil Perhitungan C4.5

Kemudian untuk hasil akurasi yang diperoleh dari *datasets* yang telah diproses oleh *rapidminer* dengan menggunakan algoritma C4.5 adalah 86%.

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Kesimpulan yang didapatkan dari penelitian yang telah dilakukan adalah mengetahui bahwa Penerapan algoritma KNN dan C4.5, untuk kelayakan kredit dapat bekerja dengan baik. Dengan algoritma KNN dan C4.5. Hasil akurasi algoritma C4.5 terbukti lebih baik dari algoritma KNN, dapat dilihat pada nilai akurasi dari pengujian menggunakan *rapidminer*. Untuk pengujian *rapidminer* algoritma KNN diperoleh nilai akurasi 80%, sedangkan algoritma C4.5 diperoleh nilai akurasi 86%.

B. Saran

Adapun saran untuk penelitian ini yaitu gunakan algoritma dan metode lain supaya didapatkan hasil prediksi terbaik, sehingga dapat diterapkan pada perusahaan untuk prediksi kelayakan pemberian kredit.

PENGAKUAN

Naskah ilmiah ini merupakan ekstraksi dari penelitian Tugas Akhir milik Edi Junaedi dengan Model Prediksi Untuk Kelayakan Pemberian Kredit Dengan Algoritma K-Nearest Neighbor Dan C4.5, yang dibimbing oleh Amril Mutoi Siregar, M.Kom dan Euis Nurlaelasari, M.Kom.

DAFTAR PUSTAKA

- [1]. Pt Federal International Finance, "Fifgroup," Multifinance, 1 Mei 2013. [Online]. Available: <https://www.fifgroup.co.id/informasi-perusahaan/informasi-umum>. [Diakses 15 Agustus 2021].
- [2]. Amrin, "Perbandingan Metode Neural Network Model Radial Basis Function Dan Multilayer Perceptron Untuk Analisa Risiko Kredit Mobil," Vol. Xx, No. 1, Pp. 31–38, 2018.
- [3]. R. Bimantara, "Penyelesaian Kredit Macet Perseroan Melalui Eksekusi Jaminan Hak Tanggungan Atas Nama Pribadi," *J. Bina Mulia Huk.*, Vol. 3, No. 4, 2019.
- [4]. F. Hadi, "Penerapan Metode Algoritma C4.5 Dalam Menganalisa Pegajian Kredit Pada Koperasi Jasa Keuangan Syariah Kelurahan Limau Manis Selatan," *Indones. J. Comput. Sci.*, Vol. 7, No. 1, Pp. 28–42, 2018.
- [5]. D. Feblian And D. U. Daihani, "Implementasi Model Crisp-Dm Untuk Menentukan Sales Pipeline Pada Pt X," *J. Tek. Ind.*, Vol. 6, No. 1, Pp. 1–12, 2017.
- [6]. N. A. Hasibuan *Et Al.*, "Implementasi Data Mining Untuk Pengaturan Layout," *J. Ris. Komput.*, Vol. 4, No. 4, Pp. 6–11, 2017.
- [7]. D. Setiawati, I. Taufik, And W. B. Z, "Klasifikasi Terjemahan Ayat Al-Quran Tentang Ilmu Sains Menggunakan Algoritma Decision Tree Berbasis Mobile," Vol. I, No. 1, Pp. 24–27, 2016.
- [8]. R. Kesuma And H. Akbar, "Algoritma K-Nearest Neighbor Dengan Euclidean Distance Dan Manhattan Distance Untuk Klasifikasi Transportasi Bus," Vol. 12, No. 2, Pp. 104–111, 2020.
- [9]. A. Budiyantra, Irwansyah, E. Prengki, P. A. Pratama, And N. Wiliani, "Komparasi Algoritma Decision Tree, Naive Bayes Dan K-Nearest Neighbor Untuk Memprediksi Mahasiswa Lulus," Vol. 5, No. 2, Pp. 265–270, 2020.
- [10]. S. Nurjanah, A. M. Siregar, And D. S. Kusumaningrum, "Penerapan Algoritma K – Nearest Neighbor (Knn) Untuk Klasifikasi Pencemaran Udara Di Kota Jakarta," *Technol. Sci.*, Vol. 1, Pp. 71–76, 2020.
- [11]. F. M. Hana, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5," *J. Sist. Komput. Dan Kecerdasan Buatan*, Vol. 4, No. 1, 2020.
- [12]. J. Sinaga, "Data Mining Klasifikasi Nasabah Dalam Pengajuan Kredit Tanpa Agunan Dengan Algoritma K-Nearest Neighbor," *Jikomsi J. Ilmu Komput. Dan Sist. Informas*, Vol. 1, No. 1, Pp. 22–28, 2018.
- [13]. Y. Widiastiwi And I. Ernawati, "Klasifikasi Penyakit Batu Ginjal Menggunakan Algoritma Decision Tree C4.5 Dengan Membandingkan Hasil Uji Akurasi," *J. Ikra-Ith Inform.*, Vol. 5, No. 2, Pp. 128–135, 2021.
- [14]. R. Nofitri And J. Eska, "Implementasi Data Mining Klasifikasi C4.5 Dalam Menentukan Kelayakan Pengambilan Kredit," *Amik R.*, Vol. 9986, No. September, Pp. 307–310, 2018.
- [15]. Rapidminer, "Rapidminer, Inc.," 2020. [Online]. Available: <https://rapidminer.com/>. [Diakses 17 Januari 2021].