

PENERAPAN ALGORITMA C4.5 UNTUK PENENTUAN KELAYAKAN KREDIT

Siti Nur Khasanah

Sistem Informasi

STMik Nusa Mandiri Jakarta

Jl. Damai No 8 Warung Jati Barat Margasatwa, Jakarta Selatan

siti.skx@nusamandiri.ac.id

Abstract— Bad credit is one of the credit risk faced by the financial industry. Bad credit happens if in the long term, financial institutions can not attract loans within the stipulated time. Credit scorings can be used to assist in determining the feasibility analysis of the credit applicant. With proper credit scoring models, banks can evaluate whether an applicant is eligible to get credit or not. In this study, carried out the classification of the customers who are categorize as good customers and bad customers by algorithm C4.5. The test results obtained for the accuracy value of C4.5 88.52%.

Keyword : algorithm classifier, c4.5, credit scoring

Intisari—Kredit macet merupakan salah satu resiko kredit yang dihadapi oleh industri keuangan. Kredit macet terjadi apabila dalam jangka panjang, lembaga keuangan tidak dapat menarik pinjaman kredit dalam waktu yang ditentukan. Penilaian kredit dapat digunakan untuk membantu analisa kredit dalam menentukan kelayakan pemohon. Dengan model penilaian kredit yang tepat, bank dapat mengevaluasi apakah pemohon layak untuk mendapatkan kredit atau tidak. Pada penelitian ini, dilakukan klasifikasi terhadap nasabah yang dikategorikan nasabah lancar dan bermasalah dengan membandingkan algoritma C4.5. Dari hasil pengujian kemudian didapatkan nilai akurasi C4.5 sebesar 88.52%.

Kata Kunci: algoritma klasifikasi, c4.5, penilaian kredit

PENDAHULUAN

Salah satu sistem pembiayaan alternative yang cukup berperan aktif dalam menunjang dunia usaha akhir-akhir ini yaitu pembiayaan konsumen atau dikenal dengan istilah customer finance. Berdasarkan pasal 1 pada Keputusan Presiden No 61 Tahun 1988 tentang Lembaga Pembiayaan, perusahaan pembiayaan konsumen adalah badan usaha yang melakukan pembiayaan pengadaan barang untuk

kebutuhan konsumen dengan sistem pembayaran berkala. Sejak diumumkan Paket Kebijakan 20 Desember 1988, mulai diperkenalkan pranata hukum di Indonesia, salah satu diantaranya adalah pembiayaan konsumen. Dengan memanfaatkan lembaga pembiayaan ini, masyarakat yang tadinya sulit untuk membeli barang kebutuhannya secara tunai, kini dengan bantuan pembiayaan konsumen kebutuhan mereka dapat terpenuhi. Konsumen yang berkepentingan menghubungi perusahaan pembiayaan konsumen agar dapat membayar secara tunai harga barang kebutuhan yang dibelinya dari pemasok (Supplier) dengan ketentuan pembayaran kembali harga barang itu kepada perusahaan konsumen dilakukan secara angsuran. Dengan cara demikian, kebutuhan masyarakat dapat terpenuhi secara wajar. Pada akhirnya banyak masyarakat yang menggunakan jasa perusahaan finansial dengan memanfaatkan jasa lembaga pembiayaan dalam bentuk pembiayaan konsumen untuk memenuhi sandang, pangan dan papan.

PT. Mitra Dana Putra Utama Finance merupakan salah satu perusahaan pembiayaan konsumen yang juga menerapkan prinsip umum yang berlaku dalam perkreditan. PT. Mitra Dana Putra Utama Finance menetapkan standar untuk menerima dan menolak konsumen. Prinsip-prinsip yang dimaksud adalah Character, Capacity, Capital, Collateral dan Condition atau yang biasa disebut sebagai 5C. Pada kenyataannya, bisnis pembiayaan konsumen ini memang bukanlah tanpa resiko. Salah satu resiko adalah macetnya pembayaran angsuran oleh konsumen, dalam hal ini berarti adanya kredit bermasalah.

Dalam proses pemberian kredit pada PT. Mitra Dana Putra Utama Finance, pemberian kredit dilakukan dengan menganalisa data-data para calon debitur. Data tersebut dianalisa oleh seorang analisa kredit. Adanya seorang analisa kredit, masih saja terdapat debitur-debitur yang macet dikarenakan kurang akuratnya (Swastina, 2013)(Iriadi & Nuraini, 2016) analisa kredit dalam memberikan analisisnya. Salah satu tahapan dalam pengevaluasian dan penilaian tersebut adalah dengan melakukan

pengklasifikasian nasabah berdasarkan kriteria-kriteria tertentu yang telah menjadi standar di setiap Bank (Rani, 2015)

Beberapa tahun terakhir banyak metode yang digunakan untuk klasifikasi data (Santoso, 2014). Klasifikasi data mining dengan model algoritma C4.5 dilakukan dengan pengujian-pengujian yang terukur melalui uji AUC, ROC dan T-Test dengan bantuan rapid miner (Sucipto, 2015). Akurasi pengujian dengan menggunakan algoritma C4.5 masih baik dan dapat dijadikan salah satu pedoman untuk deteksi kredit macet sebelum ada keputusan member atau tidak calon nasabah baru (Sucipto, 2015).

Berdasarkan Khasanah (2017) di atas maka pada penelitian ini akan menggunakan model pohon keputusan dengan algoritma C4.5. Model ini akan digunakan untuk memprediksi apakah debitur nantinya akan bermasalah dalam pembayaran kreditnya atau tidak. Model C4.5 dipilih karena proses learning dan klasifikasi pohon keputusan sederhana dan cepat. Secara umum, hasil klasifikasi metode pohon keputusan juga lebih mudah dipahami dan diinterpretasikan (Rahmayuni, 2014).

BAHAN DAN METODE

Metode penelitian kuantitatif dibagi menjadi tiga sub kategori yaitu inferentia, experimental dan simulasi. Jenis penelitian yang digunakan pada penelitian ini adalah penelitian Experiment. Penelitian eksperimen melibatkan penyelidikan perlakuan parameter/variabel tergantung pada penelitinya dan menggunakan tes yang dikendalikan oleh si peneliti itu sendiri. Dalam penelitian ini akan digunakan algoritma *C4.5 Classifier* Jenis data yang digunakan adalah data pengajuan kredit pada PT. Mitra Dana Putra Utama Finance cabang Jurumudi, sedangkan software yang digunakan dalam penelitian ini adalah Rapidminer 5.3 dan Microsoft Excell 2013.

Pada penelitian ini, teknik pengambilan sampel dengan menggunakan rumus Slovin. Rumus Slovin digunakan untuk menentukan ukuran sampel minimal (n) jika diketahui ukuran populasi (N) pada taraf signifikansi:

$$s = \frac{N}{1 + Ne^2}$$

Keterangan:

S = Jumlah Sampel

N = Jumlah Populasi

E = Taraf Signifikansi

Populasi yang terdapat pada PT. Mitra Dana Putra Utama Finance sebesar 1200

nasabah. Maka jumlah sampel dapat dihitung sebagai berikut:

$$S = \frac{1200}{1 + 1200(5\%)^2} = 300$$

Berdasarkan perhitungan di atas, maka sampel data yang digunakan sebanyak 300 data. Data yang digunakan dalam penelitian ini adalah sampel data nasabah selama tiga tahun, yaitu sampel data dari tahun 2012, 2013 dan 2014. Data yang dikumpulkan sebanyak 300 data yang terdiri dari 218 data nasabah lancar dan 82 data nasabah bermasalah.

HASIL DAN PEMBAHASAN

Pada metode penelitian eksperimen, digunakan model proses CRISP-DM (Cross- Standard Industry Process for Data Mining) yang terdiri dari 6 tahapan (Larose, 2005):

1) Tahap *Bussiness Understanding*

Berdasarkan data nasabah kredit pada PT. Mitra Dana Putra Utama Finance Tahun 2012, 2013 dan 2014 terlihat bahwa nasabah dengan status kreditnya macet cukup banyak, hal ini menjadi sebuah masalah dan kendala bagi PT. Mitra Dana Putra Utama Finance khususnya karena akan berakibat pada kurang akuratnya analisa kredit. Dalam beberapa penelitian telah dilakukan proses analisa kredit dengan menggunakan algoritma C4.5 Classifier dengan tujuan analisa yang dilakukan akan lebih akurat.

2) Tahap *Data Understanding*

Atribut atau variabel yang ada sebanyak 12 atribut (data lengkapnya bisa dilihat dilampiran). Setelah dilakukan proses data understanding, atribut atau variabel yang digunakan terdiri dari 10 atribut yang terdapat dalam data status kredit nasabah. Variabel-variabel tersebut ada yang tergolong variabel prediktor atau pemrediksi (Predictor Variable) yaitu variabel yang dijadikan dasar sebagai penentu resiko kredit (Susanto & Suryadi, 2010). Variabel yang digunakan yaitu jenis kelamin, radius survey, status domisili, pekerjaan, besar penghasilan, jumlah peminjaman, tenor, jumlah tanggungan, status perkawinan dan kategori. Sedangkan variabel tujuan adalah kategori nasabah (yang menunjukkan status kredit lancar atau bermasalah).

3) Tahap *Data Preparation*

Beberapa teknik yang dapat dilakukan untuk mendapatkan dataset yang berkualitas tinggi untuk digunakan dalam analisis data mining diantaranya adalah (Vercellis, 2009);

a. Data Validation

Data validation dilakukan untuk mengidentifikasi dan menghapus anomali dan inkonsistensi data. Anomali dan inkonsistensi data dapat terjadi dikarenakan adanya missing values dan juga karena data yang dipakai mengandung nilai-nilai yang salah atau disebut noise atau outlier.

b. Data Integration dan data transformation

Data Integration dan data transformation diterapkan pada beberapa dataset untuk meningkatkan akurasi dan efisiensi dari model pembelajaran algoritma yang digunakan. Pada tahap ini data yang akan diolah ditransformasikan ke dalam bentuk angka.

Tabel 1. Tabel Atribut yang Digunakan

No	ATRIBUT	KETERANGAN	KATEGORI NILAI
1	JENIS KELAMIN	LAKI-LAKI	1
		PEREMPUAN	2
2	RADIUS SURVEY	1-5 KM	1
		6-10 KM	2
		11-15 KM	3
		16-20 KM	4
3	STATUS DOMISILI	MILIK SENDIRI	1
		MILIK KELUARGA	2
		KONTRAK	3
4	PEKERJAAN	KARYAWAN	1
		WIRASWASTA	2
5	BESAR PENGHASILAN	1.000.000 sd 2.500.000	1
		2.600.000 SD 4.000.000	2
		4.100.000 SD 5.500.000	3
		5600000 SD 7.000.000	4
		7JT UP	5
6	JUMLAH PEMINJAMAN	1.000.000 sd 2.500.000	1
		3.000.000 sd 4.500.000	2
		5.000.000 sd 6.000.000	3
		6.500.000 sd 8.000.000	4
		8.500.000 sd 10.000.000	5
7	TENOR	6 bulan	1
		9 bulan	2
		12 bulan	3
		18 bulan	4
8	JUMLAH TANGGUNGAN	0 sd 2 orang	1
		3 sd 6 orang	2
		7 orang up	3
9	STATUS PERKAWINAN	KAWIN	1
		TIDAK KAWIN	2
10	KELAS	LANCAR	Y
		BERMASALAH	N

Sumber : Khasanah (2017) (2017)

4) Tahap Modelling

Pada tahap ini, data preparation yang telah didapatkan akan dilakukan pemodelan dengan menggunakan Algoritma C4.5. Tahap modelling dilakukan dengan dua cara yaitu dengan cara manual dan dengan menggunakan software Rapid Miner.

Perhitungan manual untuk pembuatan model algoritma C4.5 dilakukan dengan cara mencari nilai gain tertinggi dari setiap atribut. Sedangkan untuk pembuatan model Naive Bayes Classifier dilakukan dengan menghitung probabilitas prior dan probabilitas posterior. Sedangkan dalam pembuatan modelling dengan menggunakan software rapid miner. Penjelasan secara lengkap tentang model yang digunakan dan cara perhitungan manualnya.

5) Tahap Evaluation

Pada tahap evaluasi, disebut tahap klasifikasi karena pada tahap ini akan ditentukan pengujian untuk akurasi. Tahap pengujiannya adalah melihat hasil akurasi pada proses klasifikasi Algoritma C.45 classifier serta evaluasi dengan ROC Curve.

Pada tahap eksperimen berikutnya yaitu menghitung dan mendapatkan rule-rule yang ada pada model algoritma yang diusulkan. Dimana dalam penelitian ini menggunakan metode C4.5 classifier.

Dalam membuat pohon keputusan terlebih dahulu menghitung jumlah class nasabah yang memiliki kredit lancar dan kredit bermasalah. Setelah diketahui berapa total nasabah yang memiliki kredit lancar dan bermasalah kemudian hitung entropy dari masing-masing class berdasarkan atribut yang ditentukan dengan menggunakan data training.

Total data = 270
 Total Kredit Lancar = 197
 Total Kredit Bermasalah = 73
 $Entropy (Total) = (- \frac{197}{270} \log_2 (\frac{197}{270})) + (- \frac{73}{270} \log_2 (\frac{73}{270}))$
 Entropy (Total) = **0,84200**

Untuk mendapatkan gain tiap atribut, maka harus dihitung entropy tiap atribut berdasarkan tiap kasus. Hasil dari perhitungan nilai gain dan entropy pada data nasabah dengan menggunakan perbandingan data training dan data testing sebesar 90% dan 10% ditunjukkan pada tabel berikut:

Tabel 2. Perhitungan Gain Node 1

Atribut	Jumlah Kasus (S)	Lancar (Si) Y	Bermasalah (Si) N	Entropy	Gain	
Total	270	1	2	0,84200		
Jenis Kelamin	1	188	143	45	0,79398	0,00786
	2	82	54	28	0,92621	
Radius Survey	1	181	174	7	0,23618	0,41720
	2	79	22	57	0,85337	
	3	9	1	8	0,50326	
	4	1	0	1	0,00000	
Status Domisili	1	149	141	8	0,30187	0,36201
	2	49	41	8	0,64206	
	3	72	15	57	0,73828	
Pekerjaan	1	218	164	54	0,80762	0,00752
	2	52	33	19	0,94706	
Besar Penghasilan	1	81	44	37	0,99461	0,08499
	2	141	108	33	0,78499	
	3	36	34	2	0,30954	
	4	10	10	0	0,00000	
	5	2	1	1	1,00000	
Jumlah Peminjaman	1	70	53	17	0,79976	0,00924
	2	116	82	34	0,87270	
	3	75	57	18	0,79504	
	4	8	4	4	1,00000	
Tenor	5	1	1	0	0,00000	
	1	11	9	2	0,68404	0,01398
	2	25	20	5	0,72193	
	3	228	166	62	0,84422	
Jumlah Tanggungan	4	6	2	4	0,91830	
	1	114	105	9	0,39846	0,18167
	2	120	83	37	0,89124	
	3	29	9	20	0,89357	
Status Perkawinan	1	202	139	63	0,89333	0,02043
	2	68	58	10	0,60243	

Sumber: Khasanah (2017) (2017)

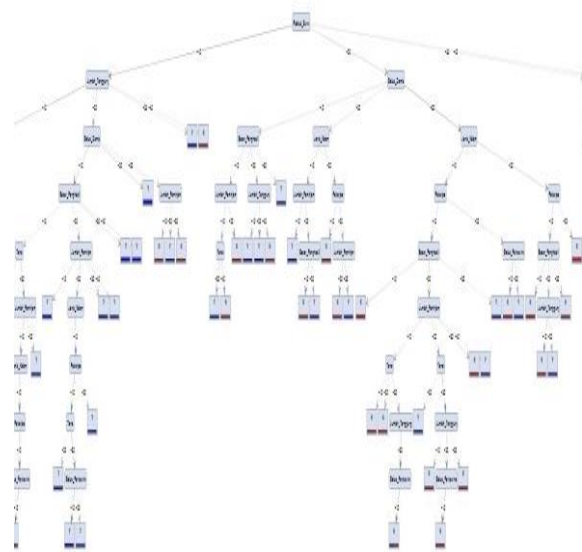
Dari hasil hitungan entropy dan gain pada Tabel 1 terlihat bahwa atribut radius survey yang diberi warna merah memiliki nilai gain paling tinggi yaitu 0,41720. Oleh karena itu atribut radius survey menjadi akar atau node pertama dari pohon keputusan yang terbentuk. Pada atribut radius survey terdapat 4 cabang, yaitu cabang 1, 2, 3 dan 4. Cabang-cabang tersebut belum terklasifikasi sehingga perlu dilakukan perhitungan nilai entropy dan gain untuk masing-masing cabang tersebut. Hasil perhitungan untuk node 1.1 dapat dilihat pada tabel 2 di bawah ini:

Tabel 3. Perhitungan Gain Node 1.1

Atribut	Jumlah Kasus (S)	Lancar (Si)		Bermasalah (Si)		Entropy	Gain
		Y	N	Y	N		
Total	181	1	2			0,23618	
Jenis Kelamin	1	131	125	6		0,26829	0,00293
	2	50	49	1		0,14144	
Status Domisili	1	132	130	2		0,11327	0,08453
	2	36	36	0		0,00000	
	3	13	8	5		0,96124	
Pekerjaan	1	151	145	6		0,24108	0,00011
	2	30	29	1		0,21084	
Besarnya Penghasilan	1	44	39	5		0,51079	0,05682
	2	94	93	1		0,08500	
	3	31	31	0		0,00000	
	4	10	10	0		0,00000	
	5	2	1	1		1,00000	
Jumlah Peminjaman	1	49	47	2		0,24602	0,01307
	2	76	74	2		0,17557	
	3	52	50	2		0,23519	
	4	3	2	1		0,91830	
	5	1	1	0		0,00000	
Tenor	1	8	8	0		0,00000	0,00921
	2	18	18	0		0,00000	
	3	154	147	7		0,26676	
	4	1	1	0		0,00000	
Jumlah Tanggungan	1	99	99	0		0,00000	0,11350
	2	71	67	4		0,31273	
	3	8	8	0		0,00000	
Status Perkawinan	1	127	120	7		0,30775	0,02024
	2	54	54	0		0,00000	

Sumber: Khasanah (2017) (2017)

Dari hasil perhitungan entropy dan gain yang terdapat pada table 4.2 terlihat bahwa atribut jumlah tanggungan mempunyai nilai paling tinggi yaitu 0,11350, Pada atribut jumlah tanggungan terdapat tiga cabang yaitu 1, 2 dan 3. Setelah dilakukan hasil perhitungan entropy dan gain, maka pohon keputusan yang terbentuk akan terlihat seperti gambar 1 berikut:



Sumber: Khasanah (2017) (2017)
Gambar 1. Pohon Keputusan yang dibentuk

Dari pohon keputusan yang terbentuk pada gambar 1 diperoleh rule model sebagai berikut:

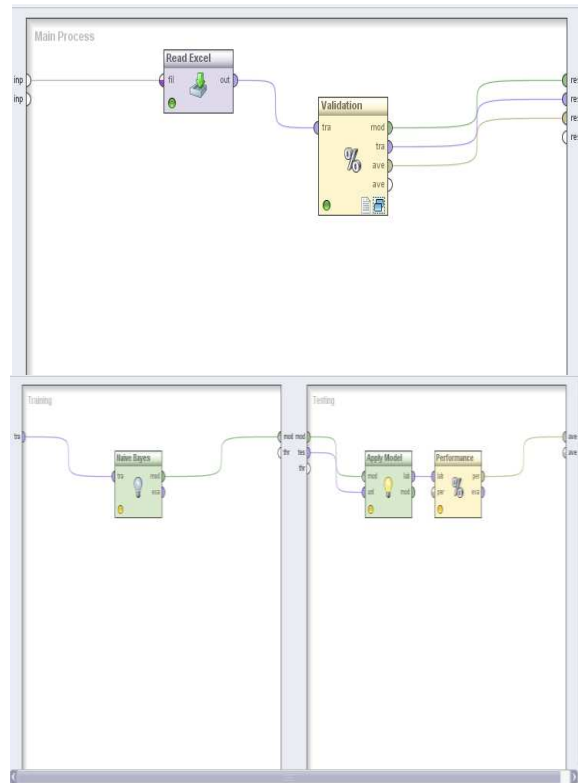
- 1) if Radius_Survey = 1.0 and Jumlah_Tanggungan = 1.0 then Y (99 / 0)
- 2) if Radius_Survey = 1.0 and Jumlah_Tanggungan = 2.0 and Status_Domisili = 1.0 and Besarnya Penghasilan = 1.0 and Tenor = 2.0 then Y (2 / 0)
- 3) if Radius_Survey = 1.0 and Jumlah_Tanggungan = 2.0 and Status_Domisili = 1.0 and Besarnya Penghasilan = 1.0 and Tenor = 3.0 and Jumlah_Peminjaman = 1.0 and Jenis_Kelamin = 1.0 and Pekerjaan = 1.0 and Status_Perkawinan = 1.0 then Y (2 / 1)
- 4) if Radius_Survey = 1.0 and Jumlah_Tanggungan = 2.0 and Status_Domisili = 1.0 and Besarnya Penghasilan = 1.0 and Tenor = 3.0 and Jumlah_Peminjaman = 2.0 then Y (1 / 0)
- 5) if Radius_Survey = 1.0 and Jumlah_Tanggungan = 2.0 and Status_Domisili = 1.0 and Besarnya Penghasilan = 2.0 and Jumlah_Peminjaman = 1.0 then Y (9 / 0)
- 6) if Radius_Survey = 1.0 and Jumlah_Tanggungan = 2.0 and Status_Domisili = 1.0 and Besarnya Penghasilan = 2.0 and Jumlah_Peminjaman = 2.0 and Jenis_Kelamin = 1.0 and Pekerjaan = 1.0 and Tenor = 2.0 then Y (2 / 0)
- 7) if Radius_Survey = 1.0 and Jumlah_Tanggungan = 2.0 and Status_Domisili = 1.0 and Besarnya Penghasilan = 2.0 and Jumlah_Peminjaman = 2.0 and Jenis_Kelamin = 1.0 and Pekerjaan = 1.0 and

- Tenor = 3.0 and Status_Perkawinan = 1.0 then Y (11 / 1)
- 8) if Radius_Survey = 1.0 and Jumlah_Tanggungan = 2.0 and Status_Domisili = 1.0 and Besar_Penghasilan = 2.0 and Jumlah_Peminjaman = 2.0 and Jenis_Kelamin = 1.0 and Pekerjaan = 1.0 and Tenor = 3.0 and Status_Perkawinan = 2.0 then Y (1 / 0)
 - 9) if Radius_Survey = 1.0 and Jumlah_Tanggungan = 2.0 and Status_Domisili = 1.0 and Besar_Penghasilan = 2.0 and Jumlah_Peminjaman = 2.0 and Jenis_Kelamin = 1.0 and Pekerjaan = 2.0 then Y (1 / 0)
 - 10) if Radius_Survey = 1.0 and Jumlah_Tanggungan = 2.0 and Status_Domisili = 1.0 and Besar_Penghasilan = 2.0 and Jumlah_Peminjaman = 3.0 then Y (6 / 0)
 - 11) if Radius_Survey = 1.0 and Jumlah_Tanggungan = 2.0 and Status_Domisili = 1.0 and Besar_Penghasilan = 2.0 and Jumlah_Peminjaman = 4.0 then Y (1 / 0)
 - 12) if Radius_Survey = 1.0 and Jumlah_Tanggungan = 2.0 and Status_Domisili = 1.0 and Besar_Penghasilan = 3.0 then Y (10 / 0)
 - 13) if Radius_Survey = 1.0 and Jumlah_Tanggungan = 2.0 and Status_Domisili = 1.0 and Besar_Penghasilan = 4.0 then Y (6 / 0)
 - 14) if Radius_Survey = 1.0 and Jumlah_Tanggungan = 2.0 and Status_Domisili = 2.0 then Y (14 / 0)
 - 15) if Radius_Survey = 1.0 and Jumlah_Tanggungan = 2.0 and Status_Domisili = 3.0 and Jumlah_Peminjaman = 1.0 then N (0 / 1)
 - 16) if Radius_Survey = 1.0 and Jumlah_Tanggungan = 2.0 and Status_Domisili = 3.0 and Jumlah_Peminjaman = 2.0 then Y (1 / 0)
 - 17) if Radius_Survey = 1.0 and Jumlah_Tanggungan = 2.0 and Status_Domisili = 3.0 and Jumlah_Peminjaman = 3.0 then N (0 / 1) if Radius_Survey = 1.0 and Jumlah_Tanggungan = 3.0 then Y (8 / 0)
 - 18) if Radius_Survey = 1.0 and Jumlah_Tanggungan = 4.0 then N (0 / 3)
 - 19) if Radius_Survey = 2.0 and Status_Domisili = 1.0 and Besar_Penghasilan = 1.0 and Jumlah_Peminjaman = 1.0 and Tenor = 3.0 then Y (1 / 0)
 - 20) if Radius_Survey = 2.0 and Status_Domisili = 1.0 and Besar_Penghasilan = 1.0 and

Jumlah_Peminjaman = 1.0 and Tenor = 4.0 then N (0 / 1)

- 21) if Radius_Survey = 2.0 and Status_Domisili = 1.0 and Besar_Penghasilan = 1.0 and Jumlah_Peminjaman = 2.0 then N (0 / 3) if Radius_Survey = 2.0 and Status_Domisili = 1.0 and Besar_Penghasilan = 2.0 and Jumlah_Tanggungan = 1.0 then Y (3 / 0)

Berikut gambar pengujian modelnya menggunakan RapidMiner :



Sumber : Khasanah (2017) (2017)
Gambar 2. Pengujian Algoritma C4.5 dengan RapidMiner

Hasil Pengujian Model Algoritma C4.5 Classifier ditunjukkan pada gambar 3 di bawah ini:

accuracy: 88.52% +/- 7.11% (mikro: 88.52%)			
	true Y	true N	class precision
pred. Y	183	17	91.50%
pred. N	14	56	80.00%
class recall	92.89%	76.71%	

Sumber : Khasanah (2017) (2017)
Gambar 3. Hasil Pengujian Algoritma C4.5 dengan RapidMiner

Jumlah True Positive (TP) adalah 183 record diklasifikasikan sebagai kredit LANCAR dan False Negative (FN) sebanyak 17 record diklasifikasikan sebagai kredit LANCAR tetapi

kredit BERMASALAH. Berikutnya 56 record untuk True Negative (TN) diklasifikasikan sebagai kredit BERMASALAH, dan 14 record False Positive (FP) diklasifikasikan sebagai kredit BERMASALAH tetapi kredit LANCAR. Berdasarkan gambar 3 tersebut menunjukkan bahwa, tingkat akurasi dengan menggunakan algoritma C4.5 Classifier adalah sebesar **88,52%** dan dapat dihitung untuk mencari nilai accuracy, sensitivity, specificity, ppv, dan npv di bawah ini:

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{183}{183 + 17} = 0.915$$

$$\text{Specificity} = \frac{TN}{TN + FP} = \frac{56}{56 + 14} = 0.800$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{183 + 56}{183 + 56 + 14 + 17} = 0.8852$$

$$\text{PPV} = \frac{TP}{TP + FP} = \frac{183}{183 + 14} = 0.9289$$

$$\text{Precision} = \frac{TN}{TN + FN} = \frac{56}{56 + 17} = 0.7671$$

Hasil perhitungan divisualisasikan dengan kurva ROC. Perbandingan kedua class bisa dilihat pada Gambar 4.3 yang merupakan kurva ROC untuk algoritma C4.5. Kurva ROC pada gambar 4 mengekspresikan confusion matrix dari gambar 3. Garis horizontal adalah false positives dan garis vertikal true positives. Menghasilkan nilai AUC (Area Under Curve) sebesar 0.799 dengan nilai akurasi klasifikasi cukup (fair classification).



Sumber : Khasanah (2017) (2017)
Gambar 4. Nilai AUC dalam grafik ROC Algoritma C4.5 Classifier

Khasanah (2017) ini memperlihatkan bahwa nilai akurasi klasifikasi dengan menggunakan algoritma C4.5 sebesar 88.52%. Hasil ini berbeda dengan penelitian sebelumnya yang dilakukan oleh Rahmayuni pada Tahun 2014, dimana pada penelitian tersebut akurasi yang dihasilkan adalah 85.61%.

KESIMPULAN

Berdasarkan hasil penerapan algoritma C4.5 classifier maka dapat disimpulkan bahwa untuk penentuan kelayakan kredit apakah calon nasabah akan menjadi nasabah yang pembayaran lancar atau bermasalah dengan menggunakan algoritma C4.5 classifier. Khasanah (2017) untuk nilai akurasi algoritma C4.5 classifier dengan menggunakan data training 270 nasabah adalah 88.52%.

Adapun saran-saran yang dapat disampaikan yaitu Khasanah (2017) dapat diimplementasikan ke dalam program aplikasi yang bisa bermanfaat untuk seorang analis kredit dalam menganalisa kelayakan kredit. Saran lain yang dapat disampaikan yaitu pengembangan penelitian selanjutnya dapat menggunakan algoritma klasifikasi lainnya seperti Naive Bayes.

REFERENSI

- Iriadi, N., & Nuraeni, N. (2016). Kajian Penerapan Metode Klasifikasi Data Mining Algoritma C4. 5 Untuk Prediksi Kelayakan Kredit Pada Bank Mayapada Jakarta. *Jurnal Teknik Komputer AMIK BSI*, 2(1), 132-137.
- Khasanah, S., N. (2017). Laporan Akhir Penelitian Mandiri. Jakarta: STMIK Nusa Mandiri Jakarta.
- Rahmayuni, I. (2014). Perbandingan Performansi Algoritma C4.5 dan Cart Dalam Klasifikasi Data Nilai Mahasiswa Prodi Teknik Komputer Politeknik Negeri Padang. *Jurnal TEKNOIF*, Vol. 2 No. 1 April 2014 : 40-46.
- Rani, L. N. (2015). Klasifikasi Nasabah Menggunakan Algoritma C4.5 Sebagai Dasar Pemberian Kredit. *Jurnal KomTekInfo Fakultas Ilmu Komputer*, Volume 2, No 2, Des 2015 : 33-38 .
- Santoso, T. B. (2014). Analisa dan Penerapan Metode C4.5 Untuk Prediksi Loyalitas Pelanggan. *Jurnal Ilmiah Fakultas Teknik LIMIT'S*, Vol. 10 No. 1, 1-10.
- Sucipto, A. (2015). Prediksi Kredit Macet Melalui Perilaku Nasabah Pada Koperasi Simpan Pinjam Dengan Menggunakan Metode Algoritma Klasifikasi C4. *Jurnal DISPROTEK*, Volume 6 No 1 : 75-87.
- Swastina, L. (2013). Penerapan Algoritma C4. 5 Untuk Penentuan Jurusan Mahasiswa. *Jurnal GEMA AKTUALITA*, 2(1), 93-98.