PENERAPAN ALGORITMA KLASIFIKASI *RANDOM FOREST* UNTUK PENENTUAN KELAYAKAN PEMBERIAN KREDIT DI KOPERASI MITRA SEJAHTERA

Nugraha Listiana Hanun¹, Achmad Udin Zailani^{2*}

^{1,2*} Program Studi Teknik Informatika Universitas Pamulang Tangerang, Indonesia
E-mail: dosen00270@unpam.ac.id

Submission date: 2020-04-10

Accepted date: 2020-05-29

ABSTRACT

Credit is the provision of money or bills which can be equalized with an agreement or deal between the bank and another parties that requires the borrower to pay off the debt after a certain period of time through interest. Before the cooperative approves the credit proposed by the debtor, the cooperative conducts a credit analysis of borrowers whether the credit application is approved or disapproved. This study objectives to predict creditworthiness by applying the Random Forest Classification Algorithm in order to provide a solution for determining the creditworthiness. This research method is absolute experimental research that leads to the impact resulting from experiments on the application of the decision tree model of the Random Forest Classification Algorithm's approach. The study results using the Random Forest Classification Algorithm's are able to analyze problem credit and disproblems debtors with an accuracy value of 87.88%. Besides that, decision tree model was able to improve the accuracy in analyzing the credit worthiness of borrowers who filed.

Keywords: *Credit analysis, decision tree model, random forest classification algorithm's*

ABSTRAK

Kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan atas persetujuan atau kesepakatan pinjam meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam melunasi utangnya setelah jangka waktu tertentu dengan pemberian bunga. Koperasi Mitra Sejahtera menghadapi masalah pembayaran pihak peminjam atas tunggakan kredit. Penelitian ini bertujuan untuk memprediksi kelayakan kredit dengan penerapan Algoritma Klasifikasi *Random Forest* agar dapat memberikan solusi untuk penentuan kelayakan pemberian kredit. Metode penelitian ini adalah riset eksperimen absolut yang mengarah kepada dampak yang dihasilkan dari eksperimen atas penerapan model pohon keputusan menggunakan pendekatan Algoritma Klasifikasi *Random Forest*. Hasil pengujian dengan algoritma klasifikasi *Random Forest* mampu menganalisis kredit yang bermasalah dan yang debitur yang tidak bermasalah dengan nilai akurasi sebesar 87,88%. Di samping itu, model pohon keputusan ternyata mampu meningkatkan akurasi dalam menganalisis kelayakan kredit yang diajukan calon debitur.

Kata Kunci: Analisis kredit, model pohon keputusan, algoritma klasifikasi *random forest*

PENDAHULUAN

Kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan atas persetujuan atau kesepakatan pinjam meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam melunasi utangnya setelah jangka waktu tertentu dengan pemberian bunga. Kredit Tanpa Agunan (KTA) adalah kredit yang diberikan bank atau lembaga keuangan dalam bentuk uang tunai, yang dapat diperoleh tanpa memberikan jaminan. Kemudahan ini banyak dimanfaatkan oleh para nasabah untuk meminjam sejumlah dana dari

koperasi untuk digunakan berbagai keperluan nasabah (Peraturan Bank Indonesia, 2002).

Kredit yang diajukan oleh debitur memiliki resiko, karena dari sekian banyak debitur yang mengajukan ke aplikasi ada kemungkinan beberapa debitur yang bermasalah dalam pembayarannya sehingga menyebabkan kredit macet. Sebelum koperasi menyetujui kredit yang diajukan oleh debitur, koperasi melakukan analisis kredit terhadap debitur apakah pengajuan kredit disetujui atau tidak disetujui. Analisis kredit adalah penyelidikan faktor-faktor yang berpengaruh pada lancarnya atau kurang lancarnya pengembalian kredit (Iriadi & Leidiyana, 2013).

Sebagai tolak ukur bahwa debitur disetujui atau ditolak, dapat digunakan data histori debitur yang telah disetujui oleh koperasi. Namun, perlu diperhatikan juga bahwa debitur yang telah disetujui juga tidak semuanya pembayar kredit yang baik, artinya ada beberapa debitur yang telah disetujui tapi beberapa bulan kemudian pembayarannya menunggak (Abdussomad, 2017).

Kredit Koperasi Mitra Sejahtera 2016 yang menggunakan data kredit pada tahun 2015 hingga 2016, dimana dari 239 debitur, 29 diantaranya bermasalah dalam pembayaran atau dengan istilah lain macet dalam pembayaran kreditnya, hal ini disebabkan karena sejumlah anggota koperasi yang menunggak pembayaran kredit. Data ini menjadi permasalahan bagi koperasi, karena jika debitur tidak mampu membayar cicilan kredit dalam jangka waktu yang telah ditentukan, maka ini akan menyebabkan kredit macet.

Penilaian kredit dapat diaplikasikan dalam berbagai bidang seperti dalam industri asuransi menggunakan penilaian kredit untuk menyederhanakan aplikasi dan proses perpanjangan, menentukan premi dan prediksi akuntabilitas konsumen yang lebih akurat (Astuty, 2015).

Lembaga keuangan menggunakan penilaian kredit untuk memutuskan kelayakan pemberian kredit. Tuan tanah dapat menggunakan penilaian kredit untuk memprediksi apakah penyewa akan membayar tepat waktu atau tidak. *Supplier* juga menggunakan penilaian kredit untuk menetapkan konsumen potensial (Sunjana, 2010).

Analisis kredit dilakukan oleh *account officer* yang dari sisi level jabatannya merupakan level seksi atau bagian, atau bahkan dapat berupa tim (*committee*) yang ditugaskan untuk melakukan analisis permohonan kredit. Prinsip dasar dalam menganalisis kredit yang lazim dikenal dengan Prinsip 6 C's, yaitu *character, capital, capacity, collateral, condition of ekonomis* dan *constrain* (Susanto, Uryani, & Hidayat, 2013).

Data mining adalah aplikasi algoritma spesifik untuk mengekstrak poladari data (Kusrini & Luthfi, 2014). *Data Mining* didefinisikan sebagai proses penemuan pola dalam data (Witten, Frank, & Hall, 2011). *Data mining* sering juga disebut analisis data eksploratif. Data dalam jumlah besar yang diperoleh dari mesin kasir, pemindaian *barcode* dan dari berbagai basis data dalam perusahaan, kemudian ditelaah, dianalisis, dihapus dan dipakai ulang.

Klasifikasi adalah proses menempatkan obyek atau konsep tertentu kedalam satu set kategori, berdasarkan sifat obyek atau konsep yang bersangkutan (Larose, 2013). Metode klasifikasi

ditujukan untuk pembelajaran fungsi-fungsi berbeda yang memetakan masing-masing data terpilih kedalam salah satu dari kelompok kelas yang telah ditetapkan sebelumnya.

Algoritma *Random Forest* didesain oleh J. Ross Quinlan, dinamakan *Random Forest* karena merupakan keturunan dari pendekatan ID3 untuk membangun pohon keputusan. *Random Forest* merupakan algoritma yang cocok digunakan untuk masalah klasifikasi pada *machine learning* dan *data mining* (Larose, 2013).

Random Forest memetakan atribut dari kelas sehingga dapat digunakan untuk menemukan prediksi terhadap data yang belum muncul. Pohon keputusan sendiri merupakan pendekatan "*divide and conquer*" dalam mempelajari masalah dari sekumpulan data independen yang digambarkan dalam bagan pohon (Witten, Frank, & Hall, 2011). Pohon keputusan juga merupakan sekumpulan pertanyaan yang tersusun secara sistematis, dimana setiap pertanyaan yang ada menentukan percabangan berdasarkan nilai atribut dan berhenti pada daun dari pohon yang merupakan prediksi dari kelas variabel (Yahya & Zuliana, 2018).

Berikut tahapan algoritma dalam pembuatan pohon keputusan dengan Algoritma *Random Forest*:

1. Perhatikan label pada data, jika sudah sama semua, maka akan dibentuk daun dengan nilai label data keseluruhan
2. Menghitung nilai informasi dengan menggunakan semua data yang ada dengan formula:

$$Info(D) = - \sum_{i=1}^m p_i \log_2(p_i) \dots \dots \dots (1)$$

Dimana merupakan probabilitas *tuple* dalam D yang menjadi kelas dengan asumsi atau disebut juga entropy dari D merupakan rata rata informasi yang diperlukan untuk identifikasi *tuple* dalam D (Kusrini & Luthfi, 2014).

Jika Nilai A adalah nilai diskrit maka data D akan dipisahkan sejumlah nilai data A sehingga nilai setiap cabang akan murni dan sejenis. Setelah percabangan pertama, jumlah percabangan yang mungkin terjadi diukur dengan persamaan:

$$Info_A(D) = \sum_j \frac{|D_j|}{|D|} \times Info_A(D_j) \dots \dots \dots (2)$$

3. Menghitung nilai informasi dengan formula

4. Untuk setiap atribut dengan memperhatikan isi data dari atribut. dimana $\frac{|D_j|}{|D|}$ merupakan bobot dari partisi j. $Info_A(D)$ merupakan informasi yang diperlukan untuk mengklasifikasikan *tuple* dari D pada partisi A. Semakin kecil hasil persamaan ini, semakin baik pula partisi yang dihasilkan. Nilai dari sebuah atribut menentukan penting tidaknya atribut tersebut dalam penyusunan pohon keputusan. Jika atribut bernilai kontinyu, maka akan dicari *split_point* dengan cara mengurutkan seluruh data menurut atribut tersebut dari kecil ke besar, lalu di rata-rata antar satu data dengan data setelahnya. Nilai informasi akan dihitung menurut satu persatu calon *split_point* dan nilai *split_point* yang akan dipilih yang terkecil. (4) Nilai *gain* untuk setiap atribut akan diperhitungkan dengan formula (2.3), nilai dengan *gain* tertinggi akan dijadikan cabang dalam pohon keputusan.

$$Gain(A) = Info(D) - Info_A(D) \dots \dots \dots (3)$$

5. Setelah cabang pohon keputusan terbentuk, perhitungan dilakukan kembali seperti pada tahap 1 sampai 4. Namun jika cabang telah mencapai maksimal cabang yang diperbolehkan, daun akan terbentuk dengan nilai mayoritas dari nilai data.

Pengujian *K-Fold Cross Validation* adalah teknik validasi dengan membagi data secara acak kedalam k bagian dan masing-masing bagian akan dilakukan proses klasifikasi (Witten, Frank, & Hall, 2011).

Dengan menggunakan *cross validation* akan dilakukan percobaan sebanyak k. Data yang digunakan dalam percobaan ini adalah data *training* untuk mencari nilai *error rate* secara keseluruhan. Secara umum, pengujian nilai k dilakukan sebanyak 10 kali untuk memperkirakan akurasi estimasi. Dalam penelitian ini nilai k yang digunakan berjumlah 10 atau *10-fold Cross Validation*, tiap percobaan akan menggunakan satu data *testing* dan k-1 bagian akan menjadi data *training*, kemudian data *testing* itu akan ditukar dengan satu buah data *training* sehingga untuk tiap percobaan akan didapatkan data *testing* yang berbeda-beda (Wibowo, 2017).

Confusion matrix memberikan keputusan yang diperoleh dalam *training* dan *testing*, *confusion matrix* memberikan penilaian *performance* klasifikasi berdasarkan objek dengan benar atau salah (Witten, Frank, & Hall, 2011). *Confusion matrix* berisi informasi aktual (*actual*) dan prediksi (*predicted*) pada sistem klasifikasi.

Kurva ROC (*Receiver Operating Characteristic*) adalah alat visual yang berguna untuk membandingkan dua model klasifikasi. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horisontal dan *true positives* sebagai garis vertikal (Witten, Frank, & Hall, 2011). Dengan kurva ROC, kita dapat melihat *trade off* antara tingkat dimana suatu model dapat mengenali *tuple* positif secara akurat dan tingkat dimana model tersebut salah mengenali *tuple* negatif sebagai *tuple* positif. Sebuah grafik ROC adalah plot dua dimensi dengan proporsi positif salah (fp) pada sumbu X dan proporsi positif benar (tp) pada sumbu Y. Titik (0,1) merupakan klasifikasi yang sempurna terhadap semua kasus positif dan kasus negatif. Nilai positif salah adalah tidak ada (fp = 0) dan nilai positif benar adalah tinggi (tp = 1). Titik (0,0) adalah klasifikasi yang memprediksi setiap kasus menjadi negatif {-1}, dan titik (1,1) adalah klasifikasi yang memprediksi setiap kasus menjadi positif {1}. Grafik ROC menggambarkan *trade-off* antara manfaat (*true positive*) dan biaya (*false positives*). Berikut tampilan dua jenis kurva ROC (*discrete* dan *continous*).

Sumber: (Witten, Frank, & Hall, 2011)

Gambar 1. Grafik ROC (*discrete* dan *continous*)

Poin diatas garis diagonal merupakan hasil klasifikasi yang baik, sedangkan point dibawah garis diagonal merupakan hasil klasifikasi yang buruk. Dapat disimpulkan bahwa, satu point pada kurva ROC adalah lebih baik dari pada yang lainnya jika arah garis melintang dari kiri bawah ke kanan atas didalam grafik. Untuk tingkat akurasi nilai AUC dalam klasifikasi *data mining* dibagi menjadi lima kelompok (Witten, Frank, & Hall, 2011), yaitu:

1. 0.90 - 1.00 = klasifikasi sangat baik (*excellent classification*)
2. 0.80 - 0.90 = klasifikasi baik (*good classification*)
3. 0.70 - 0.80 = klasifikasi cukup (*fair classification*)
4. 0.60 - 0.70 = klasifikasi buruk (*poor classification*)

5. 0.50 - 0.60 = klasifikasi salah (*failure*)

Beberapa penelitian yang menerapkan teknik analisis kredit yang berjudul klasifikasi data nasabah sebuah asuransi menggunakan Algoritma *Random Forest* untuk penentuan kelayakan pemberian asuransi dan menentukan fitur yang berpengaruh. Atribut-atribut yang ini meliputi penghasilan, premi_dasar, cara_pembayaran, mata_uang, dan status sedangkan label yang digunakan untuk pengklasifikasian adalah lancar dan tidak lancar sehingga diketahui data nasabah mana yang dikelompokkan ke kelas lancar dan tidak lancar dengan melakukan klasifikasi kelayakan pemberian kredit nasabah dengan jumlah 8 atribut menghasilkan akurasi 87,36% merupakan tingkat akurasi yang baik sehingga kelancaran proses penilaian kelayakan kredit dapat tercipta dan pembayaran terlambat (menunggak) dapat diprediksi dari awal untuk dapat diwaspadai agar dapat meminimalisir meningkatnya kredit macet (Sunjana, 2010).

Prediksi pinjaman kredit antara *Support Vector Machine* dan *K-Nearest Neighbors* pada Koperasi Serba Usaha menghasilkan nilai akurasi *Support Vector Machine* sebesar 92,67% diikuti oleh *k-Nearest Neighbors* sebesar 88,67%, dengan penilaian *cross validation*, *confusion matrix* dan kurva *ROC* (Iriadi & Leidiyana, 2013).

Analisis kredit macet melalui perilaku nasabah pada koperasi simpan pinjam dengan menggunakan metode Algoritma Klasifikasi *Random Forest* untuk mendeteksi kredit macet telah menghasilkan nilai akurasi Algoritma *Random Forest* sebesar 92,00%, dimana cukup baik keakuratannya digunakan untuk memprediksi kredit macet pada koperasi simpan pinjam (Sucipto, 2015). Penerapan Algoritma Klasifikasi C4.5 untuk menghasilkan pola kelayakan kredit leasing di Kota Karawang dengan pengujian model dengan hasil uji akurasi sebesar 90,99% dan nilai AUC sebesar 0,911 dengan tingkat diagnosa *Excellent Classification* (Abdussomad, 2017).

Berdasarkan fenomena, tinjauan pustaka dan beberapa penelitian yang direviu, maka penelitian ini menggunakan Model Pohon Keputusan Algoritma *Random Forest*. Model ini akan digunakan untuk memprediksi apakah debitur nantinya akan bermasalah dalam pembayaran kredit atau tidak. Model ini dipilih karena proses Learning dan Klasifikasi pada Algoritma *Random Forest* sangat sederhana dan cepat dan secara umum memiliki tingkat akurasi yang tinggi (Susanto, Uryani, & Hidayat, 2013).

Di samping itu, model pohon keputusan juga dapat dikonversi ke dalam aturan klasifikasi dengan mudah. Identifikasi masalah penelitian ini

adalah untuk memprediksi penetapan kelayakan kredit dengan subjek penelitian pada Koperasi Mitra Sejahtera yang memiliki data anggota sebanyak 259 sehingga penting untuk dikaji lebih lanjut. Penggunaan algoritma klasifikasi *Random Forest* diharapkan dapat memberikan solusi untuk permasalahan penentuan kelayakan pemberian kredit yang menghasilkan nilai akurasi di atas 90 persen pada beberapa penelitian terdahulu. Penelitian ini bertujuan untuk menerapkan Pohon Keputusan berbasis Algoritma *Random Forest* yang diharapkan agar dapat meningkatkan keakuratan analisis kredit pada subjek Koperasi Mitra Sejahtera dalam menentukan kelayakan pemberian kredit.

METODE PENELITIAN

Metode penelitian yang digunakan pada eksperimen ini adalah eksperimen absolut mengarah kepada dampak yang dihasilkan dari eksperimen (Sugiyono, 2016).

Penelitian ini menggunakan Algoritma C.45 untuk mengembangkan aplikasi berdasarkan model yang dibuat menggunakan *RapidMiner Studio 7*, kemudian dilakukan pengujian dengan *Cross validation* terhadap kinerja dari kedua metode tersebut, hasil ditunjukkan oleh *Confusion matrix* dan Kurva *ROC*.

Sumber : (Zailani, 2020)

Gambar 2. Model Penelitian

Beberapa tahap yang dilakukan dalam melakukan eksperimen ini menggunakan model *Cross-Standard Industry for Data Mining* (CRISP-DM) (Larose, 2013) yang terdiri dari 6 (enam) tahap, yaitu:

1. Tahap *business understanding*. Tahap ini merupakan tahap pemahaman penelitian, menentukan tujuan proyek penelitian dalam perumusan mendefinisikan masalah *data mining*.
2. Tahap *data understanding*. Tahap ini dilakukan pengumpulan data, melakukan analisis penyelidikan data (data kredit) untuk

mengenali lebih lanjut data dan pencarian pengetahuan awal kemudian mengevaluasi kualitas dari data tersebut.

3. Tahap *data preparation*. Kelanjutan dari data *understanding*, untuk mendapatkan data yang berkualitas. Beberapa teknik *preprocessing* yang digunakan yaitu:
 - a. *Data cleaning* bekerja untuk membersihkan nilai yang kosong, tidak konsisten atau mungkin tupel yang kosong (*missing values* dan *noisy*).
 - b. *Data integration* berfungsi menyatukan tempat penyimpanan (arsip) yang berbeda ke dalam satu data. Dalam hal ini, ada dua arsip yang diambil sebagai *data warehouse* yaitu data anggota dan data kredit.
 - c. *Data reduction*. Jumlah atribut dan tupel yang digunakan untuk data training mungkin terlalu besar, hanya beberapa atribut yang diperlukan sehingga atribut yang tidak diperlukan akan dihapus. Tupel dalam data set mungkin terjadi duplikasi atau terdapat tupel yang sama, sehingga untuk memperkecil jumlah tupel, tupel yang sama dijadikan akan dijadikan dalam satu tupel untuk mewakili tupel tersebut.
4. Tahap *modelling* (modeling). Tahap ini data training diklasifikasikan oleh model dan kemudian menghasilkan sejumlah aturan.
5. Tahap *evaluation* (evaluasi). Tahap ini mengklasifikasi uji akurasi dengan melihat hasil akurasi pada proses klasifikasi Algoritma C4.5 serta evaluasi dengan ROC Curve.
6. Tahap *deployment*. Tahap ini menerapkan model Algoritma *Random Forest* ke koperasi untuk penentuan kelayakan pemberian kredit dengan metode penyeleksian atribut melalui *information gain*, *gain ratio* dan *gini index* untuk mendapatkan nilai akurasi hasil *confusion matrix* menggunakan software *Rapid Miner*.

HASIL DAN PEMBAHASAN

Tahap Business Understanding

Berdasarkan laporan kredit tahun 2015 ditemukan kredit macet sebanyak 29 orang sehingga ini menjadi permasalahan koperasi dan merupakan imbas dari analisis yang kurang akurat. Agar memperoleh hasil yang akurat, maka dikembangkan model klasifikasi Algoritma

Random Forest untuk menentukan kelayakan pemberian kredit dengan tujuan analisis yang dilakukan lebih akurat.

Tahap Data Understanding

Data kredit diambil dari Koperasi Mitra Sejahtera pada tahun 2015 hingga 2016, dimana dari 239 debitur, 29 diantaranya bermasalah dalam pembayaran atau dengan istilah lain macet dalam pembayaran kreditnya. Ada beberapa atribut yang digunakan di dalam data debitur yaitu *no_anggota* (no keanggotaan yang diberikan koperasi), *jumlah tanggungan* (*jumlah_anak*), *level golongan* (tingkatan golongan pada pegawai negeri), *level pinjaman* (besar pinjaman yang diberikan kepada debitur), *jangka waktu* (lama pinjaman dalam tahun), dan *class* (label yang diberikan kepada debitur, label baik untuk yang pembayarannya lancar, dan bermasalah jika kreditnya macet). Nilai dari semua atribut yang ada di tabel merupakan nilai kategorikal dan bukan nilai angka, misalnya seperti atribut *Jumlah Tanggungan* yang mempunyai nilai sedikit merupakan kategori jika debitur mempunyai jumlah anak 1 sampai 2.

Tabel 1. Kategori Atribut

No	Nama Atribut	Nilai Angka	Kategori
1	Jumlah Tanggungan	1-2	Sedikit
2	Jurusan	3-5 > 5 0	Sedikit Banyak Kosong
3	Level Golongan	IA - ID IIA - IID IIIA - IIID IVA - IVE	I II III IV
4	Level Pinjaman	Rp1jt - Rp5jt Rp5jt - Rp10jt Rp10jt - Rp15jt	Kecil Sedang Besar
5	Jangka Waktu	1 tahun 2 tahun 3 tahun	1 2 3

Sumber: (Koperasi Mitra Sejahtera, 2016)

Tahap Preparation

Model yang telah dikembangkan akan diuji keakuratannya dengan memasukkan sejumlah data uji (*test set*) ke dalam model. Untuk mengukur keakuratan model dengan baik, data uji seharusnya bukan data yang berasal dari data training (Witten, Frank, & Hall, 2011). Data uji diambil dari laporan kredit koperasi tahun 2016. Ada 20 sampel yang diambil dari data keseluruhan, jumlah data sebanyak 239 tidak diambil karena jumlah ini sama dengan data tahun 2015 dan tidak ada perubahan, hanya 20 sampel yang bertambah sehingga totalnya menjadi 259.

Tabel 2. Laporan Kredit 20 Sampel Nasabah

No	Jumlah Tanggungan	Level Golongan	Level Pinjaman	Jangka Waktu	Class
1	Banyak	III	Sedang	2	Baik
2	Sedikit	II	Besar	3	Bermasalah
3	Kosong	III	Besar	3	Baik
4	Sedikit	IV	Sedang	2	Baik
5	Sedikit	III	Kecil	2	Baik
6	Sedang	III	Kecil	2	Baik
7	Sedang	IV	Besar	3	Baik
8	Sedang	III	Sedang	3	Baik
9	Sedang	II	Besar	3	Bermasalah
10	Sedikit	III	Sedang	3	Baik
11	Banyak	IV	Sedang	3	Baik
12	Sedang	III	Sedang	2	Baik
13	Banyak	II	Besar	2	Bermasalah
14	Kosong	II	Sedang	3	Baik
15	Banyak	III	Besar	3	Baik
16	Banyak	III	Kecil	3	Baik
17	Sedang	II	Sedang	2	Baik
18	Sedang	III	Besar	3	Bermasalah
19	Sedikit	IV	Besar	3	Baik
20	Sedikit	III	Sedang	3	baik

Sumber: (Koperasi Mitra Sejahtera, 2016)

Tahap Modelling

Model yang telah dikembangkan akan diuji keakuratannya dengan memasukkan sejumlah data uji (*test set*) ke dalam model. Data uji diambil dari laporan kredit Koperasi Mitra Sejahtera tahun 2016 sebanyak 20 sampel (Tabel 2).

Tahap Evaluation

Dari hasil pengujian di atas akan dievaluasi tingkat akurasi menggunakan 2 (dua) model yaitu menggunakan *confusion matrix* dan ROC (*Receiver Operating Characteristic*).

a. Evaluasi model *confusion matrix*

		Predicted class	
		C ₁	C ₂
Actual class	C ₁	true positives	false negatives
	C ₂	false positives	true negatives

Sumber: (Witten, Frank, & Hall, 2011)

Gambar 3. Nilai *true positives*, *false negatives*, *false positives*, dan *true negatives*

Tabel 3. Konversi ke *confusion matrix*

	Baik	Bermasalah
Baik	210	0
Bermasalah	29	0

Sumber : (Zailani, 2020)

Hitung nilai-nilai yang telah dimasukkan tersebut untuk dihitung jumlah *sensitivity*, *specificity*, *precision* dan *accuracy*. *Sensitivity* digunakan untuk membandingkan jumlah *true positives* terhadap jumlah tupel yang *positives* sedangkan *specificity* adalah perbandingan jumlah *true negatives* terhadap jumlah tupel yang *negatives* ke dalam format perhitungan ini:

Tabel 4. Klasifikasi *confusion matrix*

	pos	Neg
$accuracy = sensitivity (pos+neg) + specificity (pos+neg)$		
dimana :		
t_{pos}	= jumlah <i>true positives</i>	
t_{neg}	= jumlah <i>true negatives</i>	
pos	= jumlah tupel positif	
neg	= jumlah tupel negatif	
f_{pos}	= jumlah <i>false positives</i>	

Sumber : (Zailani, 2020)

Kemudian masukkan nilai yang ada di dalam *confusion matrix* ke dalam persamaan di atas.

Tabel 5. Nilai *sensitivity*, *specificity*, *precision*, dan *accuracy*

	Nilai (%)
<i>Sensitivity</i>	100
<i>Specificity</i>	50
<i>Precision</i>	0
<i>Accuracy</i>	87,88
<i>Recall</i>	0

Sumber : (Zailani, 2020)

Nilai *accuracy*, *recall* dan *precision* yang dihasilkan oleh software *Rapid Miner* menggunakan model *confusion matrix* menghasilkan nilai akurasi sebesar 87,88%.

	true Baik	true Bermasalah	class precision
pred. Baik	210	29	87.87%
pred. Bermasalah	0	0	0.00%
class recall	100.00%	0.00%	

Sumber: (Witten, Frank, & Hall, 2011)

Gambar 4. *Confusion matrix* menggunakan *Rapid Miner*

b. Kurva ROC

Grafik ROC dengan nilai AUC (*Area Under Curve*) dengan metode *information gain* sebesar 0,500, metode *gain ratio* juga menunjukkan hasil yang sama. sedangkan nilai AUC yang menggunakan metode *gini index* mencapai angka 0,703. Akurasi AUC dikatakan sempurna apabila nilai AUC mencapai 1,000 dan akurasi buruk jika nilai AUC dibawah 0,500

Sumber : (Zailani, 2020)

Gambar 5. Nilai AUC dalam grafik ROC dengan metode *information gain* dan *gain ratio* dengan memasukkan data uji

Sumber : (Zailani, 2020)

Gambar 6. Nilai AUC dalam grafik ROC dengan metode *gini index* dan *gain ratio* dengan memasukkan data uji.

Tahap Deployment

Penelitian ini membandingkan 3 (tiga) metode penyeleksian atribut yaitu *information gain*, *gain ratio* dan *gini index*. Hasil dari ketiga metode menghasilkan grafik yang berbeda, namun demikian nilai akurasi hasil *confusion matrix* tetap sama menggunakan *Rapid Miner*.

Sumber : (Zailani, 2020)

Gambar 7. Grafik ROC dengan perbandingan tiga metode yaitu *information gain*, *gain ratio* dan *gini index* tanpa memasukkan data uji.

Dari hasil evaluasi baik secara *confusion matrix* maupun empiris ternyata terbukti akurat

dibanding hasil analisis. Dengan hasil ini, maka kemungkinan terjadinya kredit macet dapat dikurangi dan menekan jumlah debitur yang bermasalah. Dengan demikian, *Algoritma Random Forest* dapat memberikan solusi untuk permasalahan penentuan kelayakan pemberian kredit.

KESIMPULAN

Algoritma Random Forest lebih akurat dibandingkan analisis yang dilakukan oleh analisis kredit dengan hasil evaluasi penelitian bahwa *Algoritma Random Forest* mampu menganalisis kredit yang bermasalah dan yang debitur yang tidak bermasalah yang menghasilkan nilai akurasi sebesar 87,88%. Beberapa hal yang harus ditambahkan untuk menambah akurasi *Algoritma Random Forest*, yaitu: (1) melakukan *pruning* terhadap *Algoritma Random Forest* sehingga hal ini dilakukan untuk mengoptimalkan kinerja tanpa mengurangi keakuratannya; (2) tingkat akurasi dari *Algoritma Random Forest* dapat dibandingkan pada model algoritma lainnya seperti *Random Forest* atau *Support Vector Machine*; dan (3) menerapkan *Algoritma Random Forest* ke dalam data yang lebih besar untuk menguji akurasi dari algoritma.

DAFTAR PUSTAKA

- Abdussomad. (2017). Penerapan Algoritma Klasifikasi C4.5 untuk Menghasilkan Pola Kelayakan Kredit. *Jurnal Inovasi Informatika*, II (1), 19-27.
- Astuty, H. S. (2015). Prinsip 6C (Character, Capacity, Capital, Condition Of Economy, Collateral Dan Constraint) Dalam Wirausaha Mahasiswa. *Jurnal Economia*, II (1), 56-71.
- Iriadi, N., & Leidiyana, H. (2013). Prediksi Pinjaman Kredit dengan Support Vector Machine K-Nearest Neighbors pada Koperasi Serba Usaha. *Jurnal Penelitian Ilmu Komputer, System Embedded & Logic*, I (2), 115-124.
- Koperasi Mitra Sejahtera. (2016). Laporan Koperasi Tahun 2016.
- Kusrini, & Luthfi. (2014). *Algoritma Data Mining*. Yogyakarta: Andi.
- Larose. (2013). *Discovering Knowledge in Data*. New Jersey: John Wiley.

- Peraturan Bank Indonesia. (2002). *Peraturan Bank Indonesia Nomor 4/ 11/PBI/2002 Tentang Perlakuan Khusus Terhadap Kredit Bank Umum Pasca Tregedi Bank Bali*. Jakarta: Jakarta.
- Sucipto, A. (2015). Prediksi Kredit Macet melalui Perilaku Nasabah pada Koperasi Simpan Pinjam dengan Menggunakan Metode Algoritma Klasifikasi C4.5. *Disprotek*, 6 (1), 75-87.
- Sugiyono. (2016). *Metode Penelitian Kombinasi (Mixed Methods)*. (M. Sutopo, Penyunt.) Bandung: CV. Alfabeta.
- Sunjana. (2010). Klasifikasi Data Nasabah Sebuah Asuransi Menggunakan Algoritma C4.5. *Seminar Nasional Aplikasi Teknologi Informasi 2010 (SNATI 2010)* (hal. D31.D34). Yogyakarta: Universitas Islam Indonesia.
- Susanto, Uryani, N. S., & Hidayat, C. R. (2013). Penerapan Algoritma C4.5 untuk Menganalisis Kelayakan Pemberian Kredit Nasabah. 4 (1), 1-7.
- Wibowo, A. (2017, November 24). *10 Fold - Cross Validation*. Retrieved Maret 2018, from <https://mti.binus.ac.id/2017/11/24/10-fold-cross-validation/>
- Witten, Frank, & Hall. (2011). *Data Mining : Practical Machine Learning and Tools*. Burlington: Morgan Kaufmann Publisher.
- Yahya, & Zuliana, R. (2018). Prediksi Jumlah Penggunaan BBM Perbulan Menggunakan Algoritma Decition Tree (C4.5) Pada Kantor Dinas Lingkungan Hidup dan Kebersihan Kecamatan Selong Kabupaten Lombok Timur. *Infotek: Jurnal Informatika dan Teknologi*, 1 (1), 56-63.