

# KAJIAN ALGORITMA C4.5, NAIVE BAYES, NEURAL NETWORK DAN SVM DALAM PENENTUAN KELAYAKAN PEMBERIAN KREDIT

Achmad Rifai

**Abstract**—Credit may encourage the growth of the national economy so that it has an important role in financing. However, in the implementation of the financing, the risk of loss largely borne by the lender (creditor). So we need a method to reduce the risk in lending. In this study conducted using data mining classification algorithm C4.5, Naive Bayes, Neural Network and Support Vector Machine and then do a comparison. From the test results to measure the performance of the four methods using 10-fold cross validation, confusion matrix and ROC curve. Based on the test results it is known that the C4.5 algorithm and Support Vector Machine has the same value and the highest accuracy of 85.40% as compared to the other algorithms. The fourth classification algorithm based on it can be concluded that the algorithm C4.5, Naive Bayes and Support Vector Machine is included in the Good Classification Neural Network while included in the Fair Classification.

**Intisari**—Kredit dapat mendorong pertumbuhan ekonomi nasional sehingga memiliki peran penting dalam pembiayaan. Namun dalam pelaksanaan pembiayaan, resiko kehilangan sebagian besar ditanggung oleh pemberi pinjaman (kreditur). Sehingga diperlukan suatu metode dalam mengurangi resiko dalam penyaluran kredit. Dalam penelitian ini dilakukan dengan menggunakan data mining algoritma klasifikasi C4.5, Naive Bayes, Neural Network dan Support Vector Machine kemudian dilakukan perbandingan. Dari hasil pengujian untuk mengukur kinerja dari empat metode menggunakan 10-fold cross validation, confusion matrix dan ROC curve. Berdasarkan hasil pengujian diketahui bahwa algoritma C4.5 dan Support Vector Machine memiliki nilai akurasi yang sama dan tertinggi sebesar 85,40% dibandingkan dengan algoritma yang lain. Berdasarkan klasifikasi keempat algoritma maka dapat disimpulkan bahwa algoritma C4.5, Naive Bayes dan Support Vector Machine termasuk dalam Good Clasification sedangkan Neural Network termasuk dalam Fair Clasification.

**Kata Kunci**—Komparasi, Kredit, Data Mining, algoritma C4.5, Naive Bayes, Neural Network dan Support Vector Machine.

## I. PENDAHULUAN

Kredit adalah penyediaan uang atau tagihan yang dapat disamakan dengan itu, berdasarkan persetujuan pinjam meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam melunasi hutangnya setelah jangka waktu tertentu dengan pemberian bunga [1]. Pembiayaan adalah penyediaan dana oleh pemerintah, dunia usaha, dan masyarakat melalui lembaga keuangan bank, lembaga keuangan bukan bank, atau melalui lembaga lain dalam

rangka memperkuat pemodalannya usaha kecil [2].

Agar kredit yang diberikan mencapai sasaran, yaitu aman dan lancar dalam hal pembayaran angsuran maka analisis kredit sangat perlu dilakukan oleh pihak koperasi. Analisis kredit merupakan hal yang penting dalam lingkup resiko keuangan oleh karena itu perlunya dilakukan analisa kredit.

Penelitian sebelumnya yang berkaitan dengan metode klasifikasi *data mining* tentang analisa resiko kredit telah banyak dilakukan seperti penelitian dengan melakukan komparasi dari metode *Logistic Regression*, *Neural Network*, *Radial Basis Function*, *Support Vector Machine*, *Case-base Reasoning* dan *Decision Trees* untuk menguji tingkat akurasi klasifikasi dengan data historis oleh lembaga keuangan Jerman dengan hasil *Decision Tree* secara signifikan mengungguli 0.05 dibanding model yang lainnya [3].

Penelitian dengan melakukan komparasi dari beberapa metode data diantaranya adalah *Bayes Clasification*, *Decision Tree*, *Boosting*, *Bagging* dan *Random Forest* dengan akurasi tertinggi adalah metode *Bayes Clasification* sebesar 85.20% untuk *dataset* Australia [4].

Penelitian dengan membandingkan dengan metode MDA, LR, SVM dan NN dengan menggunakan data Catest menggunakan lima kelas, dimana akurasi yang terbaik adalah NNs sebesar 96.3% mencapai klasifikasi secara signifikan lebih baik [5].

Penelitian menggunakan model *Support Vector Machine* (SVM), *Logistic Regression* (LR), *Linear Discriminant Analysis* (LDA) dan *k-Nearest Neighbours* (kNN) untuk penentuan kelayakan pemberian kredit dan menentukan fitur yang berpengaruh. Hasil menunjukkan bahwa metode *Support Vector Machine* (SVM) mengungguli ketiga metode lainnya sebagai metode yang baik dalam seleksi fitur yang berpengaruh secara signifikan terhadap dasar keputusan kelayakan pemberian kartu kredit dan juga sangat tepat dalam pengolahan data dengan jumlah besar [28].

Berdasarkan penelitian sebelumnya bahwa metode yang memiliki akurasi terbaik hasilnya berbeda-beda sehingga belum diketahui metode yang paling akurat dalam menentukan kelayakan pemberian kredit. Untuk itu dalam penelitian ini akan dilakukan perbandingan metode Algoritma C4.5, Naive Bayes, Neural Network dan Support Vector Machine sehingga diperoleh metode dengan akurasi prediksi analisis kredit yang terbaik, dimana dari keempat algoritma tersebut merupakan sepuluh klasifikasi data mining paling populer [6]. Data yang digunakan dalam penelitian diperoleh berdasarkan data koperasi pegawai Dirjen Pajak tahun 2010. Data yang diteliti merupakan hasil dari persetujuan pemberian kredit oleh koperasi dan pembayaran pinjaman oleh anggota.

*Program Studi Teknik Informatika, Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri, Jakarta, Jl. Damai No. 8 Warung Jati Barat Jakarta Selatan (email : achmad.acf@email.com)*

II. LANDASAN TEORI

A. Data Mining

Data mining adalah proses menelusuri pengetahuan baru, pola dan tren yang dipilah dari jumlah data yang besar yang disimpan dalam repositori atau tempat penyimpanan dengan menggunakan teknik pengenalan pola serta statistik dan tehnik matematika [7]. Data mining adalah sebuah proses, sehingga dalam melakukan prosesnya harus sesuai prosedur yaitu proses CRISP-DM (Cross-Industry Standard Process for Data Mining) yaitu sebagai keseluruhan proses, preprocessing data, pembentukan model, model evaluasi, dan akhirnya penyebaran model [8].

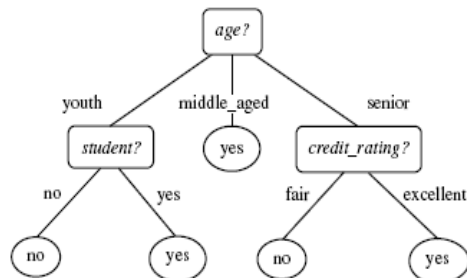
B. Algoritma Klasifikasi Data Mining

Klasifikasi adalah proses menempatkan obyek atau konsep tertentu, kedalam satu set kategori berdasarkan sifat obyek atau konsep yang bersangkutan [9]. Klasifikasi Data mining adalah suatu metode pembelajaran, untuk memprediksi nilai dari sekelompok attribut dalam menggambarkan dan membedakan kelas data atau konsep yang bertujuan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui [10].

C. Algoritma C4.5

C4.5 didesain oleh J. Ross Quinlan, dinamakan C4.5 karena merupakan keturunan dari pendekatan ID3 untuk membangun pohon keputusan. C4.5 merupakan algoritma yang cocok digunakan untuk masalah klasifikasi pada machine learning dan data mining [11]. C4.5 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 [12]. Pohon keputusan dianggap sebagai salah satu pendekatan yang paling populer, dalam klasifikasi pohon keputusan terdiri dari sebuah node yang membentuk akar, node akar tidak memiliki inputan. Node lain yang bukan sebagai akar tetapi memiliki tepat satu inputan disebut node internal atau test node, sedangkan node lainnya dinamakan daun. Daun mewakili nilai target yang paling tepat dari salah satu class [13].



Gbr. 1 Contoh Konsep Pohon Keputusan

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5, yaitu [14]:

1. Menyiapkan data training. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung gain dari atribut, hitung dahulu nilai entropy yaitu:

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

Keterangan:

S : himpunan kasus

A : atribut

n : jumlah partisi S

pi : proporsi dari Si terhadap S

3. Kemudian hitung nilai gain dengan metode informasi gain:

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

Keterangan:

S : himpunan kasus

A : atribut

n : jumlah partisi atribut A

|Si| : jumlah kasus pada partisi ke-i

|S| : jumlah kasus dalam S

4. Ulangi langkah ke-2 hingga semua tupelo terpartisi.
5. Proses partisi pohon keputusan akan berhenti saat:
  - a. Semua tupel dalam node N mendapat kelas yang sama.
  - b. Tidak ada atribut di dalam tupel yang dipartisi lagi.
  - c. Tidak ada tupel di dalam cabang yang kosong.

D. Algoritma Naïve Bayes

Klasifikasi Bayes juga dikenal dengan Naïve Bayes, memiliki kemampuan sebanding dengan dengan pohon keputusan dan Neural Network [15]. Klasifikasi Bayes adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas [14]. Naïve Bayes dapat menggunakan penduga kernel kepadatan, yang meningkatkan kinerja jika asumsi normalitas sangat tidak benar, tetapi juga dapat menangani atribut numeric menggunakan diskritisasi diawasi [12]. Teknik Naïve Bayes (NB) adalah salah satu bentuk sederhana dari Bayesian yang jaringan untuk klasifikasi. Sebuah jaringan Bayes dapat dilihat sebagai diarahkan sebagai tabel dengan distribusi probabilitas gabungan lebih dari satu set diskrit dan variabel stokastik (Pearl 1988) [16].

Klasifikasi Bayes didasarkan pada teorema Bayes, diambil dari nama seorang ahli matematika yang juga menteri Prebysterian Inggris, Thomas Bayes (1702-1761), yaitu [17]:

$$P(x|y) = \frac{P(y|x) P(x)}{P(y)} \quad (3)$$

Keterangan :

P(y)

y = data aengan keias yang belum diketahui

x = hipotesis data y merupakan suatu kelas spesifik

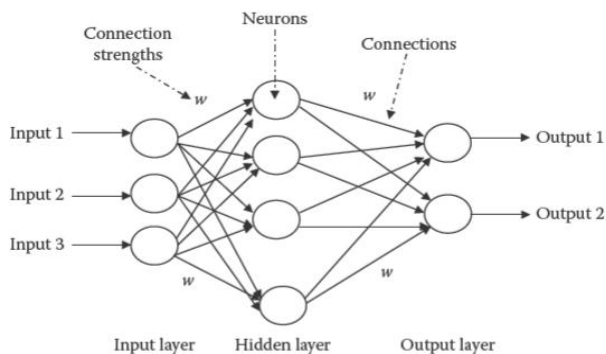
P(x|y) = probabilitas hipotesis x berdasarkan kondisi y (posteriori probability)

- P(x) = probabilitas hipotesis x (*prior probability*)
- P(y|x) = probabilitas y berdasarkan kondisi pada hipotesis x
- P(y) = probabilitas dari y

E. Algoritma Neural Network

Neural network adalah satu set unit input/output yang terhubung dimana tiap relasinya memiliki bobot, Selama fase pembelajaran, neural network menyesuaikan bobot sehingga dapat memprediksi class yang benar dari tuple [15].

Neural Network dapat dimaksudkan untuk mensimulasikan perilaku sistem biologi susunan syaraf manusia, yang terdiri dari sejumlah besar unit pemroses yang disebut neuron, yang beroperasi secara paralel [18]. Sejak tahun 1950-an, neural network telah digunakan untuk tujuan prediksi, bukan hanya klasifikasi tapi juga regresi dengan atribut target kontinu [19]. Banyak model mirip dengan otak manusia telah diusulkan [20].



Gbr. 2 Arsitektur Artificial Neural Network

Langkah pembelajaran algoritma backpropagation adalah sebagai berikut [21]:

1. Inisialisasi bobot jaringan secara acak (biasanya antara -0.1 sampai 1.0)
2. Untuk setiap data pada data training, hitung input untuk simpul berdasarkan nilai input bobot jaringan saat itu, menggunakan rumus:

$$Input_j = \sum_{i=1}^n O_i \cdot w_{ij} + \theta_j \quad (4)$$

Keterangan:

- $O_i$  = output simpul i dari layer sebelumnya
- $w_{ij}$  = bobot relasi dari simpul i pada layer sebelumnya ke simpul j
- $\theta_j$  = bias (sebagai pembatas)

3. Berdasarkan input dari langkah dua, selanjutnya membangkitkan output untuk simpul menggunakan fungsi aktivasi sigmoid :

$$Output = \frac{1}{1 + e^{-Input}} \quad (5)$$

4. Hitung nilai error antara nilai yang diprediksi dengan nilai yang sesungguhnya menggunakan rumus:

$$Error_j = Output_j \cdot (1 - Output_j) \cdot (Target_j - Output_j) \quad (6)$$

Keterangan:

- $Output_j$  = output actual dari simpul j
- $Target_j$  = nilai target yang sudah diketahui pada data training

5. Setelah nilai error dihitung, selanjutnya dibalik ke layer sebelumnya (backpropagated). Untuk menghitung nilai error pada hidden layer menggunakan rumus:

$$Error_j = Output_j(1 - Output_j) \sum_{k=1}^n Error_k w_{jk} \quad (7)$$

Keterangan:

- $Output_j$  = output actual dari simpul j
- $Error_k$  = error simpul k.
- $w_{jk}$  = bobot relasi dari simpul j ke simpul k pada layer berikutnya

6. Nilai error yang dihasilkan dari langkah sebelumnya digunakan untuk memperbaharui bobot relasi menggunakan rumus:

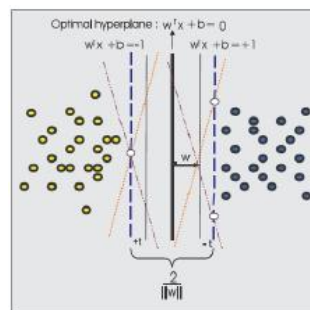
$$w_{ij} = w_{ij} + l \cdot Error_j \cdot Output_i \quad (8)$$

Keterangan:

- $w_{ij}$  = bobot relasi dari unit i pada layer sebelumnya ke unit j
- $l$  = learning rate (konstanta, nilainya antara 0 sampai dengan 1)
- $error_j$  = error pada output layer simpul j
- $output_i$  = output dari simpul

F. Algoritma Support Vector Machine

Support Vector Machine (SVM) adalah metode learning machine yang bekerja atas prinsip Structural Risk Minimization (SRM) dengan tujuan menemukan hyperplane terbaik yang memisahkan dua buah class pada input space (Bellotti & Crook, 2007). Hyperplane terbaik adalah hyperplane yang terletak ditengah-tengah antara dua set obyek dari dua class. Hyperplane pemisah terbaik antara kedua class dapat ditemukan dengan mengukur margin hyperplane tersebut dan mencari titik maksimalnya. Margin adalah jarak antara hyperplane tersebut dengan pattern terdekat dari masing-masing class. Pattern yang paling dekat ini disebut sebagai support vector [27].



Gbr. 3 Konsep SVM untuk mencari hyperplane terbaik

G. Pengujian K-Fold Cross Validation

Seluruh dokumen harus dalam Times New Roman atau Times font. Font tipe 3 tidak boleh digunakan. Jenis font lain

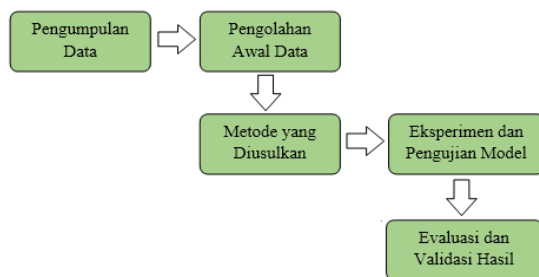
dapat digunakan jika diperlukan untuk tujuan khusus. Fitur ukuran font dapat dilihat pada Tabel 1. *Cross Validation* adalah teknik validasi dengan membagi data secara acak kedalam  $k$  bagian dan masing-masing bagian akan dilakukan proses klasifikasi [15]. Dengan menggunakan *cross validation* akan 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*.

### III. METODE PENELITIAN

Metode penelitian yang umum digunakan ada empat yaitu *Action Reserch Experiment*, *Case Study*, dan *Survey* [22]. Penelitian adalah mencari melalui proses yang metodis untuk menambahkan pengetahuan itu sendiri dan dengan yang lainnya, oleh penemuan fakta dan wawasan tidak biasa. Pengertian lainnya, penelitian adalah sebuah kegiatan yang bertujuan untuk membuat kontribusi orisinal terhadap ilmu pengetahuan [22].

Dalam sebuah penelitian, pendekatan metode yang digunakan untuk memecahkan masalah, diantaranya: mengumpulkan data, merumuskan hipotesis atau proposisi, menguji hipotesis, hasil penafsiran, dan kesimpulan yang dapat dievaluasi secara independen oleh orang lain [23]. Sedangkan terdapat empat metode penelitian yang umum digunakan, diantaranya: *Action Research*, *Experiment*, *Case Study*, dan *Survey* [22].

Penelitian ini menggunakan penelitian percobaan. Penelitian percobaan melibatkan penyelidikan perlakuan pada parameter atau variabel tergantung dari penelitiannya dan menggunakan tes yang dikendalikan oleh peneliti itu sendiri, dengan metode penelitian sebagai berikut:



Gbr. 4 Tahapan Penelitian

#### A. Pengumpulan Data

Data kredit Berdasarkan data laporan kredit dari Koperasi Pegawai Kantor Pusat Dirjen Pajak yang diambil dari bulan Januari sampai bulan Desember 2010. Setelah dilakukan proses data *preparation*, atribut atau variabel yang digunakan terdiri dari 15 atribut yang terdapat dalam data status kredit nasabah. Variabel-variabel tersebut ada yang tergolong variabel prediktor atau pemrediksi (*predictor variabel*). Variabel prediktor yaitu umur, jenis kelamin, status pernikahan, jumlah anak, pendidikan, gaji, plafon, jangka

waktu, pokok, bunga, jumlah angsuran, jaminan, tujuan dan *prosentase* Sedangkan variabel tujuan adalah status kredit.

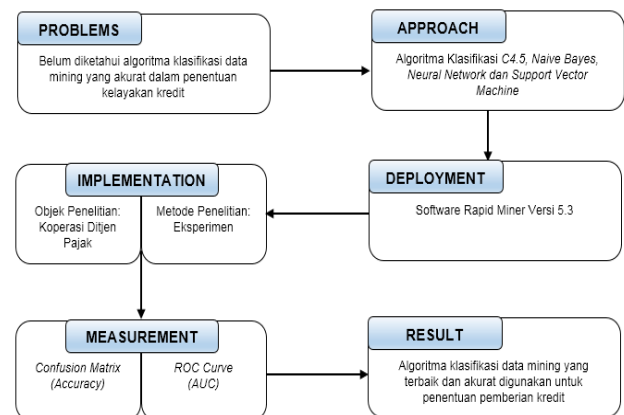
#### B. Pengolahan Awal Data

Dalam mendapatkan data yang berkualitas, beberapa teknik yang dilakukan sebagai berikut [10]:

1. Data validation, untuk mengidentifikasi dan menghapus data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*). Data nasabah atau anggota yang ada akan diidentifikasi apakah ada data yang ganjil (*outlier/noise*), data yang tidak konsisten, dan data yang tidak lengkap (*missing value*).
2. Data *integration and transformation*, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini bernilai kategorikal. Data ditransformasikan kedalam *software Rapidminer*.
3. Data *size reduction and discretization*, untuk memperoleh data set dengan jumlah atribut dan *record* yang lebih sedikit tetapi bersifat informatif. Pada tahap ini data untuk algoritma neural network digunakan data kategorikal yang telah ditransformasikan kedalam bentuk angka.

#### C. Model yang Diusulkan

Metode yang diusulkan yaitu penerapan metode *C4.5*, *naive bayes* dan *neural network* dalam penilaian kelayakan pemberian kredit. Dimulai dari pembagian dataset dengan metode *10 cross validation* yaitu data testing dan data training, kemudian data training diproses dengan membandingkan dari keempat metode tersebut, sehingga menghasilkan model *evaluation* yang diukur dengan nilai *Confusion Matrix* dan *ROC curve* dapat dilihat pada Gambar 5:



Gbr. 5 Model yang diusulkan

#### D. Eksperimen dan Pengujian Model

Pada penelitian ini metode yang digunakan adalah penelitian eksperimen (percobaan). Penelitian percobaan melibatkan penyelidikan hubungan kausal menggunakan tes dikendalikan oleh peneliti itu sendiri. Dalam penelitian percobaan digunakan spesifikasi perangkat lunak dan perangkat keras sebagai alat bantu penelitian pada Tabel 1:

TABEL I  
Spesifikasi *Hardware* dan *Software*

| <i>Hardware</i>       | <i>Software</i>               |
|-----------------------|-------------------------------|
| CPU : Pentium Core i3 | Sistem Operasi: Windows 7     |
| Memory : 4 GB         | Data Mining : Rapid Miner 5.3 |
| Hardisk : 500 GB      |                               |

E. *Evaluasi dan Validasi Hasil*

Model yang diusulkan pada penelitian tentang penentuan kelayakan pemberian kredit adalah dengan menerapkan dan membandingkan antara algoritma klasifikasi *C4.5*, *Naive Bayes*, *Neural Network* dan *Support Vector Machine*. Evaluasi terhadap model yang terbentuk akan dilakukan dengan pengukuran akurasi dan *area under curve*. Akurasi diukur dengan menggunakan *confusion matrix*, dan nilai *AUC* akan diukur dengan menggunakan *ROC Curve*. *Confusion matrix* akan menggambarkan hasil akurasi mulai dari prediksi positif yang benar (tp), prediksi positif yang salah (tn), prediksi negatif yang benar (fp), dan prediksi negatif yang salah (fn). Sehingga model yang terbentuk dapat langsung diuji dengan data yang secara acak dipisahkan dengan *10-folds cross validation*.

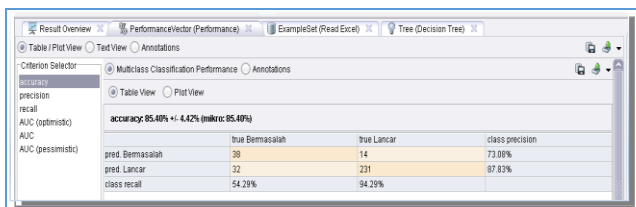
Dengan *confusion matrix*, nilai akurasi dari model akan dibandingkan antara algoritma klasifikasi *C4.5*, *Naive Bayes*, *Neural Network* dan *Support Vector Machine*. Setelah diketahui *ROC* dan *AUC* kemudian dengan menggunakan *software Rapidminer versi 5.3*. Algoritma yang terbaik dan akurat nantinya akan digunakan dalam penentuan kelayakan kredit.

IV. PEMBAHASAN PENELITIAN

Model yang telah dibentuk diuji tingkat akurasinya dengan memasukan data uji yang berasal dari data *training*. Karena data yang didapat dalam penelitian ini setelah proses *preprocessing* Setelah dilakukan *preprocessing* data yang didapat dari koperasi sebanyak 366 *record* direduksi dengan menghilangkan duplikasi menjadi 315 *record* untuk data *training* data maka digunakan metode *cross validation* untuk menguji tingkat akurasi.

A. *Algoritma C4.5*

Hasil perhitungan akurasi data *training* menggunakan Algoritma *C4.5*. Diketahui tingkat akurasinya 85.40% dapat dilihat dengan gambar 6.



Gbr. 6 *Confusion Matrix* Algoritma *C4.5*

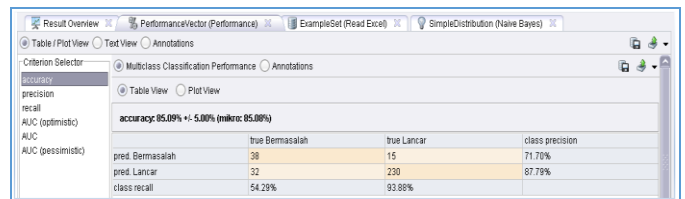
Hasil perhitungan didapat dari pengolahan menggunakan kurva *ROC* untuk algoritma *C4.5* sebesar 0.835 dapat dilihat pada gambar 7.



Gbr. 7 Grafik *AUC* Algoritma *C4.5*

B. *Algoritma Naive Bayes*

Hasil perhitungan akurasi data *training* menggunakan Algoritma *Naive Bayes*. Diketahui tingkat akurasinya 85.09% dapat dilihat dengan gambar 8.



Gbr. 8 *Confusion Matrix* Algoritma *Naive Bayes*

Hasil perhitungan didapat dari pengolahan menggunakan kurva *ROC* untuk algoritma *Naive Bayes* sebesar 0.833 dapat dilihat pada gambar 9.



Gbr. 9 Grafik *AUC* Algoritma *Naive Bayes*

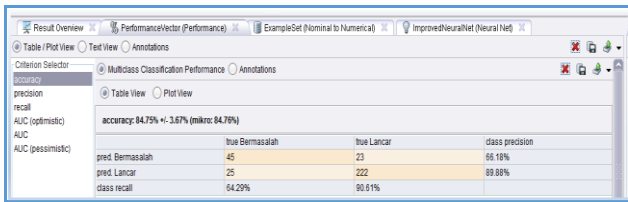
C. *Algoritma Neural Network*

Penelitian ini melakukan uji coba pada nilai *training cycles* ditentukan dengan cara memasukkan nilai *range* dari 100 sampai dengan 1500 untuk *training cycles*, serta nilai 0.1 sampai dengan nilai 0.9 untuk *learning rate* dan nilai 0.0 sampai dengan 0.9 untuk *momentum*, serta dengan percobaan penentuan nilai *neuron size* terhadap 1 hidden layer dengan menggunakan data *training*.

Pada desain model algoritma *neural network* dilakukan proses *training* model dilakukan dengan memberikan *training cycles* sebagai berikut:

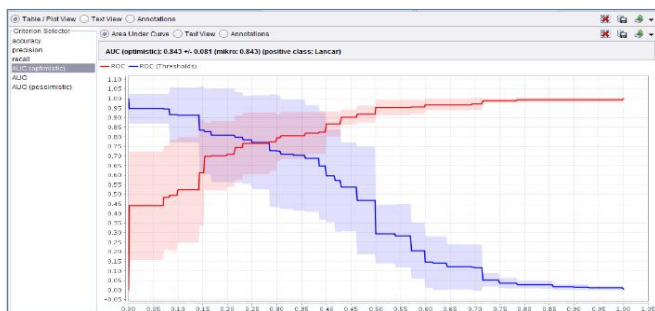
1. *Hidden Layer Size* : 10
2. *Training Cycles* : 200
3. *Learning Rate* : 0.1
4. *Momentum* : 0.0

Hasil perhitungan akurasi data *training* menggunakan Algoritma *Neural Network*. Diketahui tingkat akurasinya 84.75% dapat dilihat dengan gambar 10.



Gbr. 10 Confusion Matrix Algoritma *Neural Network*

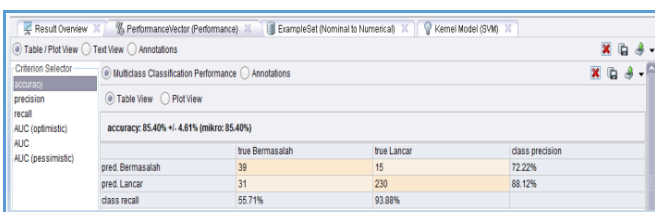
Hasil perhitungan didapat dari pengolahan menggunakan kurva ROC untuk algoritma *Neural Network* sebesar 0.764 dapat dilihat pada gambar 11.



Gbr. 11 Grafik AUC Algoritma *Neural Network*

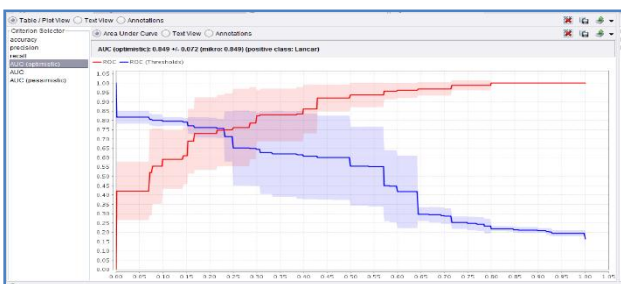
#### D. Algoritma *Support Vector Machine*

Hasil perhitungan akurasi data *training* menggunakan Algoritma *Support Vector Machine*. Diketahui tingkat akurasinya 85.40% dapat dilihat dengan gambar 12.



Gbr. 12 Confusion Matrix Algoritma *Support Vector Machine*

Hasil perhitungan didapat dari pengolahan menggunakan kurva ROC untuk algoritma *Naive Bayes* sebesar 0.8490 dapat dilihat pada gambar 13.



Gbr. 13 Grafik AUC Algoritma *Support Vector Machine*

#### E. Analisis Hasil Komparasi

Dari hasil pengujian dengan mengukur kinerja algoritma *C4.5*, *Naive Bayes*, *Neural Network* dan *Support Vector Machine* dengan menggunakan *10-fold cross validation* dapat dilihat hasilnya pada tabel 4.1, dimana algoritma *C4.5* dan *SVM* menghasilkan nilai akurasi tertinggi sebesar 85.40% diikuti metode *Naive Bayes* menghasilkan nilai akurasi 85.09% sedangkan metode *Neural Network* menghasilkan nilai akurasi 84.75%.

TABEL II  
Nilai Accuracy *10-fold cross validation*

| N-Fold | C4.5          | NB            | NN            | SVM           |
|--------|---------------|---------------|---------------|---------------|
| 1      | 80.32%        | 81.60%        | 79.04%        | 84.13%        |
| 2      | 80.32%        | 81.60%        | 79.04%        | 84.13%        |
| 3      | 82.22%        | 82.54%        | 82.54%        | 84.76%        |
| 4      | 80.64%        | 81.88%        | 81.27%        | 84.45%        |
| 5      | <b>85.40%</b> | 83.49%        | 83.81%        | 84.76%        |
| 6      | 83.51%        | 84.48%        | <b>84.75%</b> | <b>85.40%</b> |
| 7      | 83.49%        | 84.13%        | 83.17%        | 85.08%        |
| 8      | 81.59%        | <b>85.09%</b> | 82.23%        | 85.08%        |
| 9      | 82.54%        | 83.49%        | 81.27%        | 85.08%        |
| 10     | 84.76%        | 83.84%        | 83.11%        | 85.02%        |

Sedangkan perbandingan nilai *accuracy*, *auc*, *precision*, dan *recall* keempat algoritma dapat dilihat pada Tabel II dimana algoritma *SVM* memiliki nilai tertinggi berdasarkan nilai *accuracy*, *AUC* dan *precision*.

TABEL III  
Perbandingan Nilai keempat algoritma

|           | C4.5          | NB     | NN     | SVM           |
|-----------|---------------|--------|--------|---------------|
| Accuracy  | <b>85.40%</b> | 85.09% | 84.48% | <b>85.40%</b> |
| AUC       | 0.8350        | 0.8333 | 0.7640 | <b>0.8490</b> |
| Recall    | <b>94.29%</b> | 93.88% | 95.12% | 94.02%        |
| Precision | 88.11%        | 87.83% | 86.39% | <b>88.41%</b> |

Dalam *Performance* keakurasian *Area Under Curve* (AUC) dapat diklasifikasikan menjadi lima kelompok yaitu [9]:

1. 0.90 – 1.00 = *Excellent Clasification*
2. 0.80 – 0.90 = *Good Clasification*
3. 0.70 – 0.80 = *Fair Clasification*
4. 0.60 – 0.70 = *Poor Clasification*
5. 0.50 – 0.60 = *Failure Clasification*

Berdasarkan klasifikasi tersebut maka dapat disimpulkan bawah Algoritma *C4.5*, *Naive Bayes* dan *Support Vector Machine* termasuk dalam *Good Clasification* sedangkan *Neural Network* termasuk dalam *Fair Clasification*.

#### V. KESIMPULAN

Dari penelitian yang telah dilakukan dalam perbandingan menggunakan algoritma *C4.5*, *Naive Bayes*, *Neural Network* dan *Support Vector Machine* diketahui bahwa algoritma *C4.5* dan *Support Vector Machine* memiliki nilai *accuracy* tertinggi

sebesar 85,40%. Algoritma *Support Vector Machine* memiliki nilai tertinggi berdasarkan nilai *AUC* dan *Precision* sebesar 0,8490 dan 88,41%, sedangkan Algoritma *C4.5* memiliki nilai *Recall* tertinggi sebesar 94,29%.

Dengan demikian metode *Support Vector Machine* adalah metode yang terbaik dari keempat algoritma yang telah diujikan untuk pemecahan masalah kelayakan penentuan kredit.

Untuk keperluan penelitian lebih lanjut dapat dilakukan dengan menambahkan algoritma yang lain seperti *Linear Regression*, *Boosting*, *Bagging*, *Random Forest* dan sebagainya.

#### REFERENSI

- [1] Republik Indonesia, 1998. Undang-undang RI nomor 10 tahun 1998 tentang perubahan atas undang-undang nomor 7 tahun 1992 tentang perbankan. Sekretariat Negara. Jakarta
- [2] Republik Indonesia, 1995. Undang-undang RI nomor 9 tahun 1995 tentang usaha kecil. Sekretariat Negara. Jakarta
- [3] Josep, Zurada. (2010). *Could Decision Trees Improve the Classification Accuracy and Interpretability of Loan Granting Decisions?*. Hawaii: Proceedings of the 43rd Hawaii International Conference on System Sciences.
- [4] A. Sawant, Abhijit, & M. Chawan, P. (2008). *Comparison of Data Mining Techniques used for Financial Data Analysis*. Mumbai: International Journal of Emerging Technology and Advanced Engineering Volume 3 (6), 2013.
- [5] Hájek, Petr. (2011). *Municipal credit rating modelling by neural networks*. Czech Republic: Elsevier Decision Support Systems 51 (2011) 108–118.
- [6] Wu & Kumar. (2009). *The Top Ten Algorithms in Data Mining*. USA: CRC Press.
- [7] Larose, D. T. (2006). *Data Mining Methods and Models*. Hoboken, New Jersey, United State of America: John Wiley & Sons, Inc.
- [8] Larose, D. T. (2005). *Discovering Knowledge in Data*. New Jersey: John Willey & Sons, Inc.
- [9] Gorunescu, F. (2011). *Data Mining Concept, Models and Techniques*. Berlin Heidelberg: Springer Verlag.
- [10] Vercellis, C. (2009). *Data Mining and Optimization for Decision Making*. Italy: WILEY.
- [11] Wu, Xindong & Kumar, Vipin. (2009). *The Top Ten Algorithms in Data Mining*. Boca Raton: CRC Press.
- [12] Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining : Practical Machine Learning and Tools*. Burlington: Morgan Kaufmann Publisher.
- [13] Maimon & Rokach. (2010). *Data Mining and knowledge Discovery Handbook*. New York: Springer.
- [14] Kusriani, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- [15] Han, J., & Kamber, M. (2006). *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman.
- [16] Liao. (2007). *Recent Advances in Data Mining of Enterprise Data: Algorithms and Application*. Singapore: World Scientific Publishing
- [17] Bramer, Max. (2007). *Principles of Data Mining*. London: Springer
- [18] Ethem Alpaydin. (2010). *Introduction to Machine Learning: The MIT Press*.
- [19] Wu & Kumar. (2009). *The Top Ten Algorithms in Data Mining*. USA: CRC Press.
- [20] Kulyal, P., Tiwari, U.K., Shukla, A., & Gaur, A.K. (2010). Chemical Constituent Isolated from *Andrographis paniculata*, *Indian Journal of Chemistry*, 49,356-359.
- [21] Myatt, Glenn J. (2007). *Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining*. New Jersey: John Wiley & Sons, Inc.
- [22] Dawson, C. W. (2009). *Projects in Computing and Information System A Student's Guide*. England: Addison-Wesley.
- [23] Berndtsson, M., Hansson, J., Olsson, B., & Lundell, B. (2008). *A Guide for Students in Computer Science and Information Systems*. London: Springer.
- [24] Shukla, A., Tiwari, R., & Kala, R. (2010). *Real Life Application of Soft Computing*. CRC Press.
- [25] Santosa, B. (2007). *Data Mining Teknik Pemanfaat Data Untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- [26] Susanto, S., & Suryadi, D. (2010). *Pengantar Data Mining menggali Pengetahuan dari Bongkahan Data*. Yogyakarta: C.V ANDI OFFSET.
- [27] Aydin, I., Karakose, M., & Akin, E. (2011). A multi-objective artificial immune algorithm for parameter optimization in support vector machine. *Journal Applied Soft Computing*, 11, 120-129.
- [28] Bellotti, T., & Crook, J. (2007) Support vector machines for credit scoring and discovery of significant features. *Expert System with Application: An International Journal*, 36, 3302-3308.



Achmad Rifai, M.Kom, memperoleh gelar Sarjana Komputer (S.Kom), jurusan Sistem Informasi STMIK Nusa Mandiri Jakarta, lulus tahun 2011 dan memperoleh gelar Magister Komputer (M.Kom) Program Pasca Sarjana Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta, lulus tahun 2015.