

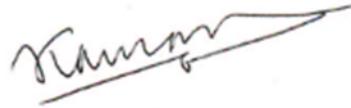
HALAMAN PENGESAHAN

Tesis ini diajukan oleh:

Nama : Saeful Bahri
NIM : 14001186
Program Studi : Magister Ilmu Komputer
Jenjang : Strata Dua (S2)
Konsentrasi : E-Business
Judul Tesis : “Optimasi Algoritma C4.5 Pada Pemilihan Atribut Untuk Analisis Kelayakan Pemberian Kredit Menggunakan Algoritma Genetika Dan Metode *Bagging*”

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar magister Ilmu Komputer pada Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK Nusa Mandiri).

Jakarta, 02 Agustus 2015
Pascasarjana Magister Ilmu Komputer
STMIK Nusa Mandiri
Direktur



Prof. Dr. Ir. Kaman Nainggolan, MS.

DEWAN PENGUJI

Penguji I : **Dr. Sularso Budilaksono**


.....

Penguji II : **Dr. Windu Gata, M.Kom**


.....

Penguji III/
Pembimbing : **Dr. Dwiza Riana, M.Kom, S.Si, M.M**


.....

**OPTIMASI ALGORITMA C4.5 PADA PEMILIHAN ATRIBUT UNTUK
ANALISIS KELAYAKAN PEMBERIAN KREDIT MENGGUNAKAN
ALGORITMA GENETIKA DAN METODE *BAGGING***



TESIS

SAEFUL BAHRI

14001186

**PROGRAM PASCASARJANA MAGISTER ILMU KOMPUTER
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER**

NUSA MANDIRI

JAKARTA

2015

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam-meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam untuk melunasi utangnya setelah jangka waktu tertentu dengan pemberian bunga (UU Perbankan No 10 Tahun 1992), analisa pemberian kredit dilakukan untuk mengevaluasi nasabah atau debitur berdasarkan data historis seperti pendapatan, usia, histori kredit sebelumnya, catatan kriminal dan sebagainya (Akbulgic, 2015). Pada umumnya bank sebagai pemberi kredit atau kreditor melakukan proses pemberian pembiayaan secara garis besar yaitu pengajuan pembiayaan, analisis usulan pembiayaan, persetujuan pihak terkait, perjanjian kredit, dan proses pencairan dana.

Setiap orang yang ingin mendapat pembiayaan atau pinjaman dari bank harus melakukan proses diatas, salah satu yang paling berperan adalah proses analisis usulan pembiayaan yang dilakukan oleh pihak berwenang karena pada proses inilah ditentukan diberikan atau tidaknya pembiayaan atau pinjaman kepada peminjam atau debitur, jika setelah analisis peminjam memiliki nilai baik hal ini akan mengurangi resiko kredit yang akan diterima oleh pihak pemberi pinjaman. Dalam pemberian kredit atau pinjaman saat proses mengembalikan dana tidak selalu berjalan sebagaimana mestinya, pasti akan mengalami ketidaksanggupan peminjam atau debitur untuk membayar angsuran yang telah disepakati. Beberapa faktor yang menentukan nasabah tidak lancar dalam membayar angsuran salah satunya adalah situasi ekonomi sehingga mengganggu usaha yang ia jalankan.

Dalam artian resiko kredit merupakan isu yang paling penting dalam dunia industri perbankan (Oreski & Oreski, 2013), karena akan merugikan terhadap kelangsungan keuangan suatu negara dan berpotensi menimbulkan kesulitan keuangan (Zurada, 2010), untuk mengurangi resiko kredit maka analisa kredit menjadi kunci utama dalam manajemen resiko kredit (Liu & Huang, 2012).

Data historis merupakan data latihan atau pengalaman dari data tersebut akan didapat sebuah pengetahuan. algoritma klasifikasi akan menemukan pengetahuan, pengetahuan untuk menggolongkan resiko kredit seorang nasabah pada masa mendatang berdasarkan variabel yang ada. Sebagai bahan acuan nasabah tersebut diterima atau ditolak.

Berikut merupakan atribut-atribut yang terdapat pada penelitian ini diantaranya digunakan yaitu nama nasabah, jenis kelamin, *Rate*, plafon pinjaman, jangka waktu, jml ansuran per bulan, *ln_type*, *main branch*, no rek, region, branch, *cif no*, sisa angsuran, tunggakan pokok, tunggakan bunga atribut tersebut yang akan dijadikan dasar dalam menentukan risiko kredit disebut variabel prediktor atau prediksi (*predictor variable*). Adapun status kredit yang akan ditentukan adalah variabel tujuan (*target variable*).

Ada beberapa penelitian yang telah dilakukan oleh para peneliti sebelumnya yang berkaitan dengan analisa kredit resiko kredit seperti Zhang & Härdle (2010) yang melakukan penelitian tentang memodelkan resiko kredit dengan menggunakan *Bayesian Additive Classification Tree*, menggunakan *data set* German *credit* dengan memakai 28 variabel atau atribut dengan tingkat akurasi 60,4%. Yu, Chen, Koronios, Zhu, & Guo, (2007) melakukan perbandingan beberapa algoritma klasifikasi salah satunya adalah *Decision tree* atau C4.5 untuk melakukan klasifikasi dalam manajemen resiko hasilnya *Decision tree* memiliki tingkat akurasi paling tinggi dibanding algoritma-algoritma yang lain. Masripah (2011) dalam tesisnya melakukan optimasi C4.5 dengan *Partical swarm Optimization* dengan *data set* yang digunakan *data set* koperasi syariah.

Pohon keputusan telah sukses diterapkan sebagai tool untuk klasifikasi, diagnosa penyakit, prediksi loyalitas pelanggan dan lain sebagainya (cho, Lee, & Jun, 2011) keuntungan pengklasifikasian menggunakan pohon keputusan memiliki kelebihan dalam memecahkan struktur kompleks menjadi struktur yang lebih sederhana sehingga lebih mudah untuk diimplementasikan (cho, Lee, & Jun, 2011), (Abbelan & Masegosa, 2012), akan tetapi pohon keputusan memiliki kelemahan dalam menangani data yang besar dan ketidakseimbangan data (Sarkar, Sana, & Caudhuri, 2011), data *noise* sering muncul pada pohon

keputusan terutama ketika sebuah data set memiliki label yang tidak benar (Abbelan & Masegosa, 2012).

Algoritma genetika merupakan salah satu algoritma untuk pemilihan atribut pada klasifikasi dan berperan dalam meningkatkan akurasi (Zhang & Yang, 2008). Algoritma genetika dapat menangani masalah pada algoritma optimasi yang lain (Machairas, Tsangrassoulis, & Axarli, 2014), lebih efisien dalam mengoptimasi lebih dari sepuluh parameter selain itu algoritma genetika juga mampu menangani masalah *non linier* lokal minimum, data *noise* pada pohon keputusan ditanggulangi dengan menerapkan *bagging bootstarp aggregating* (Breiman, 1996) .

Dari penelitian-penelitian tersebut diatas untuk menangani beberapa kelemahan yang masih ada maka akan diterapkan algoritma pohon keputusan berbasis Algoritma Genetika (AG) yang akan diterapkan untuk pemilihan atribut dan *bagging* akan diterapkan untuk menanggulangi data *noise* yang dihasilkan dari proses pengklasifikasian menggunakan *decision tree* untuk meningkatkan akurasi hasil analisa kelayakan pemberian kredit.

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya maka dapat diidentifikasi permasalahan algoritma klasifikasi C4.5 masih memiliki beberapa masalah ketidakseimbangan data. Sedangkan pertanyaan penelitian apakah algoritma C4.5 bisa dikembangkan lagi dengan memanfaatkan algoritma optimasi seleksi fitur Algoritma Genetika dan *noise* dikurangi menggunakan *bagging* untuk lebih meningkatkan tingkat akurasi untuk analisis kelayakan pemberian kredit ?.

1.3 Tujuan Penelitian

Tujuan pada penelitian ini adalah menerapkan Algoritma Genetika (AG) untuk pemilihan atribut dan penerapan *bagging* untuk mengurangi data *noise* pada algoritma klasifikasi C4.5 dalam meningkatkan tingkat akurasi untuk analisis kelayakan pemberian kredit.

1.3.1. Manfaat Penelitian

Manfaat dari penelitian ini:

1. Manfaat praktis dari penelitian ini dapat digunakan oleh penyedia pinjaman untuk melakukan analisa lebih baik lagi
2. Manfaat kebijakan dapat digunakan sebagai bahan pertimbangan dalam pengambilan keputusan dalam menganalisa kredit
3. Manfaat teoritis dari penelitian ini diharapkan memberikan pengetahuan tentang penerapan metode Data Mining khususnya algoritma C4.5.

1.3.2. Kontribusi Penelitian

Sesuai tujuan penelitian diatas maka kontribusi dari penelitian ini yaitu tercapainya penerapan Algoritma Genetika untuk mencari atribut terbaik dan *bagging* untuk mengurangi data *noise* pada algoritma klasifikasi pohon keputusan C4.5 untuk meningkatkan akurasi hasil klasifikasi.

1.4 Ruang Lingkup Penelitian

Karena analisa kredit menjadi salah satu penentu kesuksesan dalam suatu sistem perbankan. Maka dalam penelitian ini akan membahas mengenai analisa kelayakan pemberian kredit menggunakan algoritma pohon keputusan C4.5 dengan Algoritma Genetika (AG) untuk pemilihan fitur atau atribut dan *bagging* untuk mengurangi tingkat *noise* yang ada dari hasil klasifikasi.

Pada penelitian ini data yang digunakan berupa data primer yang didapat dari hasil riset pada salah satu perbankan nasional, sebagai salah satu acuan, kemudian hasil analisa akan dievaluasi menggunakan *Confusion Matrix* (CM) untuk mengetahui prediksi salah dan benar, ROC (*Receive Operation Chacacteristic*) Curve dengan nilai AUC (*Area Under Curve*) untuk mengetahui tingkat akurasi Aloritm Genetika (AG) dan *Bagging* dalam analisis kelayakan pemberian kredit.

1.4.1. Rumusan Masalah

Rumusan masalah pada penelitian ini:

1. Seberapa besar Algoritma Genetika (AG) dan *Bagging* meningkatkan akurasi hasil klasifikasi algoritma C4.5 dalam analisis kelayakan pemberian kredit ?
2. Bagaimana respon pengguna dalam pemanfaatan GUI (*Graphical User Interface*) yang tepat berdasarkan algoritma klasifikasi C4.5 yang telah di analisa.?

1.5 Hipotesis

Hipotesis merupakan dugaan awal terhadap sebuah penelitian yang dilakukan dalam penelitian ini ditentukan hipotesis sebagai berikut:

Diduga algoritma klasifikasi C4.5 dengan pemilihan atribut yang akan dioptimalkan oleh Algoritma genetika dan *bagging* akan meningkatkan akurasi dibanding dengan algoritma C4.5.

1.6 Sistematika Penulisan

Sistematika dari penulisan tesis ini terdiri dari 5 (lima) bab, dimana setiap bab terdiri dari sub bab dan sub-sub bab yang dapat diuraikan sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini dijabarkan mengenai latar belakang penelitian, rumusan dari permasalahan yang akan dibahas, tujuan yang diharapkan dari penelitian yang dilakukan, manfaat yang didapatkan dari penelitian ini, kontribusi yang diharapkan dapat diberikan melalui penelitian ini, dan sistematika penulisan yang digunakan dalam penyusunan laporan.

BAB II LANDASAN TEORI DAN KERANGKA PEMIKIRAN

Pada bab ini dijelaskan mengenai tinjauan studi yang terdiri dari penelitian sebelumnya yang berhubungan dengan penelitian yang dilakukan, tinjauan pustaka yang berisi landasan teori yang terkait dengan penelitian ini, kerangka pemikiran yang menggambarkan pola

dari penerapan metode dari penelitian yang dilakukan serta hipotesa awal sebagai gambaran harapan awal dari hasil penelitian yang dilakukan.

BAB III METODE PENELITIAN

Pada bab ini menjelaskan mengenai desain penelitian yang digunakan dalam penelitian yang kemudian dijelaskan melalui tahapan pengumpulan data, pengolahan awal data, menyajikan model atau metode yang diusulkan, pemrosesan eksperimen dan pengujian model serta evaluasi dan validasi hasil.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini menjelaskan mengenai hasil penelitian yang diperoleh dari pembahasan, proses pengujian hasil berdasarkan metode yang diterapkan dan implikasi penelitian.

BAB V PENUTUP

Pada bab ini dipaparkan mengenai kesimpulan yang diperoleh dan saran-saran yang dibutuhkan untuk bahan pertimbangan pada penelitian dan pengembangan sistem lebih lanjut.

BAB 2

LANDASAN TEORI DAN KERANGKA PEMIKIRAN

2.1. Tinjauan Pustaka

Dalam penulisan tesis ini digunakan beberapa media diantaranya buku, jurnal, baik jurnal nasional maupun jurnal internasional serta prosiding untuk penunjang teori dan referensi dalam menjelaskan kredit, data mining, model algoritma C4.5, Algoritma Genetika (AG), metode *Bagging (Bootstrap Aggregating)* metode evaluasi *confusion matrix*, *ROC curve* dan *Cross Validation*

2.1.1. Kredit

Pada pasal 1 angka 11 Undang-undang no. 10 tahun 1998 tentang perubahan undang-undang no. 7 tahun 1992 tentang perbankan yang dimaksud dengan kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan pinjam meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam melunasi hutangnya setelah jangka waktu tertentu dengan pemberian bunga. Ada beberapa kriteria yang dilakukan atau prinsip utama yang berkaitan dengan kondisi calon nasabah.

Prinsip ini dikenal dengan 5C yaitu:

1. *Character*

Penilaian terhadap karakter atau kepribadian calon peminjam untuk memperkirakan kemungkinan bahwa peminjam dapat memenuhi kewajibannya.

2. *Capacity*

Penilaian tentang kemampuan peminjam untuk melakukan pembayaran. Kemampuan diukur dengan catatan prestasi peminjam dimasa lalu yang didukung dengan mengamati dilapangan atas sarana dan usahanya seperti karyawan, mesin, sarana produksi, cara usaha, dan lain sebagainya.

3. *Capital*

Penilaian terhadap kemampuan modal yang dimiliki oleh calon peminjam, diukur dengan posisi usaha/perusahaan secara keseluruhan yang ditunjukkan oleh rasio keuangan dan penekanan pada komposisi modal.

4. *Collateral*

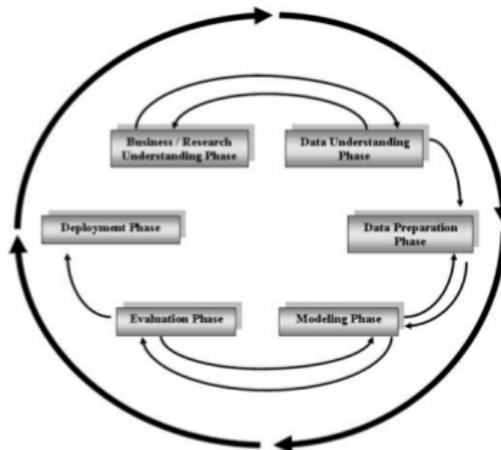
Jaminan yang dimiliki calon peminjam. Penilaian ini untuk lebih meyakinkan bahwa jika suatu risiko kegagalan pembayaran tercapai terjadi, maka jaminan dapat dipakai sebagai pengganti kewajiban.

5. *Conditions*

Pihak Bank harus melihat kondisi ekonomi yang terjadi dimasyarakat dan secara spesifik melihat adanya keterkaitan dengan jenis usaha yang dilakukan oleh calon peminjam. Hal tersebut dilakukan karena kondisi eksternal memiliki pengaruh yang cukup besar dalam proses berjalannya usaha calon peminjam dalam jangka panjang.

2.1.2. *Data Mining*

Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. *Interpretation/Evaluation* Pola informasi yang dihasilkan dari proses *data mining* diterjemahkan menjadi bentuk yang lebih mudah dimengerti oleh pihak yang berkepentingan. *Data mining* merupakan sebuah proses, sehingga dalam melakukan proses tersebut harus sesuai dengan prosedur yaitu yang disebut dengan CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yaitu sebagai keseluruhan proses, *preprocessing* data, pembentukan model, model evaluasi dan akhirnya penyebaran model (Larose, 2005). Penjelasan hal tersebut dapat digambarkan sebagai berikut:



Gambar II. 1. Proses Data Mining menurut CRISP-DM (sumber gambar : *Discovering Knowledge In Data Larose*)

Enam fase tahapan *Crisp* menurut (Larose, 2005):

1. Fase pemahaman bisnis (*Business understanding phase*).
Fase pertama dari proses CRISP-DM disebut juga tahapan pemahaman dalam penelitian, menentukan tujuan proyek penelitian dalam perumusan mendefinisikan.
2. Fase pemahaman data (*Data understanding phase*)
Dilakukan pengumpulan data, kemudian menganalisa data serta evaluasi kualitas data.
3. Fase Pengolahan data (*Data preparation phase*)
Persiapkan data mentah kemudian diatur untuk data akhir yang akan digunakan untuk fase selanjutnya, pilih kasus dan variabel yang diinginkan yang digunakan untuk menganalisa sesuai analisa masalah, lakukan transformasi pada variabel tertentu jika diinginkan, bersihkan data untuk alat pemodelan.
4. Fase Pemodelan (*Modeling phase*)
Pada tahap ini, pilih dan terapkan teknik pemodelan yang tepat, lakukan pengaturan model untuk mengoptimalkan hasil, jika diperlukan lakukan ulang ke tahap persiapan sesuai dengan persyaratan spesifikasi dari teknik *data mining* tertentu.
5. Fase Evaluasi (*Evaluation phase*)
Melakukan evaluasi satu atau lebih model, tentukan apakah model sudah mencapai tujuan yang diterapkan dalam tahap pertama, mengambil keputusan mengenai penggunaan hasil *data mining*.
6. Fase Penyebaran (*Deployment phase*)
Memanfaatkan model yang telah dibuat, *deployment* yang sederhana adalah sampai menghasilkan laporan sedangkan *deployment* yang kompleks adalah melaksanakan model untuk proses *data mining* paralel pada departemen lain.

Fungsi-fungsi yang ada pada *data mining* mengacu pada (Larose, 2005) terdapat enam fungsi *minor* (Susanto & Suryadi, 2010) yaitu:

1. Fungsi deskripsi (*description*)
Fungsi deskripsi adalah cara yang digunakan untuk menggambarkan sekumpulan data secara ringkas. Banyak cara yang digunakan dalam memberikan gambaran secara ringkas bagi sekumpulan data yang besar jumlahnya dan banyak macamnya yaitu Deskripsi Grafis, Deskripsi Lokasi, dan Deskripsi Keragaman.
2. Fungsi estimasi (*estimation*)
Fungsi estimasi adalah fungsi untuk memperkirakan suatu hal yang sudah ada datanya. Fungsi estimasi terdiri dari dua cara yaitu Estimasi Titik dan Estimasi Selang Kepercayaan.
3. Fungsi prediksi (*prediction*)
Fungsi prediksi adalah memperkirakan hasil dari hal yang belum diketahui, untuk mendapatkan hal baru yang akan muncul selanjutnya. Cara memprediksi dalam fungsi ini adalah Regresi Linier.
4. Fungsi klasifikasi (*classification*)
Fungsi klasifikasi atau menggolongkan suatu data. Cara yang digunakan terdiri dari algoritma *Mean Vector*, algoritma *K-nearest Neighbor*, algoritma ID3, algoritma C4.5, dan algoritma C5.0.
5. Fungsi penegelompokan (*cluster*)
Fungsi pengelompokan, data yang dikelompokan disebut objek atau catatan yang memiliki kemiripan atribut kemudian dikelompokan pada kelompok yang berbeda. Algoritma yang digunakan adalah algoritma *Hirarchical Clustering*, algoritma *Partitional Clustering*, algoritma *Single Linkage*, algoritma *Complete Linkage*, algoritma *Average Linkage*, algoritma *K-Means* dan lain-lain.
6. Fungsi Asosiasi (*Asosiation*)
Fungsi asosiasi adalah untuk menemukan aturan asosiasi (*association rule*) yang mampu mengidentifikasi item-item yang menjadi objek. Algoritma yang digunakan adalah algoritma *Generalized Association Rules*, *Quantitative Association Rule*, *asynchronous Parallel Mining*.

Dalam penelitian ini fungsi *data mining* yang digunakan adalah fungsi klasifikasi dengan algoritma C4.5 dalam penggunaan klasifikasi dalam penelitian ini digunakan data latih atau data pengalaman data yang akan dihasilkan berupa data pengetahuan. Pengetahuan Analisis Kelayakan pemberian kredit.

2.1.3. Algoritma C4.5

Algoritma C4.5 atau yang dikenal dengan pohon keputusan merupakan metode klasifikasi dan prediksi dengan menggunakan aturan-aturan yang mudah dipahami dengan bahasa alami (Kusrini & Luthfi, 2009), C4.5 merupakan algoritma hasil dari evolusi algoritma ID3 dalam algoritma C4.5 menggunakan *Gain Ratio* sebagai pembagi dalam pembentukan kriteria (Maimon & Rokach, 2010). Beberapa tahapan dalam pembentukan pohon keputusan menggunakan algoritma C4.5 (Larose, 2005) :

1. Mempersiapkan *data training*, *data training* biasanya diambil dari data *history* yang pernah terjadi sebelumnya atau disebut data masa lalu dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menghitung Total *Entropy* sebelum dicari masing-masing *Entropy class*

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2.1)$$

Keterangan:

S = Himpunan Kasus

n = Jumlah Partisi S

p_i = Proporsi dari S_i terhadap S

Hitung nilai gain :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2.2)$$

Keterangan:

S = Himpunan Kasus

A = Atribut

n = Jumlah Partisi Atribut A

$|S_i|$ = Jumlah Kasus pada partisi ke- i

$|S|$ = Jumlah Kasus dalam S

3. Ulangi langkah ke-2 dan ke-3 hingga semua tupel terpartisi.
4. Proses partisi pohon keputusan akan berhenti pada saat:
5. Semua tupel dalam *node n* mendapat kelas yang sama
6. Tidak ada atribut didalam tupel yang dipartisi lagi
7. Tidak ada tupel didalam cabang yang kosong.

2.1.4. Genetic Algorithm

Genetic algorithm merupakan sebuah prosedur optimasi yang diilhami oleh evolusi genetik organisme, Algoritma ini pertama kali di perkenalkan oleh John Holland 1970an, algoritma ini menggunakan metode pencarian *heuristic* yang mengubah nilai-nilai fungsi individu kedalam biner melalui penerapan secara *stokastik*, operator yang termasuk didalamnya diantaranya *crossover*, mutasi, *dropping condition* dan lain-lain operator ini disebut kromosom, kromosom dibagi-bagi lagi kedalam beberapa bagian disebut gen (Sarkar, Sana, & Caudhuri, 2011).

Beberapa struktur umum pada Algoritma Genetika (AG) seperti istilah populasi, dalam sebuah populasi terdapat individu yang disebut kromosom, kromosom merupakan sebuah solusi yang masih berbentuk simbol, pada awalnya populasi dibangun secara acak, sedangkan populasi berikutnya hasil dari evolusi kromosom-kromosom melalui iterasi yang disebut dengan istilah generasi. Pada setiap generasi *kromosom* akan dievaluasi dengan alat ukur yang dikenal dengan *fitness*. Nilai *fitness* akan menunjukkan kualitas dari sebuah kromosom, kemudian terciptalah sebuah generasi berikutnya yang disebut dengan *offspring*. *Offspring* terbentuk dari gabungan 2 kromosom generasi sekarang yang menjadi induk dari *offspring* atau *parent* dengan operator penyilangan *crossover* selain *crossover* bisa juga menggunakan operator mutasi (Kusumadewi, 2003)

2.1.5. Bagging (Bootstrap Aggregating)

Bagging merupakan kependekan dari *Bootstrap Aggregating* menggunakan *sub-data set (bootstrap)* untuk menghasilkan set pelatihan L (*Learning*) yang tidak stabil kemudian ketika *training* data mengambil hasil rata-rata (Breiman, 1996). *Bagging* adalah sebuah algoritma pembelajaran yang stabil pada perubahan kecil dalam *training set* menyebabkan perbedaan besar dalam peserta didik yang

dihasilkan, yaitu algoritma belajar pada data yang memiliki varian tinggi (*noise*). *Bagging* mampu meningkatkan akurasi secara signifikan lebih besar dibanding model individual, dan lebih kuat terhadap efek *noise* dan *overfitting* dari data pelatihan asli. (Culp, Michaelidis, & Johnson, 2011) Algoritma *Bagging* (Breiman, 1996).

Perulangan for $b = 1, 2, \dots, B$

1. Buat sampel *bootstrap* $\{(X^*1, Y^*1), (X^*2, Y^*2), \dots, (X^*n, Y^*n)\}$ dengan penggantian secara acak dari *data training* $\{(X^*1, Y^*1), (X^*2, Y^*2), \dots, (X^*n, Y^*n)\}$ mencocokkan dengan *classifier* C_b dinyalakan pada sampel yang sesuai *bootstrap*.
2. *Output classifier* akhir: $C(x) = B^{-1} \sum_{b=1}^B C_b(x)$ (2.3)

Karya (Breiman, 1994) pada Kim & Kang melaporkan bahwa *bagging* dapat meningkatkan kinerja dengan penggabungan (*ensemble*) algoritma seperti *Decision Tree* (DT), *Neural Network* (NN), dan *Support Vector Machine* (SVM) (M. Kim & Kang, 2012). Dataset dengan *noise* yang tinggi menyebabkan kesalahan dalam generalisasi pengklasifikasian, sehingga dibutuhkan algoritma yang tepat untuk digabungkan (*ensemble*) dengan *neural network* agar akurasi prediksi dapat meningkat.

2.1.6. Confusion Metric

Evaluasi kinerja model klasifikasi didasarkan pada pengujian objek yang diprediksi dengan benar dan salah, hitungan ini ditabulasikan *confusion matrix* (Gorunescu, 2011). *Confusion Matrix* adalah perangkat yang berguna untuk menganalisis seberapa baik *classifier* dapat mengenali *tupel* dari kelas yang berbeda (Han & Kamber, 2006). *Confusion matrix* memberikan rincian klasifikasi, kelas yang diprediksi akan ditampilkan di bagian atas *matrix* dan kelas yang diobservasi ditampilkan dibagian kiri (Gorunescu, 2011). Evaluasi model *confussion matrix* menggunakan tabel seperti *matrix* dibawah ini:

Tabel II. 1. *Confusion Matrix*

<i>Classification</i>	<i>Predicted Class</i>		
		<i>Class = Yes</i>	<i>Class = No</i>
<i>Observed Class</i>	<i>Class = Yes</i>	<i>(True Positive-TP)</i>	<i>(False Negative-FN)</i>
	<i>Class = No</i>	<i>(False Positive-FP)</i>	<i>(True Negative-TN)</i>

Sumber: Gorunescu (2011)

False Positive dikenal sebagai *error type 1*, terjadi ketika kasus yang seharusnya diklasifikasikan sebagai negatif diklasifikasikan sebagai *positif*. *False Negatif* dikenal sebagai *error type 2*, terjadi ketika kasus yang seharusnya diklasifikasikan sebagai positif diklasifikasi sebagai negatif (Bramer, 2007). Pada tahap evaluasi menggunakan *confusion matrix* yang dilakukan menggunakan *tool rapid miner* akan diperoleh nilai *accuracy*, *sensitivity*, *specificity*, PPV dan NPV. Akurasi dapat dihitung menggunakan rumus:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.4)$$

Dimana:

TP : Jumlah kasus positif yang diklasifikasikan sebagai positif

FP : Jumlah kasus negatif yang diklasifikasikan sebagai positif

TN : Jumlah kasus negatif yang diklasifikasikan sebagai negatif

FN : Jumlah kasus positif yang diklasifikasikan sebagai negatif

Sensitifitas dan spesifitas dapat digunakan sebagai ukuran statistik dari kinerja klasifikasi biner, sensitifitas dan spesifitas digunakan untuk mengukur model yang paling baik dan untuk memilih model yang paling efisien. Sensitifitas mengukur proporsi *true positive* yang diidentifikasi dengan benar, spesifitas mengukur proporsi *true negative* yang diidentifikasi dengan benar, dapat dihitung menggunakan rumus: $Sensitivity = \frac{NumberofTruePositives}{NumberofTruePositives+NumberofFalseNegative}$ (2.5)

$$Specificity = \frac{NumberofTrueNegatives}{NumberofTrueNegatives + NumberofFalsePositives} \quad (2.6)$$

Sensitifitas juga dapat dikatakan *true positive rate (TP rate)* atau *recall*. Nilai *sensitivity* 100% berarti menunjukkan bahwa pengklasifikasian mengakui sebuah kasus yang diamati positif. Misalnya salah nasabah yang memiliki kemungkinan lancar. Sedangkan

untuk PPV (*Prediktive Positif Value*) adalah proporsi kasus dengan hasil diagnosa positif, NPV (*Prediktif Negative Value*) adalah proporsi kasus dengan hasil diagnosa negatif, dapat dihitung menggunakan rumus:

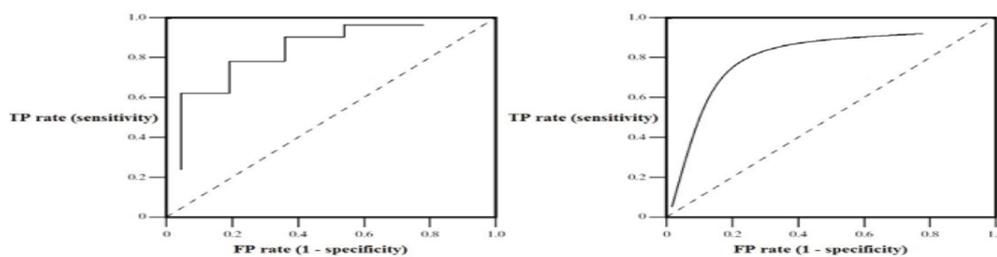
$$PPV = \frac{\text{Number of True Positives}}{\text{Number of True Positives} + \text{Number of False Positives}} \quad (2.7)$$

$$NPV = \frac{\text{Number of True Negatives}}{\text{Number of True Negatives} + \text{Number of False Negatives}} \quad (2.8)$$

2.1.7. ROC Curve

Kurva ROC (*Receiver Operating Characteristic*) banyak digunakan untuk menilai hasil prediksi, kurva ROC adalah teknik untuk memvisualisasikan, mengatur, dan memilih pengklasifikasian berdasarkan kinerja mereka (Gorunescu, 2011). Kurva ROC adalah perangkat dua dimensi yang digunakan untuk menilai kinerja klasifikasi yang menggunakan dua class keputusan, masing-masing objek dipetakan ke salah satu elemen dari himpunan pasangan, positif atau negatif.

Secara teknik, kurva ROC juga disebut grafik ROC, dua dimensi grafik yaitu TP rate diletakan pada sumbu Y sedangkan FP rate diletakan pada sumbu X. Grafik ROC menggambarkan *trade-off* antara manfaat (*true positives*) dan biaya (*false positives*). Berikut tampilan dua jenis kurva ROC (*discrete dan continuous*).



Gambar II. 2 Grafik ROC (*Discrete dan Continuous*)

Berdasarkan grafik di atas dapat diketahui bahwa titik kiri bawah pada koordinat (0,0) yaitu diantara nilai TP dan FP, titik (1,1) merupakan klasifikasi positif. Titik (0,1) adalah klasifikasi sempurna yaitu tidak ada FN dan tidak ada FP yang benar-benar acak akan memberikan titik sepanjang garis diagonal dari kiri bawah ke sudut kanan atas. Garis ini membagi ruang ROC menjadi sebagai berikut:

1. Poin diatas garis diagonal merupakan hasil klasifikasi yang baik.
2. Poin di bawah garis diagonal merupakan hasil klasifikasi yang buruk.

Dapat disimpulkan bahwa dengan melihat *ROC Curve* yang arah garisnya melintang dari kiri bawah ke kanan atas berarti evaluasi menunjukkan hasil yang baik dibandingkan bentuk grafik lainnya. Untuk klasifikasi *data mining*, nilai AUC dapat dibagi menjadi beberapa kelompok (Gorunescu, 2011).

- a. 0.90-1.00 = *Excellent Classification*
- b. 0.80-0.90 = *Good Classification*
- c. 0.70-0.80 = *Fair Classification*
- d. 0.60-0.70 = *Poor Classification*
- e. 0.50-0.60 = *Failure*.

2.1.8. *K-fold Cross Validation*

Cross Validasi adalah sebuah metode statistik untuk mengevaluasi dan membandingkan dengan cara membagi data menjadi dua segmen, satu segmen digunakan untuk data *training* dan satu segmen untuk data *testing*, ciri khas dari *cross validation* adalah data diulang sampai semua data menjadi data *testing* dan *training* (Rafaeilzadeh, Tang, & Liu, 2009), dalam *cross validasi* data dibagi kedalam partisi yang sama tergantung jumlah K, dalam *data mining cross validasi* yang paling umum digunakan ialah *10 fold validation* caranya dengan cara membagi data kedalam 10 set, ukuran set jumlah data/10 kemudian 9 set data untuk untuk training dan satu set data untuk testing ulangi langkah tersebut sampai 10 kali iterasi, pada Tabel II. 2. Ilustrasi *10 fold Cross Validation* Dibawah akan digambarkan ilustrasi dari proses *10 fold cross validation*.

Tabel II. 2. Ilustrasi 10 *fold Cross Validation*

Ekspirimen	Dataset									
1	Orange	Grey								
2	Grey	Orange	Grey							
3	Grey	Grey	Orange	Grey						
4	Grey	Grey	Grey	Orange	Grey	Grey	Grey	Grey	Grey	Grey
5	Grey	Grey	Grey	Grey	Orange	Grey	Grey	Grey	Grey	Grey
6	Grey	Grey	Grey	Grey	Grey	Orange	Grey	Grey	Grey	Grey
7	Grey	Grey	Grey	Grey	Grey	Grey	Orange	Grey	Grey	Grey
8	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Orange	Grey	Grey
9	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Orange	Grey
10	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Grey	Orange

Dari Tabel II. 2. Ilustrasi 10 *fold Cross Validation* Diatas dapat dideskripsikan sebagai berikut *Cell* yang berwarna *orange* merupakan *K-subset* atau data *testing*, proses pengujian berulang sampai 10 kali dan hasil pengukuran terbaik adalah nilai rata-rata dari 10 kali pengujian.

2.2. Tinjauan Studi

Beberapa penelitian terdahulu yang relevan dengan analisis kelayakan kredit menggunakan beberapa metode. Hal ini digunakan sebagai tolak ukur dari hasil penelitian yang telah dicapai. Penelitian dibidang analisa kredit yang menggunakan *data mining* dengan data yang bersifat historis diantaranya:

1. Penelitian yang dilakukan Mandala, Nawangpalupi, & Praktikto, (2012) Melakukan penelitian tentang penilaian resiko kredit pada salah satu BPR (Bank Perkreditan Rakyat), menggunakan algoritma pohon keputusan dengan menggunakan beberapa atribut sebagai prediktor diantaranya Jenis kelamin, Usia, jumlah kredit, penghasilan perbulan, pengeluaran tiap bulan, cicilan perbulan, simpanan, jenis agunan, nilai agunan, priode pinjaman, jenis aktifitas usaha, sumber penghasilan dan status kredit sebelumnya, merupakan atribut-atribut dalam penentuan pemberian kredit.

2. Penelitian yang dilakukan oleh Zurada (2010) tentang komparasi beberapa metode seperti *Logistic Regression (LR)*, *Neural Network (NN)* dan Pohon keputusan (*DTs*) . yang diaplikasikan untuk alasan penolakan kredit, pohon keputusan memiliki tingkat akurasi klasifikasi paling baik dari model-model yang lain, selain hasil klasifikasi yang baik tapi juga memiliki aturan-aturan yang mudah ditafsirkan.
3. Penelitian oleh Zhang & Härdle, (2010) tentang pemodelan resiko kredit menggunakan BACT (*Bayesian Additive Classification Tree*). Pada penelitiannya menggunakan *data set* German Credit dengan menggunakan 28 atribut yang menghasilkan nilai akurasi 60,4 %.
4. Penelitian yang dilakukan oleh Rao & Singh (2013) tentang mendeteksi penipuan keuangan menggunakan pohon keputusan menggunakan teknik *ensemble learning* pada penelitiannya pohon keputusan ditambahkan dengan model *AdaBoost*, sedangkan *data set* yang digunakan adalah *data set* German Credit dengan menggunakan 21 variabel termasuk 1 variable *class risk Good and bad*, dari hasil penelitian didapat nilai akurasi sebesar 79%.

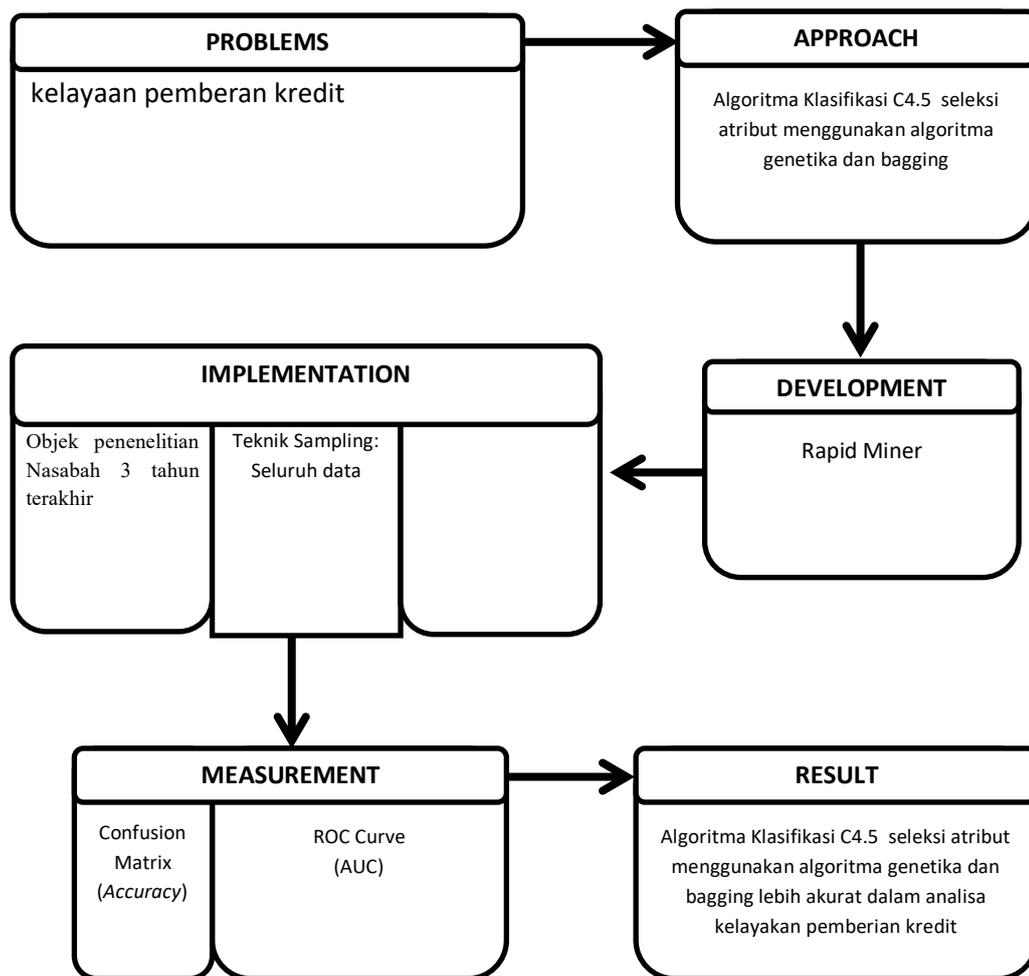
Tabel II. 3. Tinjauan Penelitian Terdahulu

Penelitian	Persamaan Penelitian yang akan datang	Perbedaan Penelitian yang akan datang
1. Mandala, Nawangpalupi, & Praktiko, (2012) Melakukan penelitian tentang penilaian resiko kredit pada salah satu BPR	Melakukan penelitian tentang penilaian resiko pemberian kredit	Algoritma pohon keputusan yang digunakan, data set untuk melakukan training data
2. Zurada (2010) tentang komparasi beberapa metode seperti <i>Logistic Regression (LR)</i> , <i>Neural Network (NN)</i> dan Pohon keputusan (<i>DTs</i>)	Algoritma klasifikasi yang digunakan, hasil berupa akurasi, ROC	Optimasi pemilihan atribut menggunakan GA dan <i>bagging</i> untuk mengurangi <i>noise</i> pada hasil klasifikasi

Penelitian	Persamaan Penelitian yang akan datang	Perbedaan Penelitian yang akan datang
3. Zhang & Härdle, (2010) tentang pemodelan resiko kredit menggunakan BACT (<i>Bayesian Additive Classification Tree</i>)	Memodelkan pohon keputusan untuk penentuan pemberian kredit	Model pohon keputusan yang digunakan, data <i>set</i> yang digunakan
4. Rao & Singh (2013) tentang mendeteksi penipuan keuangan menggunakan pohon keputusan menggunakan teknik <i>esemble learning</i>	Optimasi Pohon keputusan dengan teknik <i>esemble</i> dengan boosting	Metode <i>esemble</i> yang digunakan

2.1.1. Kerangka konsep penelitian

Kerangka konsep penelitian yang dilakukan dalam penelitian ini dijelaskan dalam bagan dan penjelasan deskriptif. Penelitian dimulai dengan mengidentifikasi masalah yang ada yaitu belum diketahuinya seberapa akurat metode klasifikasi terpilih yaitu algoritma pohon keputusan C4.5 jika dalam pemilihan atribut dioptimalkan dengan Algoritma genetika (AG) dan *bagging* dalam analisis kelayakan pemberian kredit. *CRISP-DM* sebagai proses standar *data mining* digunakan untuk mendapatkan *rule* dari masing-masing algoritma, *rule* tersebut kemudian divalidasi dengan menggunakan *confussion matrix* dan kurva *ROC*. Hasil validasi masing-masing algoritma kemudian dikomparasi untuk mengetahui *rule* yang terbaik yang memiliki akurasi yang paling tinggi. *Rule* yang terpilih akan digunakan untuk diterapkan pada data baru. Model kerangka yang digunakan pada penelitian ini dapat digambarkan sebagai berikut:



Gambar II. 3. Kerangka Penelitian

2.3. Tinjauan Objek Penelitian

Pada penelitian ini objek penelitian yang digunakan adalah data perkreditan yang diambil dari hasil riset pada salah satu bank penyalur kredit terbesar di Indonesia, bank tersebut dipilih karena bank tersebut termasuk kedalam *top 10* pemilik *asset* terbesar di Indonesia

Tabel II. 4. Daftar 10 besar Bank dengan *asset* terbesar

No	Nama Bank	Total Asset
1.	Bank Mandiri	Rp 674,74 triliun
2.	Bank Rakyat Indonesia (BRI)	Rp 621,98 triliun
3.	Bank Central Asia (BCA)	Rp 512,84 triliun
4.	BNI	Rp 388,01 triliun
5.	CIMB Niaga	Rp 224,83 triliun
6.	Bank Permata	Rp 176,57 triliun
7.	Bank Panin	Rp 156,72 triliun
8.	Bank Danamon	Rp 154,42 triliun
9.	BII	Rp 137,79 triliun
10.	BTN	Rp 135,62 triliun

Sumber: <http://bisniskeuangan.kompas.com/read/2014/08/12/095520626/Ini.10.Bank.dengan.Aset.Terbesar.di.Indonesia>

Dari Tabel II. 4. Daftar 10 besar Bank dengan *asset* terbesar bisa dilihat bahwa BRI (Bank Rakyat Indonesia), memiliki jumlah asset kedua terbesar, tidak hanya itu BRI juga menyandang gelar sebagai bank spesialis micro

BAB 3

METODE PENELITIAN

3.1. Jenis Penelitian

Dalam penelitian ada empat metode umum yang digunakan diantaranya *Action Research*, *Experiment*, *Case Study* dan *Survey* (Dawson, 2009), metode penelitian dalam penelitian ini menggunakan metode penelitian *experiment*. Penelitian model *experiment* merupakan metode penyelidikan antara hubungan sebab akibat menggunakan tes yang dikendalikan oleh peneliti (Dawson, 2009). Pada penelitian jenis ini terdiri dari :

1. Mendefinisikan hipotesis teoritis
2. Memilih sampel dari populasi yang diketahui
3. Mengalokasikan sampel untuk kondisi percobaan yang berbeda
4. Memperkenalkan perubahan yang direncanakan untuk satu atau lebih *variable*.
5. Mengukur sejumlah kecil variabel
6. Mengontrol semua *variable*.

Dalam penelitian eksperimen ini dilakukan proyek pengembangan, evaluasi dan pemecahan masalah. Beberapa spesifikasi *hardware* dan *software* yang dibutuhkan untuk membantu penelitian *eksperiment* ini adalah sebagai berikut:

<i>Hardware</i>	<i>Software</i>
CPU : Intel® Celeron(R) CPU 1007U @ 1.50GHz × 2	Sistem Operasi : Linux Ubuntu 15.04
Memory : 1,9 GiB	Data mining : Rapid Miner 5.3.013 For Linux
Hardisk : 490,1 GB	

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

1. *Bussiness understanding*
2. *Data understanding*.
3. *Data preparation*
4. *Modelling*
5. *Evaluation*
6. *Deployment*.

1. Tahap *Business Understanding*

Data set yang digunakan pada penelitian ini ialah data sekunder dengan jumlah data sebanyak 766 record, terdiri dari 15 variabel atau atribut dan 1 *class* yang bernilai MACET atau LANCAR. Atribut yang digunakan sebagai *prediktor* oleh peneliti terdahulu ada 15 atribut termasuk *class* diantaranya nama nasabah, jenis kelamin, umur, jumlah pinjaman, jangka waktu, jumlah angsuran perbulan, tipe pinjaman, jenis pinjaman, bi sektor ekonomi, col bi golongan debitur, bi golongan penjamin, saldo nominatif, *plafon* teoritis, tunggakan pokok, dan tunggakan bunga dan untuk *class* atau tujuan adalah lancar dan macet, sedangkan pada penelitian ini atribut yang digunakan yaitu nama nasabah, jenis kelamin, *Rate*, plafon pinjaman, jangka waktu, jml angsuran per bulan, *ln_type*, *main branch*, no rek, region, branch, *cif no*, sisa angsuran, tunggakan pokok, tunggakan bunga.

2. Tahap *Data Understanding*

Data merupakan data sekunder yang didapat dari hasil riset, atribut atau variabel yang ada sebanyak 15 atribut (data selengkapnya dapat dilihat pada lampiran). Variabel-variabel tersebut ada yang tergolong variabel prediktor atau pemrediksi (*predictor variabel*) yaitu variabel yang dijadikan dasar sebagai penentu lancar atau macet status dari nasabah yang bersangkutan, dan variabel tujuan yaitu variabel yang dijadikan sebagai MACET atau LANCAR. Variabel prediktor yaitu digunakan yaitu nama nasabah, jenis kelamin, *Rate*, plafon pinjaman, jangka waktu, jml angsuran per bulan, *ln_type*, *main branch*, no rek, region, branch, *cif no*, sisa angsuran, tunggakan pokok, tunggakan bunga.

3. Tahap *Data Preparation*

Data pada penelitian ini berjumlah 766 yang kemudian dibagi kedalam 10 *set* menjadi masing-masing *set* 76 tupel, dengan rincian 9 *set* untuk *data training* dan 1 *set* untuk *data testing* (Rafaeilzadeh, Tang, & Liu, 2009), proses berulang hingga 10 kali *iterasi* sehingga dari sebagai langkah persiapan penelitian untuk mendapatkan *dataset* yang berkualitas tinggi, terdapat beberapa teknik yang dapat

dilakukan digunakan dalam analisis data mining diantaranya adalah (Vercellis, 2009):

1. *Data Cleaning* untuk membersihkan nilai yang kosong atau tupel yang kosong.
2. *Data Integration* yang berfungsi menyatukan tempat penyimpanan yang berbeda kedalam satu data. Dalam kasus ini data yang diambil dari dari sistem informasi debitur, di satukan dalam sebuah file dengan format excel
3. *Data reduction* jumlah atribut yang ada pada data nasabah sebanyak 31 atribut kemudian direduksi menjadi sekitar 15 atribut yang berpengaruh langsung terhadap pengambilan keputusan dalam analisa pemberian kredit (M.akbar anugrah, Analisis kredit BRI Kanca sukabumi). Berikut atribut hasil *data reduction* digunakan yaitu nama nasabah, jenis kelamin, *Rate*, plafon pinjaman, jangka waktu, jml angsuran per bulan, *ln_type*, *main branch*, no rek, region, branch, *cif no*, sisa angsuran, tunggakan pokok, tunggakan bunga.

Tabel II. 5. Data nasabah hasil *preprocessing*

nama_nasabah	jenis_kelamin	rate	plafon_pinjaman	jangka_waktu	jml_angsuran_per_bulan	ln_tipe	main_branch	no_rek	Region	Branch	cif_no	sisa_angsuran	tunggakan_pokok	tunggakan_bunga	status_kredit
x1	P	24.00	345000	1	345000	HI	92	0000441301002959109	F	4413	ACP4854	345000	345000	0	MACET
x2	L	12.24	350000	7	55716	H5	92	0000441301003268103	F	4413	WK18661	390000	111428	0	MACET
x3	L	12.48	649926	6	108321	H5	92	0000441301003387101	F	4413	DU93186	649926	216642	0	MACET
x4	P	12.24	459168	12	38264	H5	92	0000441301003669105	F	4413	DX05418	459168	382640	0	MACET
x5	P	19.20	3055499	8	381937.41	HA	92	0000441301003684105	F	4413	ASI2271	3055499	1527749.48	0	MACET
x6	L	14.40	2000000	16	0	LI	92	0000441301003823107	F	4413	BK23284	-85000	0	0	LANCAR
x7	L	19.20	8333334	10	833333.4	HA	92	0000441301003848107	F	4413	JL51141	8333334	3333333.6	0	MACET
x8	L	12.24	4435001	8	671098	H5	92	0000441301003865109	F	4413	NAH0283	4435001	0	0	LANCAR
x9	L	19.20	560000	8	95221	HA	92	0000441301003899108	F	4413	YC26971	660800	100800	0	MACET
x10	L	14.40	1443750	15	107800	HA	92	0000441301003933106	F	4413	AQP7350	1617000	1078000	0	MACET
x11	L	12.24	3066000	10	351670	H5	92	0000441301003972100	F	4413	HT90181	3066000	613200	90140	MACET
x12	P	12.30	4071669	20	203583.45	LI	92	0000441301004029104	F	4413	MMV9839	3671669	1228667.6	0	MACET
x13	L	12.48	228655000	33	7495303.73	H5	92	0000441301004040100	F	4413	EX89325	20940492	91612122.24	7362732	MACET
x14	L	19.20	840000	4	60000	HA	92	0000441301004043108	F	4413	SRV4296	240000	240000	0	MACET
x15	L	12.24	3000000	24	147500	H5	92	0000441301004082102	F	4413	EX99485	2500000	875000	157500	MACET
x16	P	12.24	1640000	16	102500	H5	92	0000441301004101100	F	4413	ATW5646	1600000	985000	0	MACET
X766	L	14.40	930000.01	4	45000.01	HU	92	0000441301004130109	F	4413	RGD5622	180000.01	180000.01	0	MACET

Berdasarkan Tabel II. 5. Data nasabah hasil *preprocessing* tidak semua data bersifat kategorikal melainkan adapula data yang bernilai angka. Berdasarkan Tabel II. 5. Data nasabah hasil *preprocessing* kemudian ditentukan pembuatan kandidat pohon, penentuan kandidat pohon dilakukan dengan cara memasukan seluruh atribut yang untuk kemudian dilakukan penilaian pada atribut-atribut sehingga menghasilkan atribut yang mempengaruhi dalam klasifikasi

Tabel II. 6. *Candidat split* dan *rule* atribut algoritma C4.5

<i>Candidat split</i>	<i>Child Node</i>	
1	Tunggakan Pokok ≤ 84199.865 ≤ 166833.330 ≤ 313750.005 ≤ 236166.645 ≤ 22670.550 ≤ 6722	Tunggakan Pokok > 84199.865 > 166833.330 > 313750.005 > 236166.645 > 22670.550 > 6722
2	Plafon Pinjaman ≤ 1831667 ≤ 780000 ≤ 960000 ≤ 441875	Plafon Pinjamann > 1831667 > 780000 > 960000 > 441875
3	LN <i>Type</i> =H5 LN <i>type</i> = HA LN <i>Type</i> =H5 LN <i>type</i> = HA LN <i>type</i> = HI LN <i>type</i> = HU LN <i>type</i> = HY LN <i>type</i> = KB LN <i>type</i> = KJ LN <i>type</i> = LI	
4.	No Rekening ≤ 441301005318601.500	No Rekening >441301005318601.500
5.	Jml angsuran per bulan ≤ 221250.115 ≤ 44267.690 ≤ 40231.260 ≤ 15238.030	Jml angsuran per bulan > 221250.115 > 44267.690 > 40231.260 > 15238.030

Untuk *Candidat Split* dengan algoritma C4.5 dengan GA dan *Bagging* bisa dilihat pada tabel dibawah ini :

Tabel II. 7. *Candidat split* dan *rule* atribut algoritma C4.5 dengan GA dan *Bagging*

<i>Candidat split</i>	<i>Child Node</i>	
1	Tunggakan Pokok ≤ 7055.330 ≤ 166833.330 ≤ 72916.665	Tunggakan Pokok > 7055.330 > 166833.330 > 72916.665
2.	Plafon Pinjaman ≤ 925208.380 ≤ 14692499.500	Plafon Pinjaman > 925208.380 > 14692499.500
3.	Jml angsuran per bulan ≤ 12750 ≤ 280000 ≤ 1311416.475	Jml angsuran per bulan > 12750 > 280000 > 1311416.475
4.	Tunggakan Bunga ≤ 1756 ≤ 2500	Tunggakan Bunga > 1756 > 2500

4. Tahap *Modeling*.

Tahap *modelling* dilakukan untuk menerapkan teknik yang tepat guna mendapatkan hasil yang optimal dalam analisis kelayakan pemberian kredit. Pada penelitian ini model yang digunakan yaitu algoritma terpilih pohon keputusan C4.5 dan algoritma pohon keputusan C4.5 dengan Algoritma Genetika (AG) berbasis *bagging* sebagai penyeleksi atribut. Tahap *modelling* dilakukan dengan dua cara yaitu cara manual dan dengan menggunakan *software rapid miner*. Perhitungan manual untuk pembuatan model dengan menggunakan algoritma pohon keputusan C4.5 dilakukan dengan cara mencari nilai *gain* tertinggi dari setiap atribut, sedangkan perhitungan manual untuk model algoritma pohon keputusan C4.5.

a. Model Klasifikasi C4.5

Untuk dapat membuat pohon keputusan, langkah pertama adalah menghitung jumlah *class* yang terpilih dan tidak dari masing-masing *class* berdasarkan atribut yang telah ditentukan dengan menggunakan data *training*.

1. Kemudian menghitung *Entropy* (Total) menggunakan persamaan (2.1)

$$Entropy(total) = \left(-\frac{210}{766} * \log_2\left(\frac{210}{766}\right)\right) + \left(-\frac{556}{766} * \log_2\left(\frac{556}{766}\right)\right)$$

$$Entropy(Total) = 0.00291$$

2. Kemudian hitung masing-masing Gain berdasarkan atribut pada Tabel II. 6. *Candidat split* dan *rule* atribut algoritma C4.5

Sebagai contoh plafon pinjaman :

$$\leq 925208.380 = \frac{200}{766}$$

$$> 925208.380 = \frac{356}{766}$$

Untuk record plafond pinjaman ≤ 925208.380 terdiri dari 10 LANCAR dan 200 MACET, untuk plafond pinjaman > 925208.380 terdiri dari 200 LANCAR dan 356 MACET. Dapat dihitung *Entropy*nya sebagai berikut:

$$Entropy < 925208.380 = \left(-\frac{10}{788} * \log_2\left(\frac{10}{766}\right)\right) + \left(-\frac{200}{766} * \log_2\left(\frac{200}{766}\right)\right)$$

$$Entropy < 925208.380 = 0.27620$$

$$Entropy > 925208.380 = \left(-\frac{200}{766} * \log_2\left(\frac{200}{766}\right)\right) + \left(-\frac{356}{766} * \log_2\left(\frac{356}{766}\right)\right)$$

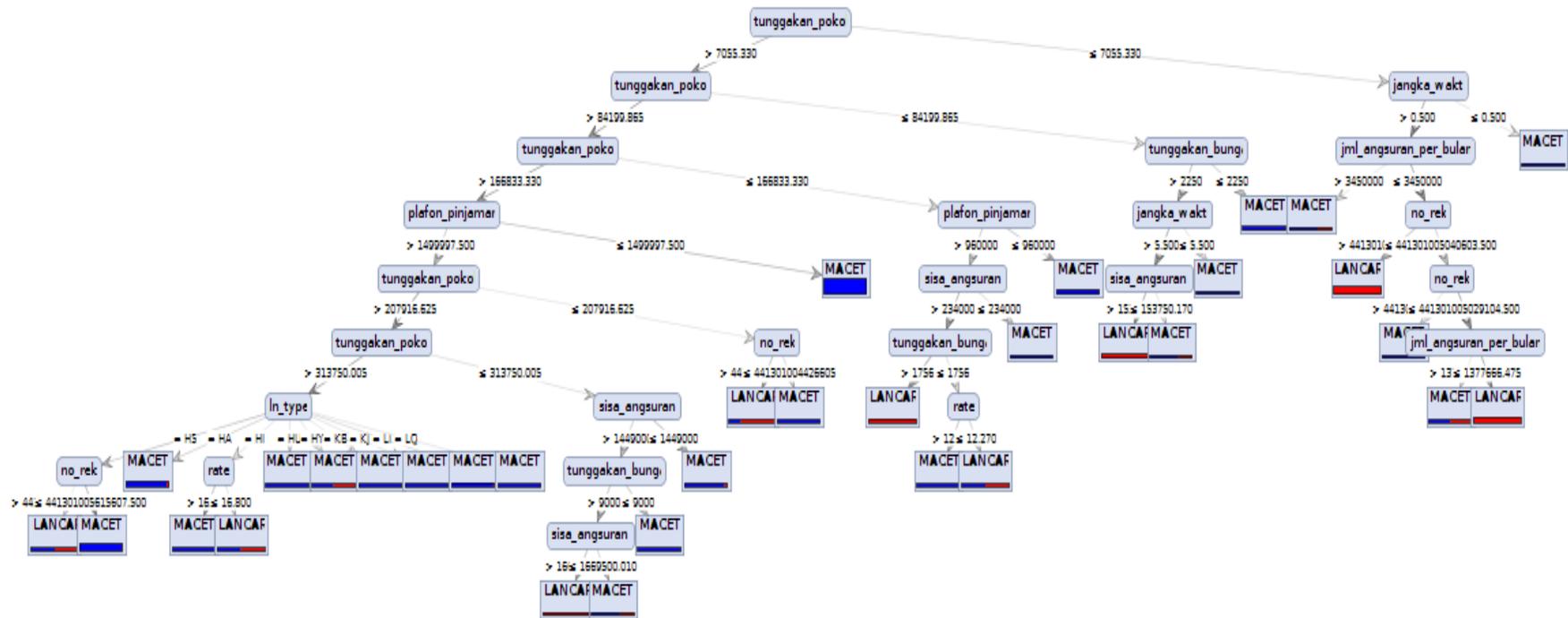
$$Entropy > 925208.380 = 0.94244$$

Selengkapnya akan disajikan dalam tabel contoh perhitungan *entropy* pada tabel dibawah ini :

Tabel. III. 1. Perhitungan *Entropy* dan *Gain* Algoritma C4.5

Node	Atribut	Jml Kasus (S)	Macet (Si)	Lancar (Si)	Entropy	Gain
1	Total	766	556	210	0.84736	
	Jenis Kelamin					0.00291
	Laki-Laki	304	210	94	0.89226	
	Perempuan	462	346	116	0.81299	
	Rate					0.43296
	10	4	3	1	0.81128	
	11	9	7	2	0.76420	
	12	49	24	25	0.99970	
	14	67	43	24	0.94119	
	19	228	164	64	0.85641	
	20	1	0	1	0.00000	
	24	1	0	1	0.00000	
	Plafon Pinjaman					0.08757
	≤ 925208.380	210	200	10	0.27620	
	>925208.380	556	356	200	0.94244	
	Jangka Waktu					0.07560
	1	10	6	4	0.97095	
	2	8	5	3	0.95443	
	3	24	14	10	0.97987	
	4	24	19	5	0.73828	
	5	47	29	18	0.96012	
	6	47	36	11	0.78499	
	7	22	21	1	0.26676	
	8	50	39	11	0.76017	
	10	84	70	14	0.65002	
	11	8	7	1	0.54356	
	12	127	84	43	0.92346	
	14	4	3	1	0.81128	
	15	7	7	0	0.00000	
	16	76	45	31	0.97538	
	20	73	50	23	0.89894	
	24	58	37	21	0.94439	
	25	2	1	1	1.00000	
	30	3	2	1	0.91830	
	32	3	2	1	0.91830	
	36	1	0	1	0.00000	
	38	1	1	0	0.00000	

Berikut pada Gambar III. 1. Pohon Keputusan Klasifikasi C4.5 kreditDi bawah ini pohon keputusan yang dihasilkan dari algoritma C4.5.



Gambar III. 1. Pohon Keputusan Klasifikasi C4.5 kredit

Pada **Gambar III. 1**. Pohon Keputusan Klasifikasi C4.5 dapat dijelaskan sebagai berikut: pada *root* *tunggakan_pokok* jika bernilai > 7055.330 dan *tunggakan_pokok* > 166833.330 dan *plafon_pinjaman* > 1831667 maka menghasilkan daun atau *class* = MACET. Untuk rincian lebih lengkap dari pohon keputusan pada diatas dapat dilihat aturan sebagai berikut:

- R1. *if* *tunggakan_pokok* > 7055.330 *and* *tunggakan_pokok* > 166833.330 *and* *plafon_pinjaman* > 1831667 *then* MACET
- R2. *if* *tunggakan_pokok* > 7055.330 *and* *tunggakan_pokok* > 166833.330 *and* *plafon_pinjaman* ≤ 1831667 *and* *no_rek* > 441301005637106.500 *and* *sisa_angsuran* > 1402125 *then* LANCAR
- R3. *if* *tunggakan_pokok* > 7055.330 *and* *tunggakan_pokok* > 166833.330 *and* *plafon_pinjaman* ≤ 1831667 *and* *no_rek* > 441301005637106.500 *and* *sisa_angsuran* ≤ 1402125 *then* MACET
- R4. *if* *tunggakan_pokok* > 7055.330 *and* *tunggakan_pokok* > 166833.330 *and* *plafon_pinjaman* ≤ 1831667 *and* *no_rek* ≤ 441301005637106.500 *then* MACET
- R5. *if* *tunggakan_pokok* > 7055.330 *and* *tunggakan_pokok* ≤ 166833.330 *and* *plafon_pinjaman* > 925208.380 *and* *jangka_waktu* > 7.500 *and* *jml_angsuran_per_bulan* > 12750 *and* *tunggakan_bunga* > 1756 *then* LANCAR
- R6. *if* *tunggakan_pokok* > 7055.330 *and* *tunggakan_pokok* ≤ 166833.330 *and* *plafon_pinjaman* > 925208.380 *and* *jangka_waktu* > 7.500 *and* *jml_angsuran_per_bulan* > 12750 *and* *tunggakan_bunga* ≤ 1756 *and* *rate* > 12.270 *then* MACET
- R7. *if* *tunggakan_pokok* > 7055.330 *and* *tunggakan_pokok* ≤ 166833.330 *and* *plafon_pinjaman* > 925208.380 *and* *jangka_waktu* > 7.500 *and* *jml_angsuran_per_bulan* > 12750 *and* *tunggakan_bunga* ≤ 1756 *and* *rate* ≤ 12.270 *then* LANCAR
- R8. *if* *tunggakan_pokok* > 7055.330 *and* *tunggakan_pokok* ≤ 166833.330 *and* *plafon_pinjaman* > 925208.380 *and* *jangka_waktu* > 7.500 *and* *jml_angsuran_per_bulan* ≤ 12750 *then* MACET.

- R9. *if* *tunggakan_pokok* > 7055.330 *and* *tunggakan_pokok* ≤ 166833.330 *and* *plafon_pinjaman* > 925208.380 *and* *jangka_waktu* ≤ 7.500 *then* MACET
- R10. *if* *tunggakan_pokok* > 7055.330 *and* *tunggakan_pokok* ≤ 166833.330 *and* *plafon_pinjaman* ≤ 925208.380 *then* MACET
- R11. *if* *tunggakan_pokok* ≤ 7055.330 *then* LANCAR.

b. Model klasifikasi C4.5 Dengan Algoritma Genetika dan *Bagging*

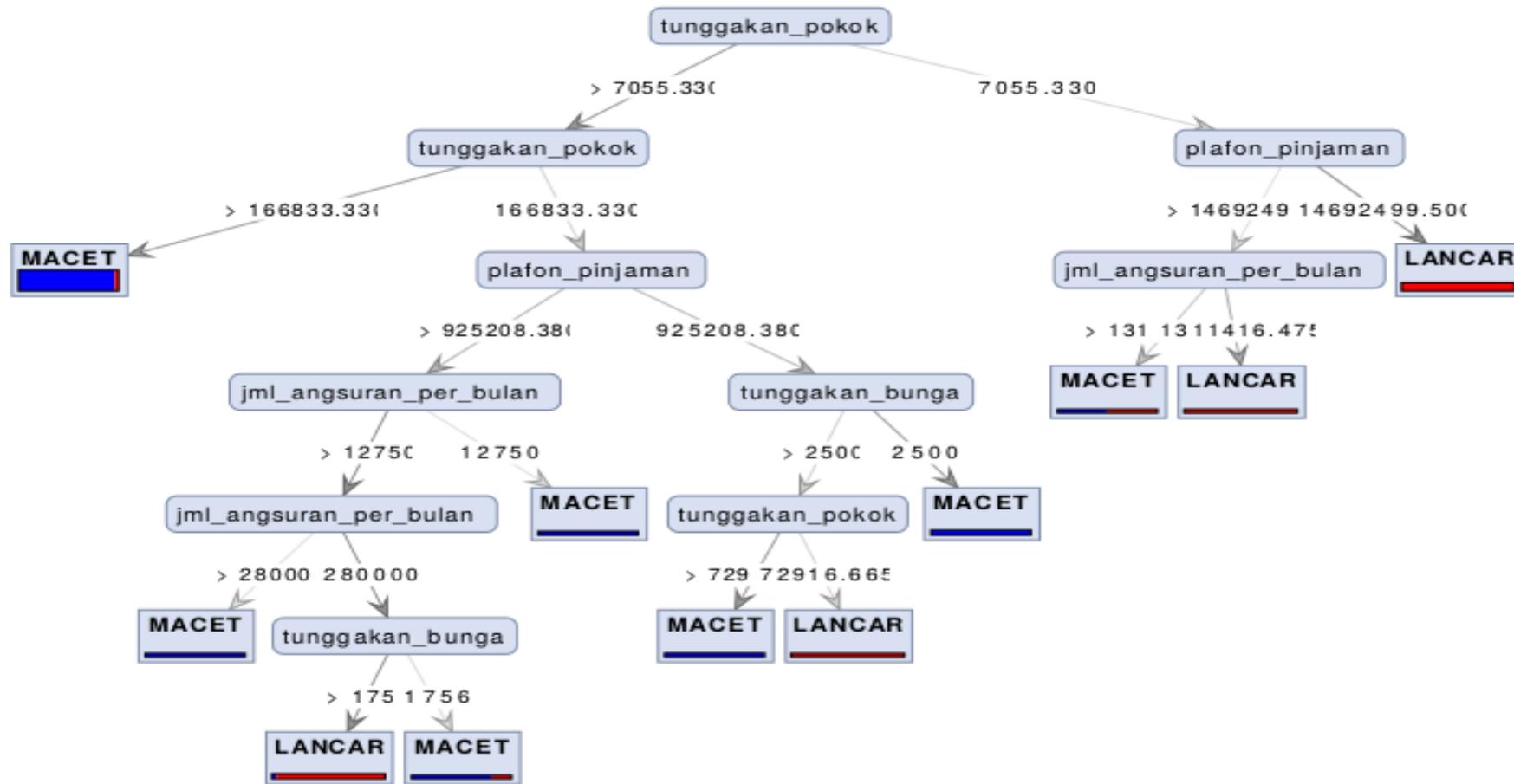
Proses yang dilakukan sama dengan proses pembuatan model klasifikasi C4.5 namun perbedaannya dari atribut yang digunakan, atribut yang digunakan merupakan atribut terpilih hasil seleksi atribut terbaik menggunakan Algoritma Genetika, untuk meningkatkan stabilitas dalam klasifikasi maka *bagging* diterapkan selain berguna untuk meningkatkan stabilitas klasifikasi, *bagging* juga berguna untuk mengurangi variansi data dan untuk menghindari *overfitting data*, untuk penjelasannya sebagai berikut

Menghitung *Entropy* (Total) menggunakan persamaan (2.1)

$$Entropy(total) = \left(-\frac{210}{766} * \log_2\left(\frac{210}{766}\right)\right) + \left(-\frac{556}{766} * \log_2\left(\frac{556}{766}\right)\right)$$

$$Entropy(Total) = 0.00291$$

Untuk mendapatkan *gain* tiap atribut, maka harus dihitung *entropy* seperti pada algoritma C4.5 sebelumnya diatas namun perbedaannya atribut yang di hitung hanya tunggakan pokok, plafon pinjaman, jumlah angsuran, tunggakan pokok, tunggakan bungan berikut pohon keputusan hasil klasifikasi C4.5 dengan Algoritma Genetika (AG) dan *bagging* :



Gambar III. 2. Pohon Keputusan Klasifikasi C4.5 Dengan GA dan *Bagging*

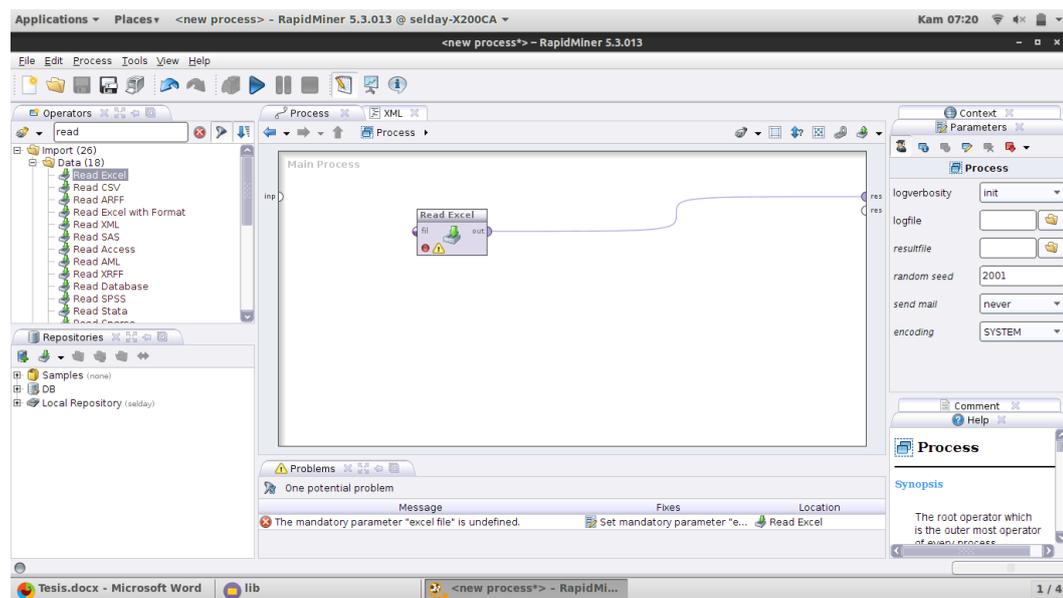
Dari **Gambar III. 2.** Pohon Keputusan Klasifikasi C4.5 Dengan GA dan *Bagging* Dapat dijelaskan sebagai berikut Jika *tunggakan_pokok* > 7055.330 dan *tunggakan_pokok* > 166833.330 Maka akan menghasilkan daun *class* MACET, selengkapnya dapat di jelaskan sebagai berikut :

- R1. *if* *tunggakan_pokok* > 7055.330 *and* *tunggakan_pokok* > 166833.330 *then* MACET
- R2. *if* *tunggakan_pokok* > 7055.330 *and* *tunggakan_pokok* ≤ 166833.330 *and* *plafon_pinjaman* > 925208.380 *and* *jangka_waktu* > 7.500 *and* *jml_angsuran_per_bulan* > 12750 *and* *tunggakan_bunga* > 1756 *then* LANCAR
- R3. *if* *tunggakan_pokok* > 7055.330 *and* *tunggakan_pokok* ≤ 166833.330 *and* *plafon_pinjaman* > 925208.380 *and* *jangka_waktu* > 7.500 *and* *jml_angsuran_per_bulan* > 12750 *and* *tunggakan_bunga* ≤ 1756 *then* MACET
- R4. *if* *tunggakan_pokok* > 7055.330 *and* *tunggakan_pokok* ≤ 166833.330 *and* *plafon_pinjaman* > 925208.380 *and* *jangka_waktu* > 7.500 *and* *jml_angsuran_per_bulan* ≤ 12750 *then* MACET
- R5. *if* *tunggakan_pokok* > 7055.330 *and* *tunggakan_pokok* ≤ 166833.330 *and* *plafon_pinjaman* > 925208.380 *and* *jangka_waktu* ≤ 7.500 *then* MACET
- R6. *if* *tunggakan_pokok* > 7055.330 *and* *tunggakan_pokok* ≤ 166833.330 *and* *plafon_pinjaman* ≤ 925208.380 *and* *tunggakan_bunga* > 2500 *and* *tunggakan_pokok* > 72916.665 *then* MACET
- R7. *if* *tunggakan_pokok* > 7055.330 *and* *tunggakan_pokok* ≤ 166833.330 *and* *plafon_pinjaman* ≤ 925208.380 *and* *tunggakan_bunga* > 2500 *and* *tunggakan_pokok* ≤ 72916.665 *then* LANCAR
- R8. *if* *tunggakan_pokok* > 7055.330 *and* *tunggakan_pokok* ≤ 166833.330 *and* *plafon_pinjaman* ≤ 925208.380 *and* *tunggakan_bunga* ≤ 2500 *then* MACET
- R9. *if* *tunggakan_pokok* ≤ 7055.330 *and* *plafon_pinjaman* > 14692499.500 *and* *jangka_waktu* > 22.500 *then* LANCAR
- R10. *if* *tunggakan_pokok* ≤ 7055.330 *and* *plafon_pinjaman* > 14692499.500 *and* *jangka_waktu* ≤ 22.500 *then* MACET

R11. *if* tunggakan_pokok \leq 7055.330 and plafon_pinjaman \leq 14692499.500
then LANCAR.

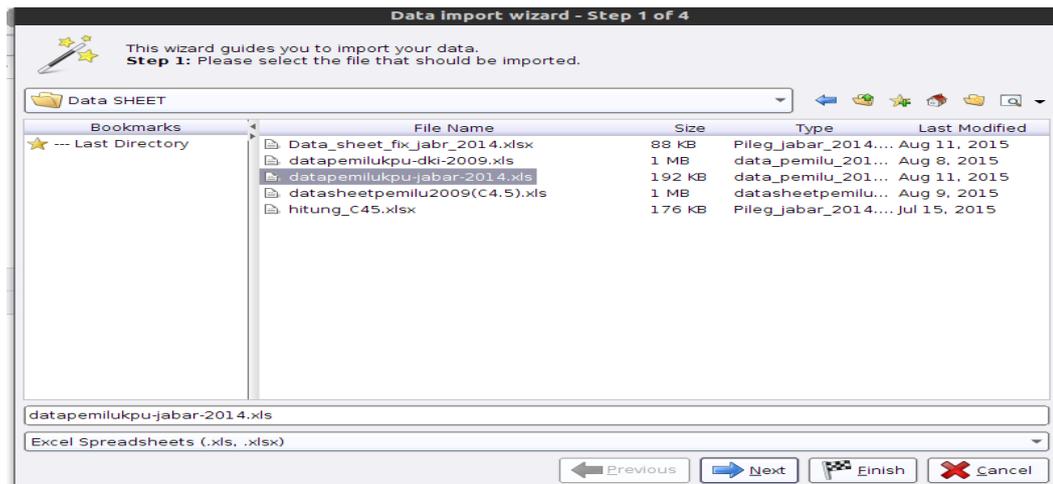
Sedangkan tahap *modelling* menggunakan *software rapid miner* dilakukan dengan cara sebagai berikut:

- a. Untuk dapat membentuk model diperlukan *data training*, *software rapid miner* mampu membaca data dengan banyak format, yang paling umum digunakan diantaranya .csv, .xls dan .mdb. Data yang digunakan pada penelitian ini adalah data dari salah satu bank BUMN terbesar di Indonesia Data awal dalam bentuk format .pdf kemudian ditransformasi menjadi format .xls.
- b. Jalankan *software rapid miner*, masukan *data training* yang telah disiapkan pada tahap *preparation*. Pada tab operators pilih Import – Data – *read Excell* kemudian *drag file* ke area *main process*. Pada area parameters klik *Import Configuration Wizard*.



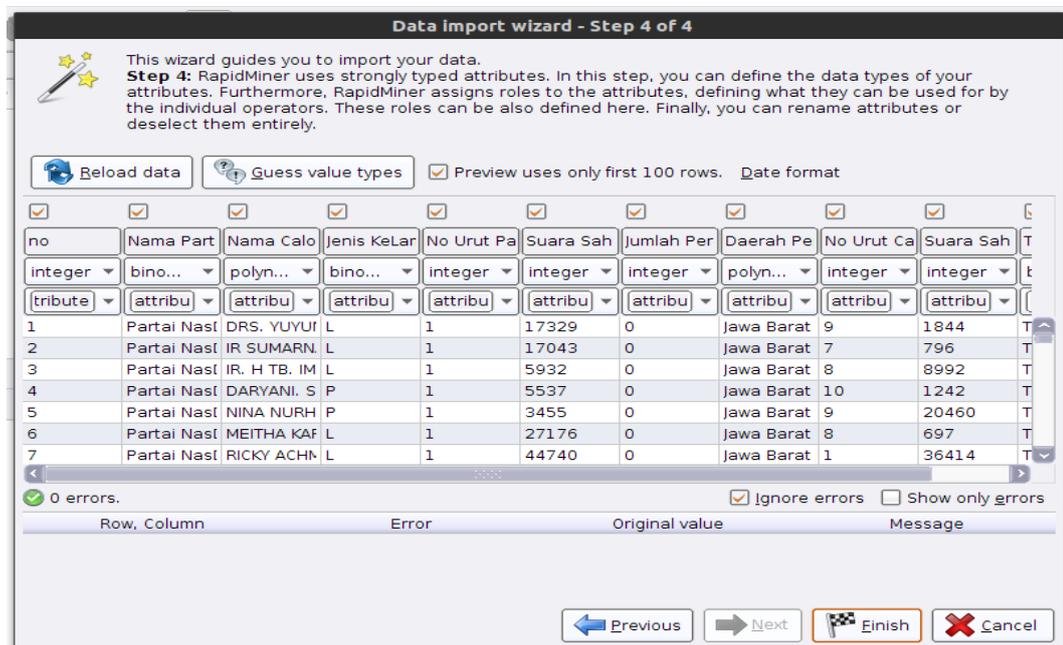
Gambar III. 3. Preparing

- c. Tampil jendela *Import Configuration Wizard*, pada langkah ini tentukan data berformat .xls yang akan digunakan sebagai *data training* kemudian klik tombol *next*.



Gambar 3.4 Import File pada Tahap Modelling Rapid Miner

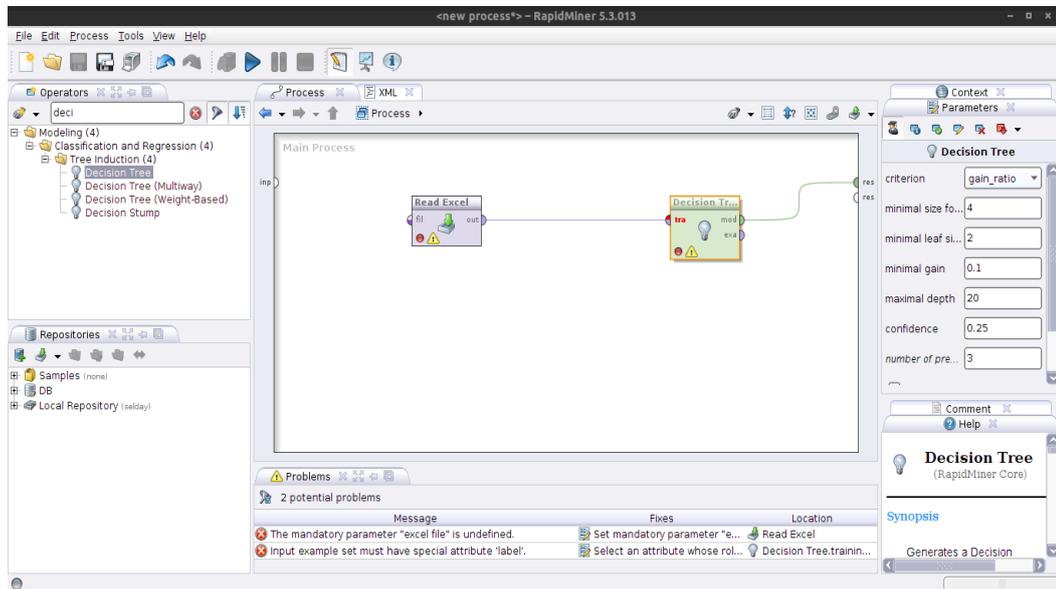
- d. Langkah ke dua yaitu memilih *set* yang akan digunakan sebagai data *training*, langkah ke tiga menentukan anotasi dari setiap *tupel*, langkah ini dapat diabaikan. Kemudian klik *Next*. Langkah ke empat yaitu menentukan nama atribut dan tipe data dari masing-masing atribut. Pada atribut *Class* ubah tipe data menjadi label. Lalu klik *finish*.



Gambar III. 4 Menentukan *set* pada Tahap Modelling Rapid Miner

- e. Setelah data *training* disiapkan, proses pembuatan model dapat dilakukan dengan cara pilih tab *Operators - Modelling – Classification and Regression*

– *Tree Induction – Decision Tree*. Geser *Decision Tree* ke area *main process* dan hubungkan seperti terlihat pada gambar 3.4.

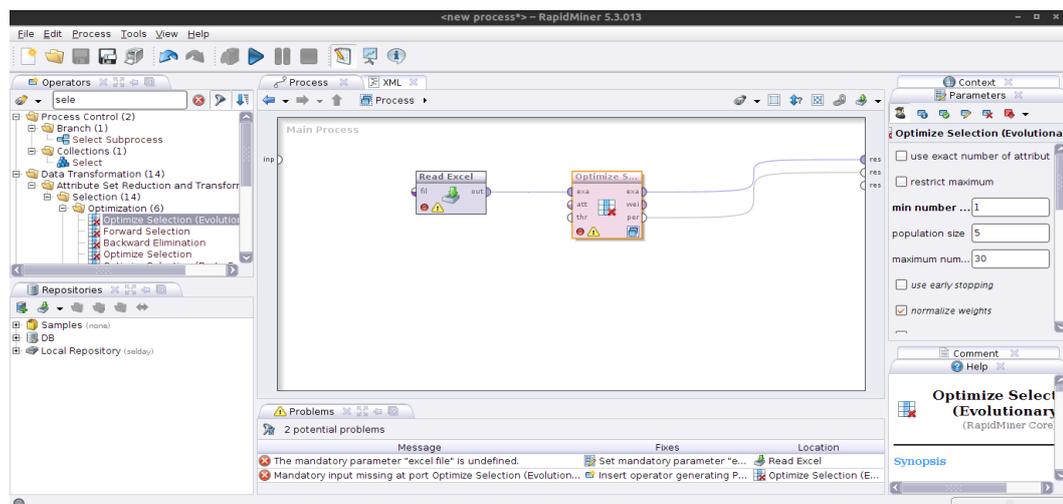


Gambar III. 5 Penggunaan *Decision Tree* pada Tahap *Modelling Rapid Miner*

Karena proses perhitungan manual dilakukan dengan menggunakan nilai gain tertinggi, maka pada tab *Decision Tree – Criterion* diganti dengan *Information gain*. Kemudian jalankan dengan memilih menu *Process – Run*, maka akan diperoleh hasil (*Result*).

Sedangkan untuk pemodelan pohon keputusan C4.5 dengan genetika algoritma dan bagging dilakukan sebagai berikut:

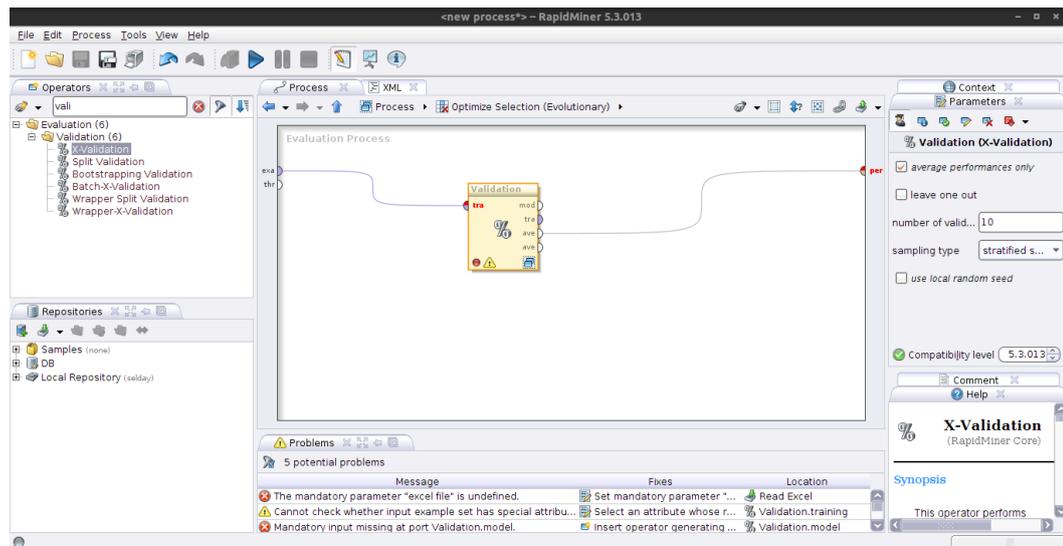
- a. Berdasarkan data *training* yang sama seperti di atas lakukan langkah



Gambar III. 6 Pemilihan *Optimize Selection*

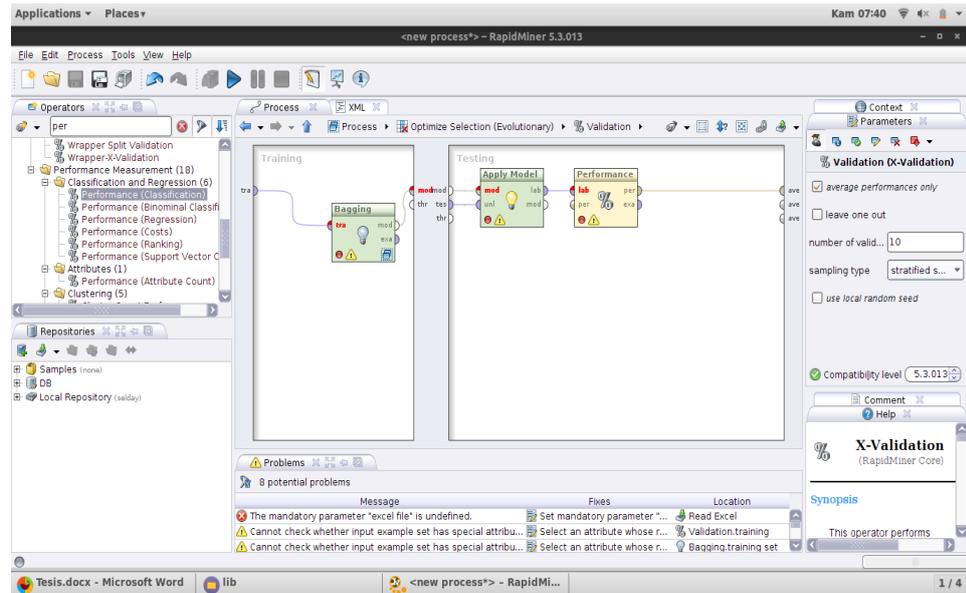
Pada Tab *Operator*, pilih *Data Transformation* lalu *Attribute Set Reduction and Transformation* kemudian *Selection – Optimization – Optimize Selection (evolutionary)*.

- b. Pada tahap *evaluation process* diawali dari double klik *Optimize Selection* pada icon kotak warna biru kemudian pilih *Evaluation - validation – cross Validation* simpan di main process dari *Validation*



Gambar III. 7. Pemilihan *Validation*

Lalu pilih model algoritma data mining yang akan digunakan, pada penelitian ini adalah model algoritma klasifikasi *decision tree* yang dimasukan kedalam *bagging (Meta Modeling – bagging)* lalu drag ke bagian Training. Sedangkan pada bagian Testing pilih Apply Model (*Modeling – Model Application – Confidences – Apply Model*) dan Performance (*Evaluation – Performance Measurement – Clustering - Performance*)



Gambar III. 8 Pilih Model Algoritma dan Performance

5. Tahap Evaluation

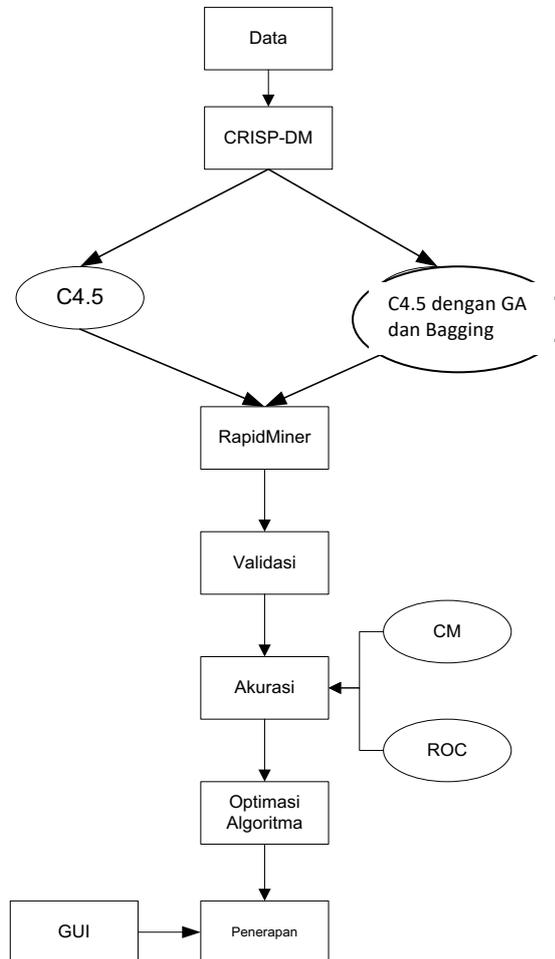
Pada tahap evaluation dilakukan evaluasi kualitas dan efektivitasnya dengan memasukan data testing. Evaluasi dan validasi menggunakan metode *Confusion Matrix* dan ROC , penjelasan lebih lengkap mengenai tahap evaluation terdapat pada bab IV

6. Tahap Deployment

Pada tahapan deployment, dilakukan penerapan model algoritma klasifikasi C4.5 dengan Algoritma Genetika (AG) dan *bagging* untuk analisis kelayakan pemberian kredit.

3.2. Langkah-langkah Penelitian

Langkah penelitian yang dilakukan dapat digambarkan pada bagan sebagai berikut:



Gambar 3.10 Bagan langkah penelitian

Berdasarkan gambar langkah-langkah pemikiran di atas, dapat dijelaskan sebagai berikut:

1. Data mentah direduksi terlebih dahulu kemudian dijadikan data *training* dan *testing* menggunakan *cross validation* seperti pada Tabel II. 2. Ilustrasi 10 *fold Cross Validation* pada penelitian ini 10 *fold cross validation* yang digunakan
2. Dengan menggunakan metode *CRISP-DM* kemudian data *training* diuji dengan algoritma klasifikasi pohon keputusan C4.5 dan C4.5 dengan genetic algoritma dan *bagging*.

3. *Tools* yang digunakan untuk menguji kedua metode tersebut menggunakan *rapid miner*.
4. Hasil dari masing-masing pengujian tersebut kemudian divalidasi menggunakan *tools* yang sama.
5. Berdasarkan evaluasi pengujian akan diperoleh nilai akurasi, dimana evaluasi yang digunakan menggunakan dua *measurement* yaitu berdasar evaluasi *Confusion Matrix (CM)* dan Evaluasi *ROC*.
6. Setelah diperoleh evaluasi dari optimasi algoritma yang diharapkan, kemudian akan diterapkan pada pembuatan *Graphical User Interface (GUI)*.
7. Dengan menggunakan data baru, kemudian *GUI* yang telah dibuat akan diujicoba kepada *user* dan dievaluasi kembali untuk mendapatkan objektivitas penelitian dan kesempurnaan program.

3.3. Uji Ketahanan *Graphical User Interface (GUI)*

Secara sederhana, *GUI* adalah suatu media *virtual* yang dapat membuat pengguna memberikan perintah tertentu pada komputer tanpa menyetik perintah tersebut, namun menggunakan gambar yang tersedia. Pengguna tidak mengetikkan perintah seperti pada komputer dengan *shell* atau teks. Dengan *GUI*, perintah dapat dikonversi menjadi ikon dalam layar monitor yang dapat diklik untuk memulai fungsinya. Uji ketahanan *GUI* merupakan salah satu bagian dari pengujian perangkat lunak. Pengujian perangkat lunak adalah elemen kritis dari jaminan kualitas perangkat lunak dan merepresentasikan kajian pokok dari spesifikasi, desain dan pengkodean.

3.3.1. Kuesioner Metode SQA (*Software Quality Assurance*).

Untuk memiliki perangkat lunak yang kualitas standar minimal, maka diterapkan salah satu metode untuk pengukuran kualitas perangkat lunak secara kuantitatif yaitu metode *SQA (Software Quality Assurance)*. Permasalahan mengukur kualitas perangkat lunak bukanlah hal yang sederhana, karena hal ini dapat dipengaruhi oleh sudut pandang penggunanya. Kualitas perangkat lunak dapat dilihat dari sudut pandang proses pengembangan perangkat lunak (*process*) dan hasil produk yang dihasilkan (*product*), (Wahono, 2006). Terdapat 5 (lima)

responden yang akan menguji ketahanan *GUI* dengan cara memberikan nilai kuesioner yang terdiri dari satu orang anggota analis kredit, dua orang dosen ilmu ekonomi, dan satu *programmer* (Bagian IT). Bentuk kuesioner yang diberikan kepada lima orang responden tersebut dapat dilihat pada Lampiran

3.3.2. Cara Hasil Evaluasi *GUI*

Pada penelitian ini dalam mengevaluasi *GUI*, menggunakan pendekatan sebagai berikut:

$$\begin{aligned} \text{Skor masing-masing responden} = & \quad <SkorAuditability> * 0.10 + \\ & \quad <SkorAccuracy> * 0.10 + \\ & \quad <SkorCompleteness> * 0.15 + \\ & \quad <SkorErrorTolerance> * 0.10 + \\ & \quad <SkorExecutionEfficiency> * 0.10 + \\ & \quad <SkorOperability> * 0.15 + \\ & \quad <SkorSimplicity> * 0.15 + \\ & \quad <SkorTraining> * 0.15 \end{aligned}$$

3.3.3. Evaluasi Berdasarkan Kriteria/Rata-Rata Skor Responden

Dalam mengevaluasi berdasarkan kriteria pada penelitian ini menggunakan persamaan sebagai berikut:

$$\text{Rata-rata Responden} : \frac{\text{Jumlah Skor Responden}}{\text{Jumlah Responden}} \quad (3.2)$$

3.4. Jadwal Penelitian

Beberapa tahapan yang dilakukan pada penelitian ini, sebagai berikut:

1. Identifikasi masalah dan analisa kebutuhan
Dalam tahapan ini dilakukan pemahaman dan penggalian informasi yang berkaitan dengan permasalahan yang berhubungan dengan analisa kelayakan pemberian kredit dan permasalahannya.
2. Pengumpulan data
Mengumpulkan data-data yang dibutuhkan untuk penelitian yang bersumber dari objek riset. Data yang didapat kemudian diolah sehingga menghasilkan data bersih yang dibutuhkan.
3. Eksperimen
Tahap ini dimulai menentukan model yang akan digunakan, memasukan data *training* dan *testing* ke dalam model dan mengujinya menggunakan *tools Rapid miner*
4. Implementasi
Menerapkan model pohon keputusan dan algoritma klasifikasi C4.5 dengan Algoritma Genetika (AG) dan bagging ke dalam sistem untuk analisis kelayakan pemberian kredit digunakan oleh pengguna.
5. Evaluasi
Mengukur apakah model yang telah dikembangkan berhasil atau tidak, evaluasi digunakan untuk mengukur keakuratan hasil yang dicapai oleh model.
6. Penulisan.
Penulisan kedalam bentuk laporan penelitian (tesis) dilakukan secara bersamaan dengan melakukan penelitian serta tahapan lainnya.

BAB 4

HASIL PENELITIAN DAN PEMBAHASAN

4. 1. Pengukuran Penelitian

Setelah dijelaskan perhitungan manual pada bab sebelumnya selanjutnya melakukan pengukuran dari model yang diajukan, pengukuran dilakukan dengan menggunakan tools, salah satu tools yang digunakan adalah rapid miner.

4. 1.1. Hasil Penelitian

Hasil penelitian bertujuan membandingkan model terpilih C4.5 dengan algoritma C4.5 yang dioptimalkan dalam pemilihan atribut dengan Algoritma genetika dan *bagging*. Data akan dianalisa kemudian model terbaik yang akan digunakan dalam perancangan GUI.

1. Hasil Pengujian dengan algoritma C4.5

Hasil pengujian menghasilkan nilai accuracy dan nilai AUC(*Area Under Curva*)

a. Evaluasi dengan *Confusion Matrix*

Model *confusion matrix* akan menghasilkan matrix yang terdiri dari *true positif* atau *true negative* setelah data *testing* di masukan kedalam model confusion matrix hasilnya akan terlihat seperti gambar dibawah:

accuracy: 93.47% +/- 2.86% (mikro: 93.47%)			
	true MACET	true LANCAR	class precision
pred. MACET	538	32	94.39%
pred. LANCAR	18	178	90.82%
class recall	96.76%	84.76%	

Gambar IV. 1. *Confusion Matrix* Algoritma Klasifikasi C4.5

Berdasar tabel diatas dari 766 *tuple* dengan rincian jumlah *True Positive (TP)* 538, *False Negative (FN)* 32, *False Positive (FP)* adalah 18 dan *True Negative (TN)* 178. Dari data tersebut maka dapat dihitung nilai *accuracy*, *sensitivity*, *specifity* dan *NPV*. Data hasil olahan dapat dilihat pada tabel di bawah:

Tabel IV. 1. Nilai *Accuracy*, *Sensitivity*, *Specificity*, *PPV* dan *NPV*

	Nilai
<i>Accuracy</i>	0,934
<i>Sensitivity</i>	0,943
<i>Specificity</i>	0,908
<i>PPV</i>	0,967
<i>NPV</i>	0,847

Secara manual data tersebut dihitung menggunakan persamaan (2.4), sebagai berikut:

$$Accuracy = \frac{538TP + 178TN}{538TP + 178TN + 18fp + 32FN} = \mathbf{0,934}$$

$$Sensitivity = \frac{538}{538 + 32} = \mathbf{0,943}$$

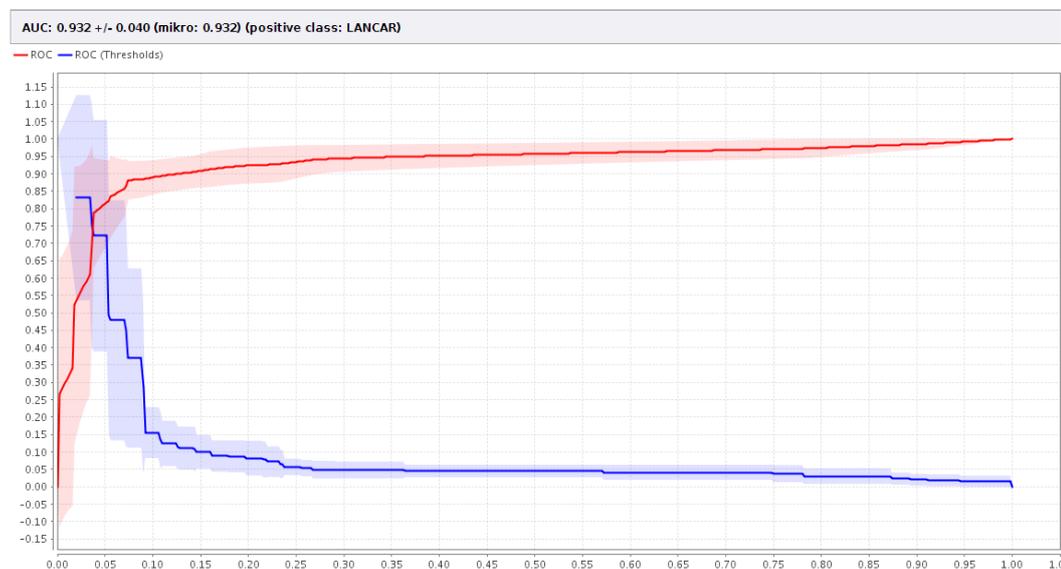
$$Specificity = \frac{178}{178 + 18} = \mathbf{0,908}$$

$$PPV = \frac{538}{538 + 18} = \mathbf{0,967}$$

$$NPV = \frac{178}{178 + 32} = \mathbf{0,847}$$

b. Evaluasi Dengan ROC Curva

Hasil pengujian terhadap data *testing* untuk algoritma klasifikasi C4.5 terhadap nilai ROC diketahui pada gambar di bawah ini:



Gambar IV. 2 ROC Curve Algoritma C4.5

2. Hasil pengujian C4.5 menggunakan Algoritma Genetika (AG) dan *Bagging*. Hasil pengujian yang dihasilkan adalah nilai *Accuracy* dan nilai *AUC (Area Under Curve)*.

a. Evaluasi dengan *Confusion Matrix*

Model *confusion matrix* akan menghasilkan *matrix* yang terdiri dari *true positif* atau *true negative* setelah data testing di masukan kedalam model *confusion matrix* hasilnya akan terlihat seperti gambar dibawah:

accuracy: 96.34% +/- 2.17% (mikro: 96.34%)			
	true MACET	true LANCAR	class precision
pred. MACET	548	20	96.48%
pred. LANCAR	8	190	95.96%
class recall	98.56%	90.48%	

Gambar IV. 3. Confusion matrix Algoritma C4.5 dengan genetic Algoritma dan Bagging

Berdasar tabel diatas dari 1066 *tuple* dengan rincian jumlah *True Positive (TP)* 548, *False Negative (FN)* 20, *False Positive (FP)* adalah 8 dan *True Negative (TN)* 190. Dari data tersebut maka dapat dihitung nilai *accuracy*, *sensitivity*, *specifity* dan *NPV*. Data hasil olahan dapat dilihat pada tabel di bawah:

Tabel IV. 2. Nilai *Accuracy*, *Sensitivity*, *Specificity*, *PPV* dan *NPV*

	Nilai
<i>Accuracy</i>	0,963
<i>Sensitivity</i>	0,964
<i>Specificity</i>	0,959
<i>PPV</i>	0,964
<i>NPV</i>	0,904

Secara manual data tersebut dihitung menggunakan persamaan sebagai berikut:

$$Accuracy = \frac{548TP + 190TN}{548TP + 190TN + 8fp + 20FN} = 0,963$$

$$Sensitivity = \frac{548TP}{548TP + 20FN} = 0,964$$

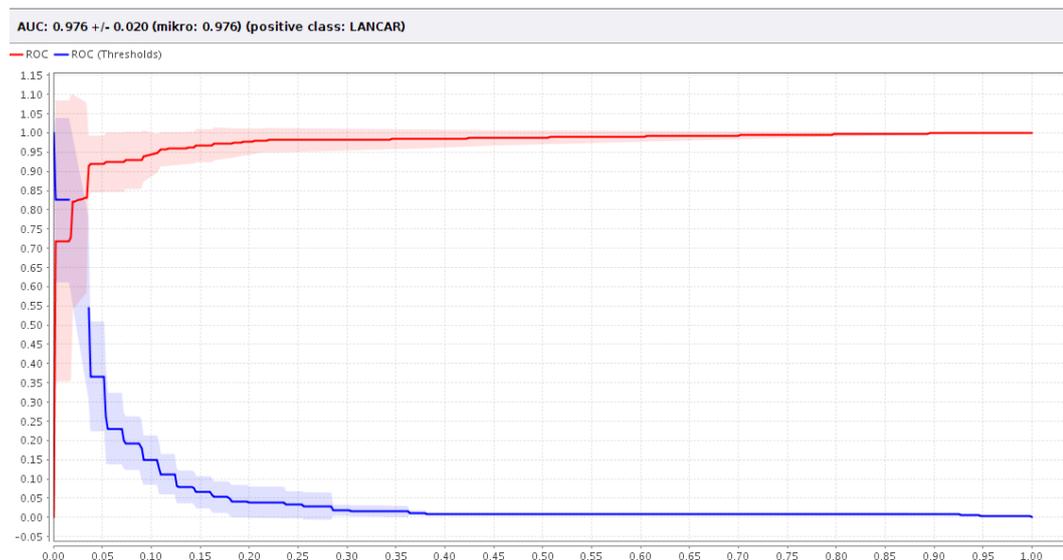
$$Specificity = \frac{190TN}{190TN + 8FP} = 0,959$$

$$PPV = \frac{548TP}{548TP + 20FN} = 0,964$$

$$NPV = \frac{190TN}{190TN + 20FN} = 0,904$$

b. Evaluasi Dengan ROC Curva

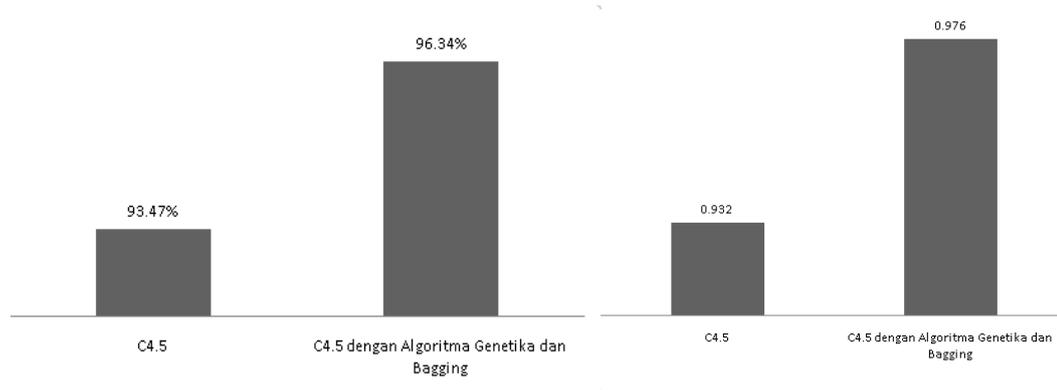
Hasil pengujian terhadap data *testing* untuk algoritma klasifikasi C4.5 terhadap nilai ROC diketahui pada gambar di bawah ini:



Gambar IV. 4. ROC Curva C4.5 dengan Genetik Algoritma dan *Bagging*

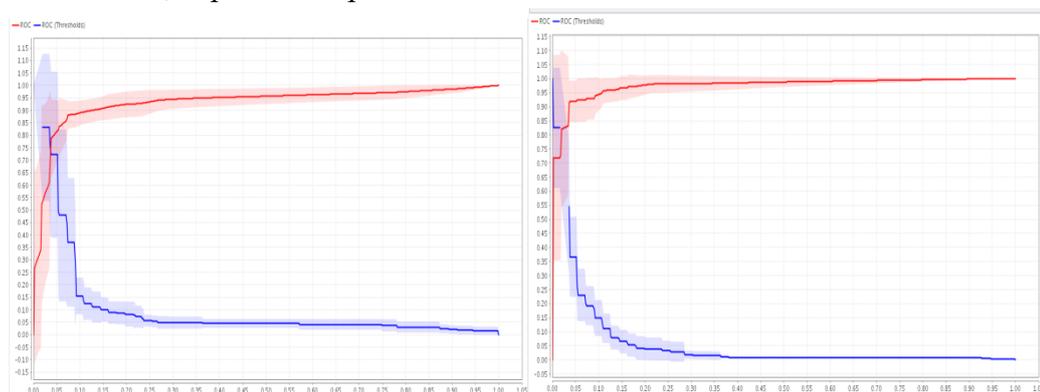
4. 2. Analisis Evaluasi dan Validasi Model

Pada bagian ini akan dibandingkan berdasarkan hasil pengujian menggunakan *confusion matrix* maupun *ROC curve* diatas terbukti bahwa, algoritma C4.5 dengan Algoritma Genetika (AG) dan *Bagging* mampu meningkatkan akurasi hasil klasifikasi C4.5. sedangkan nilai akurasi untuk C4.5 itu sendiri adalah sebesar 93,47% dan nilai akurasi C4.5 dengan Algoritma Genetika dan *Bagging* adalah sebesar 96,34 % dengan selisih akurasi 2,87% dapat dilihat pada gambar dibawah ini:



Gambar IV. 5. Perbedaan Akurasi dan AUC Algoritma C4.5 dan Algoritma C4.5 dengan Algoritma Genetika dan Bagging

Untuk evaluasi menggunakan ROC *curve* sehingga menghasilkan nilai AUC (*Area Under Curve*) untuk model algoritma klasifikasi C4.5 menghasilkan nilai 0.932 dengan nilai diagnosa *Excelent Clasificatian*, sedangkan untuk algoritma klasifikasi C4.5 dengan Algoritma Genetika dan Bagging menghasilkan nilai 0.976 dengan nilai diagnosa *Excelent Clasificatian*, dan selisih nilai keduanya sebesar 0.044, dapat dilihat pada Gambar



Gambar IV. 6. ROC Curve Algoritma klasifikasi C4.5 dan Algoritma klasifikasi C4.5 dengan Algoritma Genetika dan Bagging

Dari hasil analisa algoritma klasifikasi diatas dengan demikian C4.5 dengan Algoritma Genetika (AG) dan bagging mampu memberikan solusi terhadap permasalahan analisa kelayakan pemberian kredit.

BAB 5

PENUTUP

5.1. Kesimpulan

Dari hasil penelitian untuk akurasi algoritma klasifikasi C4.5 sebesar 93,47%, sedangkan untuk akurasi algoritma klasifikasi dengan GA dan *Bagging* sebesar 96,36%, sehingga didapat selisih peningkatan akurasi sebesar 2,87%. Hasil evaluasi keduanya menggunakan Curva ROC yaitu, algoritma klasifikasi C4.5 bernilai 0,932 dengan tingkat diagnosa *excellent clasification*, sedangkan untuk algoritma klasifikasi C4.5 dengan Algoritma Genetika dan *Bagging* senilai 0,976 dengan tingkat diagnosa *excellent clasification* maka didapatlah selisih nilai sekitar nilai AUC 0,044. Dapat disimpulkan bahwa penggunaan algoritma genetika dan bagging pada algoritma klasifikasi dapat meningkatkan akurasi.

5.2. Saran

Berdasarkan pengujian dan kesimpulan yang telah dilakukan maka ada beberapa saran dalam penelitian ini diantaranya:

1. Menggunakan algoritma klasifikasi selain C4.5 supaya bisa terlihat mana algoritma yang paling baik.
2. Menggunakan metode *esembe* lain seperti *AdaBost*
3. Menggunakan algoritma optimasi lain seperti *Ant Colony*, *bee Colony* dan lain sebagainya.
4. Melakukan komparasi dengan menggunakan pengujian metode yang sama untuk dataset *public* sebagai data sekunder dan data hasil riset sebagai data primer.

DAFTAR ISI

- Abbelan, J., & Masegosa, A. R. (2012). *Bagging schemes on the presence of class noise in classification. Expert Systems with Applications* , 6827–6837.
- Akbilgic, O. (2015). A New Supervised Classification of Credit Approval Data via the Hybridized RBF Neural Network Model Using Information Complexity. *Data Science, Learning by Latent Structures, and Knowledge Discovery, Studies in Classification, Data Analysis, and Knowledge Organization* , 13-27.
- Breiman, L. (1996). Bagging Predictor. *Machine Learning* , 123-140.
- cho, y. j., Lee, H., & Jun, C. H. (2011). Optimization of Decision Tree for clasification Using Particel Swarm. *IEMS* , 272-278.
- Culp, M., Michaelidis, G., & Johnson, K. (2011). *On Adaptive Regularization Methods in Boosting. Journal of Computational and Graphical Statistics* , 937-955.
- Dawson, C. W. (2009). *Project in Computing and information system* . Edinburgh: Pearson.
- Han, J., & Kamber, M. (2006). *Data Mining Cancept And Teknik*. Oxford: Elsvier.
- Kusrini, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- Kusumadewi, S. (2003). *Artificial Intelligence*. Yokyakarta: Graha Ilmu.
- Larose, D. T. (2005). *Discovering Knowledge In Data : An Introduction to data Mining*. New Jersey: John Wiley And Son.
- Liu, X., & Huang, J. (2012). *Genetic Algorithm-based Feature Selection method for Credit Risk Analysis. 2nd International Conference on Computer Science and Network Technology* (pp. 2233-2266). CHANGCHUN: IEEE.
- Machairas, V., Tsangrassoulis, A., & Axarli, K. (2014). *Algorithms for optimization of building design: A review. Renewable and Sustainable Energy Reviews* , 101-112.
- Maimon, O., & Rokach, L. (2010). *Data Mining And Knowlegde Dicoverly Handbook*. New York: Springer.
- Mandala, I. G., Nawangpalupi, C. B., & Praktikto, F. R. (2012). *Assessing Credit Risk: an Application of Data Mining in a Rural Bank. Procedia Economics and Finance 4* (pp. 406 – 412). Procedia Economic and Finance.
- Masripah, S. (2011). *Algoritma Klasifikasi C4.5 Berbasis Particle Swarm Optimization Untuk Evaluasi Penentuan Kelayakan Pemberian Kredit Koperasi Syariah*. Jakarta: -.
- Oreski, S., & Oreski, G. (2013). *Genetic algorithm-based heuristic for feature selection in credit risk assessment. Expert Systems with Applications* , 1-12.

- Rafaeilzadeh, P., Tang, L., & Liu, H. (2009). *Cross-Validation*. In L. Liu, & M. T. Ozsu, *Encyclopedia of Database Systems* (pp. 532-538). Arizona: Springer US.
- Rao, M. V., & Singh, Y. P. (2013). *Decision Tree Induction For Financial Fraud Detection Using Ensemble Learning Techniques*. *Proceeding of the International Conference on Artificial Intelligence in Computer Science and ICT* (pp. 321-328). Langkawi: WorldConferences.net.
- Sarkar, B. K., Sana, S. S., & Caudhuri, K. (2011). *Selecting informative rules with parallel genetic algorithm in classification problem*. *Applied Mathematics and Computation* , 3247–3264.
- Susanto, S., & Suryadi, D. (2010). *Pengantar data mining menggali pengetahuan dari bongkahan data*. Bandung: Andi.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools And technique*. Burlington: Elsevier Inc.
- Yu, L., Chen, G., Koronios, A., Zhu, S., & Guo, X. (2007). *Application and Comparison of Classification Techniques in Controlling Credit Risk*. *Recent Advances in Data Mining of Enterprise Data* , 111-145.
- Zhang, J. L., & Härdle, W. K. (2010). *The Bayesian Additive Classification Tree applied to credit risk modelling*. *Computational Statistics and Data Analysis* , 1197–1205.
- Zhang, Z., & Yang, P. (2008). *An Ensemble of Classifiers with Genetic Algorithm Based Feature Selection*. *IEEE Intelligent Informatics Bulletin* , 18-24.
- Zurada, J. (2010). *Could Decision Trees Improve the Classification Accuracy and Interpretability of Loan Granting Decisions?* *Proceedings of the 43rd Hawaii International Conference on System Sciences* (pp. 1-9). Hawaii: IEEE.