

APPLICATION OF DATA MINING ALGORITHM TO RECIPIENT OF MOTORCYCLE INSTALLMENT

Harry Dhika¹, Fitriana Destiawati²

^{1,2}Teknik Informatika Fakultas Teknik, Matematika dan IPA, Universitas Indraprasta PGRI
Jl Nangka No. 58B, Jakarta, DKI Jakarta 12530
dhikatr@yahoo.com¹⁾, honeyzone86@gmail.com²⁾

ABSTRACT

The study was conducted in the subsidiaries that provide services of finance related to the purchase of a motorcycle on credit. At the time of applying, consumers enter their personal data. Based on the personal data, it will be known whether the consumer credit data is approved or rejected. From 224 consumer data obtained, it is known that the number of consumers whose applications are approved is 87% or about 217 consumers and consumers whose application is rejected is 16% or as much as 6 consumers. Acceptance of motorcycle financing on credit by using the method of applying the algorithm through CRIS-P DM is the industry standard in the processing of data mining. The algorithm used in the decision making is the algorithm C4.5. The results obtained previously, the level of accuracy is measured with the Confusion Matrix and Receiver Operating characteristic (ROC). Evaluation of the Confusion Matrix is intended to seek the value of accuracy, precision value, and the value of recall data. While the Receiver Operating Characteristic (ROC) is used to find data tables and comparison Area Under Curve (AUC)

Keywords: motorcycle financing credit, CRISP DM, data mining, algorithms C4.5, Confusion Matrix, ROC

ABSTRAK

Penelitian dilakukan pada anak perusahaan yang memberikan pelayanan pembiayaan sepeda motor secara kredit. Pada saat konsumen mengajukan permohonan, konsumen akan memasukan data pribadi mereka. Berdasarkan data pribadi tersebut, akan diketahui apakah data kredit konsumen tersebut disetujui atau ditolak. Dari 224 data konsumen yang diperoleh, konsumen yang pembiayaan kredit motornya disetujui adalah 87% atau sekitar 217 konsumen dan konsumen yang ditolak sebesar 16% atau sebanyak 6 konsumen. Penerimaan pembiayaan sepeda motor secara kredit dengan menggunakan metode penerapan algoritma melalui proses CRIS-P DM merupakan standar industri dalam pemrosesan data mining. Algoritma yang digunakan dalam menentukan keputusan adalah algoritma C4.5. Pada hasil yang diperoleh sebelumnya, tingkat akurasi diukur dengan Confusion Matrix dan Receiver Operating Characteristic (ROC). Evaluasi dalam Confusion Matrix ditujukan untuk mencari nilai accuracy, nilai precision, dan nilai recall data. Sedangkan, Receiver Operating Characteristic (ROC) digunakan untuk mencari data dengan tabel dan perbandingan Area Under Curve (AUC)

Kata kunci: pembiayaan kredit sepeda motor, CRIS-P DM, data mining, algoritma C4.5, Confusion Matrix, ROC

PENDAHULUAN

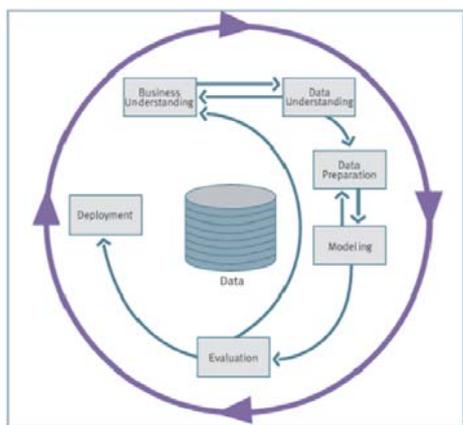
Pada saat ini, kendaraan bermotor atau motor sangat banyak sekali di jalan, tak jarang kita juga melihat banyak kemacetan. Jumlah motor yang semakin meningkat membuat kita tanpa sadar melihat motor sebagai kendaraan yang paling mudah dan efisien. Dalam mendapatkan motor, caranya juga sudah dipermudah. Mudah nya mendapatkan motor membuat salah satu perusahaan penyedia layanan kredit harus lebih selektif dikarenakan pesanan yang tinggi dan terkadang harus diikuti dengan kesiapan dalam bidang teknologi yang baik. Untuk itu dibutuhkan sebuah sistem keputusan yang sederhana dalam menganalisa pemberian kredit kendaraan bermotor. Selain itu, terdapat latar belakang lain yang mempengaruhi secara khusus pada angsuran atau kebutuhan akan pembiayaan yang baik seperti harus didukung dengan validitas data yang baik juga sehingga dalam perjalanan angsuran kedepannya masalah kredit macet dapat diminimalisir.

Adapun masalah yang dihadapi sampai saat ini yakni: belum adanya sebuah sistem komputerisasi yang dapat menangani masalah keputusan pemberian kredit pada konsumen atau pada penerima kredit motor. Kebutuhan atas komputerisasi sangat dibutuhkan karena diikuti oleh meningkatnya pesanan terhadap motor. Kebutuhan atas sistem tersebut dibuat dengan mengharapkan tercapainya beberapa tujuan dalam perancangan sistem penerapan algoritma *data mining* terhadap penerima kredit motor yakni memudahkan kinerja khususnya dalam menentukan keputusan tentang penerima atau calon konsumen yang berhak mendapat persetujuan atau penolakan terhadap kredit yang diajukan. Hal ini sejalan dengan tantangan utama dari *bussines intelligence* (BI) yang berhubungan erat dengan pola bisnis yang bersifat unik bagi tiap organisasi, begitu juga dengan kebijakan dan aturan bisnis yang diberlakukan oleh perusahaan. Hal tersebut menyebabkan perusahaan tidak dapat membeli produk *bussines intelligence* (BI) seperti barang jadi pada umumnya dan berharap dapat memenuhi setiap solusi dari kebutuhan bisnisnya sehingga *bussines intelligence* (BI) harus dikembangkan sesuai dengan kebutuhan dan proses bisnis perusahaan (Darudiato, Santoso, & Wiguna, 2010). Dalam pengajuan kredit, jika dengan cepat dapat diputuskan dengan seluruh kelengkapan data maka akan diperoleh efisiensi waktu dalam mengajukan kredit. Untuk itu, bisa lebih memaksimalkan dalam memberikan pelayanan kepada konsumen. Penelitian ini juga bertujuan untuk mengetahui dokumen dan segala kelengkapan apa saja yang dibutuhkan dalam pengajuan kredit yang sebenarnya dengan mudah dapat dilakukan.

Konsep yang digunakan dalam penelitian ini adalah konsep dari *data mining*. *Data mining* atau menambang data yang didefinisikan sebagai metode yang digunakan untuk mengekstraksi informasi prediktif tersembunyi pada *database*. Ini adalah teknologi yang sangat potensial bagi perusahaan dalam memberdayakan data *warehouse*-nya (Sulianta & Juju, 2010). *Data mining* disebut sebagai proses ekstraksi pengetahuan dari data yang besar. Sesuai fungsinya *data mining* adalah proses pengambilan keputusan dari *volume* data yang besar yang disimpan dalam basis data, data *warehouse*, atau informasi yang disimpan dalam *repository* (Han & Kamber, 2006). Istilah *data mining* berasal dari kemiripan antara pencarian informasi yang bernilai dari *basis data* yang besar dengan menambang pada sebuah gunung untuk memperoleh sesuatu yang bernilai (Sumathi & Sivanandam, 2006).

METODE

Penelitian ini didesain dengan menggunakan model CRISP-DM (*Cross Standart Industries for Data Mining*), dalam metode ini terdapat 6 tahapan (Larose, 2005).



Gambar 1. Tahap CRISP-DM
(Cross Standart Industries Process for Data Mining)

Terdapat beberapa tahapan dalam CRISP-DM (*Cross Standart Industries Process for Data Mining*).

Pertama, *Business/Research Understanding Phase*. Dalam hal ini, dilakukan pemahaman terhadap penelitian yang dilakukan. Perlunya pemahaman terhadap substansi atau inti dari penelitian yang dilakukan dimulai dari kebutuhan dan perspektif bisnis yang dilakukan. Terdapat beberapa kegiatan pada tahapan ini diantaranya adalah ditentukannya sasaran dan tujuan dari penelitian, pemahaman kondisi atau situasi bisnis, menentukan tujuan dari *data mining* dan melakukan penjadwalan atau perencanaan strategi penelitian.

Kedua, *Data Understanding Phase* (Fase Pemahaman Data). Fase ini dikenal sebagai fase pemahaman terhadap data yang diperoleh dan kemudian data awal yang dikumpulkan melalui observasi langsung, dengan melakukan analisa terhadap berkas persetujuan atau penolakan yang terdapat pada objek penelitian. Kajian perlu dilakukan sehingga diketahui data yang akan digunakan. Data yang diambil seluruhnya berjumlah 224 berkas profil pengajuan. Seluruh data berjumlah 224 sebagai populasi dan disalin setiap *field* dan bagian ke dalam tabel atribut yang kemudian dibagi menjadi 2 kelompok atau 2 bagian. Bagian pertama terdiri dari 80% data yang digunakan untuk dimasukkan dalam aplikasi *Rapidminer*. Data ini disebut sebagai data *training* data yang dijadikan acuan dalam membuat *rule* atau model algoritma. Selanjutnya, 20% data berikutnya akan digunakan dalam uji coba *rule* atau model yang dinamakan data *testing*. Pengumpulan data dilakukan sesuai dengan *attribute* yang tersedia pada objek penelitian, nilai dari form identitas, penghasilan perbulan, masa kerja, uang muka, usia, lama angsuran, penjamin hingga hasil persetujuan atas kredit yang dilakukan oleh objek penelitian. Data ini seluruhnya adalah kegiatan yang telah dilaksanakan oleh objek penelitian dalam menjalankan usaha bisnisnya. Seluruh data diambil berdasarkan berkas yang ditunjukkan oleh objek penelitian, jika data yang diperoleh semakin banyak maka hasil dari akurasi datanya juga akan maksimal. Jika diperlukan acuan pustaka juga dapat dilakukan pada tahapan ini. Dalam tahapan ini, perlu identifikasi terhadap masalah kualitas data yang baik sehingga diperoleh nilai subset yang sesuai, menarik dalam pembuatan hipotesa awal.

Ketiga, *Data Preparation Phase* (Fase Pengolahan Data). Tahapan ini merupakan tahapan pengolahan data atau dapat juga dikatakan sebagai tahapan persiapan data. Banyak persiapan yang dilakukan pada tahapan ini sehingga tak jarang fase ini juga disebut sebagai fase padat karya. Beberapa kegiatan seperti pemilihan tabel dan *field* terjadi pada fase ini. Pemilihan tabel dan *field* tersebut akan dimasukkan atau ditransformasikan kedalam *database* yang lain atau *database* baru

sebagai bahan atau *data mining* mentah. Keempat, *Modeling Phase* (Fase Pemodelan). Pada fase pemodelan dilakukan dengan penggunaan aplikasi seperti *Rapidminer*, aplikasi pengolah *data mining* mentah, dan dimasukkan juga algoritma C4.5. Dari data tersebut dipilih *attribute* yang menjadi label. Kemudian, seluruh parameter dipilih dengan penentuan nilai yang optimal.

Kelima, *Evaluation Phase* (Fase Evaluasi). Fase ini merupakan tahapan analisa dari hasil pengolahan fase sebelumnya dengan menginterpretasikan data yang kemudian diperoleh nilai perbandingan dengan proses model yang sebelumnya. Perlunya kajian mendalam pada tahapan ini adalah untuk menentukan nilai akurasi pada data model yang dihasilkan. Hal tersebut bertujuan agar dapat digunakan oleh sasaran sesuai rencana pada *domain goal* pada fase pertama. Keenam, *Deployment Phase* (Fase Penyebaran). Fase ini merupakan tahapan pembuatan laporan atau implementasi *knowledge* yang diperoleh dari fase sebelumnya.

Algoritma C4.5

Algoritma C4.5 merupakan algoritma klasifikasi dengan teknik pohon keputusan yang terkenal dan disukai karena memiliki kelebihan-kelebihan. Kelebihannya adalah dapat mengolah data numerik (kontinyu) dan diskret, dapat menangani nilai atribut yang hilang, dapat menghasilkan aturan-aturan yang mudah diinterpretasikan dan tercepat diantara algoritma-algoritma yang lain. Keakuratan prediksi adalah kemampuan model untuk dapat memprediksi label kelas terhadap data baru atau yang belum diketahui sebelumnya dengan baik. Dalam hal kecepatan atau efisiensi waktu komputasi yang diperlukan untuk membuat dan menggunakan model. Kemampuan model untuk memprediksi dengan benar walaupun ada data yang nilainya dari atribut yang hilang. Selanjutnya, skalabilitas adalah kemampuan untuk membangun model secara efisien untuk data yang berjumlah besar (aspek ini akan mendapatkan penekanan). Terakhir, interpretabilitas adalah model yang dihasilkan mudah dipahami. Selain itu dapat juga dikatakan bahwa Algoritma merupakan kumpulan perintah yang tertulis secara sistematis guna menyelesaikan permasalahan logika dari matematika. Pengertian algoritma C4.5 adalah algoritma yang digunakan untuk membentuk pohon keputusan. Sedangkan, pohon keputusan dapat diartikan sebagai suatu cara untuk memprediksi atau mengklarifikasi yang sangat kuat. Pohon keputusan dapat membagi kumpulan data yang besar menjadi himpunan-himpunan *record* yang lebih kecil dengan menerapkan serangkaian aturan keputusan.

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 (Kusrini & Luthfi, 2009), yaitu: (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 akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan rumus:

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

Keterangan:

S = himpunan kasus

n = jumlah partisi S

p_i = proporsi S_i terhadap S

1. Kemudian hitung nilai *gain* dengan menggunakan rumus:

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

Keterangan:

S = himpunan kasus

A = fitur

n = jumlah partisi atribut A

$|S_i|$ = proporsi S_i terhadap S

$|S|$ = jumlah kasus dalam S

2. Ulangi langkah ke-2 hingga semua *record* terpartisi.
3. Proses partisi pohon keputusan akan berhenti saat :
 - a. Semua *record* dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut di dalam *record* yang dipartisi lagi.
 - c. Tidak ada record di dalam cabang yang kosong.

HASIL DAN PEMBAHASAN

Perancangan dilakukan dengan menggunakan perangkat lunak *Rapidminer* ver. 5.1. Hasil meta data yang telah diinput adalah sebagai berikut:

Tabel 1 Tabel Meta Data

Role	Name	Type
<i>Id</i>	Identitas	Binomial
<i>Cost</i>	Uang Muka	Polynomial
<i>Label</i>	Hasil	Binomial
<i>Reguler</i>	Masa Kerja	Polynomial
<i>Reguler</i>	Usia	Polynomial
<i>Reguler</i>	Lama Angsuran	Polynomial
<i>Reguler</i>	Penjamin	Binomial

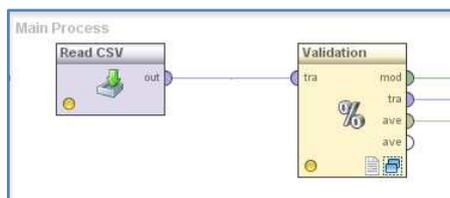
Tabel 2 Sample Data Testing Final

Identitas	Penghasilan Perbulan	Masa Kerja	Uang Muka	Usia	Lama Angsuran	Penjamin	Hasil
Lengkap	>5Juta	3-4 Tahun	>5 Juta	17-45 Tahun	3 Tahun	Ada Penjamin	Disetujui
Lengkap	1Juta-2Juta	4-5 Tahun	<1 Juta	>45 Tahun	1 Tahun	Ada Penjamin	Disetujui
Lengkap	2Juta-3Juta	4-5 Tahun	<1 Juta	<17 Tahun	2 Tahun	Ada Penjamin	Disetujui
Lengkap	3Juta-4Juta	>5 Tahun	1-2 Juta	17-45 Tahun	3 Tahun	Ada Penjamin	Disetujui
Lengkap	4Juta-5Juta	1-2 Tahun	2-3 Juta	>45 Tahun	4 Tahun	Ada Penjamin	Disetujui
Lengkap	>5Juta	2-3 Tahun	3-4 Juta	17-45 Tahun	2 Tahun	Ada Penjamin	Disetujui
Lengkap	1Juta-2Juta	3-4 Tahun	>5 Juta	>45 Tahun	3 Tahun	Ada Penjamin	Disetujui
Lengkap	2Juta-3Juta	4-5 Tahun	<1 Juta	<17 Tahun	3 Tahun	Ada Penjamin	Disetujui
Lengkap	1Juta-2Juta	>5 Tahun	1-2 Juta	17-45 Tahun	3 Tahun	Ada Penjamin	Disetujui
Lengkap	2Juta-3Juta	1-2 Tahun	2-3 Juta	<17 Tahun	2 Tahun	Ada Penjamin	Disetujui
Lengkap	3Juta-4Juta	2-3 Tahun	3-4 Juta	17-45 Tahun	3 Tahun	Ada Penjamin	Disetujui
Lengkap	1Juta-2Juta	1-2 Tahun	<1 Juta	17-45 Tahun	1 Tahun	Ada Penjamin	Disetujui
Lengkap	2Juta-3Juta	2-3 Tahun	1-2 Juta	17-45 Tahun	2 Tahun	Ada Penjamin	Disetujui
Lengkap	3Juta-4Juta	3-4 Tahun	2-3 Juta	>45 Tahun	3 Tahun	Ada Penjamin	Disetujui
Lengkap	4Juta-5Juta	4-5 Tahun	3-4 Juta	<17 Tahun	2 Tahun	Ada Penjamin	Disetujui
Lengkap	>5Juta	>5 Tahun	>5 Juta	17-45 Tahun	3 Tahun	Ada Penjamin	Disetujui
Lengkap	3Juta-4Juta	3-4 Tahun	<1 Juta	>45 Tahun	4 Tahun	Ada Penjamin	Disetujui
Lengkap	4Juta-5Juta	4-5 Tahun	1-2 Juta	<17 Tahun	5 Tahun	Ada Penjamin	Disetujui
Lengkap	>5Juta	>5 Tahun	2-3 Juta	17-45 Tahun	1 Tahun	Ada Penjamin	Disetujui

Tabel 2 Sample Data Testing Final (*lanjutan*)

Identitas	Penghasilan Perbulan	Masa Kerja	Uang Muka	Usia	Lama Angsuran	Penjamin	Hasil
Lengkap	2Juta-3Juta	1-2 Tahun	3-4 Juta	17-45 Tahun	2 Tahun	Ada Penjamin	Disetujui
Lengkap	3Juta-4Juta	2-3 Tahun	>5 Juta	>45 Tahun	1 Tahun	Ada Penjamin	Disetujui
Lengkap	4Juta-5Juta	1-2 Tahun	<1 Juta	<17 Tahun	2 Tahun	Ada Penjamin	Disetujui
Lengkap	>5Juta	2-3 Tahun	<1 Juta	17-45 Tahun	3 Tahun	Ada Penjamin	Disetujui
Lengkap	1Juta-2Juta	3-4 Tahun	1-2 Juta	17-45 Tahun	4 Tahun	Ada Penjamin	Disetujui
Lengkap	2Juta-3Juta	4-5 Tahun	2-3 Juta	>45 Tahun	3 Tahun	Ada Penjamin	Disetujui
Lengkap	3Juta-4Juta	>5 Tahun	3-4 Juta	<17 Tahun	1 Tahun	Ada Penjamin	Disetujui
Lengkap	4Juta-5Juta	1-2 Tahun	>5 Juta	17-45 Tahun	2 Tahun	Ada Penjamin	Disetujui
Lengkap	>5Juta	2-3 Tahun	<1 Juta	>45 Tahun	3 Tahun	Ada Penjamin	Disetujui
Lengkap	1Juta-2Juta	1-2 Tahun	1-2 Juta	<17 Tahun	4 Tahun	Ada Penjamin	Disetujui
Lengkap	2Juta-3Juta	2-3 Tahun	2-3 Juta	17-45 Tahun	3 Tahun	Ada Penjamin	Disetujui
Lengkap	3Juta-4Juta	1-2 Tahun	3-4 Juta	<17 Tahun	4 Tahun	Ada Penjamin	Disetujui
Lengkap	4Juta-5Juta	2-3 Tahun	<1 Juta	17-45 Tahun	3 Tahun	Ada Penjamin	Disetujui
Lengkap	>5Juta	3-4 Tahun	1-2 Juta	>45 Tahun	1 Tahun	Ada Penjamin	Disetujui
Lengkap	1Juta-2Juta	4-5 Tahun	2-3 Juta	<17 Tahun	2 Tahun	Ada Penjamin	Disetujui
Tidak Lengkap	2Juta-3Juta	4-5 Tahun	3-4 Juta	17-45 Tahun	3 Tahun	Ada Penjamin	Ditolak
Tidak Lengkap	1Juta-2Juta	>5 Tahun	>5 Juta	>45 Tahun	4 Tahun	Ada Penjamin	Ditolak
Tidak Lengkap	2Juta-3Juta	1-2 Tahun	<1 Juta	<17 Tahun	2 Tahun	Tidak ada Penjamin	Ditolak
Tidak Lengkap	3Juta-4Juta	2-3 Tahun	<1 Juta	17-45 Tahun	3 Tahun	Tidak ada Penjamin	Ditolak
Tidak Lengkap	1Juta-2Juta	3-4 Tahun	1-2 Juta	>45 Tahun	3 Tahun	Tidak ada Penjamin	Ditolak
Tidak Lengkap	2Juta-3Juta	4-5 Tahun	2-3 Juta	17-45 Tahun	3 Tahun	Tidak ada Penjamin	Ditolak
Tidak Lengkap	3Juta-4Juta	>5 Tahun	3-4 Juta	>45 Tahun	2 Tahun	Tidak ada Penjamin	Ditolak
Tidak Lengkap	4Juta-5Juta	1-2 Tahun	>5 Juta	<17 Tahun	3 Tahun	Tidak ada Penjamin	Ditolak
Tidak Lengkap	>5Juta	2-3 Tahun	<1 Juta	17-45 Tahun	1 Tahun	Tidak ada Penjamin	Ditolak
Tidak Lengkap	3Juta-4Juta	1-2 Tahun	1-2 Juta	>45 Tahun	2 Tahun	Tidak ada Penjamin	Ditolak
Tidak Lengkap	4Juta-5Juta	2-3 Tahun	2-3 Juta	<17 Tahun	3 Tahun	Tidak ada Penjamin	Ditolak

Data yang digunakan dalam aplikasi adalah data *training*. Dari seluruh populasi, data dipisah menjadi 80% data *training* yang dipakai untuk membuat model atau *rule* dan 20% sebagai data *testing* seperti pada tabel 2 diatas. Dari meta data tersebut dilakukan pengolahan pada *rapidminer* dengan *main process* sebagai berikut:



Gambar 2 Main Proses

Main proses dilakukan sesuai dengan tahapan CRISP-DM dengan menambahkan data dengan tipe csv, dan memasukan *box validation* untuk memastikan agar hasil pengukuran dapat maksimal. Selanjutnya, diperoleh tingkat pengukuran akurasi sebagai berikut:

accuracy: 85.67% +/- 4.27% (mikro: 85.65%)			
	true Disetujui	true Ditolak	class precision
pred. Disetujui	190	27	87.56%
pred. Ditolak	5	1	16.67%
class recall	97.44%	3.57%	

Gambar 3 Pengukuran Akurasi data

Hasil pengukuran akurasi data yang diperoleh bahwa secara keseluruhan nilainya mencapai 85.67%. Dari tabel tersebut diketahui prediksi disetujuinya pengajuan kredit konsumen dengan *true* disetujui mencapai 190 konsumen dan *true* ditolak sebanyak 27 konsumen, dengan hasil pencapaian nilai presisi sebesar 87.56% . Sedangkan untuk prediksi data konsumen yang ditolak untuk *true* disetujui mencapai 5 konsumen dan untuk *true* ditolak terdapat 1 konsumen atau 1 penerima kredit motor. Persentase ditolaknya konsumen adalah sebesar 16.67%. Untuk *recall* data konsumen yang sangat tinggi terdapat *class recall true* disetujui mencapai 97.44% sedangkan untuk *class recall true* ditolak mencapai 3.57% konsumen penerima kredit motor. Melihat data lebih besar kepada konsumen, pengukuran akurasi diterima dengan tingkat akurasi tinggi yang berbanding terbalik dengan pengukuran *presicion*.

Pengukuran yang berikutnya sebagai pembanding *presicion* data sebagai berikut:

precision: 16.67% (positive class: Ditolak)			
	true Disetujui	true Ditolak	class precision
pred. Disetujui	190	27	87.56%
pred. Ditolak	5	1	16.67%
class recall	97.44%	3.57%	

Gambar 4 Pengukuran Precision data

Hasil pengukuran *precision* data diperoleh bahwa secara keseluruhan nilainya mencapai 16.67%. Hasil ini sama seperti hasil pengukuran pada akurasi. Dengan hasil dari tabel tersebut diketahui prediksi disetujuinya pengajuan kredit konsumen dengan *true* disetujui mencapai 190 konsumen dan *true* ditolak sebanyak 27 konsumen, dengan hasil pencapaian nilai presisi sebesar 87.56%. Sedangkan, untuk prediksi data konsumen yang ditolak yaitu untuk *true* disetujui mencapai 5 konsumen dan *true* ditolak terdapat 1 konsumen atau 1 penerima kredit motor. Persentase ditolaknya konsumen adalah sebesar 16.67%. Untuk *recall* data konsumen sangat tinggi yaitu untuk *class recall true* disetujui mencapai 97.44%, sedangkan untuk *class recall true* ditolak mencapai 3.57%. Dalam hal ini data difokuskan pada kebalikan hasil *accuracy*.

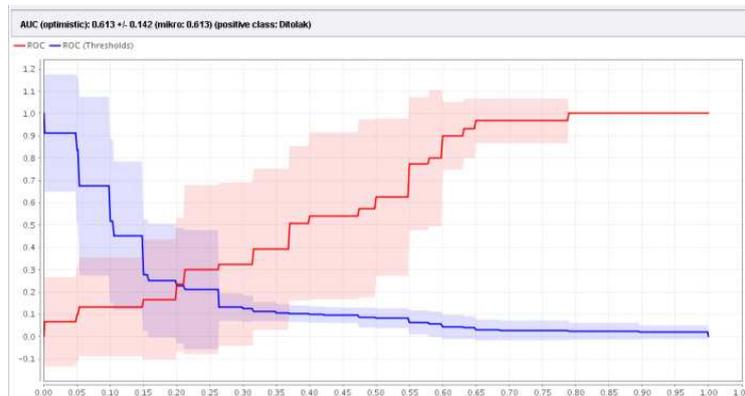
Nilai pengukuran *confusion matrix* selanjutnya yaitu *class recall* sebagai berikut:

recall: 3.33% +/- 10.00% (mikro: 3.57%) (positive class: Ditolak)			
	true Disetujui	true Ditolak	class precision
pred. Disetujui	190	27	87.56%
pred. Ditolak	5	1	16.67%
class recall	97.44%	3.57%	

Gambar 5 Pengukuran Recall data

Hasil dari pengukuran *recall* data jika diperhatikan untuk prediksi disetujui dan prediksi ditolak adalah sama dengan gambar 4 dan 5. Yang membedakannya adalah nilai dari *recall* sebesar 3.33% yang membedakan dengan lain. Nilainya mengartikan bahwa untuk nilai atau hasil keputusan dalam pemberian kredit motor sangat sedikit sekali yang ditolak sehingga yang merupakan positive class adalah “ditolak”.

Setelah dilakukan seluruh tahapan evaluasi untuk *confusion matrix* maka selanjutnya dilakukan analisa evaluasi pembandingan yakni terhadap pengukuran *Receiver Operating Characteristic* (ROC) sebagai berikut:



Gambar 6 Pengukuran *UAC Optimistic*

Nilai dari ROC diukur terlebih dahulu terhadap Area Under Curvenya (AUC) optimistic dengan nilai sebesar 0.613 terhadap kelas yang ditolak. Pengukuran dilakukan berikutnya terhadap AUC normal adalah:



Gambar 7 Pengukuran *AUC*

Nilai yang diperoleh dari pengukuran AUC adalah sebesar 0.477 untuk kelas yang ditolak. Selanjutnya pengukuran untuk *pessimistic*-nya adalah sebagai berikut:



Gambar 8 Pengukuran *AUC Pessimistic*

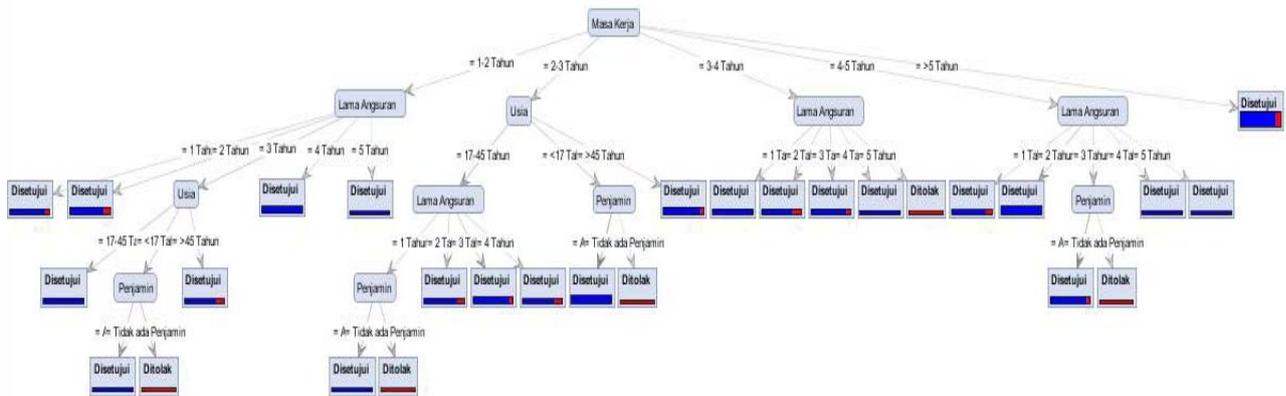
Nilai yang diperoleh dari AUC pessimistic adalah 0.341 untuk class predictive yang ditolak. Pengukuran berikutnya dilakukan untuk mengetahui *performance* terhadap *vector* yang diperoleh sebagai berikut:

```

PerformanceVector:
accuracy: 85.67% +/- 4.27% (mikro: 85.65%)
ConfusionMatrix:
True: Disetujui Ditolak
Disetujui: 190 27
Ditolak: 5 1
precision: 16.67% (positive class: Ditolak)
ConfusionMatrix:
True: Disetujui Ditolak
Disetujui: 190 27
Ditolak: 5 1
recall: 3.33% +/- 10.00% (mikro: 3.57%) (positive class: Ditolak)
ConfusionMatrix:
True: Disetujui Ditolak
Disetujui: 190 27
Ditolak: 5 1
AUC (optimistic): 0.613 +/- 0.142 (mikro: 0.613) (positive class: Ditolak)
AUC: 0.477 +/- 0.162 (mikro: 0.477) (positive class: Ditolak)
AUC (pessimistic): 0.341 +/- 0.201 (mikro: 0.341) (positive class: Ditolak)

```

Dari hasil keseluruhan maka diperoleh pohon keputusan sebagai berikut:



Gambar 9 Pohon Algoritma C4.5 Penerima Kredit Motor

Pada Gambar 9, Pohon Algoritma C4.5 Penerima Kredit Motor jika di jabarkan ke dalam bahasa yang lebih sederhana dan dibuat dalam bentuk pohon *text* maka akan diperoleh hasil sebagai berikut:

```

Masa Kerja = 1-2 Tahun
| Lama Angsuran = 1 Tahun: Disetujui {Disetujui=8, Ditolak=1}
| Lama Angsuran = 2 Tahun: Disetujui {Disetujui=11, Ditolak=2}
| Lama Angsuran = 3 Tahun
| | Usia = 17-45 Tahun: Disetujui {Disetujui=7, Ditolak=0}

```

```

| | Usia = <17 Tahun
| | | Penjamin = Ada Penjamin: Disetujui {Disetujui=4, Ditolak=0}
| | | Penjamin = Tidak ada Penjamin: Ditolak {Disetujui=0, Ditolak=3}
| | Usia = >45 Tahun: Disetujui {Disetujui=4, Ditolak=1}
| | Lama Angsuran = 4 Tahun: Disetujui {Disetujui=12, Ditolak=0}
| | Lama Angsuran = 5 Tahun: Disetujui {Disetujui=2, Ditolak=0}
Masa Kerja = 2-3 Tahun
| | Usia = 17-45 Tahun
| | | Lama Angsuran = 1 Tahun
| | | | Penjamin = Ada Penjamin: Disetujui {Disetujui=3, Ditolak=0}
| | | | Penjamin = Tidak ada Penjamin: Ditolak {Disetujui=0, Ditolak=2}
| | | Lama Angsuran = 2 Tahun: Disetujui {Disetujui=5, Ditolak=1}
| | | Lama Angsuran = 3 Tahun: Disetujui {Disetujui=10, Ditolak=1}
| | | Lama Angsuran = 4 Tahun: Disetujui {Disetujui=4, Ditolak=1}
| | Usia = <17 Tahun
| | | Penjamin = Ada Penjamin: Disetujui {Disetujui=14, Ditolak=0}
| | | Penjamin = Tidak ada Penjamin: Ditolak {Disetujui=0, Ditolak=2}
| | Usia = >45 Tahun: Disetujui {Disetujui=13, Ditolak=1}
Masa Kerja = 3-4 Tahun
| | Lama Angsuran = 1 Tahun: Disetujui {Disetujui=10, Ditolak=0}
| | Lama Angsuran = 2 Tahun: Disetujui {Disetujui=7, Ditolak=2}
| | Lama Angsuran = 3 Tahun: Disetujui {Disetujui=8, Ditolak=1}
| | Lama Angsuran = 4 Tahun: Disetujui {Disetujui=5, Ditolak=0}
| | Lama Angsuran = 5 Tahun: Ditolak {Disetujui=0, Ditolak=2}
Masa Kerja = 4-5 Tahun
| | Lama Angsuran = 1 Tahun: Disetujui {Disetujui=6, Ditolak=1}
| | Lama Angsuran = 2 Tahun: Disetujui {Disetujui=16, Ditolak=0}
| | Lama Angsuran = 3 Tahun
| | | Penjamin = Ada Penjamin: Disetujui {Disetujui=10, Ditolak=1}
| | | Penjamin = Tidak ada Penjamin: Ditolak {Disetujui=0, Ditolak=2}
| | Lama Angsuran = 4 Tahun: Disetujui {Disetujui=4, Ditolak=0}
| | Lama Angsuran = 5 Tahun: Disetujui {Disetujui=2, Ditolak=0}
Masa Kerja = >5 Tahun: Disetujui {Disetujui=30, Ditolak=4}

```

Untuk hasil pengukuran tingkat akurasi dalam *data mining*, nilai tersebut dapat dibagi menjadi beberapa kelompok (Gorunescu, 2011).

- a. 0.90-1.00 = klasifikasi sangat baik
- b. 0.80-0.90 = klasifikasi baik
- c. 0.70-0.80 = klasifikasi cukup
- d. 0.60-0.70 = klasifikasi buruk
- e. 0.50-0.60 = klasifikasi salah

Dari nilai hasil pengukuran akurasi maka dapat disimpulkan sesuai dengan pengelompokan klasifikasi termasuk pada klasifikasi baik untuk Penerapan algoritma *data mining* terhadap penerima kredit motor.

SIMPULAN

Penerapan algoritma *data mining* terhadap penerima kredit motor memiliki tingkat akurasi yang cukup tinggi yaitu 87% atau sekitar 217 konsumen. Sedangkan yang ditolak sebesar 16% atau sebanyak 6 konsumen yang ditolak pengajuan kreditnya dari 224 konsumen. Kategori penelitian ini masuk dalam klasifikasi baik dengan persentase 0.80-0.90. Dari hasil tersebut maka perlu terus dikembangkan dan dikomparasikan dengan beberapa algoritma lain sehingga hasil akurasi bisa jauh lebih tinggi lagi.

DAFTAR PUSTAKA

- Darudiato, S., Santoso, S.W. & Wiguna, S. (2010). Business Intelligence: Konsep dan Metode. *Jurnal CommIT*, 4(1), 63 – 67.
- Gorunescu, F. (2011). *Data Mining: Concepts, Models, and Techniques*. Verlag Berlin Heidelberg: Springer.
- Han, J., Kamber, M. (2006). *Data Mining Concept and Tehniques*. New York: Morgan Kauffman
- Kusrini, Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Publishing.
- Larose, D. T. (2005). *Discovering Knowledge in Data*. New Jersey: John Willey & Sons Inc.
- Sulianta, F., Juju, D. (2010). *Data Mining Meramalkan Bisnis Perusahaan*. Jakarta: PT. Elex Media Komputindo.
- Sumathi, S., Sivanandam, S. N. (2006). *Introduction to Data Mining and its Applications*. New York: Springer-Verlag Berlin Heidelberg.