

EVALUASI PEMOHON KREDIT MOBIL DI PT “X” DENGAN MENGGUNAKAN TEKNIK DATA MINING DECISION TREE

Yogi Yusuf W.¹, F. Rian Pratikto², Vivianne A. S.³

^{1,2,3}Jurusan Teknik Industri, Universitas Katolik Parahyangan
Jalan Ciumbuleuit 94, Bandung – 40141, Tlp/Fax: (022) 2032700
Email: yogi@home.unpar.ac.id

Abstrak

Kredit merupakan salah satu mekanisme pembayaran yang sangat umum di masyarakat. Fungsi pokok kredit yaitu memenuhi pelayanan terhadap kebutuhan masyarakat dalam rangka memperlancar perdagangan, produksi dan jasa-jasa bahkan konsumsi yang kesemuanya itu ditujukan untuk meningkatkan kesejahteraan manusia. PT “X” merupakan distributor mobil di kota Bandung dan telah memberikan kemudahan pembayaran kepada pelanggannya dengan menyediakan fasilitas kredit. Ada risiko yang harus dihadapi oleh perusahaan yaitu tidak tepat waktunya pembayaran atau bahkan kegagalan pembayaran dari kredit yang disalurkan. Kredit macet ini sangat berpengaruh terhadap aliran kas perusahaan. Usaha untuk mengurangi jumlah kredit yang macet harus selalu dilakukan. Banyak faktor yang dapat menyebabkan kredit macet, salah satunya adalah kesalahan dalam pengambilan keputusan penerimaan pemohon menjadi debitur perusahaan pada tahap evaluasi kredit. Disamping kesalahan dengan menerima pemohon yang buruk, evaluasi kredit juga memungkinkan membuat kesalahan dengan menolak pemohon yang baik. Oleh karena itu setiap perusahaan yang memberikan kredit harus mampu mengevaluasi pemohon kredit dengan objektif, akurat, dan konsisten. Tool yang dapat membantu dalam mengevaluasi kredit dengan objektif, akurat, dan konsisten adalah credit scoring. Penelitian ini berfokus pada pembangunan model credit scoring kredit mobil dengan teknik decision tree. Teknik decision tree dapat membangun model dengan objektif, menghasilkan model yang mudah dipahami, dan tingkat akurasi yang tinggi. Model yang dibangun melibatkan variabel target yaitu risiko kredit dan variabel prediktor yaitu penghasilan, cicilan, uang muka, jumlah periode pinjaman, rekening tabungan, umur pemohon, tagihan telepon, dan tagihan listrik. Dari hasil pengembangan model diperoleh bahwa variabel penghasilan merupakan variabel yang paling penting dalam memprediksi risiko kredit dan tingkat akurasi model sebesar 79.57 persen.

Kata kunci: credit scoring; data mining; decision tree; kredit; variabel prediktor; variabel target

Pendahuluan

Kemajuan teknologi telah memudahkan berbagai macam aktivitas yang dijalankan oleh manusia. Salah satunya adalah mobil atau motor sebagai alat transportasi yang memudahkan manusia untuk dapat bergerak dari satu tempat ke tempat lain dengan mudah, cepat, dan nyaman. Ketergantungan terhadap alat transportasi ini sangat tinggi terutama di daerah perkotaan. Hampir setiap rumah tangga di kota-kota besar memiliki mobil atau motor. Kecenderungan untuk memiliki alat transportasi pribadi dari tahun ke tahun semakin meningkat. Kondisi ini dipengaruhi oleh banyak faktor antara lain alat transportasi umum yang kurang memadai dan adanya kemudahan pembayaran dengan adanya fasilitas kredit dalam pembelian mobil ataupun motor.

PT “X” sebagai salah satu distributor mobil di kota Bandung memberikan kemudahan pembayaran kepada pelanggannya dengan menyediakan fasilitas kredit. Ada risiko yang harus dihadapi oleh perusahaan yaitu tidak tepat waktunya pembayaran atau bahkan kegagalan pembayaran dari kredit yang disalurkan. Kredit macet ini sangat berpengaruh terhadap aliran kas perusahaan. Untuk itu usaha untuk mengurangi jumlah kredit yang macet harus selalu dilakukan.

Banyak faktor yang dapat menyebabkan kredit macet, salah satunya adalah kesalahan dalam mengambil keputusan penerimaan pemohon menjadi debitur perusahaan pada tahap evaluasi kredit. Disamping kesalahan menerima pemohon yang buruk, evaluasi kredit juga memungkinkan membuat kesalahan dengan menolak pemohon yang baik. Saat ini perusahaan menerapkan evaluasi kredit kepada setiap pemohon kredit melalui serangkaian tahap, yaitu: tahap pemberian skor kredit, tahap survei lapangan, dan tahap analisa. Dari tahap-tahap tersebut akan diambil keputusan terkait dengan penerimaan atau penolakan pemohon sebagai debitur.

Tahap pemberian skor merupakan tahap penyangkapan yang memisahkan pemohon yang baik dan buruk. Tiap pemohon akan diberikan skor berdasarkan kriteria penghasilan, cicilan per bulan, uang muka, jumlah periode cicilan, rekening koran/tabungan, umur, rekening tagihan telepon, rekening tagihan listrik. Jika skor kredit 500 atau

lebih, pemohon kredit akan dievaluasi pada tahap berikutnya yaitu survei. Tahap survei merupakan tahap verifikasi informasi yang diberikan oleh pemohon dengan datang ke tempat pemohon. Jika lolos dari tahap ini, analisa terhadap pemohon dilakukan untuk melihat aspek-aspek lain yang tidak dapat terakomodasi dalam tahap penilaian kredit dan keputusan terkait penerimaan atau penolakan kredit yang diajukan.

Dari ketiga tahap tersebut, tahap pertama merupakan tahap yang krusial dengan menghasilkan informasi yang berguna bagi perusahaan terutama dalam membantu pengambilan keputusan yang dilakukan pada tahap ketiga. Hasil yang tidak akurat pada tahap pertama dapat berpengaruh terhadap kualitas keputusan pemberian atau penolakan kredit. Oleh karena itu, penelitian ini akan difokuskan pada pengembangan model penilaian kredit.

Saat ini perusahaan melakukan penilaian skor kredit dengan memberikan nilai pada tiap variabel yang dievaluasi kemudian menjumlahkan nilai-nilai tersebut menjadi nilai skor kredit. Kelemahan metode penilaian kredit sekarang adalah sebagai berikut:

1. Unsur subjektivitas sangat dominan terhadap pembagian dan penilaian kelas variabel sebagai standar penilaian pemohon kredit. Berdasarkan standar penilaian tersebut, pemohon akan dinilai. Walaupun konsistensi penilaian terjamin, tetapi standar penilaian tidak dibangun berdasarkan data objektif masa lalu. Kondisi ini dapat menimbulkan kekurangakuratan dalam menilai pemohon kredit. *Credit scoring* didasarkan pada data riil dan statistik sehingga hasilnya akan lebih andal dibandingkan dengan metode subjektif (Isaac, 2006).
2. Bobot variabel dianggap sama yang berarti bahwa tiap variabel mempunyai kontribusi yang sama terhadap skor kredit. Model *credit scoring* yang efektif didasarkan pada banyak variabel dan bobot tiap variabel bervariasi dari satu model dengan model yang lain (Glassman, 1997).
3. Setiap pemohon yang diterima menjadi debitur dianggap mempunyai peluang yang sama menjadi debitur yang tidak dapat membayar utang tepat pada waktunya atau gagal melakukan pembayaran utang.

Berdasarkan kelemahan-kelemahan di atas, pada penelitian ini akan dikembangkan model *credit scoring* dengan menggunakan teknik *decision tree*. Kemampuan dari teknik ini adalah:

1. Dapat mengklasifikasikan pemohon kedalam kelas yang baik dan buruk berdasarkan data historis debitur, sehingga unsur subjektivitas dalam pembangunan standar penilaian dapat dikurangi.
2. Dapat memberikan informasi urutan tingkat kepentingan variabel terhadap estimasi status kredit pemohon.
3. Dapat mengkarakterisasi tiap pemohon baik yang dikategorikan baik ataupun buruk berdasarkan kombinasi nilai dari tiap variabel, serta peluang pemohon tersebut menjadi debitur yang bermasalah atau tidak bermasalah.

Disamping alasan-alasan di atas, pemilihan *decision tree* juga didukung oleh hasil perbandingan beberapa teknik yang dilakukan oleh beberapa peneliti yang menunjukkan teknik *tree* mempunyai tingkat keakuratan yang lebih tinggi dibanding teknik yang lain. Perbandingan tingkat keakuratan klasifikasi dari beberapa teknik dapat dilihat pada tabel 1.

Tabel 1. Tingkat Akurasi Hasil Klasifikasi dari Berbagai Teknik Klasifikasi (dalam Persen)

<i>Authors</i>	<i>Linear Regression</i>	<i>Logistic Regression</i>	<i>Class Trees</i>	<i>Linear Programming</i>	<i>Neural Nets.</i>	<i>Genetic Algorithm</i>
Henley	43.4	43.3	43.8	-	-	-
Boyle et al	77.5	-	75	74.7	-	-
Srinivisan and Kim	87.5	89.3	93.2	86.1	-	-
Yobas et al	68.4	-	62.3	-	62.0	64.5
Desai et al	66.5	67.3	67.3	-	6.4	6.4
Chye et al	-	71.1	74.2	-	73.4	-

Berdasarkan uraian di atas, masalah yang akan diteliti dapat dirumuskan sebagai berikut:

1. Bagaimana model prediksi status kredit berdasarkan teknik *decision tree* di PT. X?
2. Bagaimana tingkat akurasi dari model yang dihasilkan?
3. Bagaimana karakteristik dari pemohon kredit dan peluang pemohon kredit tersebut digolongkan ke dalam kelas kredit lancar dan macet?

Credit Scoring

Kredit merupakan salah satu mekanisme pembayaran yang sangat umum di masyarakat. Fungsi pokok kredit yaitu memenuhi pelayanan terhadap kebutuhan masyarakat dalam rangka memperlancar perdagangan, produksi dan jasa-jasa bahkan konsumsi yang kesemuanya itu ditujukan untuk meningkatkan kesejahteraan manusia. Henry Ford dan AP Sloan telah menyadari bahwa tidaklah cukup menghasilkan produk, seperti mobil, untuk pasar massal tetapi juga harus dikembangkan cara-cara pembayaran dari produk yang dibelinya. Kesadaran tersebut mendorong munculnya lembaga pembiayaan, seperti GE Capital, GM Finance (Thomas, 2005).

Pihak pemberi dan penerima kredit harus bekerja sama untuk mencapai tujuan yang diinginkan. Pihak penerima kredit mempunyai tujuan mendapatkan sumber pembiayaan yang mudah dan cepat. Pihak pemberi kredit mempunyai tujuan mendapatkan keuntungan dari bunga yang dibayarkan oleh debitur. Disamping keuntungan yang

dapat diperoleh, kreditor juga harus menanggung risiko ketidaklancaran pembayaran oleh debitur. Pihak kreditor harus mengevaluasi setiap pemohon kredit sebelum keputusan penerimaan atau penolakan diambil.

Banyaknya permohonan kredit menuntut kreditor harus mampu mengevaluasi pemohon kredit dengan objektif, akurat, dan konsisten. Evaluasi tersebut dapat dibantu dengan *credit scoring*. Isac mendefinisikan *credit scoring* sebagai *tool* yang melibatkan penggunaan model statistik untuk mengevaluasi seluruh informasi yang tersedia dengan objektif dalam pengambilan keputusan kredit (Noe, 1997).

Manfaat yang dapat diperoleh dari penerapan *credit scoring* adalah peningkatan kecepatan dan konsistensi proses aplikasi pinjaman dan memungkinkan otomatisasi proses peminjaman (Chye, 2004); adanya kemampuan belajar sepanjang waktu karena model *credit scoring* didasarkan pada perhitungan statistik dari data masa lalu (Glassman, 1997).

Penerapan model statistik prediktif membutuhkan dua faktor (Glassman, 1997):

1. Teknologi yang memungkinkan model bekerja dengan cepat sehingga kecepatan proses memberikan waktu respon yang dapat diterima, dan
2. Basis data yang menyediakan input bagi model prediktif.

Seiring dengan kemajuan teknologi informasi yang sangat pesat disertai dengan harga yang semakin terjangkau, sangat dimungkinkan bagi perusahaan kecilpun menggunakan model statistik dalam mengevaluasi kredit.

Model *credit scoring* dibangun dengan menggunakan sampel kredit masa lalu dalam jumlah yang besar. Sampel tersebut dibagi kedalam dua kelas yaitu kredit yang baik (pembayaran dilakukan tepat waktu) dan kredit yang bermasalah (pembayaran dilakukan tidak tepat waktu atau tidak dapat melakukan pembayaran). Berdasarkan pola masa lalu, kombinasi karakteristik peminjam yang membedakan peminjam yang baik dan yang buruk menghasilkan skor sebagai estimasi risiko dari tiap peminjam baru.

Banyak teknik yang dapat membantu dalam pembangunan model *credit scoring*. Pada perkembangan terbaru, teknik-teknik yang terdapat di dalam *data mining* mulai banyak digunakan. Khususnya teknik *decision tree* telah menjadi teknik yang populer karena *decision tree* yang dihasilkan mudah diinterpretasikan dan divisualisasikan (Chye, 2004).

Data Mining

Data tidak dapat dipisahkan dari kehidupan sehari-hari. Suatu organisasi dapat dibanjiri dengan berbagai macam data. Sangatlah tidak berguna data yang terkumpul dan tersimpan jika tidak dimanfaatkan. Permasalahan yang muncul adalah bagaimana memproses data-data tersebut sehingga dapat menampilkan suatu pengetahuan yang berguna bagi manajemen dalam mengambil keputusan. Salah satu teknologi yang dapat digunakan untuk mengolah data menjadi pengetahuan adalah *data mining*.

Menurut laporan Gartner Agustus 2005, pertumbuhan pasar *data mining* sangat pesat. Di dunia pada tahun 2004 pertumbuhan pasar *data mining* mencapai 12,1 persen dengan total pasar sebesar \$2,1 miliar dollar, Asia Pasifik sendiri tumbuh 24,2 persen dengan total pasar \$137,2 juta dollar, dan di Indonesia dengan total pasar \$400 ribu dollar mencatatkan pertumbuhan 60,4 persen. Pertumbuhan pesat data mining di Asia Pasifik ini disebabkan banyaknya korporat yang sekarang membutuhkan fungsi laporan dan analisa untuk pengambilan keputusan setelah dalam beberapa tahun sebelumnya mengimplementasikan infrastruktur data melalui ERP dan *data warehousing*. Alasan yang lebih spesifik untuk negara berkembang dalam implementasi intelegensi bisnis adalah kebutuhan usaha penurunan biaya dan peningkatan efisiensi operasi.

Banyak pihak yang telah mendefinisikan *data mining*. Berikut beberapa definisi *data mining*:

- Gartner mendefinisikan *data mining* sebagai berikut: "*Data mining* merupakan suatu proses pencarian pola dari data-data dengan jumlah yang sangat banyak yang tersimpan dalam suatu tempat penyimpanan dengan menggunakan teknologi pengenalan pola, teknik statistik, dan matematik." (Larose, 2005)
- Hand et al. mendefinisikan *data mining* sebagai berikut: "*Data mining* merupakan analisis dari sekumpulan data yang diamati (sangat besar) untuk menemukan hubungan yang tidak terduga dan merangkum data dengan cara yang baru yang dapat dipahami dan berguna bagi pemilik data." (Larose, 2005)
- Berry mendefinisikan *data mining* sebagai berikut: "*Data mining* merupakan proses eksplorasi dan analisis data yang berjumlah sangat besar untuk menemukan pola dan aturan yang berarti." (Berry, 1997)

Dari definisi di atas dapat disimpulkan bahwa *data mining* merupakan suatu proses pencarian pola dengan menggunakan teknik statistik dan matematik dari *record* yang berjumlah sangat besar yang dapat memberikan manfaat bagi pengambil keputusan.

Data mining membantu perusahaan untuk mendapatkan pola dari data-data yang tersimpan di dalam basis data perusahaan. Pengetahuan yang diperoleh tersebut akan menjadi pedoman dalam mengambil tindakan-tindakan bisnis sebagai upaya pemeliharaan dan peningkatan tingkat kompetitif bisnis perusahaan. Walaupun sudah banyak perangkat lunak yang menawarkan kemampuan dalam proses *data mining*, keterlibatan manusia sangat dibutuhkan dalam setiap fase proses *data mining* itu sendiri. Pemahaman terhadap model statistik dan matematik yang digunakan dalam perangkat lunak sangat dituntut demi keberhasilan penerapan *data mining*.

Banyak teknik *data mining* yang dapat digunakan untuk mendapatkan pola atau pengetahuan. Salah satu teknik yang banyak digunakan untuk pengklasifikasian adalah *decision tree*.

Decision Tree

Decision tree merupakan salah satu teknik yang dapat digunakan untuk melakukan klasifikasi terhadap sekumpulan objek atau *record*. Teknik ini terdiri dari kumpulan *decision node*, dihubungkan oleh cabang, bergerak ke bawah dari *root node* sampai berakhir di *leaf node*. Pengembangan *decision tree* dimulai dari *root node*, berdasarkan konvensi ditempatkan di bagian atas diagram *decision tree*, semua atribut dievaluasi pada *decision node*, dengan tiap *outcome* yang mungkin menghasilkan cabang. Tiap cabang dapat masuk baik ke *decision node* yang lain ataupun ke *leaf node*.

Persyaratan yang harus dipenuhi dalam penerapan algoritma *decision tree*:

1. Algoritma *decision tree* merepresentasikan *supervised learning*, dan oleh karena itu membutuhkan variabel target *preclassified*. *Training data set* harus menyediakan nilai-nilai variabel target.
2. *Training data set* harus kaya dan bervariasi, menyediakan algoritma dengan *cross section* yang sehat dari tipe *record* dimana klasifikasinya yang mungkin dibutuhkan dimasa datang. *Decision tree* belajar melalui contoh, dan jika contoh yang tersedia secara sistematis kurang memiliki subset *record* yang *definable*, klasifikasi dan prediksi untuk subset tersebut akan sulit atau tidak mungkin.
3. Kelas atribut target harus diskrit. Variabel target harus mempunyai nilai yang jelas batasnya apakah termasuk ke dalam kelas tertentu atau bukan.

Decision tree mencoba membangun sekumpulan *pure leaf node*, tiap *record* dalam *leaf node* tersebut mempunyai klasifikasi yang sama. *Decision tree* menyediakan cara pengklasifikasian dengan ukuran *confidence* yang tertinggi.

Strategi pengembangan *decision tree* dengan menggunakan algoritma C5.0 adalah sebagai berikut:

1. Pada tahap awal, *tree* digambarkan sebagai *node* tunggal yang merepresentasikan *training data*.
2. Jika sampel seluruhnya berisi kelas yang sama, maka *node* tersebut menjadi *leaf* dan dilabeli dengan kelas tersebut.
3. Jika tidak, algoritma dengan menggunakan ukuran berbasis entropi (*information gain*) akan memilih atribut yang akan memisahkan sampel ke dalam kelas-kelas individual. Atribut tersebut menjadi atribut tes atau keputusan pada *node* tersebut.
4. Cabang dikembangkan untuk tiap nilai yang diketahui dari atribut tes, dan sampel dipartisi berdasarkan cabang tersebut.
5. Algoritma menggunakan proses yang sama secara rekursif membentuk *decision tree* untuk sampel-sampel pada tiap partisi. Jika suatu atribut telah muncul pada suatu *node*, maka atribut tersebut tidak perlu dipertimbangkan pada *node* turunan.
6. Partisi rekursif berakhir hanya ketika satu dari kondisi-kondisi berikut terpenuhi:
 - a. Seluruh sampel pada *node* tertentu memiliki kelas yang sama.
 - b. Tidak ada atribut yang tersisa pada sampel yang dapat dipartisi lebih lanjut. Dalam kasus ini suara mayoritas digunakan. *Node* tersebut menjadi *leaf node* dan dilabeli dengan kelas yang menjadi mayoritas dalam sampel yang ada.
 - c. Tidak ada sampel untuk cabang atribut tes. Dalam kasus ini, *leaf* terbentuk dengan mayoritas kelas sebagai label sampel tersebut.

Information Gain

Misalkan sebuah variabel X mempunyai k nilai yang mungkin dengan peluang p_1, p_2, \dots, p_k . Jumlah bit minimum, rata-rata per simbol, yang dibutuhkan untuk mengantarkan *stream* simbol yang merepresentasikan nilai X disebut entropi dari X . Entropi dari X diperoleh dari persamaan berikut:

$$H(X) = - \sum_j p_j \log_2(p_j) \quad (1)$$

Konsep entropi di atas digunakan untuk memilih kandidat *split* terbaik. Misalkan ada kandidat *split* sebanyak S , yang mempartisi *training data set* T kedalam beberapa subset T_1, T_2, \dots, T_k . Kebutuhan rata-rata informasi tiap kandidat *split* dihitung dengan jumlah terbobot dari entropi untuk subset individual.

$$H_S(T) = \sum_{i=1}^k P_i H_S(T_i) \quad (2)$$

$H_S(T)$ = entropi dari T untuk kandidat S

$H_S(T_i)$ = entropi dari T_i untuk kandidat S

P_i = proporsi *record* dalam subset i (T_i) terhadap jumlah *record* dalam *training data* (T)

Information gain dapat didefinisikan sebagai selisih dari entropi dari T dengan entropi dari T untuk kandidat S .

$$\text{Gain}(S) = H(T) - H_S(T) \quad (3)$$

Information gain mencerminkan peningkatan informasi yang dihasilkan dengan pemecahan *training data* T berdasarkan kandidat split S. Pada tiap *decision node*, C5 memilih split optimal yaitu split yang mempunyai *information gain*, gain (S) terbesar.

Pengembangan Model Credit Scoring dengan Teknik Decision Tree

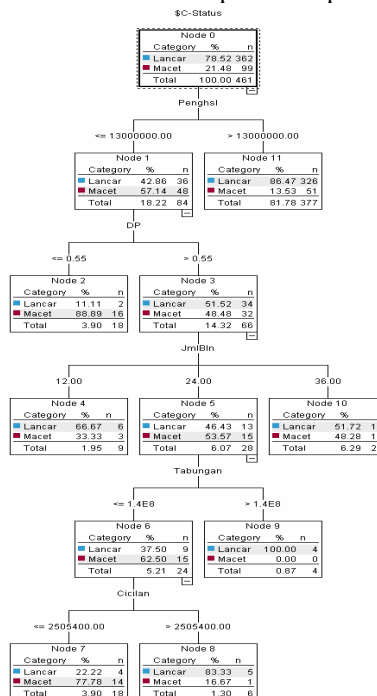
Model *credit scoring* yang akan dikembangkan adalah model yang memprediksi risiko kredit mobil sebagai kredit lancar dan kredit macet. Model ini akan digunakan pada tahap skoring kredit. Setiap pemohon kredit diharuskan mengisi informasi-informasi yang selanjutnya akan diproses pada tahap skoring kredit. Informasi-informasi yang harus diisi oleh pemohon kredit adalah:

1. Penghasilan, rata-rata penghasilan selama tiga bulan terakhir.
2. Cicilan per bulan, besarnya pembayaran yang harus dilakukan setiap bulan.
3. Uang muka, besarnya pembayaran pertama yang mampu dilakukan oleh pemohon kredit.
4. Jumlah periode pinjaman, jumlah periode pelunasan utang (pilihanannya adalah 12, 24, 36 bulan).
5. Rekening koran/tabungan, rata-rata perputaran keuangan selama tiga bulan terakhir.
6. Umur, umur pemohon kredit.
7. Rekening tagihan telepon, rata-rata tagihan telepon yang harus dibayar dalam tiga bulan terakhir.
8. Rekening tagihan listrik, rata-rata tagihan listrik yang harus dibayar dalam tiga bulan terakhir.

Variabel target dari kasus ini adalah risiko kredit dan variabel inputnya adalah 8 variabel yang telah disebutkan di atas. Sebelum model dikembangkan, sampel yang representatif yaitu debitur yang sudah melunasi kewajiban-kewajibannya dikumpulkan. Jumlah sampel yang tersedia sebanyak 750 record. Sampel tersebut dikelompokkan ke dalam dua kelas yaitu kelas lancar dan kelas macet. Karena *decision tree* mengharuskan variabel bersifat diskrit, maka variabel-variabel kontinyu didiskritkan terlebih dahulu.

Hasil Pemodelan

Pemodelan prediksi status kredit dibantu dengan software Clementine. Algoritma yang digunakan adalah algoritma C5.0. Algoritma C5.0 bekerja dengan melakukan *splitting* pada sampel berdasarkan variabel yang memberikan *information gain* tertinggi. Setiap *sub* sampel didefinisikan dengan *first split* kemudian melakukan *split* kembali yang biasanya didasarkan pada variabel yang berbeda dan proses akan terus berulang hingga *sub* sampel tidak dapat dilakukan *split*. Kelebihan dari algoritma C5.0 yaitu *quite robust* pada kehadiran masalah seperti *missing data* dan jumlah yang besar dari input variabel. Model yang dihasilkan oleh algoritma C5.0 lebih mudah dipahami daripada model lainnya untuk diinterpretasikan. Model prediksi yang dihasilkan berdasarkan dari *training data* sebanyak 60% dari total sampel yang ada. Model tersebut dapat dilihat pada gambar 1 berikut ini.



Gambar 1. Model Tree Prediksi Status Kredit

Model di atas menunjukkan bahwa tingkat kepentingan atau bobot penghasilan mempunyai tingkat kepentingan yang lebih tinggi dibandingkan dengan variabel lain dalam memprediksi status kredit, disusul dengan

variabel uang muka, jumlah bulan, tabungan, dan cicilan. Walaupun model ini dapat menunjukkan urutan tingkat kepentingan dari tiap variabel, tetapi model ini tidak dapat menunjukkan besarnya bobot tiap variabel.

Validasi Model

Model yang dihasilkan divalidasi dengan menggunakan *validation data* sebanyak 40% dari total sampel. Validasi dilakukan dengan membandingkan hasil prediksi risiko kredit dengan model dan status kredit aktual. Tabel 2 berikut merangkum hasil validasi model.

Tabel 2. Hasil Validasi

Aktual Prediksi	Lancar	Macet
Lancar	137	33
Macet	5	11

Berdasarkan tabel di atas dapat diketahui bahwa terdapat 148 *record* (79.57%) yang memiliki hasil validasi yang benar dari 186 sampel. Sedangkan untuk hasil validasi yang kurang tepat terdapat 20.43 % dimana terdapat 38 *record* yang memiliki perbedaan antara hasil prediksi dan aktual. Hasil prediksi di atas juga menunjukkan bahwa sebesar 19.4 % yang semula diprediksi memiliki status kredit lancar ternyata memiliki status kredit macet.

Aturan Keputusan

Berdasarkan model *decision tree*, aturan keputusan prediksi risiko kredit yang lebih mudah digunakan dapat dibangun. Berikut aturan keputusan prediksi risiko kredit:

- Rule 1 for Lancar
 if Penghasilan > 13000000
 then Lancar
- Rule 2 for Lancar
 if Penghasilan <= 13000000
 and DP > 55 %
 and Jumlah Bulan = 12
 then Lancar
- Rule 3 for Lancar
 if Penghasilan <= 13000000
 and DP > 55 %
 and Jumlah Bulan = 36
 then Lancar
- Rule 4 for Lancar
 if Penghasilan <= 13000000
 and DP > 55 %
 and Jumlah Bulan = 24
 and Rekening koran / Tabungan > 140.911.945
 then Lancar
- Rule 5 for Lancar
 if Penghasilan <= 13000000
 and DP > 55 %
 and Jumlah Bulan = 36
 and Rekening koran / Tabungan <= 140.911.945
 and Cicilan per Bulan > 2.505.400
 then Lancar
- Rule 6 for Macet
 if Penghasilan <= 13000000
 and DP <= 55%
 then Macet
- Rule 7 for Macet
 if Penghasilan <= 13000000
 and DP > 55%
 and Jumlah Bulan = 24
 and Rekening koran / Tabungan <= 140.911.945
 and Cicilan per Bulan <= 2.505.400
 then Macet

Aturan keputusan di atas dapat menunjukkan karakteristik pemohon kredit. Pemohon kredit yang diprediksi menjadi debitur lancar mempunyai 5 karakteristik dan yang diprediksi menjadi debitur macet mempunyai 2 karakteristik. Misalkan aturan pertama menunjukkan karakteristik dari pemohon kredit adalah pemohon yang mempunyai penghasilan di atas 13 juta rupiah. Masing-masing karakteristik mempunyai peluang menjadi debitur yang baik (kredit lancar) dan buruk (kredit macet) dengan peluang seperti yang terlihat pada tabel 3 berikut ini.

Tabel 3. Peluang Kredit Lancar dan Macet Tiap Karakteristik Pemohon Kredit

Rule	Peluang Lancar	Peluang Macet	Jumlah Sampel
1	0.86	0.14	377
2	0.67	0.33	9
3	0.52	0.48	29
4	1	0	4
5	0.83	0.17	6
6	0.11	0.89	18
7	0.22	0.78	18

Dilihat dari jumlah sampel yang ada pada tiap aturan, aturan 1 dengan karakteristik konsumen yang berpenghasilan di atas 13 juta rupiah lebih berarti dibandingkan dengan aturan yang lain. Semakin banyak jumlah sampel yang ada pada aturan, semakin besar keyakinan yang diberikan pada saat aturan tersebut digunakan.

Analisis Keterbatasan Model *Credit Scoring* dengan *Decision Tree*

Kualitas model yang dihasilkan sangat tergantung pada ketersediaan dan kualitas data. *Decision tree* dapat memberikan model yang akurat jika *training data* yang tersedia berjumlah sangat banyak dan bervariasi. Pada kasus ini jumlah *training data* relatif sedikit karena adanya keterbatasan akses data perusahaan yang lengkap. Walaupun demikian, keakuratan model tidak buruk jika dibandingkan dengan keakuratan model *tree* yang telah diteliti oleh peneliti sebelumnya. Tabel 4 menunjukkan perbandingan tingkat akurasi model *tree*.

Tabel 4. Perbandingan Tingkat Akurasi Model *Tree* (dalam persen)

Authors	Class Trees
Henley	43.8
Boyle et al	75
Srinivisan and Kim	93.2
Yobas et al	62.3
Desai et al	67.3
Chye et al	74.2
Hasil Penelitian	79.57

Model yang dibangun didasarkan pada asumsi bahwa data masa lalu dapat memprediksi kejadian di masa datang. Dalam kasus prediksi status kredit, karakteristik debitur baik dan buruk di masa lalu dapat memprediksi status kredit calon debitur baru. Karakteristik ini dapat berubah sepanjang waktu. Oleh karena itu, model harus selalu dimutakhirkan agar sampel data yang dijadikan sebagai dasar pengembangan model dapat mewakili kondisi aktual.

Sampel yang digunakan untuk pengembangan model adalah sampel yang cenderung bias karena hanya melibatkan konsumen yang diterima permohonan kreditnya. Konsumen yang ditolak tidak dapat dijadikan sebagai sampel karena konsumen tersebut tidak mempunyai kesempatan untuk menunjukkan status kreditnya.

Kesimpulan

Model *credit scoring* yang dibangun dengan menggunakan teknik *decision tree* dapat memberikan standar penilaian yang objektif dengan didasarkan pada data. Dari model dapat disimpulkan bahwa variabel penghasilan mempunyai bobot yang paling tinggi dalam memprediksi status kredit. Model *credit scoring* merupakan alat bantu dalam mengevaluasi kredit dengan tujuan untuk mengurangi subjektivitas penilai. Tingkat akurasi keseluruhan dari model yang dibangun sebesar 79.57 persen.

Decision tree untuk *credit scoring* memberikan informasi mengenai karakteristik pemohon kredit yang diprediksi memiliki status kredit lancar dan macet. Ada 5 karakteristik pemohon kredit yang diprediksi memiliki status kredit lancar, dan 2 karakteristik pemohon kredit yang dipredisiki memiliki status kredit macet. Tiap karakteristik mempunyai peluang yang berbeda-beda untuk menjadi debitur yang berstatus kredit lancar atau macet.

Daftar Pustaka

- Berry, M.J.A., dan Linoff, G., (1997), "*Data Mining Techniques: For Marketing, Sales, and Customer Support*", John Wiley & Sons
- Cyhe, K.H., Chin, T.W., dan Peng, G.C., (2004), "Credit Scoring Using Data Mining Techniques", *Singapore Management Review*, 26 (2) pp. 25-47
- Glassman, C.A., dan Wilkins, H.M., (1997), "Credit Scoring: Probabilities and Pitfalls", *Journal of Retail Banking Service*, 19 (2) pp. 53-56
- Han, J., dan Kamber, M., (2001), "*Data Mining: Concepts and Techniques*", Academic Press
- Isaac, F., (2006), "Small Business Credit Scoring", *Business Credit*, 108 (3) pp. 20-210
- Larose, D.T., (2005), "*Discovering Knowledge in Data: An Introduction to Data Mining*", Wiley-Interscience
- Noe, J., (1997), "Credit Scoring", *America's Community Banker*, 6 (8) pp. 29-33
- Thomas, L.C., Oliver, R.W., dan Hand D.J., (2005), "A Survey of the Issue in Consumer Credit Modelling Research", *Journal of the Operational Research Society*, 56 pp. 1006-1015