

IMPLEMENTASI ALGORITME *NAÏVE BAYES* BERBASIS *PARTICLE SWARM OPTIMIZATION* DALAM PENENTUAN PEMBERIAN KREDIT

Muhammad Husni Rifqo¹, Nuri David Veronica²

^{1,2}Program Studi Informatika, Fakultas Teknik, Universitas Muhammadiyah Bengkulu
Jl. Bali PO BOX 118. Telp (0736) 227665, Fax (0736) 26161, Bengkulu 38119

¹mhrifqo@umb.ac.id

²misz_nuricuby@gmail.com

Abstrak: Sewa guna usaha (*leasing*) atau sering disingkat SGU adalah kegiatan pembiayaan dalam bentuk penyediaan barang modal baik secara sewa guna usaha dengan hak opsi (*finance lease*) maupun sewa guna usaha (*lessee*) selama jangka waktu tertentu berdasarkan pembayaran secara angsuran, kegiatan ini bisa dikatakan sebagai kegiatan kredit, kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam meminjam antara lembaga keuangan dengan pihak lain yang mewajibkan pihak peminjam melunasi utangnya setelah jangka waktu tertentu dengan pemberian bunga. Pengajuan aplikasi kredit oleh calon pelanggan sekarang sangatlah mudah, hal ini dikarenakan pengajuan kredit bisa dilakukan oleh semua orang sepanjang memenuhi syarat tertentu. Persaingan perusahaan penyedia kredit menjadi sangat pesat dan prediksi konsumen kredit adalah hal yang sangat penting. Dari permasalahan ini diperlukan suatu model yang mampu mengklasifikasikan sekaligus memprediksi pelanggan mana saja yang bermasalah dan tidak bermasalah. Model *Naïve Bayes* berbasis *Particle Swarm Optimization* (PSO) ternyata mampu meningkatkan akurasi dalam menganalisa kelayakan kredit, semakin besar data set yang digunakan maka akurasi model *Naïve Bayes* berbasis *Particle Swarm Optimization* (PSO) akan semakin meningkat, akurasi yang didapat oleh model ini untuk data set yang digunakan adalah: Agiing 2010 (96,75%), Agiing 2011 (97,95%), Japan credit approval (84,77%) dan Australia credit approval (87,83%).

Kata Kunci: Analisa kredit, penilaian kredit, *Naïve Bayes* berbasis PSO.

Abstract: *Lease (leasing) or often abbreviated as SGU is a financing activity is the provision of both capital goods lease with option rights (finance lease) and lease (lessee) for a certain period based on payment in installments, this activity can be said the activities of credit, credit is the provision of money or bills that can be equated with it, based on an agreement between a lending financial institution to another party requiring the borrower to pay off the debt after a certain period with interest. Submission of loan applications by prospective customers are now very easy, this is because the loan application can be done by all people at all certain eligible. Competition credit providers to be very rapid and predictions of consumer credit is very important. Of this problem requires a model that is able to classify as well as predict which customers are problematic and not problematic. Naive Bayes Model based Particle Swarm Optimization (PSO) was able to increase the accuracy in analyzing*

the credit worthiness, the larger the dataset used the Naive Bayes model accuracy based Particle Swarm Optimization (PSO) will increase, the accuracy obtained by this model to the data set used are: Agiing 2010 (96.75%), Agiing 2011 (97.95%), Japan credit approval (84.77%) and Australia credit approval (87.83%).

Keywords: *Credit analysis, credit assessment, Naive Bayes based PSO.*

I. PENDAHULUAN

Dunia teknologi informasi berkembang sangat pesat, hal ini memudahkan orang dalam mendapatkan informasi secara mudah dan cepat. Perkembangan teknologi informasi ini juga mengakibatkan banyaknya terjadi perubahan dalam beberapa bidang, contohnya dalam bidang

perkreditan. Kegiatan kredit saat sekarang tumbuh dengan marak, tampak pada bermunculannya perusahaan-perusahaan penyedia jasa kredit yang fungsinya dapat memudahkan calon nasabah dalam memenuhi kebutuhannya. Sewa guna usaha (*leasing*) atau sering disingkat SGU adalah kegiatan pembiayaan dalam bentuk penyediaan barang modal baik secara sewa guna usaha dengan hak opsi (*finance lease*) maupun sewa guna usaha (*lessee*) selama jangka waktu tertentu berdasarkan pembayaran secara angsuran [1].

Kegiatan ini bisa dikatakan sebagai kegiatan kredit, kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam meminjam antara lembaga keuangan dengan pihak lain yang mewajibkan pihak peminjam melunasi utangnya setelah jangka waktu tertentu dengan pemberian bunga. Sedangkan Kredit Tanpa Jaminan atau Kredit Tanpa Agunan (KTA) adalah kredit yang diberikan bank atau lembaga keuangan dalam bentuk uang tunai, yang dapat diperoleh tanpa memberikan jaminan. Kemudahan ini banyak dimanfaatkan oleh para nasabah untuk meminjam sejumlah dana dari koperasi untuk digunakan berbagai keperluan nasabah [2].

Pengajuan aplikasi kredit oleh calon pelanggan sekarang sangatlah mudah, hal ini dikarenakan pengajuan kredit bisa dilakukan oleh semua orang sepanjang memenuhi syarat tertentu. Persaingan perusahaan penyedia kredit menjadi sangat pesat dan prediksi konsumen kredit adalah hal yang sangat penting [3].

Kegiatan kredit pada perusahaan sangat dipengaruhi oleh banyak faktor baik dari dalam maupun dari luar, salah satu faktor penghambatnya adalah masalah kredit yang diajukan oleh calon pelanggan memiliki risiko. Risiko kredit adalah

kemungkinan penurunan hasil kredit dari tindakan peminjam yang mempunyai reputasi yang buruk [4], risiko kredit mengancam sistem ekonomi perusahaan [5]. Oleh karena itu dibutuhkan seorang pengambil keputusan pada suatu perusahaan sewa guna usaha (*leasing*) yang mampu mengambil keputusan yang tepat dalam menganalisa kredit.

Dengan pesatnya pertumbuhan di bidang ini maka model penilaian kredit bisa digunakan untuk evaluasi [6], penilaian kredit adalah himpunan model dan teknik pengambilan keputusan yang mendasari kreditur dalam pemberian kredit kepada konsumen [7]. Penilaian kredit (*credit scoring*) adalah alat yang dapat membantu perusahaan meningkatkan kualitas dan menurunkan risiko kredit [8], dan merupakan metode yang efektif dari manajemen risiko kredit [9]. Penilaian kredit menjadi isu yang sangat penting dalam pertumbuhan industri kredit yang sangat cepat, sehingga lembaga keuangan menghadapi data kredit dengan jumlah yang besar. Teknik ini telah digunakan secara luas oleh lembaga keuangan atau perusahaan dalam menentukan apakah pelanggan termasuk kedalam kelas pemohon yang baik atau yang buruk.

Keuntungan menggunakan model penilaian kredit [10] adalah:

1. Mengurangi biaya analisis kredit
2. Memungkinkan keputusan kredit lebih cepat
3. Mengurangi risiko kredit

Untuk menganalisis data dalam jumlah yang besar inilah digunakannya teknik *data mining* [11], *data mining* digunakan untuk menemukan pola atau anomali dalam sejarah data kredit untuk membuat model pengklasifikasian dan mengenali masalah baru [10].

Dengan klasifikasi yang efektif perusahaan dapat membagi pelanggan ketingkat yang berbeda oleh informasi mereka sehingga perusahaan dapat memutuskan apakah memberikan pinjaman kepada pelanggan sesuai dengan klasifikasi [12], klasifikasi pelanggan ini dibedakan berdasarkan atribut mereka seperti usia, status perkawinan, pendapatan dan lain lain [13]. Perusahaan sangat membutuhkan teknik prediksi dalam mengambil suatu keputusan kredit, banyak penelitian yang dilakukan untuk permasalahan risiko kredit dengan menggunakan bermacam metode, misalnya metode neural network, *Classification And Regression Tree* (CART), *Naïve Bayes*, *C4.5*, *Support Vector Machine* (SVM), dan metode lain, metode itu dipilih sesuai dengan data dan tujuannya.

Setiap metode *data mining* yang digunakan memiliki kelebihan masing-masing, oleh karena itu analisa kredit sangat penting dalam perusahaan kredit. Jelas tidak mudah untuk menganalisa data yang jumlahnya besar [11] baik dalam hal ekonomi dan tenaga kerja, ditambah dengan atribut dan variabel yang sangat banyak dalam data tersebut.

Banyak metode yang diusulkan dalam penelitian kredit, dalam penelitian Heiat [7] data mining digunakan untuk memprediksi dan mengkomparasi kinerja dari tiga metode klasifikasi yaitu *Tree Augmented Naïve* (TAN) *Bayes*, *Markov Bayes* dan *Markov-Feature Selection* (FS) untuk membedakan kredit yang baik dan kredit yang buruk, hasil penelitian menunjukkan model Markov lebih baik dalam memprediksi masalah tersebut.

Kemudian Yu and Wang [14] mengusulkan algoritme penilaian kredit dengan menggunakan *Genetic Algorithm* berbasis *Bayesian Network*, hasil eksperimen terbagi dua yaitu untuk menguji

kinerja metode yang diusulkan dan yang kedua adalah membandingkan kinerja dengan metode yang lain. Hasilnya metode *Naïve Bayes* berbasis GA memiliki persentase yang lebih baik dalam hal *correctly predicted* 18,992 dan *correct rate* 53.23%. Lalu Islam et al [15] menyebutkan teori probabilitas adalah kerangka untuk membuat keputusan dibawah ketidakpastian, tujuan penelitian ini membandingkan kinerja dari pengklasifikasi *Naïve Bayes* dengan *K-Nearest Neighbor* (KNN) dalam masalah persetujuan pengajuan aplikasi kartu kredit.

Jiang and Wu [10] mengusulkan sebuah pendekatan baru untuk penilaian kredit dengan membuat sintesis *Simple Naïve Bayes Classifier* (SNBC) dan teori *Rough Set*. Percobaan dilakukan dengan menggunakan dataset kredit Jerman, SNBC adalah salah satu model Bayesian yang secara luas diakui sebagai metode sederhana dan efisien tetapi memiliki akurasi yang kurang baik. Hal ini disebabkan oleh dataset yang digunakan memiliki gangguan atau redundansi yang mempengaruhi kinerja dari pengklasifikasi. Dalam penelitian Ling et al [3] mengusulkan model *hybrid* baru dengan menggunakan *Support Vector Machine* dan *Particle Swarm Optimization* (PSO) untuk evaluasi penilaian kredit dan menggunakan dua dataset kredit Jerman dan Australia yang diambil dari UCI dataset untuk pengujian metode ini. Hasil akurasi menunjukkan bahwa pada dataset Australia mendapatkan akurasi sebesar 87,10% dan 78,70 % untuk dataset Jerman.

Dari beberapa penelitian yang pernah dilakukan belum ada peneliti yang menggunakan algoritme *Naïve Bayes* dan PSO untuk penilaian kelayakan kredit. Dan nampak bahwa akurasi yang dimiliki oleh *Naïve Bayes* sudah baik tetapi menurut [16] masih di bawah *good classification*.

Dari penelitian di atas maka peneliti menggunakan algoritme pengklasifikasi *Naïve Bayes* dan PSO untuk menentukan pelanggan yang baik dan buruk dalam analisa pemberian kredit mobil di *leasing* ACC.

Permasalahan yang muncul berdasarkan latar belakang di atas adalah adanya hasil analisis yang tidak akurat, seperti pengajuan kredit yang disetujui ternyata berisiko dan bermasalah dalam pembayaran cicilan kredit.

Tujuan dari penelitian ini adalah untuk mengetahui tingkat kredit macet yang diakibatkan oleh pelanggan, dan menerapkan algoritme *Naïve Bayes* dan *Particle Swarm Optimization* (PSO) untuk penentuan kelayakan pemberian kredit mobil dengan lebih akurat agar diperoleh kesimpulan yang dapat dijadikan dasar pengambilan keputusan terhadap kelayakan pemberian kredit tersebut.

Manfaat penelitian ini adalah:

1. Manfaat praktis dari penelitian ini adalah memudahkan analisis kredit dalam memberikan keputusan.
2. Manfaat teoritis penelitian ini adalah memberikan sumbangan penerapan model algoritme *Naïve Bayes* berbasis *Particle Swarm Optimization* (PSO).
3. Manfaat kebijakan dalam penelitian ini adalah agar algoritme *Naïve Bayes* berbasis *Particle Swarm Optimization* (PSO) menjadi alat pendukung keputusan dalam menganalisa kredit.

Batasan dalam penelitian ini adalah:

1. Data yang digunakan adalah data Agiing kredit perusahaan ACC dari tahun 2010 sampai dengan tahun 2011.
2. Untuk perbandingan akurasi model yang diusulkan maka penelitian ini juga

menggunakan *public data set* yang ada di UCI *data set*, yaitu data persetujuan kredit (*credit approval*) negara Japan dan Australia.

3. Pada pengujian model digunakan algoritme klasifikasi *Naïve Bayes* dan metode PSO

II. TINJAUAN PUSTAKA

A. Penelitian Terkait

Ada beberapa penelitian terdahulu yang terkait dengan tema penilaian kredit, secara garis besar tinjauan pustaka dalam penelitian ini meliputi:

a. Jiang and Wu [10] mengusulkan sebuah pendekatan baru untuk penilaian kredit dengan membuat sintesis *Simple Naïve Bayes Classifier* (SNBC) dan teori *Rough Set*. Percobaan dilakukan dengan menggunakan dataset kredit Jerman, SNBC adalah salah satu model Bayesian yang secara luas diakui sebagai metode sederhana dan efisien tetapi memiliki akurasi yang kurang baik. Hal ini disebabkan oleh dataset yang digunakan memiliki gangguan atau redundansi yang mempengaruhi kinerja dari pengklasifikasi.

b. Yu and Wang [14] mengusulkan algoritme penilaian kredit dengan menggunakan *Genetic Algorithm* berbasis *Bayesian Network*, hasil eksperimen menggunakan *Genetic Algorithm* terbagi dua yaitu untuk menguji kinerja metode yang diusulkan dan yang kedua adalah membandingkan kinerja dengan metode yang lain. Hasilnya metode *Naïve Bayes* berbasis GA memiliki persentase yang lebih baik dalam hal *correctly predicted* dan *correct rate* seperti tabel di bawah ini:

Tabel 1. Perbandingan Kinerja

<i>Methods</i>	<i>Predicted</i>	<i>Correctly predicted</i>	<i>Correct rate</i>
<i>Naïve Bayesian</i>	59,518	17,890	50,14%
<i>Mutual Information</i>	50,261	17,648	49,46%
<i>GA based</i>	46,195	18,992	53,23%

c. Islam et al. [15] menyebutkan teori probabilitas adalah kerangka untuk membuat keputusan dibawah ketidakpastian, tujuan penelitian ini membandingkan kinerja dari pengklasifikasi *Naïve Bayes* dengan KNN. Kesimpulan dari penelitian ini adalah:

1. Hasil penilaian Bayesian sangat bergantung pada probabilitas.
2. Pengklasifikasi *Naïve Bayes* yang diterapkan menemukan 12,43% kesalahan dari klasifikasi.
3. Contoh metode ini kadang-kadang disebut sebagai metode pembelajaran “Lazy” atau malas karena menunda proses sampai contoh baru harus diklasifikasikan.
4. Pada metode KNN, pemilihan *k-values* rumit dan tergantung dari aplikasi, klasifikasi kesalahan adalah 9,45%.

d. Huang et al. [17] melakukan komparasi tiga metode klasifikasi yaitu: *Naïve Bayes*, *Decision Tree* dan SVM untuk mencari metode mana yang akurasi paling bagus dengan menggunakan 13 dataset (data kredit, data mobil, data rumah sakit dan lain-lain). Hasil akurasi dari prediksi tersebut adalah *Naïve Bayes* 85,9%, *Decision Tree* 86,5% dan SVM 86,7% dan kurva ROC menunjukkan bahwa ketiga metode tersebut tidak ada perbedaan.

e. Lin and Yu [18] dalam penelitiannya melakukan klasifikasi dan perbandingan algoritme *Naïve Bayes* dengan *Naïve Bayes* berbasis PSO, algoritme PSO ini digunakan sebagai optimasi untuk menghitung bobot optimal setiap *record* dan rata-rata bobot seluruh sampel sehingga mendapatkan rata-rata bobot yang optimal. Penelitian ini menggunakan data set dari UCI, hasil dari eksperimen tersebut dapat dilihat pada tabel di bawah ini:

Tabel 2. Hasil Eksperimen

<i>Dataset</i>	<i>Num.of instance</i>	<i>Num. of classification</i>	<i>Num. of atributes</i>	<i>Cor.rate NBC</i>	<i>Corr rate PSOWNBC</i>
Bezdekiris	150	3	4	0.96	0.966
Wine	178	3	13	0.988	0.988
Transfusio n	748	2	4	0.751	0,771
Movement libras	360	15	90	0.752	0.758

B. Landasan Teori

1. Kredit

Undang-undang Republik Indonesia No.10 Tahun 1998 [2] menjelaskan kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam meminjam antara lembaga keuangan dengan pihak lain yang mewajibkan pihak peminjam melunasi utangnya setelah jangka waktu tertentu dengan pemberian bunga. Jasa layanan kredit yang diberikan oleh ACC adalah:

1. New Car Financing

Astra Car : Toyota, Daihatsu, Isuzu, Peugeot, BMW, Nissan Diesel Non-Astra Car.

2. Used Car Financing

ACC Used Car Distribution channels: Used Car Dealer, Direct to Customer (C2C program).

3. Retail Commercial Financing

Fasilitas pembiayaan untuk truck & pick up untuk bisnis Usaha Kecil Menengah.

4. Fleet Financing

Pembiayaan untuk *corporate*: transportasi, perminyakan, Agro dan lain-lain.

5. Other Products/Services

Astra Credit Protection (ACP), ACC Gold Card, FASTrack Financing, Guarantee Assets Protection (GAP), STNK Services, Direct to Customer (insurance).

1.1 Risiko kredit

Risiko kredit adalah kemungkinan penurunan hasil kredit dari tindakan peminjam yang mempunyai reputasi yang buruk [4], risiko kredit mengancam sistem ekonomi perusahaan [5]. Risiko kredit bisa diminimalisasi oleh perusahaan dengan menganalisa aplikasi dari pemohon kredit dengan tepat dan akurat, pengambilan keputusan yang kurang tepat berdampak pada kegiatan perkreditan diperusahaan tersebut.

1.2 Penilaian kredit

Heiat mengatakan [7] penilaian kredit adalah himpunan model dan teknik pengambilan keputusan yang mendasari kreditur dalam dalam pemberian kredit kepada konsumen. Penilaian kredit merupakan alat yang ampuh yang dapat membantu perusahaan meningkatkan kualitas kredit dan menurunkan risiko kredit [8]. Penilaian kredit merupakan teknik untuk mengklasifikasi pelanggan yang baik dan pelanggan yang buruk.

Zurada [19] menyatakan bahwa ada beberapa atribut pada data konsumen kredit seperti umur, banyaknya kredit, penjamin, jangka kredit, status pekerjaan, lama bekerja, status pernikahan, status rumah, histori kredit dan lain-lain. Tetapi dalam penelitian ini penulis menggunakan beberapa atribut penting (Tabel 3) untuk dijadikan ukuran dalam menganalisis calon konsumen/ debitur.

2. Data set

Data set yang digunakan dalam penelitian ini ada dua, yaitu:

1. Private Data set

Data yang didapat dari *leasing ACC* dari tahun 2010 sampai dengan 2011.

2. Public Data set

Data yang diambil dari UCI data set, yaitu data kredit dari negara Japan dan Australia.

Tabel 3. Atribut data debitur

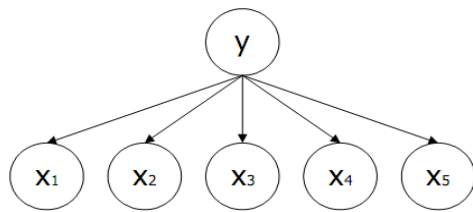
Nama Atribut	Keterangan
Over_due	Nilai Kontinu
Angka ke	Nilai Kontinu
Nilai Angka	Nilai Kontinu
Saldo	Nilai Kontinu
Wilayah	Kategorikal
Lokasi	Kategorikal
Nama Kolektor	Kategorikal
Merk	Kategorikal
Type	Kategorikal
Tahun Mobil	Nilai Kontinu
Used-New	Kategorikal
Unit	Nilai Kontinu
DP	Nilai Kontinu
Arisk	Nilai Kontinu
TLO	Nilai Kontinu
Pekerjaan	Kategorikal
Status Kontrak	Kategorikal

2.1 Algoritme Naive Bayes

Pengklasifikasi *Bayesian* adalah pengklasifikasi statistik [23], klasifikasi *Bayesian* didasarkan pada teorema *Bayes*. Teori keputusan *Bayes* adalah [24] pendekatan statistik yang fundamental dalam pengenalan pola (*pattern recognition*), teori *Bayesian* pada dasarnya adalah [15] kemungkinan kejadian di masa depan bisa dihitung dengan menentukan frekuensi pengalaman sebelumnya. Penggunaan algoritme *Bayes* dalam hal klasifikasi harus mempunyai masalah yang bisa dilihat statistiknya, misalkan X adalah set atribut data dan h kelas variabel dan jika kelas memiliki hubungan dengan atribut maka dapat diperlakukan X dan h sebagai variabel acak dan menangkap hubungan peluang $P(h|X)$ ini peluang posterior untuk h dan sebaliknya perior $P(h)$.

2.2 Struktur Naive Bayes Classifier

Struktur jaringan *Bayesian* harus diselesaikan atau dipelajari sebelum memperkirakan parameter [8], dengan asumsi bahwa variabel independen bersyarat diberi label kelas, sebagai ditunjukkan pada gambar di bawah ini:



Gambar 1. Hubungan variabel dalam Naive Bayes

Tahap-tahap algoritme *Naive Bayes*:

1. Menyiapkan data *training*
2. Setiap data dipresentasikan sebagai vektor berdimensi-n yaitu $X=(x_1,x_2,x_3,\dots,x_n)$
3. n adalah gambaran dari ukuran yang dibuat di *test* dari n atribut yaitu A_1,A_2,A_3,\dots,A_n
4. M adalah kumpulan kategori yaitu C_1,C_2,C_3,\dots,C_m
5. Diberikan data test X yang tidak diketahui kategorinya, maka *classifier* akan memprediksi bahwa X adalah milik kategori dengan posterior probability tertinggi berdasarkan kondisi X
6. *Naive Bayes classifier* menandai bahwa test X yang tidak diketahui tadi ke kategori C_i jika dan hanya jika $P(C_i|X) > P(C_j|X)$ untuk $1 \leq j \leq m, j \neq i$

7. Kemudian kita perlu memaksimalkan $P(C_i|X)$

$$P(C_i|X) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)} \quad (2)$$

8. Dimana x adalah nilai-nilai atribut dalam sampel X dan probabilitas $P(x_1/C_i), P(x_2/C_i), \dots, P(x_n/C_i)$, dapat diperkirakan dari data *training*

Contoh di bawah menggambarkan pendekatan *Bayes* untuk menentukan keputusan main atau tidak [24] dengan atribut cuaca, temperatur, kelembaban dan angin:

Tabel 1. Data cuaca dan keputusan bermain

Cuaca x1	Temperatur x2	Kelembaban x3	Angin x4	Main/tidak x5
cerah	Panas	Tinggi	kecil	Tidak
cerah	Panas	Tinggi	besar	Tidak
mendung	Panas	Tinggi	kecil	Ya
hujan	Sedang	Tinggi	kecil	Ya
hujan	Dingin	Normal	kecil	Ya
hujan	Dingin	Normal	besar	Tidak
mendung	Dingin	Normal	besar	Ya
cerah	Sedang	Tinggi	kecil	Tidak
cerah	Dingin	Normal	kecil	Ya
hujan	Sedang	Normal	kecil	Ya
cerah	Sedang	Normal	besar	Ya
mendung	Sedang	Tinggi	besar	Ya
mendung	Panas	Normal	kecil	Ya
hujan	Sedang	Tinggi	besar	Tidak

$$P(\text{main}) = 9/14 = 0.64$$

$$P(\text{tidak}) = 5/14 = 0.36$$

$$P(\text{Angin} = \text{besar}|\text{main}) = 3/9 = 0.33$$

$$P(\text{Angin} = \text{besar}|\text{tidak}) = 3/5 = 0.60$$

$$P(\text{main}) P(\text{cerah}|\text{main}) P(\text{dingin}|\text{main}) P(\text{tinggi}|\text{main}) P(\text{besar}|\text{main}) = 9/14 * 2/9 * 3/9 * 3/9 * 3/9 = 0.0059$$

$$P(\text{tidak}) P(\text{cerah}|\text{tidak}) P(\text{dingin}|\text{tidak}) P(\text{tinggi}|\text{tidak}) P(\text{besar}|\text{tidak}) = 5/14 * 3/5 * 1/5 * 4/5 * 3/5 = 0.0206$$

Sehingga probabilitasnya adalah $\frac{0.0206}{0.0206+0.0059} = 0.777$

Probabilitas posterior:

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

2.3 Particle Swarm Optimization

Algoritme *Particle Swarm Optimization* (PSO) diperkenalkan oleh Kennedy dan Eberhart pada tahun 1995 [16]. PSO didasarkan pada [20] perilaku sebuah kawanan serangga, seperti semut, rayap, lebah atau burung. PSO diterapkan untuk berbagai masalah yang besar untuk mencari solusi yang optimal, persamaan hitungan dari PSO seperti di bawah ini:

$$\begin{aligned}
 Vi_d &= w.Vi_d + c_1.rand(Pi_d - Xi_d) \\
 &\quad + c_2.rand(Pg_d - Xi_d) \\
 Xi_d &= Xi_d + Vi_d \\
 -V_{max} &\leq Vi_d \leq V_{max}
 \end{aligned} \tag{4}$$

2.4 Particle Swarm Optimization – Naive Bayes

Sesuai dengan fungsinya PSO dapat menemukan bobot yang optimal, PSO digunakan untuk menghitung bobot optimal untuk setiap record dan rata-rata seluruh sample.

Tahap-tahap algoritme Particle Swarm Optimization-Naive bayes

1. Baca sample data set, siap untuk menghitung bobot dari record pertama
2. Menginisialisasi kerumunan $P(t)$ kemudian $t = 0$, lokasi $x_i(t)$ dari setiap partikel $P_i \in P(t)$ di dalam ruang acak
3. Melalui partikel masing-masing $x_i(t)$ evaluasi kinerja dari F
4. Bandingkan kinerja masing-masing individu dengan kinerja terbaik yang dimiliki, if $F(x_i(t)) > pbest_i$, then

$$\begin{cases} pbest_i = F(x_i(t)) \\ x_{pbest_i} = x_i(t) \end{cases} \tag{5}$$

5. Membandingkan kinerja dari setiap partikel dengan kinerja partikel keseluruhan yang terbaik, if $F(x_i(t)) > gbest_i$, then

$$\begin{cases} gbest_i = F(x_i(t)) \\ x_{gbest_i} = x_i(t) \end{cases} \tag{6}$$

Mengubah kecepatan vector dari partikel $V_i(t) = V_i(t-1) + P_1(x_{pbest} - x_i(t)) + P_2(x_{gbest} - x_i(t))$

Dimana P_1 dan P_2 adalah variabel acak. P_1 dan P_2 diidentifikasi sebagai $P_1 = r_1c_1$, $P_2 = r_2c_2$, $r_1, r_2 \sim U(0,1)$ sementara c_1 dan c_2 adalah konstanta percepatan positif.

Menempatkan setiap partikel ke lokasi baru

$$\begin{cases} x_i(t) = x_i(t-1) + v_i(t) \\ t = t + 1 \end{cases} \tag{7}$$

Balik kelangkah 2 sampai mendapatkan

6. Jika perhitungan bobot dari semua record pada data set telah selesai, siap untuk menghitung dari record selanjutnya
7. Hitung rata-rata dari setiap bobot

2.3 Kerangka pemikiran

Kerangka pemikiran merupakan gambaran dari metode penelitian yang akan dilakukan dan berdasarkan CRISP-DM, adapun langkah-langkahnya sebagai berikut: menentukan masalah yang ada pada penelitian yang kemudian mengumpulkan data-data yang diperlukan dalam penelitian (data primer dan data skunder).

Setelah data terkumpul maka melakukan proses awal yang sesuai dengan CRISP-DM dan menentukan model atau algoritme apa yang akan digunakan dalam melakukan eksperimen. Tahapan selanjutnya adalah evaluasi dan validasi model menggunakan confusion matrix dan ROC curve, semakin besar persentase yang didapat dari ROC Curve maka hasil kelayakan pemberian kredit semakin akurat. Gambar 2 ini menjelaskan kerangka pemikiran untuk penentuan kelayakan pemberian kredit mobil.

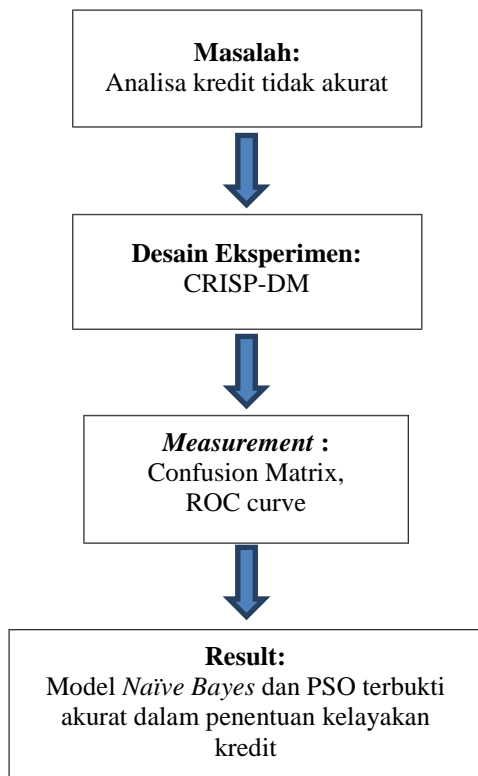
III. METODE PENELITIAN

A. Pengumpulan data

Data yang dikumpulkan ada dua macam yaitu:

1. Data primer

Data primer adalah data utama yang didapat dari objek penelitian yang terbagi dalam dua data yaitu: data agiing leasing ACC dan UCI data set credit approval (data terlampir).



Gambar 2. Kerangka Pemikiran

2. Data skunder

Data skunder adalah merupakan data yang diperoleh dari studi literatur yang dilakukan, sumber dari studi literatur yaitu jurnal, makalah ilmiah dan buku yang membahas tentang penilaian kredit.

B. Eksperimen

Jenis penelitian yang digunakan adalah penelitian eksperimen dengan proses CRISP-DM dan tahapan penelitian sebagai berikut [22]:

1. *Business Understanding*
2. *Data Understanding*
3. *Data Preparation*
4. *Tahap Modelling*

C. Model algoritme Naive Bayes

Pada model ini menghitung jumlah permasalahan *class* baik dan *class* buruk,

permasalahan dibagi berdasarkan atribut (*over due*, angka ke, nilai angka, saldo, wilayah, lokasi, nama kolektor, merk, type, tahun mobil, used-new, unit, DP, arisk, TLO, pekerjaan dan status kontrak. Pada bagian ini akan dilakukan penyeleksian atribut dengan menghitung peluang total baris kolom dengan *Naive Bayes* seperti di bawah ini:

Tabel 5. Contoh data set dan penyelesaiannya

Angka ke	Merk	Tahun	DP	Arisk	Pekerjaan	Status Kontrak
4	Daihatsu	2009	15%	1	PNS	Baik
5	Suzuki	2002	25%	0	Swasta	Baik
4	Daihatsu	2001	22,16%	0	PNS	Buruk
4	Daihatsu	2009	10%	3	PNS	Baik
17	Mitsubishi	2008	40,24	0	Swasta	Buruk
4	Mitsubishi	2009	23,76	1	Swasta	Buruk
4	Daihatsu	2009	15%	4	PNS	Buruk
10	Daihatsu	2009	15%	1	Swasta	Baik
4	Timor	1997	42,11%	0	Swasta	Buruk
10	Mitsubishi	2004	45%	0	PNS	Baik

Dari contoh tabel data training diatas terdapat 7 atribut yaitu: angka ke, nilai angka, merk, tahun, DP, Arisk, pekerjaan dan status kontrak. *Class*nya ada dua *values* yaitu *class* baik dan *class* buruk, untuk menentukan *class* data tersebut perlu dihitung nilai kemungkinannya yang dapat diketahui dari data *training*.

Hitung nilai $P(C_i)$ untuk $i =$ baik dan buruk

$$P(\text{status} = \text{baik}) = 5/10 = 0.5$$

$$P(\text{status} = \text{buruk}) = 5/10 = 0.5$$

$$P(\text{angka ke} = 4 | \text{status} = \text{baik}) = 2/5 = 0.4$$

$$P(\text{angka ke} = 4 | \text{status} = \text{buruk}) = 4/5 = 0.8$$

$$P(\text{merk} = \text{daihatsu} | \text{status} = \text{baik}) = 3/5 = 0.6$$

$$P(\text{merk} = \text{daihatsu} | \text{status} = \text{buruk}) = 2/5 = 0.4$$

$$P(\text{tahun} = 2009 | \text{status} = \text{baik}) = 3/5 = 0.6$$

$$P(\text{tahun} = 2009 | \text{status} = \text{buruk}) = 2/5 = 0.4$$

$$P(\text{DP} = 15\% | \text{status} = \text{baik}) = 2/5 = 0.4$$

$$P(\text{DP} = 15\% | \text{status} = \text{buruk}) = 1/5 = 0.2$$

$$P(\text{Arisk} = 1 | \text{status} = \text{baik}) = 2/5 = 0.4$$

$$P(\text{Arisk} = 1 | \text{status} = \text{buruk}) = 1/5 = 0.2$$

$$P(\text{Pekerjaan} = \text{PNS} | \text{status} = \text{baik}) = 3/5 = 0.6$$

$$P(\text{Pekerjaan} = \text{PNS} | \text{status} = \text{buruk}) = 2/5 = 0.4$$

Berdasarkan hasil peluang diatas, maka dapat dihitung berdasarkan rumus:

$$P(X|\text{status} = \text{baik}) = 0.4 * 0.6 * 0.6 * 0.4 * 0.4 * 0.6 = 0,013824$$

$$P(X|\text{status} = \text{buruk}) = 0.8 * 0.4 * 0.4 * 0.2 * 0.2 * 0.4 = 0,002048$$

Hitung dengan rumus:

$$P(C_i|X) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)} \quad (8)$$

$$P(C_i|X) = \frac{P(0,013824) \cdot P(0.5)}{0.002048} = 3,375$$

$$P(X|\text{status}=\text{baik}) P(\text{status}=\text{baik}) = 0,013824 * 0.5 = 0,006912$$

$$P(X|\text{status}=\text{buruk}) P(\text{status}=\text{buruk}) = 0,002048 * 0.5 = 0,001024$$

Berdasarkan penghitungan di atas maka, sample X akan masuk kategori status = buruk. Hasil perhitungan dapat dilihat di tabel berikut ini:

Tabel 6. Hasil Hitung Probabilitas

No	Catatan	Class/status	
		Baik	Buruk
1	Angka ke	0,4	0,8
2	Merk	0,6	0,4
3	Tahun	0,6	0,4
4	DP	0,4	0,2
5	Arisk	0,4	0,2
6	Pekerjaan	0,6	0,4
b		0,013824	0,002048

D. Model algoritme Naive Bayes berbasis PSO

Dari perhitungan probabilitas pada tabel 3.3.2 di atas terdapat dua class yaitu class baik dan class buruk, yang kemudian akan dihitung dengan menggunakan algoritme Naive Bayes berbasis PSO dengan cara seperti berikut:

$$F = P(X|C_i) - \sum P(X|\sim C_i) \quad (9)$$

1. Class baik

$$f_1 = 0,013824 - 0,8 = -0,78618$$

$$f_2 = 0,013824 - 0,4 = -0,38618$$

$$f_3 = 0,013824 - 0,4 = -0,38618$$

$$f_4 = 0,013824 - 0,2 = -0,18618$$

$$f_5 = 0,013824 - 0,2 = -0,18618$$

$$f_6 = 0,013824 - 0,4 = -0,38618$$

2. Class buruk

$$f_1 = 0,002048 - 0,4 = -0,39795$$

$$f_2 = 0,002048 - 0,6 = -0,59795$$

$$f_3 = 0,002048 - 0,6 = -0,59795$$

$$f_4 = 0,002048 - 0,4 = -0,39795$$

$$f_5 = 0,002048 - 0,4 = -0,39795$$

$$f_6 = 0,002048 - 0,6 = -0,59795$$

Dari hasil perhitungan di atas maka nilai $Pbest_i$ dan nilai $Gbest_i$ sudah bisa ditentukan, kemudian gunakan rumus di bawah ini untuk menghitung kecepatan vektor dari partikel:

$$V_i(t) = V_i(t-1) + P_1(x_{pbest} - x_i(t)) + P_2(x_{gbest} - x_i(t)) \quad (10)$$

1. Class baik

$$V_i(1) = V_i(1-1) + 1((-0,78618 - (-0,78618)) + 1((-0,186176 - (-0,78618))) = 0,6$$

$$V_i(2) = V_i(2-1) + 1((-0,38618 - (-0,38618)) + 1((-0,186176 - (-0,38618))) = 1,2$$

$$V_i(3) = V_i(3-1) + 1((-0,38618 - (-0,38618)) + 1((-0,186176 - (-0,38618))) = 2,2$$

$$V_i(4) = V_i(4-1) + 1((-0,18618 - (-0,18618)) + 1((-0,186176 - (-0,18618))) = 3$$

$$V_i(5) = V_i(5-1) + 1((-0,18618 - (-0,18618)) + 1((-0,186176 - (-0,18618))) = 4$$

$$V_i(6) = V_i(4-1) + 1((-0,38618 - (-0,38618)) + 1((-0,186176 - (-0,38618))) = 5,2$$

2. Class buruk

$$V_i(1) = V_i(1-1) + 1((-0,39795 - (-0,39795)) + 1((-0,0397952 - (-0,39795))) = 0$$

$$V_i(2) = V_i(2-1) + 1((-0,59795 - (-0,59795)) + 1((-0,0397952 - (-0,59795))) = 1,2$$

$$V_i(3) = V_i(3-1) + 1((-0,59795 - (-0,59795)) + 1((-0,0397952 - (-0,59795))) = 2,2$$

$$V_i(4) = V_i(4-1) + 1((-0,39795 - (-0,39795)) + 1((-0,0397952 - (-0,39795))) = 3$$

$$V_i(5) = V_i(5-1) + 1((-0,39795 - (-0,39795)) + 1((-0,0397952 - (-0,39795))) = 5$$

$$V_i(6) = V_i(4-1) + 1((-0,59795 - (-0,59795)) + 1((-0,0397952 - (-0,59795))) = 5,2$$

Untuk *class* baik kita mendapatkan rata-rata bobotnya adalah = 2,7 sedangkan untuk *class* buruk rata-rata bobotnya adalah = 2,6.

E. Tahap Evaluasi

Tahapan ini dapat disebut tahapan klasifikasi karena pada tahap ini menguji akurasi dengan memasukkan data uji. Tahap ini menjelaskan secara manual menghitung *confusion matrix*, sedangkan hasil *tools rapidminer* lebih rinci dijelaskan pada bab IV.

Model Naive Bayes berbasis PSO

Pada model *Naive Bayes* berbasis PSO didapatkan *confusion matrix*, seperti pada tabel di bawah ini:

Tabel 7. *Naive Bayes* berbasis PSO ke *confusion matrix*

	Baik	Buruk
Baik	2000	227
Buruk	276	198

Kemudian masukkan nilai yang ada di dalam *confusion matrix* ke dalam persamaan di bawah ini:

$$Accuracy = \frac{198 + 2000}{2000 + 227 + 276 + 198} = \frac{2198}{2701} = 0.81$$

$$Precision = \frac{2000}{2000 + 227} = \frac{2000}{2227} = 0.89$$

$$Recall = \frac{276}{276 + 198} = \frac{276}{474} = 0.58$$

Konversi *confusion matrix* algoritme *Naive Bayes* dan PSO akan menghasilkan akurasi seperti tabel di bawah ini:

Tabel 8. *Accuracy*, *precision* dan *recall* dalam persen

	Nilai (%)
<i>Accuracy</i>	81
<i>Precision</i>	89
<i>Recall</i>	58

Tingkat akurasi yang dihasilkan dengan algoritme *Naive Bayes* berbasis PSO sebesar 81%.

F. Tahap Deployment

Pada tahap ini menerapkan model *Naive Bayes* berbasis PSO ke perusahaan *leasing* untuk penentuan kelayakan pemberian kredit mobil.

Analisa dan validasi model

Hasil dari percobaan menunjukkan bahwa algoritme *Naive Bayes* berbasis PSO akurasi meningkat dalam memprediksi penilaian kredit perusahaan *leasing* maupun *Australia credit approval* dan *Japan credit approval*, seperti yang ditunjukkan pada tabel di bawah ini:

Tabel 9. Analisa dan Validasi (*accuracy*)

Data set	Algoritme Naïve Bayes (<i>accuracy</i>)	Algoritme Naïve Bayes berbasis PSO (<i>accuracy</i>)
Agiing 2010	96,59	96,75
Agiing 2011	97,59	97,95
<i>Japan credit approval</i>	74,89	84,77
<i>Australia credit approval</i>	80,00	87,83

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Dari hasil penelitian terbukti bahwa model *Naive Bayes* berbasis PSO mempunyai akurasi yang lebih baik dibandingkan dengan *Naive Bayes* sederhana. Hal ini tampak dari hasil evaluasi penelitian bahwa model *Naive Bayes* berbasis PSO mampu menganalisa pelanggan yang baik dan pelanggan yang buruk baik menggunakan data Agiing *leasing* ACC maupun menggunakan data *credit approval* negara Australia dan Japan dari UCI data set dengan tingkat akurasi yang lebih baik. Banyaknya *record* dan atribut pada sebuah data set mempengaruhi tingkat akurasi dari model *Naive Bayes* berbasis PSO ini.

B. Saran

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

1. Hasil penelitian ini diharapkan bisa digunakan pada *leasing* ACC, khususnya ACC Bengkulu untuk lebih meningkatkan akurasi analisa kelayakan kredit bagi para calon pelanggan yang akan mengajukan aplikasi kredit.
2. Melakukan pengembangan dengan menggunakan metode optimasi yang lain atau dengan seleksi fitur untuk ketepatan penyeleksian atribut.

REFERENSI

- [1] "Peraturan Presiden No 9," 2009.
- [2] "Undang-Undang Republik Indonesia No 10," 1998.
- [3] Y. Ling, "Application of the PSO-SVM model for Credit Scoring," *Seventh International Conference on Computational Intelligence and Security*, 2011.
- [4] Y. Ma, "Research of SVM Applying in the Risk of Bank 's Loan to Enterprises," *America*, no. 3, pp. 1-5, 2010.
- [5] D. Zhang, H. Hu, and H. Zhang, "Risk Analysis of Credit Rating Business for Commercial Banks on Small and Medium-sized Enterprise," in *2011 International Conference on Information Management, Innovation Management and Industrial Engineering*, 2011, pp. 312-315.
- [6] F. C. Li, F. L. Chen, and G. E. Wang, "Comparison of Feature Selection Approaches based on the SVM Classification 12," *Industrial Engineering*, pp. 400-404, 2008.
- [7] A. Heiat, "Modeling Consumer Credit Scoring Through Bayes Network Heiat," *Social Sciences*, vol. 1, no. 3, pp. 132-141, 2011.
- [8] T. W. Liao, *Recent Advanced in Data Mining of Enterprise Data: Algorithm and Applications*, Series on . New Jersey: World Scientific Publishing Co.Pte.Ltd, 2007.
- [9] W. Shuang-cheng, "Conditional Markov Network Hybrid Classifiers Using on Client Credit Scoring," *International Symposium on Computer Science and Computational Technology*, 2008.
- [10] Y. Jiang and L. H. Wu, "Credit Scoring Model Based on Simple Naive Bayesian Classifier and a Rough Set," *IEEE*, pp. 1-4, 2009.
- [11] A. Keramati and N. Yousefi, "A Proposed Classification of Data Mining Techniques in Credit Scoring," *International Conference on Industrial Engineering and Operations Management*, pp. 416-424, 2011.
- [12] W. Feng, "Application of SVM Based on Principal Component Analysis to Credit Risk Assessment in Commercial Banks," *Global Congress on Intelligent System*, pp. 49-52, 2009.
- [13] F.-chia Li, P.-kai Wang, and G.-en Wang, "Comparison of the Primitive Classifiers with Extreme Learning Machine in Credit Scoring," *Proceedings of the IEEE IEEM*, vol. 2, no. 4, pp. 685-688, 2009.
- [14] X. Y. Yu and A. Wang, "Genetic Algorithm Based Bayesian Network for Customers ' Behavior Analysis," *Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, pp. 406-409, 2010.
- [15] M. J. Islam, Q. M. J. Wu, M. Ahmadi, and M. A. Sid-ahmed, "Investigating the Performance of Naive- Bayes Classifiers and K- Nearest Neighbor Classifiers," *International Conference on CONvergence Information Technology*, pp. 1541-1546, 2007.
- [16] F. Gorunescu, *Data Mining Concept, Model and Techniques*. Romania: Scientific Publishing Services Pvt.Ltd, 2011.
- [17] J. Lu and C. X. Ling, "Comparing Naive Bayes , Decision Trees , and SVM with AUC and Accuracy," *Proceeding of the Third IEEE International Conference on Data Mining*, pp. 3-6, 2003.
- [18] J. Lin, "Weighted Naive Bayes Classification Algorithm Based on Particle Swarm Optimization," *IEEE*, pp. 444-447, 2011.
- [19] J. Zurada, "Could Decision Trees Improve the Classification Accuracy and Interpretability of Loan Granting Decisions?," *International Conference on System Sciences*, pp. 1-9, 2010.
- [20] B. Santosa, "Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis," 2007.
- [21] P. Machine and L. Tools, *No Title*. .
- [22] D. T. Larose, *Discovering Knowledge In Data An Introduction to Data Mining*. Canada: John Wiley & Sons, Inc., Hoboken, 2005.