

IMPLEMENTASI ALGORITMA *NAIVE BAYES* DALAM PENENTUAN PEMBERIAN KREDIT

Muhammad Husni Rifqo¹, Ardi Wijaya²

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

¹kokoahzani@gmail.com

²Ardi_wijaya18@yahoo.co.id

Abstrak: Pengajuan aplikasi kredit oleh calon nasabah sekarang sangatlah mudah, hal ini dikarenakan pengajuan kredit bisa dilakukan semua orang sepanjang memenuhi syarat tertentu. Pemberian kredit kepada nasabah adalah kegiatan rutin yang mempunyai risiko tinggi, hal ini bisa menyebabkan kerugian pada perusahaan dan mengakibatkan kredit macet. Persaingan perusahaan penyedia kredit menjadi sangat pesat dan prediksi konsumen kredit adalah hal yang sangat penting. Analisis terhadap data kredit diperlukan dengan tujuan untuk meminimalisasi risiko nasabah yang terlambat membayar kredit, kegiatan ini sangatlah penting karena salah satu penyebab terjadinya kredit macet bisa disebabkan oleh kurang cermatnya perusahaan dalam pemberian kredit. Masalah ini sebenarnya dapat diatasi dengan cara mengidentifikasi dan memprediksi nasabah dengan baik sebelum memberikan pinjaman dengan cara memperahitkan data historis pinjaman. Teknik prediksi dalam pengambilan keputusan telah banyak digunakan oleh perusahaan-perusahaan besar. Penelitian ini menerapkan algoritma naive bayes untuk memprediksi dan mengklasifikasi nasabah mana saja yang bermasalah dan tidak bermasalah, dan diharapkan mampu meningkatkan akurasi dalam menganalisa kelayakan kredit.

Kata kunci : Analisa kredit, penilaian kredit, naive bayes.

Abstract: Submission of a credit application by the prospective customer now is very simple, it is because the credit application can do all of the people all meet certain requirements. Giving credit to customers is routine at high risk, it can cause harm to the company and lead to bad credit. Competition finance providers become very rapid and prediction of consumer credit is very important. An analysis of the credit data necessary for the purpose of minimizing the risk of customers who pay late on credit, this activity is very important because one of the causes of bad loans could be caused by less conscientious companies in the provision of credit. This problem can actually be solved by identifying and predicting customers with good before granting a loan by way of historical data memperahitkan loans. A prediction technique in decision making has been widely used by large companies. This research applies Naive Bayes algorithm to predict and classify the customer anywhere problematic and is not problematic, and is expected to improve the accuracy in analyzing credit worthiness.

Keywords : Credit analysis, credit ratings, naive Bayes.

I. PENDAHULUAN

Pada saat ini kegiatan kredit sangatlah ramai, hal ini tampak pada berkembang pesatnya perusahaan-perusahaan yang menawarkan jasa pembiayaan bagi calon nasabah dalam memenuhi suatu 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].

Banyak perusahaan menawarkan kredit bagi para nasabahnya, pengajuan kredit bisa diajukan oleh pemohon dengan memenuhi semua persyaratan yang sudah ditentukan oleh perusahaan. Pemberian kredit kepada nasabah

adalah kegiatan rutin yang mempunyai risiko tinggi [2], hal ini bisa menyebabkan kerugian pada perusahaan dan mengakibatkan kredit macet. Analisis terhadap data kredit diperlukan dengan tujuan untuk meminimalisasi risiko nasabah yang terlambat membayar kredit [3].

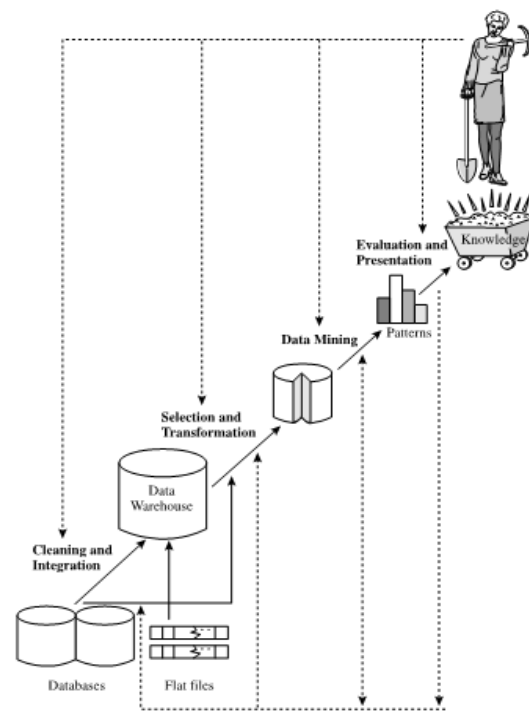
Astra Credit Companies (ACC) adalah salah satu perusahaan pembiayaan mobil terbesar di Indonesia. Perusahaan ini menyediakan pelayanan pembiayaan untuk pembelian mobil baru ataupun mobil bekas khususnya untuk merek kendaraan yang diproduksi oleh Astra. Sebagai salah satu perusahaan yang bergerak dalam bidang sewa guna usaha (*leasing*) ACC juga mengalami kendala dan hambatan dalam hal perkreditan, pemberian kredit sangatlah berisiko karena dari sekian banyak nasabah yang mengajukan aplikasi kredit ada kemungkinan beberapa pelanggan yang bermasalah dalam pembayaran.

Analisa kredit sangatlah penting, salah satu penyebab terjadinya kredit macet bisa disebabkan oleh kurang cermatnya perusahaan dalam pemberian kredit. Masalah ini sebenarnya dapat diatasi dengan cara mengidentifikasi dan memprediksi nasabah dengan baik sebelum memberikan pinjaman dengan cara memperahitkan data historis pinjaman [4]. Teknik prediksi dalam pengambilan keputusan telah banyak digunakan oleh perusahaan-perusahaan besar.

Banyak penelitian yang dilakukan untuk permasalahan risiko kredit dengan menggunakan bermacam metode data mining seperti: *Fuzzy Decision Tree*, *Logistic Regression*, *Naive Bayes*, *C4.5*, *Super Vektor Machine (SVM)* dan metode lain, metode itu dipilih sesuai dengan dataset dan tujuannya. Setiap metode ini mempunyai kelebihan masing-masing sesuai dengan fungsi dan data yang digunakan.

Data mining adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan pola dan hubungan dalam set data berukuran besar [5]. Sedangkan menurut Larose [6] data mining adalah proses menelusuri pengetahuan baru, pola dan tren yang dipilah dari jumlah data yang besar yang disimpan dalam repositori atau tempat penyimpanan dengan menggunakan teknik pengenalan pola serta statistik dan teknik matematika.

Data mining merupakan bagian dari proses *Knowledge Discovery from Data (KDD)*. Dibawah ini adalah proses dari KDD:



Gambar 1. Proses KDD [6]

Manajemen risiko adalah suatu proses untuk mengidentifikasi, mengukur, memantau dan mengendalikan risiko yang timbul serta mengambil langkah-langkah perbaikan yang dapat menyesuaikan risiko pada tingkat yang dapat diterima [7]. Dalam penerapan data mining pada data berskala besar diperlukan metodologi sistematis untuk menganalisis dan mempersiapkan data. Metodologi sistematis juga dibutuhkan untuk

melakukan interpretasi sehingga dapat menghasilkan keputusan yang bermanfaat.

Berdasarkan latar belakang diatas, maka permasalahan yang dirumuskan adalah kurang cermatnya perusahaan dalam menganalisa data calon nasabah dan tidak memaksimalkan pada manajemen risiko dalam pemberian kredit.

II. LANDASAN TEORI

A. Pengertian Implementasi

Menurut Kamus Besar Bahasa Indonesia, implementasi adalah pelaksanaan dan penerapan, dimana kedua hal ini bermaksud untuk mencari bentuk tentang hal yang disepakati terlebih dahulu. Implementasi adalah proses untuk memastikan terlaksananya suatu kebijakan dan tercapainya kebijakan tersebut. Tujuan dari implementasi sebuah sistem ialah untuk menyelesaikan desain sistem yang telah disetujui, menguji serta mendokumentasikan program-program dan prosedur sistem yang diperlukan, memastikan bahwa personil yang terlibat dapat mengoperasikan sistem yang baru dan memastikan bahwa konversi sistem lama ke sistem baru dapat berjalan dengan baik dan benar [8].

B. Data Mining

Data mining adalah proses menelusuri pengetahuan baru, pola dan tren yang dipilah dari jumlah data yang besar yang disimpan dalam repositori atau tempat penyimpanan dengan

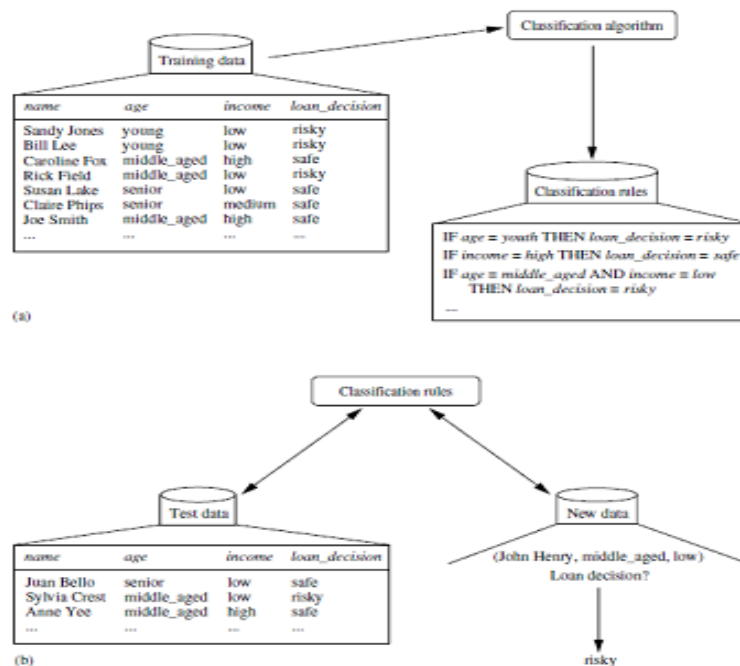
menggunakan teknik pengenalan pola serta statistik dan teknik matematika [6]. Data mining menganalisis data menggunakan *tool* untuk menemukan pola dan aturan dalam himpunan data [9]. Perangkat lunak bertugas untuk menemukan pola dengan mengidentifikasi aturan dan fitur pada data.

Budi Santosa menjelaskan bahwa data mining adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan pola dan hubungan dalam set data berukuran besar [5]. Suatu organisasi dapat dibanjiri dengan berbagai macam data, sangatlah tidak berguna data yang terkumpul dan tersimpan jika tidak dimanfaatkan [4].

Berdasarkan tugasnya [6] data mining dikelompokkan menjadi: deskripsi, estimasi, prediksi klasifikasi, clustering dan asosiasi.

1. Klasifikasi

Pada penelitian ini metode klasifikasi data mining yang digunakan adalah *naive bayes*, klasifikasi merupakan bagian algoritma dari data mining. Klasifikasi ini adalah algoritma yang menggunakan data dengan target (*class/label*) yang berupa nilai kategorikal/nominal. Contoh dari klasifikasi adalah menganalisa apakah seseorang akan membeli komputer atau tidak, dalam data mining klasifikasi dapat menghasilkan sebuah prediksi seperti gambar dibawah ini:



Gambar 2. Tahapan klasifikasi

Algoritma klasifikasi yang sering digunakan adalah: *naive bayes*, KNN, C4.5, ID3, CART dan lain-lain. Dalam penelitian ini peneliti menggunakan algoritma *naive bayes*.

2. Algoritma Naive Bayes

Pengklasifikasi *bayesian* adalah pengklasifikasi statistik dan didasarkan pada teorema *bayes*. Teori keputusan *bayes* [5] adalah pendekatan statistik yang fundamental dalam pengenalan pola (*pattern recognition*), penggunaan algoritma ini 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 diperlukan X dan h sebagai variabel acak dan menangkap hubungan peluang $P(h|X)$ ini peluang posterior untuk h dan sebaliknya prior $P(h)$.

3. Naive Bayes Classifier

Naive Bayes Classifier mengestimasi peluang kelas bersyarat dengan mengasumsikan bahwa atribut adalah independen secara bersyarat yang

diberikan dengan label kelas label kelas label kelas y . Asumsi independen bersyarat dapat dinyatakan dalam bentuk:

$$P(X|Y = y) = \prod_{i=1}^d P(X_i|Y = y)$$

dengan tiap set atribut $X = \{X_1, X_2, \dots, X_d\}$ terdiri dari d atribut.

Tahapan algoritma *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 $P(C_i/X)$

$$P(C_i/X) = \frac{P(X/C_i) \cdot P(C_i)}{P(X)}$$

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*

a. Sewa Guna Usaha (SGU)

Menurut Perpres No. 9 Tahun 2009 tentang lembaga pembiayaan [10], sewa guna usaha (*leasing*) atau sering disebut 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 *leasing* mempunyai empat tahap utama:

1. Perjanjian antara pihak *lessor* dengan pihak *lessee*.
2. Berdasarkan perjanjian sewa guna usaha, *lessor* mengalihkan hak penggunaan barang pada pihak *lessee*.
3. *Lessee* membayar kepada *lessor* uang sewa atas penggunaan barang (asset).
4. *Lessee* mengembalikan barang tersebut pada *lessor* pada akhir periode yang ditetapkan lebih dahulu dan jangka waktunya kurang dari umur ekonomi barang tersebut.

III. METODE PENELITIAN

Proses standar yang harus digunakan dalam data mining adalah CRISP-DM, yang mempunyai tahapan sebagai berikut [6]:

A. *Business Understanding*

Tahap pemahaman penelitian, untuk tujuan proyek penelitian dalam perumusan mendefinisikan masalah data mining.

B. *Data Understanding*

Tahap pengumpulan data, menganalisa data dan evaluasi kualitas data. Adapun data yang digunakan dalam penelitian ini adalah:

1. 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.

C. *Data Preparation*

- a. Siapkan data mentah (data set) yang akan digunakan untuk tahap-tahap berikutnya.
- b. Pilih variabel yang akan menjadi penentu untuk permasalahan.
- c. Lakukan transformasi pada variabel tertentu.
- d. Bersihkan data mentah sehingga data siap untuk di analisa.

D. *Tahap Modelling*

- a. Pilih model atau teknik yang akan diterapkan.
- b. Pengaturan kalibrasi model untuk optimalisasi hasil.
- c. Teknik yang berbeda dapat digunakan untuk masalah yang sama.

E. *Evaluasi*

Melakukan evaluasi terhadap model untuk mengambil keputusan mengenai penggunaan data mining.

F. *Deployment*

Memfaatkan model yang sudah dibuat serta melaksanakan model untuk proses data mining.

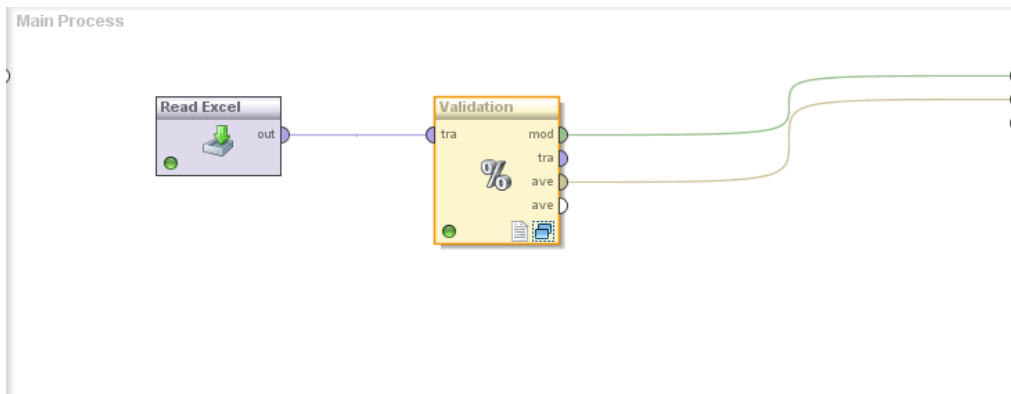
IV. HASIL DAN PEMBAHASAN

A. Pengujian Model

Penelitian ini bertujuan untuk menguji keakuratan analisa kredit mobil menggunakan algoritma *Naive Bayes*. Data yang dianalisa adalah data kredit (Agiing) perusahaan *leasing ACC* tahun 2010 sampai dengan 2011 dan data yang didapat dari UCI data set, yaitu data persetujuan kredit negara Japan dan Australia.

1. Model *Naive Bayes* untuk data set Agiing 2010 dan 2011

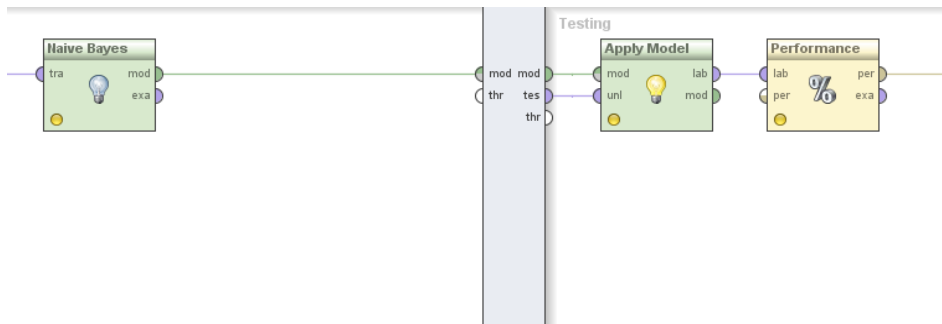
Algoritma *Naive Bayes* didasarkan pada teorema *Bayes* dimana mengukur probabilitas dengan melakukan pendekatan dari suatu ketidak-tentuan, gambar di bawah merupakan pengaplikasian model dalam *RapidMiner* untuk kasus data set Agiing 2010.



Gambar 3. Model Validation *Naive Bayes*

Dalam validation terdapat dua kolom: kolom *training* dan *testing*. Pada kolom *training* masukkan model *Naive Bayes* dan pada kolom

testing dimasukkan *apply model* dan *performance* seperti gambar berikut..



Gambar **Error! No text of specified style in document.** Model *Naive Bayes*, *Apply Model* dan *Performance* Algoritma *Naive Bayes* menghasilkan nilai *accuracy* = 96,59%, *precision* = 96,34% dan *recall* = 99,69%, seperti gambar dibawah ini:

accuracy: 96.59% +/- 0.41% (mikro: 96.59%)			
	true BURUK	true BAIK	class precision
pred. BURUK	2025	38	98.16%
pred. BAIK	462	12143	96.33%
class recall	81.42%	99.69%	

Gambar 5. Accuracy *Naive Bayes*

Gambar di bawah adalah *Distribution* model pengujian dengan *Naive Bayes*. Menghasilkan kelas dengan perincian kelas yang baik = 0.170 dan 16 distribusi, sedangkan kelas buruk mendapatkan hasil = 0.830 dan 16 distribusi.



Gambar 6. Distribution mode *Naive Bayes*

Untuk data set Agiing 2011 model *Naive Bayes* menghasilkan nilai *accuracy* = 97,59%, *precision* = 97,84% dan *recall* = 99,43%, seperti yang ditunjukkan pada gambar dibawah ini:

accuracy: 97.59% +/- 0.51% (mikro: 97.59%)			
	true BURUK	true BAIK	class precision
pred. BURUK	1752	81	95.58%
pred. BAIK	314	14236	97.84%
class recall	84.80%	99.43%	

Gambar 7. Accuracy dari *Naive Bayes*

2. Model *Naive Bayes* untuk data set Japan dan Australia *Credit Approval*

= 73,40% dan *recall* = 86,46%, seperti gambar berikut:

Pada data set ini model *Naive Bayes* menghasilkan nilai *accuracy* = 74,89%, *precision*

accuracy: 74.89% +/- 3.13% (mikro: 74.89%)			
	true baik	true buruk	class precision
pred. baik	185	52	78.06%
pred. buruk	121	331	73.23%
class recall	60.46%	86.42%	

Gambar 8. Accuracy dari *Naive Bayes*

Sedangkan untuk data set Australia model *Naive Bayes* menghasilkan nilai *accuracy* = 80,00%, *precision* = 86,47% dan *recall* = 65,45%, seperti gambar berikut:

accuracy: 80.00% +/- 4.93% (mikro: 80.00%)			
	true baik	true buruk	class precision
pred. baik	351	106	76.81%
pred. buruk	32	201	86.27%
class recall	91.64%	65.47%	

Gambar 9. Accuracy dari *Naive Bayes*

B. Evaluasi dan validasi model

Penelitian ini evaluasi dan validasi hasil menggunakan *confusion matrix* dan *ROCcurve*.

1. Evaluasi Model Confusion Matrix

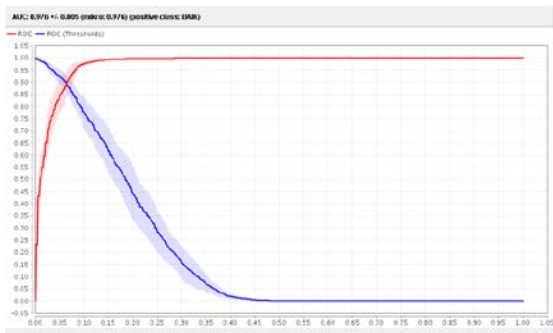
Percobaan yang telah dilakukan dengan menggunakan model *Naive Bayes* sederhana mendapatkan hasil sebagai berikut:

Tabel 1. Hasil Eksperimen Naive Bayes

Data set	Jumlah record	Jumlah atribut	Accuracy (%)	Precision (%)	Recall (%)
Agiing 2010	14668	17	96,59	96,34	96,69
Agiing 2011	16383	17	97,59	97,84	99,43
Japan credit approval	689	16	74,89	73,40	86,46
Australia credit approval	690	15	80,00	86,47	65,45

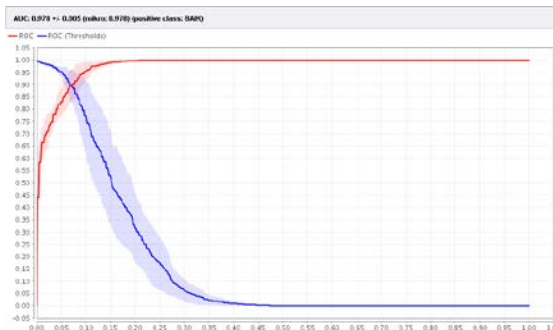
2. Kurva Receiver Operating Characteristic (ROC)

Grafik ROC dengan algoritma Naive Bayes untuk data set Agiing 2010 seperti berikut:



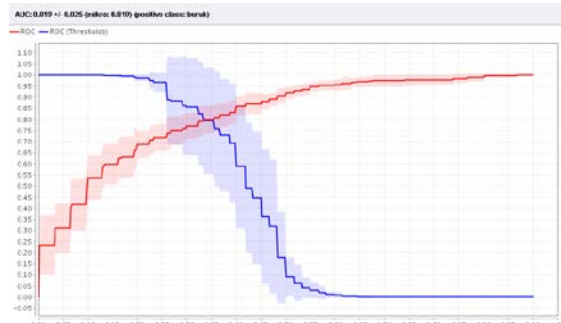
Gambar 10. Grafik ROC dengan algoritma Naive Bayes untuk data set Agiing 2010

Untuk data set Agiing 2011 model *Naive Bayes* menghasilkan 0,978 seperti yang ditunjukkan pada gambar di bawah ini:



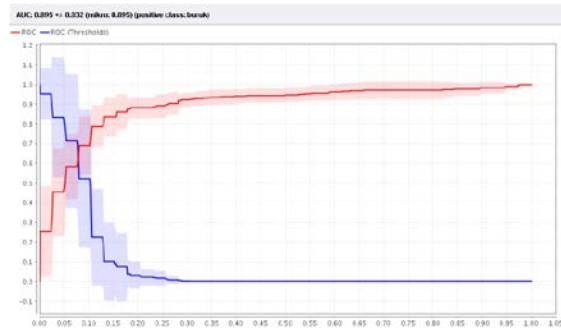
Gambar 11. Grafik ROC dengan algoritma Naive Bayes untuk data set Agiing 2011

Dan untuk data set *Japan credit approval* mendapatkan kurva ROC seperti gambar berikut:



Gambar 12. Grafik ROC dengan algoritma Naive Bayes untuk data set *Japan credit approval*

Untuk data set *Australia credit approval* mendapatkan kurva ROC seperti gambar berikut:



Gambar 13. Grafik ROC dengan algoritma Naive Bayes untuk data set *Australia credit approval*

C. Analisa dan Validasi Model

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

Tabel 2. Analisa dan Validasi (accuracy)

Data set	Algoritma Naive Bayes (accuracy)
Agiing 2010	96,59
Agiing 2011	97,59
Japan credit approval	74,89
Australia credit approval	80,00

V. PENUTUP

Dari hasil penelitian terbukti bahwa model *Naive Bayes* mempunyai akurasi yang baik, hal ini tampak dari hasil evaluasi penelitian bahwa model *Naive Bayes* 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 baik. Banyaknya record dan atribut pada sebuah data set mempengaruhi tingkat akurasi dari model *Naive Bayes* ini.

REFERENSI

- [1] *Peraturan Presiden No 9. 2009.*
- [2] *Penerapan Naive Bayes Untuk Prediksi Kelayakan Kredit.* Kurniawan, Dedy Ahmad dan Kriestanto, Danny. 2016, Jurnal Informatika dan Komputer (JIKO), Vol. 1, hal. 19-23.
- [3] *Analisis Data Pembayaran Kredit Nasabah Bank Menggunakan Metode Data Mining.* Melissa, Ira dan S. Oetama, Raymond. 2013, ULTIMA InfoSys, Vol. IV. No1, hal. 18-27.
- [4] *Aplikasi Data Mining Untuk Penilaian Kredit Menggunakan Metode Fuzzy Decision Tree.* Wahyuningtyas, Grizelda, Mukhlash, Imam dan Soetrisno. 2014, JURNAL SAINS DAN SENI POMITS, Vol. 2, No.1, hal. 1-5.
- [5] Santosa, Budi. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis.* Yogyakarta : Graha Ilmu, 2007.
- [6] Larose, Daniel T. *Data Mining Methods and Models.* Canada : A Jhon Willey & Sons, Inc., Publication, 2005.
- [7] *Analisis Terhadap Penerapan Manajemen Risiko Kredit pada PT. Bank Ekspor Indonesia.* Setiawan, Dharma. 2007, hal. 1-19.
- [8] *Desain dan Implementasi Sistem Informasi Perpustakaan Berbasis Web dengan MVC (Model View Controller).* Pertiwi, Dini Hari. 2011, Jurnal Teknologi dan Informatika (Teknomatika), Vol. 1 No.2, hal. 125-147.
- [9] *Evaluasi Algoritma Untuk Klasifikasi Nasabah Kredit Bank "X" Menggunakan Beberapa Algoritma Klasifikasi.* Marcos, Hendra dan Hidayah, Indriana. 2014, Seminar Nasional Teknologi Informasi dan Multimedia ISSN : 2302-3805, Vol. 2.02, hal. 31-36.
- [10] *Peraturan Presiden Nomor 9.* Jakarta : s.n., 2009.