

# ALGORITMA KLASIFIKASI NAÏVE BAYES UNTUK MENILAI KELAYAKAN KREDIT

Claudia Clarentia Ciptohartono

**Abstract**—Credit is a way of selling goods or loans money with no cash payment where the payment is deferred or repaid with a certain amount of the loan to the limit permitted by the bank or other entity. Along with the advancement of information technology, it is possible for companies to use statistical models to evaluate credits. Credit scoring models are built using samples of past credit in large numbers. Data mining has been proven as a tool that plays an important role for banking and retail industries, which identify useful information from a large size data. This study uses Naïve Bayesian model, this model holds the assumption that the relationship between the features or attributes are independent, which makes it simple and efficient. The result of this study proves that Naïve Bayesian algorithm can be applied to assess the credit worthiness of the BCA Finance Jakarta. And that the preprocessing is a step that greatly affects the final result to get an Excellent category for its accuracy. The credit worthiness assessment accuracy on BCA Finance Jakarta using initial data preprocessing is 85.57%, while after the initial data processing and preprocessing is 92.53%. Credit worthiness assessment using Naïve Bayesian algorithm on BCA Finance Jakarta is superior to after preprocessed initial data though Naïve Bayesian algorithm is capable of handling missing data.

**Index Terms**—Creditworthiness, Data Mining, Bayesian Classification

## I. PENDAHULUAN

Kredit adalah cara menjual barang dan atau pinjaman uang dengan pembayaran secara tidak tunai dimana pembayaran ditangguhkan atau diangsur dengan pinjaman sampai batas jumlah tertentu yang diizinkan oleh bank atau badan lain. Salah satu tugas utama dari sebuah lembaga keuangan adalah untuk mengembangkan beberapa set model dan teknik untuk memungkinkan mereka untuk memprediksi kebangkrutan dan untuk menilai kredibilitas pelanggan.

BCA Finance adalah salah satu perusahaan yang memberi jasa kegiatan kredit, khususnya pada pembiayaan mobil. Bagi perusahaan kredit, ada risiko yang harus dihadapi yaitu tidak tepat waktunya pembayaran atau bahkan kegagalan pembayaran dari kredit yang disalurkan. Masalah kredit macet ini disebabkan oleh nasabah yang berisiko. Risiko kredit adalah kemungkinan penurunan hasil kredit dari tindakan peminjam yang mempunyai reputasi yang buruk.

Seiring dengan kemajuan teknologi informasi, sangat dimungkinkan bagi perusahaan menggunakan model statistik dalam mengevaluasi kredit. Model *credit scoring* dibangun dengan menggunakan sampel kredit masa lalu dalam jumlah yang besar. Data mining telah terbukti sebagai alat yang memegang peran penting untuk industri

perbankan dan ritel, yang mengidentifikasi informasi yang berguna dari data ukuran besar. Banyak teknik yang dapat membantu pembangunan model *credit scoring* menggunakan aplikasi dari data mining.

Pada penelitian ini, penulis akan menggunakan model Naïve Bayes. Model ini memegang asumsi akan hubungan antar fitur atau atributnya yang independen sehingga menjadikannya sederhana dan efisien. Penerapan Naïve Bayes ini diharapkan dapat menilai akurasi dengan biaya terendah dalam menentukan nilai kelayakan kredit. Keuntungan lain dari model Naïve Bayes adalah mampu untuk mengoreksi diri, yang berarti bahwa ketika terjadi perubahan data, begitu juga terjadi perubahan pada hasilnya. Dalam penelitian ini, penulis menggunakan data yang baru sehingga hasilnya tentu berbeda. Dataset yang penulis gunakan adalah dataset konsumen dari perusahaan BCA Finance Jakarta 2013.

## II. ALGORITMA KLASIFIKASI NAÏVE BAYES

Pada penelitian ini, metode yang digunakan adalah Algoritma Klasifikasi Naïve Bayes untuk menilai kelayakan kredit.

### A. Algoritma Klasifikasi

Data Mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu: [1] Deskripsi, Estimasi, Prediksi, Klasifikasi, Pengklusteran, dan Asosiasi.

Klasifikasi merupakan bagian dari algoritma data

mining, klasifikasi ini adalah algoritma yang menggunakan data dengan target (*class/label*) yang berupa nilai kategorikal/nominal. Menurut Gorunescu [3] proses klasifikasi didasarkan pada empat komponen mendasar, yaitu:

### 1. Kelas (*Class*)

Variabel dependen dari model, merupakan variabel kategorikal yang merepresentasikan "label" pada objek setelah klasifikasinya. Contoh kelas semacam ini adalah: adanya kelas penyakit jantung, loyalitas pelanggan, kelas bintang (galaksi), kelas gempa bumi (badai), dll.

### 2. Prediktor (*Predictor*)

Variabel independen dari model, direpresentasikan oleh karakteristik (atribut) dari data yang akan diklasifikasikan dan berdasarkan klasifikasi yang telah dibuat. Contoh prediktor tersebut adalah: merokok, konsumsi alkohol, tekanan darah, frekuensi pembelian, status perkawinan, karakteristik (satelit) gambar, catatan geologi yang spesifik, kecepatan dan arah angin, musim, lokasi terjadinya fenomena, dll.

### 3. Pelatihan dataset (*Training dataset*)

Kumpulan data yang berisi nilai-nilai dari kedua komponen sebelumnya dan digunakan untuk melatih model dalam mengenali kelas yang cocok/sesuai, berdasarkan prediktor yang tersedia. Contoh set tersebut adalah: kelompok pasien yang diuji pada serangan jantung, kelompok pelanggan supermarket (diselidiki oleh intern dengan jajak pendapat), database yang berisi gambar untuk monitoring teleskopik dan pelacakan objek astronomi, database badai, database penelitian gempa.

### 4. Dataset Pengujian (*Testing Dataset*)

Berisi data baru yang akan diklasifikasikan oleh (*classifier*) model yang telah dibangun di atas sehingga akurasi klasifikasi (*model performance*) dapat dievaluasi.

## B. Naïve Bayes Classifier

*Naïve Bayes Classifier* (NBC) adalah salah satu algoritma dalam teknik data mining yang menerapkan teori Bayes dalam klasifikasi. [3] [5] NBC merupakan pengklasifikasian statistik [1] yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class.

Teori keputusan Bayes merupakan [3] pendekatan statistik yang fundamental dalam pengenalan pola (*pattern recognition*), teori Bayesian pada dasarnya adalah [5] kemungkinan kejadian di masa depan yang bisa dihitung dengan menentukan frekuensi pengalaman sebelumnya. Penggunaan algoritma Bayes dalam hal klasifikasi harus mempunyai masalah yang bisa dilihat statistiknya.

Setiap sampel data direpresentasi dengan set atribut  $n$ -dimensional eigenvector:  $X = \{X_1, X_2, \dots, X_n\}$ , masing-masing menggambarkan  $n$  atribut  $A_1, A_2, \dots, A_n$ . Asumsikan bahwa terdapat sejumlah kelas (*class*)  $m$   $C_1, C_2, \dots, C_m$ , dan diberi sampel data yang tidak

diketahui  $X$  (bukan *class* label).  $P(C_i|X)$  dapat ditentukan dengan Teorema Bayes berikut: [6]

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

$P(X)$  untuk semua kelas adalah konstan, hanya memerlukan  $P(X|C_i)P(C_i)$  menjadi maksimal. Jika probabilitas priori kelas  $C_i$  tidak diketahui, maka biasanya diasumsikan bahwa probabilitas kelas-kelas ini adalah sama, yakni  $P(C_1) = P(C_2) = \dots = P(C_m)$ , oleh karena pertanyaan tersebut dikonversi untuk memaksimalkan  $P(X|C_i) \cdot P(C_i)$  sering disebut sebagai kemungkinan data  $X$  ketika diberi  $C_i$ , sedangkan asumsi untuk memaksimalkan  $P(X|C_i)$  disebut kemungkinan maksimum. Jika tidak, maka perlu memaksimalkan  $P(X|C_i)P(C_i)$ . Jika diketahui asumsi probabilitas tidak sama, maka probabilitas priori dari kelas dapat dihitung dengan  $P(C_i) = \frac{S_i}{S}$ , dimana  $S_i$  adalah jumlah sampel *training* di kelas  $C_i$  sementara  $S$  adalah jumlah keseluruhan sampel *training*.

Data set dengan banyak atribut menyebabkan biaya perhitungan  $P(C_i)$  menjadi sangat besar. Untuk mereduksi perhitungan berlebihan dari ( $C_i$ ), maka *Naive bayes classifier* mengestimasi peluang kelas bersyarat dengan mengasumsikan bahwa atribut adalah independen secara bersyarat. Dengan member label *class* pada sampel, mengasumsi bahwa nilai atribut adalah independen secara bersyarat antara satu sama lain, sehingga tidak ada relasi ketergantungan antar atribut. Asumsi independen bersyarat dapat dinyatakan dalam bentuk berikut: [6]

$$P(X|C_i) = \prod_{k=1}^n P(x_k|C_i)$$

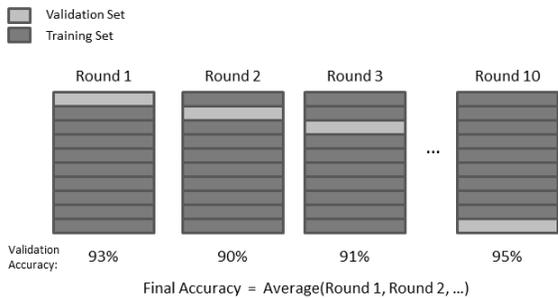
Dimana probabilitas  $P(x_1|C_i), P(x_2|C_i), \dots, P(x_n|C_i)$  dapat diestimasi dari sampel *training*.

### C. Pengujian Cross Validation

*Validation* adalah proses untuk mengevaluasi keakuratan prediksi dari model. Validasi digunakan untuk memperoleh prediksi menggunakan model yang ada dan kemudian membandingkan hasil tersebut dengan hasil yang sudah diketahui, ini mewakili langkah paling penting dalam proses membangun sebuah model. [7]

*Cross Validation* adalah teknik validasi dengan membagi data secara acak ke dalam  $k$  bagian dan masing-masing

bagian akan dilakukan proses klasifikasi. Dalam *Cross Validation*, jumlah tetap lipatan atau partisi dari data ditentukan sendiri. Cara standar untuk memprediksi *error rate* dari teknik pembelajaran dari sebuah sampel data tetap adalah dengan menggunakan *tenfold cross validation*.



II.1 Gambar tenfold cross validation

D. Alat Ukur Evaluasi dan Validasi

Untuk melakukan evaluasi pada algoritma naïve bayes dan algoritma naïve bayes dioptimasi dengan algoritma genetika dilakukan beberapa pengujian menggunakan confusion matrix dan kurva ROC (receiver operating characteristic).

1) Confusion Matrix

If you are importing your graphics into this Word template, please use the following steps:

*Confusion matrix* memberikan keputusan yang diperoleh dalam *training* dan *testing*, *confusion matrix* memberikan penilaian *performance* klasifikasi berdasarkan objek dengan benar atau salah. [3] *Confusion matrix* berisi informasi aktual (*actual*) dan prediksi (*predicted*) pada sistem klasifikasi.

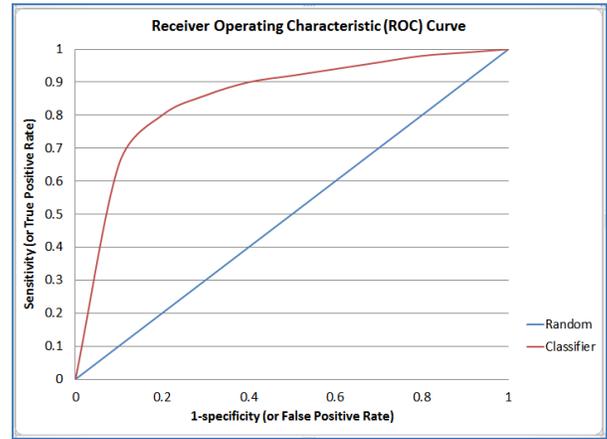
II.1 Confusion Matrix

CLASSIFICATION	PREDICTED CLASS	
	Class=Yes	Class=No
Class=Yes	A (True Positive-TP)	B (False Negative-FN)
Class=No	C (False Positive-FP)	D (True Negative-TN)

Berikut adalah persamaan model *confusion matrix* untuk menghitung akurasi. Nilai akurasi (*acc*) adalah proporsi jumlah prediksi yang benar. Dapat dihitung dengan menggunakan persamaan:

$$acc = \frac{tp + tn}{tp + tn + fp + fn}$$

2) Curve ROC



II.2 Gambar kurva ROC

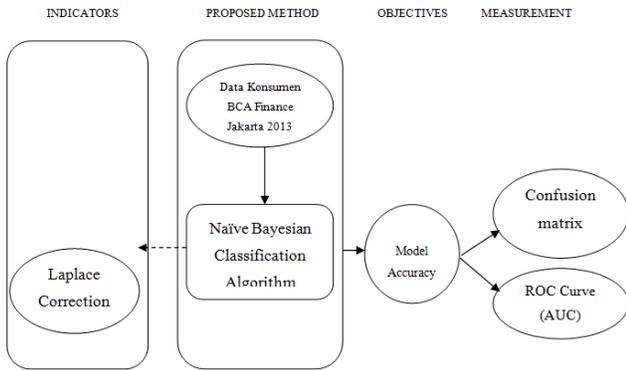
Curve ROC (*Receiver Operating Characteristic*) adalah cara lain untuk mengevaluasi akurasi dari klasifikasi secara visual [8]. Sebuah grafik ROC adalah plot dua dimensi dengan proporsi positif salah (*fp*) pada sumbu X dan proporsi positif benar (*tp*) pada sumbu Y. Titik (0,1) merupakan klasifikasi yang sempurna terhadap semua kasus positif dan kasus negatif. Nilai positif salah adalah tidak ada (*fp* = 0) dan nilai positif benar adalah tinggi (*tp* = 1). Titik (0,0) adalah klasifikasi yang memprediksi setiap kasus menjadi negatif {-1}, dan titik (1,1) adalah klasifikasi yang memprediksi setiap kasus menjadi positif {1}. Grafik ROC menggambarkan trade-off antara manfaat ('true positives') dan biaya ('false positives').

Dapat disimpulkan bahwa, satu point pada kurva ROC adalah lebih baik dari pada yang lainnya jika arah garis melintang dari kiri bawah ke kanan atas didalam grafik. Tingkat akurasi dapat di diagnosa sebagai berikut:

- Akurasi 0.90 – 1.00 = Excellent classification
- Akurasi 0.80 – 0.90 = Good classification
- Akurasi 0.70 – 0.80 = Fair classification
- Akurasi 0.60 – 0.70 = Poor classification
- Akurasi 0.50 – 0.60 = Failure

E. Kerangka Pemikiran

Pada penelitian ini, data set yang digunakan adalah *private dataset* konsumen BCA Finance cabang Jakarta yang bersifat rentet waktu (*time series*). Sedangkan metode yang diusulkan (*Proposed Method*) adalah menggunakan metode klasifikasi Naïve Bayes. Indikator (*Indicators*) yang diobservasi adalah Laplace Correction pada Naïve Bayes. Tujuan (*Objectives*) pada penelitian ini adalah adanya penilaian akurasi pada model, dimana pengukuran akurasi (*Measurements*) menggunakan Confusion matrix (Accuracy) dan ROC Curve (AUC).

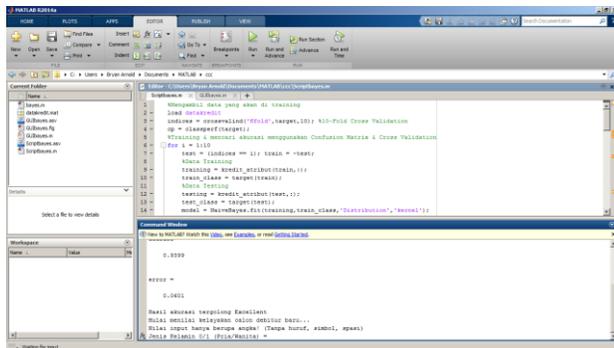


II.3 Kerangka Pemikiran

### III. IMPLEMENTASI

Model dari algoritma dalam penelitian ini juga diimplementasi menggunakan bahasa pemrograman MATLAB dengan spesifikasi sebagai berikut:

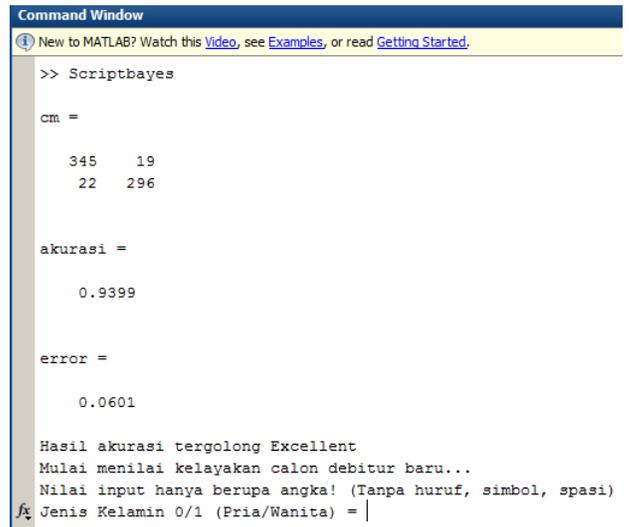
- (a) Versi : MATLAB R2014a
- (b) Pengembang : The MathWorks, Inc.
- (c) Alamat : [www.mathworks.com](http://www.mathworks.com)



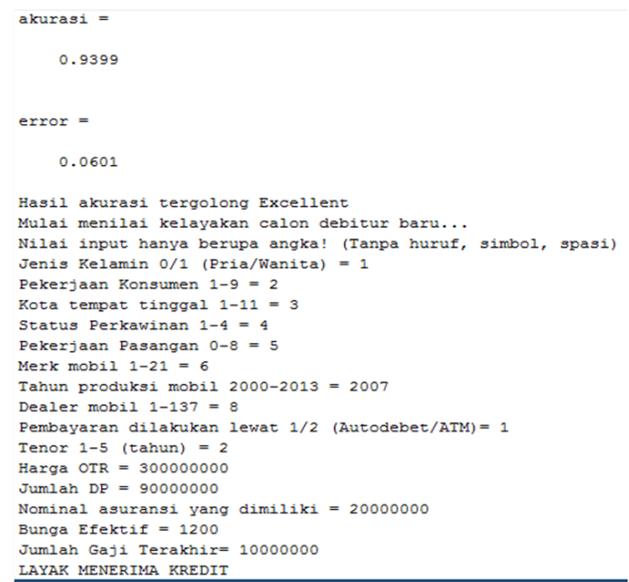
III.1 Tampilan laman kerja pada MATLAB

Program ini dapat menghitung Akurasi yang didapat dari perhitungan *Confusion Matrix* dan juga dapat menghitung laju eror. Selain mengukur performa algoritma, program ini juga dapat menilai kelayakan kredit yang sesuai dengan objek pada penelitian ini.

Berikut contoh tampilan program sederhana yang dibuat untuk implementasi dalam menilai kelayakan kredit sesuai data konsumen BCA Jakarta:



III.2 Tampilan program pada Command Window



III.3 Tampilan hasil implementasi

## IV. HASIL PENELITIAN DAN PEMBAHASAN

### A. Data Penelitian

Data penelitian merupakan data konsumen dari BCA Finance Jakarta selama periode Mei 2013 sampai dengan Februari 2014. Data awal penelitian berjumlah 682 data dengan 20 atribut. Data masih mengandung banyak data tidak penting, atribut yang tidak mendukung, dan *error*. Untuk mendapatkan data berkualitas dalam *Data Mining* maka diperlukan proses pengolahan data awal.

ID	Kategori	Nama Konsumen	Jenis Kelamin	Kota	Jenis Pekerjaan	Keperluan Pemasangan	Waktu	Salah	Status	Pembayaran	Tenor	DP	ASURANSI	Luas Lahan	Spesies
1	Pria	WALDI SURBANDI	Karyawan	Cirebon	REGULER	ibu Ruman Tagga MT/TK/TKU	2013	Selesai	USA MEDIUM	Autodebet	1	8000000	8770000	288000	0
2	Pria	WENDEL GUNO	Karyawan	Jakarta	REGULER	ibu Ruman Tagga MT/TK/TKU	2013	Selesai	USA MEDIUM	Autodebet	1	8000000	8770000	288000	0
3	Pria	WENDEL GUNO	Karyawan	Jakarta	REGULER	ibu Ruman Tagga MT/TK/TKU	2013	Selesai	USA MEDIUM	Autodebet	1	8000000	8770000	288000	0
4	Pria	WENDEL GUNO	Karyawan	Jakarta	REGULER	ibu Ruman Tagga MT/TK/TKU	2013	Selesai	USA MEDIUM	Autodebet	1	8000000	8770000	288000	0
5	Pria	WENDEL GUNO	Karyawan	Jakarta	REGULER	ibu Ruman Tagga MT/TK/TKU	2013	Selesai	USA MEDIUM	Autodebet	1	8000000	8770000	288000	0
6	Pria	WENDEL GUNO	Karyawan	Jakarta	REGULER	ibu Ruman Tagga MT/TK/TKU	2013	Selesai	USA MEDIUM	Autodebet	1	8000000	8770000	288000	0
7	Pria	WENDEL GUNO	Karyawan	Jakarta	REGULER	ibu Ruman Tagga MT/TK/TKU	2013	Selesai	USA MEDIUM	Autodebet	1	8000000	8770000	288000	0
8	Pria	WENDEL GUNO	Karyawan	Jakarta	REGULER	ibu Ruman Tagga MT/TK/TKU	2013	Selesai	USA MEDIUM	Autodebet	1	8000000	8770000	288000	0
9	Pria	WENDEL GUNO	Karyawan	Jakarta	REGULER	ibu Ruman Tagga MT/TK/TKU	2013	Selesai	USA MEDIUM	Autodebet	1	8000000	8770000	288000	0
10	Pria	WENDEL GUNO	Karyawan	Jakarta	REGULER	ibu Ruman Tagga MT/TK/TKU	2013	Selesai	USA MEDIUM	Autodebet	1	8000000	8770000	288000	0
11	Pria	WENDEL GUNO	Karyawan	Jakarta	REGULER	ibu Ruman Tagga MT/TK/TKU	2013	Selesai	USA MEDIUM	Autodebet	1	8000000	8770000	288000	0
12	Pria	WENDEL GUNO	Karyawan	Jakarta	REGULER	ibu Ruman Tagga MT/TK/TKU	2013	Selesai	USA MEDIUM	Autodebet	1	8000000	8770000	288000	0
13	Pria	WENDEL GUNO	Karyawan	Jakarta	REGULER	ibu Ruman Tagga MT/TK/TKU	2013	Selesai	USA MEDIUM	Autodebet	1	8000000	8770000	288000	0
14	Pria	WENDEL GUNO	Karyawan	Jakarta	REGULER	ibu Ruman Tagga MT/TK/TKU	2013	Selesai	USA MEDIUM	Autodebet	1	8000000	8770000	288000	0
15	Pria	WENDEL GUNO	Karyawan	Jakarta	REGULER	ibu Ruman Tagga MT/TK/TKU	2013	Selesai	USA MEDIUM	Autodebet	1	8000000	8770000	288000	0
16	Pria	WENDEL GUNO	Karyawan	Jakarta	REGULER	ibu Ruman Tagga MT/TK/TKU	2013	Selesai	USA MEDIUM	Autodebet	1	8000000	8770000	288000	0
17	Pria	WENDEL GUNO	Karyawan	Jakarta	REGULER	ibu Ruman Tagga MT/TK/TKU	2013	Selesai	USA MEDIUM	Autodebet	1	8000000	8770000	288000	0
18	Pria	WENDEL GUNO	Karyawan	Jakarta	REGULER	ibu Ruman Tagga MT/TK/TKU	2013	Selesai	USA MEDIUM	Autodebet	1	8000000	8770000	288000	0
19	Pria	WENDEL GUNO	Karyawan	Jakarta	REGULER	ibu Ruman Tagga MT/TK/TKU	2013	Selesai	USA MEDIUM	Autodebet	1	8000000	8770000	288000	0
20	Pria	WENDEL GUNO	Karyawan	Jakarta	REGULER	ibu Ruman Tagga MT/TK/TKU	2013	Selesai	USA MEDIUM	Autodebet	1	8000000	8770000	288000	0

IV.1 Potongan tampilan data awal pada Excel Sheet

B. Pengolahan Data Awal

Data ganda atau data yang hilang mungkin menyebabkan ketidakbenaran atau bahkan menyesatkan, maka untuk mendapatkan data yang berkualitas, beberapa teknik yang harus dilakukan sebagai berikut:

1) Pembersihan Data

- a) Proses menghapus data yang tidak lengkap. Terdapat 12 data dari 682 yang tergolong sangat tidak lengkap. 16 dari 702 tidak diberi nilai Bunga Efektif dan 4 dari 686 sebagai kurang lengkap. Sehingga data yang tersisa ada 682 record.
- b) Mengisi nilai-nilai yang hilang, data yang tidak lengkap (*missing value*).

2) Integrasi dan Transformasi Data

Meningkatkan akurasi dan efisiensi algoritma. Algoritma naïve bayes dapat memproses data yang bernilai nominal, ordinal, maupun kontinyu. Sehingga nilai-nilai dari setiap atribut yang terdapat pada data set tidak perlu ditransformasikan.

Terdapat 1 atribut data yang bernilai numerik dan 5 atribut data yang bernilai integer. Naïve Bayes langsung dapat menangani semua tipe data dari data penelitian.

3) Reduksi Data

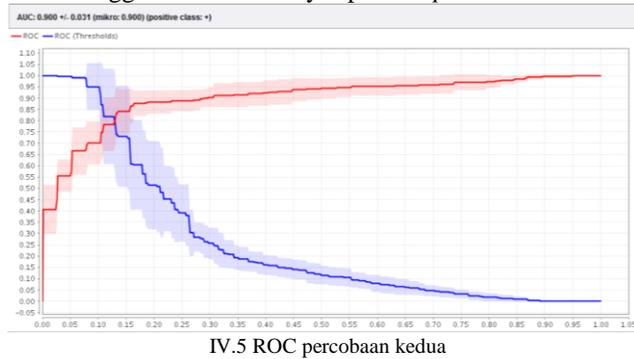
Mereduksi data set dengan mengurangi jumlah atribut dan record agar lebih sedikit namun bersifat informatif.

- a) Memecahkan redundansi dan memperhalus *data noise* dengan cara menghapus atribut Kondisi. Dikarenakan nilai dari keseluruhan atribut Kondisi selama satu periode adalah sama, berisi nilai "Bekas".
- b) Terdapat atribut *Jangka Waktu* dan *Tenor* yang nilainya redundan, disebabkan karena informasi yang disampaikan sama, yakni jumlah lama waktu pemberian kredit.
- c) Menghapus atribut dengan nilai yang sama atau yang kurang penting pada Excel Sheet:
  - Menghapus atribut NO (Nomer).
  - Menghapus atribut Nama Konsumen.

Hasil dari pembersihan data awal setelah melalui pembersihan, transformasi dan reduksi data menyisakan 16 atribut dan 682 record.

Kategori	Pelaksanaan konsumen	Data	Status Perencanaan	Pelaksanaan Pasaran	Ident	Tahun	Pencaja	Perencanaan	Tenor	OTR	DP	BB/BBK	Bunga Effektif	Tenor	Tersedia
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA AUTO	AutoBekas	4	88000000	45000000	11420000	11,00	3612000	+
Pria	Berkonsumen	Jakarta Utara	Membayar	Belum Membayar	SAUDARI	2016	2016 SAU CEPER AUTO	AutoBekas	4	88000000	28000000	5700000	14,34	2855000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA CEPER AUTO	AutoBekas	4	100000000	39000000	10807000	14,34	2488000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	13,53	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar	Belum Membayar	HONGTA	2016	2016 HONGTA HONGTA MOTOR	AutoBekas	4	110000000	44000000	11900000	14,34	2190000	+
Pria	Berkonsumen	Jakarta Barat	Membayar												

Berikut adalah gambar ROC hasil pengujian akurasi data awal menggunakan naïve bayes pada *Rapidminer*:



Akurasi dari percobaan pertama adalah 90,75% yang sudah termasuk dalam rentang 0.90 – 1.00 atau *Excellent classification*. Dan akurasi dari percobaan kedua adalah 85,57% yang termasuk dalam rentang 0.80 – 0.90 atau *Good classification*.

### 2) Algoritma Naïve Bayes pada Data Bersih

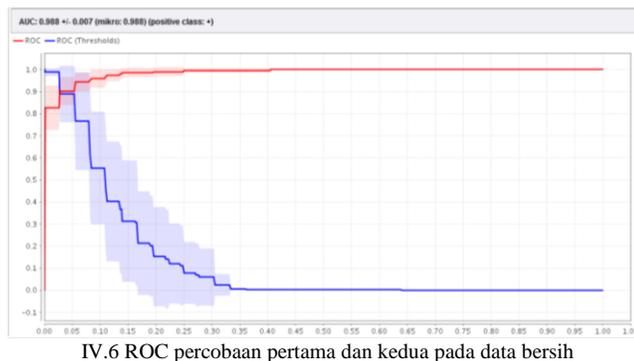
Percobaan kedua adalah menerapkan algoritma naïve bayes untuk menilai kelayakan kredit BCA Finance Jakarta dengan adanya pengolahan data awal sehingga menjadi data bersih.

Terdapat 682 data dan 16 atribut diantaranya yaitu *layak, kelamin, pekerjaan konsumen, kota, status perkawinan, pekerjaan pasangan, merk, tahun, penjual, pembayaran, gaji terakhir, asuransi, DP, bunga eff, tenor, dan OTR*.

Kemudian jika pada data bersih dilakukan pengisian nilai-nilai yang hilang dan proses diskritasi. Data yang tidak lengkap (*missing value*) pada *Rapidminer* menggunakan operator *replace missing values*.

Akurasi dari percobaan pertama dan kedua adalah sama yaitu 92,53% yang sudah termasuk dalam rentang 0.90 – 1.00 atau *Excellent classification*.

Berikut adalah gambar ROC hasil pengujian naïve bayes pada *Rapidminer*:



## V. PENUTUP

Dari hasil penelitian membuktikan bahwa algoritma Naïve Bayes dapat diterapkan untuk menilai kelayakan kredit pada BCA Finance Jakarta. Dan pengolahan data awal merupakan tahapan yang sangat mempengaruhi hasil

akurasi yang baik sehingga akurasi akhir yang dihasilkan termasuk kategori *Excellent*. Penilaian kelayakan kredit pada BCA Finance Jakarta pada data awal dengan *pre-processing* menghasilkan akurasi sebesar 85,57% sedangkan setelah dilakukan pengolahan data awal dan dengan *pre-processing* menghasilkan akurasi sebesar 92,53%.

Jadi penilaian kelayakan kredit menggunakan Algoritma Naïve Bayes pada BCA Finance Jakarta lebih unggul jika dilakukan pengolahan data awal sekalipun Naïve Bayes merupakan algoritma yang mampu menangani data yang hilang.

## REFERENCES

### VI. WORKS CITED

- [1] M. Berry, *Data Mining Techniques*, John Wiley & Sons, 2004.
- [2] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*, Hoboken: John Wiley & Sons Inc., 2005.
- [3] F. Gorunescu, *Data Mining Concept Model Technique*, Craiova, Romania: Springer, 2011.
- [4] B. Santosa, *Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Yogyakarta: Graha Ilmu, 2007.
- [5] 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.
- [6] J. Lin and J. Yu, "Weighted Naive Bayes Classification Algorithm Based on Particle Swarm Optimization," *IEEE*, pp. 444-447, 2011.
- [7] A. G. Mabrur and R. Lubis, "Penerapan Data Mining untuk Memprediksi Kriteria Nasabah Kredit," *Jurnal Komputer dan Informatika (KOMPUTA)*, vol. 1, pp. 53-57, 2012.
- [8] S. Budi, "Sunar Budi: EVERYONE'S INNER-SELF LIGHTS," [Online]. Available: <http://subud.wordpress.com>. [Accessed 15 7 2014].