Penentuan Kelayakan Kredit Dengan Algoritma Naïve Bayes *Classifier*: Studi Kasus Bank Mayapada Mitra Usaha Cabang PGC

Nia Nuraeni[1]

Abstract* — *In analyzing a credit sometimes a less accurate credit officer in credit analysis, so that it can lead to increased bad debts. Classification data mining algorithms are widely used to determine the credit worthiness of one Naive Bayes classifier, NBC superior in increasing the value of high accuracy but weak in the selection of attributes. After testing Naive Bayes classifier algorithm the results obtained is Naive Bayes classifier algorithm produces an accuracy of 89.33% and AUC values for 0.955
Keyword: Credit Analysis, Naive Bayes ClassifierAlgorithm.

Intisari — Dalam menganalisis kredit kadang-kadang petugas kredit kurang akurat dalam analisis kredit, sehingga dapat menyebabkan peningkatan kredit macet . Klasifikasi algoritma data mining secara luas digunakan untuk menentukan kelayakan kredit dari salah satu Naif Bayes classifier, BC unggul dalam meningkatkan nilai akurasi yang tinggi tetapi lemah dalam pemilihan atribut. Setelah menguji algoritma Naive Bayes menghasilkan akurasi 89,33 % dan nilai AUC 0.955.
Kata Kunci: ***Analisa Kredit, Algoritma Naive Bayes Classifier.***

I. PENDAHULUAN

Dalam proses pemberian kredit selama ini, khususnya pemberian kredit usaha mikro yang dilakukan oleh Bank Mayapada meskipun melalui analisa kredit masih saja ada permasalahan yang timbul diantaranya, para calon debitur melakukan segala macam cara agar kreditnya disetujui oleh pihak bank. Hal ini yang menyebabkan tingkat kredit macet juga meningkat. Penyebabnya antara lain kurang akuratnya *credit officer* dalam memberikan analisanya, pihak *account officer* yang kurang tepat dalam mencari calon debitur karna dikejar oleh target perusahaan.

Bank Mayapada menetapkan kebijakan dalam pemberian kredit, antara lain menetapkan standar untuk menerima atau menolaknya. Analisa kredit yaitu untuk menentukan siapa yang berhak menerima kredit yang telah memenuhi prinsip bagaimana karakter nasabah (*Character*/Data pribadi nasabah), kapasitas nasabah untuk melunasi kreditnya (*Capacity*), kemampuan modal yang dimiliki nasabah dan aktivitas usahanya (*Capital*), jaminan nasabah (*Collateral*/ Jaminan) dan kondisi usaha nasabah (*Condition*) atau biasa disebut sebagai 5C. Sehingga berdasarkan analisa kredit tersebut bisa didapatkan beberapa variabel antara lain: nama debitur, alamat, jenis usaha, status tempat tinggal, status tempat usaha, lama usaha, sistem penjualan, sistem pembelian, pemasok/supplier yang dimiliki debitur, pembelian dari supplier, *repayment capacity*, omzet perbulan, *gross profit margin*, *security coverage ratio*, jenis jaminan yang diberikan, status kepemilikan jaminan dan BI Checking. Variabel-variabel tersebut memiliki keterhubungan satu sama lain dalam penentuan kelayakan pemberian kredit

Naive Bayes, Neural Network dan C45 merupakan algoritma klasifikasi data mining yang banyak digunakan dalam setiap penelitian untuk prediksi kelayakan pemberian kredit. Naive Bayes Classifier merupakan algoritma yang memiliki tingkat akurasi lebih tinggi dibandingkan dengan algoritma klasifikasi yang lain.[1]

Persetujuan kelayakan pemberian kredit dalam *credit scoring* menggunakan *data mining* dan dihasilkan pengklasifikasikan tentang "*good credit*" dan "*bad credit*" sehingga dapat dijadikan acuan dalam pemberian kredit kepada calon debitur[2]. Meskipun Naive Bayes Classifier diunggulkan karna memiliki tingkat akurasi yang cukup tinggi, dan dapat di terapkan dalam jumlah data yang besar Naive Bayes juga memiliki kekurangan yakni lemah dalam proses penyeleksian atribut atau variabel.

Dilingkungan PGC dan sekitarnya (Plasma) Bank Mayapada memberikan kredit terhadap calon debitur yang memiliki berbagai faktor usaha dimana faktor usaha ini merupakan salah satu aspek penilaian kredit, sehingga dalam penelitian ini akan dilakukan penelitian tentang keakurasian salah satu penilaian analisa kredit faktor usaha sehingga dapat dijadikan acuan terutama untuk *credit officer* Bank Mayapada dalam memberikan kredit dimasa datang.

Berdasarkan penjelasan diatas penelitian menggunakan algoritma *data mining* dapat digunakan dalam mengatasi masalah untuk menganalisa kasus kredit, sehingga menghasilkan nilai akurasi yang cukup tinggi dibandingkan dengan algoritma yang lainnya. Penelitian ini akan menggunakan algoritma data mining Naive Bayes *Classifier* sehingga diharapkan akan menghasilkan akurasi yang lebih tinggi.

II. KAJIAN LITERATUR

a. Kredit

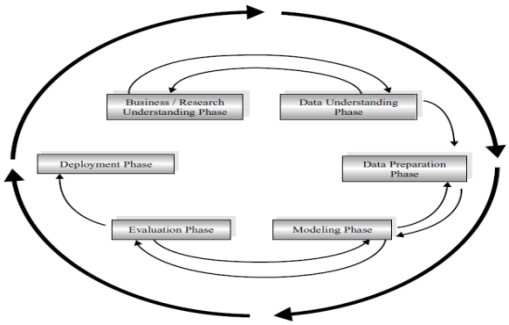
Undang-Undang Republik Indonesia Nomor 10 Tahun 1998 tentang perubahan atas Undang-Undang Nomor 7 Tahun 1992 tentang perbankan yang dimaksud dengan kredit adalah penyedian uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan pinjam meminjam antar Bank dengan pihak lain yang mewajibkan pihak peminjam melunasi hutangnya setelah jangka waktu tertentu dengan pemberian bunga[3].

[1] *Program Studi Teknik Informatika STMIK Nusa Mandiri Jakarta,Jl. Damai No 8 Warung Jati Barat (Margasatwa) Jakarta Selatan (Telp: 021-78839513; fax: (021) 78839421; e-mail: nia.nne@bsi.ac.id)*

b. Data Mining

Data Mining adalah proses menemukan korelasi baru yang bermakn, pola dan tren dengan memilah-milah sejumlah besar data yang tersimpan dalam repositori, menggunakan teknologi penalaran pola serta teknik-teknik statistik dan matematika[4].

Pada prosesnya data mining akan mengekstrak informasi yang berharga dengan cara menganalisis adanya pola-pola ataupun hubungan keterkaitan tertentu dari data-data yang berukuran besar. Data mining berkaitan dengan ilmu-ilmu lain seperti, *Database System*, *Data Warehousing*, *Statistic*, *Machine Learning*, *Information Retrievel*, dan Komputasi tingkat tinggi. Data mining adalah sebuah proses, sehingga dalam melakukan prosesnya harus sesuai dengan prosedur CRISP-DM (*Cross- Industry Standard Process for Data Mining*) [5].



Gambar 1
CRISP-DM process

c. Algoritma Klasifikasi Naive Bayes

Algoritma Naïve Bayes merupakan suatu bentuk klasifikasi data dengan menggunakan metode *probabilitas* dan *statistik.* Metode ini pertama kali dikenalkan oleh ilmuwan Inggris Thomas Bayes, yaitu digunakan untuk memprediksi peluang yang terjadi di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai *teorema Bayes*.

Metode *Teorema bayes* kemudian dikombinasikan dengan *naive* yang diasumsikan dengan kondisi antar atribut yang saling bebas. *Algoritma Naive Bayes* dapat diartikan sebagai sebuah metode yang tidak memiliki aturan, *Naive Bayes* menggunakan cabang matematika yang dikenal dengan teori probabilitas untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan cara melihat frekuensi tiap klasifikasi pada data training. *The Naive Bayes Classifier* bekerja sangat baik dibanding dengan model classifier lainnya bahwa "Naïve Bayes *Classifier* memiliki tingkat akurasi yang lebih baik dibanding model classifier lainnya" [6]. *Bayes rule* digunakan untuk menghitung probabilitas suatu class. *Algoritma Naive Bayes* memberikan suatu cara mengkombinasikan peluang terdahulu dengan syarat kemungkinan menjadi sebuah formula yang dapat digunakan untuk menghitung peluang dari tiap kemungkinan yang terjadi.

Berikut adalah bentuk umum dari *teorema bayes*:

$$P(H|X) = \frac{p(X|H|P(H)}{P(X)} \quad \text{.....(1)}$$

Keterangan:
X = Data dengan *class* yang belum diketahui.
H = *Hipotesis* data X merupakan suatu *class spesifik*.
P(H|X) = *Probabilitas hipotesis* H berdasarkan kondisi X (*posteriori probability*).
P(H) = *Probabilitas Hipotesis* H (*prior probability*).
P(X|H) = *Probabilitas* X berdasar kondisi pada *Hipotesis* H
P(X) = Probabilitas dari X.

Metode algoritna Naïve bayes merupakan penyederhanaan *metode bayes*. Untuk mempermudah pemahaman, maka *Teorema Bayes* disederhanakan menjadi:

$$P(H|X) = P(X|H)\ P(X) \quad \text{.....(2)}$$

Metode Bayes rule digunakan dan diterapkan untuk melakukan penghitungan terhadap *posterior* dan *probabilitas* dari data sebelumnya. Dalam analisis *bayesian*, fungsi klasifikasi akhir dihasilkan dengan menggabungkan kedua sumber informasi (*prior dan posterior*) untuk menghasilkan probabilitas menggunakan aturan *bayes*

Classification		Predicated Class	
		Class=YES	**Class=NO**
Observed Class	**Class = Yes**	*A (True Positive-Tp)*	*B (False Negative-Fn)*
	Class = NO	*C (False Positive-FP)*	*D (True Negative-TN)*

d. Cross Validation

Dalam pendekatan *cross validation***,** setiap *record* digunakan beberapa kali dalam jumlah yang sama untuk *training* dan tepat sekali untuk untuk *testing*. *Cross validation* adalah teknik validasi dengan membagi data secara acak kedalam k bagian dan masing-masing bagian akan dilakukan proses klasifikasi[7].

e. Confusion Matrix

Confussion matrix adalah sebuah metode untuk melakukan evaluasi dengan menggunakan tabel matrix [8].

Tabel 1
Model *Confusion Matrix* [9]

True Positive (TP) = proporsi positif dalam data set yang diklasifikasikan positif
True Negative (TN) = proporsi negative dalam data set yang diklasifikasikan negative
False Positive (FP) = proporsi negatif dalam data set yang diklasifikasikan potitif
False Negative (FN) = proporsi negative dalam data set yang diklasifikasikan *negative*

1. $Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$

2. $Sensitivity = \frac{Number\ of\ 'True\ Positives'}{Number\ of\ 'True\ Positives' + Number\ of\ 'False\ Negatives'}$

3. $Specificity = \frac{Number\ of\ 'True\ Negatives'}{Number\ of\ 'True\ Negatives' + Number\ of\ 'False\ Positives'}$

4. $PPV = \frac{Number\ of\ 'True\ Positives'}{Number\ of\ 'True\ Positives' + Number\ of\ 'False\ Positives'}$

5. $NPV = \frac{Number\ of\ 'True\ Negatives'}{Number\ of\ 'True\ Negatives' + Number\ of\ 'False\ Negatives'}$

Sensitivity juga dapat dikatakan *true positive rate* (TP *rate*) atau *recall*. Sebuah *sensitivity* 100% berarti bahwa pengklasifikasian mengakui sebuah kasus yang diamati positif

f. Nilai *Accuracy*

Nilai *accuracy* adalah *presentase* dari jumlah *record data* yang diklasifikasikan secara baik dan benar dengan menggunakan sebuah algoritma dan dapat membuat klasifikasi setelah dilakukan pengujian pada hasil klasifikasi tersebut.[10]

g. Kurva ROC

Fungsi Kurva *ROC* adalah untuk menunjukkan akurasi dan membandinngkan klasifikasi secara visual. *ROC* mengekspresikan *Confusion Matrix*, ROC adalah grafik dua dimensi dengan *false positives* sebagai garis *horizontal dan true positive* sebagai garis *vertical*.[11].

III. METODOLOGI PENELITIAN

Penelitian yang digunakan adalah penelitian *Experiment*. Penelitian eksperimen melibatkan penyelidikan perlakuan parameter/variabel tergantung pada penelitinya dan menggunakan tes yang dikendalikan oleh si peneliti itu sendiri. Pada metode penelitian eksperimen, digunakan model proses CRISP-DM (*Cross- Standard Industry Process for Data Mining*) yang terdiri dari 6 tahapan[9]:

1. *Bussiness Understanding*
2. *Data Understanding*
3. *Data Preparation*
4. *Modelling*
5. *Evaluation*
6. *Deployment*

1. Bussiness Understanding

Berdasarkan data nasabah kredit tahun 2014, terlihat bahwa nasabah dengan status kreditnya macet lebih banyak dari nasabah yang status kreditnya lancar, hal ini menjadi sebuah masalah dan kendala bagi Bank Mayapada Mitra Usaha cab PGC khususnya karna akan berakibat pada kurang akuratnya analisa kredit. Dalam beberapa penelitian telah dilakukan proses analisa kredit dengan menggunakan Naive Bayes *Classifier*.

2. Data Understanding

Data yang didapatkan dari Bank Mayapada Mitra Usaha Cab PGC dan sekitarnya (plasma) adalah data kredit nasabah tahun 2014. Atribut atau variabel yang ada sebanyak 22 atribut (data lengkapnya bisa dilihat dilampiran). Setelah dilakukan proses data *preparation*, atribut atau variabel yang digunakan terdiri dari 10 atribut yang terdapat dalam data status kredit nasabah. Variabel-variabel tersebut ada yang tergolong variabel prediktor atau pemrediksi (*Predictor Variable*) yaitu variabel yang dijadikan dasar sebagai penentu resiko kredit. Variabel prediktor yaitu jenis usaha, status tempat usaha, lama usaha, sistem penjualan, sistem pembelian, omset per bulan, *gross profit margin*, *repayment capacity*, dan fasilitas. Sedangkan variabel tujuan adalah kolek nasabah (yang menunjukan status kredit lancar atau macet)

3. Data Preparation

Pada tahapan ini data sebanyak 300 dan atribut yang terdiri dari 22 atribut, akan dilakukan beberapa penyeleksian untuk menghasilkan data yang dibutuhkan, tahapannya yaitu:

a. *Data Cleaning* untuk membersihkan nilai yang kosong atau tuple yang kosong.
b. *Data Integration* yang berfungsi menyatukan tempat penyimpanan yang berbeda kedalam satu data.

Data Reduction, jumlah atribut yang digunakan terlalu banyak dan tidak semua atribut menjadi syarat atas atribut penentu sehingga yang digunakan hanya 10 Atribut

4. Modelling

Pada tahap ini, data *preparation* yang telag didapatkan akan dilakukan pemodelan Algoritma Naive Bayes *Classifier* yang disebut dengan model probabilitas Naive Bayes dimana dalam probabilitas itu dilakukan perhitungan manual dengan cara menghitung probabilitas prior dan probabilitas posterior, kemudian dimasukan kedalam *tools* rapidminer dengan menggunakan algoritma Naive Bayes .

5. Evaluation

Pada tahap evaluasi, disebut tahap klasifikasi karena pada tahap ini akan ditentukan pengujian untuk akurasi. Tahap pengujiannya adalah melihat hasil akurasi pada proses klasifikasi Algoritma Naive Bayes dan klasifikasi Naive Bayes berbasis *Particle Swarm Optimization* serta evaluasi dengan ROC *Curve*. Penjelasan secara lengkap tentang membandingkan kedua model tersebut terdapat pada bab IV.

6. Deployment

Pada tahapan *deployment*, dilakukan penerapan model algoritma klasifikasi Naive Bayes berbasis *Particle Swarm Optimization* untuk menentukan kelayakan pemberian kredit berdasarkan faktor usaha nasabah/debitur pada Bank Mayapada atau dalam obyek penelitian ini adalah Bank Mayapada Mitra Usaha Cabang PGC.

IV. HASIL DAN PEMBAHASAN

1. Eksperimen dan Pengujian Metode.

Pada tahap ini dilakukan eksperimen dan pengujian model yaitu menghitung dan mendapatkan rule-rule yang ada pada model algoritma yang diusulkan. Dimana dalam penelitian ini bertujuan untuk mengembangkan model yang sudah terbentuk dengan algoritma Naïve Bayes *Classifier*. Data dianalisa dengan melakukan dua perbandingan yaitu menggunakan algoritma Naïve Bayes *Classifier*

1.1 Model Algoritma Naïve Bayes *Classifier*

A. Menghitung Probabilitas Prior

Eksperimen yang penulis lakukan dalam penelitian ini adalah dengan menghitung Probabilitas Prior dan Probabilitas Posterior dengan menggunakan data sebanyak 300 *record* data.

Menghitung *Probabilitas Prior* dan *Probabilitas posterior*, dalam bentuk persamaan dibawah ini:

Total data = 300
Data kredit lancar = 112
Data kredit macet = 188
P(Lancar) = 112 : 300 = **0.373**
P(Macet) = 188 : 300 = **0.627**

Setelah didapatkan nilai *probabilitas* untuk tiap hipotesis dari *class*, maka langkah selanjutnya adalah melakukan penghitungan terhadap kondisi *probabilitas* tertentu (*Probabilitas X*) dengan menggunakan data berdasarkan *probabilitas* tiap hipotesis (*Probabilitas H*) atau yang dinamakan dengan *probabilitas Prior*. Selanjutnya untuk mengetahui hasil perhitungan dari *Probabilitas Prior*, maka dilakukan penghitungan dengan cara merinci jumlah kasus dari tiap-tiap atribut variabel data, adapun hasil perhitungan *probabilitas prior* dengan menggunakan *Algoritma Naïve Bayes* dapat dilihat pada tabel 2 berikut:

Tabel 2
Perhitungan Probabilitas *Prior*

Atribut/ Variabel		Jumlah Data (S)	Kredit Lancar (Si)	Kredit Macet (Si)	P(X|Ci) Lancar	P(X|Ci) Macet
Total		300	112	188	0.373	0.627
Jenis Usaha	Usaha Mikro 1 (Primer)	164	37	127	0.33	0.675
	Usaha Mikro 0 (Sekunder)	136	75	61	0.669	0.324
Status Tempat Usaha	Milik Sendiri	156	77	79	0.687	0.42
	Kontrak	144	35	109	0.312	0.579
Lama Usaha	> 5 Tahun	186	70	116	0.625	0.617
	3 - 5 Tahun	114	42	72	0.375	0.382
Sistem Penjualan	Cash 100 %	151	92	59	0,821	0,314
	Cash ≥ 80 % dan < 100 %, Kredit ≥ 10 % dan ≤ 20 %	49	20	29	0,179	0,154
Sistem Pembelian	Cash 100 %	256	85	171	0,759	0,910
	Cash ≥ 80 % dan < 100 %, Kredit ≥ 1 % dan ≤ 20 %	44	27	17	0,241	0,090
Omset/ Bulan	> 150 Juta	121	38	83	0,339	0,441
	> 100 - 150 Juta	31	2	29	0,018	0,154
	> 50 - 100 Juta	34	6	28	0,054	0,149
	> 25 - 50 Juta	95	65	30	0,580	0,160
	< 25 Juta	19	1	18	0,009	0,096
Gross Profit Margin	> 20 %	109	12	97	0,107	0,516
	> 15 % - 20 %	40	5	35	0,045	0,186
	> 10 % - 15%	41	8	33	0,071	0,176
	> 5 % - 10 %	110	87	23	0,777	0,122
Repayment Capacity	≥ 3 X Angsuran	206	108	98	0,964	0,521
	≥ 2,5 - 3 X Angsuran	81	3	78	0,027	0,415
	≥ 2 - 2,5 Angsuran	13	1	12	0,009	0,064
Fasilitas	KU1	186	82	104	0,732	0,553
	KU2	114	30	84	0,268	0,447

Sumber : hasil Penelitian 2016

Setelah melakukan pengolahan data tersebut diatas, maka terdapat dua *class* yang dibentuk pada *Probabilitas Prior*, yaitu:

Class Kredit = Lancar
Class Kredit = Macet

B. Menghitung Probabilitas Posterior.

Tahapan selanjutnya adalah menggunakan *Probabilitas Prior* untuk menentukan *class* terhadap temuan kasus baru, dengan cara terlebih dahulu menghitung *Probabilitas Posterior*nya, hal tersebut dilakukan apabila ditemukan kasus baru dalam pengolahan data. Berikut tabel *probabilitas posterior* untuk menghitung kasus baru yang ditemukan:

Tabel 3
Penghitungan *Probabilitas Posterior*.

Data X		P(X|Ci)	
Atrbibut	**Nilai (*Value*)**	**Lancar**	**Macet**
Jenis Usaha	1 (Primer)	0,33	0,68

Status Tempat Usaha	Kontrak	0,31	0,58
Lama Usaha	>3 – 5 Tahun	0,38	0,38
Sistem Penjualan	Cash 100%	0,82	0,31
Sistem Pembelian	Cash 100%	0,76	0,91
Omzet per Bulan	> 25 – 50 Juta	0,58	0,25
Gross Profit Margin	> 20 %	0,11	0,52
Repayment Capacity	≥ 3 X Angsuran	0,96	0,52
Fasilitas	KU1	0,73	0,55

Sumber : Hasil Penelitian 2016

Selanjutnya setelah mengetahui nilai *probabilitas* dari setiap atribut terhadap *probabilitas* tiap *class* atau yang dirumuskan dalam bentuk persamaan P(X|Ci), maka langkah berikutnya adalah melakukan penghitungan terhadap total keseluruhan *probabilitas* tiap *class*.Berikut persamaan untuk menghitung probabilitas tiap *class*:

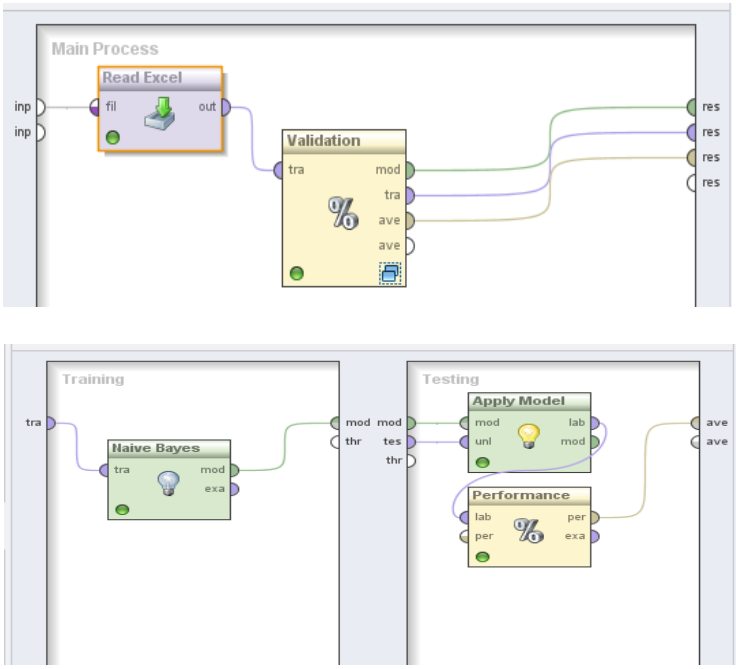
P(X| Kredit = Lancar)= P(Jenis Usaha = 1 (Primer) | Kredit = Lancar) * P(Status Tempat Usaha = Kontrak | Kredit = Lancar)* P(Lama Usaha = > 3 - 5 Tahun | Kredit = Lancar)* P(Sistem Penjualan = Cash 100 % | Kredit = Lancar)* P(Sistem Pembelian = Cash 100 % | Kredit = Lancar)* P(Omzet per Bulan = > 25 – 50 Juta | Kredit = Lancar) * P(Gross Profit Margin = > 20 % | Kredit = Lancar) * P(Repayment Capacity = ≥ 3 X Angsuran | Kredit = Lancar) * P(Fasilitas= KU1 | Kredit = Lancar)
= 0,33 * 0,31 * 0,38 * 0,82 *0,76 * 0,58 *0,11 * 0,96 *0,73
= **0,001083**

P(X| Kredit = Macet))= P(Jenis Usaha = 1 (Primer) | Kredit = Macet) * P(Status Tempat Usaha = Kontrak | Kredit = Macet)* P(Lama Usaha = > 3 - 5 Tahun | Kredit = Macet)* P(Sistem Penjualan = Cash 100 % | Kredit = Macet)* P(Sistem Pembelian = Cash 100 % | Kredit = Macet)* P(Omzet per Bulan = > 25 – 50 Juta | Kredit = Macet) * P(Gross Profit Margin = > 20 % | Kredit = Macet) * P(Repayment Capacity = ≥ 3 X Angsuran | Kredit = Macet) * P(Fasilitas= KU1 | Kredit = Macet)
= 0,68 * 0,58 * 0,38 * 0,31 * 0,91 * 0,25 * 0,52 * 0,52 *0,55
= **0,001571**

P(X|Kredit = Lancar)P(Lancar) = 0,001083 * 0,373 = **0,000403**

P(X|Kredit = Macet)P(Macet) = 0,001571 * 0,627 = **0,000985**

Hasil perhitungan terhadap *probabilitas* tiap *class* diatas, diketahui bahwa nilai P(X|Macet) lebih besar daripada nilai P(X|Lancar), sehingga dapat diambil kesimpulan bahwa dalam kasus Kredit tersebut akan masuk kedalam klasifikasi tingkat kredit **Macet**. Berikut gambar pengujian modelnya menggunakan *Tools Rapidminer*



Sumber : Hasil penelitian 2016

Gambar 2
Pengujian Model Algoritma Naïve Bayes *Classifier*

1.2 Evaluasi dan Validasi hasil.

Untuk mendapatkan nilai akurasi yang benar, maka diperlukan proses klasifikasi dan alat ukur yang tepat, karena keakuratan klasifikasi merupakan alat ukur yang bisa menunjukkan bagaimana cara untuk mengklasifikasi sehingga dapat mengidentifikasi data objek dengan benar [9]. Proses pengujian terhadap nilai akurasi dapat dilakukan dengan cara melakukan evaluasi tingkat akurasi dari algoritma. *Tools* yang digunakan untuk pengujian adalah *software rapid miner* serta mengunakan model *Confussion Matrix* dan kurva ROC (*Receiver Operating Characteristic*).

1.3 Hasil Pengujian Model Algoritma Naïve Bayes *Classifier*

Hasil dari pengujian model yang telah dilakukan adalah untuk mengukur tingkat akurasi dan AUC (*Area Under Cover*)

a. Confussion Matrix

Gambar 4 merupakan pengujian tools rapidminer dengan jumlah data 300 *record*. Berikut tabel yang didapat:

Tabel 4
Model *Confussion Matrix* algoritma Naïve Bayes *Classifier*

accuracy: 89.33% +/- 5.33% (mikro: 89.33%)

	true Lancar	true Macet	class precision
pred. Lancar	97	17	85.09%
pred. Macet	15	171	91.94%
class recall	86.61%	90.96%	

Sumber: Hasil Penelitian 2016

Jumlah *True Positive* (TP) adalah 97 record diklasifikasikan sebagai kredit LANCAR dan *False Negative* (FN) sebanyak 17 *record* diklasifikasikan sebagai kredit LANCAR tetapi kredit MACET. Berikutnya 171 record untuk *True Negative* (TN) diklasifikasikan sebagai kredit MACET, dan 15 *record False Positive* (FP) diklasifikasikan sebagai kredit MACET tetapi kredit LANCAR. Berdasarkan tabel 4.5 tersebut menunjukan bahwa, tingkat akurasi dengan menggunakan algortima Naïve Bayes *Classifier* adalah sebesar 89,33% dan dapat dihitung untuk mencari nilai *accuracy*, *sensitivity*, *specificity*, *ppv*, dan *npv* pada persamaan dibawah ini:

acc = (tp +tn) / (tp+tn+fp+fn) → acc = (97+171) / (97+171+15+17)

sensitivity = (tp) / (tp+ fn) → 97 / (97 +17)

specitivity = (tn) / (tn+ fp) → 171 / (171 +15)

ppv = (tp) / (tp+ fp) → 97 / (97 + 15)

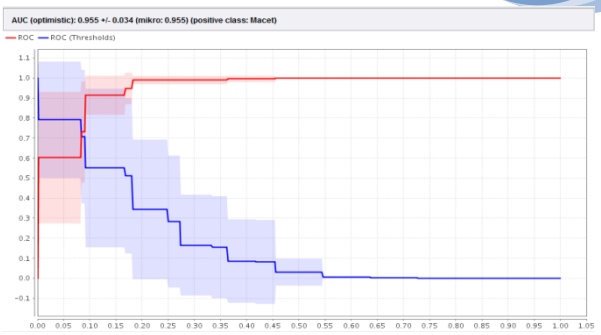
npv = (tn) / (tn+ fn) → 171 / (171 + 17)

Hasil dari perhitungan persamaan diatas terlihat pada Tabel 5 dibawah ini:

Tabel 5.
Nilai *accuracy*, *sensitivity*, *specificity*, *ppv* dan *npv*

Parameter	Nilai (%)
Accuracy	89,33%
Sensitivity	85.09%
Specitivity	91.94%
PPV	86.61 %
NPV	90.96%

b. Evaluasi ROC *Curve*
Grafik 4 merupakan terdapat grafik ROC dari tebel 5, dengan nilai AUC (*Area Under Cover*) sebesar 0.955 dengan nilai akurasi *Excellent Classification*.



Sumber : Hasil Penelitian 2016

Gambar 3.
Nilai AUC dalam grafik ROC Algoritma Naïve Bayes

V. KESIMPULAN

Hasil penelitian untuk nilai akurasi algoritma klasifikasi Naïve Bayes *Classifier* adalah 89.33%, Sementara untuk evaluasi menggunakan ROC *Curve* untuk model klasifikasi Naïve Bayes *Classifier* nilai AUC adalah 0.955 dengan tingkat diagnosa *Excellent Classification*.

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

1. Penelitian ini hanya menggunakan metode naive bayes diharapkan untuk penelitian selanjutnya bisa digunakan metode lain sehingga dapat dikomparasi
2. Untuk nilai akurasi yang lebih baik sebaiknya menggunakan metode optimasi seperti PSO (*Particle Swarm Optimiztion*), GA (*Genetic Algorithm*) dan lainnya
3. Penambahan jumlah data dan atribut dapat memungkinkan untuk meningkatkan nilai akurasi
4. Pada Bank Mayapada Mitra Usaha Cabang PGC, dapat ditingkatkan sistem analisa kredit untuk penentuan kelayakan pemberian kredit bagi calon debitur.

REFERENSI

[1] Xhemali, D., Hinde, C.J. and Stone. R.G. (2009). *Naive Bayes vs Decision Trees vs. Neural Network in the Classification of Training Web Pages*. IJCSI International Journal of Computer Science Issues, 4(1). Pp. 16-23.

[2] Yap, Bee W., Ong, Seng H., and Husain. N.H.M, (2011). *Using Data Mining to Improve Assessment of Credit Worthiness via Credit Scoring Models*. Expert System with Applications, 38(2011) 13274-13283

[3] Undang-Undang Republik Indonesia Nomor 10 Tahun 1998 tentang perubahan atas Undang-Undang Nomor 7 Tahun 1992 tentang perbankan.

[4] Larose, D.T.(2005). *Discovering Knowledge in Data*. Canada: Wiley-Interscience.

[5] Larose, D.T.(2005). *Discovering Knowledge in Data*. Canada: Wiley-Interscience.

[6] Xhemali, D., Hinde, C.J. and Stone. R.G. (2009). *Naive Bayes vs Decision Trees vs. Neural Network in the Classification of Training Web Pages*. IJCSI International Journal of Computer Science Issues, 4(1). Pp. 16-23.

[7] Han, J., and Kamber, M. (2006). *Data Mining Concept and Techniques*. San Francisco: Diane Cerra.

[8] Bramer, Max. (2007). *Principles of Data Mining*. London: Springer. ISBN-10: 1-84628-765-0, ISBN-13: 978-1-84628-765-7

[9] Gorunescu, Florin. (2011). *Data Mining Concepts, Models and Techniques*. Intelligent System Reference Library, Vol 12, ISBN 978-3-642-19721-5.

[10] Han, J., and Kamber, M. (2006). *Data Mining Concept and Techniques*. San Francisco: Diane Cerra.

[11] Vercellis, Carlo. (2009). Business Intelligence: Data Mining and Optimization for Decision Making. United Kingdom: John Willey & Son