

# PREDIKSI TINGKAT KELANCARAN PEMBAYARAN KREDIT BANK MENGUNAKAN ALGORITMA NAIVE BAYES BERBASIS FORWARD SELECTION

Maryam Hasan

maryamhasan130790@gmail.com  
Fakultas Ilmu Komputer Universitas Ichsan Gorontalo

## Abstrak

Tujuan dari Penelitian ini adalah Mengoptimalkan algoritma *Naive Bayes* dengan seleksi fitur *Forward Selection* untuk dapat meningkatkan hasil akurasi atau tingkat keberhasilan yang didapatkan dari prediksi pembayaran kredit. Data yang akan digunakan dalam penelitian ini berasal dari Bank XY yang berada di Gorontalo. Data yang diperoleh berkaitan dengan semua aspek dari nasabah kredit termasuk informasi pribadi dari nasabah. Desain eksperimen dalam penelitian ini menggunakan dataset nasabah kredit. Sedangkan analisis yang digunakan adalah Model algoritma *Naive Bayes* dengan seleksi fitur *Forward Selection*. Prediksi tingkat kelancaran pembayaran kredit menggunakan algoritma *Naive Bayes* berbasis *Forward Selection* mampu memprediksi kelancaran pembayaran kredit ke depannya hal ini terbukti dengan perolehan nilai akurasi *Naive bayes* berbasis *Forward Selection* mampu mencapai nilai akurasi 71,97 %.

**Kata kunci:** Risiko Kredit, Penilaian Kredit, Algoritma *Naive Bayes*, *Forward Selection*

## 1. PENDAHULUAN

Keuangan merupakan tulang punggung dari setiap bisnis, sebagai bisnis yang tumbuh perlu lebih banyak modal untuk membantu berbagai kegiatan operasional dan non-operasional. Untuk membiayai kegiatan tersebut setiap organisasi memiliki dua sumber utama yaitu, Sumber eksternal untuk membiaya bisnis tersebut dilakukan dengan melakukan pinjaman dari lembaga keuangan seperti bank, sedangkan sumber internal sepenuhnya menggunakan modal sendiri [1]. Bank merupakan badan usaha yang menghimpun dana dari masyarakat dalam bentuk simpanan dan menyalurkannya kepada masyarakat dalam bentuk kredit dan atau bentuk-bentuk lainnya dalam rangka meningkatkan taraf hidup rakyat banyak [2]. Perbankan menyediakan berbagai macam alternatif pinjaman uang bagi nasabah salah satunya adalah melalui pemberian pinjaman berupa kredit kepada nasabah.

Sejatinya dalam pemberian kredit kepada nasabah pihak bank mengalami berbagai masalah atau risiko. Salah satu masalah atau risiko yang dialami bank dalam pemberian kredit adalah perilaku nasabah yang tidak membayar angsuran tepat waktu ataupun menunda sampai beberapa bulan pembayaran angsuran yang pada akhirnya menyebabkan kredit macet. Hal ini merupakan masalah yang serius yang perlu diperhatikan oleh pihak bank untuk lebih berhati-hati dalam menentukan nasabah karena dalam pemberian kredit sangat berisiko. Perlu adanya suatu metode perhitungan dengan sistem komputasi cerdas untuk membantu pihak bank dalam menyeleksi calon nasabah yang akan diberikan pinjaman. Metode yang digunakan seperti *Regression Analysis*, *Neural Networks*, *Genetic Algorithms*, *Decision Tree*, dan *Support Vector Machines (SVM)* [3].

Bahramirzaee et al menyatakan bahwa berbagai macam perbankan menghadapi beberapa risiko kredit yang berbeda, diantaranya memiliki dampak sosial yang potensial karena jumlah dan keragaman stakeholder yang terkena dampak kegagalan bisnis yang mempengaruhi pemegang saham, seperti manajer, pemberi pinjaman (bank), pemasok, klien, masyarakat keuangan, pemerintah, pesaing, dan badan pengawas. Oleh karena itu agar lebih efektif mengelola masalah ini, perlu adanya system pendukung keputusan yang kuat dan canggih yang didukung oleh alat analisis untuk mengukur, memonitor, mengelola, dan kontrol, risiko dan inefisiensi keuangan dan operasional. Sistem keputusan disini merupakan system komputasi cerdas dengan berbagai metode-metode komputer atau dengan berbagai metode algoritma yang berhubungan dengan management risiko kredit [4].

Lean Yu, Xiao Yao and Shouyang Wang mengusulkan penelitian sebuah pendekatan berbobot least squares support vector Machine (LSSVM) dengan desain of experiment (DOE) untuk menunjukkan bagaimana membangun LSSVM classifier berbobot dengan DOE, bagaimana mengevaluasi risiko kredit menggunakan LSSVM classifier berbobot dengan DOE dan bagaimana menampilkan beberapa metode untuk membandingkan kinerja dalam mengevaluasi risiko kredit. Penelitian ini menggunakan data set kredit Jerman dan data set kredit Australia. Hasil penelitian

secara empiris menemukan bahwa seluruh model yang berbeda untuk menguji kelas dari dua data set kredit yang diusulkan. LSSVM classifier berbobot dengan DOE untuk pemilihan parameter menunjukkan hasil akurasi terbaik dari total akurasi [5].

S. B. Kotsiantis et al mengusulkan sebuah sistem berbasis ontology yang memprediksi risiko kredit untuk membedakan pembayar yang baik dan pembayar yang buruk. Penelitian ini menggunakan data set Jerman yang berisi pengamatan terhadap variabel untuk membedakan pembayar yang baik dan pembayar yang buruk. Adapun parameter yang digunakan seperti status account, jangka waktu kredit dalam beberapa bulan, sejarah kredit, tujuan kredit, jumlah kredit, rata-rata saldo rekening tabungan, tingkat angsuran sebagai persen dari pendapatan, status pribadi, pihaklain, besarnya property, usia, rencana pembayaran lainnya, perumahan, jumlah kredit yang di bank, pekerjaan, jumlah orang siapa yang bertanggung jawab untuk menyediakan pemeliharaan, pemohon memiliki telepon, dan tenaga kerja asing [6]. Algoritma *Naive Bayes* merupakan teknik prediksi yang berbasis probabilistik sederhana yang berdasar pada penerapan teorema Bayes dengan asumsi yang independensi (ketidak tergantungan) yang kuat (naïf) [6].

Algoritma *Naive Bayes* juga merupakan algoritma klasifikasi yang sederhana karena setiap atribut bersifat independen dan memungkinkan kontribusi terhadap keputusan akhir [7]. Meskipun algoritma *Naive Bayes* sederhana tetapi dapat mengungguli metode lain yang lebih canggih. Agar dapat membangun model algoritma *Naive Bayes* perlu menghitung probabilitas dari setiap nilai atribut yang diberikan oleh atribut class. Maka dari itu jika menggunakan atribut dengan jumlah besar (banyak) nilainya, mungkin memerlukan waktu yang lebih lama [8]. Dari permasalahan algoritma *Naive Bayes* tersebut dilakukan pengoptimalan seleksi yang di mana untuk memilih atribut yang relevan untuk diberikan [1]. Pengoptimalan seleksi yang dipilih yaitu Forward selection. Forward selection digunakan untuk menyeleksi setiap fitur yang tidak terpakai saat memulai iterasi fitur, akan ditambahkan pada subset fitur yang dipilih sebelumnya. Semua subset yang dihasilkan akan dievaluasi dengan menerapkan pembelajaran dan subset yang terbaik di gunakan sebagai dasar untuk iterasi berikutnya [9].

Untuk mengatasi masalah atau risiko tentang pemberian kredit kepada nasabah yang mungkin membayar angsuran tepat waktu penelitian ini menggunakan algoritma *Naive Bayes*. Algoritma *Naive Bayes* ini digunakan karena klasifikasinya yang sederhana dan dapat memprediksi di masadepan. Oleh karena itu penelitian ini di harapkan dapat memprediksi nasabah kredit dan dapat membantu pihak bank dalam pengambilan keputusan pemberian kredit. Dari latar belakang tersebut permasalahan yang muncul adalah Hasil akurasi atau tingkat keberhasilan yang didapatkan dari prediksi tingkat kelancaran pembayaran kredit dengan menggunakan algoritma *Naive Bayes* masih rendah sehingga hasilnya tidak akurat. Perlu penambahan fitur lain untuk mendukung model ini. Oleh karena itu penelitian ini bertujuan untuk mengoptimalkan algoritma *Naive Bayes* dengan seleksi fitur *Forward Selection* dapat meningkatkan hasil akurasi atau tingkat keberhasilan yang didapatkan dari prediksi pembayaran kredit.

## 2. LANDASAN TEORI

### 2.1 Kredit

Kredit berasal dari bahasa Yunani "*Credere*" yang berarti kepercayaan. Kredit berasal dari bahasa Latin "*Creditum*" yang berarti kepercayaan dalam kebenaran [10]. Menurut Undang-Undang Republik Indonesia nomor 10 Tahun 1998 Kredit adalah penyediaan uang atau tagihan yang dapat dipersamakan dengan itu, berdasarkan persetujuan atau kesepakatan pinjam-meminjam antara bank dengan pihak lain yang mewajibkan pihak peminjam untuk melunasi hutangnya setelah jangka waktu tertentu dengan pemberian bunga [2]. Ada beberapa prinsip penilaian kredit yaitu [11] :

- Character* merupakan watak atau perilaku seseorang yang dalam hal ini yaitu calon nasabah. Tujuannya untuk memberikan keyakinan kepada pihak bank, bahwa watak atau perilaku dari orang yang akan diberikan kredit dapat dipercaya.
- Capitalty* merupakan kemampuan calon nasabah untuk membayar kreditannya yang dihubungkan dengan kemampuan mengelola bisnisnya serta kemampuan mencari laba.
- Capital*, dapat mengetahui dari mana sumber-sumber pemberian yang dimiliki oleh nasabah terhadap usaha yang akan dibiayai oleh pihak bank.
- Collateral* merupakan jaminan yang akan diberikan oleh calon nasabah baik yang bersifat fisik atau yang non fisik.
- Cadition* merupakan dapat menilai kredit yang hendaknya dinilai dari kondisi ekonomi sekarang dan untuk masa depan yang akan datang yang sesuai sektor masing-masing.

## 2.2 Risiko Kredit

Risiko kredit adalah sebagai risiko kerugian sehubungan dengan pihak peminjam yang tidak dapat dan atau tidak mau memenuhi kewajiban untuk membayar kembali dana yang dipinjamnya secara penuh pada saat jatuh tempo atau sesudahnya. Dengan kata lain risiko ini timbul karena adanya ketidak pastian tentang pembayaran kembali pinjaman oleh debitur (nasabah). Oleh karena itu, pihak bank harus berhati-hati, cermat, dan teliti dalam memilih calon debitur (nasabah) [10]. Disisi lain, risiko ini timbul karena kinerja satu atau lebih nasabah yang buruk. Kinerja nasabah yang buruk ini dapat berupa ketidak mampuan atau ketidak mauan nasabah untuk memenuhi sebagian atau seluruh isi perjanjian kredit yang telah disepakati bersama sebelumnya. Dalam hal ini yang menjadi perhatian bank bukan hanya kondisi keuangan dan nilai pasar dari jaminan kredit termasuk *collateral* tetapi juga karakter dari nasabah [12].

Menurut kountur (2006, hal 3) risiko kredit merupakan kemungkinan kejadian yang merugikan. Risiko akan menjadi besar / kompleksnya aktifitas yang dilakukan maka semakin besar risiko yang dihadapi. Sedangkan menurut tampubolon (2007, hal 21) risiko bank merupakan sebagian kombinasi dari tingkat kemungkinan sebuah peristiwa terjadi disertai dampak dari peristiwa tersebut pada bank. Setiap kegiatan mengandung potensi sebuah peristiwa terjadi atau tidak terjadi, dengan dampak member peluang untuk hutang atau mengancam sebuah kesuksesan [13]. Ukuran yang dapat digunakan untuk mengukur risiko kredit yaitu, sebagai berikut [14] :

- a. Probabilitas gagal bayar (*default probability*)
- b. Jarak gagal bayar (*distance to default*)
- c. Tingkat pengambalian hutang pada saat perusahaan mengalami gagal bayar (*recovery rate*)

## 2.3 Data Mining

Data mining merupakan kegiatan penemuan pola-pola yang menarik dari data berukuran besar yang disimpan dalam basis data, data *warehouse*, atau sarana penyimpanan yang lain. Data mining dapat diklasifikasikan menjadi dua kategori: *descriptive data mining* dan *predictive data mining* [15]. Data mining sering disebut sebagai *Knowledge Discovery in Database (KDD)* yang bertugas untuk mengekstrak pola atau model dari data dengan menggunakan suatu algoritma yang spesifik [16]. Dari beberapa pengertian tersebut dapat ditarik kesimpulan data mining merupakan suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (database) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui. Data mining itu sendiri merupakan usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar oleh karena itu data mining memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan database [16]. Tahapan untuk melakukan data mining seperti berikut:

- a. Pembersihan data (*data cleaning*) merupakan proses menghilangkan noise dan data yang tidak konsisten atau yang tidak relevan.
- b. Integrasi data (*data integration*) merupakan penggabungan data dari berbagai database ke dalam satu database baru.
- c. Seleksi data (*data selection*) merupakan data yang tidak sesuai atau yang tidak perlu dianalisis tidak diambil sedangkan data yang sesuai untuk dianalisis yang akan diambil di database.
- d. Transformasi data (*data transformation*) merupakan data yang akan diubah atau digabungkan ke dalam format yang sesuai untuk diproses dalam data mining.
- e. Proses mining merupakan suatu proses utama saat metode diterapkan untuk menentukan pengetahuan berharga dan tersembunyi dari data.
- f. Evaluasi pola (*pattern evaluation*) merupakan pengidentifikasian pola-pola menarik ke dalam *knowledge based* yang ditemukan.
- g. Presentasi pengetahuan (*knowledge presentation*) merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

## 2.4 Naive Bayes

Algoritma *Naive Bayes* ini memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan inggris Thomas Bayes yang merupakan memprediksi probabilitas dimasa depan berdasarkan pengalaman dimasa sebelumnya. *Naive Bayes* classidier atau bisa disebut dengan multinominal *Naive Bayes* yang merupakan model sederhana dari algoritma Bayes yang cocok dalam pengklasifikasian text atau dokumen [17]. *Naive Bayes* classifier merupakan salah satu algoritma dalam teknik data mining yang menerapkan teori Bayes dalam klasifikasi [18].

Navie *Bayes* merupakan algoritma klasifikasi yang sederhana dimana setiap atribut bersifat independent dan memungkinkan berkontribusi terhadap keputusan akhir [19]. *Bayesian Classification* merupakan pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Bayesian Classification* ini juga didasarkan pada teorema *Bayes* yang memiliki kemampuan klasifikasi setupa dengan *decision tree* dan *neural network* [20]. Dasar teorema *Naive Bayes* dengan rumus *Bayes* yaitu sebagai berikut [21] :

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \dots\dots\dots (1)$$

$$P(H|X) = P(X|H) P(H) \dots\dots\dots (2)$$

Keterangan [20] :

- 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 probabilitas)
- P(H) = Probabilitas hipotesis H (prior probabilitas)
- P(X|H) = Probabilitas X berdasarkan kondisi tersebut
- P(X) = Probabilitas dari X

Dapat dijelaskan dengan dimana  $P(H|X)$  itu merupakan probabilitas dari H yang di dalam X atau bahasa lain  $P(H|X)$  merupakan persentase dengan banyaknya H yang di dalam X,  $P(X|H)$  merupakan probabilitas X di dalam H,  $P(H)$  merupakan probabilitas prior dari H dan  $P(X)$  merupakan probabilitas prior dari X. Untuk klasifikasi dengan data kontinyu digunakan rumus *Densitas Gauss* [22]:

$$P(X_i = x_i | Y = Y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \dots\dots\dots (3)$$

Keterangan :

- P = Peluang
- X<sub>i</sub> = Atribut ke i
- x<sub>i</sub> = Nilai Atribut ke i
- Y = Kelas yang dicari
- y<sub>j</sub> = Sub kelas Y yang dicari
- μ = Mean, menyatakan rata – rata dari seluruh atribut
- σ = Deviasi standar, menyatakan varian dari seluruh atribut

Adapun alur dari metode *Naive Bayes* sebagai berikut :

- a Baca data traning
- b Hitung jumlah probabilitas, namun apabila data numerik maka :
  - 1) Cari nilai mean dan standar deviasi dari masing – masing parameter yang merupakan data numerik
  - 2) Cari nilai probabilitik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut
- c Mendapatkan nilai dalam tabel mean, standar deviasi dan probabilitas

### 2.5 Forward Selection

*Forward Selection* merupakan menyeleksi variabel berdasrkan koefisien korelasi dan meregresikan variabel-variabel bebas X satu demi satu sampai diperoleh persamaan yang sempurna [23]. *Forward Selection* dimulai dengan pemilihan atribut yang kosong serta dalam setiap putaran ia menambahkan setiap atribut yang tidak terpakai sebagai contoh set. Untuk setiap atribut yang ditambahkan diperkirakan menggunakan kinerja operator batin, misalnya cross validasi. Yang dimana hanya memberikan atribut yang kinerjanya tinggi untuk ditambahkan ke seleksi. Kemudian memulai babak baru untuk pemilihan dengan memodifikasi [9]. Untuk prosedur *Forward Selection* dapat di rumuskan sebagai berikut [24] :

- a. Menentukan model awal  $\hat{y} = b_0$
- b. Memasukan variabel respon dengan setiap variabel berprediktor, misalnya X<sub>1</sub>, X<sub>2</sub>, .... X<sub>n</sub> yang terkait dengan  $\hat{y}$ . Misalkan X<sub>1</sub> sehingga membentuk model  $\hat{y} = b_0 + b_1 X_1$ .
- c. Uji F terhadap peubah pertama yang terpilih. Jika  $F_{hitung} < F_{tabel}$  maka peubah terpilih dibuang dan proses dihentikan. Apa bila  $F_{hitung} > F_{tabel}$  maka peubah terpilih memiliki pengaruh nyata terhadap peubah terkait y, sehingga layak untuk di perhitungkan di dalam model.

- d. Masukkan peubah bebas terpilih (yang paling signifikan) ke dalam model. Misalkan  $x_2$ , sehingga membentuk suatu model  $\hat{y} = b_0 + b_1x_1 + b_2x_2$
- e. Uji F, jika  $F_{hitung} < F_{tabel}$  maka proses dihentikan dan model terbaik adalah model sebelumnya. Namun jika  $F_{hitung} \geq F_{tabel}$ , variabel peubah bebas layak untuk dimasukkan ke dalam model dan kembali ke langkah c. Proses akan berakhir jika tidak ada lagi peubah yang tersisa yang bisa dimasukkan ke dalam model.

## 2.6 Confusion Matrix

Dalam bidang pembelajaran mesin *Confusion Matrix* dikenal sebagai table kemungkinan atau *matrix error* yang merupakan tata letak table khusus yang memungkinkan visualisasi kinerja algoritma yang biasanya *supervised learning* (dalam *unsupervised learning* itu biasanya disebut *matching matrix*). Yang setiap kolom dari matrix merupakan contoh di kelas prediksi, sementara setiap baris mewakili contoh di kelas sebenarnya [25]. *Confusion Matrix* juga memberikan keputusan yang diperoleh dalam training dan testing dan memberikan penilaian dalam performance klasifikasi berdasarkan objek dengan benar atau salah [3].

Tabel 1. *Confusion Matrix*

Classification		Predicted Class	
		Class = Yes / Positive	Class = No / Negatif
Obsterved Class	Class = Yes / Positive	A (True Positive = TP)	B (False Negatif = FN)
	Class = No / Negatif	C (False Positive = FP)	D (True Negatif = TN)

Untuk menghitung nilai akurasi dengan rumus perhitungan sebagai berikut :

$$Accuracy = \left( \frac{a+d}{a+b+c+d} \right) = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots (4)$$

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots (5)$$

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots (6)$$

Keterangan :

- TP = Prediksi positif yang positif
- FN = Prediksi positif yang negatif
- FP = Prediksi negatif yang positif
- TN = Prediksi negatif yang negative

## 2.7 Regresi Linier

Regresi linier adalah metode statistika yang digunakan untuk membentuk model hubungan antara variabel terikat (dependen; respon; Y) dengan satu atau lebih variabel bebas (independen, prediktor, X). Apabila banyaknya variabel bebas hanya ada satu, disebut sebagai regresi linier sederhana, sedangkan apabila terdapat lebih dari 1 variabel bebas, disebut sebagai regresi linier berganda. Analisis regresi setidaknya-tidaknya memiliki 3 kegunaan, yaitu untuk tujuan deskripsi dari fenomena data atau kasus yang sedang diteliti, untuk tujuan kontrol, serta untuk tujuan prediksi. Regresi mampu mendeskripsikan fenomena data melalui terbentuknya suatu model hubungan yang bersifat numerik. Regresi juga dapat digunakan untuk melakukan pengendalian (kontrol) terhadap suatu kasus atau hal-hal yang sedang diamati melalui penggunaan model regresi yang diperoleh. Selain itu, model regresi juga dapat dimanfaatkan untuk melakukan prediksi untuk variabel terikat [26]. Bentuk umum model regresi linier berganda, yaitu [27] :

$$Y_i = \beta_0 + \beta_1 X_{i1} + \epsilon_i \dots\dots\dots (7)$$

Keterangan :

- $Y_i$  = variabel respon atau variabel tidak bebas
- $\beta_0$  dan  $\beta_1$  = parameter
- $X_{i1}$  = nilai diketahui variabel prediktor atau variabel bebas

$\epsilon$ =standar error

### 3. METODE PENELITIAN

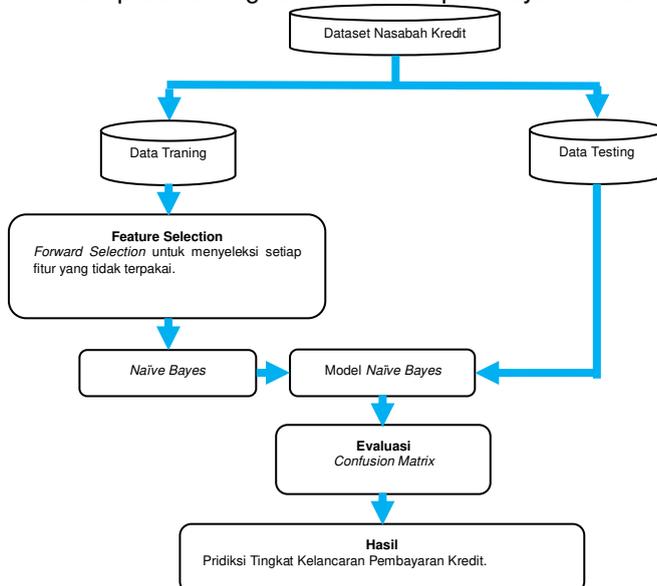
#### 3.1 Metode Pengumpulan Data

Data yang akan digunakan dalam penelitian ini berasal dari Bank XY yang berada di Gorontalo. Data yang diperoleh berkaitan dengan semua aspek dari nasabah kredit termasuk informasi pribadi dari nasabah. Parameter yang menentukan kelancaran kredit adalah usia, status perkawinan, pekerjaan, pendapatan, jaminan, jumlah tanggungan, bunga kredit, jumlah bunga kredit, jumlah pokok, jumlah angsuran, jangka waktu kredit, jumlah pinjaman, dan keterangan (lancar, dalam perhatian khusus, kurang lancar, diragukan dan macet). Jumlah dataset yang akan digunakan dalam penelitian ini 874 record.

#### 3.2 Desain Eksperimen

Desain eksperimen dalam penelitian ini menggunakan dataset nasabah kredit. Model desain eksperimen dapat dijelaskan dengan beberapa tahap yaitu :

- Pengumpulan dataset nasabah kredit yang dimana pengumpulan dataset nasabah kredit pada bank XY yang berjumlah 874 record.
- Featur Selection untuk pemilihan variabel-variabel yang relevan untuk diberikan atau yang berpengaruh terhadap dataset nasabah kredit. Featur Selection yang digunakan *Forward Selection* dimana dilakukan untuk menyeleksi variabel berdasarkan koefisien korelasi dan meregresikan variabel-variabel bebas X satu demi satu sampai diperoleh persamaan yang sempurna.
- Pembagian Dataset Dataset tersebut akan dibagi menjadi dua (2) bagian yaitu, data training dan data testing. Pembagian data akan dilakukan dengan menggunakan X-Validation (Cross-Validation)
- Algoritma *Naive Bayes* Setelah dilakukannya pembagian data, maka data training akan diproses dengan menggunakan algoritma *Naive Bayes* yang dimana untuk mendapatkan model yang terbaru.
- Model *Naive Bayes* Data testing tersebut akan diproses dengan menggunakan model *Naive Bayes* yang dimana untuk mendapatkan hasil akurasi atau tingkat keberhasilan yang maksimal dari prediksi tingkat kelancaran pembayaran kredit.



Gambar 2 Desain Eksperimen

#### 3.3 Evaluasi

Untuk proses evaluasi ini dimana akan memberikan akurasi atau tingkat keberhasilan dari prediksi yang telah dilakukan. Untuk proses evaluasi ini menggunakan *confusion matrix* yang dimana akan memudahkan peneliti untuk mendapatkan tingkat akurasi dari prediksi pembayaran kredit.

## 4. HASILDAN PEMBAHASAN

### 4.1 Hasil Pengumpulan Data

Dataset ini merupakan dataset nasabah kredit yang sudah diketahui keterangannya (lancar, dalam perhatian khusus, kurang lancar, diragukan, dan macet) dalam kredit. Jumlah dataset yang akan digunakan 874 *record*.

### 4.2 Hasil Eksperimen dan Implementasi

#### 4.2.1 Evauasi Menggunakan Algoritma *Naive Bayes*

Dalam penelitian ini, data yang digunakan adalah data nasabah kredit. Penelitian ini menggunakan model 10 *fold Cross Validation* tanpa menggunakan *feature selection*. Dari hasil eksperimen dari dataset nasabah kredit yang berjumlah 874 *record* yang berdasarkan tabel *confusion matrix* hasil *accuracy* yang didapatkan pada dataset nasabah kredit dengan menggunakan algoritma *Naive Bayes* tanpa menggunakan *feature selection* adalah 68,42%.

#### 4.2.2 Evaluasi Menggunakan Algoritma *Naive Bayes* Berbasis *Forward Selection*

Dari hasil eksperimen menggunakan Algoritma *Naive Bayes* Berbasis *Forward Selection* terpilih 5 atribut yang dianggap relevan dan dapat mempengaruhi tingkat akurasi dengan menggunakan *Forward Selection* yaitu, pekerjaan, pendapatan, jumlah tanggungan, bunga, dan jangka waktu kredit. Didapatkan hasil *accuracy* pada dataset nasabah kredit dengan menggunakan algoritma *Naive Bayes* berbasis *Forward Selection* adalah 71,97%.

### 4.3 Prediksi Tingkat Kelancaran Pemberian Kredit Dengan Menggunakan Algoritma *Naive Bayes* Berbasis *Forward Selection*

Prediksi tingkat pembayaran kredit yang dilakukan oleh bank untuk mengetahui apakah bank telah mencapai keberhasilan yang telah ditargetkan dalam memprediksi tingkat kelancaran pembayaran kredit, tujuannya untuk mengurangi risiko dalam keputusan pemberian kredit dan dapat membantu meningkatkan modal agar pihak bank tidak mengalami krisis keuangan, apabila dapat memprediksi tingkat kelancaran pembayaran kredit dan nasabah dapat membayar kreditnya. Melakukan prediksi tingkat pembayaran kredit merupakan hal penting karena menjadi landasan kebijakan masalah pembayaran kredit dan algoritma *Naive Bayes* berbasis *Forward Selection* untuk mendapatkan *accuracy* dalam memprediksi tingkat pembayaran kredit. Berdasarkan evaluasi model yang dilakukan dapat diketahui bahwa algoritma *Naive Bayes* berbasis *Forward Selection* dapat diterapkan berdasarkan atribut nasabah kredit, serta mampu secara optimal memprediksi tingkat kelancaran pembayaran kredit.

## 5. KESIMPULAN

### 5.1 Kesimpulan

1. Prediksi tingkat kelancaran pembayaran kredit bank menggunakan algoritma *Naive Bayes* berbasis *Forward Selection* dapat memprediksi tingkat pembayaran kredit. Hal ini sangat penting dilakukan karena untuk mengurangi risiko dalam keputusan pemberian kredit dan dapat membantu meningkatkan modal agar pihak bank tidak mengalami krisis keuangan, apabila dapat memprediksi tingkat kelancaran pembayaran kredit dan nasabah dapat membayar kreditnya. Dengan adanya prediksi tingkat kelancaran pembayaran kredit pihak bank dapat mengambil keputusan untuk memberikan kredit kepada nasabah.
2. Prediksi tingkat kelancaran pembayaran kredit menggunakan algoritma *Naive Bayes* berbasis *Forward Selection* mampu memprediksi kelancaran pembayaran kredit ke depannya hal ini terbukti dengan perolehan nilai akurasi *Naive bayes* berbasis *Forward Selection* mampu mencapai nilai akurasi 71,97%

### 5.2 Saran

1. Pertimbangan dalam penggunaan dataset lain dalam penelitian untuk *accuracy* yang akan dihasilkan oleh *Naive Bayes*.
2. Kiranya prediksi tingkat kelancaran pembayaran kredit ini dapat membantu pihak bank sehingga dapat memudahkan pihak bank dalam masalah risiko keputusan pemberian kredit.
3. Kiranya untuk dapat meningkatkan *accuracy* dalam prediksi tingkat pembayaran kredit dapat dilakukan beberapa penggabungan algoritma dan menggunakan *feature selection* lain misalnya algoritma genetika.

### Daftar Pustaka

- [1] Kashif Imran “Determinants of bank credit in Pakistan: A supply side approach”. *Economic Modelling* 35 (2013) 384–390.
- [2] Undang-Undang Republik Indonesia Nomor 10 Tahun 1998 Tentang Perubahan Atas Undang-Undang Nomor 7 Tahun 1992 Tentang Perbankan
- [3] Jiawei Han and Micheline Kamber “Data Mining : Concepts and Tecniques”. Department of Computer Science University of Illinois at Urbana-Champaign. [www.cs.uiuc.edu/~hanj](http://www.cs.uiuc.edu/~hanj), 2006
- [4] Arash Bahrammirzaee *et al* “Hybrid credit ranking intelligent system using expert system and artificial *Neural Networks*”. Department of Information Technology Management, Tarbiat Modares University, Tehran, Iran. *Appl Intell* (2011) 34: 28–46. DOI 10.1007/s10489-009-0177-8
- [5] Lean Yu *et al* “Credit risk evaluation using a weighted least squares SVM classifier with design of experiment for parameter selection”. *Expert Systems with Applications* 38 (2011) 15392–15399
- [6] Harry Zhang “The Optimality of Naive Bayes”. American Association for Artificial Intelligence ([www.aaai.org](http://www.aaai.org)). All rights reserved, 2004
- [7] Yi Jiang “Credit Scoring Model Based on the *Decision Tree* and the Simulated Annealing Algorithm”. World Congress on Computer Science and Information Engineering, 2008 IEEE DOI 10.1109/CSIE.2009.481
- [8] Charu C. Aggarwal and Philip S. Yu “*A General Survey of Privacy-Preserving Data Mining Models and Algorithms*”, Springer, 2008
- [9] Fareed Akthar dan Caroline Hahne “RapidMiner 5 Operator Reference”. Rapid-I GmbH. All rights reserved © 2012
- [10] Farianto Pandila, SE., MM, “Manajemen Dana dan Kesehatan Bank”, Soft Cover, Jakarta : Rineka Cipta, 2012, 169 – 204
- [11] Nani Triwahyuniati, Sh “Pelaksanaan Analisis Pemberian Kredit Di Pt Bank Huga Cabang Semarang”. Tesis, 2008
- [12] Nisa’ Mustikawati *et al* “Penerapan Manajemen Risiko Untuk Meminimalisir Risiko Kredit Macet (Studi Pada PT. Bank Tabungan Pensiunan Nasional Cabang Kediri)”. Artikel, 2013
- [13] Dharma Setiawan “Analisis Terhadap Penerapan Manajemen Risiko Kredit Pada Pt. Bank Ekspor Indonesia”. <http://papers.gunadarma.ac.id>. 2007
- [14] Harjum Muharam “Model Risiko Kredit: Pendekatan Dan Faktor-Faktor Yang Mempengaruhinya”. <http://eprints.unisbank.ac.id/id/eprint/167>. 2012
- [15] Irman Hermadi *et al* “*Clustering Menggunakan Self Organizing Maps* (Studi Kasus: Data PPMB IPB)”. Related articles, 2007
- [16] Dian Wirdasari dan Ahmad Calam “Penerapan Data Mining Untuk Mengelola Data Penempatan Buku Di Perpustakaan SMK TI PAB 7 Lubuk Pakam Dengan Metode Association Rule”. *Jurnal SAINTIKOM*. Vol. 10 / No. 2 / Mey 2011
- [17] Prasetyo Anugroho *et al* “Klasifikasi Email Spam Dengan Metode *Naive Bayes Classifier* Menggunakan Java Programming”. Related articles, 2010
- [18] Mujib Ridwan *et al* “Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier”. *Jurnal EECCIS* Vol.7, No. 1, Juni 2013
- [19] Ivan Jaya “Analisis Seleksi Atribut Pada Algoritma *Naive Bayes* Dalam Memprediksi Penyakit Jantung”. Tesis, 2013
- [20] Arief Jananto “Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa”. *Jurnal Teknologi Informasi DINAMIK* Volume 18, No.1, Januari 2013 : 09-16. ISSN : 0854-9524
- [21] Jiawei Han and Micheline Kamber “Data Mining: Concepts and Techniques (2nd edition)”. Related articles, 2006
- [22] Bustami “Penerapan Algoritma *Naive Bayes* Untuk Mengklasifikasi Data Nasabah Asuransi”. *Jurnal Informatika* Vol. 8, No. 1, Januari 2014
- [23] Khairul Saleh “Penyeleksian Variabel Dengan Prosedur Regresi Bertahap”. *Jurnal Eksakta – BIAGROTEK*, ISSN : 2085 – 0646. Volume 2, Nomor 2, Juli 2010
- [24] Rian Mulyana “Aplikasi Metode *Variable Selection* Untuk Menentukan Faktor Dominan Yang Mempengaruhi Pendidikan Dan Kesehatan”. <http://repository.ipb.ac.id>. 2012
- [25] Stephen V. Stehman “Selecting And Interpreting Measures Of Thematic Classification Accuracy”. *Remote Sens. Environ.* 62:77--89 (1997).
- [26] Deny Kurniawan “REGRESI LINIER (LINEAR REGRESSION)”. R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>
- [27] Michael H. Kutner *et al*, “*Applied Linear Statistical Models*”. 4th ed. New York: McGraw-Hill Companies, Inc. 2004