

Klasifikasi Data NAP (Nota Analisis Pembiayaan) untuk Prediksi Tingkat Keamanan Pemberian Kredit (Studi Kasus : Bank Syariah Mandiri Cabang Luwuk Sulawesi Tengah)

Sumarni Adi*¹, Edi Winarko²

¹Prodi S2/S3, Jurusan Ilmu Komputer, FMIPA UGM, Yogyakarta

²Jurusan Ilmu Komputer dan Elektronika, FMIPA UGM, Yogyakarta
Sekip Utara Yogyakarta

e-mail: *sumarniadi@gmail.com, ewin@mail.ugm.ac.id

Abstrak

Setiap bulannya bank syariah mandiri cabang luwuk menerima proposal kredit (NAP) dari nasabah dalam jumlah yang terus meningkat dan perlu respon yang cepat. Dengan demikian, perlu dikembangkan sistem untuk melakukan data mining dari tumpukan data tersebut yang akan digunakan untuk kepentingan tertentu, salah satunya adalah untuk menganalisis resiko pemberian kredit. Teknik data mining digunakan dalam penelitian ini untuk klasifikasi tingkat keamanan pemberian kredit dengan menerapkan algoritma Naïve Bayes Classification. Naive bayes classifier merupakan pendekatan yang mengacu pada teorema Bayes yang mengkombinasikan pengetahuan sebelumnya dengan pengetahuan baru, sehingga merupakan salah satu algoritma klasifikasi yang sederhana namun memiliki akurasi tinggi. Sebelum dilakukan klasifikasi, data debitur melalui preprocessing. Kemudian dari preprocessing ini dilakukan klasifikasi dengan naive bayes classifier, sehingga menghasilkan model probabilitas klasifikasi untuk prediksi kelas pada debitur selanjutnya. Teknik pengujian akurasi model diukur menggunakan bootstrap, dan menunjukkan bahwa nilai akurasi terkecil 80% dihasilkan pada sampel data 100, dan menghasilkan nilai akurasi terbesar 98,66% pada sampel data 463.

Kata kunci— akurasi, naive bayes, data mining, klasifikasi, preprocessing, NAP

Abstract

Every month the Mandiri Syariah Bank Branch Office of Luwuk receives a very large number of proposal credit. Thus, the system should be developed to perform data mining of the heap data to be used for specific purpose, one of which is for the risk analysis of credit allowance. Data mining techniques used in this study for classification level prediction of credit allowance by applying a naïve Bayes Classification algorithm . Naive bayes classifier is an approach that refers to the bayes theorem, is a combination of prior knowledge with new knowledge. So that is one of the classification algorithm is simple but has a high accuracy. Prior to classification, data of debitur has been through a preprocessing. Then the weight is to perform classification with naive bayes classifier. After the data is classified, so produce probabilitas of model classification for prediction class to next debitur. Testing techniques the accuracy of the model was measured by bosstrap, and shows that the smallest value of accuracy is 80% produced in the 100 data sample, and the largest value of accuracy 98,66% on a data sample of 463.

Keywords— accuracy, naive bayes, data mining, classification, preprocessing, NAP

1. PENDAHULUAN

Kredit merupakan sumber utama penghasilan bagi sebuah bank dan juga sekaligus sumber resiko operasi bisnis terbesar, karena sebagian besar dana operasional bank diputar dalam bentuk kredit. Proses analisa data nasabah, data jaminan hingga proses pengambilan keputusan memiliki peranan penting dalam menekan atau meminimalisasi resiko bagi pihak kreditur, termasuk dalam hal ini adalah bank syariah mandiri cabang luwuk yang menjadi obyek penelitian. Syarat-syarat perbankan menjadi indikator pemberian pinjaman kepada nasabah agar layanan yang diberikan dapat tepat guna dari sisi nasabah dan tidak memberikan efek buruk bagi kestabilan manajemen bank, misalnya terjadi kredit macet.

Namun perkembangan dunia perbankan yang sangat pesat, mendorong terbentuknya suatu timbunan data-data yang berukuran sangat besar. Data-data tersebut pada umumnya berasal dari data entry dan customer service, kemudian oleh komputer data tersebut disimpan ke dalam server. Di dalam server data diubah menjadi informasi yang disimpan dalam bentuk tabel-tabel. Informasi yang didapat dari data dalam bentuk tabel-tabel tersebut sangat sedikit yang dapat dimanfaatkan oleh pihak manajemen perbankan dalam menganalisis untuk pemberian kredit nasabah, oleh karena itu perlu adanya aktivitas penggalian (ekstraksi) data yang masih tersembunyi untuk selanjutnya diolah menjadi pengetahuan yang bermanfaat dalam pengambilan keputusan. Proses ekstraksi informasi dari kumpulan data-data yang tersimpan di server disebut dengan data mining [1].

Ada banyak teknik yang bisa dilakukan untuk melakukan klasifikasi data di antaranya adalah *decision tree*, *bayesian classifiers*, *bayesian belief network* dan *rule based classifiers* [1]. Namun dalam penelitian ini akan dikembangkan sistem dengan menggunakan teknik klasifikasi *naive bayesian classifiers*. Pendekatan ini merupakan pendekatan mengacu pada teorema Bayes yang merupakan prinsip peluang statistika untuk mengkombinasikan pengetahuan sebelumnya dengan pengetahuan baru. Penggunaan teknik ini didasari oleh keperluan dari sistem untuk mengetahui nilai probabilitas dari data hasil klasifikasi yang dihasilkan. *Naive bayesian classifier* merupakan salah satu algoritma klasifikasi yang sederhana namun memiliki kemampuan dan akurasi tinggi [2]. Dan untuk prediksi kebangkrutan perusahaan, algoritma yang memiliki kinerja yang lebih unggul adalah algoritma naive bayes dengan tingkat akurasi mencapai 100% [3].

2. METODE PENELITIAN

2.1 Persiapan Data

Pada proses komunikasi dengan pihak Bank Syariah Mandiri Cabang Luwuk Sulawesi Tengah diperoleh data Nota Analisis Pembiayaan (NAP) berupa file excel, kemudian data tersebut diekstraksi dengan menggunakan prinsip 5 C + 1 S ke bentuk file sql. Data NAP tersebut adalah data pembiayaan (kredit) jenis murabahah tahun 2005-2006. Contoh data diperlihatkan pada Tabel 1.

Tabel 1 Contoh data NAP Bank Syariah mandiri Cabang Luwuk Sulawesi Tengah

No_Rekening	Nama	Pendidikan	Pekerjaan	Gaji	Nilai_jaminan	Tempo	Jumlah_pinjam	Tujuan_pinjam	Angsuran	Margin_pembi	Kelas
1050070629	ABDUL K	Sarjana	PNS Umum	2386700.00	50000000.00	96	96500000.00	BIAYA MENIKAH	1874146.77	83418089.87	Aman
7044291538	ABD.MAL	SMA	PNS Umum	3847000.00		80	140000000.00	MODAL USAHA	2956253.89	96500311.50	Aman
1050060436	Mintje	Sarjana	PNS Umum	2767500.00	50000000.00	72	60000000.00	BARANG KONSUMTIF	1334767.87	36103286.56	Aman
1050074990	Abdul Ga	Sarjana	PNS Umum	2547700.00		60	50000000.00	BARANG KONSUMTIF	1229226.05	23753563.29	Aman
1050049194	Abdullah	Sarjana	PNS Umum	3071400.00		96	30000000.00	BARANG KONSUMTIF	582636.30	25933084.93	Aman
7043990416	ABDUL H	Sarjana	PNS Umum	2214500.00	100000000.00	60	37500000.00	RENOVASI RUMAH	921919.54	17815172.47	Aman

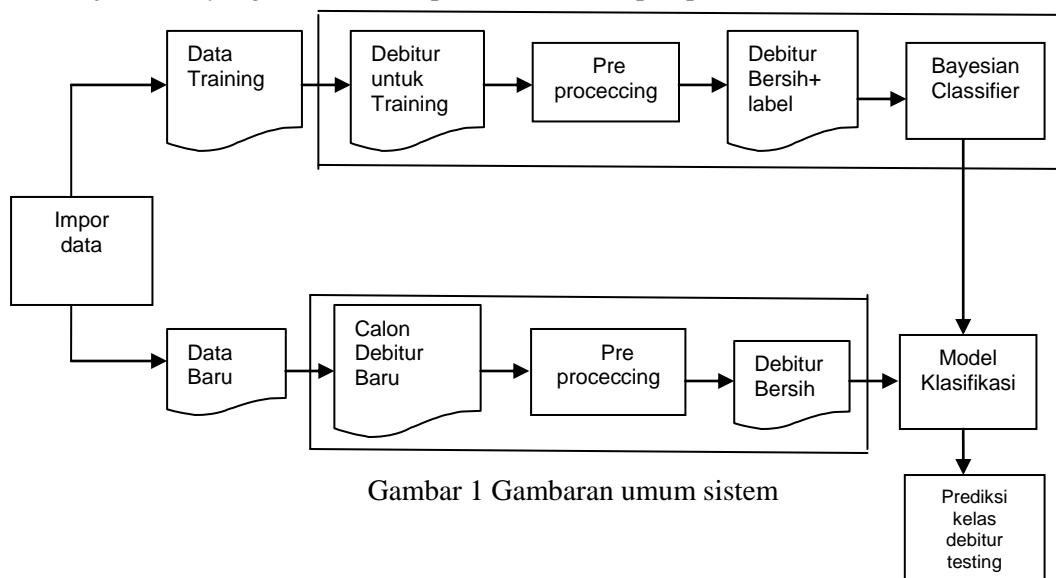
2.2 Analisis Sistem

Sistem baru yang dirancang merupakan sistem yang dapat melakukan proses *preprocessing* secara otomatis dan tersimpan dalam database berdasarkan data yang akan diklasifikasi. Proses *preprocessing* meliputi data *Transformation* yaitu melakukan transformasi data agar layak di mining, data *Reduction* yaitu untuk mengurangi representasi *dataset* sehingga berukuran lebih kecil. Data calon debitur yang sudah melewati proses *preprocessing* kemudian diklasifikasi dengan *naive bayesian classification* untuk menentukan model yang akan digunakan untuk menentukan prediksi kelas pada data *debitur* yang baru.

Data yang telah diambil diberi kelas label, kemudian ditentukan untuk dijadikan *training data* dan *testing data* dengan masing-masing porsi sejumlah 2/3 dan 1/3 dari keseluruhan data berdasarkan metode pengujian *bootstrap*. Selanjutnya untuk *training data* tersebut dilakukan proses dengan metode klasifikasi menggunakan algoritma *naive bayes* untuk menentukan model. Dari model yang dihasilkan tersebut selanjutnya akan digunakan untuk melakukan klasifikasi kelas label terhadap *testing data* sehingga diperoleh nilai akurasi. Model yang dihasilkan tersebut lalu digunakan untuk mengklasifikasi data debitur baru (data *debitur* tanpa label kelas) ke dalam kelas “aman” atau “tidak aman”. Data debitur yang sudah diklasifikasi, selanjutnya oleh sistem divisualisasikan dengan menggunakan pemrograman java. Inti dari permasalahan yang diangkat adalah bagaimana melakukan ekstraksi data *debitur* untuk digunakan sebagai data *training* dan *testing* model klasifikasi *naive bayes classifier* yang digunakan untuk prediksi kelas untuk data *debitur* yang baru. Kemudian hasil klasifikasi tersebut digunakan sebagai model untuk menentukan “aman” atau tidaknya status pinjaman calon debitur berikutnya.

2.3 Deskripsi Sistem

Secara garis besar sistem ini bertujuan untuk menentukan kelas pada *debitur* menggunakan teknik klasifikasi data mining yaitu *naive bayesian classification* untuk mencari pola prediksi debitur-debitur yang cenderung “tidak aman”, yang kemudian dijadikan sebagai model untuk penentuan status pembiayaan debitur berikutnya. Selengkapnya gambaran umum alur kerja sistem yang dibuat dalam penelitian ini tampak pada Gambar 1.



Gambar 1 Gambaran umum sistem

Pada Gambar 1 diperlihatkan langkah pertama adalah mengimpor data debitur yang sudah diekstraksi bertipe .csv secara otomatis lalu data tersebut disimpan dalam database. Jadi langkah perancangan pertama kali adalah dengan merancang program impor data debitur yang sudah diekstraksi ke database Mysql. Selanjutnya membangun model data mining dalam hal ini klasifikasi berdasarkan data ekstraksi yang telah kita masukkan ke database tersebut. Pada proses impor data ini juga dilakukan pe-label-an data, kemudian data tersebut diproses untuk

tahap persiapan (*preprocessing*) agar data siap digunakan untuk proses klasifikasi. Hal ini dilakukan karena tidak semua data *debitur* tersebut dapat digunakan pada proses klasifikasi.

Kemudian hasil *preprocessing* data ini disimpan di tempat yang berbeda dari “data kotor”-nya. Data *debitur* bersih yang sudah diberi label/kelas atau data *training* diolah oleh *naive bayes classifier* untuk menghasilkan model probabilitas. Model probabilitas tersebut terbagi menjadi dua bagian, yaitu : probabilitas kelas dan probabilitas semua variabel pada suatu kelas yang dalam sistem ini disimpan di database. Setelah itu, model probabilitas ini digunakan untuk memprediksi data *debitur* baru yang sudah dibersihkan pada tahap *preprocessing*. Kemudian hasil prediksi data *debitur* yang baru dijadikan model sistem untuk menentukan status *debitur* berikutnya.

2.4 Perancangan Sistem

Data awal yang telah dikumpulkan, selanjutnya dilakukan proses awal yaitu impor data lalu selanjutnya data mengalami *preprocessing* untuk mengurangi berbagai masalah dari data sehingga data tersebut siap digunakan untuk proses data mining. Semua kegiatan ini dilakukan secara otomatis.

2.4.1 Impor Data Debitur



Data yang diimpor adalah data yang sudah melalui proses ekstraksi. Data ekstraksi yang dihasilkan adalah *file* bertipe *.csv* tersebut tidak sepenuhnya langsung disimpan, melainkan dicek terlebih dahulu apakah *debitur* yang akan dimasukkan ke database tersebut sudah ada sebelumnya dengan mengecek kesamaan antara Nomor rekening dan nama *debitur* dengan yang sudah tersimpan di database. Apabila sudah ada, maka data tidak akan disimpan tetapi apabila data belum ada maka selanjutnya data disimpan di database. Misalkan 6 data *debitur* yang sudah diekstraksi akan diimpor yang ditunjukkan pada Tabel 2, yang hasilnya akan disimpan di database dengan nama *debitur* yang ditunjukkan pada Tabel 3.

Tabel 2 Data Debitur Hasil Ekstraksi

No_Rekening	Nama	Pendidikan	Pekerjaan	Gaji	Nilai_jaminan	Tempo	Jumlah_pinjam	Tujuan_pinjam	Angsuran	MARGIN_pembi	Kelas
1050070629	ABDUL K	Sarjana	PNS Umum	2386700.00	50000000.00	96	96500000.00	BIAYA MENIKAH	1874146.77	83418089.87	Aman
7044291538	ABD.MAL	SMA	PNS Umum	3847000.00		80	140000000.00	MODAL USAHA	2956253.89	96500311.50	Aman
1050060436	Mintje	Sarjana	PNS Umum	2767500.00	50000000.00	72	60000000.00	BARANG KONSUMTIF	1334767.87	36103286.56	Aman
1050074990	Abdul Ga	Sarjana	PNS Umum	2547700.00		60	50000000.00	BARANG KONSUMTIF	1229226.05	23753563.29	Aman
1050049194	Abdullah	Sarjana	PNS Umum	3071400.00		96	30000000.00	BARANG KONSUMTIF	582636.30	25933084.93	Aman
7043990416	ABDUL H	Sarjana	PNS Umum	2214500.00	100000000.00	60	37500000.00	RENOVASI RUMAH	921919.54	17815172.47	Aman

Tabel 3 Data Debitur Hasil Impor

id_debitur	nama	pendidikan	pekerjaan	gaji	nilai_jaminan	tempo_pinjam	jumlah_pinjam	tujuan_pinjam	angsuran	margin	id_kelas
1050049194	Abdullah Nursam	Sarjana	PNS Umum	3071400.00	0.00	96	30000000.00	BARANG KONSUMTIF	582636.30	25933084.93	1
1050074990	Abdul Gafur Laganja	Sarjana	PNS Umum	2547700.00	0.00	60	50000000.00	BARANG KONSUMTIF	1229226.05	23753563.29	1
1050060436	Mintje	Sarjana	PNS Umum	2767500.00	50000000.00	72	60000000.00	BARANG KONSUMTIF	1334767.87	36103286.56	1
7044291538	ABD.MALIK RAHMAN	SMA	PNS Umum	3847000.00	0.00	80	140000000.00	MODAL USAHA	2956253.89	96500311.50	1
1050070629	ABDUL KAMIL LASADAM	Sarjana	Pegawai Swasta	2386700.00	50000000.00	96	96500000.00	BIAYA MENIKAH	1874146.77	83418089.87	1
7043990416	ABDUL HALIK LASADAM	Sarjana	PNS Umum	2214500.00	100000000.00	60	37500000.00	RENOVASI RUMAH	921919.54	17815172.47	1

Check All / Uncheck All With selected:   

2.4.2 Preprocessing

Secara umum *preprocessing* ini melewati beberapa tahap antara lain :

1. Data *transformation* dengan metode *normalization* yaitu mengubah data ke range spesifik. Metode ini dilakukan pada variabel gaji, variabel nilai Jaminan, dan variabel angsuran. Contoh penerapan *normalization* pada ketiga variabel tersebut adalah memberikan range tinggi, sedang atau rendah pada nilai variabel-variabel tersebut.
2. Data *reduction* dengan metode *Discretization and concept hierarchy generation*, yaitu mengurangi jumlah nilai yang diberikan atribut kontinyu dengan membagi *range* pada atribut menjadi suatu interval tertentu. Beberapa atribut yang diterapkan teknik ini adalah variabel gaji, variabel nilai jaminan, dan variabel angsuran. Contoh penerapan *Discretization and concept hierarchy generation* yaitu ketiga variabel tersebut diberikan nilai batas atas dan batas bawah untuk menentukan range tinggi, sedang dan rendah.
3. Data *reduction* dengan metode *compression*, yaitu menghilangkan beberapa variabel data dengan cara mengubah data dengan melakukan fungsi agregasi agar menjadi representasi data yang lebih kecil. Contohnya adalah variabel angsuran didapatkan dengan cara menjumlahkan variabel jumlah pinjam dengan variabel margin pembiayaan dibagi dengan tempo pinjam. Sehingga setelah dilakukan metode ini, variabel yang tersisa adalah variabel angsuran yang akan digunakan pada proses selanjutnya yaitu klasifikasi.
4. Data *Reduction* dengan Metode *Dimension reduction*, dimana atribut atau dimensi yang tidak relevan dihapus. Contoh penerapan metode *Dimension reduction* adalah variabel tujuan pinjam dihapus karena semua nilai tujuan pinjam sesuai syariah, sehingga tidak digunakan dalam proses klasifikasi karena hanya akan menambah nilai komputasi namun tidak menambah nilai akurasi sistem yang dibangun.

Misalkan, 6 data debitur akan dilakukan *preprocessing*, yang hasilnya akan disimpan di database dengan nama “debitur bersih” yang ditunjukkan pada Tabel 4.

Tabel 4 Data Hasil *Preprocessing*

Id	Pendidikan	Pekerjaan	Gaji	Angsuran	Nilai Jaminan	Kelas
1050070629	Sarjana	PNS Umum	Kecil	Kecil	Sedang	Aman
7044291538	Sarjana	PNS Umum	Kecil	Kecil	Sedang	Aman
1050060436	SMA	PNS Umum	Sedang	Sedang	Kecil	Tidak Aman
1050074990	Sarjana	PNS Umum	Kecil	Kecil	Sedang	Aman
1050049194	Sarjana	PNS Umum	Kecil	Kecil	Kecil	Aman
7043990416	Sarjana	PNS Umum	Sedang	Kecil	Kecil	Aman

2.5 Naive Bayes Classifier

Salah satu metode klasifikasi yang dapat digunakan adalah metode *naive bayes* yang sering disebut sebagai *naive bayes classifier* (NBC). *Naive bayes classifier* (NBC) merupakan salah satu metode pada teknik klasifikasi dan termasuk dalam *classifier* statistik yang dapat memprediksi probabilitas keanggotaan *class*. NBC berprinsip pada teori bayes. NBC mengasumsikan bahwa nilai atribut pada sebuah *class* adalah independen terhadap nilai pada atribut yang lain. Ada dua tahap pada proses klasifikasi data. Tahap pertama adalah pelatihan terhadap himpunan contoh (*training example*). Sedangkan tahap kedua adalah proses klasifikasi data yang belum diketahui kategorinya.

Naive bayes atau simple bayesian *classifier* memiliki prosedur sebagai berikut [1] :

1. Setiap sample data direpresentasikan dengan n-dimensional *feature vector*, $X=(X_1, X_2, \dots, X_n)$, dengan n dibuat dari *sample* n atribut, berturut-turut A_1, A_2, \dots, A_n .
2. Diandaikan terdapat m *class*, C_1, C_2, \dots, C_m . Diberikan sebuah data *sample*, X (yang tidak diketahui *class* labelnya), kemudian *classifier* akan memprediksi X ke dalam *class* yang

memiliki probabilitas posterior tertinggi berdasarkan kondisi X . Naive bayes *classifier* akan menentukan bahwa *sample* X yang tidak diketahui tadi ke dalam *class* C_i jika dan hanya jika :

$$P(C_i | X) > P(C_j | X) \text{ untuk } 1 \leq j \leq m, j \neq i$$

Sehingga perlu dimaksimalkan $P(C_i | X)$. *Class* C_i pada $P(C_i | X)$ adalah nilai terbesar, yang disebut dengan maksimum posterio hypothesis dengan teorema bayes yang ditunjukkan pada persamaan (1),

$$P(C_i | X) = \frac{P(X | C_i) P(C_i)}{P(X)} \quad (1)$$

3. $P(X)$ adalah konstan untuk semua *class*, hanya $P(X/C_i)P(C_i)$ yang perlu dimaksimalkan. Jika probabilitas prior *class* tidak diketahui, secara umum diasumsikan bahwa *class-class* adalah sama, yaitu $P(C_1)=P(C_2)=\dots=P(C_m)$, dan selanjutnya menghitung nilai $P(X/C_i)$ dan menghitung nilai $P(X/C_i)P(C_i)$. Probabilitas *class* prior dapat diestimasi dengan perhitungan seperti yang ditunjukkan pada persamaan (2) :

$$P(C_i) = \frac{s_i}{s} \quad (2)$$

dimana s_i adalah jumlah *training sample* dari *class* C_i dalam data *training*, dan s adalah jumlah total data *training sample*.

4. Apabila *dataset* terdiri dari banyak atribut, akan mengakibatkan komputasi yang rumit untuk menghitung $P(X/C_i)$. Untuk mengurangi komputasi dalam evaluasi $P(X/C_i)$, naive bayes mengasumsikan pada pembuatan *class independen* seperti yang ditunjukkan pada persamaan (3). Sehingga nilai pada atribut dikondisikan bersifat independen antara atribut yang satu dengan atribut yang lain, serta diantara atribut tidak terdapat relasi depedensi. Sehingga,

$$P(X | C_i) = \prod_{k=1}^n P(X_k | C_i) \quad (3)$$

$$= P(x_1/C_i) \times P(x_2/C_i) \times \dots \times P(x_n/C_i)$$

Jadi dapat dengan mudah diestimasi probabilitas $P(x_1/C_i), P(x_2/C_i), \dots, P(x_n/C_i)$ dari data *training*. Dimana x_k merupakan nilai atribut A_k untuk data X . Untuk setiap atribut, dilihat apakah atribut tersebut bernilai kategorikal atau kontinyu, sehingga untuk menghitung $P(X/C_i)$ mengikuti aturan berikut dimana untuk atribut bernilai kategorikal menggunakan persamaan (4), sedangkan atribut bernilai kontinyu menggunakan persamaan (6) :

- a. Jika A_k adalah kategorikal, maka $P(x_k / C_i) = \frac{s_{ik}}{s_i}$ (4)

s_{ik} adalah jumlah dari *training sample* pada *class* C_i yang mempunyai nilai x_k untuk A_k dan s_i adalah jumlah *training sample* yang termasuk ke dalam *class* C_i .

- b. Jika A_k bernilai kontinyu, maka perlu dilakukan sedikit lebih banyak pekerjaan, tapi perhitungannya cukup sederhana. Sebuah atribut bernilai kontinyu biasanya diasumsikan memiliki distribusi Gaussian yang dihitung menggunakan persamaan (5) dengan *mean* (rata-rata) μ yang dihitung menggunakan persamaan (7) dan standar deviasi σ menggunakan persamaan (8) :

$$g(x_k, \mu C_i, \sigma C_i) = \frac{1}{\sqrt{2\pi} \sigma C_i} e^{-\frac{(x_k - \mu C_i)^2}{2 \sigma^2 C_i}} \quad (5)$$

sehingga

$$P(X_k | C_i) = g(x_k, \mu C_i, \sigma C_i) \quad (6)$$

dimana,

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad (7)$$

$$\sigma = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu)^2 \quad (8)$$

adalah fungsi gaussian untuk atribut A_k dengan μC_i dan σC_i masing-masing adalah *mean*(rata-rata) dan standar deviasi dari nilai atribut A_k pada *training sample class* C_i .

5. Untuk memperkirakan *class label* dari *sample* X , $P(X | C_i) P(C_i)$ dievaluasi untuk setiap *class* C_i menggunakan aturan pada persamaan (9). *Classifier* memperkirakan *class label* dari *sample* X adalah *class* C_i jika dan hanya jika

$$P(X | C_i)P(C_i) > P(X | C_j)P(C_j) \text{ untuk } 1 \leq j \leq m, j \neq i \quad (9)$$

Dengan kata lain, diperkirakan *class label* adalah *class* C_i untuk $P(X | C_i)P(C_i)$ yang bernilai maksimum.

2.6 Metode Validasi Bootstrap

Tujuan utama penggunaan *bootstrap* adalah untuk memperoleh estimasi parameter berdasarkan data yang minimal dengan bantuan komputer. Setiap sampel *bootstrap* berdistribusi sama satu dengan lainnya, atau dapat diasumsikan bahwa sampel *bootstrap* berasal dari distribusi populasi yang sama, tetapi setiap sampel *bootstrap* saling independen. Dengan kata lain, untuk menghitung akurasi populasi tersebut dilakukan dengan teknik sampel dengan membagi *training set* dan *testing set*.

Prinsip dasar pembentukan sampel *bootstrap* sebagai berikut :

- Konstruksi fungsi distribusi empiris dari sampel yaitu \hat{F}_n dengan peluang yang sama terambil yaitu $\frac{1}{n}$ untuk masing-masing X_1, X_2, \dots, X_n .
- Dengan \hat{F}_n tetap, ambil sampel acak berukuran n dari \hat{F}_n sebut $X_i^* = x_i^*, X_i^* \sim \hat{F}_n$, $i = 1, 2, \dots, n$.

Estimasi akurasi *bootstrap* adalah jumlah keseluruhan klasifikasi yang benar, dibagi dengan jumlah kasus dalam dataset [4]. Jadi untuk menghitung nilai akurasinya digunakan persamaan (10)

$$\text{Akurasi} = \frac{\text{Jumlah Klasifikasi Benar}}{\text{Jumlah Data Uji}} \times 100\% \quad (10)$$

3. HASIL DAN PEMBAHASAN

3.1 Pengujian Akurasi Klasifikasi

Pengujian yang akan dilakukan pada sistem ini adalah dengan menggunakan metode *bootstrap*. Nilai populasi dalam penelitian ini adalah 463 data. Dengan ukuran sampel 1/3 digunakan untuk training dan 2/3 digunakan untuk testing. Dengan demikian, akan dilakukan 3 kali pengujian akurasi klasifikasi dari data bersih yang dijadikan sampel. Berikut merupakan mekanisme pengujian sistem yang dilakukan pada sistem ini :

- Membagi sampel menjadi 1/3 bagian yang sama rata.
- Sebanyak 1/3 dari jumlah sampel tersebut secara bergantian dijadikan sebagai data testing dan 2/3 lainnya dijadikan sebagai data training.
- Dari 1/3 yang dijadikan data testing tersebut kemudian dibandingkan hasil klasifikasi oleh sistem dengan kelas yang sudah ditentukan sebelumnya.
- Dihitung nilai akurasinya menggunakan persamaan (10) untuk masing-masing bagian sampel.

Sistem yang sudah dibangun diuji dengan menggunakan data penelitian yang sudah terkumpul sebelumnya. Pada proses pengujian ini, dilakukan pengklasifikasian data dengan berbagai porsi jumlah sampel yang dijadikan sebagai data input dari sistem ini. Adapun jumlah porsi sampel adalah 100, 250, 350, dan 463 sampel data yang digunakan.

3.1.1 Pengujian kelas Prediksi dengan Standar deviasi

Pada pengujian ini dilakukan perhitungan jumlah kelas prediksi berdasarkan pembagian level pada variabel gaji, jaminan dan angsuran menjadi 3 level. Jumlah sampel data yang digunakan adalah 463 data debitur. Tabel 5 tertuang hasil pengujian kelas prediksi dengan standar deviasi berdasarkan besar kecilnya nilai variabel gaji, jaminan dan angsuran yang ditetapkan berdasarkan levelnya.

Tabel 5 Pengujian Kelas Prediksi Teknik Standar Deviasi

No	Level Gaji			Level Angsuran			Level Jaminan			Jumlah Sampel		
	Kecil	Sedang	Besar	Kecil	Sedang	Besar	Kecil	Sedang	Besar	Aman	Tidak Aman	Unknown
1	1.000.000- 2.000.000	2.000.000- 3.000.000	3.000.000- 4.000.000	500.000- 1.500.000	1.500.000- 2.500.000	2.500.000- 3.500.000	0- 25.000.000	25.000.000- 50.000.000	50.000.000- 75.000.000	223	135	105
2	1.000.000- 3.000.000	3.000.000- 5.000.000	5.000.000- 7.000.000	500.000- 2.500.000	2.500.000- 4.500.000	4.500.000- 6.500.000	0- 50.000.000	50.000.000- 100.000.000	100.000.000- 150.000.000	302	150	11
3	1.000.000- 4.000.000	4.000.000- 7.000.000	7.000.000- 10.000.000	500.000- 3.500.000	3.500.000- 6.500.000	6.500.000- 9.500.000	0- 75.000.000	75.000.000- 150.000.000	150.000.000- 225.000.000	312	150	1
4	1.000.000- 5.000.000	5.000.000- 9.000.000	9.000.000- 13.000.000	500.000- 4.500.000	4.500.000- 8.500.000	8.500.000- 12.500.000	0- 100.000.000	100.000.000- 200.000.000	200.000.000- 300.000.000	312	150	1

3.1.2 Pengujian kelas Prediksi dengan Logika Fuzzy

Pada pengujian ini dilakukan perhitungan jumlah kelas prediksi berdasarkan logika fuzzy, pada variabel gaji, variabel angsuran dan variabel jaminan. Ketiga variabel ini dibagi ke dalam tiga himpunan yaitu kecil, sedang dan besar. Pengguna diminta untuk mengatur derajat keanggotaan yaitu bernilai 0, 0-1 dan 1. Setelah itu menentukan batasan dari masing-masing himpunan. Setelah proses pengaturan nilai derajat keanggotaan dan batasan dari setiap himpunan pada masing-masing variabel selesai. Hasil dari pengujian kelas prediksi dengan teknik logika fuzzy ditunjukkan pada tabel 6. dengan membandingkan jumlah sampel yang dapat diolah pada Tabel 6 dan 5 menunjukkan bahwa dengan teknik fuzzy maupun teknik standar deviasi hasilnya sama.

Tabel 6 Pengujian Kelas Prediksi Teknik Logika Fuzzy

Gaji									Jumlah Sampel		
Kecil			Sedang			Besar			Aman	Tidak Aman	Unknow
0	0..1	1	0	0..1	1	0	0..1	1			
>= 7.000.000	4.000.000 <= gaji <= 7.000.000	0 <= gaji <= 4.000.000	Gaji <= 4.000.000 atau gaji >= 13.000.000	10.000.000 <= gaji <= 13.000.000	7.000.000 <= gaji <= 10.000.000	Gaji <= 10.000.000	10.000.000 <= gaji <= 11.000.000	Gaji >= 13.000.000	312	150	1
Jaminan									Jumlah Sampel		
Kecil			Sedang			Besar			Aman	Tidak Aman	Unknow
0	0..1	1	0	0..1	1	0	0..1	1			
>= 100.000.000	90.000.000 >= jaminan <= 100.000.000	0 >= jaminan <= 90.000.000	Jaminan >= 160.000,00	150.000,00 >= Jaminan <= 160.000,00	100.000.000 >= Jaminan <= 150.000,00	Jaminan <= 150.000.000	150.000,00 >= Jaminan <= 160.000,00	Jaminan >= 160,000,00	312	150	1
Angsuran									Jumlah Sampel		
Kecil			Sedang			Besar			Aman	Tidak Aman	Unknow
0	0..1	1	0	0..1	1	0	0..1	1			
Angsuran >= 3.500.000	2.000,00 <= Angsuran <= 3.500,00	0 <= Angsuran <= 2.000,00	Angsuran >= 4.000,000	3.500,00 <= Angsuran >= 4.000,000	2.000,00 <= Angsuran >= 3.500,000	Angsuran <= 3.500.000	3.500.000 <= angsuran <= 5.000.000	angsuran >= 8.500.000	312	150	1

3.2 Pengujian Akurasi Model probabilitas kelas teknik Standar Deviasi

Pada pengujian ini dilakukan pengujian akurasi model probabilitas kelas terhadap data bersih yang terbentuk dengan berbagai range dengan tingkatan nilai tertentu. Seperti pada preprocessing, range nilai level dibagi dengan interval tertentu sampai mendapat nilai terbaik. Tabel 7 merupakan rincian hasil pengujian model probabilitas kelas dengan teknik standar deviasi dengan berbagai range nilai level.

Tabel 7 Pengujian Akurasi Model Probabilitas Kelas teknik standard Deviasi

No	Level Gaji			Level Anesuran			Level Jaminan			Probabilitas Kelas	
	Kecil	Sedanz	Besar	Kecil	Sedanz	Besar	Kecil	Sedanz	Besar	Aman	Tidak Aman
1	1.000.000- 2.000.000	2.000.000- 3.000.000	3.000.000- 4.000.000	500.000- 1.500.000	1.500.001- 2.500.000	2.500.001- 3.500.000	0- 25.000.000	25.000.001- 50.000.000	50.000.001- 75.000.000	0,62	0,38
2	1.000.000- 3.000.000	3.000.000- 5.000.000	5.000.000- 7.000.000	500.000- 2.500.000	2.500.001- 4.500.000	4.500.001- 6.500.000	0- 50.000.000	50.000.001- 100.000.000	100.000.001- 150.000.000	0,67	0,33
3	1.000.000- 4.000.000	4.000.000- 7.000.000	7.000.000- 10.000.000	500.000- 3.500.000	3.500.001- 6.500.000	6.500.001- 9.500.000	0- 75.000.000	75.000.001- 150.000.000	150.000.001- 225.000.000	0,68	0,32
4	1.000.000- 5.000.000	5.000.000- 9.000.000	9.000.000- 13.000.000	500.000- 4.500.000	4.500.001- 8.500.000	8.500.001- 12.500.000	0- 100.000.000	100.000.001- 200.000.000	200.000.001- 300.000.000	0,68	0,32

3.3 Pengujian Akurasi Model probabilitas kelas teknik Logika Fuzzy

Pada pengujian ini dilakukan pengujian akurasi model probabilitas kelas terhadap data bersih yang terbentuk dengan teknik logika fuzzy. Pada variabel gaji, variabel angsuran dan variabel jaminan dibagi ke dalam tiga himpunan yaitu kecil, sedang dan besar. Setelah itu mengatur derajat keanggotaan yaitu bernilai 0, 0-1 dan 1. terakhir menentukan batasan dari masing-masing himpunan. Tabel 8 merupakan rincian hasil pengujian model probabilitas kelas dengan teknik logika Fuzzy.

Tabel 8 Pengujian Akurasi Model Probabilitas Kelas teknik standard Deviasi

Gaji									Probabilitas	
Kecil			Sedang			Besar			Aman	Tidak Aman
0	0...1	1	0	0...1	1	0	0...1	1		
>= 7.000.000	4.000.000 <= gaji <= 7.000.000	0 <= gaji <= 4.000.000	Gaji <= 4.000.000 atau gaji >= 13.000.000	10.000.000 <= gaji <= 13.000.000	7.000.000 <= gaji <= 10.000.000	Gaji <= 10.000.000	10.000.000 <= gaji <= 11.000.000	Gaji >= 13.000.000	0,68	0,32
Jaminan									Probabilitas	
Kecil			Sedang			Besar			Aman	Tidak Aman
0	0...1	1	0	0...1	1	0	0...1	1		
>= 100.000.000	90.000.000 >= jaminan <= 100.000.000	0 >= jaminan <= 90.000.000	Jaminan >= 160.000,00	150.000,00 >= Jaminan <= 160.000,00	100.000,000 >= Jaminan <= 150.000,00	Jaminan <= 150.000,000	150.000,000 >= Jaminan <= 160.000,000	Jaminan >= 160.000,000	0,68	0,32
Angsuran									Probabilitas	
Kecil			Sedang			Besar			Aman	Tidak Aman
0	0...1	1	0	0...1	1	0	0...1	1		
Angsuran >= 3.500,000	2.000,00 <= Angsuran <= 3.500,00	0 <= Angsuran <= 2.000,00	Angsuran >= 4.000,000	3.500,000 <= Angsuran <= 4.000,000	2.000,000 <= Angsuran <= 3.500,000	Angsuran <= 3.500.000	3.500.000 <= angsuran <= 5.000.000	angsuran >= 8.500.000	0,68	0,32

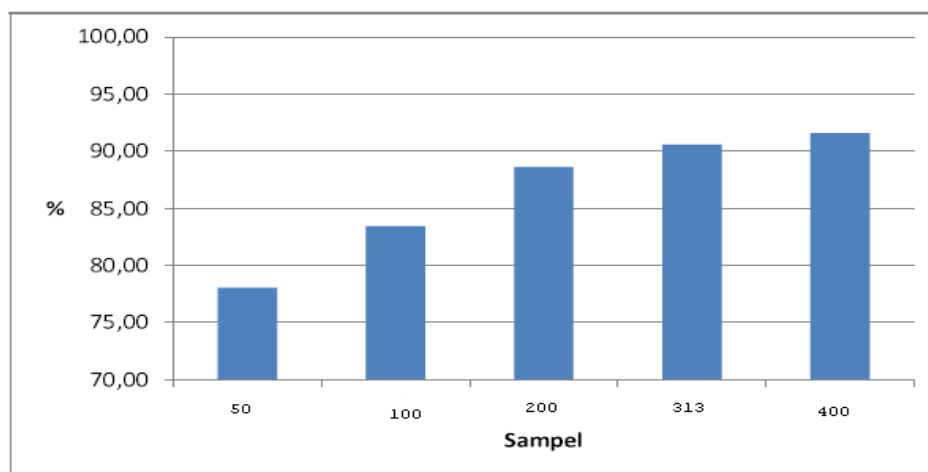
3.4 Pengujian Akurasi Model

Pengujian ini dilakukan pengujian akurasi model probabilitas terhadap data debitur dengan berbagai porsi jumlah data sampel. Adapun jumlah porsi sampel seperti pada pengujian waktu pemrosesan yaitu 100, 250, 350 dan 463 sampel data yang digunakan. Pengujian akurasi model menggunakan *Bootstrap*.

Tabel 9 Rata-rata nilai akurasi model

Sampel	Akurasi model (%)		
	Aman	Tidak Aman	Model
100	80.00	81.00	80.00
250	89.83	87.03	84.10
350	92.53	67.45	89.90
463	98.75	97.08	98.66

Tabel 9 merupakan tabel nilai akurasi model untuk masing - masing pengujian dengan berbagai porsi data. Dari tabel tersebut dapat diketahui bahwa nilai akurasi tertinggi terdapat pada pengujian dengan menggunakan sampel data sebanyak 463.



Gambar 2 Grafik rata-rata akurasi model

Berdasarkan Gambar 2 dapat dilihat bahwa semakin besar sampel data yang digunakan untuk klasifikasi, maka menghasilkan nilai akurasi model semakin besar pula.

3.5 Perbandingan validasi model dengan Rapid Miner 5.3

Pengujian ini dilakukan untuk melakukan perbandingan validasi model algoritma naive bayes yang dihasilkan dengan aplikasi yang dibuat dengan aplikasi lain menggunakan algoritma *naive bayes* dan *support vector machine*.

Pada aplikasi yang dibuat yang menggunakan algoritma *naive bayes* dilakukan pengujian validasi model yang dihasilkan dalam beberapa variasi jumlah data debitur yang digunakan.

Perangkat lunak lain yang digunakan untuk mengklasifikasi data debitur adalah aplikasi Rapid Miner 5.3. Digunakan algoritma *naive bayes* dan *support vector machine* untuk pengujian validasi model yang dihasilkan juga dalam beberapa variasi jumlah data debitur yang digunakan.

Pemrosesan ketiga pengujian tersebut menghasilkan rata-rata nilai akurasi model untuk masing-masing metode menggunakan masing-masing aplikasi. Pengujian tersebut menggunakan sumber data yang sama dengan beberapa variasi jumlah data yang digunakan yaitu 100, 250, 350 dan 463.

Berdasarkan hasil pengujian validasi model pada aplikasi yang dibuat menggunakan algoritma *naive bayes* serta pengujian validasi model pada aplikasi Rapid Miner 5.3 menggunakan algoritma *naive bayes* dan SVM, dapat dibuat rangkuman yang hasilnya ditunjukkan pada Tabel 10.

Tabel 10 Perbandingan rata-rata persentase akurasi model

Jumlah data	Model Naive Bayes (%)	Rapid Miner 5.3 (Naive Bayes) (%)	Rapid Miner 5.3 (SVM) (%)
100	80.00	92.00	72.00
250	84.10	98.70	84.10
350	89.90	98.76	89.90
463	98.66	99.11	92.66

4. KESIMPULAN

1. Dari hasil pengujian dengan menggunakan teknik standar deviasi dan logika fuzzy pada proses preprocessing, dapat disimpulkan bahwa hasil data yang dapat diolah adalah sama.
2. Dari hasil pengujian *training* yang dilakukan, dapat disimpulkan bahwa semakin banyak data yang dapat diolah maka semakin besar nilai probabilitas kelas.
3. Dari hasil pengujian akurasi model dari sistem yang dikembangkan, menghasilkan nilai akurasi terkecil sebesar 80% pada proses pengujian dengan menggunakan sampel sebanyak 100 dan menghasilkan nilai akurasi tertinggi sebesar 98,66% pada proses pengujian dengan menggunakan sampel sebanyak 463. Hal ini menunjukkan bahwa akurasi model semakin meningkat dengan bertambahnya data.
4. Hasil pengujian akurasi dengan menggunakan perangkat lunak Rapid Miner 5.3 dengan metode *naive bayes classifier* menghasilkan nilai akurasi terkecil sebesar 92% pada proses pengujian dengan menggunakan sampel sebanyak 100 dan menghasilkan nilai akurasi tertinggi sebesar 99,11% pada proses pengujian dengan menggunakan sampel sebanyak 463.
5. Hasil pengujian akurasi dengan menggunakan perangkat lunak Rapid Miner 5.3 dengan metode *support vector machine* menghasilkan nilai akurasi terkecil sebesar 72,00% pada proses pengujian dengan menggunakan sampel sebanyak 100 dan menghasilkan nilai akurasi tertinggi sebesar 92,66% pada proses pengujian dengan menggunakan sampel sebanyak 463.
6. Nilai akurasi sistem yang dikembangkan relatif hampir sama dengan nilai akurasi yang dihasilkan oleh perangkat lunak Rapid Miner 5.3 dengan metode *naive bayes classifier*.
7. Dengan menggunakan perangkat lunak Rapid Miner 5.3 maka dapat disimpulkan bahwa nilai akurasi metode *naive bayes classifier* lebih baik dari pada nilai akurasi metode *support vector machine*.

5. SARAN

1. Perlu dilakukan analisis hubungan antar variabel pada proses klasifikasi.
2. Perlu dilakukan proses otomatisasi pada proses ekstraksi data debitur, sehingga nota analisis pembiayaan dapat langsung diproses oleh sistem tanpa harus melakukan impor data terlebih dahulu.

DAFTAR PUSTAKA

- [1] Han, J., and Kamber, M., 2006, *Data Mining: Concepts and Techniques*, Second Edition, Morgan Kaufmann, San Francisco.
- [2] Rish, I., 2006, An empirical study of the Naive Bayes Classifier, *International Joint Conference on Artificial Intelligence*, California, 41-46.

- [3] Oktafia, D., dan Pardede, D.L.C., 2010, Perbandingan Kinerja Algoritma Decision Tree dan Naïve Bayes dalam Prediksi Kebangkrutan, *Proceeding Seminar Ilmiah Nasional KOMMIT 2010*, Universitas Gunadarma.
- [4] Kohavi, R., 1995, *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection*, *Proceedings of IJCAI Conferences 1995* : Stanford University, 1137-1143.