

KLASIFIKASI DATA NAP (NOTA ANALISIS PEMBIAYAAN) DENGAN 5C+1S UNTUK PENENTUAN TINGKAT KEAMANAN PEMBIAYAAN MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER PADA BANK SYARIAH

Sumarni Adi¹⁾

¹⁾ Teknik Informatika STMIK AMIKOM Yogyakarta
Jl Ring road Utara, Condongcatur, Sleman, Yogyakarta 55281
Email : sumarniadi@gmail.com¹⁾

Abstrak

Setiap bulannya Bank Syariah pasti menerima proposal pembiayaan (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 pembiayaan.

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. Untuk itu, dalam penelitian ini akan dibuktikan kemampuan naive bayes classifier untuk mengklasifikasikan data debitur yang menginformasikan resiko pemberian pembiayaan di Bank Syariah. Sebelum dilakukan klasifikasi, data debitur dengan konsep 5C+1S melalui preprocessing. Kemudian dari preprocessing ini dilakukan klasifikasi dengan Naive Bayes Classifier, sehingga menghasilkan model probabilitas klasifikasi untuk penentuan kelas pada debitur selanjutnya. Dari hasil pengujian akurasi model dari sistem yang dikembangkan, menghasilkan nilai akurasi terkecil sebesar 80% dihasilkan pada pengujian dengan sampel sebanyak 100 dan menghasilkan nilai akurasi tertinggi sebesar 98,66% dihasilkan pada pengujian dengan sampel sebanyak 463.

Kata Kunci : 5C+1S, NAP, klasifikasi, preprocessing, naive bayesian classifier, akurasi

1. Pendahuluan

1.1 Latar Belakang

Pembiayaan merupakan sumber utama penghasilan bagi sebuah bank dan juga sekaligus sumber resiko operasi bisnis terbesar, karena sebagian besar dana operasional bank diputar dalam bentuk pembiayaan. 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 pembiayaan bermasalah.

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 pemberian pembiayaan kepada nasabah, oleh karena itu perlu adanya aktivitas penggalian (ekstraksi) data yang masih tersembunyi untuk selanjutnya diolah menjadi pengetahuan yang bermanfaat dalam pengambilan keputusan. Proses yang secara otomatis untuk menemukan informasi yang berharga dari repositori data yang sangat besar disebut dengan data mining[1].

Data mining merupakan sebuah proses dari knowledge discovery (penemuan pengetahuan) dari data yang sangat besar [2]. Dengan demikian, dari tumpukan data tersebut akan didapat beragam informasi yang berharga dan penting yang sebelumnya tidak diketahui.

Ada banyak teknik yang bisa dilakukan untuk melakukan klasifikasi data diantaranya *decision tree*, *bayesian classifier*, *bayesian belief network* dan *rule based classifiers*[2]. Namun dalam penelitian ini akan dikembangkan sistem dengan menggunakan teknik klasifikasi *naive bayes classifiers*. Pendekatan ini merupakan pendekatan yang mengacu pada teorema bayes, dimana teorema ini merupakan prinsip peluang statistika untuk mengkombinasikan pengetahuan sebelumnya dengan pengetahuan baru. Prinsip ini kemudian digunakan untuk memecahkan masalah klasifikasi[1]. Data perbankan akan diolah berdasarkan data-data variabel yang terdapat dalam dokumen calon debitur yang mengajukan pembiayaan yang jumlahnya ratusan bahkan ribuan record. Dengan demikian, dalam pengolahan data akan menghasilkan dimensi yang sangat besar. Oleh sebab itu, diperlukan algoritma yang sederhana agar proses pengolahan dapat efektif dan efisien. Penggunaan algoritma ini dinilai sesuai karena *naive bayes classifier* merupakan salah satu algoritma klasifikasi yang sederhana namun memiliki kemampuan dan akurasi tinggi[3].

Dengan teknik klasifikasi menggunakan algoritma *naive bayes classification* (NBC) dalam pengolahan data tersebut, diharapkan dapat dibangun sebuah model yang mana model tersebut dapat

digunakan untuk menganalisis pemberian pembiayaan kepada calon debitur. Hal ini diharapkan dapat digunakan untuk peringatan dini bagi industri perbankan dalam hal ini Bank Syariah terhadap pemberian pembiayaan kepada calon debitur.

1.2 Rumusan Masalah

Dari latar belakang yang dijelaskan di atas, maka rumusan masalah dalam penelitian ini adalah sebagai berikut :

1. Bagaimana melakukan pemilihan data debitur berdasarkan analisis 5C+1S
2. Bagaimana melakukan proses klasifikasi terhadap data Nota Analisis Pembiayaan (NAP) nasabah untuk menghasilkan sebuah model, kemudian menggunakan model tersebut untuk menganalisis pemberian pembiayaan pada calon debitur.

1.3 Batasan Masalah

Proses data Mining untuk analisis data perbankan pada penelitian ini, di lakukan di Bagian Pembiayaan Bank Syariah, dengan batasan masalah sebagai berikut :

1. Data yang digunakan adalah datadebitur yang mengajukan pembiayaan jenis murabahah (Konsumtif) di bank syaria'h selama kurun waktu 2005-2006, dengan jumlah data yang digunakan sejumlah 463 *record*.
2. Menganalisis pemberian pembiayaan kepada nasabah yang diterima proposal pengajuan pembiayaannya melalui nota analisis pembiayaan (NAP).
3. Penelitian ini hanya meneliti tentang aman atau tidaknya seorang calon debitur diberikan pinjaman.
4. Dalam penelitian ini menggunakan metode klasifikasi dengan algoritma naive bayesian untuk mengklasifikasikan data nasabah.
5. Analisis yang digunakan untuk proses ekstraksi data debitur adalah 5 C + 1 S yaitu : *Character, Capacity, Capital, Collateral, Condition*, dan *Syariah*.

1.4 Metode Penelitian

Penelitian ini dilakukan dengan tahap-tahap sebagai berikut :

1. Pengumpulan Data
2. Analisis dan Pengembangan Sistem
 - a. Perancangan Preprocessing
 - b. Perancangan Naïve Bayesian Classification
3. Pengujian Sistem.

1.4.1 Pengumpulan Data

Pada dasarnya, prinsip yang dipakai dalam penilaian atau menganalisis calon nasabah Bank syariah merupakan prinsip pemberian pembiayaan yang sudah umum dikenal di dunia perbankan syariah dengan Prinsip 5 C + 1 S yaitu :

1. *Character* yaitu penilaian terhadap karakter atau kepribadian calon penerima pembiayaan. Data yang digunakan adalah data pendidikan
2. *Capacity* yaitu penilaian secara *subyektif* tentang kemampuan penerima pembiayaan untuk melakukan

pembayaran. Data yang digunakan adalah data pekerjaan.

3. *Capital* yaitu penilaian terhadap kemampuan modal yang dimiliki oleh calon penerima pembiayaan yang diukur dengan posisi perusahaan secara keseluruhan yang ditunjukan oleh *rasio finansial* dan penekanan pada komposisi modalnya. Data dapat berupa besarnya penghasilan dan jumlah plafon pinjaman dengan waktu pengembalian tertentu.
4. *Collateral* yaitu jaminan yang dimiliki calon penerima pembiayaan.
5. *Condition*, Data dapat berupa kebijakan bank terhadap besarnya keuntungan yang diperoleh disesuaikan dengan jenis usaha debitur. Data yang digunakan adalah margin pembiayaan
6. *Syariah*, Penilaian ini dilakukan untuk menegaskan bahwa usaha yang akan dibiayai benar-benar usaha yang tidak melanggar syariah. Data yang digunakan adalah tujuan pinjam.

1.4.2 Analisis dan Pengembangan Sistem

Setelah tahap analisis dikerjakan dan menghasilkan spesifikasi kebutuhan data dan perangkat lunak kemudian dilakukan pemodelan sistem berdasarkan hasil analisis sistem, maka sistem tersebut akan dikembangkan. Adapun proses pengembangan sistem ini merupakan proses klasifikasi yang dilakukan dengan langkah-langkah sebagai berikut :

1. Setelah data terkumpul, kemudian dilakukan ekstraksi data pada Nota Analisis Pembiayaan (NAP) yaitu mengambil beberapa variabel data berdasarkan analisis 5C+1S, lalu data tersebut diimpor ke database. Kemudian data tersebut mengalami proses preprocessing data yang meliputi *data cleaning, data integration, data transformation*, dan *data reduction* sehingga data menjadi lebih mudah untuk dilakukan proses selanjutnya.
2. Kelas dalam klasifikasi ini adalah kelas “aman” untuk calon debitur yang mengandung informasi dapat mengembalikan dana pinjaman lancar sesuai waktunya dan kelas “tidak aman” untuk calon debitur yang mengandung informasi tidak dapat mengembalikan dana pinjaman secara lancar sesuai waktunya.
3. Memisahkan data untuk digunakan sebagai data training dan data testing. Data training kemudian diklasifikasikan secara manual yang akan digunakan pada proses training. Data testing akan diklasifikasikan menggunakan sistem
4. Dilakukan implementasi algoritma Naive Bayes pada sistem sehingga komputer dapat melakukan klasifikasi dengan menggunakan data training. Proses klasifikasi dengan data training ini dilakukan untuk membentuk model klasifikasi data calon debitur.
5. Dilakukan visualisasi dari hasil klasifikasi yang sudah dilakukan oleh sistem ke dalam aplikasi antar muka berbasis desktop. Data yang ditampilkan dalam aplikasi antar muka adalah

informasi tingkat keamanan pengembalian dana bank dan nilai akurasinya .

- Dilakukan pengujian akurasi model klasifikasi pada sistem yang dikembangkan dan pengujian akurasi model klasifikasi dari hasil data bersih yang dihasilkan dengan menggunakan perangkat lunak lain sebagai pembanding.

1.4.3 Naive Bayes Classifier

Naive bayes atau simple bayesian classifier memiliki prosedur sebagai berikut [2] :

- Setiap sample data direpresentasikan dengan n-dimensional feature vector, $X=(X_1,X_2,...,X_n)$, dengan n dibuat dari sample n atribut, berturut-turut $A_1,A_2,...,A_n$.
- Diandaikan terdapat m class, $C_1,C_2,...,C_m$. Diberikan sebuah data sample, X (yang tidak diketahui class labelnya), kemudian classifier akan memprediksi X ke dalam class yang memiliki probabilitas posterior tertinggi, Naive bayes classifier akan menentukan sample X ke dalam class C_i jika dan hanya jika $P(C_i | X) > P(C_j | X)$ untuk $1 \leq j \leq m, j \neq i \dots (2)$
- Class C_i adalah nilai terbesar, yang disebut dengan maksimum poster hypothesis dengan teorema bayes : $P(C_i | X) = \frac{P(X | C_i) P(C_i)}{P(X)}$ (3)
- $P(X)$ adalah konstan untuk semua class. Jika probabilitas class prior tidak diketahui, secara umum diasumsikan bahwa class adalah sama, yaitu $P(C_1)=P(C_2)=...=P(C_m)$, dan selanjutnya menghitung nilai $P(X|C_i)$ dan menghitung nilai $P(X|C_i)P(C_i)$. Probabilitas class prior diestimasi dengan $P(C_i) = \frac{s_i}{s}$, dimana s_i adalah jumlah training sample pada class C_i , dan s adalah jumlah training sample.
- Apabila dataset terdiri dari banyak atribut, akan mengakibatkan komputasi yang rumit untuk menghitung $P(X|C_i)$. Untuk mengurangi komputasi, naive bayes mengasumsikan pada pembuatan class independen. Sehingga nilai pada atribut dikondisikan bersifat independen antara atribut yang satu dengan atribut yang lain, serta diantara atribut tidak terdapat relasi depedensi.

$$P(X | C_i) = \prod_k^n P(X_k | C_i) \dots\dots\dots (4)$$

- Probabilitas $P(X_1 | C_i), P(X_2 | C_i), \dots, P(X_n | C_i)$, dapat diestimasi dari training sample, dimana

- Jika A_k adalah kategorikal, maka $P(x_k | C_i) = \frac{s_{ik}}{s_i} \dots\dots (5)$

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 .

- Jika A_k bernilai kontinyu, maka diasumsikan mempunyai sebuah gaussian distribusi $P(x_k | C_i) = g(x_k, \mu_{C_i}, \sigma_{C_i}) =$

$$\frac{1}{\sqrt{2\pi\sigma_{C_i}}} e^{-\frac{(x_k - \mu_{C_i})^2}{2\sigma_{C_i}^2}} g(x_k, \mu_{C_i}, \sigma_{C_i}) \text{ adalah fungsi gaussian untuk atribut } A_k \text{ dengan } \mu_{C_i} \text{ dan } \sigma_{C_i}$$

adalah mean dan standard deviasi untuk atribut A_k pada training sample class C_i .

- Untuk mengklasifikasikan sample X yang tidak diketahui, $P(X | C_i)P(C_i)$ dievaluasi untuk setiap class C_i . Sample X ditetapkan untuk class C_i jika dan hanya jika $P(C_i | X) > P(C_j | X)$ untuk $1 \leq j \leq m, j \neq i$. Dengan kata lain, ditetapkan sebagai class C_i untuk $P(C_i | X)$ yang bernilai maksimum.

1.4.4 Pengujian Sistem

Pengujian hasil dengan perangkat lunak lain untuk melakukan proses klasifikasi terhadap nasabah yang mengajukan pembiayaan pada penelitian ini adalah dengan menggunakan rapid miner 5.3. Metode yang digunakan untuk mengklasifikasikan data penelitian ini adalah dengan membandingkan metode Naive Bayes Classifier (NBC) yang telah dibuat dengan NBC pada perangkat lunak lain. Pada pengujian ini akan diperoleh nilai akurasi untuk masing-masing perangkat.

Proses pengujian sistem dilakukan untuk menentukan akurasi proses klasifikasi. Analisis akurasi algoritma Naive Bayes dilakukan dengan menggunakan metode bootstrap. Pada pengujian ini, sebanyak 1/3 dari jumlah nasabah yang mengajukan pembiayaan secara bergantian dijadikan data uji sebanyak 3 kali terhadap 2/3 data lainnya yang dijadikan data training. Nilai akurasi diperoleh dari rata-rata nilai akurasi dari 3 kali pengujian tersebut. Dengan demikian, setiap data nasabah yang mengajukan pembiayaan akan menjadi data training dan data testing secara bergantian. Hal ini bertujuan untuk meminimalkan nilai akurasi yang dihasilkan oleh faktor kebetulan. Untuk menghitung nilai akurasi digunakan persamaan (1).

$$\text{akurasi} = \frac{\text{jumlah klasifikasi benar}}{\text{jumlah data uji}} \times 100\% \dots\dots\dots (1)$$

2. Pembahasan

2.1 Persiapan Data

Pada proses komunikasi dengan pihak Bank Syariah 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 jenis murabahah tahun 2005-2006. Contoh data diperlihatkan pada Tabel 1.

Tabel 1 Contoh data NAP Bank Syariah

No_Rekening	Nama	Pendidikan	Pekerjaan	Gaji	Nilai_jaminan	Tempo	Jumlah_ginjam	Tujuan_ginjam	Angsuran	MARGIN_ginjal_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
1030074990	Abdul Gal/Sarjana	PNS Umum	2547700.00		60	50000000.00	BARANG KONSUMTIF	1228226.05	23753286.29	Aman
1030049194	Abdullah/Sarjana	PNS Umum	3071400.00		96	300000000.00	BARANG KONSUMTIF	582636.30	25833084.93	Aman
7043990416	ABDUL H/Sarjana	PNS Umum	2214500.00	100000000.00	60	37500000.00	RENOVASI RUMAH	921819.54	17815172.47	Aman

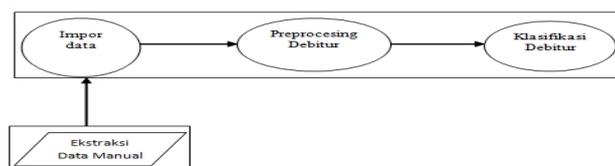
2.2 Analisis dan Pengembangan Sistem

Sistem 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".

2.2.1 Arsitektur Sistem

Secara umum sistem ini terdiri dari dua bagian utama yaitu backend dan frontend. Bagian backend adalah bagian yang dikerjakan secara manual yaitu bagian ekstraksi data manual, sedangkan frontend adalah bagian yang dilakukan secara otomatis, diantaranya impor data, *preprocessing debitur*, dan klasifikasi *debitur*. Adapun desain arsitektur sistem dapat terlihat pada Gambar 4.



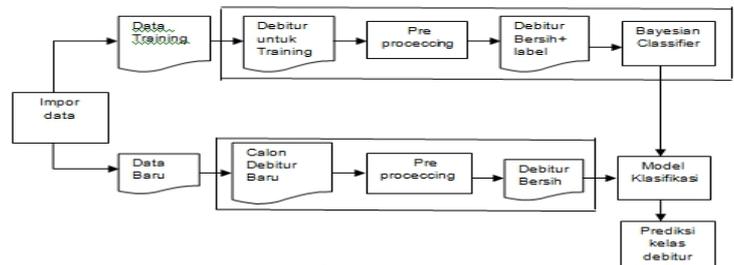
Gambar 4 Rancangan Arsitektur Sistem

Mula-mula melakukan ekstraksi data secara manual dari Nota Analisis Pembiayaan yang bertipe.xls. Ekstraksi data debitur dilakukan dengan menggunakan prinsip 5C+1S, yaitu hanya mengambil data-data debitur yang ada kaitannya dengan prinsip 5C+1S.

Kemudian data hasil ekstraksi disimpan dalam format .csv. Data-data yang sudah diekstraksi akan diimpor secara otomatis oleh sistem dan akan disimpan di database sebagai tabel debitur. Data yang sudah diekstraksi kemudian mengalami proses *preprocessing* secara otomatis untuk menghasilkan data bersih, data yang sudah bersih disimpan dalam database sebagai tabel *debitur bersih*. Setelah itu data diklasifikasikan berdasarkan klasifikasi yang telah ditentukan yang kemudian hasil klasifikasi tersebut

digunakan sebagai model dalam penentuan status debitur berikutnya.

Gambaran umum sistem *frontend* dapat dilihat pada Gambar 5.



Penelitian ini dimulai dengan 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.

2.2.2 Perancangan 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* metode *Discretization and concept hierarchy generation* dengan teknik *logika fuzzy* yaitu mengurangi jumlah nilai yang diberikan atribut kontinyudengan 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 teknik *logika fuzzy* 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.

5. Penyimpanan data “debiturbersih” ke database.

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
	SMA	Sarjana	PNS Umum	PNS	Kecil	Sedang	Kecil	Sedang	Kecil	Sedang	
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

2.2.3 Perancangan Naïve Bayesian Classification

Sebagai contoh penerapan dari naïve bayes dengan menggunakan data training yang merupakan data debitur hasil proses *preprocessing* yang terdapat pada Tabel 4.

Pada tabel tersebut yang berfungsi sebagai label adalah kelas, dengan nilai labelnya adalah “Aman” atau “Tidak Aman”. Dengan demikian kelas “aman” memiliki 5 data debitur dan kelas “tidak aman” memiliki 1 data debitur dari jumlah keseluruhan debitur. Dari data tersebut, maka proses membangun probabilitas kelasnya adalah sebagaiberikut :

1. Kelas “Aman”

Pada tahap ini dikumpulkan seluruh debitur yang mempunyai nilai kelas “aman” dan dihitung nilai probabilitasnya dengan menggunakan persamaan (5). Berikut merupakan contoh perhitungan nilai probabilitas kelas dengan nilai “aman” :

Jumlah debitur “aman” pada data training adalah 5 buah (S_{ik}),

Jumlah data debitur pada data training adalah 6 buah (S_i),

Dengan menggunakan persamaan (5) maka diperoleh:

$$P(\text{Kelas} = \text{“Aman”}) = 5/6 = 0.833$$

Setelah itu lakukan perhitungan probabilitas setiap variabel yang mempunyai kelas “aman” dengan menggunakan persamaan (3). Berikut merupakan contoh perhitungan nilai probabilitas variabel pendidikan yang mempunyai kelas “aman” :

Jumlah debitur “aman” pada data training adalah 5 buah (S_{ik}),

Jumlah data pendidikan yang bernilai “sarjana” dan kelasnya “aman” adalah 5 buah (S_i),

Dengan menggunakan persamaan (3) maka diperoleh:

$$P(\text{Pendidikan} = \text{“sarjana”} | \text{Kelas} = \text{“Aman”}) = 5/5 = 1$$

Dari perhitungan tersebut, diperoleh bahwa probabilitas pendidikan dengan nilai “sarjana” pada kelas “aman” adalah sebesar 1. Hasil perhitungan setiap variabel dengan nilai lain pada kelas “aman” tampak pada Tabel 5.

Tabel 5 Nilai Probabilitas Variabel pada Kelas

Variabel	Pendidikan		Pekerjaan		Gaji		Angsuran		Jaminan	
	SMA	Sarjana	PNS Umum		Kecil	Sedang	Kecil	Sedang	Kecil	Sedang
Nilai	0	1	1		0.8	0.2	1	0	0.4	0.6

2. Kelas “Tidak Aman”

Dengan menggunakan teknik yang sama, dilakukan perhitungan untuk probabilitas kelas dan probabilitas variabel pada kelas “tidak aman”, sehingga diperoleh nilai probabilitas kelas “tidak aman” : $P(\text{Kelas} = \text{“Tidak Aman”}) = 1/6 = 0.166$, sedangkan nilai probabilitas variabel pada kelas “tidak aman” tampak pada Tabel 6.

Tabel 6 Nilai Probabilitas Variabel pada Kelas “Tidak Aman”

Variabel	Pendidikan		Pekerjaan		Gaji		Angsuran		Jaminan	
	SMA	Sarjana	PNS Umum		Kecil	Sedang	Kecil	Sedang	Kecil	Sedang
Nilai	1	0	1		0	1	0	1	1	0

Dengan demikian, data training tersebut membentuk model probabilitas yang dihasilkan oleh algoritma naïve bayes classifier yang dapat dilihat pada Tabel 7.

Tabel 7 Nilai Probabilitas Data Training

Variabel	Pendidikan		Pekerjaan		Gaji		Angsuran		Jaminan	
	SMA	Sarjana	PNS Umum		Kecil	Sedang	Kecil	Sedang	Kecil	Sedang
Aman	0	1	1		0.8	0.2	1	0	0.4	0.6
Tidak Aman	1	0	1		0	1	0	1	1	0

2.3 Hasil Pengujian Akurasi Klasifikasi

Nilai populasi dalam penelitian ini adalah 463 data. Dengan ukuran sampel 1/3 digunakan untuk training dan 2/3 digunakan untuk testing. Secara lebih rinci, berikut merupakan mekanisme pengujian sistem yang dilakukan pada sistem ini :

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

Pengujian Kelas Prediksi dengan teknik Logika Fuzzy

Pada pengujian ini dilakukan perhitungan jumlah kelas prediksi berdasarkan logika fuzzy, pada variabel gaji, variabel angsuran dan variabel jaminan. Hasil dari pengujian kelas prediksi dengan teknik logika fuzzy ditunjukkan pada Tabel 10.

Tabel 10 Pengujian Kelas Prediksi Teknik Logika Fuzzy

										Jumlah Sampel				
Kelas	Gaji									Aman	Tidak Aman	Unknow		
	Kecil			Sedang			Besar							
0	0.1	1	0	0.1	1	0	0.1	1	0	0.1	1	312	150	1
>= 7.000.000	4.000.000 <= gaji <= 7.000.000	0 <= gaji < 4.000.000	0 <= gaji < 4.000.000	10.000.000 <= gaji < 13.000.000	7.000.000 <= gaji < 10.000.000	10.000.000 <= gaji < 11.000.000	10.000.000 <= gaji < 11.000.000	10.000.000 <= gaji < 13.000.000	0	0	0			
										Jumlah Sampel				
Kelas	Jaminan									Aman	Tidak Aman	Unknow		
	Kecil			Sedang			Besar							
0	0.1	1	0	0.1	1	0	0.1	1	0	0.1	1	312	150	1
>= 100.000.000	90.000.000 <= jaminan <= 100.000.000	0 <= jaminan < 90.000.000	150.000.000 <= jaminan < 160.000.000	100.000.000 <= jaminan < 150.000.000	100.000.000 <= jaminan < 150.000.000	150.000.000 <= jaminan < 160.000.000	0	0	0					
										Jumlah Sampel				
Kelas	Angsuran									Aman	Tidak Aman	Unknow		
	Kecil			Sedang			Besar							
0	0.1	1	0	0.1	1	0	0.1	1	0	0.1	1	312	150	1
>= 3.500.000	2.000.000 <= angsuran <= 3.500.000	0 <= angsuran < 2.000.000	3.500.000 <= angsuran < 4.000.000	2.000.000 <= angsuran < 3.500.000	2.000.000 <= angsuran < 3.500.000	3.500.000 <= angsuran < 4.000.000	0	0	0					

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 11 merupakan rincian hasil pengujian model probabilitas kelas dengan teknik logika fuzzy.

Tabel 11 Pengujian Akurasi Model Probabilitas Kelas teknik Logika Fuzzy

Gaji										Probabilitas	
Kecil			Sedang				Besar			Aman	Tidak Aman
0	0,1	1	0	0,1	1	0	0,1	1	0,68	0,32	
>= 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	0,68	0,32	
>= 100.000,000	90.000.000 >= jaminan <= 100.000.000	0 >= jaminan <= 90.000.000	Jaminan >= 150.000,000	150.000,000 <= Jaminan <= 160.000,000	0 = Jaminan <= 160.000,000	100.000,000 <= Jaminan <= 150.000,000	Jaminan <= 150.000,000	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	0,68	0,32	
>= 3.500,000	2.000,000 >= Angsuran <= 2.000,000	0 <= Angsuran <= 2.000,000	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 >= 5.500.000	0,68	0,32	

Pengujian Akurasi Model

Pengujian ini dilakukan pengujian akurasi model probabilitas terhadap data debitur dengan berbagai porsi jumlah data sampel. Pengujian akurasi model menggunakan *Bootstrap*, yang ditunjukkan pada Tabel 12.

Tabel 12 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

Perbandingan validasi model dengan Rapid Miner 5.3

Pengujian ini dilakukan untuk melakukan perbandingan validasi model algoritma naive bayes yang dihasilkan dengan aplikasi lain yaitu Rapid Miner 5.3 dengan menggunakan algoritma yang sama. Dapat dibuat rangkuman yang hasilnya ditunjukkan pada Tabel 13.

Tabel 13 Perbandingan validasi model dengan Rapid Miner

Jumlah data	Model Naive Bayes (%)	Rapid Miner 5.3 (Naive Bayes) (%)
100	80.00	64.79
250	84.10	66.58
350	89.90	74.01
463	98.66	80.06

3. Kesimpulan

Dari penelitian yang telah dilakukan terdapat beberapa kesimpulan yaitu :

1. Prinsip 5C+1S sangat membantu sistem dalam proses analisis data pemberian pembiayaan pada nasabah
2. Variabel yang mempunyai nilai sama sebaiknya direduksi saja karena tidak mempengaruhi akurasi model sebaliknya malah menambah beban komputasi

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.
5. Hasil pengujian akurasi dengan menggunakan perangkat lunak Rapid Miner 5.3 dengan metode *naive bayes classifier* menghasilkan nilai akurasi terkecil sebesar 64,79% pada proses pengujian dengan menggunakan sampel sebanyak 100 dan menghasilkan nilai akurasi tertinggi sebesar 80,06% 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*.

Saran

penelitian ini masih memiliki keterbatasan yang dapat dijadikan acuan untuk pengembangan dimasa yang akan datang, sehingga dapat disarankan beberapa hal sebagai berikut :

1. Perlu dilakukan analisis hubungan antar variabel pada proses klasifikasi.
2. Perlu dilakukan perbandingan dengan teknik lain pada proses *preprocessing* variabel kontinue
3. 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] Tan, P. N., Stenbach, M., & Kumar, V., *Introduction to Data Mining*. Boston: Pearson Education, 2006.
 [2] Han, J., & Kamber, M., *Data Mining Concept and Technique*, San Francisco: Morgan Kaufman Publisher, 2006
 [3] Rish, I., "An empirical study of the Naive Bayes Classifier", in Proc. International Joint Conference on Artificial Intelligence, pp. 41-46, 2006.
 [4] Teknomo, K. 2005. *Bootstrap Sampling Tutorial*. Diakses tanggal 7 September 2013
<http://people.revoledu.com/kardi/tutorial/bootstrap/bootstrap.htm>.
 [5] Kusriani & Luthfi, T. E., 2009, *Algoritma Data Mining*, Andi Offset, Yogyakarta.

Biodata Penulis

Sumarni Adi, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Informatika STMIK AMIKOM Yogyakarta, lulus tahun 2009. Memperoleh gelar *Master Of Computer Science (M.Cs)* Program Pasca Sarjana Ilmu Komputer Universitas Gajah Mada Yogyakarta, lulus tahun 2013. Saat ini menjadi Dosen di STMIK AMIKOM Yogyakarta.