

PENERAPAN ALGORITMA NAIVE BAYES UNTUK MENGLASIFIKASI DATA NASABAH ASURANSI

Bustami

Teknik Informatika, Universitas Malikussaleh
Reuleut, Aceh Utara, Aceh-Indonesia
E-mail : busabiel@gmail.com

ABSTRAK

Data mining adalah teknik yang memanfaatkan data dalam jumlah yang besar untuk memperoleh informasi berharga yang sebelumnya tidak diketahui dan dapat dimanfaatkan untuk pengambilan keputusan penting. Pada penelitian ini, penulis berusaha menambang data (data mining) nasabah sebuah perusahaan asuransi untuk mengetahui lancar, kurang lancar atau tidak lancarnya nasabah tersebut. Data yang ada dianalisis menggunakan algoritma Naive Bayes. Naive Bayes merupakan salah satu metode pada probabilistic reasoning. Algoritma Naive Bayes bertujuan untuk melakukan klasifikasi data pada kelas tertentu, kemudian pola tersebut dapat digunakan untuk memperkirakan nasabah yang bergabung, sehingga perusahaan bisa mengambil keputusan menerima atau menolak calon nasabah tersebut.

Kata Kunci : data mining, asuransi, klasifikasi, algoritma Naive Bayes

1. PENDAHULUAN

Premi merupakan pendapatan bagi perusahaan asuransi, yang jumlahnya ditentukan dalam suatu persentase atau tarif tertentu dari jumlah yang dipertanggungjawabkan. Bagi tertanggung premi merupakan beban karena membayar premi merupakan beban tertanggung. Pendapatan premi untuk perusahaan asuransi ditentukan oleh jumlah premi yang dibayar oleh nasabah.

Permasalahan yang sering timbul dalam perusahaan asuransi adalah banyaknya nasabah yang menunggak dalam membayar premi, sehingga diperlukan sebuah sistem yang dapat mengklasifikasikan nasabah mana yang masuk ke dalam kelompok lancar, kelompok kurang lancar dan nasabah mana yang masuk ke dalam kelompok tidak lancar dalam membayar iuran premi. Sehingga pihak asuransi bisa mengatasi sejak dini permasalahan tersebut.

Sebuah perusahaan asuransi pastilah mempunyai data yang begitu besar. Banyak yang belum menyadari bahwa dari pengolahan data – data tersebut dapat memberikan informasi berupa klasifikasi data nasabah yang akan bergabung pada perusahaan itu sendiri. Penggunaan teknik data mining diharapkan mampu memberikan informasi yang berguna tentang teknik klasifikasi data nasabah yang akan bergabung dalam kelompok lancar, kelompok kurang lancar atau tidak lancar dalam membayar premi.

2. LANDASAN TEORI

2.1. Data Mining

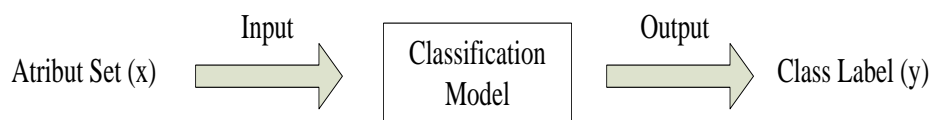
Data mining adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar. Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data.

Tahapan dari proses *Knowledge Discovery in Database* (KDD) adalah :

1. Selection
2. Pre-Processing / Cleaning.
3. Transformation
4. Data Mining
5. Interpretation / Evaluation.

2.2. Metode Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Dalam mencapai tujuan tersebut, proses klasifikasi membentuk suatu model yang mampu membedakan data kedalam kelas-kelas yang berbeda berdasarkan aturan atau fungsi tertentu. Model itu sendiri bisa berupa aturan “jika-maka”, berupa pohon keputusan, atau formula matematis.



Gambar 1. Blok Diagram Model Klasifikasi

2.3. Algoritma Naive Bayes

Algoritma *Naive Bayes* merupakan salah satu algoritma yang terdapat pada teknik klasifikasi. *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris *Thomas Bayes*, yaitu memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya sehingga dikenal sebagai *Teorema Bayes*. Teorema tersebut dikombinasikan dengan *Naive* dimana diasumsikan kondisi antar atribut saling bebas. Klasifikasi *Naive Bayes* diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri dari kelas lainnya.

Persamaan dari teorema *Bayes* adalah :

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

Keterangan :

- X : Data dengan class yang belum diketahui
- H : Hipotesis data X merupakan suatu class spesifik
- $P(H|X)$: Probabilitas hipotesis H berdasar kondisi X (posteriori probability)
- $P(H)$: Probabilitas hipotesis H (prior probability)
- $P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H
- $P(X)$: Probabilitas X

Untuk menjelaskan teorema *Naive Bayes*, perlu diketahui bahwa proses klasifikasi memerlukan sejumlah petunjuk untuk menentukan kelas apa yang cocok bagi sampel yang dianalisis tersebut. Karena itu, teorema *bayes* di atas disesuaikan sebagai berikut :

$$P(C|F_1 \dots F_n) = \frac{P(C)P(F_1 \dots F_n|C)}{P(F_1, \dots F_n)} \dots\dots\dots (2)$$

Dimana Variabel C merepresentasikan kelas, sementara variabel $F_1 \dots F_n$ merepresentasikan karakteristik petunjuk yang dibutuhkan untuk melakukan klasifikasi. Maka rumus tersebut menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas C (*Posterior*) adalah peluang munculnya kelas C (sebelum masuknya sampel tersebut, seringkali disebut *prior*), dikali dengan peluang kemunculan karakteristik – karakteristik sampel pada kelas C (disebut juga *likelihood*), dibagi dengan peluang kemunculan karakteristik – karakteristik sampel secara global (disebut juga *evidence*). Karena itu, rumus diatas dapat pula ditulis secara sederhana sebagai berikut :

$$Posterior = \frac{Prior \times likelihood}{evidence}$$

Nilai *Evidence* selalu tetap untuk setiap kelas pada satu sampel. Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai – nilai *posterior* kelas lainnya untuk menentukan ke kelas apa suatu sampel akan diklasifikasikan. Penjabaran lebih lanjut rumus *Bayes* tersebut dilakukan dengan menjabarkan $(C|F_1, \dots, F_n)$ menggunakan aturan perkalian sebagai berikut :

$$\begin{aligned} P(C|F_1, \dots, F_n) &= P(C) P(F_1, \dots, F_n|C) \\ &= P(C)P(F_1|C)P(F_2, \dots, F_n|C, F_1) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3, \dots, F_n|C, F_1, F_2) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2), P(F_4, \dots, F_n|C, F_1, F_2, F_3) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1}) \end{aligned} \dots\dots\dots (3)$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor – faktor syarat yang mempengaruhi nilai probabilitas, yang hampir mustahil untuk dianalisa satu persatu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Disinilah digunakan asumsi independensi yang sangat tinggi (*naif*), bahwa masing – masing petunjuk

$(F_1, F_2 \dots F_n)$ saling bebas (*independen*) satu sama lain. Dengan asumsi tersebut, maka berlaku suatu kesamaan sebagai berikut :

$$P(P_i|F_j) = \frac{P(F_i \cap F_j)}{P(F_j)} = \frac{P(F_i)P(F_j)}{P(F_j)} = P(F_i)$$

Untuk $i \neq j$, sehingga

$$P(F_i|C, F_j) = P(F_i|C) \dots\dots\dots (4)$$

Dari persamaan diatas dapat disimpulkan bahwa asumsi independensi naif tersebut membuat syarat peluang menjadi sederhana, sehingga perhitungan menjadi mungkin untuk dilakukan. Selanjutnya, penjabaran $P(C|F_1, \dots, F_n)$ dapat disederhanakan menjadi :

$$P(C|F_1, \dots, F_n) = P(C)P(F_1|C)P(F_2|C)P(F_3|C) \dots$$

$$= P(C) \prod_{i=1}^n P(F_i|C) \dots\dots\dots (5)$$

Persamaan diatas merupakan model dari teorema *Naive Bayes* yang selanjutnya akan digunakan dalam proses klasifikasi. Untuk klasifikasi dengan data kontinyu digunakan rumus *Densitas Gauss* :

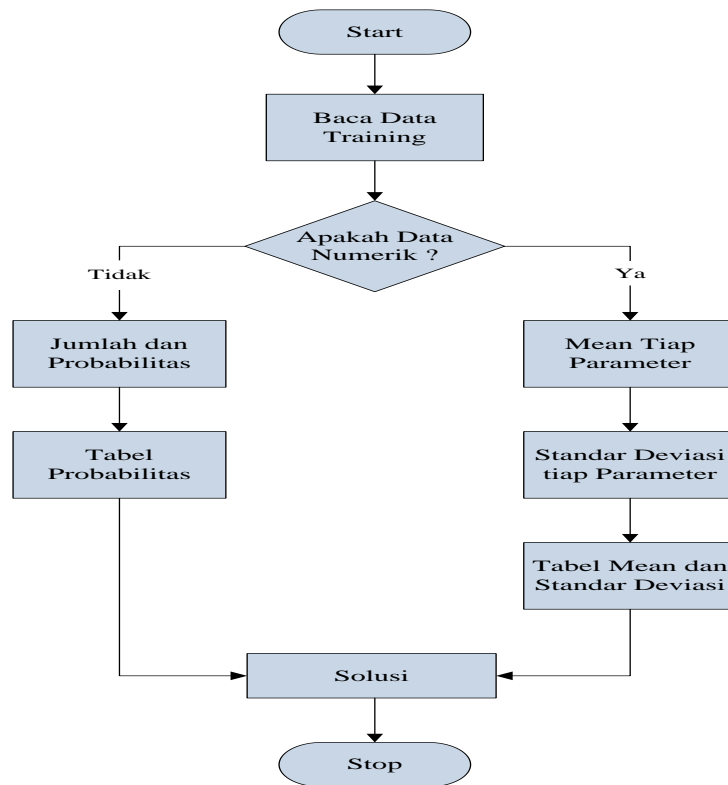
$$P(X_i = x_i|Y = y_j) = \frac{1}{\sqrt{2\pi}\sigma_{ij}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \dots\dots\dots (6)$$

Keterangan :

- P : Peluang
- X_i : Atribut ke i
- x_i : Nilai atribut ke i
- Y : Kelas yang dicari
- y_j : 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* adalah sebagai berikut :

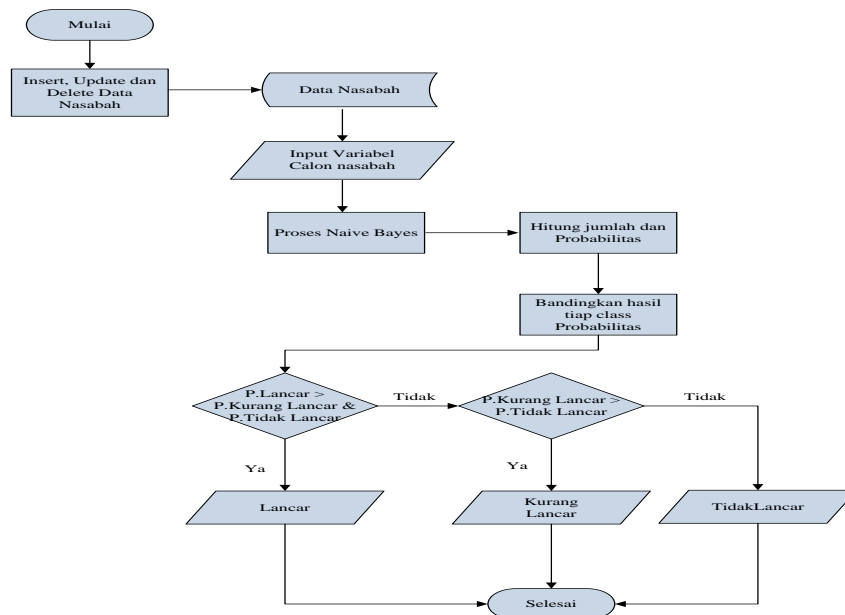
1. *Baca data training*
2. *Hitung Jumlah dan probabilitas, namun apabila data numerik maka:*
 - a. *Cari nilai mean dan standar deviasi dari masing – masing parameter yang merupakan data numerik.*
 - b. *Cari nilai probabilistik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.*
3. *Mendapatkan nilai dalam tabel mean, standart deviasi dan probabilitas.*



Gambar 2. Skema *Naive Bayes*

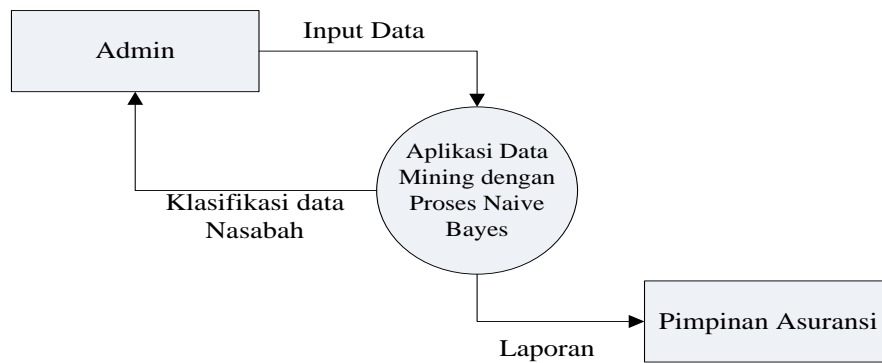
3. PERANCANGAN SISTEM

3.1. Flowchart Sistem



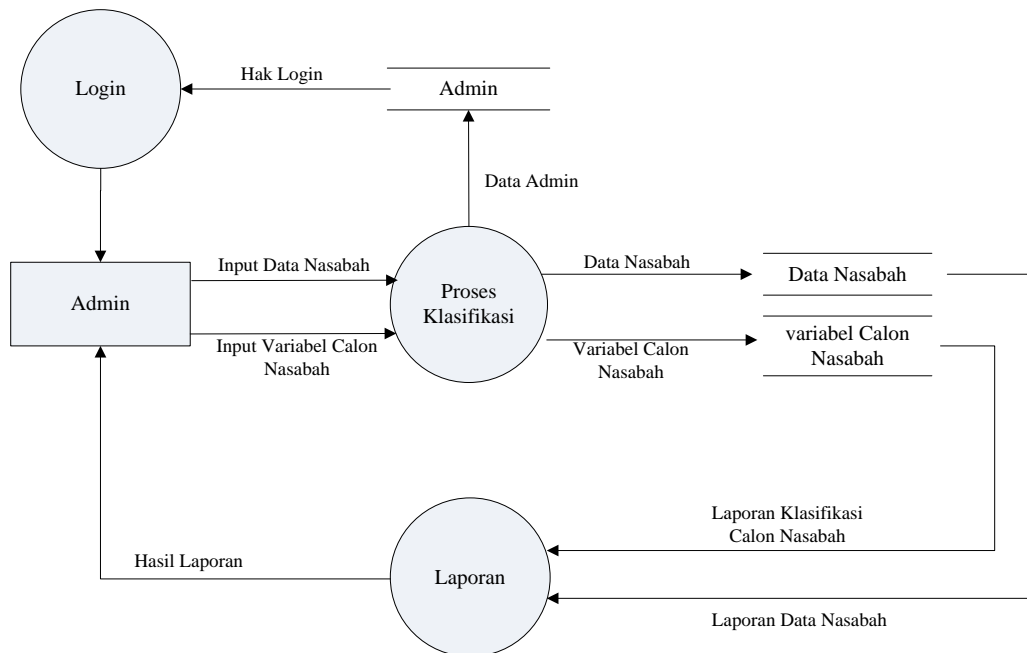
Gambar 3. Flowchart sistem

3.2. Diagram Konteks



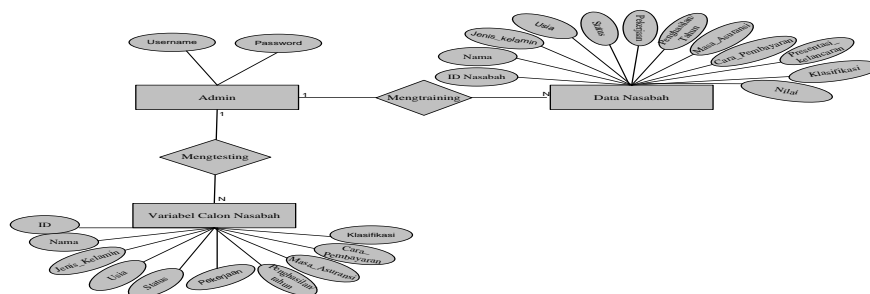
Gambar 4. Diagram Konteks

3.3. Data Flow Diagram (DFD)



Gambar 5. Data Flow Diagram (DFD) Level 0

3.4. Entity Relationship Diagram (ERD)



Gambar 6. Entity Relationship Diagram (ERD)

4. PERANCANGAN BASIS DATA

4.1. Desain Tabel Admin

Tabel ini digunakan untuk menyimpan data admin

Tabel 1. Admin Asuransi

Nama Field	Tipe Data	Lebar	Keterangan
Username	Varchar	10	Nama User
Password	Varchar	10	Password User

4.2. Desain Tabel Data Nasabah

Tabel ini digunakan untuk menyimpan data nasabah yang akan digunakan dalam sistem.

Tabel 2. Data Nasabah

Nama Field	Tipe Data	Lebar	Keterangan
ID_nasabah	Integer	5	No Id nasabah
Nama	Varchar	30	Nama nasabah
Jenis Kelamin	Varchar	10	Jenis Kelamin Nasabah
Usia	Varchar	20	Usia nasabah
Status	Varchar	15	Status nasabah
Pekerjaan	Varchar	20	Pekerjaan nasabah
Penghasilan/tahun	Varchar	20	Penghasilan nasabah
Masa_asuransi	Varchar	20	Masa asuransi
Cara_pembayaran	Varchar	15	Cara pembayaran
Persentasi_kelancaran	Integer	3	Persentasi kelancaran
Klasifikasi	Varchar	20	Klasifikasi
Nilai	Integer	4	Nilai Data

4.3. Desain Tabel Variabel Calon Nasabah

Tabel ini digunakan untuk menyimpan variabel yang akan digunakan dalam sistem.

Tabel 3. Variabel Data Calon Nasabah

Nama Field	Tipe Data	Lebar	Keterangan
ID	Integert	5	ID Calon Nasabah
Nama	Varchar	30	Nama Calon nasabah
Jenis Kelamin	Varchar	10	Kriteria
Usia	Varchar	20	Kriteria
Status	Varchar	15	Kriteria
Pekerjaan	Varchar	20	Kriteria
Penghasilan/tahun	Varchar	20	Kriteria
Masa_asuransi	Varchar	20	Kriteria
Cara_pembayaran	Varchar	15	Kriteria
Klasifikasi	Varchar	20	Kriteria

5. IMPLEMENTASI DENGAN PERHITUNGAN NAIVE BAYES

Model statistik merupakan salah satu model yang efisien sebagai pendukung pengambilan keputusan. Konsep probabilistik merupakan salah satu bentuk model statistik. Salah satu metode yang menggunakan konsep probabilistik adalah *Naive Bayes*. Algoritma *Naive Bayes* adalah salah satu algoritma dalam teknik klasifikasi yang mudah diimplementasikan dan cepat prosesnya. Pada metode ini, semua atribut akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting dan setiap atribut saling bebas satu sama lain. Apabila diberikan k atribut yang saling bebas (*independence*), nilai probabilitas dapat diberikan sebagai berikut :

$$P(x_1, \dots, x_k | C) = P(x_1 | C) \dots P(x_k | C)$$

Tahap awal cara kerja dari proses perhitungan *Naive Bayes* adalah dengan melakukan pengambilan data training dari data nasabah asuransi. Adapun variabel penentu yang digunakan dalam mengklasifikasikan data nasabah yaitu:

- a. Jenis Kelamin
Merupakan variabel jenis kelamin nasabah yang dikelompokkan dalam dua kategori yaitu laki – laki dan perempuan.
- b. Usia
Merupakan variabel usia nasabah yang di kelompokkan dalam tiga kategori yaitu 20-29 tahun, 30-40 tahun, dan diatas 40 tahun.
- c. Status
Merupakan variabel status nasabah yang dikelompokkan dalam dua kategori yaitu kawin dan belum kawin.
- d. Pekerjaan
Merupakan variabel pekerjaan nasabah yang dikelompokkan dalam tiga kategori yaitu PNS, Pegawai Swasta, Wiraswasta.
- e. Penghasilan
Merupakan variabel penghasilan dari nasabah yang dikelompokkan dalam tiga kategori yaitu 0-25 juta, 25-50 juta, dan diatas 50 juta.
- f. Cara pembayaran premi
Merupakan variabel cara pembayaran premi yang dikelompokkan dalam empat kategori yaitu bulanan, triwulan, semesteran, dan tahunan.
- g. Masa pembayaran premi
Merupakan variabel masa pembayaran premi yang dikelompokkan dalam tiga kategori yaitu 5-10 tahun, 11 -15 tahun, dan diatas 15 tahun.

Tabel 4. Data Pelatihan

No	Nama	Jenis Kelamin	Usia	Status	Pekerjaan	Penghasilan	Masa Asuransi	Cara Pembayaran	Klasifikasi
1	Dani Lukman	Laki-Laki	30 - 40 Tahun	Kawin	Pns	<25 Juta	>15 Tahun	Tahunan	Tidak Lancar
2	Evaliana	Perempuan	30 - 40 Tahun	Kawin	Pns	<25 Juta	5 - 10 Tahun	Semesteran	Lancar
3	Rasyidah	Perempuan	20 - 29 Tahun	Kawin	Pegawai Swasta	<25 Juta	5 - 10 Tahun	Triwulan	Tidak Lancar
4	Dina Saufika	Perempuan	30 - 40 Tahun	Belum Kawin	Pns	<25 Juta	5 - 10 Tahun	Triwulan	Lancar
5	Wilsa Rizki	Laki-Laki	30 - 40 Tahun	Kawin	Wiraswasta	<25 Juta	5 - 10 Tahun	Tahunan	Kurang Lancar
6	Irwanto	Laki-Laki	30 - 40 Tahun	Belum Kawin	Wiraswasta	>50 Juta	11 - 15 Tahun	Semesteran	Lancar
7	Ade Gunawan	Laki-Laki	30-40 Tahun	Kawin	Pns	25 - 50 Juta	11 - 15 Tahun	Semesteran	Tidak Lancar
8	Fauziah	Perempuan	20 - 29 Tahun	Kawin	Wiraswasta	25 - 50 Juta	11 - 15 Tahun	Tahunan	Lancar
9	Zulaikha	Perempuan	20 - 29 Tahun	Kawin	Wiraswasta	<25 Juta	11 - 15 Tahun	Triwulan	Tidak Lancar
10	Zulfahmi	Laki-Laki	20 - 29 Tahun	Kawin	Pns	<25 Juta	11 - 15 Tahun	Triwulan	Kurang Lancar
11	Hidayatullah	Laki-Laki	30 - 40 Tahun	Belum Kawin	Wiraswasta	25 - 50 Juta	11 - 15 Tahun	Tahunan	Lancar
12	Nilam Sari	Perempuan	30 - 40 Tahun	Kawin	Wiraswasta	25 - 50 Juta	>15 Tahun	Tahunan	Kurang Lancar
13	Nahari Anifin	Laki-Laki	30 - 40 Tahun	Kawin	Wiraswasta	>50 Juta	11 - 15 Tahun	Triwulan	Lancar
14	Yusnidar	Perempuan	>40 Tahun	Kawin	Pns	<25 Juta	>15 Tahun	Semesteran	Kurang Lancar
15	Rizwan Hadi	Laki-Laki	20-29 Tahun	Belum Kawin	Pns	<25 Juta	11 - 15 Tahun	Tahunan	Lancar
16	Rahmat Saputra	Laki-Laki	30-40 Tahun	Belum Kawin	Wiraswasta	<25 Juta	11 - 15 Tahun	Semesteran	Lancar
17	M. Sahril	Laki-Laki	>40 Tahun	Kawin	Pegawai swasta	<25 Juta	11 - 15 Tahun	Tahunan	Tidak Lancar
18	M. Irfan	Laki-Laki	30 - 40 Tahun	Kawin	Pegawai swasta	25 - 50 Juta	11 - 15 Tahun	Tahunan	Tidak Lancar
19	Tutri Wulandari	Perempuan	30 - 40 Tahun	Kawin	Wiraswasta	<25 Juta	11 - 15 Tahun	Triwulan	Lancar
20	Leni Syamsiah	Perempuan	20 - 29 Tahun	Belum Kawin	Wiraswasta	25 - 50 Juta	5 - 10 Tahun	Bulanan	Tidak Lancar
21	Syafi Arkan	Laki-Laki	30 - 40 Tahun	Kawin	wiraswasta	25 - 50 Juta	11 - 15 Tahun	Semesteran	????

Berdasarkan tabel diatas dapat dihitung klasifikasi data nasabah apabila diberikan input berupa jenis kelamin, usia, status, pekerjaan, penghasilan/tahun, masa asuransi dan cara pembayaran menggunakan algoritma *Naive Bayes*.

Apabila diberikan input baru, maka klasifikasi data nasabah asuransi dapat ditentukan melalui langkah berikut :

1. Menghitung jumlah class/label
 - $P(Y=Lancar) = 9/20$ “Jumlah data lancar pada data pelatihan dibagi dengan jumlah keseluruhan data”
 - $P(Y=Kurang Lancar) = 4/20$ “Jumlah data kurang lancar pada data pelatihan dibagi dengan jumlah keseluruhan data”
 - $P(Y= Tidak Lancar) = 7/20$ “Jumlah tidak lancar pada data pelatihan dibagi dengan jumlah keseluruhan data”

2. Menghitung jumlah kasus yang sama dengan class yang sama
 - $P(\text{Jenis Kelamin} = \text{Laki-laki} | Y=Lancar) = 5/9$
 - $P(\text{Jenis Kelamin} = \text{Laki-laki} | Y=Kurang Lancar) = 2/4$
 - $P(\text{Jenis Kelamin} = \text{Laki-laki} | Y=Tidak Lancar) = 4/7$

 - $P(\text{Usia} = 30 - 40 \text{ Tahun} | Y=Lancar) = 7/9$
 - $P(\text{Usia} = 30 - 40 \text{ Tahun} | Y=Kurang Lancar) = 2/4$
 - $P(\text{Usia} = 30 - 40 \text{ Tahun} | Y=Tidak Lancar) = 3/7$

 - $P(\text{Status} = \text{Kawin} | Y=Lancar) = 4/9$
 - $P(\text{Status} = \text{Kawin} | Y=Kurang Lancar) = 4/4$
 - $P(\text{Status} = \text{Kawin} | Y=Tidak Lancar) = 6/7$

 - $P(\text{Pekerjaan} = \text{Wiraswasta} | Y=Lancar) = 6/9$
 - $P(\text{Pekerjaan} = \text{Wiraswasta} | Y=Kurang Lancar) = 2/4$
 - $P(\text{Pekerjaan} = \text{Wiraswasta} | Y=Tidak Lancar) = 2/7$

 - $P(\text{Penghasilan} = 25 - 50 \text{ Juta} | Y=Lancar) = 2/9$
 - $P(\text{Penghasilan} = 25 - 50 \text{ Juta} | Y=Kurang Lancar) = 1/4$
 - $P(\text{Penghasilan} = 25 - 50 \text{ Juta} | Y=Tidak Lancar) = 3/7$

 - $P(\text{Masa_Asuransi} = 11 - 15 \text{ Tahun} | Y=Lancar) = 7/9$
 - $P(\text{Masa_Asuransi} = 11 - 15 \text{ Tahun} | Y=Kurang Lancar) = 1/4$
 - $P(\text{Masa_Asuransi} = 11 - 15 \text{ Tahun} | Y=Tidak Lancar) = 4/7$

 - $P(\text{Cara Pembayaran} = \text{Semesteran} | Y=Lancar) = 3/9$
 - $P(\text{Cara Pembayaran} = \text{Semesteran} | Y=Kurang Lancar) = 1/4$
 - $P(\text{Cara Pembayaran} = \text{Semesteran} | Y=Tidak Lancar) = 1/7$

3. Kalikan semua hasil variabel Lancar, Kurang Lancar dan Tidak Lancar
 - $P(\text{Laki} - \text{Laki} \backslash \text{Lancar}) * P(30 - 40 \text{ Tahun} \backslash \text{Lancar}) * P(\text{Kawin} \backslash \text{Lancar}).$
 - $P(\text{Wiraswasta} \backslash \text{Lancar}) * P(25 - 50 \text{ Juta} \backslash \text{Lancar}) * P(11 - 15 \text{ Tahun} \backslash \text{Lancar}). P(\text{Semesteran} \backslash \text{Lancar}) * P(\text{Lancar})$
 - $= \frac{5}{9} \times \frac{7}{9} \times \frac{4}{9} \times \frac{6}{9} \times \frac{2}{9} \times \frac{7}{9} \times \frac{3}{9} \times \frac{9}{20}$
 - $= 0,5556 \times 0,7778 \times 0,4444 \times 0,6667 \times 0,2222 \times 0,7778 \times$
 - $0,3333 \times 0,45$
 - $= 0,0033$

$$\begin{aligned}
 & P(\text{Laki - Laki} \setminus \text{Kurang Lancar}) * P(30 - 40 \text{ Tahun} \setminus \text{Kurang Lancar}) * \\
 & P(\text{Kawin} \setminus \text{Kurang Lancar}) * P(\text{Wiraswasta} \setminus \text{Kurang Lancar}) * P(25 - 50 \\
 & \text{Juta} \setminus \text{Kurang Lancar}) * P(11 - 15 \text{ Tahun} \setminus \text{Kurang Lancar}). \\
 & P(\text{Semesteran} \setminus \text{Kurang Lancar}) * P(\text{Kurang Lancar}) \\
 & = \frac{2}{4} \times \frac{2}{4} \times \frac{4}{4} \times \frac{2}{4} \times \frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} \times \frac{4}{20} \\
 & = 0,5 \times 0,5 \times 1 \times 0,5 \times 0,25 \times 0,25 \times 0,25 \times 0,2 \\
 & = 0,0004
 \end{aligned}$$

$$\begin{aligned}
 & P(\text{Laki - Laki} \setminus \text{Tidak Lancar}) * P(30 - 40 \text{ Tahun} \setminus \text{Tidak Lancar}) * \\
 & P(\text{Kawin} \setminus \text{Tidak Lancar}) * P(\text{Wiraswasta} \setminus \text{Tidak Lancar}) * P(25 - 50 \\
 & \text{Juta} \setminus \text{Tidak Lancar}) * P(11 - 15 \text{ Tahun} \setminus \text{Tidak Lancar}) * P(\text{Semesteran} \setminus \text{Tidak} \\
 & \text{Lancar}). P(\text{Tidak Lancar}) \\
 & = \frac{4}{7} \times \frac{3}{7} \times \frac{6}{7} \times \frac{2}{7} \times \frac{3}{7} \times \frac{4}{7} \times \frac{1}{7} \times \frac{7}{20} \\
 & = 0,5714 \times 0,4286 \times 0,857 \times 0,2857 \times 0,4286 \times 0,5714 \times \\
 & 0,1429 \times 0,35 \\
 & = 0,0007
 \end{aligned}$$

4. Bandingkan hasil class Lancar, Kurang Lancar dan Tidak Lancar
 Dari hasil diatas, terlihat bahwa nilai probabilitas tertinggi ada pada kelas (P|Lancar) sehingga dapat disimpulkan bahwa **status calon nasabah tersebut masuk dalam klasifikasi “Lancar”**.

6. IMPLEMENTASI SISTEM

Setelah melalui tahapan perancangan sistem, databaseselanjutnya adalah implementasi sistem. Implementasi sistem merupakan bagian akhir daripada perancangan sistem yang telah dibangun dimana tahapan ini juga merupakan testing program.

a. Form Login

Form login berfungsi sebagai form keamanan, form ini merupakan form untuk masuk ke program yang akan diakses dengan cara mengisikan username dan password. Jika hak akses telah diberikan oleh sistem maka user dapat mengakses menu utama aplikasi. Adapun tampilan form login dapat dilihat pada gambar berikut :

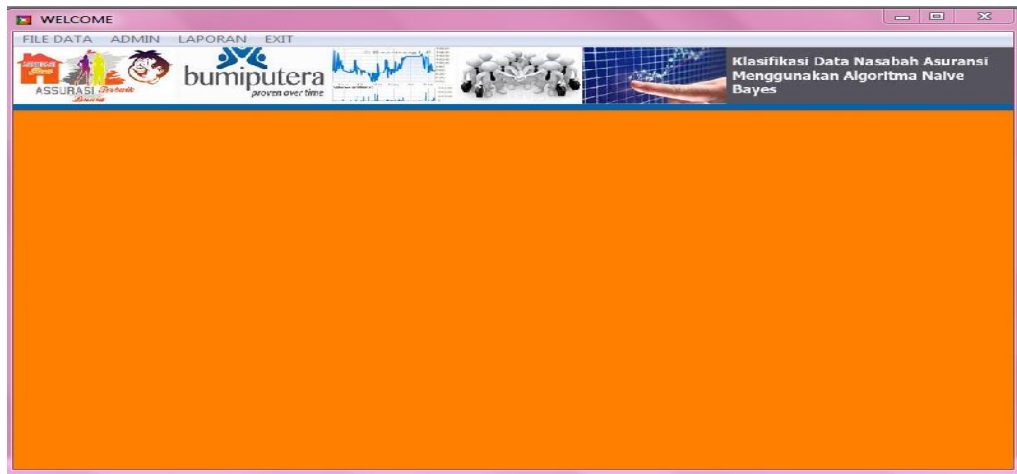


Gambar 8. Tampilan Form Login

b. Form Menu Utama

Form menu utama berfungsi untuk mengakses segala perintah yang terdapat dalam aplikasi. Form tersebut dapat diakses setelah user melakukan login. Pada form ini terdapat beberapa menu yaitu Menu File

Data yang berisi submenu data nasabah (*data training*) dan cek persentase kelancaran (*data testing*), Menu Admin, Laporan dan Exit. Adapun tampilannya dapat dilihat pada gambar berikut :



Gambar 9. Tampilan Form Menu Utama

c. Form Data Nasabah

Form ini dapat ditampilkan melalui menu form utama > file data>data nasabah. Form ini berfungsi untuk mencari data nasabah, menambah, menghapus, menyimpan data nasabah. Data nasabah inilah yang selanjutnya digunakan untuk data pelatihan (*training*) untuk proses klasifikasi. Adapun tampilannya dapat dilihat pada gambar berikut:



Gambar 10. Tampilan Form Data Nasabah

d. Form Cek Persentasi Kelancaran

Untuk memanggil form cek persentasi kelancaran dapat dilakukan melalui menu form utama > file data >cek persentasi kelancaran. Form ini adalah form data testing yang digunakan untuk mengecek tingkat kelancaran calon nasabah. Adapun tampilannya dapat dilihat pada gambar berikut :

Gambar 11. Tampilan Form Cek Persentase Kelancaran

- e. Form Hasil Input Data Calon Nasabah
- Pada form ini menampilkan hasil output dari penginputan data calon nasabah yang telah di proses dengan algoritma *Naive Bayes*. Proses klasifikasi dipengaruhi oleh atribut – atribut terpilih yang mendukung untuk menentukan kelas nasabah lancar, kurang lancar dan tidak lancar. Adapun tampilannya dapat dilihat pada gambar berikut :

Gambar 12. Tampilan Form Hasil Input Data Calon Nasabah

- f. Form Laporan Akhir
- Laporan akhir merupakan output dari proses klasifikasi data. Laporan ini menampilkan hasil akhir dari proses yang telah dilakukan yaitu output dari penginputan data calon nasabah yang telah di proses dengan algoritma *Naive Bayes*. Pada menu laporan juga terdapat submenu laporan berdasarkan klasifikasi nasabah lancar, kurang lancar atau tidak lancar. Adapun tampilan form laporan dapat dilihat pada gambar berikut :

ID	NAMA	JENIS KELAMIN	KATEGORI USIA	STATUS	PEKERJAAN	PENGHASILAN PERTAHUN	MASA ASURANSI	CARA PEMBAYARAN	KLASIFIKASI
1	Nurfiya Sari	Perempuan	20 -29 Tahun	Belum Kawin	Pegawai Swasta	< 25 Juta	11 -15 Tahun	Triwulan	Tidak Lancar
2	Teuku Fathan	Laki-Laki	30 -40 Tahun	Kawin	Pegawai Swasta	< 25 Juta	11 -15 Tahun	Triwulan	Kurang Lancar
3	Headrawan	Laki-Laki	30 -40 Tahun	Belum Kawin	Wiraswasta	< 25 Juta	11 -15 Tahun	Tahunan	Lancar
4	Hermasayih	Laki-Laki	30 -40 Tahun	Belum Kawin	Wiraswasta	< 25 Juta	11 -15 Tahun	Semester	Lancar
5	Riki Hazaini	Laki-Laki	20 -29 Tahun	Belum Kawin	Pes	< 25 Juta	11 -15 Tahun	Tahunan	Lancar
6	Yuni Zulhina	Perempuan	30 -40 Tahun	Kawin	Pes	25 -50 Juta	5 -10 Tahun	Semester	Kurang Lancar
7	Dewi Hartati	Perempuan	> 40 Tahun	Kawin	Pes	25 -50 Juta	11 -15 Tahun	Semester	Kurang Lancar

Gambar 13. Tampilan Form Laporan

7. KESIMPULAN

Berdasarkan hasil pembahasan maka penulis dapat mengambil beberapa kesimpulan antara lain :

- Sistem klasifikasi data nasabah ini digunakan untuk menampilkan informasi klasifikasi lancar, kurang lancar atau tidak lancarnya calon nasabah dalam membayar premi asuransi dengan menggunakan algoritma *Naive Bayes*.
- Dengan adanya sistem ini maka mempermudah pihak asuransi dalam memperkirakan nasabah yang bergabung, sehingga perusahaan bisa mengambil keputusan untuk menerima atau menolak calon nasabah tersebut.
- Algoritma *Naive Bayes* di dukung oleh ilmu Probabilistik dan ilmu statistika khususnya dalam penggunaan data petunjuk untuk mendukung keputusan pengklasifikasian. Pada algoritma *Naive Bayes*, semua atribut akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting dan setiap atribut saling bebas satu sama lain.
- Variabel penentu yang digunakan dalam penelitian ini adalah jenis kelamin, usia, status, pekerjaan, penghasilan per tahun, masa pembayaran asuransi, dan cara pembayaran asuransi.

DAFTAR PUSTAKA

- [1]Budi, Santoso, 2007, *Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis*, Graha Ilmu, Yogyakarta
- [2]Frieyadi, 2010, *Mudah Belajar Pemograman Database MySql dengan Microsoft Visual Basic 6.0*, Andi, Yogyakarta
- [3]Hermawati, Fajar Astuti, 2013, *Data Mining*, Andi, Yogyakarta

- [4]Jogiyanto, H.M, 2000, *Analisis dan Desain Sistem Informasi*, Andi, Yogyakarta
- [5]Kusumadewi, Sri, 2009, *Klasifikasi Status Gizi Menggunakan Algoritma Naive Bayes Classification*, Jurusan Teknik Informatika, universitas Islam Indonesia
- [6]Mulyanto, Agus, 2009, *Sistem Informasi Konsep dan Aplikasi*, Cetakan I, Pustaka Pelajar, Yogyakarta
- [7]Natalius, Samuel, 2010, *Metode Naive Bayes Classifier dan Penggunaannya Pada Klasifikasi Dokumen*, Program Studi Sistem dan teknologi Informasi, Sekolah Teknik Elektro dan Informatika, Institut Teknologi Bandung
- [8]Prakoso, Djoko, 1994, *Asuransi Indonesia*, Dahara Prize, Semarang
- [9]Rahadian, Hadi, 2004, *Membuat Laporan dengan Crystal Report 8.5 dan Visual Basic 6.0*, Cetakan 2, Elex Media Komputindo, Jakarta
- [10]Rokhmah, Dewi Pyriana, 2011, *Klasifikasi Data Menggunakan Metode K-Nearest Neighbour dan Teorema Bayes*, Program Studi Teknik Informatika, Universitas Brawijaya Malang
- [11]Supardi, Yuniar, 2006, *Microsoft Visual Basic 6.0*, Elex Media Komputindo, Jakarta
- [12]Susanto, Sani, Ph.D, Suryadi, Dedy, 2010, *Pengantar Data Mining: Menggali Pengetahuan dari Bongkahan Data*, Andi, Yogyakarta
- [13][http://www.scribd.com/doc/45017830/Algoritma - Data - Mining-desicion-tree-naive bayes-dll](http://www.scribd.com/doc/45017830/Algoritma-Da-Mining-desicion-tree-naive-bayes-dll) di unduh, 5 Maret 2013, 10:00 Wib
- [14][http://www.scribd.com/doc/55713517/Metode - Bayes](http://www.scribd.com/doc/55713517/Metode-Bayes) di unduh 5 Maret 2013, 10:10 Wib