

## Klasifikasi Data Nasabah Asuransi Dengan Menggunakan Metode *Naive Bayes* (Studi Kasus : PT. Prudential Life Jalan Mt. Haryono Samarinda)

### *Classification of Insurance Data Customers Using Naive Bayes Method* (Case Study : PT. Prudential Life MT. Haryono Street Samarinda)

Dyah Arumatica Novilla<sup>1</sup>, Rito Goejantoro<sup>2</sup>, dan Fidia Deny Tisna Amijaya<sup>3</sup>

<sup>1,2</sup>Laboratorium Statistika Komputasi FMIPA Universitas Mulawarman

<sup>3</sup>Laboratorium Matematika Komputasi FMIPA Universitas Mulawarman

<sup>1</sup>E-mail: [dyahnovilla@yahoo.com](mailto:dyahnovilla@yahoo.com)

#### **Abstract**

*Classification is the logical grouping of objects according to the characteristics of their similarities. Naive Bayes is a method for predicting future opportunities based on past experiences. This study discusses the classification of insurance customer data of PT. Prudential Life Branch of Samarinda in 2017. With the aim to know whether the method of Naive Bayes can classify data of insurance customers of PT. Prudential Life in 2017 using the R program and to determine the accuracy of the results of data testing I and data testing II. As a result, Naive Bayes method can classify data of insurance customers of PT. Prudential Life with 80% accuracy for 25 data testing I and 74.67% for 75 data testing II.*

#### **Pendahuluan**

Dengan kemajuan teknologi informasi, kebutuhan akan informasi yang akurat sangat dibutuhkan dalam kehidupan sehari-hari, sehingga informasi akan menjadi suatu elemen penting dalam perkembangan masyarakat saat ini dan juga waktu yang akan datang. *Data mining* merupakan bidang dari beberapa keilmuan yang menyatukan teknik mesin dari pembelajaran mesin, pengenalan pola, statistik, *database* dan visualisasi untuk penanganan permasalahan pengambilan informasi dari *database* yang besar (Larose, 2005).

*Naive Bayes* merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dengan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada atribut kelas (Patil, 2013). Definisi lain mengatakan *Naive Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh Thomas Bayes, yaitu memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya (Bustami, 2013). Salah satunya pengklasifikasian *naive bayes* yaitu pada bidang asuransi.

Dalam kehidupan ada masalah yang ditakuti manusia adalah kemungkinan kematian yang terjadi terlalu dini. Kematian ini merupakan hal yang pasti, namun masalah waktu atau kapan kematian itu datang adalah suatu hal yang tidak dapat ditentukan oleh manusia. Salah satu cara untuk mengurangi risiko tersebut di atas yaitu dengan mengalihkan atau melimpahkan kepada risiko tersebut pihak atau badan usaha lain. Yang

dimaksud pihak atau badan usaha lain itu ialah suatu lembaga yang menjamin sekiranya timbul suatu peristiwa yang tidak diinginkan, lembaga ini dikenal dengan apa yang disebut asuransi. Salah satu jenis asuransi yang dikenal sekarang ini adalah asuransi jiwa. Asuransi jiwa merupakan alat sosial ekonomi, yang merupakan cara dari sekelompok orang untuk dapat bekerja sama meratakan beban kerugian karena kematian sebelum waktunya dari anggota-anggota kelompok tersebut.

Pada asuransi jiwa yang dipertanggungjawabkan ialah yang disebabkan oleh kematian (*death*). Kematian tersebut mengakibatkan hilangnya pendapatan seseorang atau suatu keluarga tertentu. Risiko yang mungkin timbul pada asuransi jiwa terutama terletak pada unsur waktu (*time*), oleh karena sulit untuk mengetahui kapan seseorang meninggal dunia. Untuk memperkecil risiko tersebut, maka sebaiknya diadakan pertanggungjawaban jiwa (Abbas, 2005).

#### **Data Mining**

*Data mining* merupakan teknologi yang menggabungkan metode analisis tradisional dengan algoritma yang canggih untuk memproses data dengan volume besar. *Data mining* adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam *database*. *Data mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam *database* besar (Turban, 2005).

**Probabilitas Bersyarat**

Probabilitas terjadinya kejadian A dengan syarat bahwa B adalah terjadinya atau akan terjadi disebut *probabilitas bersyarat (contional probability)*, atau biasa di tulis  $P(A/B)$  (Supranto, 2008). Probabilitas bersyarat terjadinya B dengan syarat A telah terjadi dapat dirumuskan sebagai berikut :

$$P(C|F) = \frac{P(F \cap C)}{P(F)} \tag{1}$$

Keterangan :

- P(C|F) : Probabilitas akhir bersyarat suatu kelas C terjadi jika diberikan petunjuk F terjadi
- P(F ∩ C) : Probabilitas awal kelas C dan awal petunjuk F terjadi secara simultan
- P(F) : Probabilitas awal petunjuk F terjadi tanpa memandang kelas apapun

**Klasifikasi**

Secara Etimologi, Klasifikasi berasal dari bahasa Inggris dari kata “*classification*” dan kata ini berasal dari kata “*to classy*” yang berarti menggolongkan dan menempatkan benda-benda di suatu tempat. Istilah ini menunjuk kepada sebuah metode untuk menyusun data secara sistematis atau menurut beberapa aturan atau kaidah yang telah ditetapkan. Selain itu, *Harrolds Librarians Glossary* menyebutkan bahwa klasifikasi adalah pengelompokkan benda secara logis menurut ciri-ciri kesamaannya. Menurut Towa P. Hamakonda dan J.N.B. Tairas (2002) klasifikasi adalah pengelompokan yang sistematis dari obyek, gagasan, buku atau benda-benda lain ke dalam kelas atau golongan tertentu berdasarkan ciri-ciri yang sama.

**Metode Naive Bayes**

*Naive Bayes Classifier* merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang dimasa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Ciri utama dari *naive bayes classifier* ini adalah asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi kejadian.

Menurut David L Olson dan Dursun Delen (2008) menjelaskan *Naive Bayes* untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi objek. Algoritma ini mengasumsikan bahwa atribut objek adalah independen. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari “master” tabel keputusan.

Persamaan dari teorema Bayes adalah :

$$P(C|F) = \frac{P(C).P(F|C)}{P(F)} \tag{2}$$

Keterangan :

- P(C|F) : Probabilitas akhir bersyarat suatu kelas C terjadi jika diberikan petunjuk (atribut) F terjadi
- P(C) : Probabilitas awal kelas C terjadi tanpa memandang petunjuk (atribut) apapun
- P(F|C) : Probabilitas sebuah petunjuk (atribut) F terjadi jika bersyarat suatu kelas C terjadi
- P(F) : Probabilitas awal petunjuk (atribut) F terjadi tanpa memandang kelas apapun

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

$$P(C|F_1, \dots, F_n) = \frac{P(C).P(F_1, \dots, F_n|C)}{P(F_1, \dots, F_n)} \tag{3}$$

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

$$Posterior = \frac{Prior \times Likelihood}{Evidence} \tag{4}$$

Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai-nilai *posterior* kelas lainnya untuk menentukan kelas suatu objek akan diklasifikasikan. Mengklasifikasikan suatu objek dapat ditentukan dengan memilih kelas yang memiliki *posterior* terbesar nilai *evidence* selalu tetap dan merupakan pembagi pada setiap kelasnya sehingga dalam perhitungan *posterior* cukup mengalikan nilai *prior* dengan *likelihood*. Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan  $P(C|F_1, \dots, F_n)$  menggunakan aturan perkalian berikut :

$$\begin{aligned} P(C|F_1, \dots, F_n) &= P(C)P(F_1, \dots, F_n|C) \\ &= P(C)P(F_1|C)P(F_2, \dots, F_n|C, F_1) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3, \dots, F_n|C, F_1, F_2) \\ &= P(C)P(F_1|C)P(F_2|C, F_1)P(F_3|C, F_1, F_2) \dots P(F_n|C, F_1, F_2, F_3, \dots, F_{n-1}) \end{aligned} \tag{5}$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor-faktor syarat yang mempengaruhi nilai probabilitasnya, yang hampir mustahil untuk dianalisis satu per satu. Akibatnya, perhitungan tersebut menjadi sulit untuk dilakukan. Disinilah digunakan asumsi independensi yang sangat tinggi (*naive*), bahwa masing-masing petunjuk ( $F_1, F_2, \dots, F_n$ ) saling bebas (*independent*) satu sama lain. Dengan asumsi tersebut, berlaku suatu kesamaan sebagai berikut :

$$\begin{aligned} P(F_i|F_j) &= \frac{P(F_i \cap F_j)}{P(F_j)} \\ &= \frac{P(F_i)P(F_j)}{P(F_j)} \\ &= P(F_i) \end{aligned} \tag{6}$$

Untuk  $i \neq j$ , sehingga

$$\begin{aligned} P(F_i|C, F_j) &= \frac{P(F_i \cap F_j)}{P(F_j)} \cap C \\ &= \frac{P(F_i) \times P(F_j)}{P(F_j)} \cap C \\ &= P(F_i) \cap C \\ &= P(F_i|C) \end{aligned} \tag{7}$$

Persamaan di atas dapat disimpulkan bahwa asumsi independensi *naive* tersebut membuat syarat peluang menjadi sederhana, sehingga perhitungan menjadi mungkin untuk dilakukan. Selanjutnya, penjabaran  $P(C|F_1, \dots, F_n)$  dapat disederhanakan menjadi,

$$\begin{aligned} P(C|F_1, \dots, F_n) &= P(C)P(F_1|C)P(F_2|C)P(F_3|C) \dots P(F_n|C) \\ &= P(C)\prod_{i=1}^n P(F_i|C) \end{aligned} \tag{8}$$

Persamaan di atas merupakan model dari teorema *naive* Bayes yang selanjutnya akan digunakan dalam proses klasifikasi.

Untuk menghitung probabilitas akhir untuk setiap kelas. Adapun persamaan yang digunakan adalah sebagai berikut :

$$\prod_{i=1}^n P(F_i|C) \tag{9}$$

**Pengukuran Tingkat Akurasi**

Dalam melakukan klasifikasi diharapkan dapat melakukan klasifikasi pada semua objek dengan benar, sehingga dalam klasifikasi juga harus diukur tingkat akurasi. Untuk mendapatkan tingkat akurasi dari hasil prediksi maka digunakan rumus berikut (Sihaan, 2015):

$$\text{Akurasi} = \frac{\text{Jumlah prediksi benar}}{\text{Jumlah data}} \times 100\% \tag{10}$$

**Pengertian Asuransi**

Asuransi berasal dari kata *assurance* atau *insurance*, yang berarti jaminan atau pertanggungan. Hidup penuh dengan ketidakpastian dan manusia selalu berusaha memperkecil atau meminimumkan ketidakpastian tersebut. Misalnya pendidikan yang seseorang tempuh saat ini merupakan salah satu usaha untuk memperkecil ketidakpastian tersebut, dalam usaha mendapatkan jaminan atas pekerjaan yang makin luas (Sembiring, 1986).

**Pengertian Asuransi Jiwa**

Asuransi jiwa adalah usaha kerjasama dari sejumlah orang yang sepakat memikul kesulitan keuangan bila terjadi musibah terhadap salah satu anggotanya. Usaha kerjasama ini dilakukan melalui perusahaan asuransi. Perusahaan yang besar dengan pemegang saham yang banyak akan mudah mengatasi santunan asuransi dari anggota yang meninggal. Setiap orang yang mengansuransikan jiwanya pada suatu perusahaan asuransi berarti sepakat terhadap suatu kontrak tertulis antara dia dan perusahaan. Kontrak tersebut sering disebut polis asuransi (Fajriani, 2013).

**Hasil dan Pembahasan**

Tahap awal yang dilakukan dalam penelitian ini adalah analisis deskriptif dari 100 data, bertujuan untuk menggambarkan karakteristik data nasabah asuransi PT. Prudential Life Samarinda tahun 2017. Karakteristik yang digambarkan pada analisis deskriptif adalah jenis kelamin, umur, status, pekerjaan, penghasilan per tahun, masa pembayaran, cara pembayaran, dan pembayaran premi. Adapun jumlah nasabah setiap atribut pada setiap kelasnya dapat dilihat pada Tabel 1.

**Klasifikasi Naive Bayes**

Dalam proses menghitung klasifikasi *naive* Bayes, terdapat 2 alur yaitu menghitung probabilitas setiap atribut pada setiap kelasnya dan menentukan probabilitas akhir. Probabilitas akhir yang didapat digunakan untuk menentukan termasuk dalam kelas lancar, kurang lancar atau tidak lancar untuk data yang ingin diuji. Kasus yang akan diprediksi kelasnya adalah data pertama percobaan 1 yaitu data seseorang yang berjenis kelamin laki-laki, berumur diatas 50 tahun, berstatus belum kawin, pekerjaan Pegawai Swasta, penghasilan per tahun 11-16 juta, masa pembayaran 1-5 tahun, cara pembayaran bulanan. Selanjutnya akan ditentukan apakah orang tersebut termasuk kelas lancar (artinya orang tersebut membayar premi secara lancar), kelas kurang lancar (artinya orang tersebut membayar premi secara kurang lancar) atau tidak

lancar (artinya orang tersebut membayar premi secara tidak lancar).

Untuk menentukan data yang nantinya akan dianalisis dengan metode *naive* Bayes maka tahap pertama yang dilakukan adalah menentukan nilai probabilitas setiap atribut pada setiap kelasnya dengan memperhatikan apakah data tersebut merupakan data kualitatif atau kuantitatif. Pada penelitian ini, atribut yang termasuk dalam data kualitatif adalah atribut jenis kelamin ( $F_1$ ), umur ( $F_2$ ), status ( $F_3$ ), pekerjaan ( $F_4$ ), penghasilan pertahun ( $F_5$ ), masa pembayaran ( $F_6$ ), cara pembayaran ( $F_7$ ), dan pembayaran premi (C).

Tabel 1. Jumlah nasabah pada atribut jenis kelamin pada setiap kelasnya

| Atribut               | Kelas           | Persentase | Jumlah |
|-----------------------|-----------------|------------|--------|
| Jenis Kelamin         | Laki-laki       | 76%        | 19     |
|                       | Perempuan       | 24%        | 6      |
| Umur                  | 20-35 tahun     | 52%        | 3      |
|                       | 36-50 tahun     | 36%        | 13     |
|                       | Diatas 50 tahun | 12%        | 9      |
|                       | Kawin           | 96%        | 24     |
| Status                | Belum Kawin     | 4%         | 1      |
|                       | Wiraswasta      | 48%        | 4      |
| Pekerjaan             | Pegawai Swasta  | 36%        | 12     |
|                       | PNS             | 66%        | 9      |
|                       | 5-10 juta       | 56%        | 5      |
| Penghasilan Per Tahun | 11-16 juta      | 24%        | 14     |
|                       | Diatas 16 juta  | 20%        | 6      |
| Masa Pembayaran       | 1-5 tahun       | 52%        | 12     |
|                       | 6-10 tahun      | 48%        | 13     |
| Cara Pembayaran       | Bulanan         | 72%        | 18     |
|                       | Tahunan         | 28%        | 7      |
|                       | Lancar          | 68%        | 17     |
| Pembayaran Premi      | Kurang Lancar   | 20%        | 5      |
|                       | Tidak Lancar    | 12%        | 3      |

Adapun nilai probabilitas setiap atribut pada setiap kelasnya untuk data kualitatif sebagai berikut :

Berdasarkan Jenis Kelamin ( $F^1$ ). Pada perhitungan nilai probabilitas atribut jenis kelamin pada kelas lancar terdiri dari 14 nasabah berjenis kelamin laki-laki dan 3 nasabah yang berjenis kelamin perempuan, pada kelas kurang lancar terdiri dari 4 nasabah yang berjenis kelamin laki-laki dan 1 nasabah berjenis kelamin perempuan sedangkan untuk kelas tidak lancar terdiri dari 1 nasabah berjenis kelamin laki-laki dan 2 nasabah berjenis kelamin perempuan. Adapun jumlah nasabah atribut jenis kelamin pada setiap kelasnya dapat dilihat pada Tabel 2.

Nilai probabilitas atribut jenis kelamin pada setiap kelasnya yaitu sebagai berikut:

Probabilitas atribut jenis kelamin laki-laki untuk kelas lancar, dihitung sebagai berikut,

$$P(\text{Jenis Kelamin=L|L}) = \frac{14}{14+3} = 0,8235$$

Jadi peluang nasabah berjenis kelamin laki-laki yang membayar premi secara lancar sebesar 0,8235.

Tabel 2. Jumlah nasabah pada atribut jenis kelamin pada setiap kelasnya

| Jenis Kelamin | Kelas  |               |              |
|---------------|--------|---------------|--------------|
|               | Lancar | Kurang Lancar | Tidak Lancar |
| Laki-laki     | 14     | 4             | 1            |
| Perempuan     | 3      | 1             | 2            |

Probabilitas atribut jenis kelamin perempuan untuk kelas lancar, dihitung sebagai berikut,

$$P(\text{Jenis Kelamin=P|L}) = \frac{3}{14+3} = 0,1765$$

Jadi peluang nasabah berjenis kelamin perempuan yang membayar premi secara lancar sebesar 0,1765.

Probabilitas atribut jenis kelamin laki-laki untuk kelas kurang lancar, dihitung sebagai berikut,

$$P(\text{Jenis Kelamin=L|KL}) = \frac{4}{4+1} = 0,8$$

Jadi peluang nasabah berjenis kelamin laki-laki yang membayar premi secara kurang lancar sebesar 0,8.

Probabilitas atribut jenis kelamin perempuan untuk kelas kurang lancar, dihitung sebagai berikut,

$$P(\text{Jenis Kelamin=P|KL}) = \frac{1}{4+1} = 0,2$$

Jadi peluang nasabah berjenis kelamin perempuan yang membayar premi secara kurang lancar sebesar 0,2.

Probabilitas atribut jenis kelamin laki-laki untuk kelas tidak lancar, dihitung sebagai berikut,

$$P(\text{Jenis Kelamin=L|TL}) = \frac{1}{1+2} = 0,3333$$

Jadi peluang nasabah berjenis kelamin laki-laki yang membayar premi secara tidak lancar sebesar 0,3333.

Probabilitas atribut jenis kelamin perempuan untuk kelas tidak lancar, dihitung sebagai berikut,

$$P(\text{Jenis Kelamin=P|TL}) = \frac{2}{1+2} = 0,6667$$

Jadi peluang nasabah berjenis kelamin perempuan yang membayar premi secara tidak lancar sebesar 0,6667. Adapun nilai probabilitas atribut jenis kelamin pada setiap kelasnya dapat dilihat pada Tabel 3.

Dengan cara yang sama diperoleh probabilitas atribut umur ( $F^2$ ), status ( $F^3$ ) pekerjaan ( $F^4$ ),

penghasilan per tahun ( $F^5$ ), masa pembayaran ( $F^6$ ), cara pembayaran ( $F^7$ ), dan pembayaran premi (C). Adapun jumlah dan nilai probabilitas seluruh atribut pada setiap kelasnya terdapat pada Tabel 4 dan 5.

Tabel 3. Probabilitas Jenis Kelamin pada setiap kelasnya

| Jenis Kelamin | Probabilitas |               |              |
|---------------|--------------|---------------|--------------|
|               | Lancar       | Kurang Lancar | Tidak Lancar |
| Laki-laki     | 0,8235       | 0,8           | 0,3333       |
| Perempuan     | 0,1765       | 0,2           | 0,6667       |

Tabel 4. Jumlah nasabah pada atribut jenis kelamin pada setiap kelasnya

| Atribut               | Kelas           | Jumlah |               |              |
|-----------------------|-----------------|--------|---------------|--------------|
|                       |                 | Lancar | Kurang Lancar | Tidak Lancar |
| Jenis Kelamin         | Laki-laki       | 14     | 4             | 1            |
|                       | Perempuan       | 3      | 1             | 2            |
| Umur                  | 20-35 tahun     | 3      | 0             | 0            |
|                       | 36-50 tahun     | 8      | 3             | 2            |
|                       | diatas 50 tahun | 6      | 2             | 1            |
| Status                | Kawin           | 17     | 5             | 2            |
|                       | Belum Kawiiin   | 0      | 0             | 1            |
| Pekerjaan             | PNS             | 4      | 0             | 0            |
|                       | Wirausaha       | 8      | 1             | 3            |
|                       | Pegawai Swasta  | 5      | 4             | 0            |
| Penghasilan Per Tahun | 5-10 juta       | 4      | 1             | 0            |
|                       | 11-16 juta      | 9      | 4             | 1            |
| Masa Pembayaran       | diatas 16 juta  | 4      | 0             | 2            |
|                       | 1-5 tahun       | 9      | 2             | 1            |
| Cara Pembayaran       | 6-10 tahun      | 8      | 3             | 2            |
|                       | Bulanan         | 12     | 4             | 2            |
| Pembayaran Premi      | Tahunan         | 5      | 1             | 1            |
|                       |                 | 17     | 5             | 3            |

Setelah mengetahui probabilitas setiap atribut pada setiap kelasnya maka selanjutnya ke tahap menentukan probabilitas akhir. Perhitungan probabilitas akhir setiap kelas menggunakan Persamaan (9) sebagai berikut :

$$\begin{aligned} \prod_{i=1}^7 P(F_i | \text{Lancar}) &= P(\text{Laki-laki} | \text{Lancar}) \times P(\text{Usia} = \text{diatas 50 tahun} | \text{Lancar}) \times P(\text{Status} = \text{Kawin} | \text{Lancar}) \times P(\text{Pekerjaan} = \text{Pegawai Swasta} | \text{Lancar}) \times P(\text{Penghasilan} = \text{11-16 juta} | \text{Lancar}) \times P(\text{Masa Asuransi} = \text{1-5 tahun} | \text{Lancar}) \times P(\text{Cara Pembayaran} = \text{Bulanan} | \text{Lancar}) \\ &= 0,8235 \times 0,3530 \times 1 \times 0,2941 \times 0,2353 \times 0,5294 \times 0,7059 \\ &= 0,00751765479629645 \end{aligned}$$

$$\begin{aligned} \prod_{i=1}^7 P(F_i | \text{Kurang Lancar}) &= P(\text{Jenis Kelamin} = \text{Laki-laki} | \text{Kurang Lancar}) \times P(\text{Usia} = \text{diatas 50 tahun} | \text{Kurang Lancar}) \times P(\text{Status} = \text{Kawin} | \text{Kurang Lancar}) \times P(\text{Pekerjaan} = \text{Pegawai Swasta} | \text{Kurang Lancar}) \times P(\text{Penghasilan} = \text{11-16 juta} | \text{Kurang Lancar}) \times P(\text{Masa Asuransi} = \text{1-5 tahun} | \text{Kurang Lancar}) \times P(\text{Cara Pembayaran} = \text{Bulanan} | \text{Kurang Lancar}) \\ &= 0,8 \times 0,4 \times 1 \times 0,8 \times 0 \times 0,4 \times 0,8 \\ &= 0 \end{aligned}$$

$$\begin{aligned} \prod_{i=1}^7 P(F_i | \text{Tidak Lancar}) &= P(\text{Jenis Kelamin} = \text{Laki-laki} | \text{Tidak Lancar}) \times P(\text{Usia} = \text{diatas 50 tahun} | \text{Tidak Lancar}) \times P(\text{Status} = \text{Kawin} | \text{Tidak Lancar}) \times P(\text{Pekerjaan} = \text{Pegawai Swasta} | \text{Tidak Lancar}) \times P(\text{Penghasilan} = \text{11-16 juta} | \text{Tidak Lancar}) \times P(\text{Masa Asuransi} = \text{1-5 tahun} | \text{Tidak Lancar}) \times P(\text{Cara Pembayaran} = \text{Bulanan} | \text{Tidak Lancar}) \\ &= 0,3333 \times 0,3333 \times 0,6667 \times 0 \times 0,6667 \times 0,3333 \times 0,6667 \\ &= 0 \end{aligned}$$

Tabel 5. Probabilitas jenis kelamin pada setiap kelasnya

| Atribut               | Kelas           | Jumlah |               |              |
|-----------------------|-----------------|--------|---------------|--------------|
|                       |                 | Lancar | Kurang Lancar | Tidak Lancar |
| Jenis Kelamin         | Laki-laki       | 0,8235 | 0,8           | 0,3333       |
|                       | Perempuan       | 0,1765 | 0,2           | 0,6667       |
| Umur                  | 20-35 tahun     | 0,1764 | 0             | 0            |
|                       | 36-50 tahun     | 0,4706 | 0,6           | 0,6667       |
|                       | diatas 50 tahun | 0,3530 | 0,4           | 0,3333       |
| Status                | Kawin           | 1      | 1             | 0,6667       |
|                       | Belum Kawiiin   | 0      | 0             | 0,3333       |
| Pekerjaan             | PNS             | 0,2353 | 0             | 0            |
|                       | Wirausaha       | 0,4706 | 0,2           | 1            |
|                       | Pegawai Swasta  | 0,2941 | 0,8           | 0            |
| Penghasilan Per Tahun | 5-10 juta       | 0,2353 | 0,2           | 0            |
|                       | 11-16 juta      | 0,5294 | 0,8           | 0,3333       |
| Masa Pembayaran       | diatas 16 juta  | 0,2353 | 0             | 0,6667       |
|                       | 1-5 tahun       | 0,5294 | 0,4           | 0,3333       |
| Cara Pembayaran       | 6-10 tahun      | 0,4706 | 0,6           | 0,6667       |
|                       | Bulanan         | 0,7059 | 0,8           | 0,6667       |
| Pembayaran Premi      | Tahunan         | 0,2941 | 0,2           | 0,3333       |
|                       |                 | 0,68   | 0,20          | 0,12         |

Selanjutnya, ketiga nilai tersebut digunakan untuk menghitung probabilitas akhir menggunakan Persamaan (8) sebagai berikut :

$$P(\text{Lancar} | F_1, F_2, F_3, F_4, F_5, F_6, F_7) = P(\text{Lancar}) \times P(\text{Jenis Kelamin} = \text{Laki-})$$

$$\begin{aligned}
 & \text{laki} \mid \text{Kurang Lancar} \times P(\text{Status} = \\
 & \text{Kawin} \mid \text{Lancar}) \times P(\text{Pekerjaan} = \\
 & \text{Pegawai Swasta} \mid \text{Lancar}) \\
 & \times P(\text{Penghasilan} = 11-16 \text{ juta} \mid \text{Lancar}) \\
 & \times P(\text{Masa Asuransi} = 1-5 \text{ tahun} \mid \\
 & \text{Lancar}) \times P(\text{Cara Pembayaran} = \\
 & \text{Bulanan} \mid \text{Lancar}) \\
 & = 0,68 \times 0,00751765479629645 \\
 & = 0,00511200526148159
 \end{aligned}$$

$$\begin{aligned}
 P(\text{Kurang Lancar} \mid F_1, F_2, F_3, F_4, F_5, F_6, F_7) = \\
 P(\text{Kurang Lancar}) \times P(\text{Jenis Kelamin} = \\
 \text{Laki-laki} \mid \text{Kurang Lancar}) \times P(\text{Usia} = \\
 \text{diatas 50 tahun} \mid \text{Kurang Lancar}) \\
 \times P(\text{Status} = \text{Kawin} \mid \text{Kurang Lancar}) \\
 \times P(\text{Pekerjaan} = \text{Pegawai Swasta} \mid \\
 \text{Kurang Lancar}) \times P(\text{Penghasilan} = 11- \\
 16 \text{ juta} \mid \text{Kurang Lancar}) \times P(\text{Masa} \\
 \text{Asuransi} = 1-5 \text{ tahun} \mid \text{Kurang Lancar}) \\
 \times P(\text{Cara Pembayaran} = \text{Bulanan} \mid \\
 \text{Kurang Lancar}) \\
 = 0,20 \times 0 \\
 = 0
 \end{aligned}$$

$$\begin{aligned}
 P(\text{Tidak Lancar} \mid F_1, F_2, F_3, F_4, F_5, F_6, F_7) = \\
 P(\text{Tidak Lancar}) \times P(\text{Jenis Kelamin} = \\
 \text{Laki-laki} \mid \text{Tidak Lancar}) \times P(\text{Usia} = \\
 \text{diatas 50 tahun} \mid \text{Tidak Lancar}) \\
 \times P(\text{Status} = \text{Kawin} \mid \text{Tidak Lancar}) \\
 \times P(\text{Pekerjaan} = \text{Pegawai Swasta} \mid \\
 \text{Tidak Lancar}) \times P(\text{Penghasilan} = 11-16 \\
 \text{juta} \mid \text{Tidak Lancar}) \times P(\text{Masa Asuransi} \\
 = 1-5 \text{ tahun} \mid \text{Tidak Lancar}) \times P(\text{Cara} \\
 \text{Pembayaran} = \text{Bulanan} \mid \text{Tidak Lancar}) \\
 = 0,12 \times 0 \\
 = 0
 \end{aligned}$$

Berdasarkan perhitungan dapat diketahui bahwa kelas yang mempunyai nilai probabilitas terbesar adalah kelas lancar, sehingga dapat disimpulkan contoh kasus dengan nilai-nilai atribut jenis kelamin laki-laki, usia diatas 50 tahun, status kawin, pekerjaan Pegawai Swasta, penghasilan 11-16 juta, masa pembayaran 1-5 tahun, dan cara pembayaran bulanan diprediksi masuk dalam kelas lancar yang berarti nasabah tersebut memiliki potensi membayar premi secara lancar.

### Menghitung Tingkat Akurasi Naive Bayes

Dalam prediksi (klasifikasi) dapat melakukan klasifikasi pada semua objek dengan benar. Untuk menghitung tingkat akurasi (ketepatan) menggunakan Persamaan (10). Langkah pertama untuk mendapatkan nilai akurasi yaitu menggunakan *software R* dan hasil klasifikasi data nasabah asuransi. Selanjutnya mencocokkan hasil prediksi pembayaran premi dengan data pembayaran premi, didapat jumlah prediksi yang benar. Percobaan dilakukan sebanyak dua kali dengan data *testing* berturut-turut sebesar 25%

dan 75%. Adapun hasil akurasi dapat dilihat pada Tabel 6.

Tabel 6. Hasil Akurasi Klasifikasi

|                    | Data Testing I | Data Testing II |
|--------------------|----------------|-----------------|
| <b>Jumlah Data</b> | 25             | 75              |
| <b>Prediksi</b>    | 20             | 56              |
| <b>Akurasi</b>     | 80%            | 74,67%          |

Berdasarkan Tabel 6 diketahui bahwa percobaan 1 jumlah data *testing* sebanyak 25 data dengan prediksi yang sama 20 data menunjukkan tingkat ketepatan (akurasi) dalam memprediksi (klasifikasi) sebesar 80% dan pada percobaan 2 jumlah data *testing* sebanyak 75 data dengan prediksi yang sama 56 data menunjukkan tingkat ketepatan (akurasi) dalam memprediksi (klasifikasi) sebesar 74,67%.

### Kesimpulan

Berdasarkan hasil dan pembahasan, maka kesimpulan yang dapat diambil adalah sebagai berikut :

1. Metode *naive Bayes* dapat dipergunakan dalam klasifikasi data nasabah asuransi.
2. Ketepatan hasil prediksi terhadap hasil klasifikasi data nasabah asuransi berdasarkan data yang didapatkan dari (PT. Prudential Life Jalan MT. Haryono Samarinda) pada tahun 2017 dengan menggunakan program *R* yaitu untuk data *testing* I sebanyak 25 data dengan prediksi secara tepat klasifikasi sebanyak 20 data didapat tingkat akurasi sebesar 80% dan pada data *testing* II dengan jumlah 75 data dengan prediksi secara tepat terklasifikasi sebanyak 56 data diperoleh tingkat akurasi sebesar 74,67%.

### Daftar Pustaka

- Abbas, S. (2005). *Asuransi Dan Manajemen Resiko*. Jakarta : PT. Raja Grafindo Persada.
- Bustami. (2013). Penerapan Algoritma *Naive Bayes* Untuk Mengklasifikasi Data Nasabah Asuransi. *Jurnal Penelitian Teknik Informatika*, 3(2), 129-132.
- Fajriani, A. N., Djuwandi., Wilandari. Y. (2013). Perbandingan Nilai Tebus dan Cadangan Premi pada Asuransi Jiwa Kontinu. Eprints UNDIP.
- Larose, D. T. (2005). *Discovering Knowledge in Data : An Introduction to Data Mining*. New Jersey : John Wiley & Sons.
- Olson, David. L. & Delen, D. (2008). *Advanced Data Mining Techniques*, Springer. Amerika Serikat : University of Nebraska.
- Patil, T. R., Sherekar, M. S. (2013). Performance Analysis of *Naive Bayes* and J48 Classification Algorithm for Data Classification, *International Journal of*

- Computer Science and Applications*, 6(2), 256-261.
- Sembiring, R. K. (1986). *Buku Materi Pokok Asuransi 1*. Jakarta : Universitas Terbuka.
- Siahaan, D. (2015). *Aplikasi Classification and Regresion Tree (CART) dan Regresi Logistik Ordinal dalam Bidang Ilmu Pendidikan dengan Studi Kasus Predikat Kelulusan Mahasiswa S1 Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman*. *Jurnal Exponensial*, 7(1), 95-104
- Supranto, J. (2008). *Statistika Teori dan Aplikasi Edisi Tujuh*. Jakarta : Erlangga.
- Towa, H., Tairas, J. N. B. (2002). *Pengantar Klasifikasi Persepuluhan Dewey*. Jakarta : BPK Gunung Mulya.
- Turban, E., Aronson, J. E., & Liang, T. P. (2005). *Decision Support System and Intelligent System*. Yogyakarta : Penerbit Andi Offset.

