

**Penerapan Metode Klasifikasi Multinomial Naive Bayes (Studi Kasus: PT Prudential Life Samarinda Tahun 2019)***Application of Naive Bayes Multinomial Classification Method (Case Study: PT Prudential Life Samarinda in 2019)***Rival Rinaldi<sup>1</sup>, Rito Goejantoro<sup>2</sup>, Syaripuddin<sup>3</sup>**<sup>1,2</sup>Laboratorium Statistika Komputasi FMIPA Universitas Mulawarman<sup>3</sup>Laboratorium Matematika Komputasi FMIPA Universitas MulawarmanE-mail: [rivalemergency@gmail.com](mailto:rivalemergency@gmail.com)**Abstract**

*Life insurance is a risk management service provide payment to policyholders in the event of a disaster that has been stipulated in the agreement. A classification system needs to be done to facilitate the company in making decisions to provide policies to customers. One system that can be used is multinomial Naive Bayes. Multinomial Naive Bayes is a simple probabilistic classification that has more than two groups or categories. An algorithm using Bayes theorem assumes all independent variables. The aim of this study is to obtain an accuracy level of 5 different proportions with the Naive Bayes multinomial method used in insurance customer payment status data. The data used is the customer data of PT. Prudential Life Samarinda in 2019 with the status of current premium payment, substandard and non-current and using 5 independent variables, namely income, age, amount of premium payment, sex and employment. The results of the measurement of classification accuracy using APER status premium payment on insurance customer data of PT. Prudential Life 2019 Naive Bayes multinomial method showed 22,96% misclassification at 50:50 proportion, at the proportion of 60:40 there were 21,43% misclassification, at the proportion of 70:30 there were 19,05% misclassified, at proportions 80:20 had a misclassification of 14,29%, and a proportion of 90:10 has a misclassification of 7,14%.*

*Keywords: classification, Gaussian distribution, multinomial naive Bayes, insurance*

**Pendahuluan**

Dalam kehidupan sehari-hari, manusia tidak dapat melepaskan diri dari ketidakpastian. Baik secara individu ataupun kelompok kita selalu dikelilingi oleh ketidakpastian tersebut. Pertama, ketidakpastian ekonomi (*economic uncertainty*), yaitu kejadian yang timbul sebagai akibat dari perubahan sikap konsumen. Misalnya perubahan selera minta konsumen atau terjadinya perubahan harga, teknologi, dan lain sebagainya. Kedua, Ketidakpastian yang disebabkan oleh alam (*uncertainty of nature*), misalnya kebakaran, badai topan, banjir, dan lain-lain. Ketiga, Ketidakpastian yang disebabkan oleh perilaku manusia (*human uncertainty*), misalnya peperangan, pencurian, perampokan, dan pembunuhan. (Salim, 2007). Salah satu cara untuk mengahlikan ketidakpastian tersebut dengan mengikuti asuransi.

Asuransi adalah suatu perjanjian dimana pihak yang menjamin berjanji kepada pihak yang dijamin, untuk menerima sejumlah uang premi sebagai pengganti kerugian, yang mungkin akan diderita oleh yang dijamin, di mana akibat dari suatu peristiwa yang belum jelas. Salah satu jenis asuransi yang paling diminati masyarakat ialah 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 (Arumatica, 2019).

Perusahaan asuransi jiwa memiliki kriteria-kriteria tertentu dimana nasabah tersebut dapat memiliki polis asuransi jiwa. Kriteria calon pemilik polis asuransi jiwa dapat dilihat dari umur, pendapatan, status pekerjaan, masa pembayaran, besar pembayaran premi dan lain-lain (Prakoso, 1994). Berdasarkan kriteria tersebut, perusahaan dapat menentukan apakah calon nasabah tersebut termasuk kategori pembayaran premi lancar, kurang lancar atau tidak lancar. Hal ini dilakukan untuk mengantisipasi adanya nasabah yang mengalami pembayaran premi tidak lancar. Pengelompokan nasabah dalam kategori pembayaran premi lancar, kurang lancar atau tidak lancar dengan kriteria tertentu dapat dilakukan menggunakan analisis statistika yaitu metode klasifikasi.

Klasifikasi merupakan proses penemuan model atau fungsi yang menjelaskan atau membedakan konsep pada kelompok data dengan tujuan untuk dapat memperkirakan kelompok dari suatu obyek yang labelnya tidak diketahui (Prasetyo, 2014). Terdapat beberapa macam metode pengklasifikasian yang dapat digunakan yaitu *decision tree*, *neural network*, *k-nearest neighbor*, analisis *diskriminan* dan lain sebagainya. Selain metode analisis tersebut, terdapat metode *Naive Bayes*.

Menurut Bustami (2013), *Naive Bayes* merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes yaitu memprediksi peluang dimasa depan berdasarkan pengalaman di masa sebelumnya. *Naive Bayes* merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan peluang dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma dengan menggunakan teorema Bayes mengasumsikan semua variabel saling bebas atau tidak saling ketergantungan.

Berdasarkan uraian di atas, penulis tertarik untuk melakukan penelitian dengan judul Penerapan Metode Klasifikasi Multinomial *Naive Bayes*, Studi Kasus PT. Prudential *Life* Samarinda Tahun 2019.

**Data Mining**

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 yang tersimpan dalam *database* besar (Turban, 2005).

**Peluang Bersyarat**

Peluang terjadinya *A* dengan syarat bahwa *B* telah terjadi disebut peluang bersyarat (*conditional probability*), atau biasa di tulis  $P(A|B)$  (Supranto, 2008). Peluang bersyarat terjadinya *A* dengan syarat *B* telah terjadi dapat dirumuskan sebagai berikut:

$$P(A|B) = \frac{P(B \cap A)}{P(B)} \tag{1}$$

Dengan:

$P(A|B)$  : peluang terjadinya *A* dengan syarat *B* telah terjadi

$P(B \cap A)$  : peluang awal *A* dengan petunjuk *B* terjadi simultan

$P(B)$  : peluang terjadinya *B*

**Klasifikasi**

Klasifikasi merupakan suatu pekerjaan menilai obyek data untuk memasukkannya ke dalam kelompok tertentu dari sejumlah kelompok yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu pembangunan model sebagai prototipe untuk disimpan sebagai memori dan penggunaan model tersebut untuk melakukan pengenalan/klasifikasi/prediksi pada suatu obyek data lain agar diketahui di kelompok mana obyek data tersebut dalam model yang sudah disimpnnya (Prasetyo, 2014).

**Metode Multinomial Naive Bayes**

Multinomial *Naive Bayes* merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes. Teorema Bayes dikemukakan oleh ilmuan Inggris Thomas Bayes pada abad ke-18, yaitu memprediksi peluang dimasa depan berdasarkan pengalaman di masa sebelumnya. Ciri utama dari klasifikasi *Naive Bayes* adalah asumsi independensi yang kuat (Prasetyo, 2014).

Multinomial *Naive Bayes* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam *database* dengan data yang besar. Teori keputusan Bayes adalah pendekatan statistik *fundamental* dalam pengenalan pola (*pattern recognition*). Pendekatan ini didasarkan pada kuantifikasi *trade-off* antara berbagai keputusan klasifikasi dengan menggunakan peluang yang ditimbulkan dalam keputusan-keputusan tersebut (Santosa, 2017).

Persamaan dari teorema Bayes pada umumnya adalah sebagai berikut:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \tag{2}$$

Keterangan:

$P(A|B)$  = peluang terjadinya *A* dengan syarat *B* telah terjadi

$P(A)$  = peluang terjadinya *A*

$P(B|A)$  = peluang terjadinya *B* dengan syarat *A* telah terjadi

$P(B)$  = peluang terjadinya *B*

Pada proses klasifikasi diperlukan sejumlah petunjuk yaitu variabel-variabel untuk menentukan kelompok mana yang tepat bagi obyek yang dianalisis tersebut sehingga teorema Bayes disesuaikan sebagai berikut:

$$P(Y|X_1, X_2, \dots, X_p) = \frac{P(Y)P(X_1, X_2, \dots, X_p|Y)}{P(X_1, X_2, \dots, X_p)} \tag{3}$$

Rumus diatas dapat juga dituliskan secara sederhana sebagai berikut:

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

Nilai dari *posterior* tersebut nantinya akan dibandingkan dengan nilai-nilai *posterior* kelompok lainnya untuk menentukan kelompok suatu obyek akan diklasifikasikan. Mengklasifikasikan suatu obyek dapat ditentukan dengan memilih kelompok yang memiliki *posterior* terbesar, nilai *evidence* selalu tetap untuk setiap kelompok pada satu sampel yaitu bernilai 1 dan merupakan pembagi pada setiap kelompok sehingga dalam perhitungan *posterior* hanya cukup mengalikan nilai *prior* dengan *likelihood*. Nilai *prior* yang merupakan peluang munculnya kelompok *Y* sebelum masuknya

obyek dapat dihitung menggunakan persamaan sebagai berikut:

$$P(Y_i) = \frac{n_i}{N} \quad (5)$$

Penjabaran lebih lanjut rumus Bayes tersebut dilakukan dengan menjabarkan  $P(Y|X_1, X_2, \dots, X_p)$  menggunakan aturan perkalian sebagai berikut:

$$\begin{aligned} P(Y|X_1, X_2, \dots, X_p) &= P(Y)P(X_1|Y)P(X_2, X_3, \dots, X_p|Y, X_1) \\ &= P(Y)P(X_1|Y)P(X_2, X_3, \dots, X_p|Y, X_1, X_2) \\ &= P(Y)P(X_1|Y)P(X_2|Y, X_1)P(X_3, X_4, \dots, X_p|Y, X_1, X_2) \\ &= P(Y)P(X_1|Y)P(X_2|Y, X_1)P(X_3|Y, X_1, X_2)P(X_4, X_5, \dots, X_p|Y, X_1, X_2, X_3) \\ &= P(Y)P(X_1|Y)P(X_2|Y, X_1)P(X_3|Y, X_1, X_2)P(X_p|Y, X_1, X_2, X_3, \dots, X_{p-1}) \end{aligned} \quad (6)$$

Dapat dilihat bahwa hasil penjabaran tersebut menyebabkan semakin banyak dan semakin kompleksnya faktor-faktor syarat yang mempengaruhi nilai peluang yang hampir mustahil untuk dianalisis satu persatu. Kompleksnya faktor-faktor syarat yang mempengaruhi nilai peluang menyebabkan perhitungan tersebut menjadi sulit untuk dilakukan, maka digunakan asumsi independensi yang sangat tinggi (*naive*), bahwa masing-masing petunjuk  $(X_1, X_2, \dots, X_p)$  saling bebas (*independent*) satu sama lain sehingga berlaku suatu kesamaan sebagai berikut:

$$\begin{aligned} P(X_a|X_b) &= \frac{P(X_a \cap X_b)}{P(X_b)} \\ &= \frac{P(X_a)P(X_b)}{P(X_b)} \\ &= P(X_a) \end{aligned} \quad (7)$$

Pada Persamaan (7) dapat disimpulkan bahwa asumsi independensi *naive* tersebut membuat syarat peluang menjadi sederhana sehingga perhitungan menjadi mungkin dilakukan. Selanjutnya penjabaran  $P(Y|X_1, X_2, \dots, X_p)$  dapat disederhanakan menjadi:

$$\begin{aligned} P(Y|X_1, X_2, \dots, X_p) &= P(Y)P(X_1|Y) \dots P(X_p|Y) \\ &= P(Y) \prod_{g=1}^p P(X_g|Y) \end{aligned} \quad (8)$$

Persamaan (8) merupakan model dari teorema *Naive Bayes* yang selanjutnya akan digunakan dalam proses klasifikasi. Pada umumnya teorema *Naive Bayes* mudah dihitung untuk nilai pengamatan variabel bebas bertipe kategorik, namun untuk nilai pengamatan variabel bebas dengan tipe numerik (non kategorik) ada perlakuan khusus sebelum diproses menggunakan *naive Bayes* yaitu dengan cara sebagai berikut:

- a Melakukan diskritisasi pada setiap nilai pengamatan variabel bebas kontinu dan mengganti nilai pengamatan tersebut dengan nilai interval diskrit. Pendekatan ini dilakukan dengan mentransformasi ke dalam skala ordinal.
- b Mengasumsikan bentuk tertentu dari distribusi peluang untuk nilai pengamatan kontinu dan memperkirakan parameter distribusi dengan data *training*. Distribusi Gauss biasanya dipilih untuk merepresentasikan peluang bersyarat pada nilai pengamatan kontinu pada sebuah kelompok  $P(Y|X_g)$ , sedangkan distribusi Gauss dikarakteristikan dengan dua parameter yaitu rata-rata ( $\mu$ ) dan variansi ( $\sigma^2$ ) untuk setiap kelompok  $y_i$ , peluang bersyarat kelompok  $y_i$  untuk nilai pengamatan  $X_g$  adalah sebagai berikut:

$$P(X = x_g | Y = y_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\left(\frac{x_g - \mu_i}{2\sigma_i^2}\right)} \quad (9)$$

$, i = 1, 2, \dots, k \text{ dan } g = 1, 2, \dots, p$

dengan:

- $x_g$  = nilai variabel bebas ke- $g$
- $y_i$  = kelompok ke- $i$
- $\sigma_i^2$  = variansi kelompok ke- $i$
- $\mu_i$  = rata-rata kelompok ke- $i$

### Pengukuran Tingkat Akurasi

Pengukuran tingkat akurasi klasifikasi multinomial *naive Bayes* dapat dilakukan dengan menghitung kesalahan klasifikasi. Ukuran yang dapat digunakan adalah *Apparent Error Rate* (APER). Nilai APER menyatakan fraksi atau proporsi sampel yang salah diklasifikasikan oleh fungsi klasifikasi. Semakin banyak kesalahan klasifikasi akan berdampak pada hasil keakurasian metode pengklasifikasian. Metode pengklasifikasian yang baik akan menghasilkan probabilitas kesalahan klasifikasi yang semakin kecil (Johnson dan Wichern, 2002).

Menghitung nilai APER dapat dilakukan melalui tabel klasifikasi sebagai berikut

**Tabel 1.** Tabel Klasifikasi

Actual Membership	Predicted membership		
	$\hat{c}_1$	$\hat{c}_2$	$\hat{c}_3$
$c_1$	$f_{11}$	$f_{12}$	$f_{13}$
$c_2$	$f_{21}$	$f_{22}$	$f_{23}$
$c_3$	$f_{31}$	$f_{32}$	$f_{33}$

$$APER = \frac{\text{jumlah obyek yang salah diklasifikasi}}{\text{jumlah prediksi yang dilakukan}} \times 100\%$$

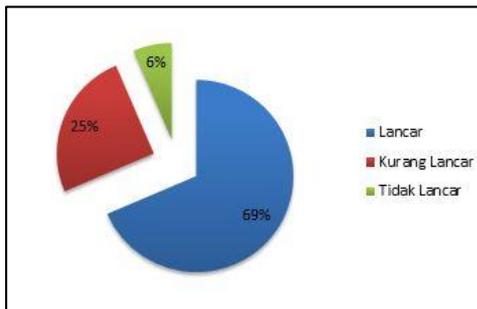
$$= \frac{f_{12} + f_{13} + f_{21} + f_{23} + f_{31} + f_{32}}{f_{11} + f_{12} + f_{13} + f_{21} + f_{22} + f_{23} + f_{31} + f_{32} + f_{33}} \times 100\% \quad (10)$$

**Asuransi**

Asuransi berasal dari kata *assurance* atau *insurance*, yang berarti jaminan atau pertanggungan. Hidup penuh dengan ketidakpastian dan jaminan selalu berusaha memperkecil atau meminimumkan ketidakpastian tersebut (Sembiring, 1986).

**Deskripsi data penelitian**

Deskriptif data dilakukan untuk mengetahui karakteristik data nasabah asuransi PT. Prudential Life Samarinda Tahun 2019. Karakteristik data nasabah asuransi berdasarkan status pembayaran premi ditampilkan pada diagram *pie* pada Gambar 1.



**Gambar 1.** Karakteristik data nasabah asuransi berdasarkan status pembayaran premi

Pada Gambar 1 menunjukkan persentase status pembayaran premi di PT. Prudential Life Samarinda Tahun 2019. Dari 140 data nasabah asuransi dapat diketahui bahwa 69% atau sebanyak 96 nasabah memiliki status pembayaran premi dengan kategori lancar, 25% atau sebanyak 35 nasabah memiliki status pembayaran premi dengan kategori kurang lancar dan 6% atau sebanyak 9 nasabah memiliki status pembayaran premi dengan kategori tidak lancar. Dari data tersebut dapat diketahui bahwa nasabah yang memiliki status pembayaran premi dengan kategori lancar memiliki persentase yang lebih besar dibandingkan dengan nasabah yang memiliki status pembayaran premi dengan kategori kurang lancar dan kategori tidak lancar.

**Klasifikasi Multinomial Naive Bayes**

Pada proses klasifikasi metode multinomial *naive Bayes* terdapat tiga alur yaitu menghitung probabilitas *prior*, menghitung probabilitas setiap variabel bebas pada setiap kelompok dan menghitung perkalian probabilitas *prior* dan probabilitas variabel bebas pada setiap kelompok. Adapun data yang digunakan untuk perhitungan klasifikasi adalah menggunakan lima proporsi. Pembahasan pada subbab ini menggunakan proporsi 80:20 yaitu data *training* 80% sebanyak 112 data dan data *testing* 20% sebanyak 28 data.

Adapun tahapan-tahapan klasifikasi metode multinomial *naive Bayes* dalam pengklasifikasian status pembayaran premi data nasabah asuransi sebagai berikut:

**1. Menghitung probabilitas awal (*prior*) setiap kelompok**

Tahapan klasifikasi metode multinomial *naive Bayes* yang pertama yaitu menghitung probabilitas awal (*prior*) pada kedua kelompok menggunakan data *training*. Adapun nilai probabilitas awal (*prior*) setiap kelompok menggunakan Persamaan (5) sebagai berikut:

a. Kelompok pertama (nasabah asuransi dengan status pembayaran premi lancar)

Pada 112 data *training* yang ada, terdapat 79 nasabah asuransi yang memiliki status pembayaran premi lancar sehingga nilai probabilitas awal (*prior*) kelompok pertama sebagai berikut:

$$P(Y_1) = \frac{79}{112} = 0,7054$$

b. Kelompok kedua (nasabah asuransi dengan status pembayaran premi kurang lancar)

Pada 112 data *training* yang ada, terdapat 26 nasabah asuransi yang memiliki status pembayaran premi kurang lancar sehingga nilai probabilitas awal (*prior*) kelompok kedua yaitu sebagai berikut:

$$P(Y_2) = \frac{26}{112} = 0,2321$$

c. Kelompok ketiga (nasabah asuransi dengan status pembayaran premi tidak lancar)

Pada 112 data *training* yang ada, terdapat 7 nasabah asuransi yang memiliki status pembayaran premi tidak lancar sehingga nilai probabilitas awal (*prior*) kelompok kedua yaitu sebagai berikut:

$$P(Y_2) = \frac{7}{112} = 0,0625$$

**2. Menghitung probabilitas setiap variabel bebas pada setiap kelompok**

Perhitungan nilai probabilitas setiap variabel bebas pada ketiga kelompok dilakukan pada data *testing* berdasarkan data *training* untuk . Pada data *testing* pertama diketahui nasabah asuransi memiliki pendapatan sebesar Rp.9.000.000,00, usia 39 tahun, besar pembayaran premi sebesar Rp.4.000.000,00 jenis kelamin laki-laki dan pekerjaan pegawai swasta.

Kelima nilai variabel bebas pada data *testing* kemudian di nilai probabilitas setiap variabel bebas pada ketiga kelompok. Perhitungan nilai probabilitas setiap variabel bebas berdatan numerik pada ketiga kelompok mengasumsikan variabel bebas berdistribusi Gauss menggunakan Persamaan (9) sebagai berikut:

a. Pendapatan ( $X_1$ )

Adapun nilai probabilitas variabel pendapatan data *testing* pertama pada setiap kelompok sebagai berikut:

Probabilitas variabel pendapatan untuk status pembayaran premi lancar dihitung sebagai berikut

$$P(X_1 = 9.000.000|Y_1) = \frac{1}{\sqrt{(2)(3,14)(3,056312 \times 10^{13})}} \times e^{-\left(\frac{(9.000.000)-(8.529.114)^2}{2(3,056312 \times 10^{13})}\right)} = 7,1913 \times 10^{-8}$$

Sehingga probabilitas variabel pendapatan sebesar Rp. 9.000.000,00 dengan syarat status pembayaran premi lancar sebesar  $7,1913 \times 10^{-8}$ .

Probabilitas variabel pendapatan untuk status pembayaran premi kurang lancar dihitung sebagai berikut,

$$P(X_1 = 9.000.000|Y_2) = \frac{1}{\sqrt{(2)(3,14)(1,119785 \times 10^{13})}} \times e^{-\left(\frac{(9.000.000)-(8.776.923)^2}{2(1,119785 \times 10^{13})}\right)} = 1,1898 \times 10^{-7}$$

Sehingga probabilitas variabel pendapatan sebesar Rp. 9.000.000,00 dengan syarat status pembayaran premi kurang lancar sebesar  $1,1898 \times 10^{-7}$ .

Probabilitas variabel pendapatan untuk status pembayaran premi tidak lancar dihitung sebagai berikut,

$$P(X_1 = 9.000.000|Y_3) = \frac{1}{\sqrt{(2)(3,14)(2,72381 \times 10^{13})}} \times e^{-\left(\frac{(9.000.000)-(13.714.286)^2}{2(2,72381 \times 10^{13})}\right)} = 5,0845 \times 10^{-8}$$

Sehingga probabilitas variabel pendapatan sebesar Rp. 9.000.000,00 dengan syarat status pembayaran premi tidak lancar sebesar  $5,0845 \times 10^{-8}$ .

#### b. Usia ( $X_2$ )

Adapun nilai probabilitas variabel usia data *testing* pertama pada setiap kelompok sebagai berikut:

Probabilitas variabel usia 39 tahun untuk status pembayaran premi lancar dihitung sebagai berikut,

$$P(X_2 = 39|Y_1) = \frac{1}{\sqrt{(2)(3,14)(37,54625)}} \times e^{-\left(\frac{(39)-(35,37975)^2}{2(37,54625)}\right)} = 0,0546937$$

Sehingga probabilitas variabel usia 39 tahun dengan syarat status pembayaran premi lancar sebesar 0,0546937.

Probabilitas variabel usia 39 tahun untuk status pembayaran premi kurang lancar dihitung sebagai berikut,

$$P(X_2 = 39|Y_2) = \frac{1}{\sqrt{(2)(3,14)(47,41538)}}$$

$$\times e^{-\left(\frac{(39)-(35,84615)^2}{2(47,41538)}\right)} = 0,0521804$$

Sehingga probabilitas variabel usia 39 tahun dengan syarat status pembayaran premi kurang lancar sebesar 0,0521804.

Probabilitas variabel usia 39 tahun untuk status pembayaran premi tidak lancar dihitung sebagai berikut,

$$P(X_2 = 39|Y_3) = \frac{1}{\sqrt{(2)(3,14)(12,66667)}} \times e^{-\left(\frac{(39)-(32)^2}{2(12,66667)}\right)} = 0,01620586$$

Sehingga probabilitas variabel usia 39 tahun dengan syarat status pembayaran premi tidak lancar sebesar 0,01620586.

#### c. Besar Pembayaran Premi ( $X_3$ )

Adapun nilai probabilitas variabel besar pembayaran premi data *testing* pertama pada setiap kelompok sebagai berikut:

Probabilitas variabel besar pembayaran premi untuk status pembayaran premi lancar dihitung sebagai berikut,

$$P(X_3 = 4.000.000|Y_1) = \frac{1}{\sqrt{(2)(3,14)(190.099.537.021)}} \times e^{-\left(\frac{(4.000.000)-(766.244,7)^2}{2(190.099.537.021)}\right)} = 1,0387 \times 10^{-18}$$

Sehingga probabilitas variabel besar pembayaran premi sebesar Rp. 4.000.000,00 dengan syarat status pembayaran premi lancar sebesar  $1,0387 \times 10^{-18}$ .

Probabilitas variabel besar pembayaran premi untuk status pembayaran premi kurang lancar dihitung sebagai berikut,

$$P(X_3 = 4.000.000|Y_2) = \frac{1}{\sqrt{(2)(3,14)(1,862435 \times 10^{12})}} \times e^{-\left(\frac{(4.000.000)-(2.271.154)^2}{2(1,862435 \times 10^{12})}\right)} = 1,31 \times 10^{-7}$$

Sehingga probabilitas variabel besar pembayaran premi sebesar Rp. 4.000.000,00 dengan syarat status pembayaran premi kurang lancar sebesar  $1,31 \times 10^{-7}$ .

Probabilitas variabel besar pembayaran premi untuk status pembayaran premi tidak lancar dihitung sebagai berikut,

$$P(X_3 = 4.000.000|Y_3) = \frac{1}{\sqrt{(2)(3,14)(5,571429 \times 10^{12})}} \times e^{-\left(\frac{(4.000.000)-(6.285.714)^2}{2(5,571429 \times 10^{12})}\right)} = 1,06 \times 10^{-7}$$

Sehingga probabilitas variabel besar pembayaran premi sebesar Rp. 4.000.000,00 dengan syarat status pembayaran premi tidak lancar sebesar  $1,06 \times 10^{-7}$ .

d. Jenis Kelamin ( $X_4$ )

Adapun nilai probabilitas variabel jenis kelamin data *testing* pertama pada setiap kelompok sebagai berikut:

Probabilitas variabel jenis kelamin laki laki untuk status pembayaran premi lancar dihitung sebagai berikut

$$P(X_4 = \text{Laki-Laki}|Y_1) = \frac{51}{79} = 0,645569$$

Sehingga probabilitas variabel jenis kelamin laki-laki untuk status pembayaran premi lancar adalah sebesar 0,645569.

Probabilitas variabel jenis kelamin laki laki untuk status pembayaran premi kurang lancar dihitung sebagai berikut:

$$P(X_4 = \text{Laki-Laki}|Y_2) = \frac{14}{26} = 0,538461$$

Sehingga probabilitas variabel jenis kelamin laki-laki untuk status pembayaran premi kurang lancar adalah sebesar 0,538461.

Probabilitas variabel jenis kelamin laki laki untuk status pembayaran premi tidak lancar dihitung sebagai berikut:

$$P(X_4 = \text{Laki-Laki}|Y_3) = \frac{3}{7} = 0,428571$$

Sehingga probabilitas variabel jenis kelamin laki-laki untuk status pembayaran premi tidak lancar adalah sebesar 0,428571.

e. Pekerjaan ( $X_5$ )

Adapun nilai probabilitas variabel jenis kelamin data *testing* pertama pada setiap kelompok sebagai berikut:

Probabilitas variabel pekerjaan pegawai swasta untuk status pembayaran premi lancar dihitung sebagai berikut:

$$P(X_5 = \text{Pegawai Swasta}|Y_1) = \frac{33}{79} = 0,41772$$

Sehingga probabilitas variabel pekerjaan pegawai swasta untuk status pembayaran premi lancar adalah sebesar 0,41772.

Probabilitas variabel pekerjaan pegawai swasta untuk status pembayaran premi kurang lancar dihitung sebagai berikut:

$$P(X_5 = \text{Pegawai Swasta}|Y_2) = \frac{13}{26} = 0,5$$

Sehingga probabilitas variabel pekerjaan pegawai swasta untuk status pembayaran premi lancar adalah sebesar 0,5.

Probabilitas variabel pekerjaan pegawai swasta untuk status pembayaran premi tidak lancar dihitung sebagai berikut:

$$P(X_5 = \text{Pegawai Swasta}|Y_3) = \frac{3}{7} = 0,428571$$

Sehingga probabilitas variabel pekerjaan pegawai swasta untuk status pembayaran premi lancar adalah sebesar 0,5.

3. Menghitung perkalian probabilitas *prior* dan probabilitas setiap variabel bebas pada setiap kelompok (*posterior*)

Tahap ketiga adalah menghitung perkalian probabilitas *prior* dan probabilitas setiap variabel bebas pada setiap kelompok. Perhitungan perkalian probabilitas *prior* dan probabilitas setiap variabel bebas pada setiap kelompok dilakukan pada data *testing*. Pada data *testing* pertama diketahui nasabah asuransi memiliki pendapatan sebesar Rp. 9.000.000,00, usia 39 tahun, besar pembayaran premi sebesar Rp. 4.000.000,00 jenis kelamin laki-laki dan pekerjaan pegawai swasta. Adapun perhitungan perkalian probabilitas *prior* dan probabilitas setiap variabel bebas pada kedua kelompok menggunakan Persamaan (6) sebagai berikut:

a. Kelompok pertama (nasabah asuransi dengan status pembayaran premi lancar)

$$\begin{aligned} P(Y_1|X_1, X_2, X_3, X_4, X_5) &= P(Y_1) \times P(X_1|Y_1) \times P(X_2|Y_1) \times P(X_3|Y_1) \times P(X_4|Y_1) \\ &\quad \times P(X_5|Y_1) \\ &= P(Y_1) \times P(X_1 = 9.000.000|Y_1) \times P(X_2 = 39|Y_1) \\ &\quad \times P(X_3 = 4.000.000|Y_1) \times P(X_4 = \text{Laki-laki}|Y_1) \\ &\quad \times P(X_5 = \text{Pegawai Swasta}|Y_1) \\ &= (0,7054)(7,1913 \times 10^{-8})(0,0546937)(1,0387 \\ &\quad \times 10^{18})(0,645569)(0,41772) \\ &= 7,772 \times 10^{-28} \end{aligned}$$

Sehingga dapat diketahui bahwa probabilitas nasabah asuransi mengalami status pembayaran premi lancar dengan syarat pendapatan sebesar Rp. 9.000.000,00, usia 39 tahun, besar pembayaran premi sebesar Rp. 4.000.000,00, jenis kelamin laki-laki, dan pekerjaan pegawai swasta sebesar  $7,772 \times 10^{-28}$ .

b. Kelompok kedua (nasabah asuransi dengan status pembayaran premi kurang lancar)

$$\begin{aligned} P(Y_2|X_1, X_2, X_3, X_4, X_5) &= P(Y_2) \times P(X_1|Y_2) \times P(X_2|Y_2) \times P(X_3|Y_2) \\ &\quad \times P(X_4|Y_2) \times P(X_5|Y_2) \\ &= P(Y_2) \times P(X_1 = 9.000.000|Y_2) \times P(X_2 = 39|Y_2) \\ &\quad \times P(X_3 = 4.000.000|Y_2) \times P(X_4 = \text{Laki-laki}|Y_2) \\ &\quad \times P(X_5 = \text{Pegawai Swasta}|Y_2) \\ &= (0,2321)(1,1898 \times 10^{-7})(0,0521804) \\ &\quad (1,31 \times 10^{-7})(0,538461)(0,5) \\ &= 5,0859 \times 10^{-17} \end{aligned}$$

Sehingga dapat diketahui bahwa probabilitas nasabah asuransi mengalami status pembayaran premi kurang lancar dengan syarat pendapatan

sebesar Rp. 9.000.000,00, usia 39 tahun, besar pembayaran premi sebesar Rp. 4.000.000,00, jenis kelamin laki-laki, dan pekerjaan pegawai swasta sebesar  $5,0859 \times 10^{-17}$ .

c. Kelompok ketiga (nasabah asuransi dengan status pembayaran premi tidak lancar)

$$\begin{aligned} &P(Y_3|X_1, X_2, X_3, X_4, X_5) \\ &= P(Y_3) \times P(X_1|Y_3) \times P(X_2|Y_3) \times P(X_3|Y_3) \\ &\quad \times P(X_4|Y_3) \times P(X_5|Y_3) \\ &= P(Y_3) \times P(X_1 = 9.000.000|Y_3) \times P(X_2 = 39|Y_3) \\ &\quad \times P(X_3 = 4.000.000|Y_3) \times P(X_4 = \text{Laki-laki}|Y_3) \\ &\quad \times P(X_5 = \text{Pegawai Swasta}|Y_3) \\ &= (0,0625)(5,0845 \times 10^{-8})(0,01620586) \\ &\quad (1,06 \times 10^{-7})(0,428571)(0,428571) \\ &= 1,0006 \times 10^{-18} \end{aligned}$$

Sehingga dapat diketahui bahwa probabilitas nasabah asuransi mengalami status pembayaran premi tidak lancar dengan syarat pendapatan sebesar Rp. 9.000.000,00, usia 39 tahun, besar pembayaran premi sebesar Rp. 4.000.000,00, jenis kelamin laki-laki, dan pekerjaan pegawai swasta sebesar  $1,0006 \times 10^{-18}$ .

Berdasarkan perhitungan perkalian probabilitas awal dan probabilitas setiap variabel bebas (*posterior*) pada ketiga kelompok dapat diketahui bahwa kelompok yang memiliki *posterior* terbesar adalah nasabah asuransi dengan kelompok status pembayaran premi kurang lancar sebesar  $5,0859 \times 10^{-17}$  dibandingkan *posterior* nasabah asuransi dengan kelompok status pembayaran premi lancar sebesar  $7,772 \times 10^{-28}$  dan status pembayaran premi tidak lancar sebesar  $1,0006 \times 10^{-18}$ , sehingga dapat disimpulkan data *testing* pertama yaitu obyek dengan nilai-nilai variabel pendapatan sebesar Rp. 9.000.000,00, usia 39 tahun, besar pembayaran premi sebesar Rp. 4.000.000,00 jenis kelamin laki-laki dan pekerjaan pegawai swasta diklasifikasikan masuk ke dalam kelompok kedua yaitu nasabah asuransi dengan status pembayaran premi kurang lancar.

Setelah nilai *posterior* pada ketiga kelompok diketahui dilakukan penentuan nilai maksimum dari *posterior* pada setiap kelompok untuk mengetahui obyek akan diklasifikasikan dalam kelompok pertama yaitu nasabah asuransi dengan status pembayaran premi lancar, kelompok kedua yaitu nasabah asuransi dengan status pembayaran premi kurang lancar dan kelompok ketiga yaitu nasabah asuransi dengan status pembayaran premi tidak lancar.

Hasil klasifikasi menggunakan metode multinomial *naive* Bayes dengan 28 data *testing* yang ada diperoleh hasil terdapat 16 nasabah memiliki status pembayaran premi lancar, 8 nasabah memiliki status pembayaran premi kurang lancar dan tidak terdapat nasabah memiliki status pembayaran premi tidak lancar.

### Menghitung Tingkat Akurasi

Pada proses klasifikasi diharapkan melakukan klasifikasi pada semua obyek dengan benar. Pada proses klasifikasi menggunakan metode multinomial *naive* Bayes, data nasabah asuransi sebanyak 140 data dibagi menjadi data *training* dan data *testing* sesuai proporsi yang ditentukan. Ada 5 proporsi yang digunakan pada penelitian ini yaitu proporsi 50:50, 60:40, 70:30, 80:20, dan 90:10. Pengukuran tingkat akurasi metode multinomial *naive* Bayes dilakukan dengan menghitung peluang kesalahan klasifikasi. Semakin kecil kesalahan klasifikasi menunjukkan bahwa semakin baik hasil klasifikasi yang didapatkan.

Pada proses klasifikasi menggunakan metode multinomial *naive* Bayes, jumlah obyek yang tepat dan salah diklasifikasikan untuk masing-masing kelompok dapat dilihat pada Tabel 2. Tanda (\*) pada angka-angka menyatakan jumlah obyek kelompok tertentu yang salah diklasifikasikan dengan menggunakan metode *naive* Bayes.

**Tabel 2.** Hasil Klasifikasi

Klasifikasi Awal Status Pembayaran Premi	Prediksi Klasifikasi			Total
	Lancar	Kurang Lancar	Tidak Lancar	
Lancar	16	1*	0*	17
Kurang Lancar	1*	8	0*	9
Tidak Lancar	0*	2*	0	2
Total	17	11	0	28

$$\begin{aligned} APER &= \frac{1 + 0 + 1 + 0 + 0 + 2}{28} \times 100\% \\ &= \frac{4}{28} \times 100\% \\ &= 14,29\% \end{aligned}$$

Jadi, nilai APER (*laju error*) pada hasil klasifikasi multinomial *naive* Bayes menggunakan proporsi 80:20 adalah sebesar 14,29%. Adapun nilai APER pada hasil klasifikasi menggunakan metode multinomial *naive* Bayes untuk semua percobaan proporsi data disajikan pada Tabel 3. berikut:

**Tabel 3.** Nilai APER Pada Metode Multinomial *Naive* Bayes untuk Setiap Proporsi

Proporsi	APER
50 : 50	22,96%
60 : 40	21,43%

**Tabel 3.** Nilai APER Pada Metode Multinomial *Naive Bayes* untuk Setiap Proporsi (Lanjutan)

Proporsi	APER
70 : 30	19,05%
80 : 20	14,29%
90 : 10	7,14%

**Kesimpulan**

Berdasarkan hasil analisis dan pembahasan, maka dapat diperoleh kesimpulan yaitu hasil pengukuran tingkat akurasi klasifikasi menggunakan APER status pembayaran premi pada data nasabah asuransi PT. Prudential *Life* Tahun 2019 dengan metode multinomial *naive Bayes* menunjukkan kesalahan klasifikasi 22,96% pada proporsi 50:50, pada proporsi 60:40 terdapat 21,43% kesalahan klasifikasi, pada proporsi 70:30 terdapat 19,05% kesalahan klasifikasi, pada proporsi 80:20 memiliki kesalahan klasifikasi sebesar 14,29%, dan proporsi 90:10 memiliki kesalahan klasifikasi sebesar 7,14% . Proporsi 90:10 merupakan proporsi dengan hasil klasifikasi yang paling baik karena memiliki kesalahan klasifikasi yang paling kecil dari kelima proporsi.

**Daftar Pustaka**

Arumatica, D. N., Goejantoro, R., dan Amijaya, F. D. T. (2019). Klasifikasi Data Nasabah Asuransi dengan Menggunakan Metode *Naive Bayes* (Studi Kasus: PT. Prudential

*Life* Jalan MT. Haryono Samarinda). *Jurnal Eksponensial*, vol 10 (2), 95-101.  
 Bustami. (2013). Penerapan Algoritma *Naive Bayes* Untuk Mengklasifikasikan Data Nasabah Asuransi. *Jurnal Penelitian Teknik Informatika*, vol 3(2), 129-132.  
 Johnson, R. A., and Winchern, D. W. (2002). *Applied Multivariate Statistical Analysis*. New Jersey: Prentice Hall.  
 Prakoso, D. (1994). *Asuransi Indonesia*. Semarang: Dahara Prize.  
 Prasetyo, E. (2014). *Data Mining: Konsep dan Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Offset.  
 Salim, A. (2007). *Asuransi dan Manajemen Risiko*. Jakarta: Raja Grafindo Persada.  
 Santosa, B. (2007). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.  
 Sembiring, R. K. (1986). *Buku Materi Pokok Asuransi I Modul 1-5*. Jakarta: Universitas Terbuka.  
 Supranto, J. (2008). *Statistika Teori dan Aplikasi Edisi Tujuh*. Jakarta: Erlangga  
 Turban, E., Aronson, J. E., and Liang, T.P. (2005). *Decision Support System and Intelligent System*. Yogyakarta: Andi Offset.