

PERBANDINGAN KLASIFIKASI ANALISIS DISKRIMINAN FISHER DAN METODE NAIVE BAYES

Aliftha Ainurrochmah¹, Memi Nor Hayati², Andi M. Ade Satriya³

^{1,2,3}Universitas Mulawarman

e-mail: ¹aliftha.aaa@gmail.com, ²meminorhayati@fmipa.unmul.ac.id, ³andiade06@yahoo.com

Abstrak

Klasifikasi adalah suatu teknik pembentukan model dari data yang telah diketahui kelompok klasifikasinya. Model tersebut digunakan untuk mengklasifikasikan obyek baru. Analisis diskriminan Fisher merupakan teknik multivariat untuk memisahkan obyek-obyek dalam kelompok yang berbeda. Naive Bayes merupakan pengklasifikasian berdasarkan probabilitas dan teorema Bayes dengan asumsi independensi yang kuat. Penelitian ini bertujuan melakukan perbandingan tingkat akurasi klasifikasi dengan analisis diskriminan Fisher dan metode naive Bayes pada status pembayaran premi nasabah asuransi. Data yang digunakan memiliki 4 variabel bebas yaitu pendapatan, usia, masa pembayaran premi dan besar pembayaran premi. Hasil kesalahan akurasi dengan menggunakan nilai APER (*Apparent Rate Error*) menunjukkan bahwa metode naive Bayes memiliki tingkat akurasi yang lebih tinggi sebesar 15,38% daripada analisis diskriminan Fisher sebesar 46,15% dalam menganalisis status pembayaran premi nasabah asuransi.

Kata kunci: klasifikasi, analisis diskriminan Fisher, naive Bayes, asuransi.

Abstract

Classification is a technique to form a model of data that is already known to its classification group. The model was formed will be used to classify new objects. Fisher discriminant analysis is multivariate technique to separate objects in different groups. Naive Bayes is a classification technique based on probability and Bayes theorem with assumption of independence. This research has a goal to compare the level of classification accuracy between Fisher's discriminant analysis and Naive Bayes method on the insurance premium payment status customer. The data used four independent variables that is income, age, premium payment period and premium payment amount. The results of misclassification using the APER (Apparent Rate Error) indicate that the naive Bayes method has a higher level of accuracy is 15,38% than Fisher's discriminant analysis is 46,15% on the insurance premium payment status customer.

Keywords: classification, Fisher discriminant analysis, naive Bayes, insurance.

PENDAHULUAN

Perkembangan yang pesat di bidang ilmu dan teknologi menuntut adanya kemampuan manusia dalam mempertimbangkan segala kemungkinan sebelum mengambil keputusan dan tindakan. Salah satu pertimbangan pengambilan keputusan yang penting adalah persoalan manajemen pemilihan nasabah bagi perusahaan asuransi. Pemilihan nasabah bagi perusahaan asuransi sangat penting karena akan berdampak pada kelancaran administrasi perusahaan tersebut.

Salah satu jenis asuransi yang paling diminati masyarakat ialah asuransi jiwa. Menurut Pasal 1 Undang-Undang Nomor 40 Tahun 2014, asuransi jiwa adalah jasa penanggulangan risiko yang memberikan pembayaran kepada pemegang polis, tertanggung atau pihak lain yang berhak ketika yang tertanggung meninggal dunia atau tetap hidup dalam jangka waktu tertentu yang telah diatur dalam perjanjian. Fungsi dari asuransi jiwa sendiri yaitu memberikan jaminan tersedianya sejumlah uang tunai atau memberi penghasilan bagi keluarga (ahli waris) tertanggung jika yang tertanggung tersebut meninggal dunia.

Pada proses administrasi pengelolaan asuransi jiwa, pihak perusahaan asuransi jiwa pun tidak ingin mengalami kerugian dari nasabah. 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 atau tidak lancar. Hal ini dilakukan untuk mengantisipasi adanya nasabah yang mengalami pembayaran premi tidak lancar. Pengelompokan nasabah dalam kategori pembayaran premi lancar atau tidak lancar dengan kriteria tertentu dapat dilakukan menggunakan analisis statistika yaitu dengan metode klasifikasi.

Terdapat beberapa macam metode pengklasifikasian yang dapat digunakan yaitu *decision tree*, *neural network*, *k-*

nearest neighbor, *naive Bayes* dan lain sebagainya. Selain metode analisis tersebut, terdapat metode analisis lain dimana pada proses pengklasifikasiannya memperhatikan asumsi tertentu yaitu analisis diskriminan (Artha, 2016).

Berdasarkan uraian tersebut, penulis tertarik melakukan penelitian ilmiah dengan judul “Perbandingan Klasifikasi Analisis Diskriminan Fisher dan Metode *Naive Bayes*”.

METODE

1. Tinjauan Referensi

Penelitian mengenai analisis diskriminan Fisher dan metode *naive Bayes* telah banyak dilakukan oleh peneliti. Hal ini terkait dengan pentingnya mengetahui klasifikasi suatu obyek dengan beberapa karakteristik tertentu dan obyek tersebut sebelumnya belum diketahui klasifikasi. Banyaknya permintaan untuk menjadi nasabah asuransi menuntut adanya pertimbangan perusahaan asuransi dalam memilih nasabah. Pemilihan nasabah bagi perusahaan asuransi sangat penting karena akan berdampak pada kelancaran administrasi perusahaan tersebut. Kriteria calon pemilik polis asuransi jiwa dapat dilihat dari umur, pendapatan, status pekerjaan, masa pembayaran, besar pembayaran premi dan lain-lain.

Menurut Widarjono (2015) analisis multivariat berasal dari kata *multy* dan *variate* yang berarti alat analisis lebih dari dua variabel. Analisis multivariat berhubungan dengan semua teknik statistik yang secara simultan menganalisis sejumlah pengukuran pada suatu objek. Analisis multivariat sering digunakan untuk memecahkan permasalahan dalam penelitian yang sifatnya sangat kompleks.

Menurut Towa dan Tairas (2002) klasifikasi adalah pengelompokan yang sistematis dari obyek, gagasan, buku dan benda-benda lain ke dalam kelompok atau golongan tertentu berdasarkan ciri-ciri yang sama. Klasifikasi merupakan suatu pekerjaan menilai obyek data untuk memasukkannya ke dalam kelompok tertentu dari sejumlah kelompok yang tersedia. (Prasetyo, 2014).

Tujuan dari penelitian ini adalah mengetahui bagaimana tingkat akurasi dan hasil klasifikasi status pembayaran premi nasabah asuransi PT. Prudential Life Tahun 2018 dengan status pembayaran premi lancar atau tidak lancar menggunakan analisis diskriminan Fisher dan metode *naive* Bayes. Menurut Triandaru (2008), pada dasarnya asuransi jiwa dipengaruhi oleh pendapatan, usia, masa pembayaran premi, dan besar pembayaran premi. Oleh karena itu pada penelitian ini variabel bebas yang digunakan adalah pendapatan, usia, masa pembayaran premi, dan besar pembayaran premi.

2. Metode Analisis

Data Training dan Data Testing

Data *training* digunakan oleh algoritma klasifikasi untuk membentuk sebuah model *classifier*. Model ini merupakan representasi pengetahuan yang akan digunakan untuk memprediksi kelompok dari data baru yang belum pernah ada. Data *testing* digunakan untuk mengukur sejauh mana *classifier* berhasil melakukan klasifikasi dengan benar. Data yang ada pada data *testing* seharusnya tidak boleh ada pada data *training* sehingga dapat diketahui apakah model *classifier* sudah tepat atau belum dalam melakukan klasifikasi (Witten dan Eibe, 2011).

Jumlah data *training* dan data *testing* dapat dihitung sebagai berikut:

$$\text{Jumlah data } training = \text{proporsi data } training \times N \quad (1)$$

$$\text{Jumlah data } testing = N - \text{jumlah data } training \quad (2)$$

dengan:

N = jumlah seluruh data (data *training* + data *testing*)

Analisis Diskriminan

Analisis diskriminan adalah teknik multivariat untuk memisahkan obyek-obyek dalam himpunan yang berbeda dan mengalokasikan obyek baru dalam himpunan-himpunan tersebut. Analisis diskriminan merupakan teknik menganalisis data yang mempunyai satu variabel terikat merupakan kategori

sedangkan variabel bebas merupakan metrik.

Asumsi pada analisis diskriminan menurut Supranto (2004) adalah variabel bebas berdistribusi normal (*multivariate normality*) dan matriks varian kovarian antar kelompok adalah sama.

1. Distribusi Normal Multivariat

Menurut Johnson dan Wichern (2002), pengujian data berdistribusi normal multivariat dapat menggunakan sebuah metode yang didasarkan pada ukuran jarak kuadrat tergeneralisasi yaitu jarak mahalanobis dengan membuat Q-Q *plot*. Pengujian data berdistribusi multivariat normal dengan membuat Q-Q *plot* antara jarak mahalanobis d_j^2 dengan kuantil chi-kuadrat $q_j = \chi^2_{\left(p, \frac{(j-0,5)}{n}\right)}$. Apabila *plot*

d_j^2 dan q_j membentuk garis lurus berarti data pengamatan memenuhi asumsi distribusi normal multivariat. Selain menggunakan Q-Q *plot*, pengujian distribusi normal multivariat dapat dilakukan dengan pengujian Kolmogorov Smirnov dengan hipotesis sebagai berikut:

H_0 : Data berdistribusi normal multivariat

H_1 : Data tidak berdistribusi normal multivariat

menggunakan taraf signifikansi yaitu α dengan statistik uji sebagai berikut:

$$D_{hitung} = \text{Sup} |S(d_j^2) - F_0(d_j^2)| \quad (3)$$

dengan:

$S(d_j^2)$ = fungsi distribusi kumulatif dari

$$d_j^2$$

$F_0(d_j^2)$ = fungsi distribusi kumulatif

Chi-Kuadrat

Kriteria pengujian distribusi normal multivariat ini yaitu H_0 ditolak jika nilai D_{hitung} lebih besar dari nilai D pada tabel Kolmogorov Smirnov atau jika p -value kurang dari α .

2. Kesamaan Matriks Varians Kovarian

Menurut Johnson dan Wichern (2002), uji yang digunakan untuk mengetahui kesamaan matriks varian

kovarian adalah uji *Box's M*. Prosedur pada uji *Box's M* yaitu menggunakan pendekatan sebaran *Chi-Square* dengan derajat bebas $\frac{1}{2}(k-1)k(p+1)$ dengan p adalah banyaknya variabel bebas dan k adalah banyaknya kelompok. Hipotesis matriks varian kovarian sebagai berikut:

Hipotesis :

$H_0 : \Sigma_1 = \Sigma_2 = \dots = \Sigma_k$ (matriks varian kovarian antar kelompok sama)

H_1 : sedikitnya ada sepasang matriks varian kovarian yang berbeda antar kelompok

Statistik uji :

$$C = (1-u)M$$

$$= (1-u) \sum_{i=1}^k (n_i - 1) \ln |\mathbf{S}| - \sum_{i=1}^k (n_i - 1) \ln |\mathbf{S}_i| \quad (4)$$

Jika nilai $C \geq \chi^2_{\alpha, \frac{1}{2}(k-1)p(p+1)}$ artinya

tolak H_0 pada taraf nyata α yang berarti matriks varian kovarian dari k kelompok adalah berbeda

Analisis Diskriminan Fisher

Analisis diskriminan yang menggunakan pendekatan Fisher merupakan analisis diskriminan memiliki matriks varian kovarian antar kelompok yang sama (Johnson dan Wichern, 2007).

Fungsi diskriminan Fisher dapat ditunjukkan sebagai berikut:

$$\hat{y} = \mathbf{b}^T \mathbf{X} \quad (5)$$

(1×1) (1×p) (p×1)

Pada klasifikasi menggunakan analisis diskriminan Fisher, jika fungsi diskriminan Fisher $\hat{y} = \mathbf{b}^T \mathbf{X} \geq \hat{m}$ maka obyek pengamatan akan masuk ke dalam kelompok pertama. Jika $\hat{y} = \mathbf{b}^T \mathbf{X} < \hat{m}$ maka obyek pengamatan akan masuk ke dalam kelompok kedua, dengan nilai \hat{m} sebagai berikut:

$$\hat{m} = \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \mathbf{S}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2) \quad (6)$$

Metode Naive Bayes

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:

$$P(A|B) = \frac{P(A)P(B|A)}{P(B)} \quad (7)$$

dengan:

$P(A|B)$ = 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

Asumsi independensi *naive* membuat syarat peluang menjadi sederhana sehingga perhitungan menjadi mungkin dilakukan. Selanjutnya penjabaran $P(Y | X_1, X_2, \dots, X_p)$ dapat disederhanakan menjadi:

$$P(Y | X_1, X_2, \dots, X_p) = P(Y) \times P(X_1 | Y) \times P(X_2 | Y) \times P(X_3 | Y) \times \dots \times P(X_p | Y) \quad (8)$$

$$= P(Y) \prod_{g=1}^p P(X_g | Y)$$

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.

b Mengasumsikan bentuk tertentu dari distribusi peluang untuk nilai pengamatan kontinu dan memperkirakan parameter

distribusi dengan data *training*. Distribusi Gaussian biasanya dipilih untuk memrepresentasikan peluang bersyarat pada nilai pengamatan kontinu pada sebuah kelompok $P(Y | X_g)$ adalah sebagai berikut:

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

dengan:

$i = 1, 2, \dots, k$

$g = 1, 2, \dots, p$

$x_g =$ nilai variabel bebas ke- g

$y_i =$ kelompok ke- i

$\sigma_i^2 =$ variansi kelompok ke- i

$\mu_i =$ rata-rata kelompok ke- i

3. Pengukuran Tingkat Akurasi

Evaluasi fungsi klasifikasi dilakukan dengan cara membuat tabulasi antara *actual group* dan *predicted group* yang diperoleh dari fungsi klasifikasi dari analisis diskriminan dan metode *naive Bayes* yang selanjutnya dilakukan perhitungan proporsi pengamatan yang salah klasifikasi. Diharapkan proporsi pengamatan yang salah diklasifikasikan tersebut bisa sekecil mungkin. Menghitung nilai APER dapat dilakukan melalui tabel klasifikasi sebagai berikut:

keterangan :

$f_{11} =$ jumlah obyek dari c_1 tepat

Tabel 1. Tabel Klasifikasi

Actual Membership	Predicted membership		Total
	\hat{c}_1	\hat{c}_2	
c_1	f_{11}	f_{12}	A
c_2	f_{21}	f_{22}	B
Total	C	D	E

diklasifikasikan sebagai \hat{c}_1

$f_{12} =$ jumlah obyek dari c_1 salah diklasifikasikan sebagai \hat{c}_2

$f_{21} =$ jumlah obyek dari c_2 salah

diklasifikasikan sebagai \hat{c}_1

$f_{22} =$ jumlah obyek dari c_2 tepat

diklasifikasikan sebagai \hat{c}_2

dengan perhitungan nilai APER sebagai berikut:

$$\begin{aligned} \text{APER} &= \frac{\text{jumlah obyek salah klasifikasi}}{\text{jumlah prediksi dilakukan}} \times 100\% \\ &= \frac{f_{12} + f_{21}}{f_{11} + f_{12} + f_{21} + f_{22}} \times 100\% \end{aligned} \quad (10)$$

HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini merupakan data nasabah asuransi PT. Prudential *Life* Jalan MT Haryono Samarinda pada Tahun 2018. Adapun variabel terikat adalah status pembayaran premi nasabah asuransi dengan kategori lancar dan tidak lancar dengan variabel bebas adalah pendapatan, usia, masa pembayaran premi dan besar pembayaran premi.

1. Membagi Data *Training* dan Data *Testing*

Data nasabah asuransi dibagi menjadi data *training* dan data *testing* dengan proporsi pembagian 80:20. Adapun perhitungan pembagian data *training* dan data *testing* dengan menggunakan Persamaan (1) dan Persamaan (2) sebagai berikut:

$$\begin{aligned} \text{Jumlah data } \textit{training} &= \text{Proporsi data } \textit{training} \times N \\ &= 80\% \times 130 \\ &= 104 \end{aligned}$$

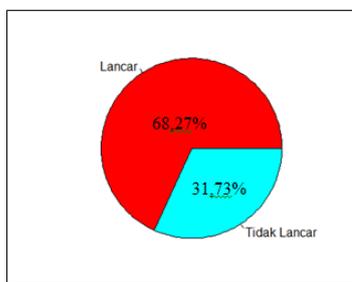
$$\begin{aligned} \text{Jumlah data } \textit{testing} &= N - \text{jumlah data } \textit{training} \\ &= 130 - 104 \\ &= 26 \end{aligned}$$

Berdasarkan perhitungan diperoleh hasil bahwa 104 data pertama hasil randomisasi data untuk masing-masing variabel penelitian berfungsi sebagai data *training*, sedangkan 26 data terakhir untuk masing-masing variabel penelitian berfungsi sebagai data *testing* yang nantinya digunakan untuk melakukan

proses klasifikasi dan pengukuran tingkat akurasi klasifikasi analisis diskriminan Fisher dan metode *naive* Bayes.

2. Analisis Statistik Deskriptif

Analisis statistika deskriptif ini dilakukan untuk mengetahui gambaran umum data nasabah asuransi PT. Prudential *Life* Tahun 2018. Data yang digunakan pada proses analisis statistika deskriptif adalah data *training*. Gambaran umum pada mengenai status pembayaran premi nasabah asuransi dengan status pembayaran premi lancar dan tidak lancar dapat dilihat pada Gambar 1.



Gambar 1. Persentase status pembayaran premi

Berdasarkan Gambar 1 menunjukkan persentase status pembayaran premi nasabah asuransi PT. Prudential *Life* Tahun 2018. Dari 104 data nasabah asuransi dapat diketahui bahwa 68,27% atau sebanyak 71 orang nasabah asuransi memiliki status pembayaran premi lancar dan 31,73% atau sebanyak 33 orang nasabah asuransi memiliki status pembayaran premi tidak lancar. Dari data tersebut dapat diketahui bahwa nasabah asuransi dengan status pembayaran premi lancar lebih banyak dibandingkan dengan nasabah asuransi dengan status pembayaran premi tidak lancar.

3. Klasifikasi Status Pembayaran Analisis Diskriminan Fisher

Analisis diskriminan adalah teknik multivariat untuk memisahkan obyek-obyek dalam kelompok berbeda dan mengalokasikan obyek baru dalam kelompok-kelompok tersebut. Adapun tahapan-tahapan dalam proses analisis diskriminan sebagai berikut:

a. Pengujian Asumsi Normal Multivariat

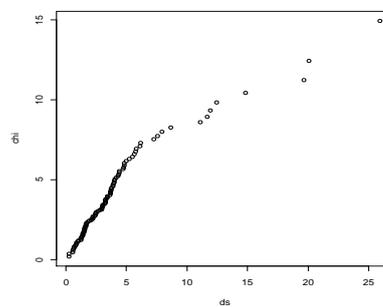
Uji normal multivariat digunakan untuk mengetahui apakah data status pembayaran premi nasabah asuransi dengan keempat variabel bebas berdistribusi normal multivariat atau tidak. Adapun pengujian hipotesis normal multivariat pada Persamaan (3) dengan taraf signifikansi sebesar 5% dengan daerah kritis H_0 ditolak jika nilai $D_{hitung} < D_{(104;0,05)}$ pada tabel Kolmogorov Smirnov atau jika $p - value < 0,05$.

Hasil pengujian menggunakan uji Kolmogorov Smirnov dapat dilihat pada Tabel 2.

Berdasarkan Tabel 2 diperoleh keputusan bahwa H_0 gagal ditolak yang artinya data nasabah asuransi dengan Tabel 2. Uji Kolmogorov Smirnov

D_{hitung}	D_{tabel}	$p - value$	Keputusan
0,125	0,133	0,3908	H_0 gagal ditolak

keempat variabel bebas berdistribusi normal multivariat.



Gambar 2. Q-Q plot uji normal multivariat

Pada Gambar 2 terlihat bahwa *plot* cenderung membentuk garis lurus, sehingga dapat disimpulkan bahwa data nasabah asuransi berdistribusi normal multivariat.

b. Pengujian Kesamaan Varians Kovarians

Uji kesamaan matriks varian kovarian antar kelompok 1 yaitu nasabah asuransi dengan status pembayaran premi lancar dan kelompok 2 yaitu nasabah asuransi dengan status pembayaran premi tidak lancar dilakukan dengan menggunakan uji *Box's M*. Adapun pengujian kesamaan matriks varian kovarian pada Persamaan (4) dengan daerah Kritis H_0 ditolak jika nilai

$C \geq \chi^2_{0,05;\frac{1}{2}(2-1)4(4+1)}$ atau jika $p\text{-value} < 0,05$.

Hasil pengujian menggunakan uji *Box's M* dapat dilihat pada Tabel 3.

Berdasarkan Tabel 3 diperoleh keputusan bahwa H_0 gagal ditolak yang artinya matriks varian kovarian antar Tabel 3. Uji *Box' M*

C	$\chi^2_{0,05;10}$	Keputusan
15,96	18,31	H_0 gagal ditolak

kelompok data nasabah asuransi adalah sama.

c. Analisis Diskriminan Fisher

Klasifikasi status pembayaran premi data nasabah asuransi menggunakan analisis diskriminan Fisher dilakukan dengan mengetahui fungsi diskriminan Fisher menggunakan data *training* terlebih dahulu. Jika fungsi diskriminan Fisher sudah diketahui, kemudian dapat dilakukan proses pengklasifikasian status pembayaran premi data nasabah asuransi menggunakan data *testing*.

Nilai koefisien dari masing-masing variabel sebagai berikut:

$$\mathbf{b} = \begin{bmatrix} 6,4458 \times 10^{-7} \\ -1,8449 \times 10^{-1} \\ 1,7133 \times 10^{-1} \\ -2,2186 \times 10^{-6} \end{bmatrix}$$

Nilai batasan kelompok \hat{m} sebesar $\hat{m} = -6,2615$ sehingga diperoleh nilai fungsi diskriminan Fisher sebagai berikut:

$$y = 6,4458 \times 10^{-7} X_1 - 1,8449 \times 10^{-1} X_2 + 1,7133 \times 10^{-1} X_3 - 2,2186 \times 10^{-6} X_4$$

Pada data *testing* pertama diperoleh nilai fungsi diskriminan Fisher sebagai berikut:

$$\begin{aligned} \hat{y}_1 &= 6,4458 \times 10^{-7} (6.000.000) \\ &\quad - 1,8449 \times 10^{-1} (40) \\ &\quad + 1,7133 \times 10^{-1} (6) \\ &\quad - 2,2186 \times 10^{-6} (3.000.000) \\ \hat{y}_1 &= -9,14 \end{aligned}$$

Sehingga dapat diketahui nilai fungsi diskriminan Fisher pada data *testing* pertama sebesar -9,14. Berdasarkan aturan

analisis diskriminan Fisher, jika $\hat{y} = \mathbf{b}^T \mathbf{X} \geq -6,2615$ maka data obyek pengamatan akan masuk ke dalam kelompok pertama yaitu nasabah asuransi memiliki status pembayaran premi lancar dan jika $\hat{y} = \mathbf{b}^T \mathbf{X} < -6,2615$ maka obyek pengamatan akan masuk ke dalam kelompok kedua yaitu nasabah asuransi memiliki status pembayaran premi tidak lancar. Pada data *testing* pertama dapat diketahui nilai $\hat{y}_1 = -9,14 < \hat{m} = -6,2615$ maka obyek pengamatan data *testing* pertama dapat diklasifikasikan ke dalam kelompok kedua yaitu nasabah asuransi memiliki status pembayaran premi tidak lancar.

4. Klasifikasi Status Pembayaran Premi Metode *Naive Bayes*

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

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

Pada 104 data *training* yang ada, terdapat 71 nasabah asuransi yang memiliki status pembayaran premi lancar sehingga nilai probabilitas awal (*prior*) kelompok pertama sebagai berikut:

$$P(Y_1) = \frac{71}{104} = 0,6827$$

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

Pada 104 data *training* yang ada, terdapat 33 nasabah asuransi yang memiliki status pembayaran premi tidak lancar sehingga nilai probabilitas awal (*prior*) kelompok kedua yaitu sebagai berikut:

$$P(Y_2) = \frac{33}{104} = 0,3173$$

Perhitungan nilai probabilitas setiap variabel bebas pada kedua kelompok dilakukan pada data *testing* berdasarkan rata-rata dan variansi dari data *training*. Pada data *testing* pertama diketahui nasabah asuransi memiliki pendapatan (X_1) sebesar

Rp.6.000.000,00, usia (X_2) 40 tahun, masa pembayaran premi (X_3) selama 6 tahun dan besar pembayaran premi (X_4) sebesar Rp.3.000.000,00. Keempat nilai variabel bebas pada data *testing* kemudian di nilai probabilitas setiap variabel bebas pada kedua kelompok. Perhitungan nilai probabilitas setiap variabel bebas pada kedua kelompok mengasumsikan variabel bebas berdistribusi Gaussian menggunakan Persamaan (9) sebagai berikut:

a. Pendapatan (X_1)

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

Berdasarkan data *training* pembayaran premi dengan status pembayaran premi lancar pada variabel pendapatan diketahui memiliki nilai rata-rata sebesar Rp. 10.239.437 dan nilai variansi sebesar $1,8813 \times 10^{13}$ maka probabilitas variabel pendapatan sebesar Rp.6.000.000,00 dengan syarat status pembayaran premi lancar dapat dihitung sebagai berikut,

$$P(X_1 = 6.000.000 | Y_1) = \frac{1}{\sqrt{(2)(3,14)(1,8813 \times 10^{13})}} \times \exp\left(-\frac{(6.000.000)-(10.239.437)^2}{2(1,8813 \times 10^{13})}\right) = 5,7061 \times 10^{-8}$$

Sehingga probabilitas variabel pendapatan sebesar Rp.6.000.000,00 dengan syarat status pembayaran premi lancar sebesar $5,7061 \times 10^{-8}$.

Berdasarkan data *training* pembayaran premi dengan status pembayaran premi tidak lancar pada variabel pendapatan diketahui memiliki nilai rata-rata sebesar Rp. 10.696.970 dan nilai variansi sebesar $1,5968 \times 10^{13}$ maka probabilitas variabel pendapatan sebesar Rp.6.000.000,00 dengan syarat status pembayaran premi tidak lancar dapat dihitung sebagai berikut,

$$P(X_1 = 6.000.000 | Y_2) = \frac{1}{\sqrt{(2)(3,14)(1,5968 \times 10^{13})}} \times \exp\left(-\frac{(6.000.000)-(10.696.970)^2}{2(1,5968 \times 10^{13})}\right) = 5,0047 \times 10^{-8}$$

Sehingga probabilitas variabel pendapatan sebesar Rp.6.000.000,00 dengan syarat status pembayaran premi tidak lancar sebesar $5,0047 \times 10^{-8}$.

b. Usia (X_2)

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

Berdasarkan data *training* pembayaran premi dengan status pembayaran premi lancar pada variabel usia diketahui memiliki nilai rata-rata sebesar 37,8591 dan nilai variansi sebesar 72,6942, maka probabilitas variabel usia 40 tahun dengan syarat status pembayaran premi lancar dihitung sebagai berikut,

$$P(X_2 = 40 | Y_1) = \frac{1}{\sqrt{(2)(3,14)(72,6942)}} \times \exp\left(-\frac{(40)-(37,8591)^2}{2(72,6942)}\right) = 0,0453$$

Sehingga probabilitas variabel usia 40 tahun dengan syarat status pembayaran premi lancar sebesar 0,0453.

Berdasarkan data *training* pembayaran premi dengan status pembayaran premi tidak lancar pada variabel usia diketahui memiliki nilai rata-rata usia sebesar 52,1515 dan nilai variansi sebesar 52,4451, maka probabilitas variabel usia 40 tahun dengan syarat status pembayaran premi tidak lancar dihitung sebagai berikut,

$$P(X_2 = 40 | Y_2) = \frac{1}{\sqrt{(2)(3,14)(52,4451)}} \times \exp\left(-\frac{(40)-(52,1515)^2}{2(52,4451)}\right) = 1,3483 \times 10^{-2}$$

Sehingga probabilitas variabel usia 40 tahun dengan syarat status pembayaran premi tidak lancar sebesar $1,3483 \times 10^{-2}$.

c. Masa Pembayaran Premi (X_3)

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

Berdasarkan data *training* pembayaran premi dengan status pembayaran premi lancar pada variabel masa pembayaran premi diketahui memiliki nilai rata-rata sebesar 8,5352 dan nilai variansi sebesar 5,0237 maka probabilitas variabel masa pembayaran premi 6 tahun dengan syarat status pembayaran premi lancar dihitung sebagai berikut,

$$P(X_3 = 6 | Y_1) = \frac{1}{\sqrt{(2)(3,14)(5,0237)}} \times \exp\left(-\frac{(6)-(8,5352)^2}{2(5,0237)}\right) = 0,0939$$

Sehingga probabilitas masa pembayaran premi 6 tahun dengan syarat status pembayaran premi lancar sebesar 0,0939.

Berdasarkan data *training* pembayaran premi dengan status pembayaran premi tidak lancar pada variabel masa pembayaran premi diketahui memiliki nilai rata-rata sebesar 9,0909 dan nilai variansi sebesar 1,8352 maka probabilitas variabel masa pembayaran premi 6 tahun dengan syarat status pembayaran premi tidak lancar dihitung sebagai berikut,

$$P(X_3 = 6 | Y_2) = \frac{1}{\sqrt{(2)(3,14)(1,8352)}} \times \exp\left(-\frac{(6)-(9,0909)^2}{2(1,8352)}\right) = 0,0218$$

Sehingga probabilitas variabel masa pembayaran premi 6 tahun dengan syarat status pembayaran premi tidak lancar sebesar 0,0218.

d. Besar Pembayaran Premi (X_4)

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

Berdasarkan data *training* pembayaran premi dengan status

pembayaran premi lancar pada variabel besar pembayaran premi diketahui memiliki nilai rata-rata sebesar Rp. 1.633.803 dan nilai variansi sebesar $1,5968 \times 10^{13}$ maka probabilitas variabel besar pembayaran premi sebesar Rp.3.000.000,00 dengan syarat status pembayaran premi lancar dihitung sebagai berikut,

$$P(X_4 = 3.000.000 | Y_1) = \frac{1}{\sqrt{(2)(3,14)(1,5968 \times 10^{13})}} \times \exp\left(-\frac{(3.000.000)-(1.633.803)^2}{2(1,5968 \times 10^{13})}\right) = 1,7494 \times 10^{-7}$$

Sehingga probabilitas variabel besar pembayaran premi sebesar Rp.3.000.000,00 dengan syarat status pembayaran premi lancar sebesar $1,7494 \times 10^{-7}$.

Berdasarkan data *training* pembayaran premi dengan status pembayaran premi tidak lancar pada variabel besar pembayaran premi diketahui memiliki nilai rata-rata sebesar Rp. 3.969.697 dan nilai variansi sebesar $3,8428 \times 10^{12}$ maka probabilitas variabel besar pembayaran premi sebesar Rp.3.000.000,00 dengan syarat status pembayaran premi tidak lancar, dihitung sebagai berikut,

$$P(X_4 = 3.000.000 | Y_2) = \frac{1}{\sqrt{(2)(3,14)(3,8428 \times 10^{12})}} \times \exp\left(-\frac{(3.000.000)-(3.969.697)^2}{2(3,8428 \times 10^{12})}\right) = 1,8012 \times 10^{-7}$$

Sehingga probabilitas variabel besar pembayaran premi sebesar Rp.3.000.000,00 dengan syarat status pembayaran premi tidak lancar sebesar $1,8012 \times 10^{-7}$.

Adapun perhitungan perkalian probabilitas *prior* dan probabilitas setiap variabel bebas pada kedua kelompok sebagai berikut:

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

$$P(Y_1 | X_1, X_2, X_3, X_4)$$

$$\begin{aligned}
&= P(Y_1) \times P(X_1 | Y_1) \times P(X_2 | Y_1) \times P(X_3 | Y_1) \times \\
&\quad P(X_4 | Y_1) \\
&= P(Y_1) \times P(X_1 = 6.000.000 | Y_1) \times \\
&\quad P(X_2 = 40 | Y_1) \times P(X_3 = 6 | Y_1) \times \\
&\quad P(X_4 = 3.000.000 | Y_1) \\
&= (0,6827) \times (5,7061 \times 10^{-8}) \times (0,0453) \times \\
&\quad (0,0939) \times (1,7494 \times 10^{-7}) \\
&= 2,9022 \times 10^{-17}
\end{aligned}$$

Sehingga dapat diketahui bahwa probabilitas nasabah asuransi mengalami status pembayaran premi lancar dengan syarat pendapatan sebesar Rp. 6.000.000,00, usia 40 tahun, masa pembayaran premi selama 6 tahun dan besar pembayaran premi Rp 3.000.000,00 adalah sebesar $2,9022 \times 10^{-17}$

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

$$\begin{aligned}
&P(Y_2 | X_1, X_2, X_3, X_4) \\
&= P(Y_2) \times P(X_1 | Y_2) \times P(X_2 | Y_2) \times P(X_3 | Y_2) \times \\
&\quad P(X_4 | Y_2) \\
&= P(Y_2) \times P(X_1 = 6.000.000 | Y_2) \times \\
&\quad P(X_2 = 40 | Y_2) \times P(X_3 = 6 | Y_2) \times \\
&\quad P(X_4 = 3.000.000 | Y_2) \\
&= (0,3173) \times (5,0047 \times 10^{-8}) \times (1,3483 \times 10^{-2}) \\
&\quad (0,0218) \times (1,8012 \times 10^{-7}) \\
&= 8,4135 \times 10^{-19}
\end{aligned}$$

Sehingga dapat diketahui bahwa probabilitas nasabah asuransi mengalami status pembayaran premi tidak lancar dengan syarat pendapatan sebesar Rp. 6.000.000,00, usia 40 tahun, masa pembayaran premi selama 6 tahun dan besar pembayaran premi Rp. 3.000.000,00 adalah sebesar $8,4135 \times 10^{-19}$

Berdasarkan perhitungan perkalian probabilitas awal dan probabilitas setiap variabel bebas (*posterior*) pada kedua kelompok dapat diketahui bahwa kelompok yang memiliki *posterior* terbesar adalah nasabah asuransi dengan kelompok status pembayaran premi lancar sebesar $2,9022 \times 10^{-17}$ dibandingkan *posterior* nasabah asuransi dengan kelompok status pembayaran premi tidak lancar sebesar $8,4135 \times 10^{-19}$, sehingga dapat disimpulkan data *testing* pertama yaitu obyek dengan

nilai-nilai variabel pendapatan sebesar Rp. 6.000.000,00, usia 40 tahun, masa pembayaran premi selama 6 tahun dan besar pembayaran premi Rp. 3.000.000,00 diklasifikasikan masuk ke dalam kelompok pertama yaitu nasabah asuransi dengan status pembayaran premi lancar.

5. Pengukuran Tingkat Akurasi Klasifikasi

Pada proses klasifikasi menggunakan analisis diskriminan Fisher, jumlah obyek yang tepat dan salah diklasifikasikan untuk masing-masing kelompok dapat dilihat pada Tabel 4. Tanda (*) pada angka-angka menyatakan jumlah obyek kelompok tertentu yang salah diklasifikasikan dengan menggunakan analisis diskriminan Fisher.

Tabel 4. Hasil Klasifikasi Analisis Diskriminan Fisher

Klasifikasi Awal Status Pembayaran Premi	Prediksi Klasifikasi Analisis Diskriminan Fisher		Total
	Lancar	Tidak Lancar	
Lancar	8	12*	20
Tidak Lancar	0*	6	6
Total	8	18	26

Berdasarkan Tabel 4 dapat diketahui bahwa status pembayaran premi nasabah asuransi PT. Prudential *Life* Tahun 2018 menggunakan analisis diskriminan Fisher diperoleh hasil yaitu dari 20 nasabah yang memiliki status pembayaran premi lancar terdapat 8 nasabah tepat diklasifikasikan memiliki status pembayaran premi lancar dan sisanya 12 nasabah tidak tepat diklasifikasikan. Terdapat 6 nasabah yang memiliki status pembayaran premi tidak lancar serta tepat diklasifikasikan memiliki status pembayaran premi tidak lancar sehingga diperoleh nilai APER sebagai berikut:

$$\begin{aligned}
\text{APER} &= \frac{12+0}{26} \times 100\% \\
&= \frac{12}{26} \times 100\% \\
&= 46,15\%
\end{aligned}$$

Pada proses klasifikasi menggunakan metode *naive* Bayes, jumlah obyek yang tepat dan salah diklasifikasikan untuk masing-masing kelompok dapat dilihat pada Tabel 5. Tanda (*) pada angka-angka menyatakan jumlah obyek kelompok tertentu yang salah diklasifikasikan dengan menggunakan metode *naive* Bayes.

Nilai multikolinieritas (VIF) variabel bebas yang ada diperoleh bahwa pada variabel pendapatan (X_1) sebesar 1,1493, variabel usia (X_2) sebesar 1,1035, variabel masa pembayaran premi (X_3) sebesar 1,1897 dan variabel besar pembayaran premi (X_4) sebesar 1,0155, karena nilai VIF pada keempat variabel yang ada kurang dari 10 maka diindikasikan tidak mengalami multikolinieritas antar variabel bebas yang ada serta di peroleh standar error residual sebesar 0,3551.

Tabel 5. Hasil Klasifikasi Metode *Naive* Bayes

Klasifikasi Awal Status Pembayaran Premi	Prediksi Klasifikasi Metode <i>Naive</i> Bayes		Total
	Lancar	Tidak Lancar	
Lancar	18	2*	20
Tidak Lancar	2*	4	6
Total	20	6	26

Berdasarkan Tabel 5 dapat diketahui bahwa status pembayaran premi nasabah asuransi PT. Prudential *Life* Tahun 2018 menggunakan metode *naive* Bayes diperoleh hasil yaitu dari 20 nasabah yang memiliki status pembayaran premi lancar terdapat 18 nasabah tepat diklasifikasikan memiliki status pembayaran premi lancar dan sisanya 2 nasabah tidak tepat diklasifikasikan. Pada 6 nasabah yang memiliki status pembayaran premi tidak lancar terdapat 4 nasabah tepat diklasifikasikan memiliki status pembayaran premi tidak lancar dan sisanya 2 nasabah tidak tepat diklasifikasi sehingga diperoleh nilai APER sebagai berikut:

$$\begin{aligned} \text{APER} &= \frac{2+2}{26} \times 100\% \\ &= \frac{4}{26} \times 100\% \\ &= 15,38\% \end{aligned}$$

Tabel 6. Hasil Pengukuran Tingkat Akurasi

Metode	APER
Analisis Diskriminan Fisher	46,15%
Metode <i>Naive</i> Bayes	15,38%

Pada Tabel 6 dapat diketahui bahwa pada analisis diskriminan Fisher menunjukkan kesalahan klasifikasi dalam memprediksi klasifikasi sebesar 46,15% dan pada metode *naive* Bayes menunjukkan kesalahan klasifikasi dalam memprediksi klasifikasi sebesar 15,38%. Hal ini menunjukkan bahwa tingkat akurasi pengklasifikasian metode *naive* Bayes lebih baik dibandingkan dengan analisis diskriminan Fisher.

KESIMPULAN DAN SARAN

1. Kesimpulan

1. Pengklasifikasian status pembayaran premi pada data nasabah asuransi PT. Prudential *Life* Tahun 2018 menggunakan analisis diskriminan Fisher diperoleh hasil yaitu dari 26 nasabah, terdapat 12 nasabah yang tidak tepat diklasifikasikan.
2. Pengklasifikasian status pembayaran premi pada data nasabah asuransi PT. Prudential *Life* Tahun 2018 menggunakan metode *naive* Bayes diperoleh hasil yaitu dari 26 nasabah, terdapat 4 nasabah yang tidak tepat diklasifikasikan.
3. Hasil pengukuran tingkat akurasi klasifikasi menggunakan APER status pembayaran premi pada data nasabah asuransi PT. Prudential *Life* Tahun 2018 dengan analisis diskriminan Fisher diperoleh hasil kesalahan klasifikasi sebesar 46,15% dan metode *naive* Bayes sebesar 15,38%. Hal ini menunjukkan bahwa metode *naive* Bayes memberikan ketepatan prediksi klasifikasi yang lebih baik

2. Saran

Sebaiknya dalam penelitian selanjutnya dapat menggunakan metode pengklasifikasian lainnya seperti *Support Vector Machine* (SVM), *Probabilistic Neural Network* (PNN), analisis diskriminan Kernel atau metode pengklasifikasian lainnya serta dapat dikembangkan dengan data yang memiliki kategori lebih dari dua pada variabel terikat dan variabel bebas yang lebih banyak.

DAFTAR PUSTAKA

- Artha, C. A. 2016. *Perbandingan Hasil Klasifikasi Menggunakan Regresi Logistik dan Analisis Diskriminan Kuadratik Pada Kasus Pengklasifikasian Jurusan Di SMA Negeri 8 Samarinda Tahun Ajaran 2014/2015*. Jurnal Eksponensial, Vol. 7(2), 179-186.
- Johnson, R. A., dan D. W Winchern. 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.
- Santosa, B. 2000. *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Supranto, J. 2004. *Analisis Multivariat Arti dan Interpretasi*. Jakarta: PT. Rineka Cipta.
- Towa, H. dan Tairas, J. N. B. (2002). *Pengantar Klasifikasi Persepuluhan Dewey*. Jakarta: BPK Gunung Mulya.
- Triandaru, S. dan Budisantoso, T. (2008). *Bank dan Lembaga Keuangan Lainnya*. Jakarta: Salemba Empat.
- Widarjono, A. 2015. *Analisis Multivariat Terapan dengan Program SPSS, AMOS, dan SMARTPLS Edisi Kedua*. Yogyakarta: UPM STIM YKPN.
- Witten, dkk. 2011. *Data Mining: Pratical Machine Learning Tools and Techniques, Third Edition*. Burlington: Morgan Kaufmann Publishers.