ISSN: 2339-2541
JURNAL GAUSSIAN, Volume 5, Nomor 1, Tahun 2016, Halaman 1 - 10
Online di: http://ejournal-s1.undip.ac.id/index.php/gaussian

PERBANDINGAN ANALISIS DISKRIMINAN FISHER DAN NAIVE BAYES UNTUK KLASIFIKASI RISIKO KREDIT (Studi Kasus Debitur di Koperasi Jateng Amanah Mandiri Cabang Sukorejo Kendal)

Abdur Rofiq[1], Triastuti Wuryandari[2], Rita Rahmawati[3]
[1]Mahasiswa Jurusan Statistika FSM Universitas Diponegoro
[2,3]Staff Pengajar Jurusan Statstika FSM Universitas Diponegoro
rofiqdr@yahoo.com, triastuti@undip.ac.id, ritarahmawati@gmail.com

ABSTRACT

Credit is a form of money lending to debitors conducted by financial institutions such as cooperatives. In practice there are obstacles in the form of bad credit. Analyze by Fisher discriminant analysis method and Naive Bayes is used to classify the debitors fall into the category bad debitorr or not. This study uses data from the Debitors of Cooperative of Central Java Amanah Independent in Sukorejo Kendal Branch. The data obtained is used for classification by Fisher discriminant analysis and Naive Bayes method. Data obtained has multivariate normal distribution, has the same of variance-covariance matrix and has metric scale. Fisher discriminant analysis and Naive Bayes calculated and compared to the level of accuracy. From this research, the degree of accuracy of each method, namely 90% for Fisher Discriminant Analysis and 83.33% for the Naive Bayes. Having tested using the proportion test, Fisher discriminant analysis method is no different accuracy when compared with Naive Bayes to classify credit risk.

Keywords: debitors, credit risk, Fisher discriminant analysis, Naive Bayes.

1. PENDAHULUAN

Program penyaluran kredit melalui lembaga keuangan informal seperti koperasi, dalam menjalankan kegiatan usahanya dituntut untuk menjaga keberlangsungan usahanya *(survive)*. Koperasi menghadapi kendala dalam pemberian kredit bagi anggotanya. Salah satu kendala dalam penyaluran kredit adalah adanya kredit macet. Untuk mengatasi masalah tersebut perlu adanya analisa terlebih dahulu terhadap calon debitur. Data debitur diperoleh dari Koperasi Jateng Amanah Mandiri Cabang Sukorejo Kendal kemudian dianalisis untuk mengetahui risiko kreditnya[4]. Data yang diperoleh berdistribusi normal multivariat dan mempunyai matriks varian dan kovarian sama antar kedua kelompok macet dan lancar. Variabel bebas yang digunakan adalah pendapatan, lama pinjam, jumlah keluarga dan jumlah pinjaman. Variabel tersebut diambil berdasarkan faktor-faktor yang berpengaruh terhadap risiko kredit dari penelitian-penelitian sebelumnya dan memiliki skala metrik. Dari alasan tersebut maka penelitian ini bertujuan untuk mengklasifikasi debitur yang berisiko lancar dan macet berdasarkan risiko kreditnya. Metode yang digunakan dalam pengkalasifikasian ini adalah Analisis Diskriminan Fisher dan Klasifikasi Naive Bayes.

2. TINJAUAN PUSTAKA

2.1 Pengertian Kredit

Kredit adalah pemberian prestasi (misalnya uang, barang) dengan balas prestasi yang akan terjadi pada waktu yang akan datang[4]. Analisis kredit adalah suatu proses yang dimaksudkan untuk menganalisis atau menilai suatu permohonan kredit yang diajukan oleh calon debitur kredit sehingga dapat memberikan keyakinan kepada pihak debitur bahwa proyek yang akan dibiayai dengan kredit tersebut cukup layak *(feasible)*. Dengan adanya analisis kredit ini dapat dicegah secara dini kemungkinan terjadinya kegagalan debitur dalam memenuhi kewajibannya untuk melunasi kredit yang diterimanya[4].

2.2 Analisis Diskriminan

Analisis diskriminan digunakan untuk mengklasifikasikan individu ke dalam salah satu dari dua kelompok atau lebih. Analisis diskriminan digunakan pada kasus dimana variabel bebas berupa data metrik (interval atau rasio) dan variabel terikat berupa data nonmetrik (nominal atau ordinal)[3]. Model analisis diskriminan berkenaan dengan kombinasi linear memiliki bentuk sebagai berikut:

$$Y_{ji} = \mathbf{b}^T\mathbf{X}_{jig} = b_1X_{ji1} + b_2X_{ji2} + ...+ b_gX_{jkg}$$
$$g = 1,2,..., p \quad , j = 1, 2, ..., n_i \ , \text{dan } i = 1, 2, ..., k \qquad (1)$$

2.3 Klasifikasi

Klasifikasi mempunyai tujuan untuk memasukkan observasi baru ke dalam kelompok yang telah mempunyai label kelompok[3]. Aturan klasifikasi berdasarkan dua kelompok (populasi) normal dengan matriks kovariansi yang sama dijelaskan sebagai berikut[3]:

$$R_1 = \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \left(\frac{C(1|2)}{C(2|1)}\right)\frac{p_2}{p_1} \qquad (2)$$

$$R_2 = \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \left(\frac{C(1|2)}{C(2|1)}\right)\frac{p_2}{p_1} \qquad (3)$$

Dengan

R_1 = Regional (daerah) kelompok 1, R_2 =Regional (daerah) kelompok 2

$C(1|2)$ = data yang sebenarnya berada pada kelompok 2 tetapi diprediksi masuk kedalam kelas 1.

$C(2|1)$ = data yang sebenarnya berada pada kelompok 1 tetapi diprediksi masuk kedalam kelas 2.

p_1 = Peluang prediksi untuk kelas 1. p_2 = Peluang prediksi untuk kelas 2.

$$R_1 = \left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)^t \boldsymbol{\Sigma}^{-1}\mathbf{x} \geq \frac{1}{2}(\bar{\mathbf{w}}_1 + \bar{\mathbf{w}}_2), \qquad (4)$$

$$R_2 = \left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)^t \boldsymbol{\Sigma}^{-1}\mathbf{x} < \frac{1}{2}(\bar{\mathbf{w}}_1 + \bar{\mathbf{w}}_2) \qquad (5)$$

Dengan

$$\hat{m} = \frac{1}{2}(\bar{\boldsymbol{w}}_1 + \bar{\boldsymbol{w}}_2) = \frac{1}{2}\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\right)^t \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2\right) \qquad (6)$$

$$\underset{(pxp)}{\boldsymbol{\Sigma}_i} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} & \sigma_{34} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_{44} \end{bmatrix}, \text{ i=1,2 dan p=1, 2, 3, 4.} \qquad (9)$$

$$\underset{(px1)}{\boldsymbol{\mu}_i} = \begin{bmatrix} \mu_{1i} \\ \mu_{2i} \\ \mu_{3i} \\ \mu_{4i} \end{bmatrix}, \text{ i=1,2 dan p=1, 2, 3, 4.} \qquad (10)$$

$f_1(x)$ = fungsi densitas untuk kelompok 1

$f_2(x)$ = fungsi densitas untuk kelompok 2

2.4 Asumsi Normal Multivariat

Pengujian distribusi normal multivariat dapat dilakukan dengan cara perhitungan secara formal, dengan Kolmogorov Smirnov sebagai berikut[3]:

a. Hipotesis:
H_0 : Data observasi berdistribusi normal multivariat.
H_1 : Data observasi tidak berdistribusi normal multivariat.
b. Taraf Signifikansi : α
c. Statistik Uji:

$$D_{hitung} = Sup\left|S\left(d_{ij}^2\right) - F_0\left(d_{ij}^2\right)\right| \tag{11}$$

$S\left(d_{ij}^2\right)$ = fungsi distribusi kumulatif dari d_{ij}^2
$F_0\left(d_{ij}^2\right)$ = fungsi distribusi kumulatif dari Chi-Kuadrat.
d. Kriteria Pengujian:
H_0 ditolak jika nilai D_{hitung} lebih besar dari nilai *D* pada tabel Kolmogorov Smirnov.

2.5 Uji Kesamaan Matriks Varian Kovarian

Terpenuhinya asumsi kesamaan matriks varian kovarian merupakan salah satu syarat digunakannya fungsi diskriminan Fisher[3]. Pengujian kesamaan matriks varian kovarian dapat dilakukan dengan menggunakan uji *Box's M*[3].

Uji kesamaan matriks varian kovarian adalah sebagai berikut:
a. Hipotesis:
H_0 : matriks varian-kovarian dari kedua kelompok yang diamati adalah sama ($\boldsymbol{\Sigma_1} = \boldsymbol{\Sigma_2}$)
H_1 : matriks varian-kovarian dari kedua kelompok yang diamati adalah berbeda
b. Taraf signifikansi = α
c. Statistik Uji:

$$C = (1-\mu)M = (1-\mu) \sum_{i=1}^{k}(n_i - 1)ln|\boldsymbol{\Sigma}| - \sum_{i=1}^{k}(n_i - 1)ln|\boldsymbol{\Sigma_i}| \tag{12}$$

Dengan $\boldsymbol{\Sigma}$ adalah matriks gabungan varian kovarian kelompok ke-i dan μ yang dirumuskan sebagai berikut:

$$\boldsymbol{\Sigma} = \left(\frac{n_1-1}{(n_1-1)+(n_2-1)}\right)\boldsymbol{\Sigma_1} + \left(\frac{n_2-1}{(n_1-1)+(n_2-1)}\right)\boldsymbol{\Sigma_2} \tag{13}$$

$$\mu=\left[\sum_{i=1}^{k}\frac{1}{(n_i-1)} - \frac{1}{\sum_{i=1}^{l}(n_i-1)}\right]\left[\frac{2p^2+3p-1}{6(p+1)(k-1)}\right] \tag{14}$$

sedangkan $\boldsymbol{\Sigma_i}$ adalah matriks varian kovarian kelompok ke-i.
d. Kriteria pengujian:
H_0 ditolak pada taraf signifikansi α apabila nilai C > $X^2_{\alpha;p(p+1)(k-1)/2}$ yang berarti bahwa dua matriks varian kovarian berbeda.

2.6 Analisis Diskriminan Fisher untuk Klasifikasi dengan Dua Kelompok

Analisis Diskriminan Fisher ini memperhatikan syarat bahwa data yang digunakan harus memiliki matriks kovarians yang sama untuk setiap kelompok yang diberikan[3]. Untuk kasus k=2 yang artinya ada dua kelompok, π_1 sebagai kelompok 1 dan π_2 sebagai kelompok 2. Probabilitas priornya tidak diketahui, maka $p_1 = p_2 = \cdots = p_k = 1/k$[3]. Fungsi diskriminan Fisher ditunjukkan dengan persamaan sebagai berikut:

$$\underset{(1x1)}{y} = \underset{(1xp)}{\boldsymbol{b}^t}\underset{(px1)}{\boldsymbol{X}} \tag{15}$$

dengan

$$\underset{(px1)}{\mathbf{b}} = \underset{(pxp)}{\boldsymbol{\Sigma}^{-1}}\underset{(px1)}{(\boldsymbol{\mu_1} - \boldsymbol{\mu_2})} \tag{16}$$

Aturan klasifikasi fungsi Diskriminan Fisher adalah jika $\boldsymbol{y} = \boldsymbol{b^t X} \geq \widehat{m}$ maka nasabah masuk ke dalam kelompok 1. Jika $\boldsymbol{y} = \boldsymbol{b^t X} < \widehat{m}$ maka nasabah masuk dalam kelompok 2.

2.7 Klasifikasi Naive Bayes

Misalkan peristiwa $A_1, A_2, \dots, A_p$ membentuk partisi dalam Ω sedemikian sehingga P(A_g) > 0 untuk g = 1,2,...,p dan misalkan B sembarang peristiwa sedemikian P(B) > 0 untuk g=1,2,...,p maka [5]:

$$P(\mathrm{A_g}|B) = \frac{P(\mathrm{A_g})P(B|\mathrm{A_g})}{\sum_{g=1}^{p} P(\mathrm{A_g})P(B|\mathrm{A_g})} \quad (22)$$

Klasifikasi Naive Bayes merupakan salah satu metode klasifikasi yang berakar pada teorema Bayes. Prediksi Bayes didasarkan pada teorema Bayes.

Jika **X** merupakan vektor yang berisi fitur dan **Y** adalah label kelompok, Naive Bayes dituliskan dengan $P(\mathbf{Y}|\boldsymbol{X})$. Nilai tersebut berarti probabilitas label kelompok **Y** didapatkan setelah fitur-fitur **X** diamati. Notasi ini disebut juga probabilitas akhir untuk **Y**, sedangkan P(**Y**) disebut probabilitas awal untuk **Y**.

Formula Naive Bayes untuk klasifikasi adalah sebagai berikut[5]:

$$P(\boldsymbol{Y}|\boldsymbol{X_g}) = \frac{P(\boldsymbol{Y})\prod_{g=1}^{p} P(\boldsymbol{X_g}|\boldsymbol{Y})}{P(\boldsymbol{X_g})} \quad (24)$$

Dengan **Y** = Y_i ,i = 1,2, ... ,k

$P(\boldsymbol{Y}|\boldsymbol{X_g})$ adalah probabilitas data dengan vektor **X** pada kelompok **Y**

$P(\boldsymbol{Y})$ adalah probabilitas awal kelompok **Y**

$\prod_{g=1}^{p} P(\boldsymbol{X_g}|\boldsymbol{Y})$ adalah probabilitas independen kelompok **Y** dari semua fitur dalam vektor **X** dan $P(\boldsymbol{X_g})$ adalah probabilitas dari $\mathbf{X_g}$. Probabilitas P(**X**) selalu tetap sehingga dalam perhitungan prediksi dapat dihilangkan dan hanya menghitung bagian $P(\boldsymbol{Y})\prod_{\boldsymbol{g=1}}^{\boldsymbol{p}} \boldsymbol{P}(\boldsymbol{X_g}|\boldsymbol{Y})$ dengan memilih nilai yang terbesar sebagai kelompok yang dipilih sebagai hasil prediksi. Sementara Probabilitas independen $\prod_{g=1}^{p} P(\boldsymbol{X_g}|\boldsymbol{Y})$ merupakan pengaruh semua fitur dari data terhadap setiap kelompok **Y**, yang dinotasikan dengan:

$$P(\boldsymbol{X_g}|\boldsymbol{Y} = y) = \prod_{g=1}^{p} P(\boldsymbol{X_g}|\boldsymbol{Y} = y) \quad (25)$$

Jika data bertipe numerik (non kategorik) ada perlakuan khusus sebelum diproses menggunakan Naive Bayes. Caranya adalah: Mengasumsikan bentuk tertentu dari distribusi probabilitas untuk fitur kontinu dan memperkirakan parameter distribusi dengan data training. Distribusi Gaussian biasanya dipilih untuk mempresentasikan probabilitas bersyarat dari fitur kontinu pada sebuah kelompok $P(\boldsymbol{X_g}|\boldsymbol{Y})$, sedangkan distribusi Gaussian dikarakteristikkan dengan dua parameter yaitu mean (µ) dan varian (σ^2) untuk setiap kelompok Y_i, probabilitas bersyarat kelompok Y_i untuk fitur X_g adalah

$$P(\boldsymbol{X} = x_g|\boldsymbol{Y} = y_i) = g\left(x_g, \mu_{gi}, \sigma_{gi}\right), dimana$$

$$g\left(x_g, \mu_{gi}, \sigma_{gi}\right) = \frac{1}{\sqrt{2\pi}\sigma_{gi}}\, exp^{-\frac{(x_g-\mu_{gi})^2}{2\sigma_{gi}^2}} \quad \text{, i=1,2 dan g=1,2,3,4} \quad (30)$$

Pada data yang kontinyu dimisalkan data berada pada interval $x \le X \le x + \Delta$, sehingga $p(x \le X \le x + \Delta) = \int_{x}^{x+\Delta} g(x;\mu;\sigma)dx$. Berdasarkan fungsi turunan diperoleh $lim_{\Delta\to 0} p(x \le X \le x + \Delta)/\Delta = g(x;\mu;\sigma)$. Untuk Δ yang kecil, $p(X = x) \approx g(x;\mu;\sigma).\Delta$. Faktor Δ muncul pada setiap kelas sebagai pembilang dan mempunyai nilai yang sama pada setiap kelas sehingga dihilangkan ketika melakukan normalisasi, sehingga digunakan persamaan (30)[2].

2.8 Kesalahan Klasifikasi *(Misclassification)* dan APER *(Apparent Error Rate)*

Misclassification merupakan kesalahan dari pengklasifikasian suatu observasi baru ke dalam suatu kelompok [6]. Penilaian kinerja dari setiap prodesur klasifikasi adalah dengan cara menghitung tingkat kesalahan atau probabilitas kesalahan klasifikasi. Salah satu metode yang dapat digunakan untuk menghitung probabilitas kesalahan klasifikasi adalah APER (*Apparent Error Rate*) yaitu perhitungan kinerja hasil klasifikasi yang dilakukan dengan menggunakan matriks konfusi (*confusion matric*)[6]. Untuk sebanyak n_1 observasi kelompok pertama (π_1), sebanyak n_2 observasi kelompok kedua (π_2) sampai sebanyak π_k observasi kelompok ke-k (π_k). Untuk analisis diskriminan 2 kelompok (k=2), persentase kesalahan klasifikasi dapat dihitung dari matriks yang menunjukkan nilai sebenarnya atau aktual dan nilai prediksi dari setiap kelompok, seperti terlihat pada tabel berikut:

Tabel 1. Klasifikasi Dua Kelompok

Aktual	Prediksi		
	π_1	π_2	Total
π_1	n_{11}	n_{12}	$n_{11}+n_{12}$
π_2	n_{21}	n_{22}	$n_{21}+n_{22}$
Total	$n_{11}+n_{21}$	$n_{12}+n_{22}$	$N=n_{11}+n_{12}+n_{21}+n_{22}$

Dengan

n_{11} = jumlah objek dari π_1 yang tepat diklasifikasikan sebagai π_1

n_{12} = jumlah objek dari π_1 yang salah diklasifikasikan sebagai π_2

n_{21} = jumlah objek dari π_2 yang salah diklasifikasikan sebagai π_1

n_{22} = jumlah objek dari π_1 yang tepat diklasifikasikan sebagai π_2

$$\text{Maka } APER = \frac{n_{12}+n_{21}}{N} \tag{31}$$

2.9 Menilai Keakuratan Prediksi Keanggotaan Kelompok

Langkah terakhir untuk menilai model secara keseluruhan adalah dengan menentukan tingkat akurasi prediksi dari fungsi diskriminan dan Naive bayes. Penentuan ini dilakukan dengan menggunakan uji statistik yang dinamakan *Press's Q*[1].

Hipotesis:

H_o : Pengklasifikasian tidak akurat

H_1 : Pengklasifikasian akurat

Taraf signifikansi = α

Statistik uji:

$$\text{Press's Q} = \frac{[N-(qk)]^2}{N(k-1)} \tag{32}$$

Dengan N adalah Banyaknya sampel=$n_{11}+n_{12}+n_{21}+n_{22}$

q adalah banyaknya kasus yang diklasifikasi secara tepat=$n_{11}+n_{22}$

k adalah banyaknya kelompok

Kriteria Uji: menolak H_0 jika Press's Q >$\chi^2_{\alpha,1}$

2.10 Perbandingan Akurasi

Akurasi merupakan nilai yang diperhatikan untuk mendapatkan hasil klasifikasi yang terbaik. Semakin besar nilai akurasi maka semakin tepat hasil klasifikasinya. Untuk menguji apakah dua metode mempunyai tingkat akurasi yang berbeda maka digunakan uji proporsi. Uji proporsi adalah pengujian hipotesis mengenai proporsi populasi yang di dasarkan atas informasi sampelnya. Nilai proporsi didapatkan dengan cara menghitung nilai akurasi dari masing-masing metode.

Hipotesis

H_0 : PR_1= PR_2 (tidak ada perbedaan tingkat akurasi dari kedua metode)

H_1:$PR_1 \neq PR_2$ (ada perbedaan tingkat akurasi dari kedua metode)

Rumus uji bedaproporsi sebagai berikut[6]:

$$Z_{hitung} = \frac{PR_1 - PR_2}{\sqrt{\left(PR_{gab.}(1-PR_{gab})\right).\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \tag{33}$$

Dengan,

PR_1 =Proporsi metode pertama (Akurasi metode pertama)

PR_2 =Proporsi metode kedua (Akurasi metode kedua)

PR_{gab}= Proporsi gabungan yaitu $\frac{n_1 PR_1 + n_2 PR_2}{n_1 + n_2}$

n_1 = Ukuran sampel pertama dan n_2= Ukuran sampel kedua

Kriteria Uji : Ho ditolak jika $Z_{hitung} > Z_{\alpha/2}$ atau $Z_{hitung} < -Z_{\alpha/2}$

3. METODOLOGI PENELITIAN

3.1. Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yaitu data Debitur di Koperasi Jateng Amanah Mandiri Cabang Sukorejo berdasarkan tanggal jatuh tempo bulan Desember 2014 – Februari 2015, dengan ukuran populasi 240 orang.

3.2. Variabel Penelitian

Variabel-variabel yang digunakan dalam penelitian ini adalah:

1. Variabel Tak Bebas (Y) : Status /kriteria Debitur,

   Dikelompokkan menjadi dua kelompok, yaitu : Kode 1, untuk kelompok debitur berisiko macet dan Kode 2, untuk kelompok debitur berisiko lancar.
2. Variabel Bebas (X):

   X_1=Pendapatan (rupiah) , X_2=Lama pinjam (bulan), X_3= Jumlah Keluarga (orang), X_4=Jumlah Pinjaman (rupiah).

3.3. Metode Sampling dan Ukuran Sampel

Sampel adalah bagian dari jumlah dan karakteristik yang dimiliki oleh populasi (Sugiyono, 2012). Pada penelitian ini peneliti menentukan ukuran sampel menggunakan Nomogram Herry King (Sugiyono,2012).

3.4. Langkah-langkah Analisis

Data yang telah dikumpulkan dalam penelitian ini kemudian diolah dengan menggunakan analisis diskriminan Fisher dan klasifikasi Naive Bayes. Pengolahan data dilakukan dengan bantuan program Microsoft Excel 2007 dan software R-14.

Adapun langkah-langkah yang dilakukan pada metode Analisis Diskriminan Fisher adalah sebagai berikut:

a. Melakukan pengujian asumsi normal multivariat dan uji kesamaan matriks varian-kovarian.
b. Membagi data training dan testing.
c. Menghitung penaksir untuk analisis Diskriminan Fisher.
d. Menentukan fungsi diskriminan dua kelompok berdasarkan penaksir diskriminan.
e. Meghitung skor diskriminan dan mengklasifikasikan ke dalam kelompok pertama atau kelompok kedua.
f. Menghitung jumlah pengamatan yang salah dalam pengklasifikasian (*APER*).
g. Menilai Keakuratan Prediksi Keanggotaan Kelompok dengan uji Press's Q.

Langkah-langkah yang dilakukan pada metode Naive Bayes adalah sebagai berikut:

a. Membagi data menjadi 2 yaitu data testing dan data training.
b. Menghitung probabilitas prior ($\mathbf{P(Y)}$) dari data testing berdasarkan data training.
c. Menghitung probabilitas atribut terhadap masing-masing kelas ($\boldsymbol{P(X_g|Y)}$) pada data testing berdasarkan data training.
d. Menghitung perkalian probabilitas dengan probabilitas atribut pada masing-masing kelas ($\boldsymbol{P(Y)P(X_g|Y)}$).
e. Mencari nilai maksimal dari ($\boldsymbol{P(Y)P(X_g|Y)}$) pada kedua kelas.
f. Menghitung jumlah pengamatan yang salah dalam pengklasifikasian (*APER*).
g. Menilai Keakuratan Prediksi Keanggotaan Kelompok dengan uji Press's Q.

Setelah langkah-langkah dipenuhi maka dilakukan perbandingan tingkat akurasi pada kedua metode.

4. HASIL DAN PEMBAHASAN

4.1. Uji Nomal Multivariat

Data yang diperoleh dengan software R. Plot data membentuk garis lurus berarti bahwa data observasi memenuhi asumsi distribusi normal multivariate. Pengujian distribusi normal multivariat secara formal dengan Kolmogorov Smirnov dipeoleh kesimpulan p-value=0,222 > $\alpha = 5\%$ dan $D_{hitung} = 0{,}0856$ < D=0,111 maka H_0 diterima sehingga Data berdistribusi Normal Multivariat.

4.2. Uji Kesamaan Matriks Varian Kovarian

Pada taraf signifikansi α=5%, C = 11,44338< $\chi^2_{0,05;4(4+1)(2-1)/2}$=18,31 maka H_0 diterima. Maka matriks varian-kovarian dari kedua kelompok yang diamati adalah sama ($\boldsymbol{\Sigma_1} = \boldsymbol{\Sigma_2}$)

4.3. Analisis Diskriminan Fisher untuk Klasifikasi Resiko Kredit

Setelah asumsi normal multivariat, dengan matriks varian kovarian sama, maka digunakan Analisis Diskriminan Fisher dapat digunakan untuk mengklasifikasikan sekumpulan obyek ke dalam dua kelompok. Berdasarkan perhitungan dengan menggunakan data training sebanyak 120 data diperoleh hasil sebagai berikut:

a. Nilai koefisien dari masing masing variabel (**b**)

$$\mathbf{b} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}) = \begin{bmatrix} -0{,}000005740395 \\ -0{,}1976804 \\ 2{,}220286 \\ 0{,}000001253012 \end{bmatrix}$$

b. Nilai batasan kelompok ($\widehat{m}$)

$\widehat{m} = \frac{1}{2}(\mu_1 - \mu_2)^t \boldsymbol{\Sigma}^{-1}(\mu_1 + \mu_2)$ =-2,33137

c. Menghitung nilai y dan melakukan klasifikasi

$$y = \boldsymbol{b^t x}$$

Sehingga persamaan yang diperoleh adalah sebagai berikut:

y = $-0{,}000005740395$ (X1) + $-0{,}1976804$ (X2) + $2{,}220286$ (X3) + . 0,000001253012 (X4).

Berdasarkan aturan klasifikasi yaitu:

Kelompok 1 : $\boldsymbol{y = b^t X} \geq$ ($\widehat{m}$=-2,33137) maka status debitur masuk kelompok 1

Kelompok 2 : $\boldsymbol{y} = \boldsymbol{b^t X} < (\hat{m}$=-2,33137) maka ststus debitur masuk kelompok 2

Tabel 2. Klasifikasi Dua Kelompok untuk Analisis Diskriminan Fisher

Aktual	Prediksi		
	1 (macet)	2 (lancar)	Total
1 (macet)	9	0	9
2 (lancar)	3	18	21
Total	12	18	$N = 30$

d. Kesalahan Klasifikasi (*Misclasification*)

Bedasakan klasifikasi diperoleh APER = ($\frac{3+0}{30}$) = 3/30 = 0,1

Akurasi = (1- APER) x 100% = (1-0,1) x 100% = 90%

Dengan demikian klasifikasi resiko kredit (studi kasus debitur di koperasi Jateng Amanah Mandiri cabang Sukorejo Kendal) dengan metode Analisis Diskriminan Fisher mempunyai nilai akurasi 90% dengan tingkat kesalahan 10%.

e. Menilai Keakuratan Prediksi Keanggotaan Kelompok

Nilai Press's Q = 19,2 > $\chi^2_{0,05;1}$ =3,84 maka pada taraf signifikansi α=5%, Pengklasifikasian dengan analisis Diskriminan Fisher akurat.

4.4. Klasifikasi Naive Bayes

Variabel-variabel signifikan yang diperoleh dalam analisis Diskriminan Fisher yaitu Pendapatan, Jumlah keluarga, Jumlah pinjaman digunakan untuk klasifikasi Naive Bayes.

4.4.1 Data Testing dan Data Training

Proporsi data yang digunakan adalah 120 (80%) digunakan sebagai data training dan 30 (20%) digunakan sebagai data testing

4.4.2. Probabilitas Awal dari Data Testing Berdasarkan Data Training.

a. Kelompok 1

Nilai Probabilitas setiap Variabel pada setiap kelompok: Untuk variabel X_1 (pendapatan), kelompok 1 (macet), Rata-rata = 2072580,645 dan Variansi = 354973000000. Peluang seorang debitur dengan pendapatan Rp 3.000.000,00 jika diketahui status kreditnya 1 (macet) adalah

$$P(X1 = 3000000|1) = \frac{1}{\sqrt{(2)(3,14)(354973000000)}}\, exp^{-\frac{(3000000-2072580,645)^2}{(2)(354973000000)}} = 0,000000199421$$

Dengan cara yang sama, perhitungan digunakan untuk nilai peluang yang lain.

b. Kelompok 2

Nilai Probabilitas setiap Variabel pada setiap kelompok: Untuk variabel X_1 (pendapatan), kelompok 2 (lancar), Rata-rata = 2589887,64 dan Variansi = 328476000000. Peluang seorang debitur dengan pendapatan Rp 3.000.000,00 jika diketahui status kreditnya 2 (lancar) adalah

$P(X1 = 3000000|2) =$

$$\frac{1}{\sqrt{(2)(3{,}14)(328476000000)}}\, exp^{-\frac{(3000000-2589887{,}64\,)^2}{(2)(\,328476000000)}} = 0{,}000000539$$

Dengan cara yang sama, perhitungan digunakan untuk nilai peluang yang lain.

4.4.3. Perkalian Probabilitas Awal dari Data Testing Berdasarkan Data Training.

Perkalian Probabilitas awal dari data testing berdasarkan data training digunakan untuk menentukan probabilitas gabungan $P(X_1, X_2, \ldots, X_p|Y) = P(X_1|Y)P(X_2|Y)P(X_3|Y) \ldots P(X_p|Y)$ pada masing-masing kelompok. Kemudian Mencari nilai maksimal dari kedua kelompok untuk prediksi kelompok.

a. $P(X_1 = 3000000, X_2 = 12, X_3 = 3, X_4 = 5000000)=$

$$((\,P(X_1 = 3000000|1)(P(X_2 = 12|1))(P(X_3 = 3|1))(P(X_4 = 5000000|1)))$$

$$=(0{,}000000199421)(0{,}050664)(0{,}170592)(0{,}000000134986)$$

$$= 0{,}00000000000000232659$$

Dengan cara yang sama maka digunakan untuk menghitung perkalian probabilitas yang lain.

b. $P(X_1 = 3000000, X_2 = 12, X_3 = 3, X_4 = 5000000)$

$$= ((P(X_1 = 3000000|2))(P(X_2 = 12|2))(P(X_3 = 3|2))(P(X_4 = 5000000|2)$$

$$= (0{,}00000053899)(0{,}074375)(0{,}332538)(0{,}0000000912747)$$

$$= 0{,}00000000000000121675$$

Dengan cara yang sama maka digunakan untuk menghitung perkalian probabilitas yang lain.

Tabel 3. Klasifikasi Dua Kelompok untuk Naive Bayes

Aktual	Prediksi		
	1 (macet)	2 (lancar)	Total
1 (macet)	8	1	9
2 (lancar)	4	17	21
Total	12	18	$N = 30$

4.4.4. Kesalahan Klasifikasi (*Misclasification*)

Dari Klasifikasi Naive Bayes diperoleh APER = $(\frac{4+1}{30})$ = 5/30 = 0,16667.

Akurasi = (1- APER) x 100% = (1-0,16667) x 100% = 83,33%

Dengan demikian klasifikasi resiko kredit (studi kasus debitur di koperasi Jateng Amanah Mandiri cabang Sukorejo Kendal) dengan metode Naive Bayes mempunyai nilai akurasi 83,33% dengan tingkat kesalahan 0,16667.

4.4.5. Menilai Keakuratan Prediksi Keanggotaan Kelompok

Diperoleh nilai Press's Q = 13,33 > $\chi^2_{0,05;1}$ =3,84 maka *H0* pada taraf signifikansi α =5%, Pengklasifikasian dengan Naive Bayes akurat.

4.5. Perbandingan Ketepatan Klasifikasi Metode Analisis Diskriminan Fisher dan Metode Naive Bayes.

Semakin besar nilai akurasi maka semakin tepat hasil klasifikasinya. Tabel.1 menunjukkan hasil perbandingan ketepatan klasifikasi dari metode Analisis Diskriminan Fisher dan Naive Bayes.

Tabel 1. Perbandingan Ketepatan Klasifikasi

Training : Testing	Analisis Diskriminan Fisher	Naive Bayes
80% : 20%	90%	83,33%

Dari Tabel 1 dapat diperoleh bahwa Analisis Diskriminan Fisher mempunyai tingkat keakurasian sebesar 90% dan Naive Bayes mempunyai tingkat keakurasian 83,33%. Untuk mengetahui manakah yang lebih baik diantara kedua metode tersebut maka dilakukan uji perbedaan dua proporsi. Dari perhitungan diperoleh nilai Z_{hitung}=0,762949 < $Z_{0,05/2}$=1,96 maka pada taraf signifikansi α =5%, PR_1= PR_2 (Tidak ada perbedaan tingkat akurasi antara metode Analisis Diskriminan Fisher dan metode Naive Bayes).

5. KESIMPULAN

Berdasarkan Pembahasan mengenai klasifikasi pada debitur Koperasi Jateng Amanah Mandiri Cabang Sukorejo Kendal dapat diperoleh kesimpulan bahwa:

1. Klasifikasi dengan analisis Diskriminan Fisher, diperoleh model $y = \boldsymbol{b^t x} = -0{,}000005740395\,(X1) + \left(-0{,}1976804\,(X2)\right) + 2{,}220286(X3) + 0{,}000001253012\,(X4)$ dan diperoleh nilai batasan klasifikasi (m) = -2,33137. Jika nilai y dari seorang debitur lebih besar atau sama dengan m maka debitur tersebut masuk ke dalam kelompok 1 (macet), jika kurang dari m maka masuk dalam kelompok 2 (lancar)
2. Penelitian dengan 120 data training dan 30 data testing diperoleh akurasi masing-masing 90 % untuk Analisis Diskriminan Fisher dan 83,33% untuk Naive Bayes. Berdasarkan uji beda dua proporsi diperoleh kesimpulan tidak ada perbedaan tingkat akurasi antara metode Analisis Diskriminan Fisher dan Naive Bayes.

6. DAFTAR PUSTAKA

[1] Hair, J. F. , Black, W. C. , Babin, B. J. , et al. 2006. *Multivariate Data Analysis.* Sixth Ed., Prentice Hall. New Jersey.

[2] John G.H. dan Langley P. 1995. *Estimating Continuous Distributions in Bayesian Classifiers. UAI 95 Proceedings of the Eleventh Conference on Uncertainity in Artificial Intelegence.*

[3] Johnson, R. A. dan Wichern, D. W. 2007. *Applied Multivariate Statistical Analysis.* Sixth Ed. , Prentice Hall. New Jersey.

[4] Kasmir. 2005. *Pemasaran Bank.* Jakarta: Kencana.

[5] Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi Menggunakan Matlab.* Yogyakarta: ANDI Yogyakarta.

[6] Sugiarto, D.S. 2000. *Metode Statistika.* Jakarta:gramedia pustaka utama

[7] Sugiyono. 2012. Statistika Untuk Penelitian. Bandung: ALFABETA.