

PERBANDINGAN ANALISIS DISKRIMINAN DAN *NAÏVE BAYES* DALAM PENGKLASIFIKASIAN STATUS PENERIMA BANTUAN PROGRAM KELUARGA HARAPAN DI NTB

LISA HARSYIAH*, MUSTIKA HADIJATI, NURUL FITRIYANI

*Program Studi Statistika, FMIPA Universitas Mataram
Jl Majapahit No. 62, Mataram, NTB, Indonesia, 83125,
email : lisa_harsyiah@unram.ac.id*

Diterima 18 Maret 2024 Direvisi 12 Juli 2024 Dipublikasikan 31 Oktober 2024

Abstrak. Permasalahan dalam penyaluran bantuan sosial Program Keluarga Harapan (PKH) adalah ketidaktepatan penyaluran bantuan PKH. Upaya yang dapat dilakukan untuk mengatasi permasalahan tersebut adalah dengan memastikan kriteria penerima bantuan PKH sudah benar dan sesuai dengan kriteria KPM. Berdasarkan kriteria KPM, perlu dilakukan klasifikasi status rumah tangga penerima bantuan PKH dan yang tidak. Hal ini dilakukan dengan tujuan untuk mengetahui apakah bantuan sosial PKH yang disalurkan tepat sasaran atau tidak. Proses klasifikasi dapat dilakukan dengan menggunakan analisis diskriminan dan metode *Naïve Bayes*. Hasil penelitian menunjukkan bahwa ketika melakukan klasifikasi menggunakan analisis diskriminan terhadap status penerima bantuan PKH di NTB diperoleh tingkat kesalahan klasifikasi sebesar 24,5%. Sedangkan hasil klasifikasi menggunakan metode *Naïve Bayes* memperoleh tingkat kesalahan sebesar 27,6%. Hasil pengklasifikasian status penerima bantuan PKH dengan menggunakan kedua metode ini tergolong akurat dan analisis diskriminan memiliki kinerja yang lebih baik dibandingkan metode *Naïve Bayes* untuk kasus pengklasifikasian status penerima bantuan PKH di NTB.

Kata Kunci: Analisis Diskriminan, Klasifikasi, *Naïve Bayes*, Program Keluarga Harapan

Abstract. The problem in distributing PKH social assistance is the inaccuracy of distributing PKH assistance. Efforts that can be made to resolve this problem are to ensure that the criteria for receiving PKH assistance are correct and in accordance with the KPM criteria. Based on the KPM criteria, it is necessary to classify the status of households that receive PKH assistance and those that do not. This is done with the aim of identifying whether the PKH social assistance distributed is on target or not. The classification process can be carried out using discriminant analysis and *Naïve Bayes* methods. The results of the research show that when classifying using discriminant analysis for the status of PKH aid recipients in NTB, a classification error rate of 24.5% was obtained. Meanwhile, the classification results using the *Naïve Bayes* method obtained an error rate of 27.6%. The results of classifying the status of PKH aid recipients using these two methods are classified as accurate and discriminant

*Penulis korespondensi

analysis has better performance than the Naïve Bayes method for the case of classifying the status of PKH aid recipients in NTB.

Keywords: Discriminant Analysis, Classification, Naïve Bayes, PKH

1. Pendahuluan

Salah satu program yang ditujukan untuk meningkatkan kesejahteraan rakyat adalah bantuan sosial PKH. PKH merupakan Upaya pemerintah berbasis bantuan sosial yang bertujuan untuk mendukung percepatan penanganan kemiskinan di Indonesia. Keluarga Penerima Manfaat (KPM) PKH diberikan secara merata hampir ke seluruh wilayah di Indonesia. Salah satu provinsi yang sudah dapat merasakan dan menerima bantuan PKH adalah Nusa Tenggara Barat (NTB). Provinsi NTB sudah memiliki KPM PKH sejumlah 310.623 KPM pada tahun 2022. Hal ini dikarenakan NTB memiliki tingkat kemiskinan yang cukup tinggi sehingga terdapat banyak rumah tangga yang diberi bantuan sosial untuk meningkatkan kesejahteraan hidupnya.

Namun, permasalahan yang muncul dalam distribusi bantuan sosial PKH adalah ketidaktepatan pendistribusian bantuan PKH [1]. Salah satu solusi untuk menyelesaikan masalah ketidaktepatan pendistribusian ini adalah dengan memastikan kriteria penerima bantuan PKH sudah tepat serta sesuai dengan kriteria KPM. Berdasarkan kriteria KPM itulah maka perlu dilakukan klasifikasi status rumah tangga yang menerima bantuan PKH dan tidak. Hal ini dilakukan dengan tujuan agar teridentifikasi apakah bantuan sosial PKH yang didistribusi sudah tepat sasaran atau tidak.

Teknik pengklasifikasian dapat menggunakan beberapa metode diantaranya dengan menggunakan metode analisis diskriminan dan *Naïve Bayes*. Menurut Rofiq dkk. [2], kedua metode ini mampu mengklasifikasikan berdasarkan variabel yang mempengaruhinya. Adapun kelebihan analisis diskriminan adalah proses perhitungan yang lebih efisien. Sedangkan *Naïve Bayes* merupakan suatu metode pengklasifikasian yang digunakan dalam data mining [3]. Metode ini terbukti memiliki akurasi tinggi, komputasi yang cepat saat diaplikasikan ke dalam data yang besar. Kelebihan dari metode *Naïve Bayes* yaitu implementasinya yang tidak rumit sehingga cocok untuk mengevaluasi probabilitas bersyarat.

Penelitian terkait klasifikasi status gizi menggunakan analisis diskriminan dilakukan oleh Diana dkk. [4]. Tingkat ketepatan klasifikasi yang diperoleh pada penelitian ini sebesar 74%. Selanjutnya penelitian yang dilakukan oleh Adzy dkk. [5] tentang Algoritma *Naïve Bayes* untuk klasifikasi kelayakan penerima Bantuan Iuran Jaminan Kesehatan Pemerintah Daerah Kabupaten Sukabumi. Hasil penelitian menunjukkan bahwa ketepatan klasifikasi memiliki tingkat akurasi sebesar 96,83%. Adapun penelitian terkait PKH menggunakan metode *Weighted Naïve Bayes* dengan *Laplace Smoothing* telah dilakukan oleh Utami dan Devi [1]. Hasil klasifikasi menunjukkan nilai akurasi sebesar 91,67%. Berdasarkan beberapa penelitian tersebut, dapat dilihat bahwa belum terdapat penelitian terkait PKH yang membandingkan analisis diskriminan dan metode *Naïve Bayes*. Oleh karena itu, perlu dilakukan penelitian yang mengklasifikasi status penerima bantuan

PKH di Provinsi NTB. Berdasarkan keakuratan penelitian sebelumnya, perlu dilakukan penelitian yang mengklasifikasi status penerima bantuan PKH di NTB menggunakan metode Analisis Diskriminan dan *Naïve Bayes* sehingga diperoleh metode klasifikasi yang lebih akurat untuk studi kasus tersebut.

2. Metode Penelitian

Penelitian ini memanfaatkan dua teknik klasifikasi, yakni analisis diskriminan dan *Naïve Bayes*. Data yang diperoleh untuk penelitian ini bersumber dari Badan Pusat Statistik (BPS) hasil Survei Sosial dan Ekonomi (SUSENAS) pada wilayah Nusa Tenggara Barat tahun 2022 yang bisa diakses melalui <https://silastik.bps.go.id>. Variabel yang digunakan dalam penelitian ini adalah sebanyak 12 variabel independen yang mempengaruhi status penerima bantuan PKH.

Adapun langkah pengklasifikasian menggunakan analisis diskriminan adalah sebagai berikut [5]:

(1) Uji asumsi diskriminan.

- (a) Uji distribusi normal multivariat. Dalam analisis multivariat, uji normalitas dapat dilakukan pada masing-masing variabel secara terpisah. Jika setiap variabel secara individu memenuhi asumsi normalitas, maka secara keseluruhan (multivariat) variabel-variabel tersebut dapat dianggap memenuhi asumsi normalitas. Pengujian kenormalan dapat dilakukan dengan menghitung nilai jarak kuadrat setiap observasi menggunakan rumus berikut.

$$d_i^2 = (X_i - \bar{X})^T S^{-1} (X_i - \bar{X}). \quad (2.1)$$

Langkah berikutnya dibuat plot antara d_i^2 dengan q_i . Nilai q_i dihitung menggunakan persamaan berikut:

$$q_i = \left(\frac{i - 0.5}{n} \right). \quad (2.2)$$

Apabila hasil plot mendekati garis lurus maka dapat disimpulkan bahwa data berdistribusi normal [6].

- (b) Uji homogenitas matriks kovarian dilakukan menggunakan uji Box's M. Formula untuk uji Box's M adalah sebagai berikut [7]:

$$\text{Statistik Uji: } \chi^2 = MC^{-1}, \quad (2.3)$$

dengan:

$$\begin{aligned} M &= \sum_{k=1}^g n_k - 1 \ln |\mathbf{S}_G| - \sum_{k=1}^g n_k - 1 \ln |\mathbf{S}_k|, \\ \mathbf{S}_G &= \frac{1}{\sum_{k=1}^g n_k} \sum_{k=1}^g n_k \mathbf{S}_k, \\ C^{-1} &= 1 - \frac{2p^2 + 3p - 1}{6(p+1)} \left(\sum_{i=1}^g \frac{1}{n_k - 1} - \frac{1}{\sum_{i=1}^g n_k - 1} \right) \end{aligned}$$

Hipotesis nol ditolak jika $\chi^2 > \chi_{\alpha; \frac{1}{2}(k-1)p(p+1)}^2$.

- (2) Membentuk fungsi diskriminan.
Terdapat dua metode dalam pembentukan fungsi diskriminan yaitu *simultaneous method* dan *stepwise method* [8].
- (3) Menentukan signifikansi variabel diskriminan.
Tahap ini dilakukan untuk melihat variabel penelitian dapat digunakan untuk menguji apakah terdapat perbedaan antar kelompok. Kriteria keputusan untuk uji signifikansi variabel adalah H_0 ditolak jika nilai *p-value* kurang dari 0.05.
- (4) Uji signifikansi Fungsi Diskriminan. [9]
Pengujian fungsi diskriminan dilakukan dengan menggunakan uji Wilks Lambda sebagai berikut [10]:

$$U = \frac{|W|}{|B + W|}. \quad (2.4)$$

Hipotesis H_0 ditolak jika $U \leq F_{p,(n-p-1)}$.

- (5) Menghitung ketepatan klasifikasi.
Perhitungan ketepatan klasifikasi dapat dilakukan menggunakan nilai *Apparent Error Rate* (APER) dengan rumus sebagai berikut [11]:

$$APER = \frac{n_{01} + n_{10}}{n_0 + n_1}. \quad (2.5)$$

Sedangkan langkah pengklasifikasian menggunakan metode Naïve Bayes adalah sebagai berikut.

- (1) Pembagian dataset menjadi *data training* dan *data testing*. Pembagian *data training* dan *data testing* menggunakan ketentuan yaitu antara 60%-90% untuk *data training* dan sisanya untuk *data testing*. Ketentuan pembagian ini berdasarkan acuan pada pengklasifikasian penelitian sebelumnya [12].
- (2) Dilakukan proses Klasifikasi Naïve Bayes pada *data training*. Langkah ini melibatkan perhitungan probabilitas prior dan posterior. Total jumlah dan probabilitas setiap kategori dan label dihitung untuk menghasilkan model terbaik yang digunakan dalam klasifikasi data pengujian.
- (3) Prediksi dilakukan menggunakan model yang telah dibangun dengan menggunakan data pengujian. Proses ini dieksekusi menggunakan *software*, seperti pada persamaan berikut [13]:

$$p(B_r|A) = \frac{P(B_r)P(A|B_r)}{\sum_{i=1}^k P(B_i)P(A|B_i)}. \quad (2.6)$$

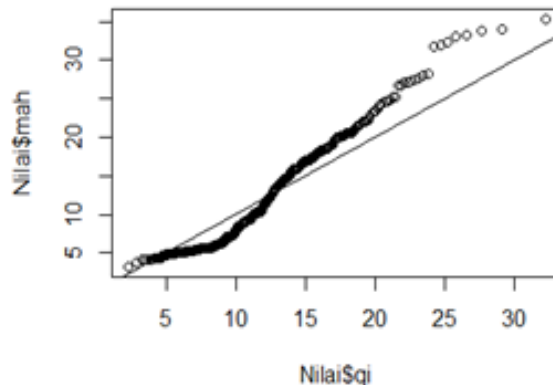
- (4) Diperoleh tabel *confusion matrix* dari data pengujian, dan kemudian dilakukan penilaian ketepatan dan kinerja klasifikasi. Hasil evaluasi ini digunakan untuk menyimpulkan hasil dari klasifikasi tersebut.

3. Hasil dan Pembahasan

3.1. Uji Normal Multivariat

Uji normalitas dilakukan dengan menggunakan persamaan (2.1) dan (2.2). Jarak kuadrat untuk setiap observasi ke-*i* yang telah dihitung dengan menggunakan

persamaan (2.1) kemudian hasilnya diurutkan dari yang terkecil hingga yang terbesar. Selanjutnya dibuat plot antara d_i^2 pada persamaan (2.1) dengan q_i pada persamaan (2.2). Pada Gambar 1 diberikan plot uji normalitas data yang diperoleh.



Gambar 1. Uji Normal Multivariat Data

Dapat dilihat pada Gambar 1 bahwa plot tersebut memiliki titik-titik yang cenderung mendekati garis lurus sehingga asumsi normalitas terpenuhi.

3.2. Uji Homogenitas Matriks Varian Kovarians

Berdasarkan proses perhitungan menggunakan uji Boxs M diperoleh nilai yang tersaji pada Tabel 1.

Tabel 1. Hasil Uji Homogenitas Matriks Varians Kovarians

Uji	χ^2_{hitung}	Derajat bebas	P-value
Boxs M	461.0294	78	2.210^{-16}

Dari hasil pengujian diperoleh nilai Boxs M sebesar 461.0294 lebih dari nilai χ^2_{table} sebesar 99.6169. Dapat disimpulkan bahwa H_0 ditolak, yang mengartikan bahwa matriks varians kovarians antar grup tidak homogen. Akan tetapi menurut Johnson dan Wichern [10], asumsi kehomogenan matriks varian kovarian di dalam praktiknya sering dilanggar karena asumsi homogenitas matriks varians-kovarians tidak terlalu sensitif mempengaruhi model Analisis diskriminan [14].

3.3. Estimasi Parameter Fungsi Diskriminan

Adapun nilai estimasi parameter fungsi diskriminan tersaji pada Tabel 2.

Berdasarkan Tabel 2 terlihat bahwa ada delapan variabel yang berpengaruh positif, sedangkan empat variabel lainnya berpengaruh negatif sehingga model

Tabel 2. Nilai Estimasi Parameter Fungsi Diskriminan

Parameter	Estimasi	Parameter	Estimasi
a	-0.3430	W_7	0.6787
W_1	1.0636	W_8	0.5167
W_2	-0.4703	W_9	-0.0961
W_3	-0.0766	W_{10}	-0.1783
W_4	0.2405	W_{11}	0.4485
W_5	0.1925	W_{12}	0.6870
W_6	0.2634		

diskriminan yang terbentuk adalah:

$$\begin{aligned}
 Z = & -0.3430 + 1.0636X_1 - 0.4703X_2 - 0.0766X_3 + 0.2405X_4 + 0.1925X_5 \\
 & + 0.2634X_6 + 0.6787X_7 + 0.5167X_8 - 0.0961X_9 - 0.1783X_{10} + 0.4485X_{11} \\
 & + 0.6870X_{12}.
 \end{aligned}$$

3.4. Uji Signifikansi Variabel Diskriminan

Uji signifikansi variabel bertujuan untuk melihat apakah terdapat perbedaan kelompok dengan melibatkan variabel-variabel terkait. Adapun hasil perhitungan diberikan pada Tabel 3.

Tabel 3. Uji Signifikansi Variabel Diskriminan

Variabel	Wilks Lambda	P-value
Jumlah anggota rumah tangga	0.990	0.041
Status kepemilikan rumah tinggal	0.977	0.002
Jenis atap rumah	1.000	0.695
Jenis dinding rumah	0.983	0.008
Jenis lantai rumah	0.959	0.000
Sumber air utama	0.938	0.000
Kepemilikan kulkas	0.959	0.000
Kepemilikan komputer/laptop	0.974	0.001
Kepemilikan emas	0.978	0.003
Kepemilikan motor	0.997	0.308
Kepemilikan mobil	0.973	0.001
Kepemilikan televisi	0.956	0.000

Berdasarkan Tabel 3, dapat dilihat bahwa terdapat sepuluh variabel dengan nilai *p-value* kurang dari 0.05 dan nilai Wilks Lambda mendekati angka nol. Hasil ini mengartikan bahwa sepuluh variabel tersebut berpengaruh secara signifikan terhadap status penerima bantuan sosial PKH di NTB. Sedangkan variabel lainnya tidak berpengaruh secara signifikan.

3.5. Uji Signifikansi Fungsi Diskriminan

Statistik uji yang digunakan adalah Wilks Lambda, atau bisa disebut statistik U . Berdasarkan hasil perhitungan diperoleh hasil uji signifikansi fungsi diskriminan seperti pada Tabel 4.

Tabel 4. Uji Signifikansi Fungsi Diskriminan

Wilks Lambda	Derajat Bebas 1	Derajat Bebas 2	P-value
0.8558	12	387	8.43510^{-7}

Berdasarkan Tabel 4 dapat dilihat bahwa nilai Wilks Lambda sebesar 0.8558 dan nilai F_{tabel} sebesar 1.7772. Ini mengartikan bahwa Wilks Lambda $< F_{tabel}$ sehingga H_0 ditolak, dan dapat diambil kesimpulan bahwa fungsi diskriminan yang diperoleh signifikan.

3.6. Ketepatan Klasifikasi

Pengujian keakuratan klasifikasi dihitung dengan menggunakan *Apparent Error Rate* (APER). Pada Tabel 5 diberikan klasifikasi analisis diskriminan.

Tabel 5. Tabel Klasifikasi Analisis Diskriminan

<i>Actual Group</i>	<i>Predicted Group</i>	
	0	1
0	15	86
1	12	287

Berdasarkan Tabel 5 dapat dihitung nilai APER sebagai berikut.

$$APER = \frac{n_{01} + n_{10}}{n_0 + n_1} = \frac{86 + 12}{98 + 302} = 0.2450.$$

Nilai APER yang diperoleh dari hasil perhitungan menyatakan bahwa fungsi diskriminan yang terbentuk memiliki tingkat kesalahan sebesar 0.2450 atau 24.5%.

3.7. Perhitungan Probabilitas Prior

Probabilitas prior adalah probabilitas berdasarkan informasi yang tersedia (sebelum ada tambahan informasi). Berdasarkan 360 *data training* status penerima bantuan sosial PKH, terdapat sejumlah status penerima yang diklasifikasikan ke dalam status menerima dan tidak menerima. Berikut contoh perhitungan probabilitas status klasifikasi menerima, yaitu:

$$P(\text{Tidak Menerima}) = \frac{n(\text{Tidak Menerima})}{n(\text{training})} = \frac{279}{360} = 0.775.$$

Hasil perhitungan untuk setiap kelas status klasifikasi di *data training* dapat dilihat pada Tabel 6.

Tabel 6. Perhitungan Probabilitas Prior

Status Klasifikasi	Jumlah Keluarga	Probabilitas Status
Menerima	81	0.225
Tidak Menerima	279	0.775
Total	360	1

3.8. Probabilitas Kategori terhadap Masing-masing Kelas ($P(X_i|Y_r)$)

Probabilitas status klasifikasi dengan jumlah anggota rumah tangga dapat dihitung dengan cara sebagai berikut.

$$P(> 4|Menerima) = \frac{P(Menerima \cap > 4)}{P(Menerima)} = \frac{\frac{19}{360}}{\frac{81}{360}} = 0.234.$$

Adapun probabilitas untuk setiap status klasifikasi status penerima PKH diberikan pada Tabel 7.

Tabel 7. Probabilitas Status Klasifikasi dengan Jumlah Anggota Rumah Tangga

Jumlah Anggota Rumah Tangga	Jumlah Status Klasifikasi		Probabilitas Status Klasifikasi	
	Tidak Menerima	Menerima	Tidak Menerima	Menerima
> 4	40	19	0.143	0.234
≤ 4	239	62	0.856	0.765
Total	279	81	1	1

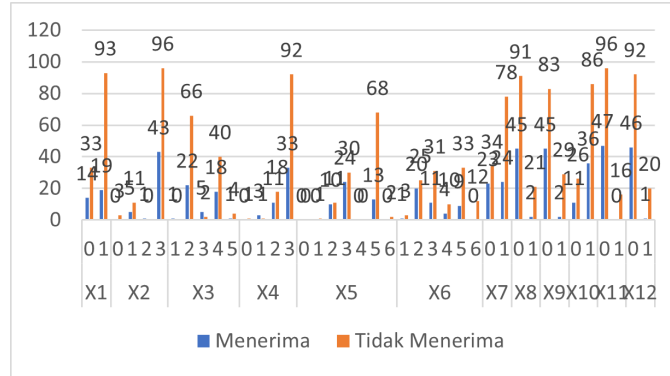
Adapun perhitungan probabilitas status klasifikasi dengan variabel lainnya dapat dilakukan seperti cara di atas.

3.9. Perhitungan Data Testing dengan Metode Naïve Bayes

Data testing yang digunakan berjumlah 159 data. Jumlah data testing dan grafik untuk setiap variabel dalam data testing yang diklasifikasikan ke masing-masing status klasifikasi secara berturut-turut disajikan dalam Tabel 8 dan Gambar 2 berikut.

Tabel 8. Data Testing

Jumlah Anggota Rumah Tangga	Jumlah
>4	47
$\leq$	112
Total	159



Gambar 2. Data Testing Status Penerima Bantuan PKH

3.10. Perhitungan Data Testing dengan Metode Naïve Bayes

Setelah diketahui sebaran data terhadap masing-masing status klasifikasi di atas, hal yang dilakukan selanjutnya adalah melakukan prediksi status klasifikasi pada *data testing*.

- (1) Probabilitas status klasifikasi status penerima bantuan PKH.

Probabilitas tiap status klasifikasi status penerima bantuan PKH dapat dilihat pada Tabel 1.

$$P(\text{Tidak Menerima}) = 0.775,$$

$$P(\text{Menerima}) = 0.225.$$

- (2) Probabilitas contoh kasus dengan tiap status klasifikasi.

Probabilitas status klasifikasi berdasarkan Tabel 7 diberikan sebagai contoh untuk jumlah anggota rumah tangga berdasarkan status klasifikasi, maka diperoleh:

$$P(X_{1(0)} = \text{Lebih besar dari 4} | \text{Menerima}) = 0.234,$$

$$P(X_{1(0)} = \text{Lebih besar dari 4} | \text{Tidak Menerima}) = 0.143.$$

Kalikan semua hasil variabel status klasifikasi menerima dan klasifikasi tidak menerima ($P(X_i|Y_r)$).

- (3)

$$\begin{aligned}
 &P(X_{1(0)} \cap X_{2(3)} \cap X_{3(4)} \cap X_{4(3)} \cap X_{5(3)} \cap X_{6(2)} \cap X_{7(0)} \cap X_{8(0)} \cap X_{9(0)} \cap \\
 &X_{10(1)} \cap X_{11(0)} \cap X_{12(0)} | Y_0) \\
 &= P(X_{1(0)} | Y_0) \times P(X_{2(0)} | Y_0) \times P(X_{3(0)} | Y_0) \times P(X_{4(0)} | Y_0) \times P(X_{5(0)} | Y_0) \\
 &\times P(X_{6(0)} | Y_0) \times P(X_{7(0)} | Y_0) \times P(X_{8(0)} | Y_0) \times P(X_{9(0)} | Y_0) \times P(X_{10(0)} | Y_0) \\
 &\times P(X_{11(0)} | Y_0) \times P(X_{12(0)} | Y_0) \\
 &= 0.234 \times 0.987 \times 0.370 \times 0.691 \times 0.345 \times 0.567 \times 0.691 \times 0.950 \times 0.925 \\
 &\times 0.654 \times 0.987 \times 1.000 \\
 &= 0.0045.
 \end{aligned}$$

(4)

$$\begin{aligned}
& P(X_{1(0)} \cap X_{2(3)} \cap X_{3(4)} \cap X_{4(3)} \cap X_{5(3)} \cap X_{6(2)} \cap X_{7(0)} \cap X_{8(0)} \cap X_{9(0)} \cap \\
& X_{10(1)} \cap X_{11(0)} \cap X_{12(0)} | Y_1) \\
& = 0.143 \times 0.835 \times 0.369 \times 0.788 \times 0.369 \times 0.258 \times 0.412 \times 0.856 \times 0.838 \\
& \times 0.702 \times 0.906 \times 0.810 \\
& = 0.0005.
\end{aligned}$$

(5) Perhitungan Probabilitas Gabungan ($\sum_{i=1}^k P(Y_i)P(X|Y_i)$).

Probabilitas gabungan suatu kasus dengan tiap variabel, maka dapat dilakukan perhitungan sebagai berikut:

$$\begin{aligned}
\sum_{i=1}^k P(Y_i)P(X|Y_i) &= P(Y_0)P(X|Y_0) + P(Y_1)P(X|Y_1) \\
&= P(\text{Menerima})P(X_{1(0)} \cap X_{2(3)} \cap X_{3(4)} \cap X_{4(3)} \\
& \cap X_{5(3)} \cap X_{6(2)} \cap X_{7(0)} \cap X_{8(0)} \cap X_{9(0)} \cap X_{10(1)} \cap X_{11(0)} \\
& \cap X_{12(0)} | \text{menerima}) + P(\text{Tidak Menerima})P(X_{1(0)} \cap X_{2(3)} \\
& \cap X_{3(4)} \cap X_{4(3)} \cap X_{5(3)} \cap X_{6(2)} \cap X_{7(0)} \cap X_{8(0)} \cap X_{9(0)} \\
& \cap X_{10(1)} \cap X_{11(0)} \cap X_{12(0)} | \text{Tidak menerima}) \\
&= 0.00103 + 0.000393 \\
&= 0.001424.
\end{aligned}$$

Jadi probabilitas gabungan kasus tersebut untuk setiap kasus klasifikasi status penerima bantuan sosial PKH adalah sebesar 0.001424.

(6) Tentukan status klasifikasi dengan teorema Bayes ($P(Y_r|X)$).

Langkah berikutnya adalah menentukan status klasifikasi dari data tersebut (*probabilitas posterior*) dengan cara sebagai berikut.

$$\begin{aligned}
& \frac{P(\text{Menerima})P(X_{1(0)} \cap X_{2(3)} \cap X_{3(4)} \cap X_{4(3)} \cap X_{5(3)} \cap X_{6(2)} \cap X_{7(0)} \cap X_{8(0)} \cap \\
& X_{9(0)} \cap X_{10(1)} \cap X_{11(0)} \cap X_{12(0)} | \text{Menerima})}{P(Y_0)P(X|Y_0) + P(Y_1)P(X|Y_1)} \\
&= \frac{0.225 \times 0.0045}{0.001424} = 0.711. \\
& \frac{P(\text{Tidak Menerima})P(X_{1(0)} \cap X_{2(3)} \cap X_{3(4)} \cap X_{4(3)} \cap X_{5(3)} \cap X_{6(2)} \cap X_{7(0)} \cap X_{8(0)} \cap \\
& X_{9(0)} \cap X_{10(1)} \cap X_{11(0)} \cap X_{12(0)} | \text{Tidak Menerima})}{P(Y_0)P(X|Y_0) + P(Y_1)P(X|Y_1)} \\
&= \frac{0.775 \times 0.0005}{0.001424} = 0.272.
\end{aligned}$$

Dari hasil simulasi perhitungan data pengujian di atas, dapat disimpulkan bahwa nilai probabilitas klasifikasi untuk penerimaan lebih tinggi daripada nilai probabilitas penolakan. Oleh karena itu, dapat disimpulkan bahwa data mengenai status penerima bantuan sosial PKH ke-35, yang mencakup jumlah anggota rumah tangga lebih dari empat orang, kepemilikan rumah sendiri dengan

atap berbahan kayu/bambu, dinding dari anyaman bambu, lantai menggunakan ubin/tegel/teraso, sumber air utama dari leding, tidak memiliki kulkas, tidak memiliki komputer/laptop, tidak memiliki emas, memiliki sepeda motor, tidak memiliki mobil, dan tidak memiliki televisi, setelah melalui proses perhitungan, diklasifikasikan sebagai penerima. Proses serupa juga dilakukan pada data pengujian lainnya.

3.11. *Confusion Matrix Data Testing*

Berikutnya, *data testing* diprediksi apakah klasifikasi sudah benar atau sebaliknya. Hasil perhitungan *confusion matrix* untuk data testing disajikan pada Tabel 9 berikut.

Tabel 9. *Confusion Matrix Data Testing*

Aktual	Prediksi		Total
	Menerima	Tidak Menerima	
Menerima	15	12	27
Tidak Menerima	32	100	132
Total	47	112	159

Berdasarkan Tabel 9 dihasilkan 15 data yang berhasil diklasifikasikan ke status menerima dan 100 data yang berhasil diklasifikasikan ke status tidak menerima. Dengan demikian dapat disimpulkan bahwa dari 159 data testing yang digunakan, sebanyak 115 data berhasil diklasifikasikan dengan benar.

3.12. *Ketepatan Klasifikasi*

Berikut pengujian ketepatan klasifikasi menggunakan APER pada persamaan (2.5).

$$APER = \frac{2 + 32}{47 + 112} = \frac{44}{159} = 0.276.$$

Berdasarkan nilai APER yang diperoleh yaitu 0.276 yang berarti bahwa kesalahan klasifikasi status penerima bantuan sosial PKH sebesar 27.6%.

4. Kesimpulan dan Saran

Hasil pengklasifikasian menggunakan analisis diskriminan untuk kasus status penerima bantuan PKH di Provinsi NTB diperoleh nilai APER sebesar 0.2450 yang artinya tingkat kesalahan klasifikasi sebesar 24.5%. Sedangkan hasil pengklasifikasian menggunakan metode *Naïve Bayes* diperoleh nilai APER sebesar 0.276 atau memiliki tingkat kesalahan sebesar 27.6%. Hasil klasifikasi status penerima bantuan PKH menggunakan kedua metode tersebut menunjukkan tingkat akurasi yang memuaskan. Berdasarkan nilai APER yang diperoleh, dapat disimpulkan bahwa analisis diskriminan memiliki kinerja yang lebih baik daripada metode *Naïve Bayes* dalam kasus klasifikasi status penerima bantuan PKH di NTB.

Berdasarkan kesimpulan yang yang diperoleh maka saran untuk penelitian selanjutnya untuk membandingkan metode analisis diskrimina dengan metode klasifikasi lainnya sehingga diperoleh tingkat kesalahan klasifikasi yang lebih rendah.

5. Ucapan Terima Kasih

Kami mengucapkan terima kasih kepada Universitas Mataram atas kesempatan dan fasilitas yang diberikan, yang memungkinkan pelaksanaan kegiatan penelitian ini. Penghargaan khusus kami sampaikan kepada tim PNBPM FMIPA Unram yang telah memberikan dana untuk mendukung kelancaran penelitian ini.

Daftar Pustaka

- [1] Utami, D., Devi, P, A, R., 2022, Klasifikasi Kelayakan Penerima Bantuan PKH Menggunakan Metode Weighted dengan Laplace Smoothing, *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)* Vol. **7**(4): 1373 – 1384
- [2] Rofiq, A., Wuryandari, T., Rahmawati, R., 2016, Perbandingan Analisis Diskriminan Fisher dan *Naïve Bayes* untuk Klasifikasi Risiko Kredit (Studi Kasus Debitur di Koperasi Jateng Amanah Mandiri Cabang Sukorejo Kendal), *Jurnal Gaussian* Vol. **5**(1): 1 – 10
- [3] Kartika, S., Sairam, N., 2015, A *Naïve Bayesian* Classifier for Educational Qualification, *Indian Journal of Science and Technology* Vol. **8**(16): 1 – 5
- [4] Diana., Susilawati, M., Sukarsa, I, K, G., Kencana, I, P, E, N., 2023, Pengklasifikasian Status Gizi Balita Menggunakan Analisis Diskriminan, *E-Jurnal Matematika* Vol. **12**(1): 22 – 30
- [5] Adzy, L. B., Asriyanik., dan Pambudi, A, 2023, Algoritma *Naïve Bayes* untuk Klasifikasi Kelayakan Penerima Iuran Jaminan Kesehatan Pemerintah Daerah Kabupaten Sukabumi, *Jurnal MNEMONIC* Vol. **6**(1): 1– 10
- [6] Maulidya., Susanto. H. T., Oktaviarina. A., 2015, Perbandingan Analisis Diskriminan dan Regresi Logistik (Studi Kasus Klasifikasi Konsumen Berdasarkan Tempat Berbelanja di Wilayah Taman Sidoarjo), *MATHunesa* Vol. **3**(1): 1 – 8
- [7] Tjahaya, M.S., Raupong, Tinungki, G. M., 2022, Analisis Diskriminan Linear Robust dengan Metode Winsorized One-Step M-Estimator, *Estimasi Journal of Statistics and Its Applications* Vol. **3**(1): 1 – 13
- [8] Ndangi, W. R. A., Resmawan., Djakaria, I., 2019, Perbandingan Analisis Diskriminan Dan Regresi Logistik Multinomial, *Jambura Journal Of Mathematics*, **1**(2): 54 – 63
- [9] Flury, Riedyl, 1998, *Multivariate Statistics A Practical Approach*, London: Chapman and Hall
- [10] Johnson, R. A., Wichern D. W., 2015, *Applied Multivariate Statistical Analysis*, Prentice Hall, Inc. 2015, New Jersey
- [11] Ananda, R. F., Harsyiah, L., Alfian, M. R., 2023, Classification of Perceptions of The Covid-19 Vaccine Using Multivariate Adaptive Regression Spline, *Jurnal Varian* Vol. **6**(2): 137 – 148
- [12] Han, J., Kamber, M., Pei, J., 2006, *Data Mining: Concepts and Techniques*, 2nd Edition, Waltham: Morgan Kaufmann Publishers
- [13] Siregar, V. M., 2017, Perancangan Sistem Pendukung Keputusan Penentuan

Jurusan Siswa/I SMA Swasta Binaguna Tanah Jawa dengan Metode *Naïve Bayes*, *Prosiding SenNasMUDI*: 66 – 75

- [14] Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., 2014, *Multivariate Data Analysis*, Seventh Edition, Pearson Prentice Hall, New Jersey
- [15] Lembang, F. K., Talukua, M. W., Hasanudin, M. S., 2019, Misklasifikasi Penjurusan Mahasiswa Fmipa Universitas Pattimura Tahun Akademik 2016/2017 Menggunakan Metode Analisis Diskriminan Berganda, *Journal of Statistics and Its Applications* Vol. 1(2): 64 – 74