UNP JOURNAL OF STATISTICS AND DATA SCIENCE
Volume I Nomor 2 Tahun 2023 Halaman 82-89

ISSN(*Online*): 2985-475X

Comparison of Naive Bayes Method and Binary Logistics Regression on Classification of Social Assistance Recipients Program Keluarga Harapan (PKH)

Fanni Rahma Sari, Fadhilah Fitri*, Atus Amadi Putra, Dony Permana
Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia
*Corresponding author: fadhilahfitri@fmipa.unp.ac.id

Submitted : 09 November 2022
Revised : 09 Januari 2023
Accepted : 10 Februari 2023

ABSTRACT

Population density is one cases of economic inequality in Indonesia. One of the solutions provided by the government is to distribute social assistance. In 2007 the government created a social assistance program called the "Program Keluarga Harapan" (PKH) with the aim of alleviating poverty. There are several problems in the distribution of social assistance, one of which is receiving aid that is not right on target. Therefore, an appropriate method is needed in classifying the recipients of social assistance properly. This study will use two methods, namely Naive Bayes and Binary Logistic Regression to compare which method is better on the data used. The data used is the DTKS data for PKH assistance recipients in the Anduring Village in 2020. The accuracy results obtained from the Binary Logistic Regression is 75% and Naive Bayes method is 70%. So the best method in measuring classification is Binary Logistic Regression.

Keywords: *Social Assistance, PKH, Classification, Binary Logistic Regression, Naive Bayes*

I. PENDAHULUAN

Indonesia adalah negara berkembang yang memiliki jumlah penduduk terbanyak ke-empat di dunia. Melihat hal tersebut akan sulit untuk mendapatkan kesetaraan ekonomi masyarakat di Indonesia nantinya. Pemerintah terus melakukan upaya dalam mencapai kesetaraan tersebut dengan melaksanakan berbagai program salah satunya bantuan sosial. Bantuan sosial merupakan suatu program pemberian bantuan kepada masyarakat secara selektif berupa uang atau barang yang bertujuan untuk meningkatkan kesejahteraan masyarakat. Penyediaan anggaran dana bantuan sosial harus dijabarkan dengan rinci agar penerimaan dan sasaran penggunaannya jelas serta tepat sasaran. Pada tahun 2007 pemerintah membuat suatu program bantuan sosial yang bernama Program Keluarga Harapan (PKH), program ini bertujuan untuk memutus rantai kemiskinan dan meningkatkan kesejahteraan serta kualitas sumber daya manusia (Nataya & Supriyadi, 2017). Tujuan dari PKH adalah mengatasi kemiskinan dan kelaparan yang terjadi melalui akses pendidikan dan kesehatan. Penerima bantuan PKH yang telah terdaftar berasal dari Data Terpadu Kesejahteraan Sosial (DTKS) yaitu basis data yang digunakan untuk penyaluran bantuan sosial PKH. Menurut Peraturan Menteri Sosial Nomor 3 Tahun 2021, DTKS adalah sekumpulan data informasi pelayanan kesejahteraan sosial yang berisi daftar penerima bantuan sosial.

Pada tahun 2022 ditemukan kesalahan penyaluran bansos dari pemerintah oleh Badan Pemeriksa Keuangan (BPK) yang mengakibatkan kerugian mencapai Rp 6,9 triliun. Dari hasil Laporan Hasil Pemeriksaan (LHP) tahun 2021 menyebutkan bahwa kesalahan penyaluran bansos terjadi pada beberapa program bantuan sosial salah satunya Program Keluarga Harapan (PKH). Melihat hal tersebut diperlukan suatu metode yang dapat mengukur akurasi dari data klasifikasi penerima bantuan sosial dengan tepat dan akurat. Tujuannya yaitu untuk mempermudah pihak yang berwenang dalam membagikan bantuan agar tepat sasaran dan sampai kepada masyarakat sesuai yang diharapkan.

Sementara itu, klasifikasi adalah metode pengelompokkan data yang disusun dengan sistematis. Klasifikasi data dalam jumlah besar dan beragam akan menghasilkan tingkat akurasi yang rendah, maka diperlukan metode yang dapat mengatasi hal tersebut yaitu metode *Naive Bayes* dan Regresi Logistik (Salim, 2017). *Naive Bayes* adalah suatu metode klasifikasi yang sederhana dan mudah untuk diterapkan sehingga sangat efektif ketika diuji ke dalam data terutama

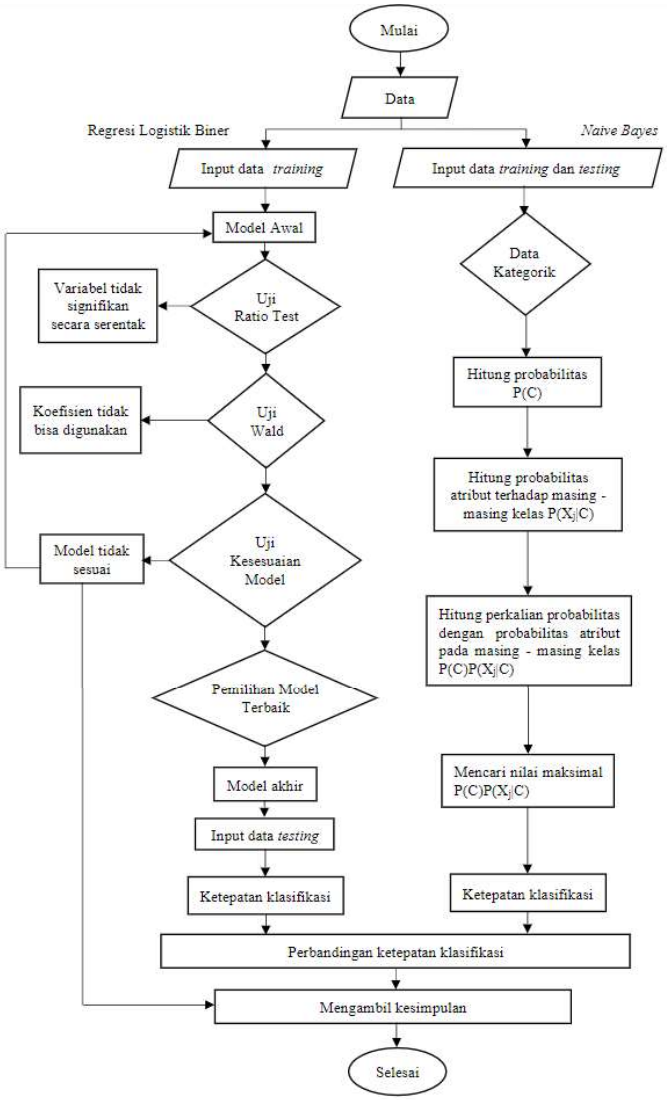
DOI: 10.24036/ujsds/vol1-iss2/24

ISSN(*Online*): 2985-475X

menggunakan seleksi fitur (Witten, dkk, 2011). Menurut Han and Kamber (2006) klasifikasi menggunakan *Naive Bayes* telah terbukti memiliki tingkat akurasi dan ketepatan yang tinggi ketika diaplikasikan ke dalam data berjumlah besar. Klasifikasi dengan menggunakan Regresi Logistik adalah suatu metode pengklasifikasian dengan setiap pengamatan memiliki variabel independen yang berkaitan dengan variabel dependen. Metode ini efektif dalam pengklasifikasian data dan mampu menjelaskan faktor-faktor yang berpengaruh terhadap variabel dependen dengan baik (Samosir, 2015).

II. METODE PENELITIAN

Penelitian ini menggunakan data sekunder yang didapatkan dari Kelurahan Anduring dengan menggunakan data DTKS penerima bantuan sosial tahun 2020 yang berjumlah 629 orang. Data yang digunakan dibagi menjadi data *training* dan data *testing* masing-masing sebesar 80% dan 20%. Variabel pada data ini terdiri dari variabel dependen (Y) dan independen (X). Variabel dependen dikategorikan sebagai "tidak menerima" dan "menerima"dan variabel independen terdiri dari 15 variabel diantaranya status bangunan, kondisi dinding, kondisi atap, sumber air minum, sumber penerangan, bahan bakar masak, ketersediaan tabung gas, ketersediaan lemari es, ketersediaan AC, ketersediaan telepon, ketersediaan TV, ketersediaan emas, ketersediaan motor, ketersediaan mobil, dan ketersediaan aset tak bergerak. Pengolahan data menggunakan bantuan *software R studio*. Berikut *flowchart* untuk melihat tahapan analisis data pada Gambar 1.



Gambar 1. *Flowchart* Tahapan Analisis Data

DOI: 10.24036/ujsds/vol1-iss2/24

ISSN(*Online*): 2985-475X

III. HASIL DAN PEMBAHASAN

A. Regresi Logistik Biner

Berikut adalah langkah-langkah dalam menganalisis data menggunakan metode Regresi Logistik Biner manggunakan data *training*.

a. Model Regresi Logistik Biner

Menurut Hosmer dan Lemeshow (2000), model Regresi Logistik merupakan model yang digunakan dengan menggunakan besaran $\pi(x) = E(Y = 1|x)$ untuk mewakili rata-rata *kodisional* dari Y ketika digunakan adalah sebagai berikut.

$$\pi(x) = \frac{exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki})}{1 + exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki})}$$

Model regresi terbaik dipilih dengan menggunakan metode AIC (*Akaike Information Criterion*). Metode ini dilakukan dengan mengevaluasi semua kemungkinan regresi yang dapat dibuat berdasarkan nilai atau kriteria tertentu salah satunya berdasarkan nilai AIC terkecil. Setelah melakukan pengujian, diperoleh niai AIC terkecil yaitu 560,330 maka didapatkan model terbaik sebagai berikut.

$$\pi(x) = \frac{\exp(64{,}226 - 0{,}270 - 0{,}842 - 0{,}348 + 0{,}001 + 0{,}048 - 0{,}121 - 16{,}001 - 0{,}311 - 0{,}234 - 14{,}914 - 0{,}161 + 0{,}241 + 0{,}128 + 0{,}389 + 0{,}745)}{1 + \exp(64{,}226 - 0{,}270 - 0{,}842 - 0{,}348 + 0{,}001 + 0{,}048 - 0{,}121 - 16{,}001 - 0{,}311 - 0{,}234 - 14{,}914 - 0{,}161 + 0{,}241 + 0{,}128 + 0{,}389 + 0{,}745)}$$

b. Uji G (*Likelihood Ratio Test*)

Uji ini bertujuan untuk mengetahui pengaruh variabel independen terhadap variabel dependen dengan melihat nilai G yang didapatkan. Nilai ini diperoleh dari L_0 (*Likelihood* tidak dengan variabel independen) dan L_1 (*Likelihood* dengan variabel independen) yang berdistribusi χ^2 (*Chi Square*) dengan derajat bebas p yang didefinisikan sebagai berikut (Hosmer dan Lameshow, 2000).

$$G = -2\left[\frac{L_0}{L_1}\right] = -2ln\frac{\left(\frac{n_1}{n}\right)n_1\left(\frac{n_0}{n}\right)n_o}{\sum_{i=1}^{n}\hat{\pi}_i{}^{y_i}(1-\hat{\pi}_i)^{(1-y_i)}}$$

Setelah melakukan perhitungan, diperoleh hasil dari nilai hitung G sebesar 39,86 dengan nilai tabel *Chi Square* yaitu 24,99. Berdasarkan hasil tersebut, disimpulkan nilai hitung G lebih besar dibandingkan nilai tabel *Chi Square*, maka terjadi tolak H_0. Artinya bahwa variabel independen memberikan pengaruh terhadap variabel dependen.

c. Uji *Wald*

Uji ini dilakukan untuk mengetahui variabel yang berpengaruh signifikan antara satu variabel independen dengan variabel dependen dengan melihat nilai *P-value* dari masing-masing variabel.

Tabel 1. Uji Wald

Variabel	*P-value*	Keputusan
X_1	0,078	Tidak Signifikan
X_2	0,010	Signifikan
X_3	0,309	Tidak Signifikan
X_4	0,959	Tidak Signifikan
X_5	0,920	Tidak Signifikan
X_6	0,284	Tidak Signifikan
X_7	0,983	Tidak Signifikan
X_8	0,200	Tidak Signifikan
X_9	0,838	Tidak Signifikan
X_{10}	0,987	Tidak Signifikan
X_{11}	0,606	Tidak Signifikan

DOI: 10.24036/ujsds/vol1-iss2/24

ISSN(*Online*): 2985-475X

X_{12}	0,641	Tidak Signifikan
X_{13}	0,593	Tidak Signifikan
X_{14}	0,747	Tidak Signifikan
X_{15}	0,010	Signifikan

Berdasarkan nilai pada Tabel 1 diperoleh bahwa variabel X_2 dan X_{15} berpengaruh secara signifikan, hal tersebut dikarenakan nilai *P-value* < 0,05 dan variabel X_1, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{14} tidak berpengaruh secara signifikan karena didapatkan nilai *P-value* > 0,05. Disimpulkan bahwa variabel kondisi dinding dan ketersediaan aset tak bergerak berpengaruh signifikan dan variabel lainnya tidak berpengaruh signifikan terhadap variabel dependen.

d. Uji Kesesuaian Model

Pengujian kesesuaian terhadap model dilakukan untuk mengetahui model yang diperoleh telah sesuai atau tidak terhadap data. Statistik uji yang digunakan adalah Uji Hosmer dan Lemeshow (Hosmer dan Lemeshow, 2000).

$$\hat{C} = \sum_{k=1}^{g} \frac{(o_k - n\hat{\pi}_k)^2}{n_k\hat{\pi}_k(1-\hat{\pi}_k)}$$

Setelah melakukan perhitungan, diperoleh nilai hitung *Chi Square* sebesar 12.419 dengan nilai tabel *Chi Square* yaitu 15.507. Dari hasil tersebut disimpulkan bahwa nilai hitung *Chi Square* lebih besar dari nilai tabel *Chi Square*. Artinya bahwa model yang didapat telah sesuai dengan data yang digunakan.

e. *Odds Ratio*

Pengujian *Odds Ratio* bertujuan untuk mengetahui kecenderungan terjadinya suatu kejadian.

Tabel 2. *Odds Ratio*

Variabel	Exp (B)
X_2	4,304
X_{15}	2,106

Berdasarkan Tabel 2 diperoleh informasi bahwa

a. Variabel kondisi dinding diketahui memiliki nilai *odds ratio* sebesar 4,304. Dari angka tersebut dijelaskan bahwa variabel kondisi dinding "layak" berpeluang menerima bantuan PKH sebesar 4,304 kali daripada kondisi dinding "tidak layak".

b. Variabel ketersediaan aset tak bergerak diketahui memiliki nilai *odds ratio* sebesar 2,106. Dari angka tersebut dijelaskan bahwa yang memiliki asset tak bergerak berpeluang menerima bantuan PKH sebesar 2,106 kali daripada yang tidak memiliki aset tak bergerak.

f. Klasifikasi Regresi Logistik Biner dengan *Confussion Matrix*

Menurut Han dan Kamber (2001) *Confussion Matrix* merupakan metode perhitungan untuk melihat apakah pengklasifikasian yang sudah dilakukan baik atau tidak. Dengan menggunakan *Confussion Matrix*, dapat dianalisa seberapa baik pengklasifikasian yang telah dilakukan.

Tabel 3. *Confussion Matrix*

		Kelas Prediksi	
		Positif	Negatif
Kelas Sebenarnya	Positif	TP	FN
	Negatif	FP	TN

Berdasarkan Tabel 3 dapat dilakukan perhitungan akurasi dan eror yang ditunjukkan dengan persamaan berikut.

$$Akurasi = \frac{TP+T}{TP+FN+FP+TN} \times 100\%$$

DOI: 10.24036/ujsds/vol1-iss2/24

ISSN(*Online*): 2985-475X

$$APER = \frac{FP+}{TP+FP+TN+FN} \times 100\%$$

Dengan menggunakan data *testing* sebanyak 20% dari keseluruhan data yang digunakan, diperoleh tabel nilai *confussion matrix* dengan metode Regresi Logistik Biner sebagai berikut.

Tabel 4. *Confussion Matrix* Regresi Logistik Biner

Kelas Prediksi	Kelas Sebenarnya	
	0	1
0	3	30
1	1	92

Berdasarkan Tabel 4 dapat dilakukan perhitungan akurasi dan eror klasifikasi data dengan metode Regresi Logistik Biner. Dari hasil perhitungan yang telah didapatkan, diperoleh sebanyak 95 orang diklasifikasikan dengan benar dan 31 orang diklasifikasikan dengan tidak benar. Dapat diketahui juga hasil persentase tingkat akurasi klasifikasi penerima bantuan sosial PKH dengan Regresi Logistik Biner yaitu sebesar 75% dan nilai eror yang dihasilkan yaitu sebesar 25%.

B. *Naive Bayes*

Perhitungan probabilitas dengan metode *Naive Bayes* menggunakan data *training* sebesar 80% dari penerima bantuan sosial PKH di Kelurahan Anduring tahun 2020. Berikut adalah langkah-langkah dalam menganalisisnya.

a. Perhitungan Probabilitas P(Ci)

P(Ci) atau probabilitas *prior* yaitu nilai probabilitas dari suatu hipotesis sebelum melakukan pengamatan terhadap suatu kondisi.

Tabel 5. Probabilitas Penerima Bantuan PKH

Menerima	Tidak Menerima
0,716	0,284

Berdasarkan Tabel 5. Didapatkan hasil perhitungan nilai probabilitas "menerima" dan "tidak menerima" bantuan PKH, disimpulkan bahwa peluang masyarakat menerima bantuan PKH sebesar 0,716 dan peluang tidak menerima bantuan PKH sebesar 0,284.

b. Perhitungan Probabilitas X bersyarat Ci (P(X|Ci)

P(X|Ci) atau probabilitas *posterior* yaitu nilai probabilitas dari suatu hipotesis setelah melakukan pengamatan terhadap suatu kondisi. Hasil dari perhitungan nilai probabilitas ini menjelaskan bahwa peluang masyarakat menerima atau tidak menerima bantuan PKH berdasarkan dengan kondisi dari variabel X.

Tabel 6. Probabilitas Penerima Bantuan PKH Berdasarkan Kondisi Masing-Masing Variabel

Variabel	Kondisi	P(X)	**P(X|Ci)**	
			Menerima	Tidak Menerima
Status Bangunan	Milik Sendiri	0,765	**0,772**	**0,774**
	Sewa/Kontrak	0,046	**0,042**	**0,056**
	Usaha	0,189	**0,186**	**0,196**
Kondisi Dinding	Layak	0,239	**0,264**	**0,175**
	Tidak Layak	0,761	**0,736**	**0,825**
Kondisi Atap	Layak	0,239	**0,242**	**0,140**
	Tidak Layak	0,761	**0,758**	**0,860**
Sumber Air Minum	Air Isi Ulang	0,656	**0,625**	**0,734**
	PAM/PDAM	0,183	**0,219**	**0,091**
	Lainnya	0,161	**0,156**	**0,175**
Sumber Penerangan	Lampu	0,974	**0,978**	**0,965**
	Lilin	0,006	**0,008**	**0**

DOI: 10.24036/ujsds/vol1-iss2/24

ISSN(*Online*): 2985-475X

	Lainnya	0,02	**0,014**	**0,035**
Bahan Bakar Masak	Gas	0,567	**0,586**	**0,517**
	Minyak Tanah	0,008	**0,011**	**0**
	Lainnya	0,425	**0,403**	**0,483**
Ketersediaan Tabung Gas	Ada	0,992	**0,992**	**0,993**
	Tidak Ada	0,008	**0,008**	**0,007**
Ketersediaan Lemari Es	Ada	0,583	**0,614**	**0,503**
	Tidak Ada	0,417	**0,386**	**0,497**
Ketersediaan AC	Ada	0,01	**0,006**	**0,021**
	Tidak Ada	0,99	**0,994**	**0,979**
Ketersediaan Telepon	Ada	0,008	**0,003**	**0,021**
	Tidak Ada	0,992	**0,997**	**0,979**
Ketersediaan Televisi	Ada	0,823	**0,814**	**0,846**
	Tidak Ada	0,177	**0,186**	**0,154**
Ketersediaan Emas	Ada	0,048	**0,047**	**0,049**
	Tidak Ada	0,952	**0,953**	**0,951**
Ketersediaan Motor	Ada	0,646	**0,617**	**0,720**
	Tidak Ada	0,354	**0,383**	**0,280**
Ketersediaan Mobil	Ada	0,01	**0**	**0,035**
	Tidak Ada	0,99	**1**	**0,965**
Ketersediaan Aset Tak Bergerak	Ada	0,654	**0,614**	**0,755**
	Tidak Ada	0,346	**0,386**	**0,245**

Dari Tabel 6 diperoleh hasil perhitungan yang telah dilakukan dari masing-masing variabel yang ditampilkan. Nilai P(X) adalah nilai peluang dari masing-masing variabel X dan nilai P(X|Ci) adalah nilai peluang masyarakat menerima atau tidak menerima bantuan PKH berdasarkan kondisi dari variabel X.

c. Perhitungan Manual Menggunakan Data *Testing*

Berikut adalah langkah-langkah perhitungan manual menggunakan data *testing* dengan metode *Naive Bayes*. Diketahui seseorang dengan kondisi sebagai berikut.

Tabel 7. Data Uji

No	Variabel	Kondisi	Tidak Menerima P(X|C0)	Menerima P(X|C1)
1	Status Bangunan	Sewa/Kontrak	0,056	0,042
2	Kondisi Dinding	Tidak Layak	0,825	0,736
3	Kondisi Atap	Tidak Layak	0,860	0,758
4	Sumber Air Minum	Air Isi Ulang	0,734	0,625
5	Sumber Penerangan	Lampu	0,965	0,978
6	Bahan Bakar Masak	Gas	0,517	0,586
7	Ketersediaan Tabung Gas	Ada	0,993	0,992
8	Ketersediaan Lemari Es	Tidak Ada	0,497	0,386
9	Ketersediaan AC	Tidak Ada	0,979	0,994
10	Ketersediaan Telepon	Tidak Ada	0,979	0,997
11	Ketersediaan Televisi	Ada	0,846	0,814
12	Ketersediaan Emas	Tidak Ada	0,951	0,953
13	Ketersediaan Motor	Ada	0,720	0,617
14	Ketersediaan Mobil	Tidak Ada	0,965	1
15	Ketersediaan Aset Tak Bergerak	Ada	0,775	0,614

DOI: 10.24036/ujsds/vol1-iss2/24

Berdasarkan Tabel 7, dilakukan perhitungan untuk menentukan pengklasifikasian menerima atau tidak menerima bantuan. Untuk kelas menerima bantuan, masing-masing nilai probabilitas akan dikalikan sehingga menghasilkan nilai probabilitas P(X|C1), didapatkan hasil sebesar 0.001 dan untuk kelas tidak menerima bantuan masing-masing nilai probabilitas dikalikan dan menghasilkan nilai probabilitas P(X|C0), didapatkan hasil sebesar 0.003. Selanjutnya, untuk mengetahui klasifikasi menerima atau tidak menerima bantuan adalah dengan mengalikan hasil masing-masing nilai probabilitas P(X|C1) dengan P(C1) dan P(X|C0) dengan P(C0).

Dari perkalian sebelumnya yang telah dilakukan, dihasilkan bahwa nilai probabilitas P(C1|X) = 0.0007 dan probabilitas P(C0|X) = 0.0009. Dapat disimpulkan bahwa nilai P(C0|X) > P(C1|X). Maka data uji tersebut diklasifikasikan kedalam kelas **tidak menerima** bantuan PKH di Kelurahan Anduring.

d. Klasifikasi *Naive Bayes* dengan *Confussion Matrix*

Dengan menggunakan data *testing* sebesar 20% dari keseluruhan data, diperoleh tabel C*onfussion Matrix* yang dapat dilihat pada Tabel 8 sebagai berikut.

Tabel 8. *Confussion Matrix Naive Bayes*

Kelas Prediksi	Kelas Sebenarnya	
	0	1
1	4	9
0	29	84

Berdasarkan Tabel 8 dilakukan perhitungan akurasi dan eror terhadap klasifikasi data dengan menggunakan metode *Naive Bayes*. Dari hasil perhitungan yang didapatkan, diperoleh sebanyak 88 orang diklasifikasikan dengan benar dan 38 orang diklasifikasikan dengan salah. Dapat diketahui hasil persentase tingkat akurasi pengklasifikasian penerima bantuan sosial PKH dengan *Naive Bayes* yaitu sebesar 70% dan nilai eror yang dihasilkan yaitu sebesar 30%.

C. Perbandingan Ketepatan Klasifikasi Metode Regresi Logistik Biner dan *Naive Bayes*

Setelah melakukan perhitungan ketepatan klasifikasi menggunakan tabel *confussion matrix* dengan menggunakan metode Regresi Logistik Biner dan *Naive Bayes*, diperoleh hasil sebagai berikut.

Tabel 9. Perbandingan Ketepatan Klasifikasi dengan Regresi Logistik Biner dan *Naive Bayes*

Metode	Hasil Ketepatan Klasifikasi
Regresi Logistik Biner	75%
Naive Bayes	70%

Berdasarkan Tabel 9 dapat dinyatakan bahwa setelah melakukan perhitungan dengan menggunakan dua metode, diperoleh hasil yaitu metode terbaik dalam pengklasifikasian penerima bantuan sosial PKH di Kelurahan Anduring tahun 2020 adalah metode Regresi Logistik Biner dengan tingkat akurasi sebesar 75%.

IV. KESIMPULAN

Setelah melakukan pengujian dengan data *testing* sebesar 20% menggunakan kedua metode, diperoleh hasil akurasi ketepatan klasifikasi dengan Regresi Logistik Biner sebesar 75% dan *Naive Bayes* sebesar 70%. Berdasarkan hasil tersebut diperoleh bahwa Regresi Logistik Biner memiliki tingkat akurasi lebih tinggi dibandingkan dengan *Naive Bayes*. Maka disimpulkan bahwa metode Regresi Logistik Biner lebih baik dalam mengukur ketepatan klasifikasi penerima bantuan Program Keluarga Harapan (PKH) di Kelurahan Anduring tahun 2020.

DAFTAR PUSTAKA

Han, J & Kamber, M. (2001). *Data Mining Concepts and Techniques 2nd Edition*. San Fransisco: Morgan Kaufmann publisher.

Han, J & Kamber, M. (2006). *Data Mining Concepts and Techniques 2nd Edition*. San Fransisco: Morgan Kaufmann Publisher.

DOI: 10.24036/ujsds/vol1-iss2/24

ISSN(*Online*): 2985-475X

Hosmer, D.W & Lemeshow, S. (2000). Apllied Logistic Regression. New York: John Wiley dan Sons

Nataya, E. J, & Supriyadi. (2017). "Pemberdayaan Keluarga Penerima Manfaat Melalui Program Keluarga Harapan Di Kelurahan Kelun Kecamatan Kartoharjo Kota Madiun", *Jurnal Sosiologi DILEMA*, Vol. 32, No. 2, hal. 1-9.

Salim, Abdurrahman. (2017). "Pengoptimalan Naive Bayes Dan Regresi Logistik Menggunakan Algoritma Genetika Untuk Data Klasifikasi (Studi Kasus: Pembuangan Limbah Domestik di Surabaya Timur)", *Skripsi*, Surabaya: Institut Tenologi Sepuluh Nopember

Samosir, Riama Oktaviyani. (2015). "Perbandingan Klasifikasi Metode Regresi Logistik Biner Dan Radial Basis Function Neural Network Pada Berat Bayi Lahir Rendah", *Jurnal Gaussian*, Vol. 4, No. 4, hal. 997-1005.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining Practical Machine Learning Tools and Techniques*, Burlington: Elsevier.

DOI: 10.24036/ujsds/vol1-iss2/24