

PERBANDINGAN REGRESI LOGISTIK BINER DAN PROBIT BINER DALAM PEMODELAN TINGKAT PARTISIPASI ANGKATAN KERJA

Hendra H. Dukalang^{1*}

¹Jurusan Perbankan Syariah, IAIN Sultan Amai Gorontalo. Kabupaten Gorontalo 96210, Indonesia

*Penulis Korespondensi. Email: hendra.statistika@iaingorontalo.ac.id

Abstrak

Regresi merupakan suatu metode analisis data yang digunakan untuk memodelkan hubungan antara satu variabel respon dan satu atau lebih variabel prediktor. Dalam pemodelan regresi sering digunakan data pada umumnya model regresi yang sering digunakan adalah sederhana maupun regresi berganda dalam pemodelan yang variabel responnya adalah data kuantitatif. Perbedaan mendasar dari model regresi dengan menggunakan data kuantitatif tujuan utamanya adalah mengestimasi nilai rata-rata dari variabel dependen dengan menggunakan nilai-nilai tertentu dari variabel independen. Sedangkan dalam model regresi dengan variabel dependen kualitatif, tujuan utamanya adalah menemukan probabilitas terjadinya sesuatu (probability model). Salah satu metode pengembangan dari model regresi untuk data dengan variabel respon kualitatif adalah regresi Logistik dan Probit. Tujuan Penelitian ini adalah membandingkan Model terbaik dengan menggunakan regresi logistic biner dengan regresi probit biner pada kasus Tingkat Partisipasi Angkatan Kerja (TPAK) di Kota Gorontalo. Metode penelitian yang digunakan adalah metode penelitian kuantitatif, dengan pemodelan regresi logistic biner dan regresi probit biner. Hasil penelitian menunjukkan bahwa variabel yang berpengaruh signifikan pada TPAK Kota Gorontalo adalah tingkat pengangguran terbuka, dan model terbaik antara model regresi logistic biner dengan nilai AIC sebesar 1,289 lebih kecil dari nilai AIC regresi Probit biner 1,318, demikian pula dari nilai R² diperoleh nilai R² untuk regresi logistic biner sebesar 12,74%, lebih besar dari nilai R² regresi probit biner sebesar 10,70%.

Kata Kunci: Regresi; Logistik Biner; Probit Biner; TPAK

Abstract

Regression is a data analysis method used to model the relationship between one response variable and one or more predictor variables. In regression modelling, data is often used. In general, the regression model that is often used is simple or multiple regression in modelling where the response variable is quantitative data. The fundamental difference from regression models using quantitative data is the main objective is to estimate the average value of the dependent variable using certain values of the independent variable. Whereas in a regression model with a qualitative dependent variable the main objective is to find the probability of something happening (probability model). One of the development methods of the regression model for data with qualitative response variables is Logistic and Probit regression. The purpose of this study was to compare the best model using binary logistic regression with binary probit regression in the case of Labor Force Participation Rate (TPAK) in Gorontalo City. The research method used is quantitative research methods, with binary logistic regression modelling and binary probit regression. The results showed that the variable that has a significant effect on TPAK Gorontalo City is the open unemployment rate, and the best model between the binary logistic regression model with an AIC value of 1.289 is smaller than the AIC value of the binary Probit regression 1.318, likewise from the R² value the R² value for regression is obtained. binary logistic of 12.74%, greater than the R² value of binary probit regression of 10.70%.

Keywords: Regression; Binary Logistics; Binary Probit; TPAK

1. Pendahuluan

Hubungan antar variabel independen dan dependen dapat dimodelkan dengan menggunakan Analisis Regresi. Metode regresi merupakan analisis data yang mendeskripsikan hubungan antara sebuah respon dan satu atau lebih prediktor [1]. Akan tetapi pada umumnya model regresi yang sering digunakan adalah sederhana maupun regresi berganda dalam pemodelan yang variabel responnya adalah data kuantitatif. Namun terkadang dalam beberapa kasus, sering di temui variabel responnya merupakan variabel dengan data kualitatif. Sehingga di perlukan suatu metode yang dapat digunakan dalam menyelesaikan permasalahan tersebut.

Perbedaan mendasar dari model regresi dengan menggunakan data kuantitatif tujuan utamanya adalah mengestimasi nilai rata-rata dari variabel dependen dengan menggunakan nilai-nilai tertentu dari variabel independen. Sedangkan dalam model regresi dengan variabel dependen kualitatif, tujuan utamanya adalah menemukan probabilitas terjadinya sesuatu (*probability model*). Salah Satu metode pengembangan dari model regresi untuk data dengan variabel respon kualitatif adalah regresi Logistik dan Probit.

Regresi logistik merupakan salah satu metode yang dapat digunakan untuk mencari hubungan respon yang bersifat *dichotomous* (berskala nominal atau ordinal dengan dua kategori) atau *polychotomous* prediktor [2]. Regresi logistik mirip dengan analisis diskriminan untuk menguji probability terjadinya variabel dependen yang diprediksi dengan menggunakan variabel dependennya. Dalam pemodelan regresi logistik sering dikatakan model regresi binary respon. Sehingga jika di kaitkan dengan analisis diskriminan Masalah dalam model regresi logistik dapat di selesaikan dengan analisis diskriminan, akan tetapi asumsi normal multivariat dalam analisis diskriminasi tidak dapat terpenuhi, karena variabel dependen dalam model merupakan campuran antara variabel metrik (kontinyu) dan non metrik (kategorikal). Dalam pemodelan regresi logistik tidak diperlukan pemenuhan asumsi normal multivariat sehingga regresi logistik dapat menjadi solusi jika asumsi normal multivariat tidak terpenuhi.

Dalam mengestimasi model regresi logistik digunakan *Cumulative logistic function* (Fungsi kumulatif logistik). Sedangkan fungsi kumulatif logistik bukan merupakan satu-satunya *Cumulative distribution function* (CDF). Dalam perkembangannya model estimasi dengan menggunakan *Normal Cumulative distribution function* lebih sesuai untuk pemodelan variabel dependen kualitatif dan variabel prediktor bersifat kualitatif, kuantitatif, maupun gabungan. Model regresi ini disebut dengan regresi probit. Regresi probit singkatan dari *Probability Unit* merupakan model berdasarkan fungsi sebaran peluang normal kumulatif baku. Model regresi logistik dan regresi probit merupakan pendekatan alternatif untuk memodelkan hubungan antara variabel respon kategorik dan variabel bebas, dimana variabel respon berdistribusi Bernoulli atau multinomial [3].

Penggunaan model logit sering kali digunakan dalam data klasifikasi [4]. Dalam beberapa penelitian yang menggunakan pemodelan regresi logistik dan regresi probit diantaranya adalah Penggunaan metode *Least Absolute Deviation* (LAD) untuk mengestimasi parameter model [5], Pengujian hipotesis dengan parameter *robust* [6] penggunaan metode *Maximum Likelihood Estimation* (MLE) untuk mengestimasi model, pendekatan numerik dan estimasi parameter serta diagnosis terhadap model [7]. Pemilihan model terbaik dengan kriteria *Cp Mallows* dan AIC (*Akaike's Information Criterion*)[8]. Penelitian untuk pemodelan regresi probit telah digunakan dalam Pemodelan dan prediksi kebangkrutan bank di Slovakia [9], Regresi probit untuk variabel pengontrol hubungan antara variabel prediktor dan variabel respon [10]. Pemodelan regresi probit dan aplikasinya untuk data mining [11] Pemodelan Ketahanan Pangan di Indonesia dengan pendekatan regresi probit ordinal [12].

Dalam penelitian ini regresi logistik dan regresi probit di implementasikan dalam pemodelan Pemodelan Angkatan Kerja Di Kota Gorontalo. Menurut BPS Kota Gorontalo laju pertumbuhan penduduk Kota Gorontalo dapat menimbulkan berbagai tekanan dalam berbagai aspek kehidupan dan berimplikasi dalam penyediaan lapangan kerja bagi angkatan kerja yang ada, karena bila penduduk

bertambah maka jumlah angkatan kerja bertambah sehingga dapat menimbulkan masalah bagi penyediaan lapangan kerja baru. Penelitian ini bertujuan untuk mendapatkan model regresi terbaik pada pemodelan Angkatan Kerja di Kota Gorontalo, sehingga mendapatkan faktor-faktor yang mempengaruhi Tingkat partisipasi angkatan kerja di Kota Gorontalo.

2. Metode Penelitian

Penelitian ini menggunakan metode penelitian kuantitatif. Data yang digunakan dalam penelitian ini adalah data sekunder yang di peroleh dari Dinas tenaga Kerja, UKM dan Koperasi di Kota Gorontalo. Variabel yang digunakan meliputi Tingkat Partisipasi Angkatan Kerja (TPAK) sebagai Variabel Y, dengan kategori 0 apabila kelurahan masuk dalam tingkat partisipasi Angkatan kerja rendah, serta kategori 1 apabila kelurahan masuk dalam tingkat partisipasi Angkatan kerja tinggi. Variabel prediktor yang digunakan dalam penelitian ini sebanyak 6 variabel dengan tipe kontinu, yaitu presentase usia produktif (X1), Tingkat pengangguran terbuka (X2), presentase pendidikan SD (X3), presentase Pendidikan SMP (X4), presentase Pendidikan SMA (X5), presentase Pendidikan PT (X6). Unit pengamatan yang digunakan dalam penelitian ini adalah kelurahan di Kota Gorontalo yang terdiri dari 50 Kelurahan. Sehingga jumlah keseluruhan dari unit pengamatan adalah sebanyak 50 kelurahan.

Metode analisis data untuk pemodelan tingkat partisipasi angkatan kerja (TPAK) dengan regresi logistik dan regresi probit adalah sebagai berikut :

1. Melakukan analisis statistika deskriptif terhadap variabel respon dan variabel prediktor
2. Melakukan pengujian multikolinieritas pada variabel prediktor, untuk mengetahui interdependensi antar variabel prediktor. Apabila terjadi multikolinieritas maka diatasi dengan *backward eliminasi*.

$$VIF = \frac{1}{1 - R_s^2}, \text{ untuk } s = 1, 2, \dots, p, \text{ dimana } R_s^2 = 1 - \frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{\sum_{i=1}^N (x_i - \bar{x}_i)^2}$$

3. Mengestimasi parameter regresi logistik untuk mengetahui hubungan antar variabel prediktor dan variabel respon menggunakan Metode *Maximum Likelihood*. Metode tersebut mengestimasi parameter β dengan cara memaksimalkan fungsi *likelihood*
4. Melakukan pengujian signifikansi parameter dengan melakukan uji serentak menggunakan persamaan

$$G = -2 \ln \frac{\left[\frac{n_1}{n} \right]^{n_1} \left[\frac{n_0}{n} \right]^{n_0}}{\prod_{j=1}^n \hat{\pi}_j^{y_j} [1 - \hat{\pi}_j]^{1 - y_j}}$$

5. Melakukan pengujian signifikansi parameter dengan melakukan uji parsial menggunakan persamaan.

$$W_i = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$$

6. Melakukan uji kesesuaian model untuk mengetahui apakah model yang didapatkan telah sesuai atau tidak.

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

7. Menghitung ketepatan klasifikasi model dengan persamaan $1 - APER$, dimana *APER* (*apparent error rate*) menggunakan persamaan berikut:

$$APER = \frac{n_{1M} + n_{2M}}{n_1 + n_2} \times 100\%$$

8. Pemilihan Model terbaik menggunakan *Pseudo R² McFadden* dengan persamaan sebagai berikut:

$$R_{MF}^2 = 1 - \frac{\text{Log } L_1}{\text{Log } L_0} = 1 - \ln \left(\frac{\left[\frac{n_1}{n} \right]^{n_1} \left[\frac{n_0}{n} \right]^{n_0}}{\prod_{j=1}^n \hat{\pi}_j^{y_j} [1 - \hat{\pi}_j]^{1-y_j}} \right)$$

9. Membuat kesimpulan dan saran dari hasil penelitian.

Pengolahan dan analisis data untuk estimasi dan pengujian hipotesis parameter menggunakan *software views* 10.

3. Hasil dan Pembahasan

3.1 Karakteristik Variabel Respon Dan Variabel Prediktor.

3.1.1 Karakteristik Variabel Respon

Variabel respon yang digunakan dalam penelitian ini adalah partisipasi Angkatan kerja (TPAK) di Kota Gorontalo. TPAK yang tersebar di 50 kelurahan di Kota Gorontalo. Status tingkat partisipasi Angkatan kerja dikategorikan menjadi dua kategori, yaitu kategori yakni kategori tinggi dan kategori rendah. Berikut hasil pengujian deskriptif untuk tingkat partisipasi angkatan kerja di Kota Gorontalo. Dari 50 Kelurahan yang ada di Kota Gorontalo, terdapat 25 kelurahan yang masuk dalam kategori tingkat partisipasi Angkatan kerja kategori rendah. Sedangkan 25 kelurahan lainnya masuk dalam kategori kategori tingkat partisipasi Angkatan kerja kategori tinggi.



Gambar 1. Presentase Tingkat Partisipasi Angkatan Kerja

3.1.2 Karakteristik Variabel Prediktor

Variabel prediktor yang digunakan dalam penelitian ini sebanyak 6 variabel dengan tipe kontinu, yaitu presentase usia produktif (X1), Tingkat pengangguran terbuka (X2), presentase pendidikan Sekolah Dasar (X3), presentase Pendidikan SMP (X4), presentase Pendidikan SMA (X5), presentase Pendidikan Tinggi (X6), Karena data yang digunakan dalam variabel prediktor merupakan data interval maka Analisis statistika deskriptif yang digunakan untuk mendapatkan informasi dari seluruh variabel menggunakan nilai mean, standar deviasi, minimum, dan maksimum.

Tabel 1. Deskriptif Variabel Prediktor

Variable	Mean	StDev	Minimum	Maximum
X1	40.179	1.499	36.42	43.37
X2	20.392	4.465	10.14	37.82
X3	12.064	4.645	5.4	26.15
X4	11.588	1.866	6.06	15.91
X5	26.023	5.274	12.01	36.66
X6	7.366	3.025	2.13	14.55

Tabel 1 menunjukkan bahwa usia produktif pekerja yaitu 45-45 tahun di kota Gorontalo memiliki rata-rata 40,179 persen, dengan standar deviasi sebesar 1,499 dengan nilai minimum sebesar 36,42 persen dan nilai maksimum 43,37 persen. Serta Tingkat pengangguran terbuka di kota Gorontalo memiliki rata-rata 20,392 persen, dengan standar deviasi sebesar 4,465 dengan nilai minimum sebesar 10,14 persen dan nilai maksimum 37,82 persen. Tingkat pendidikan SD di kota Gorontalo memiliki rata-rata 12,064 persen, dengan standar deviasi sebesar 4,645 dengan nilai minimum sebesar 5,4 persen dan nilai maksimum 26,15 persen. Tingkat Pendidikan SMP kota Gorontalo memiliki rata-rata 11,588 persen, dengan standar deviasi sebesar 1,866 dengan nilai minimum sebesar 6,06 persen dan nilai maksimum 15,91 persen. Untuk Tingkat Pendidikan SMA di kota Gorontalo memiliki rata-rata 22,023 persen, dengan standar deviasi sebesar 5,274 dengan nilai minimum sebesar 12,01 persen dan nilai maksimum 36,66 persen. Selanjutnya tingkat Pendidikan perguruan tinggi di kota Gorontalo memiliki rata-rata 7,366 persen, dengan standar deviasi sebesar 3,025 dengan nilai minimum sebesar 2,13 persen dan nilai maksimum 14,55 persen.

3.2 Pengujian Asumsi Multikolinieritas

Uji multikolinieritas merupakan suatu pengujian yang dilakukan untuk mengetahui ada atau tidaknya hubungan secara linier antar variabel prediktor. Dalam Analisis regresi logistic biner dan regresi probit biner tidak diperkenankan adanya hubungan secara linier antara variabel prediktor sehingga perlu dilakukan uji multikolinieritas sebelum melakukan pemodelan menggunakan regresi logistic biner maupun regresi probit biner.

Berdasarkan hasil pengujian korelasi antar variabel prediktor, tidak terdapat korelasi yang lebih dari 0,80. Sehingga dapat disimpulkan bahwa tidak terdapat multikolinieritas antar variabel prediktor. Hasil ini juga sesuai dengan pengujian menggunakan nilai *Variance Inflating Factor (VIF)*, sehingga diperoleh hasil pengujian multikolinieritas yang disajikan pada Tabel 2.

Tabel 2. Hasil Uji Multikolinieritas

Variabel	Nilai VIF	Keterangan
X1	1,18	Tidak terjadi multikolinieritas
X2	1,50	Tidak terjadi multikolinieritas
X3	3,14	Tidak terjadi multikolinieritas
X4	1,70	Tidak terjadi multikolinieritas
X5	3,64	Tidak terjadi multikolinieritas
X6	2,23	Tidak terjadi multikolinieritas

Berdasarkan Tabel 2 ditunjukkan bahwa hasil pengujian asumsi multikolinieritas tidak terjadi masalah dalam multikolinieritas, Karena Nilai VIF kurang dari 10,00. Sehingga dapat dilanjutkan pada pemodelan dengan menggunakan regresi logistic biner dan regresi probit biner.

3.3 Model Regresi Logistik Biner

3.3.1 Pengujian parameter secara serentak Regresi Logistik Biner

Pengujian parameter secara serentak digunakan untuk memeriksa keberartian koefisien β secara keseluruhan dengan variabel prediktor Uji serentak pada penelitian ini menggunakan likelihood ratio test (G^2) dengan α (0,10).. Berdasarkan hasil pengujian parameter secara serentak diperoleh p-value sebesar 0,033. Nilai tersebut kurang dari nilai alpha, serta nilai chi square sebesar $13,70 > 2,706$ sehingga dapat dikatakan bahwa H_0 ditolak yang berarti bahwa pada tingkat kepercayaan sebesar 90 persen minimal terdapat satu parameter yang signifikan pada model.

3.3.2 Pengujian parameter secara parsial Regresi Logistik Biner

Pengujian parameter secara parsial digunakan untuk mengetahui keberartian koefisien β dari masing-masing variabel prediktor yaitu X_1 dan X_6 secara individu. hipotesis yang digunakan

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

Kriteria pengujian, Tolak H_0 jika nilai $W_s^2 > \chi_{(df,\alpha)}^2$ atau p-value $< \alpha (0,10)$.

Pengujian parameter secara parsial pada penelitian ini menggunakan uji Wald. Hasil pengujian parameter secara parsial adalah variable predictor dengan nilai p-value sebesar $0,0154 < \alpha (0,10)$., dengan nilai chi-square sebesar 2,822. Lebih besar dari chi-square table 2,706. Sehingga hanya variabel tingkat pengangguran terbuka yang dimasukkan dalam model regresi logistik biner.

3.4 Model Regresi Logistik Biner

Variabel prediktor yang signifikan yang merupakan hasil yang diperoleh dari uji serentak dan uji parsial yaitu X digunakan untuk membentuk model regresi logisti biner terbaik. Berikut adalah model regresi logistik biner yang dapat di bentuk berdasarkan persamaan:

$$\ln\left(\frac{p}{1-p}\right) = g(x) = -5,074 + 0,2512X_2$$

Sehingga dapat di peroleh formula untuk mencari nilai probabilitas adalah sebagai berikut:

$$P = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

Dimana $g(x) = -5,074 + 0,2512X_2$

Berdasarkan model yang terbentuk variabilitas variabel respon yang dapat dijelaskan oleh variabel prediktor sangat kecil yaitu sebesar 12,74% yang menunjukkan pemodelan yang terbentuk kurang bagus, hal ini dapat terlihat dari nilai *R-Square* sebesar 12,74%. Selanjutnya untuk menguji apakah terdapat perbedaan antara hasil observasi dengan kemungkinan hasil prediksi model. Maka digunakan hasil pengujian *Hosmer and Lemeshow*.

Hipotesis

H_0 : Model sesuai (Tidak terdapat perbedaan antara hasil observasi dengan kemungkinan hasil prediksi)

H_1 : Model tidak sesuai (terdapat perbedaan antara hasil observasi dengan kemungkinan hasil prediksi)

Kriteria pengujian, Tolak H_0 jika nilai $\hat{C} > \chi_{(df,\alpha)}^2$ atau p-value $< \alpha (0,10)$.

Berdasarkan hasil pengujian *Hosmer and Lemeshow* diperoleh bahwa nilainya sebesar $6,396 < 62,038$ tidak terdapat perbedaan antara hasil observasi dengan hasil prediksi yang diperoleh menggunakan model yang telah terbentuk dimana hal ini dapat diketahui dari nilai p-value $(0,603) >$ nilai $\alpha (0,10)$. sehingga gagal tolak H_0 .

3.5 Ketepatan Klasifikasi Model Regresi Logistik Biner

Maka model yang terbentuk berdasarkan variabel yang ada sudah sesuai. Kesesuaian model juga dapat ditunjukkan dari persentase ketepatan klasifikasi. Berikut ini ketepatan klasifikasi model yang terbentuk pada Tabel 3.

Tabel 3 Ketepatan Klasifikasi Model Regresi Logistik Biner

Observasi		Prediksi		
		TPAK		Presentase Kebenaran
		Rendah	Tinggi	
TPAK	Rendah	22	9	88,0%
	Tinggi	3	16	64,0%
Presentase Keseluruhan				76,0%

Berdasarkan tabel ketepatan klasifikasi pada tabel 3 diketahui tingkat persentase kebenaran pada tingkat rendah sangat baik yaitu sebesar 88% sedangkan tingkat persentase ketepatan klasifikasi pada tingkat tinggi juga baik yaitu 64%. Secara keseluruhan tingkat persentase kebenaran cukup tinggi yaitu sebesar 76%.

3.6 Model Regresi Probit Biner

3.6.1 Pengujian parameter secara serentak Regresi Probit Biner

Pengujian parameter secara serentak digunakan untuk memeriksa keberartian koefisien β secara keseluruhan dengan variabel prediktor. Uji serentak pada penelitian ini menggunakan likelihood ratio test (G^2) dengan α (0,10). Berdasarkan hasil pengujian parameter secara serentak diperoleh p-value sebesar 0,043. Nilai tersebut kurang dari nilai alpha, serta nilai chi square sebesar 12,97 sehingga dapat dikatakan bahwa H_0 ditolak yang berarti bahwa pada tingkat kepercayaan sebesar 90 persen minimal terdapat satu parameter yang signifikan pada model.

3.6.2 Pengujian parameter secara parsial Regresi Probit Biner

Pengujian parameter secara parsial digunakan untuk mengetahui keberartian koefisien β dari masing-masing variabel prediktor yaitu X_1 dan X_6 secara individu. hipotesis yang digunakan

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

Kriteria pengujian, Tolak H_0 jika nilai $W^2 > \chi^2_{(df, \alpha)}$ atau p-value $< \alpha$ (0,10). Pengujian parameter secara parsial pada penelitian ini menggunakan uji *Wald*. Hasil pengujian parameter secara parsial adalah variable predictor dengan nilai p-value $< \alpha$ (0,10) adalah variabel tingkat pengangguran terbuka, yakni sebesar 0,013 dengan nilai *chi-square* sebesar 2,901. Lebih besar dari *chi-square* table 2,706 Sehingga hanya variable tersebut yang dimasukkan dalam model regresi probit biner.

3.6.3 Model Regresi Probit Biner

Variabel prediktor yang signifikan yang merupakan hasil yang diperoleh dari uji serentak dan uji parsial yaitu X digunakan untuk membentuk model regresi probit biner terbaik. Berikut adalah model regresi probit biner yang dapat di bentuk adalah:

$$\hat{p}(Y = 0) = \Phi(-2,287 + 0,1116X_2)$$

$$\hat{p}(Y = 1) = 1 - \Phi(-2,287 + 0,1116X_2)$$

Berdasarkan model yang terbentuk variabilitas variabel respon yang dapat dijelaskan oleh variabel prediktor sangat kecil yaitu sebesar 10,70% yang menunjukkan pemodelan yang terbentuk kurang bagus, hal ini dapat terlihat dari nilai R-Square sebesar 10,70%.

Selanjutnya untuk menguji apakah terdapat perbedaan antara hasil observasi dengan kemungkinan hasil prediksi model. Maka digunakan hasil pengujian *Hosmer and Lemeshow*. Hipotesis yang digunakan dalam pengujian model adalah sebagai berikut.

H_0 : Model sesuai (Tidak terdapat perbedaan antara hasil observasi dengan kemungkinan hasil prediksi)

H_1 : Model tidak sesuai (terdapat perbedaan antara hasil observasi dengan kemungkinan hasil prediksi)

Kriteria pengujian, Tolak H_0 jika nilai $\hat{C} > \chi^2_{(df,\alpha)}$ atau p-value $< \alpha (0,10)$

Berdasarkan hasil pengujian *Hosmer and Lemeshow* diperoleh bahwa tidak terdapat perbedaan antara hasil observasi dengan hasil prediksi yang diperoleh menggunakan model yang telah terbentuk dimana hal ini dapat diketahui dari nilai p-value $(0,4515) > \alpha (0,10)$ sehingga gagal tolak H_0 atau model sudah sesuai.

3.6.4 Ketepatan Klasifikasi Model Regresi Probit Biner

Untuk Menunjukkan kesesuaian model dapat dikuatkan dengan menggunakan ketepatan klasifikasi model yang terbentuk, sebagaimana ditunjukkan pada Tabel 4.

Tabel 4. Ketepatan Klasifikasi Model

Observasi		Prediksi		
		TPAK		Presentase Kebenaran
		Rendah	Tinggi	
TPAK	Rendah	22	9	88,0%
	Tinggi	3	16	64,0%
Presentase Keseluruhan				76,0%

Berdasarkan tabel ketepatan klasifikasi pada tabel 4 diketahui tingkat persentase kebenaran pada tingkat rendah sangat baik yaitu sebesar 88% sedangkan tingkat persentase ketepatan klasifikasi pada tingkat tinggi juga baik yaitu 64%. Secara keseluruhan tingkat persentase kebenaran cukup tinggi yaitu sebesar 76%

3.7 Pemilihan Model Terbaik

Pemilihan model terbaik dalam penelitian ini menggunakan kriteria pemilihan model terbaik AIC (*Aikake Information Criterion*) sebagaimana ditunjukkan pada Tabel 5.

Tabel 5. Perbandingan Model Terbaik

Model	AIC	R ²
Regresi Logistik Biner	1,289	12,74%
Regresi Probit Biner	1,318	10,70%

Berdasarkan tabel 5 diperoleh bahwa nilai AIC terkecil adalah model regresi logistic biner dengan nilai AIC sebesar 1,289 sedangkan nilai AIC untuk model regresi probit biner sebesar 1,318. Hal ini dapat disimpulkan bahwa model terbaik adalah model regresi logistic biner. Hal ini sesuai dengan nilai dari R², di mana nilai R² terbesar merupakan R² dari model regresi logistic biner yaitu sebesar 12,74%. Sedangkan nilai R² dari model regresi probit biner yaitu sebesar 10,70%.

4. Kesimpulan

Berdasarkan hasil analisis dan pembahasan di simpulkan bahwa TPAK dengan kategori tinggi sebanyak 50% dan kategori rendah 50%, dengan rata-rata usia adalah 40 tahun, dan tingkat pengangguran terbuka adalah 20%. Variabel yang berpengaruh signifikan pada model regresi logistik biner adalah variabel tingkat pengangguran terbuka dengan koefisien 0,2512. Variabel yang berpengaruh signifikan pada model regresi probit biner adalah variabel tingkat pengangguran terbuka dengan koefisien 0,1116, sedangkan model terbaik adalah regresi logistik biner dengan nilai AIC lebih kecil dari regresi probit biner, dan nilai R^2 regresi logistik biner lebih besar dari nilai R^2 regresi probit biner.

Referensi

- [1] D.W. Hosmer, S. Lemeshow, R.X. Sturdivant, *Applied logistic regression, 3rd Edition*. John Wiley & Sons, 2013.
- [2] A. Agresti, *Categorical data analysis, 3rd Edition*. John Wiley & Sons, 2013.
- [3] D.J. Finney, *Probit analysis, 3rd Edition*. Cambridge University Press, 1971
- [4] Gujarati, Damodar. *Dasar-dasar Ekonometrika*, Edisi Kelima. Mangunsong, R. C. penerjemah. Jakarta: Salemba Empat. 2013
- [5] S. Hosseinian, E. Martinez, Robust binary regression. *Journal of Statistical Planning and Inference*, 141, 1497-1509, 2011
- [6] A.M. Bianco, E. Martinez, Robust testing in the logistic regression model. *Computational Statistics and Data Analysis*, 53, 4095-4105, 2009.
- [7] D. Pregibon, Logistic regression diagnostics. *The Annals of Statistics*, 9, 705-724, 1981.
- [8] D.W. Hosmer, B. Jovanovic, S. Lemeshow, Best subsets logistic regression. *Biometrics*, 45, 1265-1270, 1989.
- [9] T. Kliestik, K. Kocisova, M. Misankova, Logit and probit model used for prediction of financial health of company. *Procedia Economics and Finance*, 23, 850-855, 2015.
- [10] K.B. Karlson, A. Holm, R. Breen, Comparing regression coefficients between same-sample nested models using logit and probit: a new method. *Sociological Methodology*, 42, 286-313, 2012.
- [11] M. Razzaghi, The probit link function in generalized linear models for data mining applications. *Journal of Modern Applied Statistical Methods*, 12, 164169, 2013.
- [12] Permatasari, D.L. dan Ratnasaril V. Pemodelan Ketahanan Pangan di Indonesia dengan Pendekatan Regresi Probit Ordinal. *Jurnal Sains dan Seni ITS*, Vol.5, No.2, Hal:151-156. 2016.