



ANALISIS FAKTOR_FAKTOR KETENAGAKERJAAN DI INDONESIA DENGAN PENDEKATAN REGRESI LOGISTIK BINER

Aprillia Elsada Rinindah¹, Muhammad Azra Firdaus², Sakila Armayani³, Erlyne Nadhilah Widyaningrum^{4,*}, Wiwit Pura Nurmayanti⁵

^{1,2,3,4,5}Program Studi Statistika, Universitas Mulawarman, 75124, Indonesia

*Correspondance author: erlynenadhilah@fmipa.unmul.ac.id

ABSTRACT

Labor Force Participation Rate (TPAK) determines the economic growth of a country with variables that affect TPAK having two values or are binary. To analyze the TPAK, binary logistic regression is used which can analyze the relationship between predictor variables and response variables that have two value categories or are binary. The variables that affect TPAK are the proportion of formal labor, open unemployment rate (TPT), Gross Regional Domestic Product (PDRB) and average hourly wages of workers. The results of the analysis of the best model with an AIC value of 34.04 indicate that TPT and average hourly have an influence on TPAK, with significance levels of 0.00978 and 0.06910, respectively. The coefficient of determination value of 39.90% indicates that the model is able to explain most of the variation in TPAK, while other factors are influenced by variables outside the scope of the model. The classification accuracy level reaches 82.35% indicating that the model is quite reliable in predicting labor force participation status.

Keywords: *Binary Logistics, Open Unemployment Rate, Labor Force Participation Rate*

ABSTRAK

Tingkat Partisipasi Angkatan Kerja (TPAK) menentukan pertumbuhan ekonomi suatu negara dengan variabel yang mempengaruhi TPAK memiliki dua nilai atau bersifat biner. Untuk menganalisis TPAK tersebut digunakan regresi logistik biner yang bisa menganalisis keterkaitan antara variabel prediktor dan variabel respon yang memiliki dua kategori nilai atau bersifat biner. Variabel yang mempengaruhi TPAK yaitu proporsi tenaga kerja formal, tingkat pengangguran terbuka (TPT), Produk Domestik Regional Bruto (PDRB) dan upah rata rata per jam pekerja. Hasil analisis terhadap model terbaik dengan nilai AIC 34,04 menunjukkan bahwa TPT dan rata-rata upah memiliki pengaruh terhadap TPAK, dengan tingkat signifikansi masing-masing sebesar 0,00978 dan 0,06910. Nilai koefisien determinasi sebesar 39,90% mengindikasikan bahwa model mampu menjelaskan sebagian besar variasi TPAK, Sedangkan faktor lainnya dipengaruhi oleh variabel-variabel di luar cakupan model. Tingkat akurasi klasifikasinya mencapai 82,35% menunjukkan bahwa model cukup andal dalam memprediksi status partisipasi angkatan kerja.

Keywords: *Logistik Biner, Tingkat Pengangguran Terbuka, Tingkat Partisipasi Angkatan Kerja*

ARTICLE INFO

Submission received: 06 June 2025

Accepted: 31 August 2025

Revised: 15 August 2025

Published: 31 August 2025

Available on: <https://doi.org/10.32493/sm.v7i2.xxxx>

StatMat: Jurnal Statistika dan Matematika is licenced under a Creative Commons Attribution-ShareAlike 4.0 International License.



1. PENDAHULUAN

Regresi adalah metode statistik dasar yang biasa digunakan untuk mengetahui pengaruh variabel prediktor terhadap variabel respon dalam suatu hubungan sebab-akibat (1). Terdapat beberapa kasus dimana variabel respon yang digunakan menggunakan data kualitatif. Sehingga diperlukan metode statistika untuk menyelesaikan permasalahan tersebut (2). Pemodelan Regresi dengan data kualitatif pada variabel prediktor digunakan untuk mengestimasi nilai variabel respon berdasarkan input tertentu. Sebaliknya, regresi dengan variabel respon yang bersifat kualitatif dimaksudkan untuk menganalisis kemungkinan terjadinya suatu fenomena dalam penelitian (2). Sebagai salah satu bentuk pengembangan metode regresi, regresi logistik digunakan untuk memodelkan variabel respons kategorikal, dengan prediktor yang bisa berasal dari data numerik maupun kategorikal. (3).

Dalam statistika, regresi logistik berfungsi untuk mempelajari variabel respons kualitatif, baik yang diklasifikasikan ke dalam dua kategori maupun lebih dari dua. Tidak seperti regresi linier yang mengasumsikan hubungan langsung antara variabel prediktor dan respons, regresi logistik mengasumsikan keterkaitan linier melalui transformasi logit dari peluang suatu peristiwa. Regresi logistik biner diterapkan ketika variabel respon hanya memiliki dua kemungkinan nilai (4). Berbagai studi sebelumnya telah memanfaatkan pendekatan ini, salah satunya dalam menganalisis faktor-faktor yang mempengaruhi tingkat pencemaran udara (5). Metode regresi logistik biner juga memiliki potensi besar untuk diterapkan dalam sektor ketenagakerjaan dan ekonomi misalnya dalam menganalisis faktor-faktor yang memengaruhi TPAK sebagai salah satu indikator pertumbuhan ekonomi.

Salah satu permasalahan dalam pembangunan ekonomi indonesia adalah masalah ketenagakerjaan. Permasalahan dalam pengembangan lapangan kerja adalah tingginya angka pengangguran karena banyak dunia usaha yang bangkrut. Selain itu, tingkat kualitas dan aktivitas kerja juga rendah. Hal ini terlihat dalam konteks dimana pembangunan ketenagakerjaan bertujuan untuk menciptakan lapangan kerja dan kesempatan berusaha sehingga pekerja dapat mencapai pekerjaan serta pencarian akan wajar guna manusiawi. Meningkatnya sektor ketenagakerjaan berdampak baik pada pembangunan nasional dan pertumbuhan ekonomi suatu negara khususnya di Indonesia sehingga tidak tertinggal oleh negara maju dan berkembang lainnya (6).

Tingkat partisipasi angkatan kerja (TPAK) menentukan maju atau mundurnya suatu negara. Hal ini dikarenakan semakin tinggi tingkat-partisipasi angkatan kerja maka secara tidak langsung pertumbuhan ekonomi suatu negara akan semakin meningkat. Beberapa faktor yang memengaruhi TPAK dapat dilihat dari variabel ekonomi dan ketenagakerjaan yang saling berkaitan. Persentase tenaga kerja formal mencerminkan stabilitas pekerjaan, di mana semakin tinggi persentase ini, semakin banyak pekerja yang memiliki perlindungan hukum dan jaminan sosial (6). Tingkat pengangguran terbuka (TPT) yaitu perbandingan penduduk yang tidak memiliki pekerjaan tetapi aktif dalam mencari kerja, di mana jika nilai TPT tinggi maka dapat menurunkan TPAK secara keseluruhan (7). Di sisi lain, rata-rata upah tenaga kerja berperan signifikan dalam mendorong partisipasi individu ke pasar kerja, karena upah yang kompetitif cenderung meningkatkan minat masyarakat untuk bekerja (8).



Berdasarkan latar belakang yang telah dijelaskan serta mengacu pada sejumlah penelitian sebelumnya, maka peneliti akan melakukan analisis TPAK di Indonesia pada tahun 2023 dengan menggunakan metode regresi logistik biner. Hal ini didasari oleh kenyataan bahwa salah satu isu penting dalam bidang ekonomi, khususnya di Indonesia, adalah pertumbuhan ekonomi yang dipengaruhi oleh dinamika dalam sektor ketenagakerjaan.

2. METODE PENELITIAN

Data yang digunakan dalam studi ini merupakan data Tingkat Partisipasi Angkatan Kerja (TPAK) yang diperoleh dari Badan Pusat Statistik (<https://www.bps.go.id/id>). Adapun tahapan dalam penerapan metode analisis regresi logistik biner adalah sebagai berikut (9):

1. Mengidentifikasi variabel dalam penelitian
2. Melakukan pengumpulan data
3. Melakukan perhitungan statistika deskriptif terhadap data yang telah dikumpulkan
4. Melakukan uji multikolinieritas dengan batas toleransi nilai sebesar 10
5. Melakukan estimasi parameter pada model awal regresi logistik biner
6. Menentukan model terbaik dengan membandingkan nilai AIC terkecil dari beberapa model regresi
7. Melakukan estimasi parameter pada model akhir regresi logistik biner
8. Melakukan pengujian koefisien determinasi dengan menggunakan fungsi logit
9. Melakukan pengujian *Hosmer and Lemeshow* untuk kesesuaian model
10. Melakukan evaluasi ketepatan klasifikasi menggunakan nilai APER
11. Menarik kesimpulan

3. HASIL DAN PEMBAHASAN

3.1 Identifikasi Data

Klasifikasi variabel respon didasarkan pada nilai rata-rata TPAK, yang kemudian dikodekan sesuai kategori tertentu. Kode 0 untuk Status Bekerja (Nilai mean variabel TPAK ≥ 70). Sedangkan kode 1 untuk Status Tidak Bekerja (Nilai mean variabel TPAK < 70), perhitungan statistika deskriptif dari data TPAK ditunjukkan pada Tabel 1.

Tabel 1. Statistika Deskriptif

Variabel	Min	Max	Mean
TPAK (Y)	0,0	1,0	0,4412
Persentase Tenaga Kerja Formal (X_1)	15,57	66,33	40,65
TPT (X_2)	2,270	7,520	4,614
PDRB (X_3)	23078	322615	81948
Upah Rata-Rata Per Jam Pekerja (X_4)	12933	42354	19662

3.2 Analisis Regresi Logistik Biner

3.2.1 Pendekripsi Multikolinieritas



Keberadaan multikolinieritas dalam model dapat mengakibatkan hasil estimasi parameter menjadi tidak stabil serta menurunkan tingkat akurasi model secara keseluruhan. Salah satu cara yang umum dipakai untuk mendeteksi multikolinieritas adalah menggunakan nilai *Variance Inflation Factor* (VIF) (10). Hasil nilai VIF untuk mengetahui terjadi multikolinieritas ditampilkan pada Tabel 2 berikut.

Tabel 2. Multikolinieritas

VIF	X_1	X_2	X_3	X_4
1,438050	1,120680	1,561541	1,429160	

Pada Tabel 2 menunjukkan bahwa seluruh nilai VIF pada variabel memiliki nilai kurang dari 10. Hal ini menunjukkan bahwa di antara variabel variabel prediktor tidak terjadi multikolinieritas.

3.2.2 Estimasi Parameter Model Awal

Setelah melakukan uji asumsi dasar, termasuk multikolinieritas, langkah selanjutnya adalah memperkirakan atau mengestimasi parameter yang digunakan pada model regresi logistik biner yang mencakup seluruh variabel prediktor.

Tabel 3. Estimasi Parameter

Estimasi Parameter	
$\hat{\beta}_0$	$1,347 \times 10^1$
$\hat{\beta}_1$	$-8,691 \times 10^{-2}$
$\hat{\beta}_2$	$-1,177$
$\hat{\beta}_3$	$4,176 \times 10^{-6}$
$\hat{\beta}_4$	$-2,877 \times 10^{-4}$

Berdasarkan Tabel 3 maka model regresi logistik biner dapat dituliskan pada persamaan 1 sebagai berikut.

$$\pi(x) = \frac{\exp(1,347 \times 10^1 - 8,691 \times 10^{-2}X_{1i} - 1,177X_{2i} + 4,176 \times 10^{-6}X_{3i} - 2,877 \times 10^{-4}X_{4i})}{1 + \exp(1,347 \times 10^1 - 8,691 \times 10^{-2}X_{1i} - 1,177X_{2i} + 4,176 \times 10^{-6}X_{3i} - 2,877 \times 10^{-4}X_{4i})} \quad (1)$$

Tabel 4. Nilai *Odds Ratio* Model Awal

Nilai <i>Odds Ratio</i>	
$\exp(\hat{\beta}_1)$	$9,167619 \times 10^{-1}$
$\exp(\hat{\beta}_2)$	$3,082463 \times 10^{-1}$
$\exp(\hat{\beta}_3)$	1,000004
$\exp(\hat{\beta}_4)$	$9,997123 \times 10^{-1}$

Pada Tabel 4 didapatkan bahwa nilai koefisien dari variabel persentase tenaga kerja formal



bernilai negatif terhadap TPAK yaitu sebesar $-8,691 \times 10^{-2}$. Dari nilai *odds rasio* dapat dikatakan bahwa setiap peningkatan persentase tenaga kerja formal, maka kemungkinan status TPAK ≥ 70 akan turun $9,167619 \times 10^{-1}$ kali lipat dengan asumsi variabel TPT, PDRB, dan upah rata-rata per jam pekerja tetap. Nilai koefisien dari variabel TPT bernilai negatif terhadap TPAK yaitu sebesar $-1,177$. Dari nilai *odds rasio* dapat dikatakan bahwa setiap peningkatan TPT, maka kemungkinan status TPAK ≥ 70 akan turun $3,082463 \times 10^{-1}$ kali lipat dengan asumsi variabel persentase tenaga kerja formal, PDRB, dan upah rata-rata per jam pekerja tetap. Nilai koefisien dari variabel PDRB (X_3) bernilai positif terhadap TPAK yaitu sebesar $4,176 \times 10^{-6}$. Dari nilai *odds rasio* dapat dikatakan bahwa setiap peningkatan PDRB, maka kemungkinan status TPAK ≥ 70 akan meningkat 1,000004 kali lipat dengan asumsi variabel persentase tenaga kerja formal, TPT, dan upah rata-rata per jam pekerja tetap. Nilai koefisien dari variabel Upah Rata-rata Per Jam Pekerja (X_4) bernilai negatif terhadap TPAK yaitu sebesar $-2,877 \times 10^{-4}$. Dari nilai *odds rasio* dapat dikatakan bahwa setiap peningkatan upah rata-rata per Jam pekerja, maka kemungkinan status TPAK ≥ 70 akan turun $9,997123 \times 10^{-1}$ kali lipat dengan asumsi variabel persentase tenaga kerja formal, TPT, dan PDRB tetap.

3.2.3 Pemilihan Model Terbaik

Nilai AIC memberikan ukuran relatif kualitas suatu model statistik untuk sekumpulan data tertentu, di mana model dengan nilai AIC yang lebih kecil dianggap lebih baik karena menunjukkan keseimbangan antara kompleksitas model dan kecocokan terhadap data (11). Tabel 5 berikut menyajikan nilai AIC untuk menentukan model yang terbaik.

Tabel 5. Nilai AIC pada Model

Model	Nilai AIC
$Y \sim X_1 + X_2 + X_3 + X_4$	36,46
$Y \sim X_1 + X_2 + X_4$	34,54
$Y \sim X_2 + X_4$	34,04

Berdasarkan Tabel 5. nilai AIC terkecil yaitu $Y \sim X_2 + X_4$ dengan nilai AIC sebesar 34,04 sehingga model tersebut merupakan model terbaik.

3.2.4 Estimasi Parameter Model Terbaik

Proses selanjutnya adalah menghitung estimasi parameter pada model terbaik yaitu dengan menggunakan variabel prediktor X_2 dan X_4 . Estimasi parameter yang terpilih ditampilkan pada Tabel 6 berikut.

Tabel 6. Estimasi Parameter Pada Model Terbaik

Estimasi Parameter	
$\hat{\beta}_0$	10,3010081
$\hat{\beta}_2$	-1,3540230
$\hat{\beta}_4$	-0,0002433

Berdasarkan tabel 6 maka dapat dituliskan persamaan regresi logistik biner model terbaik dengan menggunakan variabel prediktor X_2 dan X_4 yaitu sebagai berikut :



$$\pi(x) = \frac{\exp(10,3010081 - 1,3540230X_{2i} - 0,0002433X_{4i})}{1 + \exp(10,3010081 - 1,3540230X_{2i} - 0,0002433X_{4i})} \quad (2)$$

Tabel 7. Nilai *Odds Ratio* Model Terbaik
Nilai *Odds Ratio*

$\exp(\hat{\beta}_0)$	$2,97621 \times 10^4$
$\exp(\hat{\beta}_1)$	$2,581994 \times 10^{-1}$
$\exp(\hat{\beta}_2)$	$9,997568 \times 10^{-1}$

Nilai koefisien dari variabel TPT bernilai negatif terhadap TPAK yaitu sebesar $-1,3540230$. Dari nilai *odds ratio* dapat dikatakan bahwa setiap peningkatan TPT, maka kemungkinan status TPAK ≥ 70 akan turun $2,581994 \times 10^{-1}$ kali lipat dengan asumsi variabel upah rata-rata per jam pekerja tetap. Nilai koefisien variabel upah rata-rata per jam pekerja bernilai negatif terhadap TPAK yaitu sebesar $-0,0002433$. Dari nilai *odds rasio* dapat dikatakan bahwa setiap peningkatan upah rata-rata per Jam pekerja, maka kemungkinan status TPAK ≥ 70 akan turun $9,997568 \times 10^{-1}$ kali lipat dengan asumsi variabel TPT tetap.

3.2.5 Koefisien Determinasi Model Terbaik

Koefisien determinasi (R^2) ini menggambarkan proporsi variasi total dalam kemampuan variabel prediktor dalam menjelaskan variabel respon (12) pada regresi logistik biner dapat dilihat melalui hasil koefisien determinasi yang tercantum pada Tabel 8.

Tabel 8. Koefisien Determinasi Model

Koefisien Determinasi	
R^2	0,3990495

dapat diketahui koefisien determinasi sebesar 0,3990495 yang menyatakan bahwa 39,90% variasi yang terjadi pada TPAK disebabkan oleh TPT dan Upah Rata-rata Per Jam Pekerja. Sehingga sebesar 60,1% variabel lain yang memengaruhi TPAK di Indonesia masih belum dimasukkan dalam penelitian.

3.2.6 Uji Kesesuaian Model

Pengujian kesesuaian pada model digunakan uji *Hosmer-Lemeshow*, dimana nilai statistik uji ini berdistribusi Chi-Square dengan derajat kebebasan $df = g - 2$ (13). Dengan menggunakan *Hosmer and Lemeshow Test* diperoleh nilai p-value sebesar 0,8675. Karena p-value (0,8675) $> \alpha$ (0,05), maka keputusan yang diambil adalah gagal tolak H_0 . Dengan demikian, dapat disimpulkan bahwa model yang digunakan sudah sesuai karena tidak terdapat perbedaan yang signifikan antara data observasi. Artinya, model ini dapat dikatakan baik dalam menggambarkan data yang diamati.

3.2.7 Ketepatan Klasifikasi

Evaluasi ini dilakukan dengan membandingkan hasil prediksi model regresi logistik



biner terhadap data aktual yang digunakan, evauluasi ini disajikan dalam bentuk tabel klasifikasi (*confusion matrix*) (14). Dari tabel ini, dapat dihitung nilai *Apparent Error Rate* (APER) dan akurasi untuk menilai sejauh mana model mampu mengklasifikasikan data dengan benar. Tabel 9 dan Tabel 10 berikut menyajikan hasil klasifikasi serta nilai ketepatan akurasi model.

Tabel 9. Klasifikasi

Aktual	Prediksi	
	0	1
0	16	3
1	3	12

Tabel 10. Ketepatan Akurasi

APER	Akurasi
17,64706	82,35294

Dilihat dari Tabel 10 banyaknya data yang diklasifikasikan dengan benar adalah 19 dari total data keseluruhan sebanyak 34 data. Adapun nilai akurasi yang diperoleh sebesar 82,35294 sehingga model yang terbentuk dapat memprediksi status TPAK secara tepat dengan tingkat akurasi sebesar 82,35%.

4. SIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, diperoleh model terbaik dari regresi logistik biner pada data ketenagakerjaan yaitu:

$$\pi(x) = \frac{\exp (10,3010081 - 1,3540230X_{1i} - 0,0002433X_{2i})}{1 + \exp (10,3010081 - 1,3540230X_{1i} - 0,0002433X_{2i})}$$

Dari keempat variabel prediktor yaitu variabel persentase tenaga terja formal (X_1), tingkat pengangguran terbuka (X_2), produk domestik regional bruto (X_3) dan upah rata-rata pekerja (X_4) diperoleh kesimpulan bahwa variabel tingkat pengangguran terbuka (X_2) dan upah rata-rata pekerja (X_4) berpengaruh secara signifikan terhadap tingkat partisipasi angkatan kerja. Dari model yang terbentuk memiliki nilai akurasi sebesar 82,35% untuk memprediksi status tingkat partisipasi angkatan kerja secara tepat.

5. DAFTAR PUSTAKA

1. Sulistiowati W, Astuti CC. Statistika Dasar Konsep dan Aplikasinya. Sidoarjo: UMSIDA Press; 2017.
2. Dukalang HH. Perbandingan regresi logistik biner dan probit biner dalam pemodelan tingkat partisipasi angkatan kerja. Euler J Ilm Mat Sains dan Teknol. 2019;7(2):62–70.
3. Afifah DN. Penerapan metode regresi logistik biner pada kesejahteraan rumah tangga di Kabupaten Mojokerto. Universitas Islam Negeri Maulana Malik Ibrahim; 2020.
4. Kotimah MK, Wulandari SP. Model regresi logistik biner stratifikasi pada



- partisipasi ekonomi perempuan di Provinsi Jawa Timur. *J Sains dan Seni ITS*. 2014;3(1):D1–6.
5. Kartikasari D. Analisis Faktor-Faktor yang Mempengaruhi Level Polusi Udara dengan Metode Regresi Logistik Biner. *Mathunesa J Ilm Mat*. 2020;8(1):55–9.
 6. Seventina AF, Kholijah G. Analisis Faktor-Faktor Yang Mempengaruhi Partisipasi Angkatan Kerja Di Provinsi Jambi Menggunakan Metode Regresi Logistik Biner. *JEDMA J Edukasi Mat*. 2024;4(2):53–62.
 7. Sisnita A, Prawoto N. Analisis faktor-faktor yang mempengaruhi tingkat pengangguran terbuka di Provinsi Lampung (Periode 2009-2015). *J Econ Res Soc Sci*. 2017;1(1):1–7.
 8. Pratiwi TH, Wijaya RS. Pengaruh Tingkat Pendidikan, Produk Domestik Regional Bruto, Tingkat Pengangguran Terbuka dan Upah Minimum Terhadap TPAK di Kota Semarang. *Jambura Econ Educ J*. 2024;6(2):581–92.
 9. Ghozi S, Ramli R, Setyani A. Analisis keputusan nasabah dalam memilih jenis bank: Penerapan model regresi logistik biner (Studi kasus pada bank Bri cabang balikpapan). *Media Stat*. 2018;11(1):17–26.
 10. Madany N, Ruliana R, Rais Z. Regresi data panel dan aplikasinya dalam kinerja keuangan terhadap pertumbuhan laba perusahaan IDX LQ45 Bursa Efek Indonesia. *VARIANSI J Stat Its Appl Teach Res*. 2022;4(2):79–94.
 11. Widyaningrum EN, Putri RA, Fathan MA, Safitriani NR. Stock Price Forecasting Using Autoregressive With Exogenous Variable Support Vector Regression (ARX-SVR). *J Mat Stat dan Komputasi*. 2025;21(3):847–54.
 12. Matahelumual NP, Kawet R. Pengaruh budaya organisasi dan komitmen organisasi terhadap produktivitas kerja pegawai pada Biro Organisasi Sekretariat Daerah Provinsi Sulawesi Utara. *J EMBA J Ris Ekon Manajemen, Bisnis Dan Akunt*. 2019;7(1).
 13. Hosmer DW, Hjort NL. Goodness-of-fit processes for logistic regression: simulation results. *Stat Med*. 2002;21(18):2723–38.
 14. Martanti DM, Magdalena F, Ariska NPD, Setiyawati N, Rumboirusi WCB. Dampak pandemi Covid-19 terhadap tenaga kerja formal di Indonesia. *Populasi*. 2021;28(2):52–69.