



SKRIPSI

PENDEKATAN REGRESI *HECKMAN PROBIT TWO-STEP* UNTUK ANALISIS FAKTOR-FAKTOR PENGANGGURAN TERBUKA DI PROVINSI JAWA BARAT

HOLLY PATRYCIA
NPM 21083010024

DOSEN PEMBIMBING

Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU., Asean. Eng
Kartika Maulida Hindrayani, S.Kom., M.Kom

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI SAINS DATA
SURABAYA
2026**

LEMBAR PENGESAHAN

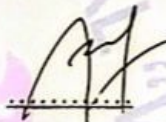
PENDEKATAN REGRESI *HECKMAN PROBIT TWO-STEP* UNTUK ANALISIS FAKTOR-FAKTOR PENGANGGURAN TERBUKA DI PROVINSI JAWA BARAT

Oleh:
HOLLY PATRYCIA
NPM. 21083010024

Telah dipertahankan di hadapan dan diterima oleh Tim Penguji Sidang Skripsi Program Studi
Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur
Pada tanggal 12 Januari 2026

Menyetujui,

Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU., Asean. Eng.
NIP. 19801205 200501 1 002



(Pembimbing I)

Kartika Maulida Hindrayani, S.Kom., M.Kom.
NIP. 19920909 202203 2 009



(Pembimbing II)

Wahyu Syaifullah J. S, S.Kom., M.Kom.
NIP. 19860825 202121 1 003



(Ketua Penguji)

Aviolla Terza Damaliana, S.Si., M.Stat.
NIP. 19940802 202203 2 015



(Penguji I)

Mengetahui,
Dekan Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT.
NIP. 19681126 199403 2 001

LEMBAR PERSETUJUAN

**PENDEKATAN REGRESI *HECKMAN PROBIT TWO-STEP* UNTUK
ANALISIS FAKTOR-FAKTOR PENGANGGURAN TERBUKA DI
PROVINSI JAWA BARAT**

Oleh:
HOLLY PATRYCIA
NPM. 21083010024

Telah disetujui untuk mengikuti Ujian Skripsi

Menyetujui,

**Koordinator Program Studi Sains Data
Fakultas Ilmu Komputer**

Dr. Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU., Asean, Eng.
NIP. 19801205 200501 1 002

SURAT PERNYATAAN BEBAS PLAGIASI

Yang bertanda tangan di bawah ini:

Nama Mahasiswa : Holly Patrycia
NPM : 21083010024
Program : Sarjana (S1)
Program Studi : Sains Data
Fakultas : Fakultas Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila dikemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan dipergunakan sebagaimana mestinya.



Surabaya, 19 Januari 2026
Yang Membuat Pernyataan,



HOLLY PATRYCIA
NPM. 21083010024

ABSTRAK

Nama Mahasiswa / NPM : Holly Patrycia / 21083010024
Judul Skripsi : Pendekatan Regresi *Heckman Probit Two-Step* untuk Analisis Faktor-Faktor Pengangguran Terbuka di Provinsi Jawa Barat
Dosen Pembimbing : 1. Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU., Asean. Eng.
2. Kartika Maulida Hindrayani, S.Kom., M.Kom.

Pengangguran terbuka merupakan permasalahan sosial dan ekonomi yang signifikan di Indonesia, khususnya di Provinsi Jawa Barat yang mencatat tingkat pengangguran terbuka tertinggi secara nasional pada Agustus 2024 sebesar 6,75%. Tingginya angka pengangguran tidak hanya mencerminkan ketidakseimbangan pasar kerja, tetapi juga dipengaruhi oleh faktor demografis, pendidikan, dan kondisi sosial-ekonomi masyarakat. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang memengaruhi pengangguran terbuka dengan menggunakan model *Heckman Probit Two-Step*, yang secara khusus dirancang untuk mengatasi permasalahan *sample selection bias*. Data yang digunakan bersumber dari Survei Angkatan Kerja Nasional (SAKERNAS) BPS tahun 2024 dengan cakupan individu usia kerja di Provinsi Jawa Barat. Tahap pertama analisis memodelkan partisipasi angkatan kerja, sedangkan tahap kedua menganalisis status pengangguran terbuka di antara mereka yang masuk angkatan kerja dengan menambahkan *Inverse Mills Ratio* (IMR) sebagai koreksi bias seleksi. Hasil penelitian menunjukkan bahwa faktor usia, jenis kelamin, tingkat pendidikan, status pernikahan, serta pengalaman pelatihan keterampilan memiliki pengaruh signifikan terhadap probabilitas pengangguran terbuka. Pendekatan ini tidak hanya menghasilkan estimasi yang lebih akurat, tetapi juga memberikan kontribusi metodologis dalam penelitian ketenagakerjaan di Indonesia. Temuan penelitian diharapkan dapat menjadi acuan bagi pemerintah daerah dalam merancang kebijakan ketenagakerjaan yang berbasis data, sehingga strategi pengurangan pengangguran dapat lebih efektif, relevan dengan kebutuhan industri, serta mendukung pertumbuhan ekonomi inklusif di Jawa Barat. Sebagai luaran praktis, penelitian ini juga berhasil merancang Antarmuka Pengguna (*User Interface*) berbasis Streamlit untuk mempermudah analisis data oleh pengguna non-teknis.

Kata kunci : Antarmuka Pengguna, Bias Seleksi, *Heckman Probit Two-Step*, Partisipasi Angkatan kerja, Pengangguran Terbuka.

ABSTRACT

Student Name / NPM : Holly Patrycia / 21083010024
Thesis Title : *Heckman Probit Two-Step Regression Approach for Analyzing the Factors of Open Unemployment in West Java Province*
Advisor : 1. Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU., Asean. Eng.
2. Kartika Maulida Hindrayani, S.Kom., M.Kom.

Open unemployment is a significant social and economic problem in Indonesia, particularly in West Java Province, which recorded the highest national unemployment rate in August 2024 at 6.75%. The high unemployment rate not only reflects labor market imbalances but is also influenced by demographic factors, education, and socio-economic conditions. This study aims to analyze the factors influencing open unemployment using the Heckman Probit Two-Step model, specifically designed to address sample selection bias. The data used are sourced from the 2024 National Labor Force Survey (SAKERNAS) by Statistics Indonesia (BPS), covering working-age individuals in West Java Province. The first stage of the analysis models labor force participation, while the second stage analyzes the status of open unemployment among those entering the labor force by adding the Inverse Mills Ratio (IMR) as a correction for selection bias. The results show that age, gender, education level, marital status, and skills training experience significantly influence the probability of open unemployment. This approach not only produces more accurate estimates but also provides a methodological contribution to labor research in Indonesia. The research findings are expected to serve as a reference for local governments in designing data-driven employment policies, thereby making unemployment reduction strategies more effective, relevant to industry needs, and supporting inclusive economic growth in West Java. As a practical outcome, this research also successfully designed a Streamlit-based User Interface to facilitate data analysis by non-technical users.

Keywords: *Heckman Probit Two-Step, Labor Force Participation, Open Unemployment, Selection Bias, User Interface.*

KATA PENGANTAR

Puji syukur kehadiran Tuhan Yang Maha Esa atas segala rahmat, hidayah dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“Pendekatan Regresi *Heckman Probit Two-Step* untuk Analisis Faktor-Faktor Pengangguran Terbuka di Provinsi Jawa Barat”** dapat terselesaikan dengan baik.

Penulis mengucapkan terima kasih kepada Bapak Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU., Asean. Eng. selaku Dosen Pembimbing utama yang telah meluangkan waktunya untuk memberikan bimbingan, nasehat serta motivasi kepada penulis. Dan penulis juga banyak menerima bantuan dari berbagai pihak, baik itu berupa moril, spiritual maupun materiil. Untuk itu penulis mengucapkan terima kasih kepada:

1. Tuhan Yang Maha Esa yang telah melimpahkan berkat, kasih, dan karunianya sehingga penulis dapat menyelesaikan skripsi ini dengan baik.
2. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Bapak Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU., Asean. Eng. selaku Ketua Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ibu Kartika Maulida Hindrayani, S.Kom., M.Kom. Selaku dosen pembimbing kedua yang telah meluangkan waktu, tenaga, dan pikirannya untuk memberikan arahan, saran, serta bimbingan yang sangat berharga selama proses penyelesaian skripsi ini.
5. Dosen-dosen Program Studi Sains Data Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur, yang dengan penuh dedikasi telah membagikan ilmu dan pengetahuan dari awal perkuliahan hingga sampai ditahap ini.
6. Ibu tercinta, terima kasih atas doa, cinta, dan dukungan tanpa batas. Skripsi ini penulis persembahkan sebagai ungkapan terima kasih atas segala pengorbanan dan kasih sayang yang tak ternilai.

7. Ayah yang telah berpulang, terima kasih atas perjuangan dan cinta semasa hidup. Semoga karya ini menjadi salah satu bentuk bakti dan kebanggaan untukmu.
8. Keluarga tercinta, terima kasih sudah memberikan motivasi, semangat, dan dukungan tiada henti kepada penulis selama penyelesaian skripsi ini.
9. Teman-teman SMP: Jihan, Nathasya, Rania, Shinta dan Tsalas, yang selalu menemani penulis suka maupun duka dari masa SMP hingga sekarang dan seterusnya. Terima kasih atas dukungan dan semangat yang berarti bagi penulis, serta tempat penulis untuk berbagi cerita.
10. Teman-teman SMA: Gading, Shessa, dan Shila, yang juga selalu mendukung dan memberi semangat kepada penulis. Terima kasih atas kebaikan yang berarti yang telah diberikan kepada penulis.
11. Muhimmatul Arofah, Alvi Yuana, dan Alfi Hidayatur, teman-teman seperjuangan kuliah. Terima kasih karena telah menemani, membantu, dan mendukung penulis selama masa perkuliahan dan proses penyelesaian skripsi ini.
12. Seseorang yang namanya tidak bisa disebutkan di sini, terima kasih karena telah memberikan dukungan, perhatian, kesabaran, dan semangat yang sangat berharga yang selalu diberikan kepada penulis dari masa awal perkuliahan hingga penyelesaian skripsi ini.
13. Mark Lee (NCT), terima kasih telah menjadi sumber inspirasi dan motivasi melalui karya, dedikasi, serta energi positif yang diberikan. Semangat dan ketekunanmu menjadi pengingat bagi penulis untuk terus berjuang dan pantang menyerah dalam menyelesaikan skripsi ini.
14. *Last but not least*, kepada diri sendiri, terima kasih atas usaha, keteguhan, dan semangat untuk terus berjuang, meskipun sering merasa lelah namun dapat bertahan hingga sejauh ini dan tidak berhenti mencoba. Penulis bangga dengan pencapaian ini dan berharap ini menjadi awal dari perjalanan yang lebih besar lagi.

Penulis menyadari bahwa di dalam penyusunan skripsi ini banyak terdapat kekurangan. Untuk itu kritik dan saran yang membangun dari semua pihak sangat diharapkan demi kesempurnaan penulisan skripsi ini. Akhirnya, dengan segala

keterbatasan yang penulis miliki semoga laporan ini dapat bermanfaat bagi semua pihak umumnya dan penulis pada khususnya.

Surabaya, Januari 2026

Penulis

DAFTAR ISI

HALAMAN SAMPUL	i
LEMBAR PENGESAHAN	ii
LEMBAR PERSETUJUAN	iii
SURAT PERNYATAAN BEBAS PLAGIASI	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	x
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
DAFTAR LAMPIRAN	xv
DAFTAR NOTASI	xvi
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	5
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian	6
1.5. Manfaat Penelitian	7
BAB II TINJAUAN PUSTAKA	9
2.1. Penelitian Terdahulu	9
2.2. Kerangka Teori.....	12
2.2.1. Pengangguran Terbuka.....	12
2.2.2. Kondisi Pengangguran di Jawa Barat	13
2.2.3. Partisipasi Angkatan Kerja dan Bias Seleksi	15
2.2.4. Regresi dalam Statistik Ekonomi	16
2.2.5. Model Regresi <i>Probit</i>	17
2.2.6. Bias Seleksi dan Model <i>Heckman Probit Two-Step</i>	18
2.2.7. Perbandingan <i>Heckman</i> dan <i>Heckman Probit Two-Step</i>	20
2.2.8. Pendugaan Parameter Model.....	21
2.2.9. Uji Korelasi <i>Pearson</i>	22

2.2.10.	Uji Chi-Square	23
2.2.11.	Uji Mann-Whitney	24
2.2.12.	Uji Multikolinearitas (VIF)	25
2.2.13.	<i>Robust Standard Error</i>	26
2.2.14.	<i>Likelihood Ratio Test</i>	27
2.2.15.	<i>Pseudo R-Square</i>	29
2.2.16.	Efek Marginal.....	30
2.2.17.	Streamlit	31
2.2.18.	Python	32
BAB III METODOLOGI PENELITIAN		35
3.1	Variabel Penelitian dan Sumber Data	35
3.2	Langkah Analisis.....	42
3.2.1.	Studi Literatur	43
3.2.2.	Pengumpulan Data	43
3.2.3.	Prapemrosesan Data	43
3.2.4.	Eksplorasi dan Deskriptif Data	46
3.2.5.	Uji Korelasi	48
3.2.6.	Uji Multikolinearitas (VIF)	49
3.2.7.	Penerapan Model <i>Heckman Probit Two-Step</i> Tahap 1 (Seleksi)..	49
3.2.8.	Penerapan Model <i>Heckman Probit Two-Step</i> Tahap 2 (Hasil)	49
3.2.9.	Penggunaan <i>Robust Standard Error</i>	50
3.2.10.	Uji Signifikansi Model.....	50
3.2.11.	Evaluasi Model.....	51
3.2.12.	Efek Marginal.....	51
3.2.13.	Pengembangan Antarmuka	51
3.3	Desain Sistem.....	52
BAB IV HASIL DAN PEMBAHASAN		57
4.1	Hasil Penelitian	57
4.1.1.	Prapemrosesan Data	57
4.1.2.	Eksplorasi dan Deskripsi Data	80
4.1.3.	Uji Korelasi	106
4.1.4.	Uji Multikolinearitas (VIF)	113

4.1.5.	Penerapan Model.....	116
4.1.6.	Penerapan <i>Robust SE</i>	123
4.1.7.	Uji Signifikansi Model.....	128
4.1.8.	Evaluasi Model.....	135
4.1.9.	Efek Marginal.....	137
4.1.10.	Interpretasi Efek Marginal	139
4.2	Tampilan Antar Muka (<i>Graphical User Interface</i>).....	142
BAB V PENUTUP.....		147
5.1	Kesimpulan	147
5.2	Saran Pengembangan	149
DAFTAR PUSTAKA		151
LAMPIRAN.....		157

DAFTAR GAMBAR

Gambar 2.1 Infografik Pengangguran Terbuka	12
Gambar 2.2 Tingkat Pengangguran Setiap Provinsi RI Agustus 2024	14
Gambar 2.3 Logo Streamlit	31
Gambar 2.4 Logo Python	32
Gambar 3.1 Workflow	42
Gambar 3.2 Desain Sistem	52
Gambar 4.1 Boxplot Jumlah_ART	66
Gambar 4.2 Boxplot Usia	67
Gambar 4.3 Boxplot Durasi_Mencari_Kerja_Usaha_Bulan	67
Gambar 4.4 Boxplot Visualisasi Perbandingan Sebelum dan Sesudah Penanganan Outlier	71
Gambar 4.5 Histogram Variabel Numerik	93
Gambar 4.6 Barplot Variabel Kategori	98
Gambar 4.7 <i>Heatmap</i> Uji Korelasi Variabel Numerik	100
Gambar 4.8 Halaman Beranda <i>User Interface</i>	143
Gambar 4.9 Halaman Dashboard <i>User Interface</i>	144
Gambar 4.10 Halaman <i>Input Data User Interface</i>	144
Gambar 4.11 Halaman Prediksi <i>User Interface</i>	145
Gambar 4.12 Halaman Tentang <i>User Interface</i>	145

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu.....	9
Tabel 3.1 Contoh Dataset Awal Pertama	35
Tabel 3.2 Contoh Dataset Awal Kedua	35
Tabel 3.3 Contoh Dataset Awal Ketiga	35
Tabel 3.4 Metadata Dataset Awal	36
Tabel 3.5 Metadata Dataset Analisis Penelitian	39
Tabel 4.1 Hasil Import Data	58
Tabel 4.2 Hasil Penggabungan Data	59
Tabel 4.3 Statistika Deskriptif Data Awal.....	62
Tabel 4.4 Data Duplikat Dataset Awal.....	63
Tabel 4.5 <i>Missing Value</i> Data Awal.....	64
Tabel 4.6 Statistika Deskriptif Setelah menangani <i>Outlier</i>	70
Tabel 4.7 Statistika Deskriptif pada Dataset Baru.....	82
Tabel 4.8 Distribusi Variabel Target	84
Tabel 4.9 Hasil Uji Korelasi Variabel Numerik dan Variabel Target	108
Tabel 4.10 Hasil Uji Signifikansi Model.....	130
Tabel 4.11 Hasil Evaluasi Model	136

DAFTAR LAMPIRAN

Lampiran 1 Dataset setelah Prapemrosesan Data.....	157
Lampiran 2 Dataset setelah mengubah Tipe Data.....	160
Lampiran 3 Jumlah Nilai Unik Setiap Variabel pada Dataset Baru.....	163
Lampiran 4 Jumlah <i>Missing Value</i> Dataset Baru	166
Lampiran 5 Hasil Eksplorasi Variabel Kategori (Angka)	169
Lampiran 6 Hasil Eksplorasi Variabel Kategori (Presntase).....	171
Lampiran 7 Hasil Uji Korelasi Awal antar Variabel Kategori.....	173
Lampiran 8 Hasil Analisis <i>Crosstab</i>	214
Lampiran 9 Hasil Uji Korelasi Variabel Kategori PAK dan Variabel Target ..	224
Lampiran 10 Hasil Uji Korelasi Variabel Kategori PT dan Variabel Target....	226
Lampiran 11 Hasil Uji Multikolinearitas (VIF) PAK	228
Lampiran 12 Hasil Uji Multikolinearitas (VIF) PT.....	229
Lampiran 13 Hasil Estimasi Model PAK.....	230
Lampiran 14 Hasil Estimasi Model PT	232
Lampiran 15 Hasil Estimasi Model dengan <i>Robust SE</i> PAK	234
Lampiran 16 Hasil Estimasi Model dengan <i>Robust SE</i> PT.....	236
Lampiran 17 Hasil Efek Marginal PAK.....	238
Lampiran 18 Hasil Efek Marginal PT	240
Lampiran 19 <i>Letter of Acceptence</i>	242
Lampiran 20 Kode <i>Script</i>	243

DAFTAR NOTASI

Y_i, Y_j, D_j	:	Variabel terikat (<i>outcome</i>) untuk observasi ke- i
β	:	Parameter yang diestimasi
Z_{1i}, X_{2i}	:	Variabel bebas
ε_i	:	Variabel <i>error</i> acak
δ	:	Parameter regresi
IMR	:	Koefisien koreksi seleksi
λ	:	Koreksi seleksi
$u_{1j} \sim N(0,1)$	:	<i>Error term</i> normal standar
$\hat{\beta}$	:	Nilai estimasi dari koefisien
$SE(\hat{\beta})$	:	Standard <i>error</i>
$ln(f_{full}^{max})$ dan $n(f_{null}^{max})$	:	Nilai likelihood maksimal untuk model penuh dan model konstan saja.
β_k dan a_k	:	Koefisien pada persamaan seleksi dan hasil
ρ	:	Korelasi antar <i>error</i>
σ_n	:	<i>Error</i> standard outcome
D_{pr}	:	<i>Inverse Mills ratio</i> plus probabilitas seleksi