

PENGGUNAAN ANALISIS REGRESI LOGISTIK
MULTINOMIAL UNTUK KLASIFIKASI PEMILIHAN
LAPANGAN PEKERJAAN DI DKI JAKARTA
(Studi Kasus: Survei Angkatan Kerja DKI Jakarta tahun 2015)

Skripsi

Disusun untuk melengkapi syarat-syarat
guna memperoleh gelar Sarjana Sains



JENNYFER ROMAITO

3125121969

PROGRAM STUDI MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI JAKARTA

2017

LEMBAR PERSETUJUAN HASIL SIDANG SKRIPSI

PENGUNAAN ANALISIS REGRESI LOGISTIK MULTINOMIAL UNTUK KLASIFIKASI PEMILIHAN LAPANGAN PEKERJAAN DI DKI JAKARTA

(STUDI KASUS : SURVEI ANGKATAN KERJA DKI JAKARTA
TAHUN 2015)

Nama : Jennyfer Romaito

No. Registrasi : 3125121969

	Nama	Tanda Tangan	Tanggal
Penanggung Jawab			
Dekan	: Prof. Dr. Suyono, M.Si. NIP. 19671218 199303 1 005
Wakil Penanggung Jawab			
Pembantu Dekan I	: Dr. Muktiningsih, M.Si. NIP. 19640511 198903 2 001
Ketua	: Dr. Makmuri, M.Si. NIP. 19640715 198903 1 006
Sekretaris	: Ibnu Hadi, M.Si. NIP. 19810718 200801 1 017
Penguji	: Prof. Dr. Suyono, M.Si. NIP. 19671218 199303 1 005
Pembimbing I	: Dra. Widyanti Rahayu, M.Si. NIP. 19661103 200112 2 001
Pembimbing II	: Vera Maya Santi, M.Si. NIP. 19790531 200501 2 006

Dinyatakan lulus ujian skripsi tanggal: 10 Februari 2017

LEMBAR PENGESAHAN

Dengan ini saya mahasiswa Fakultas Matematika dan Ilmu Pengetahuan
Alam, Universitas Negeri Jakarta

Nama : Jennyfer Romaito

No. Registrasi : 3125121969

Judul : Penggunaan Analisis Regresi Logistik Multinomial Untuk Klasifikasi
Pemilihan Lapangan Pekerjaan di DKI Jakarta

Studi Kasus: Survei Angkatan Kerja DKI Jakarta tahun 2015

Menyatakan bahwa skripsi ini telah siap diajukan untuk sidang skripsi.

Menyetujui,

Dosen Pembimbing I

Dosen Pembimbing II

Drs. Sudarwanto, M.Si., DEA

NIP. 19650325 199303 1 003

Yudi Mahatma, M.Si.

NIP. 19761020 200812 1 001

Mengetahui,

Ketua Jurusan Matematika

Drs. Makmuri, M.Si.

NIP. 19640715 198903 1 006

ABSTRACT

JENNYFER ROMAITO, 3125121969. Application Multinomial Logistic Regression On Career Opportunity Classification. Thesis. Faculty of Mathematics and Natural Science Jakarta State University. 2017.

Multinomial logistic regression can be used for the selection of the classification of career opportunity. The multinomial logit model has a variable response that divided into three categories, such as industry, services, and trade. In this thesis, application on the data using SAKERNAS DKI Jakarta 2015, and industry used as comparison / control. Significance test models simultaneously and partially performed in order to determine the existence of significant influence of the explanatory variables. The multinomial logit model is tested simultaneously and partially significance. Deviation used for goodness of fit model test. Then it takes the best model using the smallest AIC value. Based on multinomial logit models that have been tested can be seen that variables, such as gender, education level, age group, employment status, and the type of work are significant influence on the carrer opportunity classification result.

Keywords : *Career Oppurtunity, multinomial logistic, significance test, AIC, classification result.*

ABSTRAK

JENNYFER ROMAITO, 3125121969. Penggunaan Analisis Regresi Logistik Multinomial Untuk Klasifikasi Pemilihan Lapangan Pekerjaan di DKI Jakarta. Skripsi. Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta. 2017.

Analisis regresi logistik multinomial dapat digunakan untuk klasifikasi pemilihan lapangan pekerjaan. Model multinomial logit mempunyai variabel respon yang terbagi atas tiga kategori, yaitu sektor industri, jasa, dan perdagangan. Pada penulisan skripsi ini penerapan pada data menggunakan Survei Angkatan Kerja (SAKERNAS) DKI Jakarta tahun 2015, kategori sektor industri dijadikan pembanding/kontrol. Uji signifikansi model secara serentak dan parsial dilakukan agar mengetahui adanya pengaruh yang signifikan dari variabel-variabel penjelas. Selanjutnya nilai AIC terkecil digunakan untuk mendapatkan model yang terbaik. Berdasarkan model multinomial logit yang telah diuji dapat diketahui bahwa variabel-variabel, seperti jenis kelamin, tingkat pendidikan, golongan umur, status pekerjaan, dan jenis pekerjaan berpengaruh signifikan terhadap ketepatan klasifikasi pemilihan lapangan pekerjaan di DKI Jakarta.

Kata kunci : pemilihan lapangan pekerjaan, regresi logistik multinomial, uji signifikansi, AIC, ketepatan klasifikasi

PERSEMBAHANKU

”Janganlah hendaknya kamu kuatir tentang apapun juga, tetapi nyatakanlah dalam segala hal keinginanmu kepada Allah dalam doa dan permohonan dengan ucapan syukur”

Filipi 4:6

Skripsi ini penulis persembahkan untuk Tuhan Yesus Kristus.

”Terima kasih atas berkat, pengetahuan, hikmat, dan perlindunganMu”.

KATA PENGANTAR

Puji syukur kepada Tuhan Yesus Kristus atas pengetahuan dan kemampuan sehingga penulis dapat menyelesaikan skripsi yang berjudul "Penggunaan Analisis Regresi Logistik Multinomial Untuk Klasifikasi Pemilihan Lapangan Pekerjaan di DKI Jakarta" yang merupakan salah satu syarat dalam memperoleh gelar Sarjana Sains Program Studi Matematika Universitas Negeri Jakarta.

Skripsi ini dapat diselesaikan tidak terlepas dari adanya bantuan dari berbagai pihak. Oleh karena itu, dalam kesempatan ini penulis ingin menyampaikan terima kasih terutama kepada:

1. Ibu Widyanti Rahayu, M.Si selaku Dosen Pembimbing I dan Vera Maya Santi, M.Si. selaku Dosen Pembimbing II, yang telah meluangkan waktunya dalam memberikan bimbingan, saran, nasehat serta arahan sehingga skripsi ini dapat menjadi lebih baik dan terarah.
2. Ibu Dr. Lukita Ambarwati, S.Pd., M.Si selaku Koordinator Program Studi Matematika FMIPA UNJ yang telah banyak membantu penulis.
3. Ibu Ratna Widyati, S.Si., M.Kom selaku Pembimbing Akademik penulis. Terima kasih atas segala bimbingan dan kerja sama Ibu selama perkuliahan, dan seluruh Bapak/Ibu dosen atas pengajarannya yang telah diberikan, serta karyawan/karyawati FMIPA UNJ yang telah memberikan informasi yang penulis butuhkan dalam menyelesaikan skripsi.
4. Ibu Dr. Lukita Ambarwati, S.Pd., M.Si, Ir. Fariani Hermin, M.T dan Ratna Widyati, S.Si., M.Kom selaku dosen penguji seminar proposal penulis. Terima kasih atas saran yang telah diberikan

5. Bapak Dr. Makmuri, M.Si, Prof. Dr. Suyono, M.Si dan Ibnu Hadi, M.Si selaku dosen penguji skripsi penulis. Terima kasih atas saran yang telah diberikan
6. Mama dan Bapak yang selalu mendoakan dan mendukung penulis.
7. Kedua kakak perempuan penulis, Caroline dan Golda yang selalu memberikan semangat dan mengganggu ketentraman penulis.
8. Keluarga besar penulis. Opung, Tulang, Tante Dedek, Selma, Ilin, dan Taripar yang selalu mendoakan dan mendukung penulis
9. Wilujeng Mugi Sri Mulyaningsih. Teman SMA penulis yang membantu penulis untuk mendapatkan referensi yang penulis butuhkan serta menjadi teman yang dapat berbagi suka dan duka.
10. Teman-teman Matematika UNJ 2012. Terkhusus kepada kalian semua yang selalu membantu penulis dan berbagi suka maupun duka bersama. Terima kasih untuk kebersamaannya. Kalian semua tak akan terlupakan dan tergantikan. Bersama kalian, penulis merasakan indahnyanya pertemanan semasa kuliah.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Saran dan kritik akan sangat berarti. Semoga skripsi ini dapat bermanfaat bagi pembaca sekalian.

Jakarta, Februari 2017

Jennyfer Romaito

DAFTAR ISI

ABSTRACT	i
ABSTRAK	ii
KATA PENGANTAR	iv
DAFTAR ISI	vii
DAFTAR TABEL	viii
DAFTAR GAMBAR	ix
I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Perumusan Masalah	3
1.3 Pembatasan Masalah	3
1.4 Tujuan Penulisan	4
1.5 Manfaat Penulisan	4
1.6 Metode Penulisan	4
II LANDASAN TEORI	5
2.1 Angkatan Kerja DKI Jakarta	5
2.2 Analisis Regresi Linier Berganda	7
2.2.1 Penduga Parameter	8
2.2.2 Pengujian Pada Analisis Regresi Linier Berganda	11
2.2.3 Koefisien Determinasi	13
2.3 Distribusi Multinomial	14
2.4 Analisis Regresi Logistik	16

2.4.1	Uji Signifikansi	18
2.4.2	Uji Kebaikan Model (<i>Goodness Of Fit</i>)	20
2.5	Pemilihan Model Terbaik	21
2.6	Prosedur Klasifikasi	21
III PEMBAHASAN		24
3.1	Model Regresi Logistik Multinomial	24
3.2	Penduga Parameter	26
3.3	Interpretasi Model	29
3.4	Sumber Data	29
3.5	Deskripsi Data	30
3.6	Definisi Operasional	30
3.7	Analisis Deskriptif	33
3.8	Uji Multikolinieritas	39
3.9	Analisis Regresi Logistik Multinomial	40
3.9.1	Uji Signifikansi	40
3.9.2	Uji Kesesuaian Model	42
3.10	Pemilihan Model Terbaik	42
3.11	Interpretasi Model	43
3.12	Ketepatan Klasifikasi	47
IV PENUTUP		48
4.1	Kesimpulan	48
4.2	Saran	49
DAFTAR PUSTAKA		50
LAMPIRAN-LAMPIRAN		51

DAFTAR TABEL

2.1	Ketepatan Klasifikasi	21
3.1	Pilihan Berdasarkan Jenis Kelamin	33
3.2	Pilihan Berdasarkan Tingkat Pendidikan	34
3.3	Pilihan Berdasarkan Golongan Umur	35
3.4	Pilihan Berdasarkan Status Pekerjaan	37
3.5	Pilihan Berdasarkan Jenis Pekerjaan	38
3.6	Nilai VIF	39
3.7	Model Fitting Information	40
3.8	Pendugaan Parameter	41
3.9	Uji Kesesuaian Model	42
3.10	Nilai AIC	43
3.11	Rasio <i>Odd</i>	43
3.12	Ketepatan Klasifikasi	47

DAFTAR GAMBAR

2.1	Diagram Alir Penelitian	23
3.1	Persentase Jenis Kelamin	33
3.2	Persentase Tingkat Pendidikan	34
3.3	Persentase Golongan Umur	35
3.4	Persentase Status Pekerjaan	36
3.5	Persentase Jenis Pekerjaan	38
4.1	Model Fitting Information	51
4.2	Goodness of Fit	51
4.3	Likelihood Ratio Test	52
4.4	Klasifikasi	52

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Setiap manusia membutuhkan pekerjaan untuk kelangsungan hidupnya. Pekerjaan yang dipilih tersebut mempengaruhi berbagai aspek, yaitu kebutuhan ekonomi, sosial, dan psikologis. Lapangan pekerjaan yang beragam membuat seseorang untuk mengambil keputusan dalam memilih pekerjaan yang sesuai dengan kompetensi yang dimiliki. Banyak faktor yang mempengaruhi pemilihan lapangan pekerjaan yang sesuai. Seseorang yang menentukan pilihannya dengan tepat diharapkan dapat memanfaatkan kemampuannya dengan optimal pada setiap sektor lapangan pekerjaan. Dengan demikian, sumber daya manusia yang ada dapat menjadi penggerak faktor-faktor produksi. Jika penggerak faktor-faktor produksi sudah berperan maka terjadilah pembangunan ekonomi. Indeks Pembangunan Manusia (IPM) dapat menjelaskan bagaimana penduduk dapat mengakses hasil pembangunan dalam memperoleh pendapatan, kesehatan, pendidikan, dan lain-lain. Pada tahun 2015, IPM dengan kategori tinggi berada di delapan propinsi, yaitu Riau, DKI Jakarta, DI Yogyakarta, Banten, Bali, Kalimantan Timur, dan Sulawesi Utara (Statistik Indonesia Merdeka, 2015). Salah satu propinsi dengan capaian pembangunan manusia tingkat tinggi adalah DKI Jakarta, yaitu sebesar 78,99%. Jumlah penduduk DKI Jakarta sebanyak 10.187.595 jiwa dengan jumlah angkatan kerja sebanyak 4.724.029 jiwa (BPS, 2015). Sumber daya manusia yang besar menjadi modal bagi pembangunan ekonomi karena menyediakan tenaga kerja

berlimpah sehingga mampu menciptakan nilai tambah bagi produksi daerah jika kualitasnya baik.

Analisis regresi linier berganda digunakan untuk memprediksi/meramalkan nilai variabel respon (Y), lebih baik kalau diperhitungkan variabel-variabel lain yang ikut mempengaruhi. Dengan demikian, diperoleh hubungan antara satu variabel respon (Y) dengan beberapa variabel penjelas (Supranto, 2008). Sehingga analisis regresi linier berganda adalah salah satu metode yang dapat digunakan untuk memilih lapangan pekerjaan utama dengan beberapa variabel yang mempengaruhi.

Namun, terdapat pula metode lain yang dapat membuat model pemilihan lapangan pekerjaan utama dengan beberapa variabel yang mempengaruhi. Metode tersebut adalah analisis regresi logistik. Perbedaan antara regresi logistik dan regresi linier berganda terletak pada beberapa asumsi yang harus dipenuhi. Variabel respon (Y) memiliki tiga kategori sehingga lebih tepat untuk menggunakan regresi logistik multinomial (Hosmer & Lemeshow, 2000). Regresi logistik multinomial tepat digunakan karena variabel respon (Y) berskala nominal. Skala nominal adalah skala yang hanya membedakan kategori berdasarkan jenis atau macamnya. Skala ini tidak membedakan urutan atau tingkatan.

Penelitian sebelumnya yang berkaitan dengan analisis regresi logistik multinomial dan lapangan pekerjaan, diantaranya Lenggogeni (2001) yang telah melakukan penelitian mengenai pemilihan lapangan pekerjaan di wilayah JABOTABEK menggunakan analisis regresi logistik dan Fitriany (2013) yang telah melakukan penelitian mengenai analisis regresi logistik biner dan multinomial pada faktor-faktor yang mempengaruhi IPK mahasiswa suatu universitas. Lenggogeni (2001) membahas mengenai pemilihan lapangan pekerjaan di DKI Jakarta, Bogor, Tangerang, dan Bekasi. Variabel respon terbagi menjadi

tiga kategori, yaitu pertanian, jasa, dan industri, sedangkan pada penulisan ini menggunakan variabel respon yang terbagi menjadi tiga kategori, yaitu industri, jasa, dan perdagangan. Berdasarkan uraian di atas, akan dikaji penggunaan analisis regresi logistik untuk klasifikasi pemilihan lapangan pekerjaan utama DKI Jakarta tahun 2015.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan tersebut, maka perumusan masalah sebagai berikut :

1. Bagaimana model regresi logistik multinomial untuk klasifikasi pemilihan lapangan pekerjaan utama wilayah DKI Jakarta tahun 2015 ?
2. Bagaimana ketepatan klasifikasi pemilihan lapangan pekerjaan utama wilayah DKI Jakarta tahun 2015 ?

1.3 Pembatasan Masalah

Pembatasan masalah dari penulisan ini adalah sebagai berikut :

1. Variabel respon (Y) merupakan variabel multinomial dengan 3 kategori, yaitu sektor industri, jasa dan perdagangan
2. Variabel penjelas (X) sebanyak lima, yaitu jenis kelamin, tingkat pendidikan, golongan umur, status pekerjaan utama, dan jenis pekerjaan utama
3. Data yang digunakan berdasarkan SAKERNAS BPS DKI Jakarta tahun 2015

1.4 Tujuan Penulisan

Tujuan penulisan ini adalah sebagai berikut :

1. Mengidentifikasi model regresi logistik multinomial untuk klasifikasi pemilihan lapangan pekerjaan utama wilayah DKI Jakarta tahun 2015
2. Mengidentifikasi ketepatan klasifikasi pemilihan lapangan pekerjaan utama wilayah DKI Jakarta tahun 2015 ?

1.5 Manfaat Penulisan

Penulisan skripsi ini diharapkan dapat menambahkan pengetahuan tentang regresi logistik multinomial khususnya bagi penulis dan umumnya bagi pembaca. Skripsi ini juga dapat menjadi referensi tambahan bagi pihak lain yang ingin melakukan penelitian mengenai regresi logistik dan ketepatan klasifikasi.

1.6 Metode Penulisan

Metode penulisan pada skripsi ini merupakan aplikasi dari analisis regresi logistik multinomial. Pembahasan dan penggunaan terhadap data yang diberikan merupakan hasil pemikiran yang logis dan sistematis dari beberapa referensi. Referensi utama yang digunakan adalah *Introduction to Linear Regression Analysis* (2012).

BAB II

LANDASAN TEORI

Pada bab ini, akan dijelaskan mengenai tingkat kejahatan dan analisis regresi linier berganda sebagai pendahuluan. Selain itu, dijelaskan pula mengenai distribusi multinomial dan analisis regresi logistik.

2.1 Angkatan Kerja DKI Jakarta

Daerah Khusus Ibukota Jakarta (DKI Jakarta) adalah ibukota negara Republik Indonesia. Menurut Statistik Indonesia Merdeka tahun 2015, DKI Jakarta mempunyai capaian pembangunan manusia tingkat tinggi sebesar 78,39%. Sebagai pusat bisnis, politik, dan kebudayaan, DKI Jakarta merupakan tempat berdirinya kantor-kantor pusat BUMN, perusahaan swasta, dan perusahaan asing. DKI Jakarta merupakan kota dengan tingkat pertumbuhan ekonomi yang cukup pesat. Saat ini, lebih dari 70% uang negara beredar di DKI Jakarta. Perekonomian DKI Jakarta terutama ditunjang oleh sektor perdagangan, jasa, properti, industri, dan keuangan.

Angkatan kerja adalah penduduk usia kerja (15 tahun dan lebih) yang bekerja atau punya pekerjaan namun sementara tidak bekerja dan pengangguran. Indikator ini bermanfaat untuk mengetahui jumlah penduduk yang berpotensi untuk bekerja. Jumlah angkatan kerja di DKI Jakarta sebanyak 4.724.029 orang. Besarnya jumlah angkatan kerja mencerminkan besarnya jumlah penduduk yang masuk ke pasar kerja untuk bekerja atau mencari kerja.

Penduduk bekerja adalah salah satu faktor yang sangat berperan terhadap

pertumbuhan ekonomi, baik dalam ruang lingkup regional maupun nasional. Penduduk bekerja mencapai puncaknya pada umur antara 25-34 tahun. Kemudian terjadi penurunan jumlah penduduk bekerja setelah usia 35 tahun. Selanjutnya, jika dilihat menurut kelompok umur dan jenis kelamin, persentase penduduk bekerja laki-laki lebih tinggi daripada perempuan pada semua kelompok umur, kecuali kelompok umur 15-19 tahun. Penduduk bekerja di DKI Jakarta pada tahun 2015, didominasi oleh mereka yang berpendidikan Sekolah Menengah Atas, SLTA Umum dan SLTA Kejuruan masing-masing sebesar 23,37% dan 18,96%. Sedangkan penduduk bekerja yang berpendidikan Diploma hanya mencapai 5,77% dan pekerja yang berpendidikan S1 ke atas sebesar 19,02%.

Tiga sektor utama yang banyak menyerap pekerja di DKI Jakarta dari tahun 2013-2015 adalah sektor perdagangan, jasa, dan industri. Beberapa fakta yang telah disebutkan berdasarkan SAKERNAS BPS DKI Jakarta tahun 2015.

2.2 Analisis Regresi Linier Berganda

Analisis regresi adalah suatu teknik statistika untuk menyelidiki dan memodelkan hubungan antar variabel. Analisis regresi linier digunakan untuk menyelidiki dan memodelkan hubungan antar satu variabel penjelas dan variabel respon. Akan tetapi, untuk regresi linier berganda digunakan untuk menyelidiki dan memodelkan hubungan antar beberapa variabel penjelas dan variabel respon.

Pada umumnya, model regresi linier berganda dengan n pengamatan sebagai berikut:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (2.1)$$

dengan desain matrik sebagai berikut:

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}; \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \dots \\ \beta_k \end{pmatrix}; \mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}; \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

dengan

\mathbf{y} adalah vektor berordo $n \times 1$

\mathbf{X} adalah vektor koefisien regresi berordo $n \times p$

$\boldsymbol{\beta}$ adalah vektor berordo $p \times 1$

$\boldsymbol{\varepsilon}$ adalah vektor galat acak berordo $n \times 1$

Asumsi-asumsi pada model regresi linier berganda adalah sebagai berikut:

1. $E(\varepsilon) = 0$ atau rata-rata dari galat adalah nol untuk setiap himpunan nilai variabel penjelas
2. $Var(\varepsilon) = \sigma^2$ atau variansi dari galat adalah konstan
3. Galat tak berkorelasi

4. Galat berdistribusi normal

5. Tidak terjadi multikolinearitas antar variabel bebas ($VIF < 10$)

Persamaan (2.1) dapat dituliskan pula sebagai berikut:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} \quad (2.2)$$

atau

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i, i = 1, 2, \dots, n \quad (2.3)$$

2.2.1 Penduga Parameter

Metode estimasi yang digunakan adalah metode *least square* dan *maximum likelihood*.

Metode *least square* dapat digunakan untuk menduga koefisien regresi dengan persamaan (2.1). Asumsikan bahwa galat ε pada model mempunyai $E(\varepsilon) = 0$, $Var(\varepsilon) = \sigma^2$, dan galat tidak berkorelasi.

Dapat dituliskan fungsi *least square* dengan mengacu pada persamaan (2.3) adalah sebagai berikut:

$$S(\beta_0, \dots, \beta_1, \dots, \beta_k) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2 \quad (2.4)$$

Fungsi S diminimumkan terhadap $\beta_0, \dots, \beta_1, \dots, \beta_k$ sehingga penduga *least square* dari $\beta_0, \dots, \beta_1, \dots, \beta_k$ memenuhi

$$\left. \frac{\partial S}{\partial \beta_0} \right|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j x_{ij} \right) = 0 \quad (2.5)$$

dan

$$\left. \frac{\partial S}{\partial \beta_j} \right|_{\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k} = -2 \sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j x_{ij} \right) x_{ij} = 0, j = 1, 2, \dots, k \quad (2.6)$$

Akan ditentukan vektor penduga *least square*, $\hat{\beta}$, yang meminimumkan

$$\begin{aligned}
 S(\beta) &= \sum_{i=1}^n \varepsilon_i^2 \\
 &= \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} \\
 &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\
 &= \mathbf{y}'\mathbf{y} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\
 &= \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}
 \end{aligned}$$

Penduga *least square* memenuhi

$$\left. \frac{\partial S}{\partial \boldsymbol{\beta}} \right|_{\hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = 0 \quad (2.7)$$

yang mana menjadi sederhana sebagai **persamaan normal *least square*** berikut:

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y} \quad (2.8)$$

sehingga untuk mendapatkan penduga *least square* dari $\boldsymbol{\beta}$ dengan mengalikan kedua sisi dari (2.8) dengan $(\mathbf{X}'\mathbf{X})^{-1}$. Matriks $(\mathbf{X}'\mathbf{X})^{-1}$ akan selalu ada jika variabel penjelas bebas linier (Montgomery., et all, 2012). Jika variabel penjelas bebas linier maka invers matriks X ada. Matriks $(\mathbf{X}'\mathbf{X})$ adalah matriks simetris dengan orde $p \times p$. Jika $(\mathbf{X})^{-1}$ ada maka $(\mathbf{X}')^{-1}$ ada dengan $(\mathbf{X}')^{-1} = (\mathbf{X}^{-1})'$ (Anton & Rorres, 2004). Hal tersebut mengakibatkan bahwa $\mathbf{X}^{-1}(\mathbf{X}')^{-1} = (\mathbf{X}'\mathbf{X})^{-1}$ ada. Dengan demikian, penduga *least square* dari $\boldsymbol{\beta}$ adalah

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (2.9)$$

Pada metode *maximum likelihood* akan ditunjukkan penduga untuk pa-

parameter model regresi linier berganda didapatkan ketika galat berdistribusi normal independen.

Model yang digunakan adalah sebagai berikut:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.10)$$

dan galat berdistribusi normal independen dengan variansi konstan σ^2 , atau $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$. Fungsi kepadatan normal galat adalah sebagai berikut:

$$f(\varepsilon_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}\varepsilon_i^2}$$

Fungsi *likelihood* adalah fungsi kepadatan bersama dari $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ atau $\prod_{i=1}^n f(\varepsilon_i)$. Oleh karena itu, fungsi *likelihood* menjadi sebagai berikut:

$$L(\boldsymbol{\varepsilon}, \boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^n f(\varepsilon_i) = \frac{e^{-\frac{1}{2\sigma^2}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}}{(2\pi)^{\frac{n}{2}}\sigma^n}$$

Anggap $\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, sehingga fungsi *likelihood* menjadi sebagai berikut:

$$L(\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2) = \frac{e^{-\frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})}}{(2\pi)^{\frac{n}{2}}\sigma^n} \quad (2.11)$$

sehingga

$$\ln L(\mathbf{y}, \mathbf{X}, \boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\ln(2\pi) - n\ln(\sigma) - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Jelas bahwa *loglikelihood* maksimum dengan meminimumkan $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$. Dengan demikian, penduga *maximum likelihood* dari $\boldsymbol{\beta}$ adalah

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

2.2.2 Pengujian Pada Analisis Regresi Linier Berganda

Jika telah dilakukan pendugaan parameter pada regresi linier berganda maka diperlukan adanya pengujian parameter tersebut. Terdapat dua pengujian pada analisis regresi linier berganda, yaitu uji serentak dan uji parsial.

Pengujian serentak terhadap model dilakukan untuk mengetahui apakah model yang terbentuk cukup baik, artinya terdapat sedikitnya satu variabel penjelas yang memberikan kontribusi yang cukup dalam memprediksi variabel respon (Y). Prosedur untuk melakukan pengujian model adalah sebagai berikut :

Hipotesis :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

$$H_1 : \beta_j \neq 0 \text{ untuk setidaknya satu } j$$

Statistik Uji :

$$F_{hitung} = \frac{SS_{Reg}/k}{SS_{Res}/(n-(k+1))}$$

dimana

$$SS_T = SS_{Reg} + SS_{Res}$$

$$SS_T = \mathbf{y}'\mathbf{y} - \frac{(\sum_{i=1}^n y_i)^2}{n}$$

$$SS_{Reg} = \hat{\boldsymbol{\beta}}\mathbf{X}'\mathbf{y} - \frac{(\sum_{i=1}^n y_i)^2}{n}$$

$$SS_{Res} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}\mathbf{X}'\mathbf{y}$$

Daerah Penolakan :

$$F_{hitung} > F_{\alpha, k, n-k-1}$$

Aturan Keputusan :

Jika $F_{hitung} > F_{\alpha, k, n-k-1}$, maka H_0 ditolak pada tingkat signifikansi α

Pengujian untuk koefisien regresi (β) secara parsial dilakukan untuk mengetahui variabel-variabel penjelas yang dapat memberikan pengaruh yang cukup dalam memprediksi variabel respon (Y). Prosedur untuk melakukan pengujian koefisien regresi (β) secara parsial adalah sebagai berikut :

Hipotesis :

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0$$

Statistik Uji :

$$t_0 = \frac{\widehat{\beta}_j}{\sqrt{\widehat{\sigma}^2 C_{jj}}}$$

dimana

C_{jj} adalah elemen diagonal dari matriks $(\mathbf{X}'\mathbf{X})^{-1}$

Daerah Penolakan :

$$|t_0| > t_{\alpha/2, n-k-1}$$

Aturan Keputusan :

Jika $|t_0| > t_{\alpha/2, n-k-1}$, maka H_0 ditolak pada tingkat signifikansi α

2.2.3 Koefisien Determinasi

Koefisien determinasi (R^2) merupakan besaran untuk menunjukkan tingkat kekuatan hubungan antara dua variabel atau lebih dalam bentuk persen (menunjukkan besar persentase variasi variabel respon yang dapat dijelaskan oleh variabel penjelas) atau dengan kata lain besar pengaruh variabel penjelas terhadap variabel respon (Montgomery, Peck & Vining, 2012).

Koefisien determinasi dapat didefinisikan sebagai berikut:

$$R^2 = \frac{SS_{Reg}}{SS_T}$$

dengan

$$0 \leq R^2 \leq 1$$

Apabila koefisien determinasi bernilai nol berarti bahwa tidak ada hubungan antara variabel penjelas dengan variabel respon, sedangkan apabila koefisien determinasi bernilai satu berarti bahwa model yang sesuai dapat menjelaskan semua variabilitas dalam variabel respon. Dengan kata lain, semakin besar R^2 maka menunjukkan model regresi linier berganda yang terbentuk semakin baik.

2.3 Distribusi Multinomial

Andaikan pada percobaan binom terdapat setiap ulangan menghasilkan lebih dari dua kemungkinan hasil, maka percobaan itu disebut dengan percobaan multinomial. Distribusi multinomial merupakan bentuk umum dari distribusi binomial untuk lebih dari dua kategori. Andaikan bahwa terdapat percobaan acak diulang sebanyak n kali yang identik dan saling bebas (iid). Pada setiap pengulangan diasumsikan percobaan berakhir dalam salah satu k cara yang *mutually exclusive* dan *exhaustive*, sebut saja C_1, C_2, \dots, C_k . Maksud dari *mutually exclusive* dan *exhaustive* adalah $C_i \cap C_j = \emptyset, i \neq j$, dimana $i, j = 1, 2, \dots, k$ dan $P(C_1) + P(C_2) + \dots + P(C_k) = 1$. X_1, \dots, X_k berdistribusi multinomial jika mempunyai fungsi kepadatan peluang sebagai berikut:

$$f(x_1, x_2, \dots, x_k) = \frac{n!}{x_1! \dots x_{k-1}! x_k!} p_1^{x_1} \dots p_{k-1}^{x_{k-1}} p_k^{x_k} \quad (2.12)$$

dengan :

$$\sum_{i=1}^k x_i = n$$

Mean dan variansi dari distribusi multinomial didapatkan dengan menggunakan fungsi pembangkit momen.

$$\begin{aligned} M(t) &= M(t_1, \dots, t_k) \\ &= E[e^{\sum_{i=1}^k t_i x_i}] \\ &= \left[\sum_{i=1}^k p_i e^{t_i} \right]^n \end{aligned}$$

dan

$$\begin{aligned} K(t) &= K(t_1, \dots, t_k) \\ &= \ln[M(t_1, \dots, t_k)] \\ &= n \ln \left[\sum_{i=1}^k p_i e^{t_i} \right] \end{aligned}$$

Maka diperoleh $E(X_i)$ dan $\text{Var}(X_i)$ untuk distribusi multinomial sebagai berikut:

$$\begin{aligned} E(X_i) &= \left. \frac{\partial K(t_1, \dots, t_k)}{\partial t_i} \right|_{t=0} \\ &= \left. \left(\frac{np_i e^{t_i}}{\sum_{i=1}^k p_i e^{t_i}} \right) \right|_{t=0} \\ &= np_i \end{aligned}$$

dan

$$\begin{aligned} \text{Var}(X_i) &= \left. \frac{\partial^2 K(t_1, \dots, t_k)}{\partial t_i^2} \right|_{t=0} \\ &= \left. \left[\frac{np_i e^{t_i}}{\sum_{i=1}^k p_i e^{t_i}} - \frac{np_i^2 e^{2t_i}}{\left(\sum_{i=1}^k p_i e^{t_i} \right)^2} \right] \right|_{t=0} \\ &= np_i - np_i^2 \\ &= np_i(1 - np_i) \end{aligned}$$

2.4 Analisis Regresi Logistik

Analisis regresi adalah salah satu metode yang sangat populer dalam menggambarkan hubungan antara suatu variabel respon dan satu atau lebih variabel penjelas. Seringkali kasus yang terjadi adalah variabel respon yang berupa variabel diskrit dengan dua atau lebih nilai (Hosmer & Lemeshow 2000). Pada analisis regresi logistik tidak diasumsikan adanya hubungan linier antara variabel respon dan variabel penjelas. Variabel penjelas tidak perlu berupa interval dan galat tidak harus berdistribusi normal (Tabanick., et all, 2001). Tujuan dari menggunakan analisis regresi logistik adalah untuk menentukan model yang tepat untuk menggambarkan hubungan antara suatu variabel respon dan suatu himpunan variabel penjelas. Apabila variabel respon berskala biner maka disebut regresi logistik biner, jika variabel respon berskala ordinal maka disebut regresi logistik ordinal, sedangkan variabel respon berskala nominal maka disebut regresi logistik nominal. Pada regresi logistik multinomial terdapat variabel respon yang terdiri dari beberapa kategori. Kategori tersebut salah satunya dijadikan pembanding/kontrol. Penentuan kategori pembanding/kontrol tidak membutuhkan aturan khusus atau dapat dilakukan secara bebas (Hosmer & Lemeshow, 2000).

Andaikan terdapat model dengan bentuk sebagai berikut:

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i \quad (2.13)$$

dimana

y_i adalah variabel respon yang terdiri atas dua kategori, yaitu $y = 1$ (sukses) dan $y = 0$ (gagal). Nilai ekspektasi dari variabel respon adalah sebagai berikut:

$$E(y_i) = 1(\pi_i) + 0(1 - \pi_i) = \pi_i$$

dengan

$$0 \leq E(y_i) = \pi_i \leq 1$$

Bentuk umum model regresi logistik dapat diformulasikan sebagai berikut:

$$E(y) = \pi = \frac{\exp(\mathbf{x}'\boldsymbol{\beta})}{1 + \exp(\mathbf{x}'\boldsymbol{\beta})} \quad (2.14)$$

Fungsi logit pada persamaan (2.14) merupakan fungsi non linier sehingga perlu dilakukan transformasi untuk memperoleh fungsi yang linier agar hubungan antara variabel respon dan variabel penjelasnya dapat terlihat (Hosmer & Lemeshow, 2000). Selanjutnya akan didefinisikan transformasi dari $E(y)$, yaitu:

$$\begin{aligned} 1 - \pi &= 1 - \frac{\exp(\mathbf{x}'\boldsymbol{\beta})}{1 + \exp(\mathbf{x}'\boldsymbol{\beta})} \\ &= \frac{1}{1 + \exp(\mathbf{x}'\boldsymbol{\beta})} \\ \frac{\pi}{1 - \pi} &= \frac{\frac{\exp(\mathbf{x}'\boldsymbol{\beta})}{1 + \exp(\mathbf{x}'\boldsymbol{\beta})}}{\frac{1}{1 + \exp(\mathbf{x}'\boldsymbol{\beta})}} \\ &= \exp(\mathbf{x}'\boldsymbol{\beta}) \end{aligned}$$

sehingga

$$\eta = \ln\left(\frac{\pi}{1 - \pi}\right) = \mathbf{x}'\boldsymbol{\beta} \quad (2.15)$$

Fungsi $\eta = \ln\left(\frac{\pi}{1 - \pi}\right)$ pada persamaan (2.15) merupakan fungsi transformasi logit. Fungsi logit ini menjelaskan hubungan linier antara variabel penjelas dan variabel respon.

2.4.1 Uji Signifikansi

Setelah mendapatkan dugaan parameter dari model logit maka perlu dilakukan uji signifikansi model dan parameter. Pada analisis regresi logistik, untuk mendapatkan model sederhana yang cocok dengan data perlu dilakukan dua pengujian, yaitu uji rasio *likelihood* (*likelihood ratio test*) dan uji Wald.

Untuk menguji signifikansi model diperlukan suatu pengujian. Untuk menguji signifikansi model digunakan uji rasio *likelihood* (*likelihood ratio test*) yang dapat melihat pengaruh seluruh variabel penjelas dalam model secara serentak. Uji rasio *likelihood* didasarkan pada perbandingan antara model yang menyertakan variabel penjelas dengan model tanpa variabel penjelas. Prosedur untuk melakukan pengujian model sebagai berikut:

Hipotesis:

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0$$

Artinya tidak ada variabel penjelas yang berpengaruh terhadap variabel respon

$$H_1 : \text{Minimal ada satu } \beta_k \neq 0$$

Artinya minimal ada satu variabel penjelas yang berpengaruh terhadap variabel respon

Statistik Uji:

$$LR = 2 \left\{ \sum_{i=1}^n y_i \ln \hat{\pi}_i + \sum_{i=1}^n (n_i - y_i) \ln (1 - \hat{\pi}_i) - [y \ln (y) + (n - y) \ln (n - y) - n \ln (n)] \right\}$$

Daerah Penolakan:

$$LR > \chi_{\alpha, k}^2$$

Aturan Keputusan:

Jika $LR > \chi_{\alpha, k}^2$, maka H_0 ditolak pada tingkat signifikansi α . Artinya seluruh variabel penjelas dikatakan signifikan terhadap model.

Uji Wald digunakan untuk menguji signifikansi parameter secara parsial (Hosmer & Lemeshow, 2000). Prosedur untuk melakukan pengujian parameter adalah sebagai berikut:

Hipotesis:

$$H_0 : \beta_k = 0$$

Artinya tidak ada pengaruh antara variabel penjelas dengan variabel respon

$$H_1 : \beta_k \neq 0$$

Artinya ada pengaruh antara variabel penjelas dengan variabel respon

Statistik Uji:

$$W_k = \left(\frac{\hat{\beta}_k}{SE(\hat{\beta}_k)} \right)^2$$

dimana

$\hat{\beta}_k$ adalah penduga parameter β_k

$SE(\hat{\beta}_k)$ adalah galat baku untuk penduga parameter β_k

Daerah Penolakan:

$$W_k > \chi_\alpha^2$$

Aturan Keputusan:

Jika $W_k > \chi_\alpha^2$, maka H_0 ditolak pada tingkat signifikansi α . W_k berdistribusi *Chi-square* dengan derajat bebas 1. Artinya variabel penjelas ke- k secara parsial berpengaruh terhadap variabel respon.

2.4.2 Uji Kebaikan Model (*Goodness Of Fit*)

Ukuran kesesuaian antara sebuah pengamatan dari suatu variabel respon dengan nilai prediksinya dinamakan residual. Ukuran ini memberikan informasi seberapa baik model cocok dengan setiap pengamatan dalam data.

Residual dapat digunakan untuk memeriksa dan menyelidiki kebaikan model regresi logistik. Residual didefinisikan sebagai berikut:

$$e_i = y_i - \hat{y}_i = y_i - n_i \hat{\pi}_i, i = 1, 2, \dots, n \quad (2.16)$$

Hipotesis yang digunakan untuk menguji kebaikan model sebagai berikut:

H_0 : model sesuai

H_1 : model tidak sesuai

Residual yang digunakan dalam model logistik adalah residual deviansi. Residual deviansi dibentuk dari Deviansi yang diperoleh setelah mencocokkan sebuah model logistik terhadap data binomial. Untuk menguji kebaikan model regresi logistik dapat menggunakan deviansi sebagai berikut:

$$D = 2 \sum_{i=1}^n \left[y_i \ln \left(\frac{y_i}{n_i - \pi_i} \right) + (n_i - y_i) \ln \left(\frac{n_i - y_i}{n_i (1 - \hat{\pi}_i)} \right) \right] \quad (2.17)$$

dimana

D mempunyai distribusi χ_{n-p}^2

Jika $D \leq \chi_{n-p}^2$ maka H_0 diterima. Artinya model sesuai.

2.5 Pemilihan Model Terbaik

Pemilihan model terbaik dilakukan dengan melihat nilai *Akaike Information Criterion*(AIC). Model yang terbaik adalah model yang memiliki nilai AIC terkecil (Yanagihara, H.,et all,2012). Adapun persamaan yang digunakan untuk menghitung nilai AIC adalah sebagai berikut:

$$AIC = 2\ln L(\beta) + 2k$$

dimana

k adalah banyak parameter yang digunakan dalam model.

2.6 Prosedur Klasifikasi

Prosedur klasifikasi adalah peluang kebenaran dan kesalahan klasifikasi dengan ukuran *Hit Ratio* dan *Apparent error rate* (APER). Nilai *Hit Ratio* menyatakan nilai proporsi sampel yang benar diklasifikasi oleh fungsi klasifikasi dan nilai APER menyatakan nilai proporsi sampel yang salah diklasifikasikan (Johnson & Wichern, 1992).

Tabel 2.1: Ketepatan Klasifikasi

Actual Membership	Predicted Membership			Total
	$\hat{\pi}_1$	$\hat{\pi}_2$	$\hat{\pi}_3$	
π_1	n_{11}	n_{12}	n_{13}	A
π_2	n_{21}	n_{22}	n_{23}	B
π_3	n_{31}	n_{32}	n_{33}	C
Total	D	E	F	G

Keterangan :

n_{11} = Jumlah observasi dari kelompok 1 tepat diprediksi sebagai kelompok 1

n_{12} = Jumlah observasi dari kelompok 1 diprediksi sebagai kelompok 2

n_{13} = Jumlah observasi dari kelompok 1 diprediksi sebagai kelompok 3

n_{21} = Jumlah observasi dari kelompok 2 diprediksi sebagai kelompok 1

n_{22} = Jumlah observasi dari kelompok 2 tepat diprediksi sebagai kelompok 2

n_{23} = Jumlah observasi dari kelompok 2 diprediksi sebagai kelompok 3

n_{31} = Jumlah observasi dari kelompok 3 diprediksi sebagai kelompok 1

n_{32} = Jumlah observasi dari kelompok 3 diprediksi sebagai kelompok 2

n_{33} = Jumlah observasi dari kelompok 3 tepat diprediksi sebagai kelompok 3

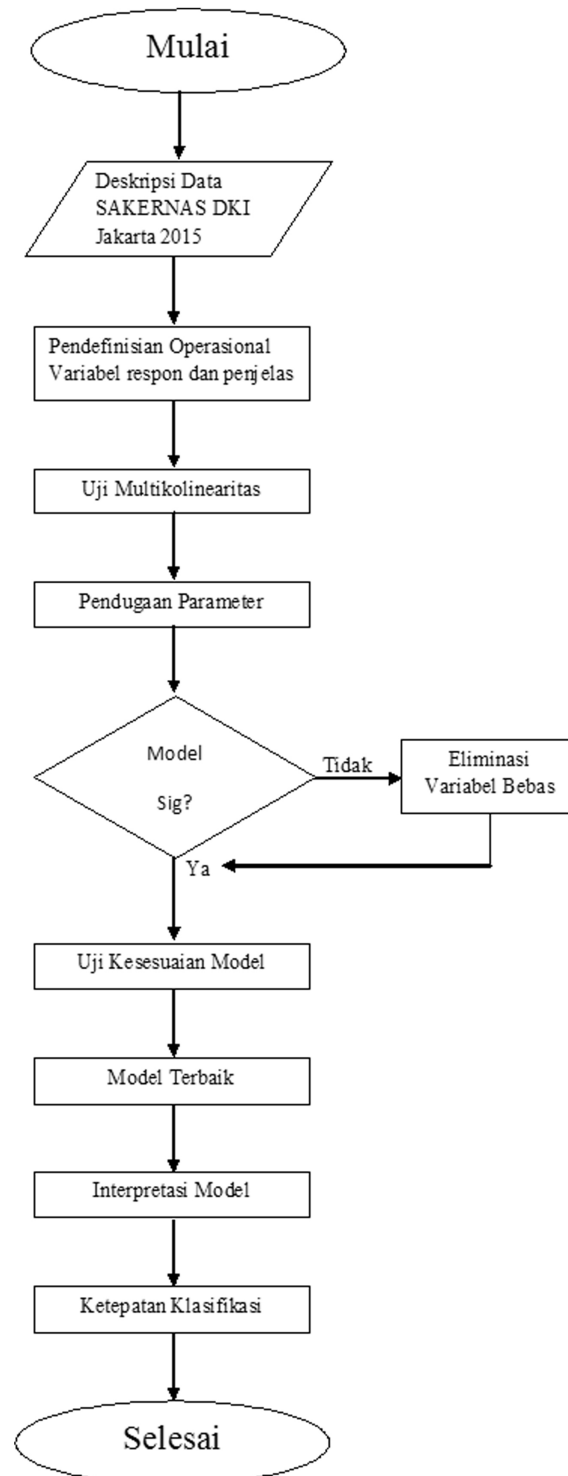
A, B, C = Jumlah observasi dari kelompok 1,2,3

D, E, F = Jumlah prediksi dari kelompok 1,2,3

G = Total dari ketiga kelompok

Rumus ketepatan klasifikasi tersebut adalah sebagai berikut :

1. Ketepatan pada π_1 (dalam persen) untuk nilai *Hit Ratio* adalah $\frac{n_{11}}{A}$ dan nilai APERnya adalah $1 - (\frac{n_{11}}{A} \times 100\%)$
2. Ketepatan pada π_2 (dalam persen) untuk nilai *Hit Ratio* adalah $\frac{n_{22}}{B}$ dan nilai APERnya adalah $1 - (\frac{n_{22}}{B} \times 100\%)$
3. Ketepatan pada π_3 (dalam persen) untuk nilai *Hit Ratio* adalah $\frac{n_{33}}{C}$ dan nilai APERnya adalah $1 - (\frac{n_{33}}{C} \times 100\%)$
4. Ketepatan secara keseluruhan untuk nilai *Hit Ratio* adalah $\frac{n_{11}+n_{22}+n_{33}}{G} \times 100\%$



Gambar 2.1: Diagram Alir Penelitian

BAB III

PEMBAHASAN

3.1 Model Regresi Logistik Multinomial

Model regresi logistik multinomial adalah model yang digunakan jika variabel respon mempunyai lebih dari dua kategori. Generalisasi untuk variabel respon dengan lebih dua kategori dapat dilakukan. Bentuk umum model regresi logistik dapat diformulasikan seperti pada persamaan (2.14).

Model dengan tiga kategori mempunyai dua buah fungsi logit. Fungsi logit dalam $Y = 1$ banding $Y = 0$ dan fungsi logit dalam $Y = 2$ banding $Y = 0$, dimana kategori kode $Y = 0$ berfungsi sebagai nilai pembandingnya. Fungsi logit yang membandingkan $Y = 2$ dengan $Y = 1$ didapat sebagai selisih dari fungsi logit dalam $Y = 2$ banding $Y = 0$ dan fungsi logit dalam $Y = 1$ banding $Y = 0$.

Andaikan \mathbf{x} sebagai vektor variabel penjelas yang berukuran $k + 1$ dengan $x_0 = 1$. Dua buah fungsi logitnya dapat dituliskan sebagai berikut:

$$\begin{aligned} g_1(x) &= \ln \left(\frac{\pi_1(x)}{\pi_0(x)} \right) \\ &= \ln \left[\frac{P(Y = 1|x)}{P(Y = 0|x)} \right] \\ &= \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \dots + \beta_{1k}x_k \\ &= \mathbf{x}'\boldsymbol{\beta}_1 \end{aligned}$$

dan

$$\begin{aligned}
g_2(x) &= \ln\left(\frac{\pi_2(x)}{\pi_0(x)}\right) \\
&= \ln\left[\frac{P(Y=2|x)}{P(Y=0|x)}\right] \\
&= \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \dots + \beta_{2k}x_k \\
&= \mathbf{x}'\boldsymbol{\beta}_2
\end{aligned}$$

Bentuk peluang bersyarat dari setiap kategori respon atau model regresi logistik multinomial sebagai berikut:

$$P(Y=0|x) = \pi_0(x) = \frac{1}{1+e^{g_1(x)}+e^{g_2(x)}}$$

$$P(Y=1|x) = \pi_1(x) = \frac{e^{g_1(x)}}{1+e^{g_1(x)}+e^{g_2(x)}}$$

$$P(Y=2|x) = \pi_2(x) = \frac{e^{g_2(x)}}{1+e^{g_1(x)}+e^{g_2(x)}}$$

Mengikuti ketentuan dalam model biner, ambil $\pi_j(x) = P(Y=j|x)$ untuk $j = 0, 1, 2$ yang masing-masing adalah sebuah fungsi dalam bentuk vektor dengan sebanyak $2(k+1)$ parameter $\boldsymbol{\beta}' = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')$ dimana vektor $\boldsymbol{\beta}_0 = 0$ dan $g_0(x) = 0$. Bentuk umum untuk peluang bersyarat pada ketiga kategori model tersebut dapat dituliskan sebagai berikut:

$$P(Y=j|\mathbf{x}) = \pi_j(x) = \frac{e^{g_j(x)}}{\sum_{j=0}^2 e^{g_j(x)}}$$

3.2 Penduga Parameter

Penduga parameter model regresi logistik multinomial dilakukan dengan metode maksimum *likelihood*. Prinsip utama dari metode maksimum *likelihood* adalah mencari β yang dapat memaksimumkan fungsi *likelihood*. Fungsi ini menyatakan peluang bersama dari data hasil pengamatan yang masih merupakan fungsi dari parameter yang tidak diketahui. Jika variabel respon Y mempunyai lebih dari dua hasil yang mungkin, maka variabel respon Y berdistribusi multinomial. Variabel respon dikode sebagai berikut:

1. Jika $Y = 0$ maka $Y_0 = 1, Y_1 = 0, Y_2 = 0$
2. Jika $Y = 1$ maka $Y_0 = 0, Y_1 = 1, Y_2 = 0$
3. Jika $Y = 2$ maka $Y_0 = 0, Y_1 = 0, Y_2 = 1$

Diketahui bahwa $\sum_{j=0}^2 Y_j = 1$. Fungsi *likelihood* untuk suatu sampel dari n pengamatan bebas sebagai berikut:

$$l(\beta) = \prod_{i=1}^n [\pi_0(x_i)^{y_{0i}} \pi_1(x_i)^{y_{1i}} \pi_2(x_i)^{y_{2i}}] \quad (3.1)$$

Dengan memberi nilai $\ln \sum_{j=0}^2 y_{ij} = 1, \forall i = 1, 2, \dots, n$ maka

$$\begin{aligned}
L(\beta) &= \ln l(\beta) \\
&= \ln \prod_{i=1}^n [\pi_0(x_i)^{y_{0i}} \pi_1(x_i)^{y_{1i}} \pi_2(x_i)^{y_{2i}}] \\
&= \ln[\pi_0(x_i)^{y_{01}} \pi_1(x_i)^{y_{11}} \pi_2(x_i)^{y_{21}}] + \dots + \ln[\pi_0(x_i)^{y_{0n}} \pi_1(x_i)^{y_{1n}} \pi_2(x_i)^{y_{2n}}] \\
&= \ln[\pi_0(x_i)^{y_{01}}] + \ln[\pi_1(x_i)^{y_{11}}] + \ln[\pi_2(x_i)^{y_{21}}] + \dots + \ln[\pi_0(x_i)^{y_{0n}}] + \ln[\pi_1(x_i)^{y_{1n}}] \\
&\quad + \ln[\pi_2(x_i)^{y_{2n}}] \\
&= \sum_{i=1}^n y_{0i} \ln \left[\frac{1}{1 + e^{g_1(x_i)} + e^{g_2(x_i)}} \right] + y_{1i} \ln \left[\frac{e^{g_1(x_i)}}{1 + e^{g_1(x_i)} + e^{g_2(x_i)}} \right] \\
&\quad + y_{2i} \ln \left[\frac{e^{g_2(x_i)}}{1 + e^{g_1(x_i)} + e^{g_2(x_i)}} \right] \\
&= \sum_{i=1}^n y_{0i} \cdot 0 - y_{0i} \ln(1 + e^{g_1(x_i)} + e^{g_2(x_i)}) + y_{1i} g_1(x_i) \\
&\quad - y_{1i} \ln(1 + e^{g_1(x_i)} + e^{g_2(x_i)}) + y_{2i} g_1(x_i) - y_{2i} \ln(1 + e^{g_1(x_i)} + e^{g_2(x_i)}) \\
&= \sum_{i=1}^n y_{1i} g_1(x_i) + y_{2i} g_2(x_i) - (y_{0i} + y_{1i} + y_{2i}) \ln(1 + e^{g_1(x_i)} + e^{g_2(x_i)}) \\
&= \sum_{i=1}^n y_{1i} g_1(x_i) + y_{2i} g_2(x_i) - \ln(1 + e^{g_1(x_i)} + e^{g_2(x_i)})
\end{aligned}$$

Persamaan *likelihood* diperoleh dengan menurunkan parsial pertama dari $L(\beta)$ terhadap masing-masing parameter yang tidak diketahui dengan menjadikan persamaan turunan pertama sama dengan nol dari sebanyak $2(k+1)$ parameter yang ada. Penyederhanaan notasi dibuat $\pi_{ij} = \pi_j(x_i)$ sehingga bentuk umum persamaan dapat dituliskan sebagai berikut:

$$\begin{aligned}
\frac{\partial L(\beta)}{\partial \beta_{jk}} &= \frac{\partial}{\partial \beta_{jk}} \left[\sum_{i=1}^n y_{1i} g_1(x_i) + y_{2i} g_2(x_i) - \ln(1 + e^{g_1(x_i)} + e^{g_2(x_i)}) \right] \\
0 &= \sum_{i=1}^n x_{ki} (y_{ij} - \pi_{ij})
\end{aligned}$$

Persamaan di atas akan menghasilkan persamaan nonlinier dalam parameter β_{jk} sehingga penyelesaian untuk $\hat{\beta} = (\hat{\beta}_{j0}, \hat{\beta}_{j1}, \dots, \hat{\beta}_{jk})$ menggunakan metode numerik *Newton Raphson*. Penggunaan metode numerik *Newton Raphson*

membutuhkan turunan parsial kedua dari fungsi $L(\beta)$. Matriks dari turunan parsial kedua dibutuhkan untuk mendapatkan matriks keluaran dan matriks kovariansi dari dugaan maksimum *likelihood*. Bentuk umum dari elemen-elemen dalam matriks turunan parsial kedua adalah sebagai berikut:

$$\begin{aligned} \frac{\partial^2 L(\beta)}{\partial \beta_{kj} \partial \beta_{k'j'}} &= \frac{\partial}{\partial \beta_{k'j'}} \sum_{i=1}^n y_{ji} x_{ik} - \pi_{ji} x_{ik} \\ &= \frac{\partial}{\partial \beta_{k'j'}} \sum_{i=1}^n -x_{ik} \pi_{ji} \\ &= - \sum_{i=1}^n x_{ik} \pi_{ji} (1 - \pi_{ji}) x_{ik'}, \quad j' = j \\ &= \sum_{i=1}^n x_{ik} \pi_{ji} \pi_{j'i} x_{ik'}, \quad j' \neq j \end{aligned}$$

dimana

$j, j' = 1, 2$ dan $k, k' = 0, 1, 2, \dots, k$

Matriks informasi $I(\beta)$ adalah matriks $2(k+1)$ yang elemen-elemennya adalah negatif dari nilai-nilai harapan yang ditunjukkan pada persamaan turunan parsial kedua di atas. Matriks kovariansi dari dugaan maksimum *likelihood* adalah invers dari matriks informasi, $\widehat{Var}(\widehat{\beta}) = I(\widehat{\beta})^{-1}$. Dugaan-dugaan dari matriks informasi dan kovariansi didapat dengan mengganti semua parameter yang tidak diketahui dengan dugaan-dugaan maksimum *likelihood*.

Misal matriks \mathbf{X} berorde $n \times (k + 1)$ yang mengandung nilai-nilai dari variabel-variabel penjelas untuk tiap amatan. Misal matriks diagonal \mathbf{V}_j berorde $n \times n$ dengan elemen $\widehat{\pi}_{ji}(1 - \widehat{\pi}_{ji})$ dimana $j = 1, 2$ dan $i = 1, 2, \dots, n$, dan misal matriks informasi dapat ditulis sebagai berikut:

$$\begin{bmatrix} \widehat{\mathbf{I}}(\widehat{\beta}_{11}) & \widehat{\mathbf{I}}(\widehat{\beta}_{12}) \\ \widehat{\mathbf{I}}(\widehat{\beta}_{21}) & \widehat{\mathbf{I}}(\widehat{\beta}_{22}) \end{bmatrix}$$

dimana

$$\widehat{\mathbf{I}}(\widehat{\beta}_{11}) = (\mathbf{X}'\mathbf{V}_1\mathbf{X})$$

$$\widehat{\mathbf{I}}(\widehat{\beta}_{22}) = (\mathbf{X}'\mathbf{V}_2\mathbf{X})$$

$$\widehat{\mathbf{I}}(\widehat{\beta}_{12}) = \widehat{\mathbf{I}}(\widehat{\beta}_{21}) = -(\mathbf{X}'\mathbf{V}_3\mathbf{X})$$

3.3 Interpretasi Model

Model terbaik berdasarkan kriteria pengujian parameter yang telah diperoleh, selanjutnya dilakukan interpretasi koefisien berdasarkan parameter tersebut. *Odd Ratio* dapat dipergunakan untuk memudahkan interpretasi model. *Odd Ratio* adalah ukuran asosiasi yang memprediksi besar kemungkinan variabel-variabel penjelas terhadap variabel respon (Hosmer & Lemeshow, 2000). *Odd Ratio* untuk $Y=j$ terhadap $Y=k$ yang dihitung pada dua nilai (misal $x = a$ dan $x = b$) adalah sebagai berikut :

$$OR(a, b) = \frac{(Y = j|x = a)/(Y = k|x = a)}{(Y = j|x = b)/(Y = k|x = b)} = e^{\beta_i(a-b)}$$

sehingga jika $a - b = 1$ maka $OR=e^{\beta_i}$.

Untuk $OR=1$ berarti bahwa $x = a$ memiliki risiko yang sama dengan $x = b$ untuk menghasilkan $Y = j$. Jika variabel penjelas kontinu diperlukan unit perubahan sebesar c , maka *odd Ratio* diperoleh dari $e^{c\beta}$ (Hosmer & Lemeshow, 2000).

3.4 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Badan Pusat Statistik (BPS), tepatnya dari Survei Angkatan Kerja Nasional (SAKERNAS) untuk wilayah DKI Jakarta tahun 2015.

3.5 Deskripsi Data

Berdasarkan SAKERNAS 2015 untuk wilayah DKI Jakarta tercatat seluruhnya berjumlah 4.724.029 pekerja. Komposisi pekerja yang bekerja di sektor perdagangan sebanyak 1.659.551 orang. Komposisi pekerja yang bekerja di sektor jasa sebanyak 1.159.277 orang. Komposisi pekerja yang bekerja di sektor industri sebanyak 660.892 orang. Jadi populasi pada ketiga sektor tersebut berjumlah 3.479.720 pekerja, sedangkan sisanya sebanyak 1.244.309 pekerja yang bekerja pada sektor lainnya.

Pengambilan sampel dengan menggunakan *propotional random sampling*. Perhitungan pengambilan sampel dengan *propotional random sampling* dapat ditunjukkan sebagai berikut :

Sektor industri :

$$\frac{660.892}{3.479.720} \times 1.000 = 190 \text{ pekerja}$$

Sektor jasa :

$$\frac{1.159.277}{3.479.720} \times 1.000 = 334 \text{ pekerja}$$

Sektor perdagangan :

$$\frac{1.244.309}{3.479.720} \times 1.000 = 476 \text{ pekerja}$$

3.6 Definisi Operasional

Variabel yang digunakan dalam studi kasus ini terdiri atas 5 variabel pen- jelas, diantaranya jenis kelamin, tingkat pendidikan, golongan umur, status pekerjaan utama, dan jenis pekerjaan utama berdasarkan pada *Key Indicators of The Labour Market* yang dipublikasikan oleh *International Labour Organia- tion* (ILO). Definisi operasional variabel-variabel yang digunakan dalam studi kasus ini adalah sebagai berikut :

Variabel Respon

Variabel respon dalam studi kasus ini adalah lapangan pekerjaan utama yang terdapat di DKI Jakarta, terdiri atas tiga kategori, yaitu:

Kategori 0 = industri

Kategori 1 = jasa

Kategori 2 = perdagangan

Variabel Penjelas

Terdapat lima variabel penjelas yang digunakan dalam studi kasus ini, yaitu:

1. Jenis Kelamin (jk)

Menyatakan jenis kelamin pekerja

Kategori 0 = Laki-laki, Kategori 1 = Perempuan

2. Tingkat Pendidikan (didik)

Menyatakan tingkat pendidikan yang dicapai pekerja setelah mengikuti pelajaran pada kelas tertinggi suatu tingkatan sekolah dengan mendapatkan ijazah

Kategori 0 = SLTP ke bawah, Kategori 1 = SLTA,

Kategori 2 = Diploma, Kategori 3 = Universitas

3. Golongan Umur (umur)

Menyatakan umur pekerja

Kategori 0 = Remaja, Kategori 1 = Dewasa, Kategori 2 = Lansia

4. Status Pekerjaan Utama (status)

Menyatakan kedudukan pekerja dalam melakukan pekerjaan di suatu unit usaha/kegiatan

Kategori 0 = Berusaha sendiri, Kategori 1 = Berusaha dibantu buruh,

Kategori 2 = Buruh/karyawan/pegawai, Kategori 3 = Pekerja bebas di non pertanian, Kategori 4 = Pekerja bebas tak dibayar

5. Jenis Pekerjaan Utama (jpu)

Menyatakan macam pekerjaan yang dilakukan pekerja atau ditugaskan kepada pekerja yang sedang bekerja atau sementara tidak bekerja

Kategori 0 = Tenaga Profesional,

Kategori 1 = Tenaga Kepemimpinan dan ketatalaksanaan,

Kategori 2 = Tenaga tata usaha,

Kategori 3 = Tenaga usaha jasa,

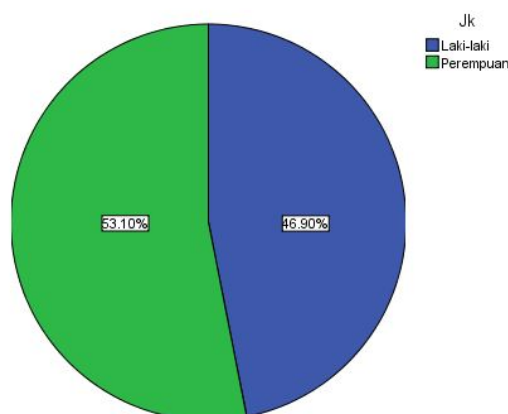
Kategori 4 = Tenaga produksi, operator alat angkutan, dan pekerja kasar

3.7 Analisis Deskriptif

Grafik *pie* dan tabulasi silang (*crosstabulation*) menunjukkan komposisi pekerja terhadap tiap-tiap variabel penjelas.

Jenis Kelamin

Pekerja dengan jenis kelamin laki-laki sebanyak 46,9% dan perempuan sebanyak 53,10%.



Gambar 3.1: Persentase Jenis Kelamin

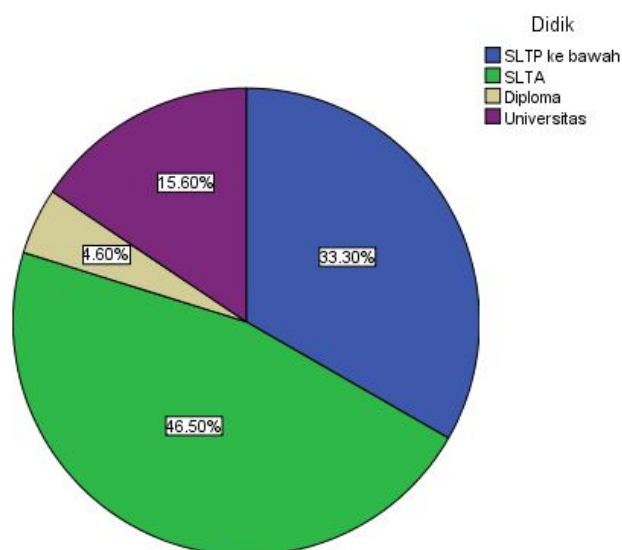
Tabel 3.1: Pilihan Berdasarkan Jenis Kelamin

Jenis Kelamin	Lapangan Pekerjaan			Total
	Industri	Jasa	Perdagangan	
Laki-laki	106	157	206	469
Perempuan	84	177	270	531
Total	190	334	476	1000

Berdasarkan tabel di atas dapat dilihat bahwa komposisi laki-laki yang memilih sektor industri sebanyak 106 orang, jasa sebanyak 157 orang, dan perdagangan sebanyak 206 orang. Komposisi perempuan yang memilih sektor industri sebanyak 84 orang, jasa sebanyak 177 orang, dan perdagangan sebanyak 270 orang.

Tingkat Pendidikan

Pekerja dengan tingkat pendidikan SLTP ke bawah sebanyak 33,30%, SLTA sebanyak 46,50%, diploma sebanyak 4,6%, dan universitas sebanyak 15,5%



Gambar 3.2: Persentase Tingkat Pendidikan

Tabel 3.2: Pilihan Berdasarkan Tingkat Pendidikan

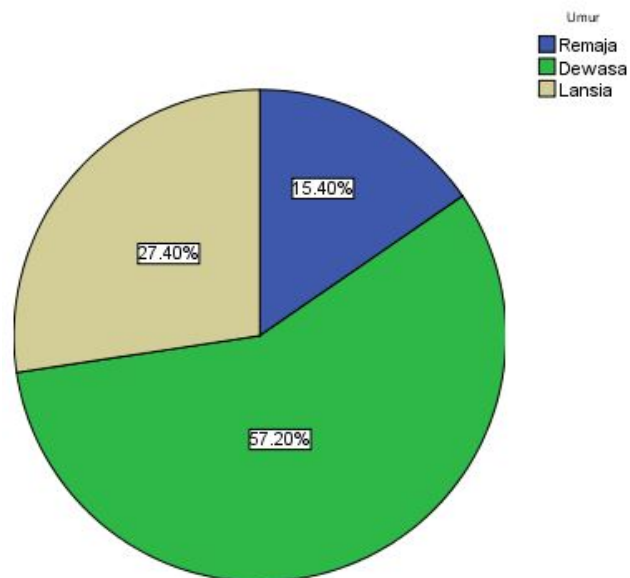
Tingkat Pendidikan	Lapangan Pekerjaan			Total
	Industri	Jasa	Perdagangan	
SLTP ke bawah	57	99	177	333
SLTA	87	147	231	465
Diploma	11	15	20	46
Universitas	35	73	48	156
Total	190	334	476	1000

Berdasarkan tabel di atas dapat dilihat bahwa komposisi pekerja dengan tingkat pendidikan SLTP ke bawah yang memilih sektor industri sebanyak 57 orang, jasa sebanyak 99 orang, dan perdagangan sebanyak 177 orang. Pekerja dengan tingkat pendidikan SLTA yang memilih sektor industri sebanyak 87 orang, jasa sebanyak 147 orang, dan perdagangan sebanyak 231 orang. Pekerja

dengan tingkat pendidikan diploma yang memilih sektor industri sebanyak 11 orang, jasa sebanyak 15 orang, dan perdagangan sebanyak 20 orang. Pekerja dengan tingkat pendidikan universitas yang memilih sektor industri sebanyak 35 orang, jasa sebanyak 73 orang, dan perdagangan sebanyak 48 orang.

Golongan Umur

Pekerja dengan golongan umur remaja sebanyak 15,40% dewasa sebanyak 57,2%, dan lansia sebanyak 27,4%.



Gambar 3.3: Persentase Golongan Umur

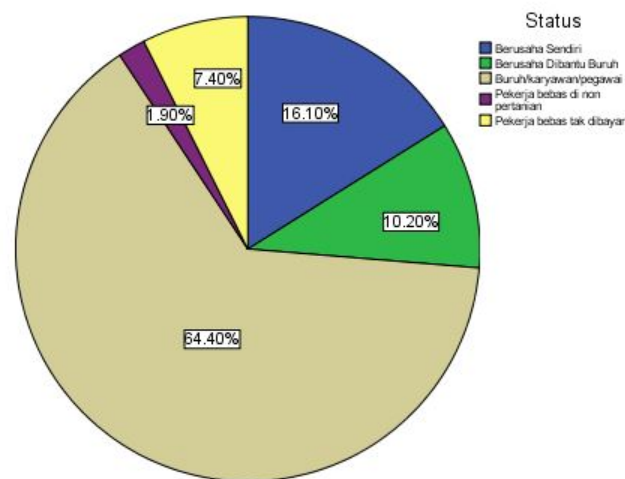
Tabel 3.3: Pilihan Berdasarkan Golongan Umur

Golongan Umur	Lapangan Pekerjaan			Total
	Industri	Jasa	Perdagangan	
Remaja	28	47	79	154
Dewasa	120	188	264	572
Lansia	42	99	133	274
Total	190	334	476	1000

Berdasarkan tabel di atas dapat dilihat bahwa pekerja dengan golongan umur remaja yang memilih sektor industri sebanyak 28 orang, jasa sebanyak 47 orang, dan perdagangan sebanyak 79 orang. Pekerja dengan golongan umur dewasa yang memilih sektor industri sebanyak 120 orang, jasa sebanyak 188 orang, dan perdagangan sebanyak 264 orang. Pekerja dengan golongan umur lansia yang memilih sektor industri sebanyak 42 orang, jasa sebanyak 99 orang, dan perdagangan sebanyak 133 orang.

Status Pekerjaan

Pekerja dengan status pekerjaan berusaha sendiri sebanyak 16,10%, berusaha dibantu buruh sebanyak 10,20%, buruh/karyawan/pegawai sebanyak 64,40%, pekerja bebas di non pertanian sebanyak 1,90%, dan pekerja tak dibayar sebanyak 7,4%.



Gambar 3.4: Persentase Status Pekerjaan

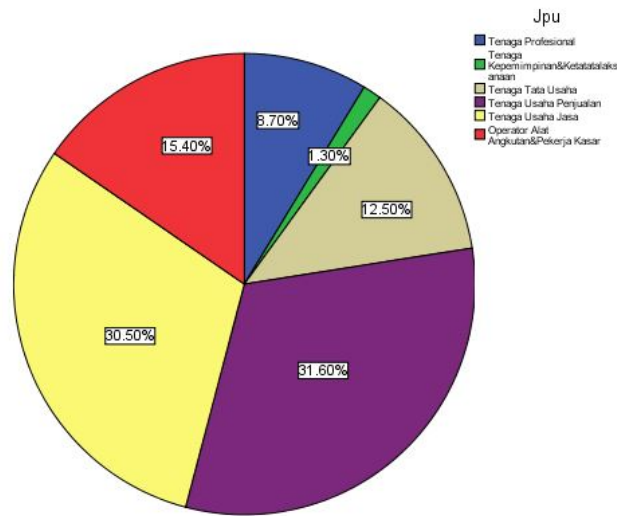
Tabel 3.4: Pilihan Berdasarkan Status Pekerjaan

Status Pekerjaan	Lapangan Pekerjaan			Total
	Industri	Jasa	Perdagangan	
Berusaha sendiri	9	33	119	161
Berusaha dibantu buruh	13	14	75	102
Buruh/karyawan/pegawai	160	270	214	644
Pekerja bebas di nonpertanian	1	15	3	19
Pekerja tak dibayar	7	2	65	74
Total	190	334	476	1000

Berdasarkan tabel di atas dapat dilihat bahwa pekerja dengan status pekerjaan berusaha sendiri yang memilih sektor industri sebanyak 9 orang, jasa sebanyak 33 orang, dan perdagangan sebanyak 119 orang. Pekerja dengan status pekerjaan berusaha dibantu buruh yang memilih sektor industri sebanyak 13 orang, jasa sebanyak 14 orang, dan perdagangan sebanyak 75 orang. Pekerja dengan status pekerjaan buruh/karyawan/pegawai yang memilih sektor industri sebanyak 160 orang, jasa sebanyak 270 orang, dan perdagangan sebanyak 214 orang. Pekerja dengan status pekerjaan pekerja bebas di nonpertanian yang memilih sektor industri sebanyak 1 orang, jasa sebanyak 15 orang, dan perdagangan sebanyak 3 orang. Pekerja dengan status pekerjaan pekerja tak dibayar yang memilih sektor industri sebanyak 7 orang, jasa sebanyak 2 orang, dan perdagangan sebanyak 65 orang.

Jenis Pekerjaan

Pekerja dengan jenis pekerjaan tenaga profesional sebanyak 8,7%, tenaga kepemimpinan & ketatalaksanaan sebanyak 1,30%, tenaga tata usaha sebanyak 12,5%, tenaga usaha penjualan sebanyak 31,60%, tenaga usaha jasa sebanyak 30,50%, dan operator alat angkutan & pekerja kasar sebanyak 15,40%.



Gambar 3.5: Persentase Jenis Pekerjaan

Tabel 3.5: Pilihan Berdasarkan Jenis Pekerjaan

Jenis Pekerjaan	Lapangan Pekerjaan			Total
	Industri	Jasa	Perdagangan	
Tenaga profesional	8	74	5	87
Tenaga kepemimpinan & ketatalaksanaan	6	5	2	13
Tenaga tata usaha	35	45	45	125
Tenaga usaha penjualan	8	3	305	316
Tenaga usaha jasa	32	177	96	305
Operator alat angkutan & pekerja kasar	101	30	23	154
Total	190	334	476	1000

Berdasarkan tabel di atas dapat dilihat bahwa pekerja dengan jenis pekerjaan tenaga profesional yang memilih sektor industri sebanyak 8 orang, jasa sebanyak 74 orang, dan perdagangan sebanyak 5 orang. Pekerja dengan jenis pekerjaan tenaga kepemimpinan & ketatalaksanaan yang memilih sektor industri sebanyak 6 orang, jasa sebanyak 5 orang, dan perdagangan sebanyak 2 orang. Pekerja dengan jenis pekerjaan tenaga tata usaha yang memilih sektor industri sebanyak 35 orang, jasa sebanyak 45 orang, dan perdagangan sebanyak 45 orang. Pekerja dengan jenis pekerjaan tenaga usaha penjualan yang

memilih sektor industri sebanyak 8 orang, jasa sebanyak 3 orang, dan perdagangan sebanyak 305 orang. Pekerja dengan jenis pekerjaan tenaga usaha jasa yang memilih sektor industri sebanyak 32 orang, jasa sebanyak 177 orang, dan perdagangan sebanyak 96 orang. Pekerja dengan jenis pekerjaan operator alat angkutan & pekerja kasar yang memilih sektor industri sebanyak 101 orang, jasa sebanyak 30 orang, dan perdagangan sebanyak 23 orang.

3.8 Uji Multikolinieritas

Jika didapatkan nilai $VIF < 10$ maka dikatakan bahwa terjadi multikolinieritas. Nilai VIF masing-masing variabel dapat ditunjukkan sebagai berikut:

Tabel 3.6: Nilai VIF

Variabel Penjelas	VIF
JK	1,081
DIDIK	2,971
UMUR	3,028
STATUS	2,210
JPU	1,703

Ternyata setelah dilakukan perhitungan nilai VIF masing-masing variabel penjelas tidak ditemukan adanya multikolinieritas.

3.9 Analisis Regresi Logistik Multinomial

Berdasarkan uji multikolinieritas di atas maka dapat diketahui bahwa tidak terjadi multikolinieritas antar variabel penjelas sehingga tahapan selanjutnya adalah analisis regresi logistik multinomial.

3.9.1 Uji Signifikansi

Terdapat dua uji signifikansi pada analisis regresi logistik multinomial, yaitu uji signifikansi model secara serentak dan parsial.

Uji Serentak

Uji signifikansi model secara serentak dilakukan dengan menggunakan uji rasio *likelihood*. Adapun hipotesis yang digunakan sebagai berikut:

H_0 : Tidak ada variabel penjelas yang berpengaruh terhadap variabel respon

H_1 : Minimal ada satu variabel penjelas yang berpengaruh terhadap variabel respon

Hasil uji rasio *likelihood* dapat ditunjukkan sebagai berikut:

Tabel 3.7: Model Fitting Information

Model	-2logLikelihood	Chi-Square	df	Sig
Intercept Only	$1,570 \times 10^3$			
Final	171,354	$1,399 \times 10^3$	30	0,000

Hasil pengujian diperoleh nilai $\text{Sig} = 0,000 < \alpha$ sehingga didapatkan keputusan tolak H_0 dan diketahui secara serentak model signifikan pada tingkat signifikansi 5%. Artinya variabel jenis kelamin, tingkat pendidikan, golongan umur, status pekerjaan, dan jenis pekerjaan secara bersama-sama memberikan pengaruh nyata terhadap peluang seseorang dalam memilih lapangan pekerjaan utama.

Berikutnya untuk mengetahui variabel penjelas yang signifikan secara parsial dan layak masuk ke dalam model digunakan uji signifikansi parsial.

Uji Parsial

Variabel respon terdiri atas tiga sektor lapangan pekerjaan sehingga ada dua fungsi logit yang terbentuk. Kategori yang dijadikan pembanding adalah kategori sektor industri. Informasi mengenai nilai koefisien untuk masing-masing fungsi logit dapat diperoleh pada tabel di bawah sebagai berikut:

Tabel 3.8: Pendugaan Parameter

Jasa	β	Sig	Perdagangan	β	Sig
JK[0]	-1,148	0,000	JK[0]	0,958	0,147
DIDIK[0]	-1,843	0,065	DIDIK[0]	-2,984	0,067
DIDIK[1]	0,725	0,371	DIDIK[1]	2,509	0,005
DIDIK[2]	-0,585	0,430	DIDIK[2]	0,098	0,910
UMUR[0]	-5,897	0,000	UMUR[0]	5,398	0,997
UMUR[1]	-1,958	0,008	UMUR[1]	-5,947	0,000
STATUS[0]	0,593	0,680	STATUS[0]	-3,081	0,998
STATUS[1]	-1,120	0,407	STATUS[1]	-34,625	0,978
STATUS[2]	0,542	0,542	STATUS[2]	-55,084	0,000
STATUS[3]	4,587	0,001	STATUS[3]	-18,610	0,997
JPU[0]	11,224	0,000	JPU[0]	-0,499	0,562
JPU[1]	7,967	0,000	JPU[1]	14,710	0,991
JPU[2]	4,685	0,000	JPU[2]	23,861	0,997
JPU[3]	1,698	0,046	JPU[3]	61,346	0,000
JPU[4]	3,991	0,000	JPU[4]	54,780	0,978

Adapun hipotesis yang digunakan sebagai berikut:

H_0 : Tidak ada pengaruh antara variabel penjelas dengan variabel respon

H_1 : Ada pengaruh antara variabel penjelas dengan variabel respon

Berdasarkan hasil pengujian untuk sektor jasa dapat dituliskan fungsi logitnya sebagai berikut :

$$g_1 = -1,234 - 1,148Jk_0 - 5,897Umur_0 - 1,958Umur_1 + 4,587Status_3 \\ + 11,2224Jpu_0 + 7,967Jpu_1 + 4,685Jpu_2 + 1,698Jpu_3 + 3,991Jpu_4$$

Variabel yang berpengaruh secara parsial adalah variabel jenis kelamin, go-

longan umur, status pekerjaan, dan jenis pekerjaan karena mempunyai nilai $\text{Sig} < \alpha = 0,05$.

Bentuk fungsi logit sektor perdagangan sebagai berikut :

$$g_2 = 0,497 + 2,509\text{Didik}_1 - 5,947\text{Umur}_1 - 55,084\text{Status}_2 + 61,346\text{Jpu}_3$$

Variabel yang berpengaruh secara parsial adalah variabel tingkat pendidikan, golongan umur, status pekerjaan, dan jenis pekerjaan karena mempunyai nilai $\text{Sig} < \alpha = 0,05$.

3.9.2 Uji Kesesuaian Model

Uji kesesuaian model dilakukan untuk mengetahui adanya perbedaan antara hasil observasi dengan hasil prediksi setelah model serentak terbentuk.

Adapun hipotesis yang digunakan sebagai berikut:

H_0 : Tidak ada perbedaan antara hasil observasi dengan hasil prediksi model

H_1 : Ada perbedaan antara hasil observasi dengan hasil prediksi model

Tabel 3.9: Uji Kesesuaian Model

	Chi-Square	df	Sig
Devians	104,013	98	0,320

Berdasarkan tabel di atas diketahui bahwa nilai $\text{Sig} > \alpha = 0,05$ sehingga terima H_0 menunjukkan bahwa tidak ada perbedaan antara hasil observasi dengan hasil prediksi model.

3.10 Pemilihan Model Terbaik

Pemilihan model terbaik dilakukan dengan membandingkan nilai *Akaike Information Criterion*(AIC). Model yang terbaik adalah model yang memiliki nilai AIC terkecil. Nilai AIC setiap model dapat ditunjukkan sebagai berikut:

Tabel 3.10: Nilai AIC

Model	AIC
Jasa	191,354
Perdagangan	181,354

Berdasarkan perhitungan nilai AIC diketahui bahwa model perdagangan mempunyai nilai terkecil sehingga model perdagangan merupakan model terbaik.

3.11 Interpretasi Model

Nilai rasio *odds* untuk seluruh variabel penjelas pada masing-masing sektor lapangan pekerjaan dapat ditunjukkan sebagai berikut:

Tabel 3.11: Rasio *Odd*

Jasa	Exp(β)	Perdagangan	Exp(β)
Jk[0]	0,317	JK[0]	2,606
Didik[0]	0,158	Didik[0]	0,051
Didik[1]	2,064	Didik[1]	12,298
Didik[2]	0,557	Didik[2]	1,103
Umur[0]	0,003	Umur[0]	220,998
Umur[1]	0,141	Umur[1]	0,003
Status[0]	1,809	Status[0]	0,046
Status[1]	0,326	Status[1]	$9,17 \times 10^{-16}$
Status[2]	1,719	Status[2]	$1,19 \times 10^{-24}$
Status[3]	98,162	Status[3]	$8,273 \times 10^9$
Jpu[0]	$7,491 \times 10^4$	Jpu[0]	0,607
Jpu[1]	$2,88 \times 10^3$	Jpu[1]	$2,446 \times 10^6$
Jpu[2]	108,351	Jpu[2]	$2,305 \times 10^{10}$
Jpu[3]	5,464	Jpu[3]	$4,389 \times 10^{26}$
Jpu[4]	54,132	Jpu[4]	$6,172 \times 10^{23}$

Berdasarkan perhitungan nilai rasio *odds* maka interpretasi model untuk fungsi logit jasa sebagai berikut:

Jenis kelamin (0) mempunyai koefisien yang bernilai negatif, yaitu 1,148 de-

ngan nilai rasio *odds* sebesar 0,317. Apabila terjadi kenaikan total skor sebesar satu poin maka peluang untuk memilih lapangan pekerjaan jasa berkurang sebesar 0,317 kali jika dibandingkan dengan peluang untuk memilih lapangan pekerjaan industri. Golongan umur (0) mempunyai koefisien yang bernilai negatif, yaitu -5,897 dengan nilai rasio *odds* sebesar 0,003. Apabila terjadi kenaikan total skor sebesar satu poin maka peluang untuk memilih lapangan pekerjaan jasa berkurang sebesar 0,003 kali jika dibandingkan dengan peluang untuk memilih lapangan pekerjaan industri. Golongan umur (1) mempunyai koefisien yang bernilai negatif, yaitu -5,947 dengan nilai rasio *odds* sebesar 0,141. Apabila terjadi kenaikan total skor sebesar satu poin maka peluang untuk memilih lapangan pekerjaan jasa berkurang sebesar 0,141 kali jika dibandingkan dengan peluang untuk memilih lapangan pekerjaan industri. Status pekerjaan (3) mempunyai koefisien yang bernilai positif, yaitu 4,587 dengan nilai rasio *odds* sebesar 98,162. Apabila terjadi kenaikan total skor sebesar satu poin maka peluang untuk memilih lapangan pekerjaan jasa meningkat sebesar 98,162 kali jika dibandingkan dengan peluang untuk memilih lapangan pekerjaan industri. Jenis pekerjaan (0) mempunyai koefisien yang bernilai positif, yaitu 11,2224 dengan nilai rasio *odds* sebesar $7,491 \times 10^4$. Apabila terjadi kenaikan total skor sebesar satu poin maka peluang untuk memilih lapangan pekerjaan jasa meningkat sebesar $7,491 \times 10^4$ kali jika dibandingkan dengan peluang untuk memilih lapangan pekerjaan industri. Jenis pekerjaan (1) mempunyai koefisien yang bernilai positif, yaitu 3,579 dengan nilai rasio *odds* sebesar $2,88 \times 10^3$. Apabila terjadi kenaikan total skor sebesar satu poin maka peluang untuk memilih lapangan pekerjaan jasa meningkat sebesar $2,88 \times 10^3$ kali jika dibandingkan dengan peluang untuk memilih lapangan pekerjaan industri. Jenis pekerjaan (2) mempunyai koefisien yang bernilai positif, yaitu 4,685 dengan nilai rasio *odds* sebesar 108,351. Apabila terjadi kenaikan

total skor sebesar satu poin maka peluang untuk memilih lapangan pekerjaan jasa meningkat sebesar 108,351 kali jika dibandingkan dengan peluang untuk memilih lapangan pekerjaan industri. Jenis pekerjaan (3) mempunyai koefisien yang bernilai positif, yaitu 1,698 dengan nilai rasio *odds* sebesar 5,464. Apabila terjadi kenaikan total skor sebesar satu poin maka peluang untuk memilih lapangan pekerjaan jasa meningkat sebesar 5,464 kali jika dibandingkan dengan peluang untuk memilih lapangan pekerjaan industri. Jenis pekerjaan (4) mempunyai koefisien yang bernilai positif, yaitu 3,991 dengan nilai rasio *odds* sebesar 5,464. Apabila terjadi kenaikan total skor sebesar satu poin maka peluang untuk memilih lapangan pekerjaan jasa meningkat sebesar 54,132 kali jika dibandingkan dengan peluang untuk memilih lapangan pekerjaan industri.

Berdasarkan perhitungan nilai rasio *odds* maka interpretasi model untuk fungsi logit perdagangan sebagai berikut:

Tingkat pendidikan (1) mempunyai koefisien yang bernilai positif, yaitu 2,509 dengan nilai rasio *odds* sebesar 12,298. Apabila terjadi kenaikan total skor sebesar satu poin maka peluang untuk memilih lapangan pekerjaan perdagangan meningkat sebesar 12,298 kali jika dibandingkan dengan peluang untuk memilih lapangan pekerjaan industri. Golongan umur (1) mempunyai koefisien yang bernilai negatif, yaitu 5,947 dengan nilai rasio *odds* sebesar 0,003. Apabila terjadi kenaikan total skor sebesar satu poin maka peluang untuk memilih lapangan pekerjaan perdagangan berkurang sebesar 0,003 kali jika dibandingkan dengan peluang untuk memilih lapangan pekerjaan industri. Status pekerjaan (1) mempunyai koefisien yang bernilai negatif, yaitu 55,084 dengan nilai rasio *odds* sebesar $1,19 \times 10^{-24}$. Apabila terjadi kenaikan total skor sebesar satu poin maka peluang untuk memilih lapangan pekerjaan perdagangan berkurang sebesar $1,19 \times 10^{-24}$ kali jika dibandingkan dengan peluang untuk memilih lapangan pekerjaan industri. Jenis pekerjaan (3) mempunyai koefisien yang

bernilai positif, yaitu 61,346 dengan nilai rasio *odds* sebesar $4,389 \times 10^{26}$. Apabila terjadi kenaikan total skor sebesar satu poin maka peluang untuk memilih lapangan pekerjaan perdagangan meningkat sebesar 413775,160 kali jika dibandingkan dengan peluang untuk memilih lapangan pekerjaan industri.

3.12 Ketepatan Klasifikasi

Ketepatan klasifikasi yang diperoleh dari model yang telah dibentuk dapat ditunjukkan sebagai berikut :

Tabel 3.12: Ketepatan Klasifikasi

Observed	Predicted Membership			Total
	INDUSTRI	JASA	PERDAGANGAN	
INDUSTRI	122	53	15	64,2%
JASA	12	302	20	90,4%
PERDAGANGAN	1	24	451	94,7%
Total	13,5%	37,9%	48,6%	87,5%

Ketepatan klasifikasi dari model yang telah dibentuk, yaitu sebesar 87,5%. Hal ini menunjukkan bahwa banyaknya prediksi yang tepat diklasifikasikan sesuai dengan kondisi sesungguhnya (observasi) adalah sebesar 87,5%. Banyaknya observasi lapangan pekerjaan sektor industri yang tepat diprediksi pada lapangan pekerjaan sektor industri sebesar 64,2%, sedangkan banyaknya prediksi lapangan pekerjaan sektor industri yang tepat diklasifikasikan sesuai dengan kondisi sesungguhnya (observasi) adalah sebesar 13,5%. Banyaknya observasi lapangan pekerjaan sektor jasa yang tepat diprediksi pada lapangan pekerjaan sektor jasa sebesar 90,4%, sedangkan banyaknya prediksi lapangan pekerjaan sektor industri yang tepat diklasifikasikan sesuai dengan kondisi sesungguhnya (observasi) adalah sebesar 37,9%. Diketahui bahwa banyaknya observasi lapangan pekerjaan sektor perdagangan yang tepat diprediksi pada lapangan pekerjaan sektor perdagangan sebesar 94,7%, sedangkan banyaknya prediksi lapangan pekerjaan sektor perdagangan yang tepat diklasifikasikan sesuai dengan kondisi sesungguhnya (observasi) adalah sebesar 48,6%.

BAB IV

PENUTUP

4.1 Kesimpulan

1. Persamaan model regresi logistik multinomial untuk tiga kategori respon dengan kategori $Y = 0$ sebagai kategori pembanding adalah sebagai berikut:

$$\pi_j(x) = \frac{e^{g_j(x)}}{\sum_{j=0}^2 e^{g_j(x)}} \text{ dan } \pi_0(x) = \frac{e^{g_0(x)}}{\sum_{j=0}^2 e^{g_j(x)}}$$

2. Berdasarkan analisis data yang dihasilkan, dapat diketahui bahwa variabel yang berpengaruh terhadap pemilihan lapangan pekerjaan sektor jasa adalah jenis kelamin laki-laki, golongan umur kategori remaja dan dewasa, status pekerjaan kategori pekerja bebas di nonpertanian, jenis pekerjaan kategori tenaga profesional, tenaga kepemimpinan&tatalaksanaan, tata usaha, dan jenis pekerjaan kategori tenaga, tenaga usaha penjualan, dan tenaga usaha jasa. Variabel yang berpengaruh terhadap pemilihan lapangan pekerjaan sektor perdagangan adalah tingkat pendidikan SLTA, golongan umur dewasa, status pekerjaan buruh/karyawan/pegawai, dan tenaga usaha penjualan

Fungsi logit untuk sektor lapangan pekerjaan jasa adalah sebagai berikut:

$$g_1 = -1,234 - 1,148Jk_0 - 5,897Umur_0 - 1,958Umur_1 + 4,587Status_3 + 11,2224Jpu_0 + 7,967Jpu_1 + 4,685Jpu_2 + 1,698Jpu_3 + 3,991Jpu_4$$

Fungsi logit untuk sektor lapangan pekerjaan perdagangan adalah sebagai berikut:

$$g_2 = 0,497 + 2,509Didik_1 - 5,947Umur_1 - 55,084Status_2 + 61,346Jpu_3$$

Model regresi logistik multinomial yang terbentuk menghasilkan ketepatan klasifikasi untuk pemilihan lapangan pekerjaan sebesar 87,5%.

4.2 Saran

Model yang dihasilkan dapat ditingkatkan ketepatan klasifikasinya dengan metode lain, seperti metode *bagging* regresi logistik multinomial atau metode nonparametrik seperti *Multivariate Adaptive Regression Spline* (MARS).

DAFTAR PUSTAKA

- Anton, H dan Chris Rorres. 2004. Aljabar Linear Elementer Versi Aplikasi. Edisi ke-8. Jakarta: Erlangga.
- Hosmer, D.W dan Lemeshow. 2002. *Applied Logistic Regression*. Edisi ke-2. New York: John Wiley,Inc.
- International Labour Organization.2016. *Key Indicators of The labour Market*. Edisi ke-9. Geneva: International Labour Organization.
- Johnson, R.A dan Dean Wichern. 2000. *Applied Multivariate Statistic Analysis*. Edisi ke-6. New Jersey: Prentice Hall.
- Montgomery, Douglas., E. A Peck., dan G Geoffrey. 2012. *Introduction To Linier Regression Analysis*. Edisi ke-5. New York: John Wiley,Inc.
- Supranto, J. 2008. Statistik Teori dan Aplikasi. Edisi ke-7. Jakarta: Erlangga.
- Statistik Indonesia Merdeka. 2015. Jakarta: BPS.
- Survei Angkatan Kerja. 2015. Jakarta: BPS.<https://www.bps.go.id/index.php>.
- Tabanick, B.G., L. S Fidell., dan Osterlind, S.J. 2001. *Using Multivariate Statistics*. Edisi ke-5. US: Allyn & Bacon Boston.
- Yanagihara, Hirozoku.,et all. Bias-corrected AIC for selecting variables in multinomial logistic regression models, *The Mathematical Journals*, 2012. <http://www.sciencedirect.com>.

LAMPIRAN-LAMPIRAN

Lampiran 1: *Output* perhitungan dengan *Software SPSS*

Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	1.570E3			
Final	171.354	1.399E3	30	.000

Gambar 4.1: Model Fitting Information

Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	2870.638	98	.000
Deviance	104.013	98	.320

Gambar 4.2: Goodness of Fit

Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	1.714E2 ^a	.000	0	.
Jk	197.997	26.643	2	.000
Status	443.613	272.259	8	.000
Jpu	1.107E3	935.678	10	.000
Didik	234.924	63.570	6	.000
Umur	264.677	93.322	4	.000

The chi-square statistic is the difference in -2 log-likelihoods between the final model and a reduced model. The reduced model is formed by omitting an effect from the final model. The null hypothesis is that all parameters of that effect are 0.

a. This reduced model is equivalent to the final model because omitting the effect does not increase the degrees of freedom.

Gambar 4.3: Likelihood Ratio Test

Classification

Observed	Predicted			Percent Correct
	Industri	Jasa	Perdagangan	
Industri	122	53	15	64.2%
Jasa	12	302	20	90.4%
Perdagangan	1	24	451	94.7%
Overall Percentage	13.5%	37.9%	48.6%	87.5%

Gambar 4.4: Klasifikasi

SURAT PERNYATAAN KEASLIAN SKRIPSI

Dengan ini saya yang bertanda tangan di bawah ini, mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta:

Nama : Jennyfer Romaito
No. Registrasi : 3125121969
Program Studi : Matematika

Menyatakan bahwa skripsi ini yang saya buat dengan judul "**Penggunaan Analisis Regresi Logistik Multinomial Untuk Klasifikasi Pemilihan Lapangan Pekerjaan di DKI Jakarta (Studi Kasus: Angkatan Kerja DKI Jakarta tahun 2015)**" adalah :

1. Dibuat dan diselesaikan oleh saya sendiri.
2. Bukan merupakan duplikat skripsi yang pernah dibuat oleh orang lain atau jiplakan karya tulis orang lain.

Pernyataan ini dibuat dengan sesungguhnya dan saya bersedia menanggung segala akibat yang timbul jika pernyataan saya tidak benar.

Jakarta, Februari 2017

Yang membuat pernyataan

Jennyfer Romaito

DAFTAR RIWAYAT HIDUP



JENNYFER ROMAITO. Lahir di Jakarta, 14 Januari 1995. Anak ketiga dari pasangan Bapak W. Manullang dan Ibu Berliana Simamora. Saat ini bertempat tinggal di Jalan Kaca Jendela No 6 RT 008/ RW 08, Jakarta Selatan .

No. Ponsel : 085781591761

Email : jennyferromaito@gmail.com

Riwayat Pendidikan : Penulis mengawali pendidikan di TK Widyastuti selama 1 tahun, dan kemudian melanjutkan pendidikan di SDN 03 Pagi Rawajati pada tahun 2000 - 2006. Setelah itu, penulis melanjutkan ke SMPN 115 Tebet Jakarta Selatan hingga tahun 2009. Kemudian kembali melanjutkan ke SMAN 14 Jakarta Timur dan lulus tahun 2012. Di Tahun yang sama penulis melanjutkan ke Universitas Negeri Jakarta (UNJ), program studi Matematika, melalui jalur SBMPTN tulis. Di awal tahun 2017 penulis telah memperoleh gelar Sarjana Sains untuk Program Studi Matematika, FMIPA, UNJ.

Riwayat Organisasi : Selama di bangku perkuliahan, penulis pernah menjadi *volunteer* pada *team* pembuat soal Calculus Cup 2014.

Riwayat Pekerjaan : Penulis mulai menjadi pengajar privat matematika sejak tahun 2013. Pada tahun 2015, penulis pernah praktek kerja lapangan pada bidang penelitian dan pengembangan di Badan Perencanaan dan Pembangunan Daerah DKI Jakarta.