



TUGAS AKHIR - SS141501

**PEMODELAN INDEKS PEMBANGUNAN MANUSIA (IPM)
PROVINSI JAWA TIMUR DENGAN MENGGUNAKAN
METODE REGRESI LOGISTIK RIDGE**

Dwi Maumere Putra
NRP 1311 100 108

Dosen Pembimbing
Dr. Vita Ratnasari, S.Si., M.Si.

Program Studi S1 Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember
Surabaya 2015



FINAL PROJECT - SS141501

MODELLING HUMAN DEVELOPMENT INDEX (HDI) OF EAST JAVA PROVINCE USING LOGISTIC RIDGE REGRESSION METHOD

Dwi Maumere Putra
NRP 1311 100 108

Supervisor
Dr. Vita Ratnasari, S.Si., M.Si.

Undergraduate Programme of Statistics
Faculty of Mathematics and Natural Science
Institut Teknologi Sepuluh Nopember
Surabaya 2015

LEMBAR PENGESAHAN

**PEMODELAN INDEKS PEMBANGUNAN MANUSIA (IPM)
PROVINSI JAWA TIMUR DENGAN MENGGUNAKAN
METODE REGRESI LOGISTIK RIDGE**

TUGAS AKHIR

Diajukan Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Sains
pada

Program Studi S-1 Jurusan Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember

Oleh :

DWI MAUMERE PUTRA
NRP. 1311 100 108

Disetujui oleh Pembimbing Tugas Akhir :

Dr. Vita Ratnasari, S.Si., M.Si.
NIP. 19700910 199702 2 001

(Ratnasari)

Mengetahui,
Ketua Jurusan Statistika FMIPA-ITS


Dr. Muhammad Mashuri, MT.
NIP. 19620408 198701 1 001

SURABAYA, JULI 2015



Pemodelan Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur Dengan Menggunakan Metode Regresi Logistik Ridge

Nama : Dwi Maumere Putra
NRP : 1311100108
Jurusan : Statistika FMIPA – ITS
Pembimbing : Dr. Vita Ratnasari, S.Si., M.Si

ABSTRAK

Pembangunan manusia di suatu daerah merupakan upaya yang dilakukan oleh pemerintah daerah terkait dengan kesejahteraan masyarakat yang diukur dengan Indeks Pembangunan Manusia (IPM). Terdapat tiga sektor pembentuk IPM yaitu kesehatan, pendidikan dan ekonomi dimana faktor-faktor dalam setiap sektor cenderung memiliki kolinieritas yang tinggi yang menyebabkan adanya kasus multikolinieritas. Apabila kasus multikolinieritas tidak diatasi, maka dapat menyebabkan variansi dari hasil estimasi parameter menjadi besar yang dapat berakibat pada banyaknya variabel prediktor yang tidak signifikan meskipun nilai koefisien determinasi (R^2) tinggi. Sehingga untuk mengatasinya dilakukan pemodelan terhadap Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur menggunakan metode Regresi Logistik Ridge. Terdapat tiga variabel yang berpengaruh signifikan, yaitu angka kematian bayi (X_1), angka buta huruf (X_4) dan angka partisipasi sekolah (X_5). Dengan metode backward elimination, didapatkan model terbaik dengan ketepatan klasifikasi sebesar 97,37% yang menghasilkan 5 kabupaten/kota tergolong IPM menengah bawah, yaitu Kabupaten Bangkalan, Sampang, Probolinggo, Situbondo dan Jember. Pada 33 kabupaten/kota yang lain tergolong dalam IPM menengah atas.

Kata Kunci : Indeks Pembangunan Manusia (IPM), Jawa Timur, Klasifikasi, Regresi Logistik Ridge.

(Halaman ini sengaja dikosongkan)

Modelling Human Development Index (HDI) Of East Java Province Using Logistic Ridge Regression Method

Name : Dwi Maumere Putra
NRP : 1311100108
Department : Statistics FMIPA – ITS
Supervisor : Dr. Vita Ratnasari, S.Si., M.Si

ABSTRACT

The human development in an area is one of the local governments duty related to all society welfare as measured by the Human Development Index (HDI). There are three sectors forming the HDI namely health, education and economy in which the factors in each sector tend to have a high collinearity caused multicollinearity case. If the multicollinearity is not be solved, it can causes to the variance of the parameter estimation which can result in a number of predictor variables were not significant although has the high coefficient determination (R^2) value. So as to overcome modeling of the late Human Development Index (HDI) of East Java Province using Logistic Ridge Regression method. There are three variables that have a significant effect, the infant mortality rate (X_1), illiteracy (X_4) and enrollment (X_5). By using backward elimination method, the best model with is obtained with a classification accuracy of 97.37%, which resulted that five districts/cities belong to the lower middle HDI, there are Bangkalan, Sampang, Probolinggo, Situbondo and Jember. As for the 33 districts/cities which others belong to the upper middle HDI.

Keyword : *Human Development Index (HDI), East Java, Classification, Logistic Ridge Regression.*

(Halaman ini sengaja dikosongkan)

KATA PENGANTAR

Alhamdu lillahi rabbil ‘alamin, puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat, taufik dan hidayah-Nya sehingga penulis dapat menyelesaikan laporan Tugas Akhir yang berjudul **“Pemodelan Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur Dengan Menggunakan Metode Regresi Logistik Ridge“**.

Laporan Tugas Akhir ini tidak akan terselesaikan dengan baik tanpa bantuan dan dukungan dari Ibu dan Kakak serta berbagai pihak lain yang terkait. Untuk itu penulis mengucapkan banyak terima kasih kepada :

1. Ibu Dr. Vita Ratnasari, S.Si, M.Si, selaku dosen pembimbing yang memberikan waktu, ilmu dan saran untuk membimbing penulis.
2. Bapak Dr. Muhammad Mashuri, MT selaku Ketua Jurusan Statistika ITS, Ibu Dra. Lucia Aridinanti, MT selaku Ketua Prodi S1 Jurusan Statistika ITS dan Ibu Santi Wulan Purnami, S.Si, M.Si., Ph.D selaku dosen wali penulis.
3. Ibu Dr. Ismaini Zain, M.Si dan Bapak Wahyu Wibowo, S.Si M.Si selaku dosen penguji yang telah memberikan saran dan kritik untuk kebaikan Tugas Akhir ini.
4. Semua dosen Statistika ITS yang telah memberikan ilmu kepada penulis.
5. Anisa Nurindah yang telah bersedia meluangkan waktu untuk berbagi senyuman indah dengan penulis.
6. Teman-teman sekontrakan yaitu Wildan, Iلمان, Zul, Nanda, Shu dan Epa yang selalu berbagi tawa dan canda dengan penulis.
7. Mentari dan Mbak Fanni yang menjadi teman seperjuangan dengan penulis.
8. Semua teman-teman Statistika Angkatan 2011 khususnya yang terus bergerak untuk meraih PW 112.
9. Dan semua pihak yang telah membantu dalam penyelesaian Tugas Akhir ini yang tidak bisa disebutkan satu persatu.

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari sempurna. Oleh karena itu penulis memohon maaf serta mengharapkan kritik dan saran yang bersifat membangun untuk perbaikan di masa yang akan datang. Harapannya semoga Tugas Akhir ini dapat memberikan manfaat bagi ilmu pengetahuan.

Surabaya, Juli 2015

Dwi Maumere Putra

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
LEMBAR PENGESAHAN	iii
ABSTRAK	v
ABSTRACT	vii
KATA PENGANTAR	ix
DAFTAR ISI	xi
DAFTAR NOTASI	xv
DAFTAR TABEL	xvii
DAFTAR GAMBAR	xix
DAFTAR LAMPIRAN	xxi
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	4
1.3 Tujuan.....	4
1.4 Manfaat.....	5
BAB II TINJAUAN PUSTAKA	7
2.1 Multikolinieritas	7
2.2 Regresi Logistik	8
2.3 Regresi Ridge	13
2.4 Regresi Logistik Ridge.....	16
2.5 Pengujian Signifikansi Parameter	19
2.6 Indeks Pembangunan Manusia (IPM)	21
2.7 Penelitian Sebelumnya	22
BAB III METODOLOGI PENELITIAN	25
3.1 Sumber Data.....	25
3.2 Variabel Penelitian dan Definisi Operasional	25
3.3 Metode Analisis Data.....	28
BAB IV ANALISIS DAN PEMBAHASAN	31
4.1 Statistika Deskriptif Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Indeks Pembangunan Manusia (IPM)31	
4.1.1 Deskriptif Kabupaten/Kota di Provinsi Jawa Timur Menurut Sektor Kesehatan	32

4.1.2	Deskriptif Kabupaten/Kota di Provinsi Jawa Timur Menurut Sektor Pendidikan	34
4.1.3	Deskriptif Kabupaten/Kota di Provinsi Jawa Timur Menurut Sektor Ekonomi.....	35
4.1.4	Identifikasi Kasus Multikolinieritas	37
4.2	Pemodelan Regresi Logistik Ridge	38
4.2.1	Pengujian Signifikansi Parameter Regresi Logistik Ridge Secara Serentak	40
4.2.2	Pengujian Signifikansi Parameter Regresi Logistik Ridge Secara Parsial	41
4.2.3	Model Lengkap Regresi Logistik Ridge IPM Provinsi Jawa Timur	41
4.2.4	Ketepatan Klasifikasi Model Lengkap Regresi Logistik Ridge	42
4.2.5	Interpretasi Model Lengkap Regresi Logistik Ridge.....	42
4.3	Pemilihan Model Regresi Logistik Ridge Terbaik	44
4.3.1	Pengujian Signifikansi Parameter Regresi Logistik Ridge Secara Serentak.....	45
4.3.2	Pengujian Signifikansi Parameter Regresi Logistik Ridge Secara Parsial	46
4.3.3	Model Terbaik Regresi Logistik Ridge IPM Provinsi Jawa Timur	47
4.3.4	Ketepatan Klasifikasi Model Terbaik Regresi Logistik Ridge	47
4.3.5	Interpretasi Model Terbaik Regresi Logistik Ridge.....	48
4.3.6	Pemetaan Model Terbaik Regresi Logistik Ridge	49
BAB V KESIMPULAN DAN SARAN.....		51
5.1	Kesimpulan	51
5.2	Saran	52
DAFTAR PUSTAKA		53
LAMPIRAN		57

BIODATA PENULIS73

(Halaman ini sengaja dikosongkan)

DAFTAR TABEL

	Halaman
Tabel 3.1 Variabel Penelitian Indikator IPM.....	25
Tabel 4.1 Deskripsi Kabupaten/Kota di Provinsi Jawa Timur Menurut Sektor Kesehatan.....	33
Tabel 4.2 Deskripsi Kabupaten/Kota di Provinsi Jawa Timur Menurut Sektor Pendidikan	34
Tabel 4.3 Deskripsi Kabupaten/Kota di Provinsi Jawa Timur Menurut Sektor Ekonomi.....	35
Tabel 4.4 Nilai <i>Variance Inflation Factors (VIF)</i>	37
Tabel 4.5 Pemilihan Nilai Ridge Paramater untuk Model Lengkap	38
Tabel 4.6 Signifikansi Parameter Model Lengkap	39
Tabel 4.7 Ketepatan Klasifikasi Hasil Model Lengkap Regresi Logistik Ridge	42
Tabel 4.8 <i>Odds Ratio</i> Hasil Model Lengkap Regresi Logistik Ridge.....	43
Tabel 4.9 Pemilihan Nilai Ridge Paramater untuk Model Terbaik.....	44
Tabel 4.10 Signifikansi Parameter Model Terbaik	45
Tabel 4.11 Ketepatan Klasifikasi Hasil Model Terbaik Regresi Logistik Ridge	47
Tabel 4.11 <i>Odds Ratio</i> Hasil Model Terbaik Regresi Logistik Ridge.....	48

(Halaman ini sengaja dikosongkan)

DAFTAR GAMBAR

	Halaman
Gambar 4.1 Jumlah Kabupaten/Kota untuk Setiap Kategori IPM.....	32
Gambar 4.2 Pemetaan Hasil Prediksi Model Terbaik Regresi Logistik Ridge.....	49

(Halaman ini sengaja dikosongkan)

DAFTAR NOTASI

Notasi	Keterangan
i	Indeks yang menunjukkan banyaknya observasi
n	Banyaknya observasi
j	Indeks yang menunjukkan banyaknya parameter
p	Banyaknya parameter/variabel prediktor
y_i	Variabel respon ke- i
\mathbf{y}	Vektor pengamatan dari variabel respon yang berukuran $n \times 1$
\mathbf{x}_i	Vektor pengamatan dari variabel prediktor ke- i yang berukuran $n \times 1$
\mathbf{X}	Matriks pengamatan dari variabel prediktor yang berukuran $n \times (h + 1)$
$\pi_i(\mathbf{x}_i)$	Probabilitas dari kejadian ke- i
β_j	Koefisien regresi ke- j
$\boldsymbol{\beta}$	Vektor dari koefisien regresi logistik
\mathbf{W}	Matriks diagonal yang komponen diagonalnya $\hat{\pi}_i(1 - \hat{\pi}_i)$
\mathbf{z}	Vektor berukuran $n \times 1$ dengan komponen $z_i = \text{Logit}[\hat{\pi}_i(\mathbf{x}_i)] + \frac{y_i - \hat{\pi}_i(\mathbf{x}_i)}{\hat{\pi}_i(\mathbf{x}_i)[1 - \hat{\pi}_i(\mathbf{x}_i)]}$
λ_j	<i>Eigen value</i> ke- j dari matriks $\mathbf{X}^T \mathbf{X}$
\mathbf{I}	Matriks identitas pada ruang dimensi tertentu
θ	Ridge parameter untuk regresi ridge, yaitu bilangan positif kecil antara 0 hingga 1
$\boldsymbol{\beta}^*$	Vektor dari koefisien regresi ridge
L_0	Nilai fungsi likelihood untuk model yang semua parameternya bernilai nol
L_1	Nilai fungsi likelihood untuk model lengkap

Notasi	Keterangan
k	Indeks yang menunjukkan komponen IPM ke- k
β^\oplus	Vektor dari koefisien regresi logistik ridge
θ^\oplus	Ridge parameter untuk regresi logistik ridge, yaitu bilangan positif kecil antara 0 hingga 1
\mathbf{Q}	Matriks <i>eigen vektor</i> dari $\mathbf{X}^T\mathbf{X}$
$\mathbf{\Lambda}$	Matriks diagonal yang komponen diagonalnya λ_j
\mathbf{Z}	<i>Principal component</i> dari \mathbf{X}
α	Koefisien <i>principal component regression</i> (PCR)
r	Indeks yang menunjukkan jumlah <i>principal component</i> regresi logistik

BAB I

PENDAHULUAN

1.1 Latar Belakang

Indonesia merupakan negara besar dengan jumlah penduduk terbesar ke empat di dunia setelah negara China, India dan Amerika Serikat. Menurut Dickson (2013) yang mengutip dari CIA World Factbook tahun 2013, menyatakan bahwa negara Indonesia menyumbang 3,5% dari jumlah penduduk dunia. Berdasarkan hasil sensus penduduk yang dilakukan oleh Badan Pusat Statistik (BPS) pada tahun 2010, jumlah penduduk Indonesia mencapai angka 237 juta jiwa lebih dengan laju pertumbuhan penduduk dari tahun 2000 hingga 2010 mencapai 1,49 juta jiwa per tahun. Hal tersebut menunjukkan bahwa bidang kependudukan di Indonesia harus sangat diperhatikan mengingat besarnya jumlah penduduk Indonesia. Jawa Timur merupakan salah satu provinsi besar di Indonesia dengan jumlah penduduk mencapai 37 juta jiwa lebih pada tahun 2010 dan pada tahun 2013 meningkat menjadi 38 juta jiwa lebih (BPS, 2010).

Besarnya jumlah penduduk Jawa Timur tersebut dapat menjadi suatu asset untuk dapat memajukan pembangunan dengan meningkatkan produktivitas apabila sumber daya manusia (SDM) yang ada dapat dimanfaatkan dengan baik, sehingga dapat meningkatkan pendapatan daerah. Hal tersebut diungkapkan oleh Darwis (2011) pada artikel BKKBN (Badan Kependudukan dan Keluarga Berencana Nasional) yang menyatakan bahwa terdapat beberapa faktor yang mempengaruhi pertumbuhan ekonomi, dimana faktor-faktor tersebut terbagi menjadi dua, yaitu faktor ekonomi dan faktor nonekonomi. Sumber Daya Manusia (SDM) merupakan salah satu faktor ekonomi yang menentukan keberhasilan pembangunan nasional melalui jumlah dan kualitas penduduk. Namun, di lain sisi jika pertumbuhan penduduk tidak terkendali, maka dapat menjadi suatu beban dalam pembangunan manusia di suatu wilayah.

Pembangunan di suatu daerah merupakan suatu upaya yang dilakukan oleh pemerintah daerah terkait kesejahteraan masyarakat serta mewujudkan kemakmuran masyarakat. Salah satu pembangunan yang menjadi perhatian pemerintah daerah adalah pembangunan manusia yang diukur menggunakan suatu indikator yang disebut sebagai Indeks Pembangunan Manusia (IPM). IPM merupakan suatu indeks komposit yang digunakan untuk mengukur capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup manusia. IPM dibangun dari tiga dimensi dasar yang digunakan sebagai ukuran kualitas hidup manusia, dimana ketiga dimensi dasar tersebut adalah indeks harapan hidup, indeks pendidikan dan indeks standar hidup layak (BPS, 2014).

IPM Provinsi Jawa Timur pada tahun 2013 sebesar 73,54. BPS mengkategorikan nilai IPM Provinsi Jawa Timur tahun 2013 dalam kategori menengah atas, namun nilai tersebut masih berada di bawah rata-rata IPM Indonesia yang bernilai 73,81 dengan selisih sebesar 0,17. Pembangunan manusia pada setiap kabupaten/ kota di Provinsi Jawa Timur masih belum merata dan terdapat kesenjangan yang cukup tinggi. Hal tersebut dapat dilihat dari nilai IPM Kota Surabaya sebesar 78,97 yang merupakan IPM tertinggi di Provinsi Jawa Timur, tetapi di sisi lain masih terdapat nilai IPM yang cukup jauh dari rata-rata IPM Jawa Timur, yaitu Kabupaten Sampang yang hanya sebesar 62,39 (BPS, 2013). Sehingga perlu dilakukan penelitian untuk mengetahui faktor-faktor yang memengaruhi nilai IPM untuk setiap kabupaten/kota di Provinsi Jawa Timur.

Yunitasari (2007) melakukan penelitian terhadap IPM Jawa Timur, dimana faktor-faktor yang mempengaruhinya adalah PDRB perkapita, tingkat kemiskinan, pengeluaran pemerintah untuk sektor pendidikan dan kesehatan. Dengan menggunakan data *cross section* dan *time series*, Melliana (2013) melakukan penelitian terhadap IPM Jawa Timur menggunakan Regresi Panel. Faktor-faktor yang digunakan adalah rasio guru-siswa, rasio sekolah-murid, angka partisipasi sekolah, jumlah sarana

kesehatan, rumah tangga dengan akses air bersih, kepadatan penduduk, tingkat partisipasi angkatan kerja dan PDRB perkapita. Di tahun yang sama, Astri (2013) melakukan penelitian terhadap IPM Indonesia menggunakan Regresi Panel dimana faktor yang digunakan adalah pengeluaran pemerintah daerah pada sektor pendidikan dan kesehatan.

Analisis terhadap IPM menggunakan respon berupa data kategorik dilakukan oleh Yurvianny (2007), dimana penelitian terhadap IPM Jawa Timur dilakukan menggunakan analisis Regresi Logistik yang mendapatkan model dengan ketepatan klasifikasi yang tinggi, yaitu sebesar 94,7%. Nur (2010) juga melakukan penelitian terhadap IPM Jawa Timur, Jawa Tengah, Jawa Barat dan Sumatera Utara menggunakan Regresi Logistik Ordinal. Pada tahun berikutnya, Pradita (2011) melakukan penelitian terhadap IPM Jawa Timur menggunakan analisis *Geographically Weighted Logistic Regression* (GWLR), tetapi Provinsi Jawa Timur hanya terdiri dari 29 kabupaten dan 9 kota, sehingga analisis GWLR tidak cukup baik digunakan pada data yang jumlahnya kecil. Sedangkan penggunaan metode Regresi Logistik Ridge dilakukan oleh Sunyoto (2009) pada analisis keberhasilan siswa SMA Negeri 1 Kediri diterima di Perguruan Tinggi Negeri (PTN).

Faktor-faktor di setiap sektor akan cenderung memiliki hubungan yang kuat satu sama lainnya, karena antar faktor-faktor tersebut saling mempengaruhi. Hal tersebut dapat mengakibatkan adanya kasus multikolinieritas yang serius. Kasus multikolinieritas merupakan suatu kondisi dimana terjadi korelasi yang tinggi di antara variabel prediktor atau dapat dikatakan antar variabel prediktor tidak bersifat saling bebas (Yan & Su, 2009). Apabila kasus multikolinieritas tidak diatasi, maka dapat menyebabkan variansi dari hasil estimasi parameter menjadi besar yang akan berakibat pada besarnya standar error serta interval kepercayaan juga akan menjadi lebar. Selain itu dapat berakibat pada banyaknya variabel prediktor yang tidak signifikan meskipun nilai koefisien determinasi (R^2) tinggi. Untuk dapat

menyelesaikan kasus tersebut dapat dilakukan dengan menambahkan suatu bilangan positif kecil yang disebut *ridge parameter* pada estimasi parameter saat pembentukan pemodelan. Sehingga berdasarkan penyelesaian kasus multikolinieritas dan struktur data IPM Provinsi Jawa Timur yang telah dikategorikan, maka digunakan metode Regresi Logistik Ridge untuk memodelkan Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur.

1.2 Rumusan Masalah

Sektor-sektor pembentuk IPM adalah pendidikan, kesehatan dan ekonomi dimana faktor-faktor dalam setiap sektor tersebut cenderung memiliki hubungan/korelasi yang tinggi satu sama lain. Sehingga antar faktor yang mempengaruhi IPM cenderung tidak saling bebas atau memiliki kolinieritas yang tinggi. Hal tersebut akan menyebabkan adanya kasus multikolinieritas pada faktor-faktor yang mempengaruhi IPM. Apabila kasus multikolinieritas tidak diatasi, maka dapat menyebabkan variansi dari hasil estimasi parameter menjadi besar yang dapat berakibat pada banyaknya variabel prediktor yang tidak signifikan meskipun nilai koefisien determinasi (R^2) tinggi. Oleh karena itu dalam penelitian ini dilakukan pemodelan IPM Provinsi Jawa Timur dengan menggunakan metode Regresi Logistik Ridge untuk mengatasi adanya kasus multikolinieritas tersebut dengan rumusan masalah sebagai berikut.

1. Bagaimana karakteristik Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur beserta faktor-faktor yang mempengaruhinya?
2. Bagaimana pemodelan Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur menggunakan metode Regresi Logistik Ridge?
3. Bagaimana model terbaik Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur menggunakan metode *backward elimination* pada Regresi Logistik Ridge?

1.3 Tujuan

Berdasarkan latar belakang dan rumusan masalah, maka tujuan yang ingin dicapai dalam penelitian ini adalah.

1. Mendeskripsikan karakteristik Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur beserta faktor-faktor yang mempengaruhinya.
2. Memodelkan Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur menggunakan metode Regresi Logistik Ridge.
3. Mendapatkan model terbaik Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur menggunakan metode *backward elimination* pada Regresi Logistik Ridge.

1.4 Manfaat

Manfaat dari penelitian ini adalah dapat mengembangkan wawasan ilmu pengetahuan yang berkaitan dengan penyelesaian kasus multikolinieritas menggunakan Regresi Logistik Ridge. Selain itu dapat memberikan informasi kepada Pemerintah Provinsi Jawa Timur terkait wilayah-wilayah yang memiliki Indeks Pembangunan Manusia (IPM) rendah untuk segera dilakukan kebijakan-kebijakan sebagai upaya dalam peningkatan IPM di wilayah tersebut berdasarkan faktor-faktor terkait.

(Halaman ini sengaja dikosongkan)

BAB II

TINJAUAN PUSTAKA

Pada bab ini dibahas mengenai beberapa kajian pustaka yang berkaitan dengan teori dasar Regresi Logistik Ridge. Di bagian akhir dibahas mengenai Indeks Pembangunan Manusia (IPM) beserta indeks komposit yang digunakan oleh Badan Pusat Statistik dalam penghitungan nilai IPM.

2.1 Multikolinieritas

Multikolinieritas merupakan suatu kondisi dimana terjadi korelasi yang tinggi di antara variabel prediktor atau dapat dikatakan antar variabel prediktor tidak bersifat saling bebas. Salah satu kriteria yang dapat digunakan untuk mendeteksi adanya kasus multikolinieritas pada regresi linier berganda adalah *Variance Inflation Factor (VIF)*. *VIF* dapat digunakan sebagai kriteria untuk mendeteksi kasus multikolinieritas pada regresi linier yang memiliki lebih dari dua variabel prediktor. Nilai *VIF* untuk parameter regresi ke- j dirumuskan pada persamaan (2.1) sebagai berikut (Yan & Su, 2009).

$$VIF_j = \frac{1}{1 - R_j^2} \quad (2.1)$$

Dengan R_j^2 merupakan koefisien determinasi antara X_j dengan variabel prediktor lainnya pada persamaan atau model regresi, dimana $j = 1, 2, \dots, p$. Apabila nilai *VIF* lebih dari 5, maka dapat diindikasikan terdapat kasus multikolinieritas. Sedangkan jika nilai *VIF* lebih besar dari 10, maka pada variabel prediktor terjadi kasus multikolinieritas yang serius dan harus diatasi.

Hal-hal yang akan terjadi apabila kasus multikolinieritas tidak diatasi adalah sebagai berikut (Yan & Su, 2009).

1. Variansi estimasi menjadi besar
2. Interval kepercayaan menjadi lebar, dikarenakan variansi dan standar error besar
3. Pengujian signifikansi secara parsial menjadi tidak signifikan

4. Koefisien determinasi (R^2) tinggi, tetapi tidak banyak variabel prediktor yang signifikan.

2.2 Regresi Logistik

Model regresi logistik digunakan untuk menggambarkan hubungan antara variabel respon dengan satu atau beberapa buah variabel prediktor. Menurut Agresti (2007), dalam model regresi logistik dapat menggunakan variabel independen yang berupa kualitatif (berskala pengukuran nominal atau ordinal) atau kuantitatif (berskala pengukuran interval atau rasio) atau gabungan dari keduanya. Variabel dependen dalam regresi logistik merupakan variabel kualitatif. Apabila kategori dalam variabel dependen berupa biner/dikotomis, maka variabel dependen dapat mengambil nilai 0 dengan suatu kemungkinan gagal atau nilai 1 dengan kemungkinan sukses. Variabel respon Y berdistribusi Binomial dengan parameter π_i , dimana untuk setiap pengamatan ke- i ditulis pada persamaan (2.2) (Agresti, 2007).

$$y_i \sim \text{Binomial}(1, \pi_i) \quad (2.2)$$

Dengan fungsi probabilitas pada persamaan (2.3).

$$f(y_i) = (\pi_i(\mathbf{x}_i))^{y_i} (1 - \pi_i(\mathbf{x}_i))^{1-y_i}, y_i = 0, 1 \quad (2.3)$$

dimana $\pi_i(\mathbf{x}_i)$ adalah probabilitas dari kejadian ke- i . Apabila $y_i = 1$, maka $f(y_i) = \pi_i(\mathbf{x}_i)$ dan apabila $y_i = 0$, maka $f(y_i) = 1 - \pi_i(\mathbf{x}_i)$.

Dalam regresi logistik, hubungan antara variabel prediktor dan variabel respon bukanlah suatu fungsi linier. Bentuk persamaan regresi logistik ditunjukkan pada persamaan (2.4).

$$\pi_i(\mathbf{x}_i) = \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}$$

$$\pi_i(\mathbf{x}_i) = \frac{\exp(\mathbf{x}\boldsymbol{\beta})}{1 + \exp(\mathbf{x}\boldsymbol{\beta})} \quad (2.4)$$

dimana β_0 = konstanta, β_j = koefisien regresi dan j = banyaknya variabel prediktor. Terdapat suatu bentuk alternatif dari persamaan regresi logistik seperti pada persamaan (2.5) yang merupakan transformasi logit dari $\pi(\mathbf{x}_i)$ (Yan & Su, 2009).

$$\begin{aligned}
\text{Logit}[\pi_i(\mathbf{x}_i)] &= \ln \left[\frac{\pi_i(\mathbf{x}_i)}{1 - \pi_i(\mathbf{x}_i)} \right] \\
\text{Logit}[\pi_i(\mathbf{x}_i)] &= \ln \left[\frac{\frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}}{1 - \frac{\exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}{1 + \exp(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})}} \right] \\
\text{Logit}[\pi_i(\mathbf{x}_i)] &= \ln \left[\frac{\frac{\exp(\mathbf{x}\boldsymbol{\beta})}{1 + \exp(\mathbf{x}\boldsymbol{\beta})}}{1 - \frac{\exp(\mathbf{x}\boldsymbol{\beta})}{1 + \exp(\mathbf{x}\boldsymbol{\beta})}} \right] = \ln \left[\frac{\frac{\exp(\mathbf{x}\boldsymbol{\beta})}{1 + \exp(\mathbf{x}\boldsymbol{\beta})}}{\frac{1 + \exp(\mathbf{x}\boldsymbol{\beta}) - \exp(\mathbf{x}\boldsymbol{\beta})}{1 + \exp(\mathbf{x}\boldsymbol{\beta})}} \right] \\
\text{Logit}[\pi_i(\mathbf{x}_i)] &= \ln \left[\frac{\frac{\exp(\mathbf{x}\boldsymbol{\beta})}{1 + \exp(\mathbf{x}\boldsymbol{\beta})}}{1} \right] = \ln[\exp(\mathbf{x}_i \boldsymbol{\beta})] \\
\text{Logit}[\pi_i(\mathbf{x}_i)] &= \mathbf{x}_i \boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \tag{2.5}
\end{aligned}$$

Untuk menaksir parameter dalam model regresi logistik digunakan metode *Maximum Likelihood Estimator (MLE)*. Metode *MLE* digunakan karena distribusi dari variabel respon telah diketahui. *MLE* didapatkan dengan cara memaksimalkan logaritma fungsi likelihood. Dari persamaan (2.3) didapatkan fungsi likelihood pada persamaan (2.6) (Agresti, 2007).

$$\begin{aligned}
L(\mathbf{X}\boldsymbol{\beta}) &= \prod_{i=1}^n [\pi_i(\mathbf{x}_i)]^{y_i} [1 - \pi_i(\mathbf{x}_i)]^{1-y_i} \\
\ln L(\mathbf{X}\boldsymbol{\beta}) &= \ln L(\boldsymbol{\beta}) = \sum_{i=1}^n y_i \ln[\pi_i(\mathbf{x}_i)] + \sum_{i=1}^n (1 - y_i) \ln[1 - \pi_i(\mathbf{x}_i)] \\
\ln L(\mathbf{X}\boldsymbol{\beta}) &= \sum_{i=1}^n y_i \ln[\pi_i(\mathbf{x}_i)] + \sum_{i=1}^n (\ln[1 - \pi_i(\mathbf{x}_i)] - y_i \ln[1 - \pi_i(\mathbf{x}_i)]) \\
\ln L(\mathbf{X}\boldsymbol{\beta}) &= \sum_{i=1}^n y_i \ln[\pi_i(\mathbf{x}_i)] - \sum_{i=1}^n y_i \ln[1 - \pi_i(\mathbf{x}_i)] + \sum_{i=1}^n \ln[1 - \pi_i(\mathbf{x}_i)]
\end{aligned}$$

$$\begin{aligned} \ln L(\mathbf{X}\boldsymbol{\beta}) &= \sum_{i=1}^n y_i (\ln[\pi_i(\mathbf{x}_i)] - \ln[1 - \pi_i(\mathbf{x}_i)]) + \sum_{i=1}^n \ln \left[1 - \frac{\exp(\mathbf{x}_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i \boldsymbol{\beta})} \right] \\ \ln L(\mathbf{X}\boldsymbol{\beta}) &= \sum_{i=1}^n y_i \ln \left[\frac{\pi_i(\mathbf{x}_i)}{1 - \pi_i(\mathbf{x}_i)} \right] + \sum_{i=1}^n \ln \left[\frac{1}{1 + \exp(\mathbf{x}_i \boldsymbol{\beta})} \right] \\ \ln L(\mathbf{X}\boldsymbol{\beta}) &= \sum_{i=1}^n [y_i \ln(\pi_i(\mathbf{x}_i)) + \ln(1 + \exp(\mathbf{x}_i \boldsymbol{\beta}))^{-1}] \\ \ln L(\mathbf{X}\boldsymbol{\beta}) &= \sum_{i=1}^n [y_i \ln(\pi_i(\mathbf{x}_i)) - \ln(1 + \exp(\mathbf{x}_i \boldsymbol{\beta}))] \end{aligned} \quad (2.6)$$

Berdasarkan persamaan (2.6) dilakukan penurunan terhadap $\boldsymbol{\beta}$ menjadi persamaan (2.7) (Ryan, 1997).

$$\begin{aligned} \frac{\partial \ln L(\mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^n \left[y_i \mathbf{x}_i - \frac{\mathbf{x}_i \exp(\mathbf{x}_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i \boldsymbol{\beta})} \right] = \sum_{i=1}^n \mathbf{x}_i \left[y_i - \frac{\exp(\mathbf{x}_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i \boldsymbol{\beta})} \right] \\ \frac{\partial \ln L(\mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^n \mathbf{x}_i [y_i - \pi_i(\mathbf{x}_i)] \\ \frac{\partial \ln L(\mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \mathbf{X}^T (\mathbf{y} - \hat{\boldsymbol{\pi}}) \end{aligned}$$

dimana \mathbf{y} = vektor pengamatan pada variabel respon yang berukuran $n \times 1$, sedangkan \mathbf{X} = matriks variabel prediktor yang berukuran $n \times (p + 1)$.

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

Jika $\frac{\partial \ln L(\mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0$ dan $\hat{\boldsymbol{\pi}} = \boldsymbol{\pi}^*$, maka didapatkan persamaan (2.8)

sebagai berikut.

$$\mathbf{X}^T (\mathbf{y} - \hat{\boldsymbol{\pi}}) = 0 \quad (2.8)$$

Hasil penurunan kedua ditunjukkan pada persamaan (2.9) (Ryan, 1997).

$$\frac{\partial^2 \ln L(\mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = \frac{\partial}{\partial \boldsymbol{\beta}^T} \left(\frac{\partial \ln L(\mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right) = \frac{\partial}{\partial \boldsymbol{\beta}^T} \left(\sum_{i=1}^n \left[y_i \mathbf{x}_i - \frac{\mathbf{x}_i \exp(\mathbf{x}_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i \boldsymbol{\beta})} \right] \right)$$

$$\begin{aligned}
\frac{\partial^2 \ln L(\mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} &= 0 - \sum_{i=1}^n \frac{\mathbf{x}_i \mathbf{x}_i^T \left[\exp(\mathbf{x}_i \boldsymbol{\beta}) - \frac{\exp(\mathbf{x}_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i \boldsymbol{\beta})} \right]^2}{[1 + \exp(\mathbf{x}_i \boldsymbol{\beta})]^2} \\
\frac{\partial^2 \ln L(\mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} &= - \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T \left(\frac{\exp(\mathbf{x}_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i \boldsymbol{\beta})} - \left[\frac{\exp(\mathbf{x}_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i \boldsymbol{\beta})} \right]^2 \right) \\
\frac{\partial^2 \ln L(\mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} &= - \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T \left(\frac{\exp(\mathbf{x}_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i \boldsymbol{\beta})} \right) \left(1 - \frac{\exp(\mathbf{x}_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i \boldsymbol{\beta})} \right) \\
\frac{\partial^2 \ln L(\mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} &= - \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T \pi_i(\mathbf{x}_i) (1 - \pi_i(\mathbf{x}_i)) \\
\frac{\partial^2 \ln L(\mathbf{X}\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} &= -\mathbf{X}^T \mathbf{W} \mathbf{X} \tag{2.9}
\end{aligned}$$

dimana :

$$\mathbf{W} = \begin{bmatrix} \hat{\pi}_1(1-\hat{\pi}_1) & 0 & \dots & 0 \\ 0 & \hat{\pi}_2(1-\hat{\pi}_2) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{\pi}_n(1-\hat{\pi}_n) \end{bmatrix}$$

Apabila dilakukan ekspansi berdasarkan Deret Taylor di sekitar nilai $\hat{\boldsymbol{\beta}}$, maka didapatkan persamaan (2.10).

$$\frac{\partial L(\mathbf{X}\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} = \frac{\partial L(\mathbf{X}\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} \Big|_{\hat{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}_0} + \frac{\partial^2 L(\mathbf{X}\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}} \partial \hat{\boldsymbol{\beta}}^T} \Big|_{\hat{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}_0} (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_0)$$

Jika $\frac{\partial L(\mathbf{X}\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} = 0$, maka :

$$\begin{aligned}
\frac{\partial L(\mathbf{X}, \hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} \Big|_{\hat{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}_0} + \frac{\partial^2 L(\mathbf{X}, \hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}} \partial \hat{\boldsymbol{\beta}}^T} \Big|_{\hat{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}_0} (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_0) &= 0 \\
\frac{\partial L(\mathbf{X}, \hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} \Big|_{\hat{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}_0} &= - \frac{\partial^2 L(\mathbf{X}, \hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}} \partial \hat{\boldsymbol{\beta}}^T} \Big|_{\hat{\boldsymbol{\beta}}=\hat{\boldsymbol{\beta}}_0} (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_0) \tag{2.10}
\end{aligned}$$

Hasil substitusi persamaan (2.7) dan (2.9) ke dalam persamaan (2.10) menghasilkan estimasi parameter $\hat{\boldsymbol{\beta}}$ yang ditunjukkan pada persamaan (2.11) (Hosmer & Lemeshow, 2000).

$$\begin{aligned}
\mathbf{X}^T(\mathbf{y} - \hat{\boldsymbol{\pi}}) &= \mathbf{X}(\mathbf{W}\mathbf{X}^T \boldsymbol{\beta} - \hat{\boldsymbol{\pi}}_0) \\
\mathbf{X}^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) &= \mathbf{W}\mathbf{X}^T \boldsymbol{\beta} - \hat{\boldsymbol{\pi}}_0 \\
\mathbf{X}^T(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) &= \mathbf{W}\mathbf{X}^T \boldsymbol{\beta} - \mathbf{X}^T \mathbf{W}\mathbf{X} \hat{\boldsymbol{\beta}}_0 \\
\mathbf{X}^T \mathbf{W} \hat{\boldsymbol{\beta}} &= \mathbf{X}^T \mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}}_0 + \mathbf{X}^T \mathbf{y} - \hat{\boldsymbol{\pi}} \\
\mathbf{X}^T \mathbf{W} \hat{\boldsymbol{\beta}} &= \mathbf{X}^T \mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}}_0 + \mathbf{X}^T \mathbf{W} \mathbf{W}^{-1} \mathbf{y} - \hat{\boldsymbol{\pi}} \\
\mathbf{X}^T \mathbf{W} \hat{\boldsymbol{\beta}} &= \mathbf{X}^T \mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}}_0 + \mathbf{W}^{-1} \mathbf{y} - \hat{\boldsymbol{\pi}} \\
\hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} [\mathbf{X} \hat{\boldsymbol{\beta}}_0 + \mathbf{W}^{-1} (\mathbf{y} - \hat{\boldsymbol{\pi}})] \\
\hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{z} \tag{2.11}
\end{aligned}$$

dimana \mathbf{z} merupakan vektor $n \times 1$ dengan :

$$z_i = \text{Logit}[\hat{\pi}_i(\mathbf{x}_i)] + \frac{y_i - \hat{\pi}_i(\mathbf{x}_i)}{\hat{\pi}_i(\mathbf{x}_i)[1 - \hat{\pi}_i(\mathbf{x}_i)]} \tag{2.12}$$

Matriks kovarian untuk $\hat{\boldsymbol{\beta}}$ ditampilkan pada persamaan (2.13) (Sunyoto, 2009).

$$\begin{aligned}
\text{Var}(\hat{\boldsymbol{\beta}}) &= \text{Var}\{(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{z}\} \\
\text{Var}(\hat{\boldsymbol{\beta}}) &= [(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}] \text{Var}(\mathbf{z}) [(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}]^T \\
\text{Var}(\hat{\boldsymbol{\beta}}) &= [(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}] \text{Var}(\mathbf{X} \hat{\boldsymbol{\beta}}_0 + \mathbf{W}^{-1} (\mathbf{y} - \hat{\boldsymbol{\pi}})) [(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}]^T
\end{aligned}$$

dimana :

$$\begin{aligned}
\text{Var}(\mathbf{X} \hat{\boldsymbol{\beta}}_0 + \mathbf{W}^{-1} (\mathbf{y} - \hat{\boldsymbol{\pi}})) &= \text{Var}(\mathbf{X} \hat{\boldsymbol{\beta}}_0) + \text{Var}(\mathbf{W}^{-1} (\mathbf{y} - \hat{\boldsymbol{\pi}})) - \text{Cov}(\mathbf{X} \hat{\boldsymbol{\beta}}_0, \mathbf{W}^{-1} (\mathbf{y} - \hat{\boldsymbol{\pi}})) \\
\text{Var}(\mathbf{X} \hat{\boldsymbol{\beta}}_0 + \mathbf{W}^{-1} (\mathbf{y} - \hat{\boldsymbol{\pi}})) &= 0 + \text{Var}(\mathbf{W}^{-1} (\mathbf{y} - \hat{\boldsymbol{\pi}})) - 0 \\
\text{Var}(\mathbf{X} \hat{\boldsymbol{\beta}}_0 + \mathbf{W}^{-1} (\mathbf{y} - \hat{\boldsymbol{\pi}})) &= \text{Var}(\mathbf{W}^{-1} (\mathbf{y} - \hat{\boldsymbol{\pi}})) \\
\text{Var}(\mathbf{X} \hat{\boldsymbol{\beta}}_0 + \mathbf{W}^{-1} (\mathbf{y} - \hat{\boldsymbol{\pi}})) &= \mathbf{W}^{-1} \text{Var}(\mathbf{y} - \hat{\boldsymbol{\pi}}) (\mathbf{W}^{-1})^T
\end{aligned}$$

Sehingga :

$$\begin{aligned}
\text{Var}(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{W}^{-1} \text{Var}(\mathbf{y} - \hat{\boldsymbol{\pi}}) (\mathbf{W}^{-1})^T [(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}]^T \\
\text{Var}(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \text{Var}(\mathbf{y} - \hat{\boldsymbol{\pi}}) (\mathbf{W}^{-1})^T \mathbf{W}^T \mathbf{X} [(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}]^T \\
\text{Var}(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \text{Var}(\mathbf{y} - \hat{\boldsymbol{\pi}}) \mathbf{X} [(\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}]^T \\
\text{Var}(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \text{Var}(\mathbf{y} - \hat{\boldsymbol{\pi}}) \mathbf{X} (\mathbf{X}^T \mathbf{W}^T \mathbf{X})^{-1} \\
\text{Var}(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^T \mathbf{X} (\mathbf{X}^T \mathbf{W}^T \mathbf{X})^{-1} \\
\text{Var}(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1}
\end{aligned}$$

$$\text{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T \text{diag}[\hat{\pi}_i(1-\hat{\pi}_i)]\mathbf{X})^{-1} \quad (2.13)$$

2.3 Regresi Ridge

Regresi ridge merupakan suatu metode dalam analisis regresi yang dikembangkan berdasarkan metode kuadrat terkecil (*least square*) untuk data yang mempunyai kasus multikolinieritas (Masruroh, 2011). Regresi Ridge adalah salah satu metode yang dapat digunakan untuk mengatasi kasus multikolinieritas, dimana parameter regresi akan distabilkan dan efek multikolinieritas dapat teratasi. Metode ini ditujukan untuk mengatasi kondisi buruk yang diakibatkan oleh korelasi yang tinggi antara beberapa variabel prediktor dalam model regresi, sehingga menyebabkan matriks $\mathbf{X}^T\mathbf{X}$ hampir singular yang dapat menghasilkan hasil estimasi dari parameter model regresi menjadi tidak stabil (Draper & Smith, 1998). Estimasi parameter regresi ridge menggunakan metode *least square* (LS) dengan menambahkan bilangan positif kecil θ pada diagonal matriks $\mathbf{X}^T\mathbf{X}$, sehingga bias yang terjadi dapat dikendalikan. Bilangan positif kecil θ bernilai antara 0 dan 1, apabila nilainya nol estimasi regresi ridge akan sama dengan estimasi LS pada Regresi Linier. Apabila nilai $\theta > 0$, maka estimasi ridge akan bias terhadap parameter β , tetapi cenderung lebih stabil (Sunyoto, 2009).

Pemilihan nilai θ dilakukan untuk mendapatkan nilai yang menghasilkan bias relatif kecil dan juga dapat menghasilkan estimasi yang relatif stabil. Menurut Hoerl dan Kennard (1970), cara terbaik untuk mendapatkan estimator ridge adalah dengan memperhatikan tiga hal sebagai berikut.

1. Pada nilai θ tertentu, sistem menjadi stabil dan mempunyai karakter umum sistem orthogonal.
2. Koefisien dengan tanda yang tidak sesuai pada saat $\theta = 0$, akan berubah menjadi sesuai.
3. *Sum Square Error* (*SSE*) tidak akan meningkat terlalu tinggi dibandingkan dengan *SSE* regresi pada umumnya.

Kebaikan estimator ridge dapat diukur dengan menggunakan fungsi *Mean Square Error (MSE)* pada persamaan (2.14).

$$MSE(\hat{\boldsymbol{\beta}}) = E[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^2] = E[\hat{\boldsymbol{\beta}}^2 - 2\hat{\boldsymbol{\beta}}\boldsymbol{\beta} + \boldsymbol{\beta}^2] \quad (2.14)$$

$$MSE(\hat{\boldsymbol{\beta}}) = Var(\hat{\boldsymbol{\beta}}) + (E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta})^2$$

Secara aljabar, *MSE* terbagi menjadi dua bagian yaitu :

$$MSE(\hat{\boldsymbol{\beta}}) = Var(\hat{\boldsymbol{\beta}}) + [\text{bias}(\hat{\boldsymbol{\beta}})]^2$$

Dimana $\text{bias}(\hat{\boldsymbol{\beta}}) = E(\hat{\boldsymbol{\beta}}) - \boldsymbol{\beta}$

Apabila $\boldsymbol{\beta}$ merupakan vektor parameter, maka *MSE* dinyatakan dalam persamaan (2.15) (Vago & Kemény, 2006).

$$MSE(\hat{\boldsymbol{\beta}}) = Tr[Var(\hat{\boldsymbol{\beta}})] + [\text{bias}(\hat{\boldsymbol{\beta}})]^T [\text{bias}(\hat{\boldsymbol{\beta}})]$$

Model regresi linier : $y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i$ atau $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. *Ordinary Least Square (OLS)* estimator didapatkan dengan meminimumkan fungsi dari persamaan (2.16).

$$\phi(\hat{\boldsymbol{\beta}}) = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (2.16)$$

$$\begin{aligned} \phi(\hat{\boldsymbol{\beta}}) &= \sum_{i=1}^n (y_i - \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip})^2 \\ &= \sum_{i=1}^n (y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}})^2 \text{ diturunkan parsial terhadap } \hat{\boldsymbol{\beta}} \\ \frac{\partial \phi(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} &= \frac{\partial}{\partial \hat{\boldsymbol{\beta}}} \sum_{i=1}^n (y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}})^2 = \sum_{i=1}^n 2(y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}})(-\mathbf{x}_i) \\ &= -2 \sum_{i=1}^n (\mathbf{x}_i (y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}})) = -2 \sum_{i=1}^n (\mathbf{x}_i^T \mathbf{x}_i \hat{\boldsymbol{\beta}} - y_i \mathbf{x}_i^T) \end{aligned}$$

Jika $\frac{\partial \phi(\hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} = 0$, maka :

$$\begin{aligned}
-2 \sum_{i=1}^n (\mathbf{x}_i^T \mathbf{y} \boldsymbol{\beta} - \mathbf{x}_i^T \mathbf{x}_i \hat{\boldsymbol{\beta}}) &= 0 \\
\sum_{i=1}^n \mathbf{x}_i^T \mathbf{y} \boldsymbol{\beta} &= \sum_{i=1}^n \mathbf{x}_i^T \mathbf{x}_i \hat{\boldsymbol{\beta}} \\
\hat{\boldsymbol{\beta}} &= \left(\sum_{i=1}^n \mathbf{x}_i^T \mathbf{x}_i \right)^{-1} \sum_{i=1}^n \mathbf{x}_i^T \mathbf{y}_i
\end{aligned}$$

Kemudian dituliskan dalam bentuk matriks pada persamaan (2.17) (Draper & Smith, 1998).

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (2.17)$$

Kovariansi matriks $\hat{\boldsymbol{\beta}}$ ditunjukkan pada persamaan (2.18) (Yan & Su, 2009).

$$\begin{aligned}
\text{Var}(\hat{\boldsymbol{\beta}}) &= \text{Var}[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}] = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{Var}(\mathbf{y}) [(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T]^T \\
\text{Var}(\hat{\boldsymbol{\beta}}) &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \sigma^2 \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\
\text{Var}(\hat{\boldsymbol{\beta}}) &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \quad (2.18)
\end{aligned}$$

Dengan menggunakan persamaan (2.15), didapatkan *MSE* untuk regresi linier berganda pada persamaan (2.19) (Yan & Su, 2009).

$$MSE(\hat{\boldsymbol{\beta}}) = \text{Tr}[\sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}] = \sigma^2 \text{Tr}[(\mathbf{X}^T \mathbf{X})^{-1}] = \sigma^2 \sum_{j=1}^p \frac{1}{\lambda_j} \quad (2.19)$$

dimana λ_j merupakan *eigen value* ke- j dari matriks $\mathbf{X}^T \mathbf{X}$. Apabila nilai λ_j relatif kecil, nilai dari *MSE* akan meningkat. Hal tersebut menunjukkan bahwa estimasi parameter tidak mendekati nilai yang sebenarnya. Kemudian pada regresi ridge, estimasi parameter ditambahkan dengan *ridge parameter* (θ) pada elemen diagonal matriks $\mathbf{X}^T \mathbf{X}$, dimana *ridge parameter* (θ) merupakan bilangan positif kecil, sehingga nilai koefisien untuk parameter Regresi Ridge dalam bentuk matriks dituliskan pada persamaan (2.20) (Draper & Smith, 1998).

$$\hat{\boldsymbol{\beta}}^* = (\mathbf{X}^T \mathbf{X} + \theta \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y} \quad (2.20)$$

yang didapatkan dengan meminimumkan fungsi obyektif pada persamaan (2.21) (Yan & Su, 2009).

$$\phi(\hat{\boldsymbol{\beta}}^*) = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^*)^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^*) + \theta \hat{\boldsymbol{\beta}}^{*T} \hat{\boldsymbol{\beta}}^* \quad (2.21)$$

$$\phi(\hat{\boldsymbol{\beta}}^*) = \sum_{i=1}^n (y_i - \hat{\beta}_0^* + \hat{\beta}_1^* x_{i1} + \hat{\beta}_2^* x_{i2} + \dots + \hat{\beta}_p^* x_{ip})^2 + \theta \hat{\boldsymbol{\beta}}^{*T} \hat{\boldsymbol{\beta}}^*$$

$$\phi(\hat{\boldsymbol{\beta}}^*) = \sum_{i=1}^n (y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}^*)^2 + \theta \hat{\boldsymbol{\beta}}^{*T} \hat{\boldsymbol{\beta}}^*$$

$$\frac{\partial \phi(\hat{\boldsymbol{\beta}}^*)}{\partial \hat{\boldsymbol{\beta}}^*} = \sum_{i=1}^n 2[(y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}^*)(-\mathbf{x}_i)] + 2\theta \hat{\boldsymbol{\beta}}^* = \sum_{i=1}^n 2[-\mathbf{x}_i (y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}^*)] + 2\theta \hat{\boldsymbol{\beta}}^*$$

$$\text{Jika } \frac{\partial \phi(\hat{\boldsymbol{\beta}}^*)}{\partial \hat{\boldsymbol{\beta}}^*} = 0, \text{ maka } \sum_{i=1}^n 2[-\mathbf{x}_i (y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}^*)] + 2\theta \hat{\boldsymbol{\beta}}^* = 0$$

$$\sum_{i=1}^n [-\mathbf{x}_i^T y_i + \mathbf{x}_i^T \hat{\boldsymbol{\beta}}^*] + \theta \hat{\boldsymbol{\beta}}^* = 0$$

$$\sum_{i=1}^n \mathbf{x}_i^T \hat{\boldsymbol{\beta}}^* + \theta \hat{\boldsymbol{\beta}}^* = \sum_{i=1}^n y_i$$

$$\left(\sum_{i=1}^n \mathbf{x}_i^T \hat{\boldsymbol{\beta}}^* + \theta \right) \hat{\boldsymbol{\beta}}^* = \sum_{i=1}^n y_i$$

$$\hat{\boldsymbol{\beta}}^* = \left(\sum_{i=1}^n \mathbf{x}_i^T \mathbf{x}_i + \theta \right)^{-1} \sum_{i=1}^n \mathbf{x}_i^T y_i$$

$$\hat{\boldsymbol{\beta}}^* = (\mathbf{X}^T \mathbf{X} + \theta \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y}$$

Kemudian nilai *MSE* regresi ridge diformulasikan pada persamaan (2.22).

$$MSE = Tr[\sigma^2 (\mathbf{X}^T \mathbf{X} + \theta \mathbf{I})^{-1}] = \sigma^2 Tr[(\mathbf{X}^T \mathbf{X} + \theta \mathbf{I})^{-1}] = \sigma^2 \sum_{j=1}^p \frac{1}{(\lambda_j + \theta)} \quad (2.22)$$

$\hat{\boldsymbol{\beta}}$ merupakan estimator dari regresi linier berganda, sedangkan $\hat{\boldsymbol{\beta}}^*$ merupakan estimator untuk regresi ridge. Metode regresi ridge akan memperbesar nilai *eigen value*, sehingga dapat mengurangi *MSE* atau dapat dikatakan bahwa perhitungan nilai $MSE(OLS) > MSE(Ridge)$ (Vago & Kemény, 2006).

2.4 Regresi Logistik Ridge

Fungsi obyektif untuk Regresi Ridge dan Regresi Logistik yang didapatkan dari model linier $\mathbf{y} = \beta \mathbf{X} + \boldsymbol{\varepsilon}$ pada persamaan di subbab sebelumnya masing-masing adalah :

$$\phi(\hat{\boldsymbol{\beta}}^*) = (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^*)^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}^*) + \theta \hat{\boldsymbol{\beta}}^{*T} \hat{\boldsymbol{\beta}}^* \text{ dan}$$

$$\phi(\boldsymbol{\beta}) = \sum_{i=1}^n y_i \ln[\pi_i(\mathbf{x}_i)] + \sum_{i=1}^n (1 - y_i) \ln[1 - \pi_i(\mathbf{x}_i)]$$

Kemudian dengan menerapkan kedua teknik pada Regresi Ridge dan Regresi Logistik, didapatkan fungsi obyektif untuk Regresi Logistik Ridge pada persamaan (2.23) (Sunyoto, 2009).

$$\begin{aligned} \phi(\hat{\boldsymbol{\beta}}^{\oplus}) &= \sum_{i=1}^n y_i \ln[\pi_i(\mathbf{x}_i)] + \sum_{i=1}^n (1 - y_i) \ln[1 - \pi_i(\mathbf{x}_i)] - \theta \hat{\boldsymbol{\beta}}^{\oplus T} \hat{\boldsymbol{\beta}}^{\oplus} \\ \phi(\hat{\boldsymbol{\beta}}^{\oplus}) &= \sum_{i=1}^n y_i \ln[\pi_i(\mathbf{x}_i)] + \sum_{i=1}^n \ln[1 - \pi_i(\mathbf{x}_i)] - \sum_{i=1}^n y_i \ln[1 - \pi_i(\mathbf{x}_i)] - \theta \hat{\boldsymbol{\beta}}^{\oplus T} \hat{\boldsymbol{\beta}}^{\oplus} \\ \phi(\hat{\boldsymbol{\beta}}^{\oplus}) &= \sum_{i=1}^n y_i \ln \left[\frac{\pi_i(\mathbf{x}_i)}{1 - \pi_i(\mathbf{x}_i)} \right] + \sum_{i=1}^n \ln[1 - \pi_i(\mathbf{x}_i)] - \theta \hat{\boldsymbol{\beta}}^{\oplus T} \hat{\boldsymbol{\beta}}^{\oplus} \\ \phi(\hat{\boldsymbol{\beta}}^{\oplus}) &= \sum_{i=1}^n \left[y_i (\mathbf{x}_i \hat{\boldsymbol{\beta}}^{\oplus}) - \ln(1 + \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^{\oplus})) \right] - \theta \hat{\boldsymbol{\beta}}^{\oplus T} \hat{\boldsymbol{\beta}}^{\oplus} \end{aligned} \quad (2.23)$$

Dimana $\hat{\boldsymbol{\beta}}^{\oplus}$ merupakan koefisien parameter untuk Regresi Logistik Ridge. Sedangkan y_i merupakan respon berupa kategorik yang mengikuti distribusi Binomial $(1, \pi_i)$ dan \mathbf{x}_i merupakan vektor untuk setiap observasi yang diambil dari matriks variabel prediktor berukuran $n \times (p + 1)$.

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

Selanjutnya diturunkan secara parsial terhadap $\hat{\boldsymbol{\beta}}^{\oplus}$ (Sunyoto, 2009) :

$$\begin{aligned}
\frac{\partial \phi(\hat{\boldsymbol{\beta}}^\oplus)}{\partial \hat{\boldsymbol{\beta}}^\oplus} &= \frac{\partial}{\partial \hat{\boldsymbol{\beta}}^\oplus} \left[\sum_{i=1}^n \left[y_i (\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus) - \ln(1 + \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus)) \right] - \boldsymbol{\Phi}^{\oplus T} \hat{\boldsymbol{\beta}}^\oplus \right] \\
&= \sum_{i=1}^n \left[y_i \mathbf{x}_i - \frac{\mathbf{x}_i \boldsymbol{\Phi} \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus)}{1 + \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus)} \right] - 2\theta^\oplus \\
&= \sum_{i=1}^n \mathbf{x}_i \left[y_i - \frac{\exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus)}{1 + \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus)} \right] - 2\theta^\oplus \\
&= \sum_{i=1}^n \mathbf{x}_i [y_i - \pi_i(\mathbf{x}_i)] - 2\theta^\oplus \\
&= \mathbf{X}^T (\mathbf{y} - \boldsymbol{\pi}) - 2\theta^\oplus
\end{aligned}$$

Kemudian dilanjutkan pada turunan kedua (Sunyoto, 2009) :

$$\begin{aligned}
\frac{\partial^2 \phi(\hat{\boldsymbol{\beta}}^\oplus)}{\partial \hat{\boldsymbol{\beta}}^\oplus \partial \hat{\boldsymbol{\beta}}^{\oplus T}} &= \frac{\partial}{\partial \hat{\boldsymbol{\beta}}^{\oplus T}} \left[\frac{\partial \phi(\hat{\boldsymbol{\beta}}^\oplus)}{\partial \hat{\boldsymbol{\beta}}^\oplus} \right] = \frac{\partial}{\partial \hat{\boldsymbol{\beta}}^{\oplus T}} \left[\sum_{i=1}^n \left[\mathbf{x}_i y_i - \frac{\mathbf{x}_i^T \boldsymbol{\Phi} \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus)}{1 + \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus)} \right] - 2\theta^\oplus \right] \\
&= \frac{\partial}{\partial \hat{\boldsymbol{\beta}}^{\oplus T}} \left[\sum_{i=1}^n \left[\mathbf{x}_i y_i - \frac{\mathbf{x}_i^T \boldsymbol{\Phi} \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus)}{1 + \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus)} \right] - 2\theta^\oplus \right] \\
&= - \sum_{i=1}^n \frac{\mathbf{x}_i^T \mathbf{x}_i \boldsymbol{\Phi} \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus) [\mathbf{I} + \boldsymbol{\Phi} \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus) - \boldsymbol{\Phi}^T \mathbf{x}_i \boldsymbol{\Phi} \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus)]^2}{[1 + \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus)]^2} - 2\theta \\
&= - \sum_{i=1}^n \mathbf{x}_i^T \mathbf{x}_i \left[\frac{\exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus)}{1 + \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus)} - \left[\frac{\exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus)}{1 + \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus)} \right]^2 \right] - 2\theta \\
&= - \sum_{i=1}^n \mathbf{x}_i^T \mathbf{x}_i \left[\frac{\exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus)}{1 + \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus)} \right] \left[1 - \frac{\exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus)}{1 + \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^\oplus)} \right] - 2\theta \\
&= - \sum_{i=1}^n \mathbf{x}_i^T \mathbf{x}_i \pi_i(\mathbf{x}_i) [1 - \pi_i(\mathbf{x}_i)] - 2\theta \\
&= -\mathbf{X}^T \mathbf{W} \mathbf{X} - 2\theta \mathbf{I}
\end{aligned}$$

dimana : $\mathbf{W} = \text{diag}[\hat{\pi}_i(x_i)[1 - \hat{\pi}_i(x_i)]]$.

Estimasi parameter Regresi Logistik Ridge menggunakan metode *MLE* dengan iterasi Newton-Raphson yang akan digunakan untuk memaksimalkan fungsi obyektif pada persamaan (2.23). Kemudian diekspansikan di sekitar $\hat{\boldsymbol{\beta}}^\oplus$ menurut Deret Taylor dan didapatkan persamaan (2.24).

$$\frac{\partial \phi(\hat{\beta}^{\oplus})}{\partial \hat{\beta}^{\oplus}} = \frac{\partial \phi(\hat{\beta}^{\oplus})}{\partial \hat{\beta}^{\oplus}} \Big|_{\hat{\beta}^{\oplus} = \hat{\beta}_0^{\oplus}} + \frac{\partial^2 \phi(\hat{\beta}^{\oplus})}{\partial \hat{\beta}^{\oplus} \partial \hat{\beta}^{\oplus T}} \Big|_{\hat{\beta}^{\oplus} = \hat{\beta}_0^{\oplus}} (\hat{\beta}^{\oplus} - \hat{\beta}_0^{\oplus})$$

Jika $\frac{\partial \phi(\hat{\beta}^{\oplus})}{\partial \hat{\beta}^{\oplus}} = 0$, maka :

$$\frac{\partial \phi(\hat{\beta}^{\oplus})}{\partial \hat{\beta}^{\oplus}} \Big|_{\hat{\beta}^{\oplus} = \hat{\beta}_0^{\oplus}} + \frac{\partial^2 \phi(\hat{\beta}^{\oplus})}{\partial \hat{\beta}^{\oplus} \partial \hat{\beta}^{\oplus T}} \Big|_{\hat{\beta}^{\oplus} = \hat{\beta}_0^{\oplus}} (\hat{\beta}^{\oplus} - \hat{\beta}_0^{\oplus}) = 0$$

$$\frac{\partial \phi(\hat{\beta}^{\oplus})}{\partial \hat{\beta}^{\oplus}} \Big|_{\hat{\beta}^{\oplus} = \hat{\beta}_0^{\oplus}} = - \frac{\partial^2 \phi(\hat{\beta}^{\oplus})}{\partial \hat{\beta}^{\oplus} \partial \hat{\beta}^{\oplus T}} \Big|_{\hat{\beta}^{\oplus} = \hat{\beta}_0^{\oplus}} (\hat{\beta}^{\oplus} - \hat{\beta}_0^{\oplus})$$

Hasil penurunan disubstitusikan ke dalam persamaan (2.24) menghasilkan persamaan (2.25) (Sunnyoto, 2009).

$$\mathbf{X}^T (\mathbf{y} - \mathbf{X}\beta) = 2\theta \hat{\beta}^{\oplus} = \mathbf{X}^T (\mathbf{W}\mathbf{X} - \mathbf{I}) \hat{\beta}^{\oplus} (\hat{\beta}^{\oplus} - \hat{\beta}_0^{\oplus})$$

$$\mathbf{X}^T (\mathbf{y} - \mathbf{X}\hat{\beta}^{\oplus}) = \mathbf{X}^T (\mathbf{W}\mathbf{X} - \mathbf{I}) (\hat{\beta}^{\oplus} - \hat{\beta}_0^{\oplus})$$

$$\mathbf{X}^T (\mathbf{y} - \mathbf{X}\hat{\beta}^{\oplus}) = \mathbf{X}^T (\mathbf{W}\mathbf{X} - \mathbf{I}) (\hat{\beta}^{\oplus} - \hat{\beta}_0^{\oplus})$$

$$(\mathbf{X}^T \mathbf{W}\mathbf{X} - 2\theta \mathbf{I}) \hat{\beta}^{\oplus} = \mathbf{X}^T (\mathbf{W}\mathbf{X} - \mathbf{I}) (\hat{\beta}^{\oplus} - \hat{\beta}_0^{\oplus}) + \mathbf{X}^T (\mathbf{y} - \mathbf{X}\hat{\beta}_0^{\oplus})$$

$$(\mathbf{X}^T \mathbf{W}\mathbf{X} - 2\theta \mathbf{I}) \hat{\beta}^{\oplus} = \mathbf{X}^T (\mathbf{W}\mathbf{X} - \mathbf{I}) (\hat{\beta}^{\oplus} - \hat{\beta}_0^{\oplus}) + \mathbf{X}^T (\mathbf{y} - \mathbf{X}\hat{\beta}_0^{\oplus})$$

$$(\mathbf{X}^T \mathbf{W}\mathbf{X} - 2\theta \mathbf{I}) \hat{\beta}^{\oplus} = \mathbf{X}^T (\mathbf{W}\mathbf{X} - \mathbf{I}) (\hat{\beta}^{\oplus} - \hat{\beta}_0^{\oplus}) + \mathbf{X}^T (\mathbf{y} - \mathbf{X}\hat{\beta}_0^{\oplus})$$

$$\hat{\beta}^{\oplus} = (\mathbf{X}^T \mathbf{W}\mathbf{X} + 2\theta \mathbf{I})^{-1} [(\mathbf{X}^T \mathbf{W}\mathbf{X}) \hat{\beta}_0^{\oplus} + \mathbf{X}^T \mathbf{W}\mathbf{W}^{-1} (\mathbf{y} - \boldsymbol{\pi})]$$

$$\hat{\beta}^{\oplus} = (\mathbf{X}^T \mathbf{W}\mathbf{X} + 2\theta \mathbf{I})^{-1} \mathbf{X}^T \mathbf{W} [\mathbf{X} \hat{\beta}_0^{\oplus} + \mathbf{W}^{-1} (\mathbf{y} - \boldsymbol{\pi})]$$

$$\hat{\beta}^{\oplus} = (\mathbf{X}^T \mathbf{W}\mathbf{X} + 2\theta \mathbf{I})^{-1} \mathbf{X}^T \mathbf{W}\mathbf{z}$$

dimana θ^{\oplus} adalah ridge parameter untuk Regresi Logistik Ridge yang merupakan bilangan positif kecil, sehingga estimasi parameter Regresi Logistik Ridge dituliskan pada persamaan (2.26) (Vago & Kemény, 2006).

$$\hat{\beta}^{\oplus} = (\mathbf{X}^T \mathbf{W}\mathbf{X} + 2\theta^{\oplus} \mathbf{I})^{-1} \mathbf{X}^T \mathbf{W}\mathbf{z} \quad (2.26)$$

Dengan \mathbf{z} merupakan vektor berukuran $n \times 1$, dimana :

$$z_i = \text{Logit}[\hat{\pi}_i(x_i)] + \frac{y_i - \hat{\pi}_i(x_i)}{\hat{\pi}_i(x_i)[1 - \hat{\pi}_i(x_i)]}$$

Dengan menambahkan *ridge parameter* untuk Regresi Logistik Ridge pada elemen diagonal matriks kovarian dari Regresi Logistik pada persamaan (2.13), maka variansi Regresi Logistik Ridge dapat dihitung dengan formula pada persamaan (2.27) (Vago & Kemény, 2006).

$$\text{Var}(\hat{\beta}^{\oplus}) = (\mathbf{X}^T \text{diag}[\hat{\pi}_i(1 - \hat{\pi}_i)]\mathbf{X} + \theta^{\oplus}\mathbf{I})^{-1} \quad (2.27)$$

2.5 Pengujian Signifikansi Parameter

Pengujian signifikansi parameter secara serentak dilakukan dengan menggunakan *Likelihood Ratio Test*, dimana hipotesis yang digunakan adalah :

$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$ (model tidak berpengaruh signifikan)

$H_1 : \text{Minimal ada satu } \beta_j \neq 0, j=1, 2, \dots, p$ (model berpengaruh signifikan)

Statistik uji yang digunakan dalam *Likelihood Ratio Test* ditampilkan pada persamaan (2.28).

$$G = -2 \log \left(\frac{L_0}{L_1} \right) = -2[\ln(L_0) - \ln(L_1)] \quad (2.28)$$

dimana :

L_0 = Nilai fungsi likelihood untuk model yang semua parameter-nya sama dengan nol.

L_1 = Nilai fungsi likelihood untuk model lengkap.

Nilai $-2[\ln(L_0) - \ln(L_1)]$ mengikuti distribusi *chi-square* dengan $df = p$. Apabila $-2[\ln(L_0) - \ln(L_1)] \geq \chi^2_{(p, \alpha)}$, maka diputuskan untuk menolak H_0 yang artinya model berpengaruh signifikan, sehingga uji signifikansi parameter dapat dilanjutkan dengan pengujian secara parsial untuk mengetahui variabel-variabel prediktor yang signifikan terhadap peluang sukses (Agresti, 2007).

Pengujian signifikansi parameter secara parsial dilakukan dengan menggunakan *Wald test*. Pengujian ini dilakukan untuk mengetahui variabel prediktor yang signifikan atau layak untuk masuk ke dalam model. Hipotesis yang digunakan adalah :

$H_0 : \beta_j = 0$ (variabel ke- j tidak berpengaruh signifikan)

$H_1 : \beta_j \neq 0, j=1, 2, \dots, p$ (variabel ke- j berpengaruh signifikan)

Statistik uji yang digunakan dalam *Wald test* ditampilkan pada persamaan (2.29) (Hosmer & Lemeshow, 2000).

$$t = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (2.29)$$

Dimana : $SE(\hat{\beta}_j) = \sqrt{Var(\hat{\beta}_j)}$

Diputuskan untuk menolak H_0 apabila nilai $t > t_{(\alpha,p)}$ atau $p\text{-value} < \alpha$, yang artinya variabel ke- j berpengaruh signifikan terhadap pembentukan model.

2.6 Indeks Pembangunan Manusia (IPM)

Pembangunan manusia merupakan suatu proses untuk memperbanyak pilihan-pilihan yang dimiliki oleh manusia. Dimana pilihan-pilihan tersebut terdiri dari tiga komponen dasar, yaitu untuk berumur panjang dan sehat, untuk memiliki ilmu pengetahuan dan yang ketiga untuk mempunyai akses terhadap sumber daya yang dibutuhkan sehingga dapat menjalani kehidupan yang layak (HDR, 2014). Sehingga Indeks Pembangunan Manusia (IPM) diartikan sebagai suatu indeks komposit yang digunakan untuk mengukur capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup manusia. Ketiga komponen dasar yang digunakan sebagai ukuran kualitas hidup tersebut diukur dengan menggunakan suatu indeks untuk masing-masing komponen, yaitu indeks harapan hidup, indeks pendidikan dan indeks standar hidup layak (BPS, 2014).

Ketiga dimensi dasar pembangunan IPM memiliki pengertian yang sangat luas, hal tersebut dikarenakan masing-masing dimensi memiliki keterkaitan dengan banyak faktor yang mempengaruhinya. Pada pengukuran dimensi berumur panjang dan sehat, digunakan Angka Harapan Hidup sebagai ukurannya. Angka Harapan Hidup (AHH) merupakan rata-rata perkiraan jumlah tahun yang dapat ditempuh oleh seseorang selama hidup atau dapat dikatakan sebagai rata-rata perkiraan usia seseorang

(BPS, 2008). Nilai AHH dihitung berdasarkan dua komponen, yaitu Anak Lahir Hidup (ALH) dan Anak Masih Hidup (AMH).

Dimensi pengetahuan penduduk suatu wilayah, diukur berdasarkan indeks pendidikan yang didapatkan berdasarkan dua indikator yaitu Rata-Rata Lama Sekolah (*Mean Years Schooling*) dan Angka Melek Huruf. Rata-rata lama sekolah merupakan jumlah tahun yang digunakan oleh penduduk usia 15 tahun ke atas dalam menjalani pendidikan formal. Sedangkan angka melek huruf merupakan prosentase penduduk usia 15 tahun ke atas yang dapat membaca dan menulis huruf latin atau huruf lainnya (BPS, 2008).

Indeks standar hidup layak adalah indeks untuk mengukur dimensi kehidupan yang layak. Standar hidup layak menggambarkan tingkat kesejahteraan yang dinikmati oleh penduduk sebagai dampak semakin membaiknya kondisi ekonomi suatu wilayah. United Nations Development Programme (UNDP) mengukur standar hidup layak menggunakan Produk Domestik Regional Bruto (PDRB) riil yang disesuaikan, sedangkan Badan Pusat Statistik (BPS) menggunakan rata-rata pengeluaran per kapita riil yang disesuaikan. Perhitungan daya beli dilakukan berdasarkan 27 komoditas kebutuhan pokok masyarakat.

2.7 Penelitian Sebelumnya

Penelitian-penelitian sebelumnya mengenai analisis Indeks Pembangunan Manusia (IPM) menjadi rujukan dalam penentuan metode yang akan digunakan, serta dapat menjadi acuan dalam penentuan variabel yang akan digunakan sebagai faktor-faktor yang mempengaruhi IPM. Penelitian sebelumnya mengenai IPM dengan menggunakan respon IPM dalam skala pengukuran rasio dilakukan oleh Yunitasari (2007) yang melakukan penelitian terhadap IPM Jawa Timur dimana didapatkan faktor-faktor yang mempengaruhinya adalah PDRB perkapita, tingkat kemiskinan, pengeluaran pemerintah untuk sektor pendidikan dan kesehatan. Penelitian Melliana (2013) terhadap IPM Jawa Timur dengan metode regresi panel didapatkan tujuh variabel yang berpengaruh

signifikan terhadap IPM antara lain variabel rasio siswa terhadap guru, angka partisipasi sekolah, jumlah sarana kesehatan, persentase RT dengan akses air bersih, kepadatan penduduk, tingkat partisipasi angkatan kerja, dan PDRB perkapita. Sedangkan dengan metode yang sama. Di tahun yang sama, Astri (2013) melakukan penelitian terhadap IPM Indonesia (Nasional) dengan menggunakan regresi panel dimana faktor yang digunakan adalah pengeluaran pemerintah daerah pada sektor pendidikan dan kesehatan.

Penelitian terhadap IPM dengan menggunakan respon IPM dalam skala pengukuran bersifat kategorik dilakukan oleh Yurviyany (2007) dimana penelitian terhadap IPM Jawa Timur dilakukan menggunakan metode regresi logistik dengan faktor-faktor yang mempengaruhi IPM adalah angka harapan hidup, angka kematian bayi, angka buta huruf penduduk umur 10 tahun ke atas, rata-rata lama sekolah dan pengeluaran perkapita. Nur (2010) juga melakukan penelitian terhadap IPM Jawa Timur, Jawa Tengah, Jawa Barat dan Sumatera Utara menggunakan regresi logistik ordinal dengan memperhatikan unsur tingkatan pada kategori IPM. Dengan memperhatikan efek spasial, Pradita (2011) menggunakan metode *Geographically Weighted Logistic Regression* (GWLR) dalam penelitiannya mengenai IPM Jawa Timur, dimana didapatkan faktor-faktor yang berpengaruh terhadap IPM adalah prosentase penduduk yang tinggal di daerah perkotaan, rata-rata pendapatan perkapita, rasio ketergantungan penduduk, prosentase penduduk miskin dan jumlah sarana kesehatan. Tetapi metode GWLR tidak cukup baik digunakan pada data yang jumlahnya kecil. Karena Provinsi Jawa Timur hanya memiliki 38 kabupaten/kota, sehingga analisis GWLR tidak cukup baik digunakan pada analisis IPM Jawa Timur. Penggunaan metode regresi logistik ridge dilakukan oleh Sunyoto (2009) pada analisis keberhasilan siswa SMA Negeri 1 Kediri diterima di Perguruan Tinggi Negeri (PTN).

Berdasarkan penelitian-penelitian sebelumnya tersebut, dapat diketahui bahwa secara umum IPM dibentuk berdasarkan

tiga sektor yang digunakan untuk mengukur kualitas hidup manusia, yaitu kesehatan, pendidikan dan ekonomi, sehingga dalam penelitian ini akan digunakan faktor-faktor di masing-masing sektor tersebut untuk memodelkan Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur.

(Halaman ini sengaja dikosongkan)

BAB III METODOLOGI PENELITIAN

Pada bab ini dibahas mengenai sumber data, variabel penelitian yang digunakan dalam analisis serta metode analisis data.

3.1 Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder mengenai Indeks Pembangunan Manusia (IPM) beserta faktor-faktornya untuk setiap Kabupaten/Kota di Jawa Timur yang diambil secara langsung di Badan Pusat Statistik (BPS) Jawa Timur.

3.2 Variabel Penelitian dan Definisi Operasional

Pada penelitian ini variabel respon yang digunakan adalah Indeks Pembangunan Manusia (IPM) untuk setiap Kabupaten/Kota di Jawa Timur Tahun 2012 yang telah dikategorikan ke dalam empat kategori, yaitu kategori IPM rendah ($IPM \leq 50$), IPM menengah bawah ($50 < IPM \leq 66$), IPM menengah atas ($66 < IPM \leq 90$) dan IPM tinggi ($IPM > 90$), sedangkan variabel prediktor berdasarkan masing-masing sektor ditampilkan pada Tabel 3.1.

Tabel 3.1 Variabel Penelitian Indikator IPM

Sektor	Simbol	Variabel	Skala
-	Y	Indeks Pembangunan Manusia Jawa Timur	Nominal
Kesehatan	X ₁	Angka kematian bayi	Rasio
	X ₂	Prosentase keluhan kesehatan	Rasio
	X ₃	Jumlah sarana kesehatan	Rasio
Pendidikan	X ₄	Angka buta huruf (Usia 10 tahun ke atas)	Rasio

Tabel 3.1 Variabel Penelitian Indikator IPM (Lanjutan)

Sektor	Simbol	Variabel	Skala
Pendidikan	X ₅	Angka partisipasi sekolah (SMA)	Rasio
	X ₆	Rasio guru-siswa (SMA)	Rasio
	X ₇	Rasio sekolah-murid (SMA)	Rasio
Ekonomi	X ₈	Prosentase penduduk miskin	Rasio
	X ₉	PDRB perkapita	Rasio
	X ₁₀	Pertumbuhan ekonomi	Rasio
	X ₁₁	Prosentase penduduk usia 15 tahun ke atas yang bekerja	Rasio
	X ₁₂	Tingkat pengangguran terbuka	Rasio
	X ₁₃	Tingkat partisipasi angkatan kerja	Rasio

Berdasarkan variabel penelitian yang ditampilkan pada Tabel 3.1 tersebut, maka definisi operasional untuk masing-masing variabel adalah.

1. Indeks Pembangunan Manusia (IPM) [Y]
Indikator komposit yang menggabungkan tiga aspek penting, yaitu peningkatan kualitas fisik (kesehatan), intelektualitas (pendidikan) dan kemampuan ekonomi (daya beli) seluruh komponen masyarakat dalam kurun waktu tertentu.
2. Angka Kematian Bayi (AKB) [X₁]
Banyaknya bayi yang meninggal sebelum mencapai usia 1 tahun per 1000 kelahiran hidup pada tahun yang sama. Nilai normatif AKB yang kurang dari 40 sangat sulit diupayakan penurunannya, antara 40-70 tergolong sedang namun sulit untuk diturunkan, dan lebih besar dari 70 tergolong mudah untuk diturunkan.
3. Prosentase keluhan kesehatan [X₂]
Jumlah penduduk yang mempunyai keluhan kesehatan dibandingkan dengan jumlah total penduduk untuk masing-masing Kabupaten/Kota selama kurun waktu tertentu.

4. Jumlah sarana kesehatan [X_3]
Jumlah sarana kesehatan di masing-masing Kabupaten/Kota yang meliputi rumah sakit, puskesmas, dsb.
5. Angka buta huruf (Usia 10 tahun ke atas) [X_4]
Ketidakmampuan untuk membaca dan menulis penduduk usia 10 tahun ke atas.
6. Angka partisipasi sekolah (SMA) [X_5]
Proporsi dari semua anak yang masih sekolah pada suatu kelompok umur tertentu terhadap penduduk dengan kelompok umur yang sesuai. Digunakan kelompok umur SMA (16-18 tahun) karena sejak tahun 2012 sudah mulai dirintis wajib belajar 12 tahun.
7. Rasio guru-siswa [X_6]
Perbandingan antara jumlah guru dan siswa dalam suatu jenjang pendidikan tertentu.
8. Rasio sekolah-murid [X_7]
Perbandingan antara jumlah sekolah dan murid dalam suatu jenjang pendidikan tertentu.
9. Prosentase penduduk miskin [X_8]
Jumlah penduduk miskin dibandingkan dengan jumlah total penduduk untuk masing-masing Kabupaten/Kota selama kurun waktu tertentu.
10. PDRB perkapita [X_9]
Produk Domestik Regional Bruto (PDRB) dibagi dengan jumlah penduduk.
11. Pertumbuhan Ekonomi [X_{10}]
Besarnya prosentase kenaikan PDRB pada tahun berjalan terhadap PDRB pada tahun sebelumnya
12. Prosentase penduduk usia 15 tahun ke atas yang bekerja [X_{11}]
Jumlah penduduk usia 15 tahun ke atas yang bekerja dibandingkan dengan jumlah total penduduk untuk masing-masing Kabupaten/Kota selama kurun waktu tertentu.

13. Tingkat Pengangguran Terbuka (TPT) [X_{12}]
Persentase jumlah orang yang sedang mencari pekerjaan terhadap jumlah angkatan kerja untuk masing-masing Kabupaten/Kota selama kurun waktu tertentu.
14. Tingkat partisipasi angkatan kerja (TPAK) [X_{13}]
Persentase jumlah angkatan kerja terhadap penduduk usia kerja untuk masing-masing Kabupaten/Kota selama kurun waktu tertentu.

3.3 Metode Analisis Data

Langkah-langkah dalam analisis data dalam penelitian ini adalah sebagai berikut.

1. Pengumpulan data mengenai Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur tahun 2012 beserta faktor-faktor yang mempengaruhinya untuk setiap sektor pembentuk IPM yang terdiri dari empat sektor, yaitu sektor kesehatan, pendidikan dan ekonomi.
2. Melakukan analisis deskriptif pada Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur dan faktor-faktor yang mempengaruhinya, dimana analisis dilakukan pada setiap sektor, yaitu sektor kesehatan pendidikan dan ekonomi.
3. Pemeriksaan kasus multikolinieritas dengan menggunakan nilai *Variance Inflation Factor* (*VIF*) untuk masing-masing variabel prediktor (X).
4. Pemilihan nilai *ridge parameter* (θ^{\oplus}):
 - a. Menghitung nilai *eigen vektor* dan *eigen value* dari matriks $\mathbf{X}^T\mathbf{X}$, sehingga $\mathbf{X}^T\mathbf{X}=\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$. Dimana \mathbf{Q} adalah *eigen vektor* dari $\mathbf{X}^T\mathbf{X}$ dan $\mathbf{\Lambda} = \text{diag}(\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_{p-1} \geq \lambda_p)$ yang merupakan matriks diagonal dengan diagonalnya adalah *eigen value* dari matriks $\mathbf{X}^T\mathbf{X}$.
 - b. Menghitung *Principal Component* dari \mathbf{X} , yaitu $\mathbf{Z} = \mathbf{X}\mathbf{Q}$ dan koefisien dari *Principal Component Regression* (PCR), yaitu pada persamaan (3.1).

$$\hat{\boldsymbol{\alpha}} = \mathbf{\Lambda}^{-1}\mathbf{Z}^T\mathbf{Y} \quad (3.1)$$

- c. Untuk $r = 1, \dots, p$ dapat dihitung nilai *ridge parameter* pada persamaan (3.2).

$$\theta_r^{\oplus} = \frac{r}{\hat{\mathbf{a}}_r^T \hat{\mathbf{a}}_r} \quad (3.2)$$

Dimana $\hat{\mathbf{a}}_r$ merupakan vektor berukuran r dari sejumlah r koefisien pertama *Principal Component Logistic Regression* (PCLR). Koefisien PCLR dihitung dari *Principal Component*, dimana variabel prediktor sebagai kovariat dalam model Regresi Logistik.

- d. Menghitung derajat bebas dari model Regresi Logistik Ridge menggunakan θ_r^{\oplus} sebagai parameter penyusut dan kemudian dipilih jumlah dari *Principal Component* (r) untuk meminimumkan perbedaan antara r dan derajat bebas dari model Regresi Logistik Ridge yang telah didapatkan.
- e. Digunakan parameter penyusutan $\theta_{r^*}^{\oplus}$ dalam pembentukan model Regresi Logistik Ridge.
5. Pemodelan menggunakan metode Regresi Logistik Ridge.
 - a. Pengujian signifikansi parameter Regresi Logistik Ridge secara serentak.
 - b. Pengujian signifikansi parameter Regresi Logistik Ridge secara parsial.
 - c. Interpretasi hasil model lengkap Regresi Logistik Ridge.
 6. Pemilihan model terbaik Regresi Logistik Ridge dengan menggunakan metode *backward elimination*.
 - a. Pengujian signifikansi parameter Regresi Logistik Ridge secara serentak.
 - b. Pengujian signifikansi parameter Regresi Logistik Ridge secara parsial.
 - c. Interpretasi hasil pemodelan Regresi Logistik Ridge terbaik yang didapatkan dari hasil *backward elimination*.
 - d. Pemetaan hasil prediksi dari model terbaik Regresi Logistik Ridge.

(Halaman ini sengaja dikosongkan)

BAB IV

ANALISIS DAN PEMBAHASAN

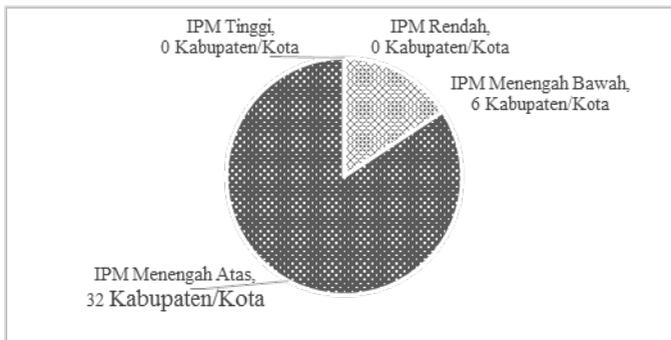
Pada bab ini dibahas mengenai deskripsi kabupaten/kota di Provinsi Jawa Timur menurut Indeks Pembangunan Manusia (IPM) untuk masing-masing sektor. Selanjutnya dilakukan identifikasi adanya kasus multikolinieritas pada faktor-faktor yang mempengaruhi IPM, serta penyelesaian kasus multikolinieritas dengan menggunakan metode Regresi Logistik Ridge. Kabupaten/ kota di Provinsi Jawa Timur meliputi :

- | | |
|-----------------|----------------------|
| 1. Pacitan | 20. Magetan |
| 2. Ponorogo | 21. Ngawi |
| 3. Trenggalek | 22. Bojonegoro |
| 4. Tulung Agung | 23. Tuban |
| 5. Blitar | 24. Lamongan |
| 6. Kediri | 25. Gresik |
| 7. Malang | 26. Bangkalan |
| 8. Lumajang | 27. Sampang |
| 9. Jember | 28. Pamekasan |
| 10. Banyuwangi | 29. Sumenep |
| 11. Bondowoso | 30. Kota Kediri |
| 12. Situbondo | 31. Kota Blitar |
| 13. Probolinggo | 32. Kota Malang |
| 14. Pasuruan | 33. Kota Probolinggo |
| 15. Sidoarjo | 34. Kota Pasuruan |
| 16. Mojokerto | 35. Kota Mojokerto |
| 17. Jombang | 36. Kota Madiun |
| 18. Nganjuk | 37. Kota Surabaya |
| 19. Madiun | 38. Kota Batu |

4.1 Statistika Deskriptif Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Indeks Pembangunan Manusia (IPM)

Analisa deskriptif pada kabupaten/kota di Provinsi Jawa Timur berdasarkan Indeks Pembangunan Manusia (IPM)

bertujuan untuk melihat karakteristik kabupaten/kota di Jawa Timur menurut masing-masing sektor, yaitu sektor kesehatan, pendidikan, dan ekonomi. Provinsi Jawa Timur memiliki 38 kabupaten/kota yang terdiri dari 29 kabupaten dan 9 kota. Pada tahun 2012, IPM Jawa Timur sebesar 72,83 tergolong dalam kategori menengah atas. Jumlah kabupaten/kota yang tergolong dalam setiap kategori ditampilkan pada diagram lingkaran pada Gambar 4.1.



Gambar 4.1 Jumlah Kabupaten/Kota untuk Setiap Kategori IPM

Kabupaten/kota di Provinsi Jawa Timur yang memiliki IPM dengan kategori menengah atas dan menengah bawah masing-masing berjumlah 6 dan 32 kabupaten/kota. Sedangkan pada kategori yang lain, tidak terdapat kabupaten/kota yang tergolong dalam IPM rendah maupun tinggi. Kota Malang memiliki nilai IPM tertinggi di antara kabupaten/kota yang lain dengan nilai IPM sebesar 78,43 pada kategori IPM menengah atas, sedangkan IPM terendah di Provinsi Jawa Timur adalah Kabupaten Sampang dengan nilai IPM sebesar 61,67 pada kategori IPM menengah bawah.

4.1.1 Deskriptif Kabupaten/Kota di Provinsi Jawa Timur Menurut Sektor Kesehatan

Pada Tabel 4.1 ditampilkan deskripsi mengenai kabupaten/kota di Provinsi Jawa Timur menurut sektor kesehatan.

Tabel 4.1 Deskripsi Kabupaten/Kota di Provinsi Jawa Timur
Menurut Sektor Kesehatan

Variabel	Y	Rataan	Varians	Min	Maks
Angka Kematian Bayi	0	56,29	13,16	53,93	63,51
	1	29,64	72,57	19,50	51,07
Keluhan Kesehatan (%)	0	30,00	41,69	17,82	37,18
	1	27,33	34,32	16,89	45,02
Jumlah Sarana Kesehatan	0	1641	602105	1050	3184
	1	1414	675816	186	3249

Keterangan kategori IPM :

0 = IPM menengah bawah

1 = IPM menengah atas

Tabel 4.1 menunjukkan bahwa rata-rata angka kematian bayi pada kategori IPM menengah bawah dan menengah atas masing-masing adalah 56,29 dan 29,64. Angka kematian bayi antar kabupaten/kota di Provinsi Jawa Timur masih memiliki perbedaan yang cukup jauh, dimana angka kematian bayi terendah adalah Kota Blitar, yaitu sebesar 19,5, sedangkan angka kematian bayi tertinggi adalah Kabupaten Probolinggo, yaitu sebesar 63,51.

Rata-rata prosentase keluhan kesehatan di Provinsi Jawa Timur pada kategori IPM menengah bawah dan menengah atas masing-masing sebesar 30% dan 27,33%. Kabupaten Lumajang memiliki prosentase keluhan kesehatan terendah, yaitu 16,89%, sedangkan Kota Mojokerto memiliki prosentase keluhan kesehatan tertinggi, yaitu 45,02%. Dalam hal penanganan keluhan kesehatan, Kota Mojokerto harus mendapat perhatian khusus agar dapat mengatasi tingginya angka keluhan kesehatan.

Jumlah sarana kesehatan di Provinsi Jawa Timur memiliki rata-rata pada kategori IPM menengah bawah dan menengah atas masing-masing adalah 1641 dan 1414 dengan varians yang sangat besar pada kedua kategori IPM. Hal tersebut dapat dikarenakan perbedaan luas wilayah dari setiap kabupaten/kota di Provinsi Jawa Timur. Kota Mojokerto memiliki jumlah sarana kesehatan

terendah dengan jumlah 186 sarana, sebaliknya Kabupaten Malang memiliki jumlah sarana kesehatan tertinggi, yaitu 3249 sarana kesehatan.

4.1.2 Deskriptif Kabupaten/Kota di Provinsi Jawa Timur Menurut Sektor Pendidikan

Tabel 4.2 menunjukkan deskripsi mengenai kabupaten/kota di Provinsi Jawa Timur menurut sektor pendidikan.

Tabel 4.2 Deskripsi Kabupaten/Kota di Provinsi Jawa Timur Menurut Sektor Pendidikan

Variabel	Y	Rataan	Varians	Min	Maks
Angka Buta Huruf	0	18,630	9,650	15,35	24,01
	1	7,654	21,336	1,50	19,52
Angka Partisipasi Sekolah (SMA)	0	47,95	58,74	38,61	61,21
	1	67,80	86,15	47,84	82,07
Rasio Guru-Siswa (SMA)	0	863,2	26689,0	658	1047
	1	1011,7	176776,3	553	2245
Rasio Sekolah-Murid (SMA)	0	22,50	50,70	12	31
	1	41,16	1468,01	11	151

Keterangan kategori IPM :

0 = IPM menengah bawah

1 = IPM menengah atas

Pada Tabel 4.2 dapat dilihat bahwa rata-rata angka buta huruf relatif kecil pada kategori IPM menengah atas, yaitu sebesar 7,654, tetapi pada kategori IPM menengah bawah sebesar 18,63. Angka buta huruf terendah sebesar 1,5 yaitu Kota Malang, sedangkan tertinggi adalah Kabupaten Sampang yaitu 24,01. Sehingga Kabupaten Sampang harus mendapatkan perhatian khusus mengenai kemampuan membaca dan menulis untuk dapat menurunkan angka buta huruf.

Rata-rata angka partisipasi sekolah (SMA) kabupaten/kota di Provinsi Jawa Timur pada kategori IPM menengah bawah dan menengah atas masing-masing adalah 47,95 dan 67,8 dimana

angka partisipasi sekolah terendah adalah Kabupaten Sampang, yaitu 38,61, sedangkan angka partisipasi sekolah tertinggi adalah Kota Mojokerto, yaitu 82,07.

Rasio guru-siswa (SMA) pada kategori IPM menengah bawah dan menengah atas masing-masing adalah 863,2 dan 1011,7. Kemudian pada rasio sekolah-murid (SMA) kabupaten/kota di Provinsi Jawa Timur untuk kategori IPM menengah bawah dan menengah atas masing-masing adalah 22,5 dan 41,16. Rasio guru-siswa tertinggi adalah Kota Madiun, yaitu 2245, sedangkan yang terendah adalah Kabupaten Kediri, yaitu sebesar 553. Kemudian pada rasio sekolah-murid tertinggi adalah Kota Mojokerto dengan nilai 151, sedangkan yang terendah adalah Kabupaten Malang yang sebesar 11. Dalam kasus ini Kabupaten Kediri dan Kabupaten Malang harus diperhatikan dalam hal ketersediaan guru/pengajar dan jumlah sekolah untuk memenuhi kebutuhan pendidikan bagi siswa/murid.

4.1.3 Deskriptif Kabupaten/Kota di Provinsi Jawa Timur Menurut Sektor Ekonomi

Pada Tabel 4.3 ditampilkan deskripsi terkait kabupaten/kota di Provinsi Jawa Timur menurut sektor ekonomi.

Tabel 4.3 Deskripsi Kabupaten/Kota di Provinsi Jawa Timur Menurut Sektor Ekonomi

Variabel	Y	Rataan	Varians	Min	Maks
Rumah Tangga Miskin (%)	0	23,80	100,96	9,94	36,03
	1	13,85	102,24	4,30	52,96
PDRB Perkapita (juta)	0	13,42	11,89	8,69	17,70
	1	30,97	2528,76	8,32	290,79
Pertumbuhan Ekonomi (%)	0	6,6120	0,1320	6,19	7,27
	1	6,9881	0,2917	5,82	8,26
Penduduk Usia > 15 Tahun Yang Bekerja (%)	0	68,710	25,660	61,630	75,330
	1	66,651	25,014	58,332	78,801
Tingkat Pengangguran Terbuka (%)	0	3,342	1,739	1,78	5,32
	1	4,397	2,980	1,16	7,85

Tabel 4.3 Deskripsi Kabupaten/Kota di Provinsi Jawa Timur
Menurut Sektor Ekonomi (Lanjutan)

Variabel	Y	Rataan	Varians	Min	Maks
Tingkat Partisipasi	0	71,050	20,320	64,13	76,69
Angkatan Kerja (%)	1	69,678	17,951	62,53	79,73

Keterangan kategori IPM :

0 = IPM menengah bawah

1 = IPM menengah atas

Rata-rata prosentase rumah tangga miskin pada kategori IPM menengah bawah dan menengah atas masing-masing adalah 23,8 dan 13,85 dimana prosentase rumah tangga terendah adalah Kabupaten Sidoarjo yang sebesar 4,3% dan Kota Probolinggo memiliki prosentase rumah tangga miskin tertinggi, yaitu sebesar 52,96%.

PDRB perkapita pada kategori IPM menengah bawah dan menengah atas masing-masing sebesar 13,42 juta dan 30,97 juta. Kemudian pada pertumbuhan ekonomi kabupaten/kota di Provinsi Jawa Timur pada kategori IPM menengah bawah dan menengah atas masing-masing adalah 6,612% dan 6,9881%. Terjadi ketimpangan pada PDRB perkapita kabupaten/kota di Provinsi Jawa Timur, dimana Kota Kediri memiliki PDRB Perkapita tertinggi, yaitu 290,79 juta, sedangkan PDRB perkapita terendah adalah Kabupaten Pacitan yang hanya sebesar 8,32 juta. Pada pertumbuhan ekonomi, menunjukkan bahwa pertumbuhan ekonomi yang terjadi pada kabupaten/kota di Provinsi Jawa Timur memiliki varians yang cukup kecil. Selisih antara pertumbuhan ekonomi tertinggi dan terendah juga cukup kecil, yaitu sebesar 2,44%. Pertumbuhan ekonomi tertinggi adalah Kota Batu yang mencapai angka 8,26%, sedangkan pertumbuhan ekonomi terendah adalah Kabupaten Bojonegoro, yaitu sebesar 5,82%.

Rata-rata prosentase penduduk usia 15 tahun ke atas yang bekerja pada kategori IPM menengah bawah dan menengah atas masing-masing adalah 68,71% dan 66,651%. Kota Madiun memiliki prosentase penduduk bekerja yang terendah dan

Kabupaten Pacitan memiliki prosentase penduduk bekerja yang tertinggi dengan prosentase masing-masing sebesar 58,332% dan 78,801%.

Tingkat pengangguran terbuka memiliki rata-rata yang relatif kecil, pada kategori IPM menengah bawah dan menengah atas masing-masing sebesar 3,342% dan 4,397%. Kabupaten Pacitan merupakan kabupaten/kota dengan tingkat pengangguran terkecil dan Kota Kediri merupakan kabupaten/kota dengan tingkat pengangguran tertinggi yang memiliki prosentase masing-masing sebesar 1,16% dan 7,85%.

Rata-rata tingkat partisipasi angkatan kerja kabupaten/kota di Provinsi Jawa Timur pada kategori IPM menengah bawah dan menengah atas masing-masing adalah 71,05% dan 69,678%. Kabupaten/kota dengan tingkat partisipasi angkatan kerja terendah adalah Kota Madiun, sedangkan tingkat partisipasi angkatan kerja tertinggi adalah Kabupaten Pacitan yang memiliki prosentase masing-masing sebesar 62,53% dan 79,73%.

4.1.4 Identifikasi Kasus Multikolinieritas

Pada Tabel 4.4 ditampilkan nilai *Variance Inflation Factors (VIF)* pada masing-masing variabel prediktor terhadap variabel prediktor yang lain.

Tabel 4.4 Nilai *Variance Inflation Factors (VIF)*

Sektor	Variabel	VIF
Kesehatan	Angka Kematian Bayi [X ₁]	4,122
	Keluhan Kesehatan [X ₂]	1,436
	Jumlah Sarana Kesehatan [X ₃]	4,935
Pendidikan	Angka Buta Huruf [X ₄]	6,205
	Angka Partisipasi Sekolah (SMA) [X ₅]	2,591
	Rasio Guru-Siswa (SMA) [X ₆]	4,539
	Rasio Sekolah-Murid (SMA) [X ₇]	9,436
Ekonomi	Rumah Tangga Miskin [X ₈]	1,766

PDRB Perkapita [X_9]	1,841
--------------------------	-------

Tabel 4.4 Nilai *Variance Inflation Factors (VIF)* (Lanjutan)

Sektor	Variabel	VIF
	Pertumbuhan Ekonomi [X_{10}]	2,652
Ekonomi	Penduduk Usia > 15 Tahun Yang Bekerja [X_{11}]	5243,225
	Tingkat Pengangguran Terbuka [X_{12}]	294,646
	Tingkat Partisipasi Angkatan Kerja [X_{13}]	3530,604

Tabel 4.4 menunjukkan bahwa diindikasikan terjadi kasus multikolinieritas pada faktor-faktor di sektor pendidikan dan ekonomi. Pada sektor pendidikan diindikasikan terdapat kasus multikolinieritas pada variabel angka buta huruf (X_4) dan rasio sekolah-murid (X_7) karena memiliki nilai *VIF* lebih dari 5, sedangkan kasus multikolinieritas yang serius terdapat pada sektor ekonomi, khususnya pada variabel prosentase penduduk usia 15 tahun ke atas yang bekerja (X_{11}), tingkat pengangguran terbuka (X_{12}) dan tingkat partisipasi angkatan kerja (X_{13}) yang memiliki nilai *VIF* lebih dari 10.

4.2 Pemodelan Regresi Logistik Ridge

Pemodelan menggunakan metode Regresi Logistik Ridge bertujuan untuk mengatasi adanya kasus multikolinieritas yang terjadi pada faktor-faktor yang mempengaruhi IPM. Tabel 4.5 menunjukkan selisih antara derajat bebas (*df*) dan r untuk setiap jumlah *Principal Component* (r) yang akan digunakan dalam penentuan nilai *ridge parameter* untuk model lengkap.

Tabel 4.5 Pemilihan Nilai Ridge Paramater untuk Model Lengkap

r	<i>df</i> model	selisih <i>df</i> dan r	θ_r^{\oplus}
1	3,050508	2,050508	0,188904
2	2,411640	0,411640	0,369302
3	3,614199	0,614199	0,110148

4*	3,780055	0,219945	0,094133
5	4,212069	0,787931	0,062276

Tabel 4.5 Pemilihan Nilai Ridge Paramater untuk Model Lengkap (Lanjutan)

<i>r</i>	<i>df model</i>	<i>selisih df dan r</i>	θ_r^{\oplus}
6	4,344199	1,655801	0,054692
7	4,232640	2,767360	0,061039

Berdasarkan pada Tabel 4.5, diketahui bahwa jumlah dari *Principal Component* (*r*) yang memiliki selisih antara *r* dan derajat bebas (*df*) paling kecil adalah 4. Sehingga digunakan 4 *Principal Component* (*r*) dan didapatkan nilai *ridge parameter* (θ^{\oplus}) sebesar 0,094133. Dengan menggunakan *Software R*, hasil koefisien model lengkap Regresi Logistik Ridge ditunjukkan pada Tabel 4.6.

Tabel 4.6 Signifikansi Parameter Model Lengkap

Variabel	Koefisien	Koefisien (scaled)	Std. error	t-value (scaled)	P-value
Konstan	-0,1771	-	-	-	-
X ₁ *	-0,0442	-3,3999	0,8601	-3,9530	0,0001
X ₂	-0,0412	-1,4890	1,0355	-1,4380	0,1504
X ₃	0,0000	-0,0527	0,9210	-0,0572	0,9544
X ₄ *	-0,0754	-2,7364	0,7910	-3,4594	0,0005
X ₅ *	0,0371	2,6081	0,9219	2,8292	0,0047
X ₆	0,0003	0,6892	0,7965	0,8652	0,3869
X ₇	0,0016	0,3440	0,6406	0,5371	0,5912
X ₈	-0,0162	-1,0453	0,9847	-1,0615	0,2885
X ₉	0,0009	0,2585	0,6053	0,4270	0,6694
X ₁₀	0,1653	0,5336	0,8760	0,6091	0,5424
X ₁₁	0,0108	0,3284	0,7000	0,4692	0,6389

X_{12}	0,0340	0,3516	0,8021	0,4383	0,6612
X_{13}	0,0191	0,4937	0,7613	0,6485	0,5166

Dengan menggunakan nilai pada Tabel 4.6, dilakukan pengujian signifikansi parameter secara serentak dan parsial untuk mengetahui pengaruh dari masing-masing variabel prediktor dalam pembentukan model Regresi Logistik Ridge untuk IPM Provinsi Jawa Timur.

4.2.1 Pengujian Signifikansi Parameter Regresi Logistik Ridge Secara Serentak

Pengujian signifikansi parameter regresi secara serentak bertujuan untuk mengetahui pengaruh koefisien $\hat{\beta}^{\oplus}$ hasil estimasi parameter Regresi Logistik Ridge secara serentak. Pengujian ini menggunakan statistik uji *likelihood ratio test* dengan hipotesis sebagai berikut.

$$\text{Hipotesis : } H_0 : \hat{\beta}_1^{\oplus} = \hat{\beta}_2^{\oplus} = \dots = \hat{\beta}_p^{\oplus} = 0$$

$$H_1 : \text{Minimal ada satu } \hat{\beta}_j^{\oplus} \neq 0, j=1, 2, \dots, p$$

Nilai statistik uji *likelihood ratio test* ditunjukkan pada persamaan (4.1).

$$\begin{aligned} G &= -2 \log \left(\frac{L_0}{L_1} \right) = -2 [\ln(L_0) - \ln(L_1)] \\ &= -2 [-28,7974 - (-6,0986)] \\ &= -2 [-22,6988] \\ &= 45,39758 \end{aligned} \tag{4.1}$$

Daerah penolakan adalah nilai yang didapatkan dari tabel $\chi^2_{(p;\alpha)}$ dengan $p = 13$ dan $\alpha = 0,05$, sehingga didapatkan nilai $\chi^2_{(13;0,10)} = 22,362$. Didapatkan keputusan menolak H_0 karena nilai statistik uji $[G = 45,398] \geq [\chi^2_{(13;0,05)} = 22,362]$, sehingga dapat

disimpulkan bahwa paling sedikit terdapat satu variabel prediktor yang berpengaruh signifikan terhadap variabel respon IPM untuk masing-masing kabupaten/kota di Provinsi Jawa Timur.

4.2.2 Pengujian Signifikansi Parameter Regresi Logistik Ridge Secara Parsial

Pengujian signifikansi parameter regresi secara parsial bertujuan untuk mengetahui pengaruh masing-masing koefisien $\hat{\beta}^{\oplus}$ hasil estimasi parameter Regresi Logistik Ridge secara parsial. Pengujian ini menggunakan statistik uji *Wald* dengan hipotesis sebagai berikut.

$$\text{Hipotesis : } H_0 : \hat{\beta}_j^{\oplus} = 0$$

$$H_1 : \hat{\beta}_j^{\oplus} \neq 0, j=1, 2, \dots, p$$

Dengan melihat nilai statistik uji *p-value* pada Tabel 4.6, dapat dilihat bahwa terdapat tiga variabel prediktor yang memiliki *p-value* $< [\alpha = 0,05]$. Sehingga dapat disimpulkan bahwa variabel angka kematian bayi (X_1), angka buta huruf (X_4) dan angka partisipasi sekolah (X_5) berpengaruh signifikan terhadap Indeks Pembangunan Manusia (IPM) pada taraf $\alpha = 0,05$.

4.2.3 Model Lengkap Regresi Logistik Ridge IPM Provinsi Jawa Timur

Model lengkap IPM kabupaten/kota Provinsi Jawa Timur menggunakan Regresi Logistik Ridge ditampilkan pada persamaan (4.2).

$$\pi_i(x_i) = \frac{\exp(\mathbf{x}\hat{\boldsymbol{\beta}}^{\oplus})}{1 + \exp(\mathbf{x}\hat{\boldsymbol{\beta}}^{\oplus})} \quad (4.2)$$

Dimana :

$$\begin{aligned} \mathbf{x}\hat{\boldsymbol{\beta}}^{\oplus} = & -0,1771 - 0,0442x_1 - 0,0412x_2 + 0,0000x_3 - 0,0754x_4 + \\ & 0,0371x_5 + 0,0003x_6 + 0,0016x_7 - 0,0162x_8 + \\ & 0,0009x_9 + 0,1653x_{10} + 0,0108x_{11} + 0,0340x_{12} + \\ & 0,0191x_{13} \end{aligned}$$

4.2.4 Ketepatan Klasifikasi Model Lengkap Regresi Logistik Ridge

Perhitungan ketepatan klasifikasi digunakan untuk melihat seberapa baik model Regresi Logistik Ridge dalam mengelompokkan suatu observasi/pengamatan ke dalam kelompok tertentu. Didapatkan hasil ketepatan klasifikasi yang ditampilkan pada Tabel 4.7.

Tabel 4.7 Ketepatan Klasifikasi Hasil Model Lengkap Regresi Logistik Ridge

		Logistik Ridge		
		Prediksi	0	1
Observasi	0	5	1	6
	1	0	32	32
Total		5	33	0,9737

Tabel 4.7 menunjukkan bahwa hasil akurasi (ketepatan klasifikasi) dari pemodelan Regresi Logistik Ridge dengan melibatkan 13 variabel prediktor sebesar 97,37%. Hal tersebut menunjukkan bahwa hasil pemodelan Regresi Logistik Ridge dengan melibatkan 13 variabel prediktor cukup efektif dalam mengelompokkan kabupaten/kota di Provinsi Jawa Timur ke dalam kategori IPM yang telah ditentukan oleh Badan Pusat Statistik.

4.2.5 Interpretasi Model Lengkap Regresi Logistik Ridge

Interpretasi dari hasil model lengkap Regresi Logistik Ridge bertujuan untuk mengetahui kecenderungan masing-masing variabel prediktor terhadap Indeks Pembangunan Manusia (IPM) untuk setiap kabupaten/kota di Provinsi Jawa Timur. Tetapi pada model lengkap, hanya dilakukan interpretasi terhadap variabel-variabel yang signifikan, yaitu angka kematian bayi (X_1), angka buta huruf (X_4) dan angka partisipasi sekolah (X_5). Digunakan nilai *odds ratio* yang ditampilkan pada Tabel 4.8.

Tabel 4.8 Odds Ratio Hasil Model Lengkap Regresi Logistik Ridge

Variabel	Koefisien	$OR = \text{Exp}(10 \times \hat{\beta}^{\oplus})$	1/OR
Konstan	-0,1771	-	-
Angka Kematian Bayi*	-0,0442	0,6425	1,5564
Prosentase Keluhan Kesehatan	-0,0412	0,6625	1,5094
Jumlah Sarana Kesehatan	0,0000	0,9999	1,0001
Angka Buta Huruf *	-0,0754	0,4707	2,1246
Angka Partisipasi Sekolah*	0,0371	1,4485	0,6904
Rasio Guru-Siswa	0,0003	1,0029	0,9971
Rasio Sekolah-Murid	0,0016	1,0159	0,9843
Prosentase Penduduk Miskin	-0,0162	0,8506	1,1756
PDRB Perkapita	0,0009	1,0092	0,9909
Pertumbuhan Ekonomi	0,1653	5,2226	0,1915
Prosentase Penduduk Usia 15 Tahun Ke Atas yang Bekerja	0,0108	1,1140	0,8976
Tingkat Pengangguran Terbuka	0,0340	1,4054	0,7116
Tingkat Partisipasi Angkatan Kerja	0,0191	1,2106	0,8260

Dengan melihat *odds ratio* pada Tabel 4.8, didapatkan bahwa setiap kenaikan 10% dari angka kematian bayi, maka suatu kabupaten/kota berpeluang untuk memiliki IPM dengan kategori menengah ke bawah 1,5 kali lipat dibandingkan IPM dengan kategori menengah ke atas. Kemudian pada variabel angka buta huruf, dimana setiap kenaikan 10% dari angka buta huruf, maka suatu kabupaten/kota akan memiliki peluang untuk memiliki IPM dengan kategori menengah ke bawah dua kali lipat dibandingkan

IPM dengan kategori menengah ke atas. Tetapi sebaliknya pada angka partisipasi sekolah, dimana setiap kenaikan 10% dari angka partisipasi sekolah, maka suatu kabupaten/kota akan berpeluang untuk memiliki IPM dengan kategori menengah ke atas 1,4 kali lipat dibandingkan IPM dengan kategori menengah bawah.

4.3 Pemilihan Model Regresi Logistik Ridge Terbaik

Pemilihan model Regresi Logistik Ridge terbaik bertujuan untuk mendapatkan model regresi dengan keseluruhan parameter telah signifikan terhadap variabel respon Indeks Pembangunan Manusia (IPM). Dalam tujuan prediksi dan pertimbangan efisiensi penggunaan variabel, seringkali digunakan model dengan variabel prediktor yang berpengaruh signifikan terhadap variabel respon. Sehingga dilakukan pemilihan model terbaik agar mendapatkan model yang lebih sederhana dengan ketepatan klasifikasi yang tinggi.

Pada *Software R* dilakukan *backward elimination* untuk memilih variabel yang akan dilibatkan dalam pemodelan Regresi Logistik Ridge. Didapatkan model terbaik dengan tiga variabel prediktor yang berpengaruh signifikan terhadap variabel respon, yaitu variabel angka kematian bayi (X_1), angka buta huruf (X_4) dan angka partisipasi sekolah (X_5). Pada Tabel 4.9 ditampilkan selisih antara derajat bebas (df) dan r untuk setiap jumlah *Principal Component* (r) yang akan digunakan dalam penentuan nilai *ridge parameter* untuk model terbaik.

Tabel 4.9 Pemilihan Nilai Ridge Paramater untuk Model Terbaik

r	df model	selisih df dan r	θ_r^{\otimes}
1	2,241553	1,241553	0,024942
2*	2,029940	0,029940	0,048791

Pada Tabel 4.9, didapatkan bahwa jumlah dari *principal component* (r) yang memiliki selisih antara r dan derajat bebas (df) paling kecil adalah 2. Perhitungan nilai ridge parameter pada model terbaik digunakan 2 *Principal Component* (r) dan

didapatkan nilai *ridge parameter* (θ^{\oplus}) sebesar 0,048791. Dengan menggunakan *Software R*, hasil koefisien model lengkap Regresi Logistik Ridge ditampilkan pada Tabel 4.10.

Tabel 4.10 Signifikansi Parameter Model Terbaik

Variabel	Koefisien	Koefisien (scaled)	Std. Error	t-value (scaled)	P-value
Konstan	2,62665	-	-	-	-
X ₁ *	-0,06379	-4,90166	1,350	-3,631	0,000283
X ₄ *	-0,10456	-3,79675	1,380	-2,751	0,005942
X ₅ *	0,04869	3,42683	1,433	2,391	0,016792

Hasil estimasi parameter pada Tabel 4.10 digunakan dalam pengujian signifikansi parameter secara serentak dan parsial untuk mengetahui pengaruh dari masing-masing variabel prediktor dalam pembentukan model Regresi Logistik Ridge terbaik untuk IPM Provinsi Jawa Timur.

4.3.1 Pengujian Signifikansi Parameter Regresi Logistik Ridge Secara Serentak

Pengaruh koefisien $\hat{\beta}^{\oplus}$ hasil estimasi parameter Regresi Logistik Ridge secara serentak dapat diketahui dari pengujian signifikansi parameter regresi secara serentak. Pengujian ini menggunakan statistik uji *likelihood ratio test* dengan hipotesis sebagai berikut.

$$\text{Hipotesis : } H_0 : \hat{\beta}_1^{\oplus} = \hat{\beta}_2^{\oplus} = \dots = \hat{\beta}_p^{\oplus} = 0$$

$$H_1 : \text{Minimal ada satu } \hat{\beta}_j^{\oplus} \neq 0, j=1, 2, \dots, p$$

Nilai statistik uji *likelihood ratio test* ditunjukkan pada persamaan (4.3).

$$\begin{aligned}
 G &= -2 \log \left(\frac{L_0}{L_1} \right) = -2 [\ln(L_0) - \ln(L_1)] \\
 &= -2 [-19,0643 - (-6,0137)] \\
 &= -2 [-13,05061] \\
 &= 26,10122
 \end{aligned} \tag{4.3}$$

Daerah penolakan yang digunakan adalah nilai yang didapatkan dari tabel $\chi^2_{(p;\alpha)}$ dengan $p = 3$ dan $\alpha = 0,05$, sehingga didapatkan nilai $\chi^2_{(3;0,05)} = 7,815$ yang artinya diputuskan menolak H_0 karena nilai statistik uji $[G = 26,10122] \geq [\chi^2_{(3;0,05)} = 7,815]$, sehingga dapat disimpulkan bahwa paling sedikit terdapat satu variabel prediktor yang berpengaruh signifikan terhadap variabel respon IPM untuk masing-masing kabupaten/kota di Provinsi Jawa Timur.

4.3.2 Pengujian Signifikansi Parameter Regresi Logistik Ridge Secara Parsial

Pada pengujian signifikansi parameter regresi secara parsial, ingin diketahui pengaruh masing-masing koefisien $\hat{\beta}^{\oplus}$ hasil estimasi parameter Regresi Logistik Ridge terhadap variabel respon secara parsial. Pengujian ini menggunakan statistik uji *Wald* dengan hipotesis sebagai berikut.

$$\text{Hipotesis : } H_0 : \hat{\beta}_j^{\oplus} = 0$$

$$H_1 : \hat{\beta}_j^{\oplus} \neq 0, j=1, 2, \dots, p$$

Dengan melihat nilai statistik uji *p-value* pada Tabel 4.9, dapat diketahui bahwa ketiga variabel prediktor memiliki *p-value* $< [\alpha = 0,05]$. Sehingga dapat disimpulkan bahwa variabel angka kematian bayi (X_1), angka buta huruf (X_4) dan angka partisipasi sekolah (X_5) berpengaruh signifikan terhadap Indeks Pembangunan Manusia (IPM) pada taraf $\alpha = 0,05$.

4.3.3 Model Terbaik Regresi Logistik Ridge IPM Provinsi Jawa Timur

Model terbaik Regresi Logistik Ridge IPM kabupaten/kota Provinsi Jawa Timur menggunakan metode *backward elimination* ditampilkan pada persamaan (4.4).

$$\pi_i(x_i) = \frac{\exp(\mathbf{x}\hat{\boldsymbol{\beta}}^{\oplus})}{1 + \exp(\mathbf{x}\hat{\boldsymbol{\beta}}^{\oplus})} \quad (4.4)$$

Dimana :

$$\mathbf{x}\hat{\boldsymbol{\beta}}^{\oplus} = 2,62665 - 0,06379x_1 - 0,10456x_4 + 0,04869x_5$$

4.3.4 Ketepatan Klasifikasi Model Terbaik Regresi Logistik Ridge

Perhitungan ketepatan klasifikasi digunakan untuk melihat seberapa baik model Regresi Logistik Ridge dalam mengelompokkan suatu observasi/pengamatan ke dalam kelompok tertentu. Didapatkan hasil ketepatan klasifikasi pada Tabel 4.11.

Tabel 4.11 Ketepatan Klasifikasi Hasil Model Terbaik Regresi Logistik Ridge

Prediksi \ Observasi	0	1	Total
0	5	1	6
1	0	32	32
Total	5	33	0,9737

Pada Tabel 4.11 ditampilkan hasil akurasi (ketepatan klasifikasi) dari model Regresi Logistik Ridge terbaik dengan melibatkan tiga variabel prediktor dimana akurasi (ketepatan klasifikasi) sebesar 97,37%. Hal tersebut menunjukkan bahwa hasil pemodelan Regresi Logistik Ridge dengan melibatkan tiga variabel prediktor lebih efektif dalam hal penggunaan variabel prediktor untuk mengelompokkan kabupaten/kota di Provinsi Jawa Timur ke dalam kategori IPM yang telah ditentukan oleh

Badan Pusat Statistik dibandingkan dengan menggunakan 13 variabel prediktor. Sehingga dapat disimpulkan model Indeks Pembangunan Manusia (IPM) dengan menggunakan model Regresi Logistik Ridge terbaik adalah dengan melibatkan tiga variabel prediktor, yaitu variabel angka kematian bayi (X_1), angka buta huruf (X_4) dan angka partisipasi sekolah (X_5).

4.3.5 Interpretasi Model Terbaik Regresi Logistik Ridge

Interpretasi dari hasil pemodelan Regresi Logistik Ridge terbaik bertujuan untuk mengetahui kecenderungan masing-masing variabel prediktor terhadap Indeks Pembangunan Manusia (IPM) untuk setiap kabupaten/kota di Provinsi Jawa Timur. Digunakan nilai *odds ratio* yang ditampilkan pada Tabel 4.12.

Tabel 4.12 *Odds Ratio* Hasil Model Terbaik Regresi Logistik Ridge

Variabel	Koefisien	$OR = \text{Exp}(10 \times \hat{\beta}^{\oplus})$	1/OR
Konstan	2,626650	-	-
Angka Kematian Bayi	-0,063790	0,52840	1,89250
Angka Buta Huruf	-0,104560	0,35148	2,84511
Angka Partisipasi Sekolah	0,048690	1,62726	0,61453

Pada Tabel 4.12 didapatkan bahwa setiap kenaikan 10% dari angka kematian bayi, maka suatu kabupaten/kota berpeluang untuk memiliki IPM dengan kategori menengah ke bawah dua kali lipat dibandingkan IPM dengan kategori menengah ke atas. Begitu juga halnya pada angka buta huruf, dimana setiap kenaikan 10% dari angka buta huruf, maka suatu kabupaten/kota akan memiliki peluang untuk memiliki IPM dengan kategori menengah ke bawah tiga kali lipat dibandingkan IPM dengan kategori menengah ke atas. Tetapi pada angka partisipasi sekolah berlaku sebaliknya, yaitu setiap kenaikan 10% dari angka partisipasi sekolah, maka suatu kabupaten/kota akan berpeluang untuk memiliki IPM dengan kategori menengah ke atas 1,6 kali lipat dibandingkan IPM dengan kategori menengah bawah.

4.3.6 Pemetaan Model Terbaik Regresi Logistik Ridge

Pemetaan hasil prediksi dari model Regresi Logistik Ridge terbaik dilakukan untuk dapat mengetahui bagaimana kondisi pembangunan manusia pada setiap kabupaten/kota di Provinsi Jawa Timur berdasarkan pembagian kategori Indeks Pembangunan Manusia (IPM). Hasil pemetaan dengan menggunakan *Software Arcview* ditampilkan pada Gambar 4.2 berikut.



Keterangan :

- IPM Menengah Bawah
- IPM Menengah Atas

Gambar 4.2 Pemetaan Hasil Prediksi Model Terbaik Regresi Logistik Ridge

Gambar 4.2 menunjukkan bahwa hasil prediksi dengan menggunakan model Regresi Logistik Ridge terbaik menghasilkan 5 kabupaten/kota tergolong dalam kategori IPM menengah bawah, yaitu Kabupaten Bangkalan, Sampang,

Probolinggo, Situbondo dan Jember. Sedangkan untuk 33 kabupaten/kota yang lain tergolong dalam IPM menengah atas. Dapat diketahui dari Tabel 4.10 mengenai ketepatan klasifikasi model Regresi Logistik Ridge terbaik bahwa terdapat satu kesalahan prediksi, yaitu pada Kabupaten Bondowoso. Pada observasi yang dilakukan, Kabupaten Bondowoso tergolong dalam IPM menengah bawah, sedangkan hasil prediksi menunjukkan bahwa Kabupaten Bondowoso tergolong IPM menengah atas. Nilai IPM Kabupaten Bondowoso sebesar 64,98 dimana nilai tersebut mendekati batas antara kategori IPM menengah bawah dan IPM menengah atas yang sebesar 66. Sehingga terjadi kesalahan prediksi pada Kabupaten Bondowoso.

BAB V

KESIMPULAN DAN SARAN

Pada bab ini dibahas mengenai kesimpulan dari hasil analisa data dan pembahasan yang telah dilakukan pada bab IV serta memuat saran yang diberikan penulis agar penelitian selanjutnya dapat lebih baik.

5.1 Kesimpulan

Berdasarkan pada hasil analisis dan pembahasan, maka dapat diambil kesimpulan sebagai berikut.

1. Terdapat kasus multikolinieritas pada faktor-faktor yang mempengaruhi indeks pembangunan manusia (IPM) untuk kabupaten/kota di Provinsi Jawa Timur, khususnya pada sektor pendidikan dan ekonomi, yaitu variabel angka buta huruf (X_4), rasio sekolah-murid (X_7), prosentase penduduk usia 15 tahun ke atas yang bekerja (X_{11}), tingkat pengangguran terbuka (X_{12}) dan tingkat partisipasi angkatan kerja (X_{13}) karena kelima variabel tersebut memiliki nilai $VIF > 5$.
2. Model terbaik dari Indeks Pembangunan Manusia (IPM) untuk kabupaten/kota di Provinsi Jawa Timur dengan menggunakan metode *backward elimination* dari Regresi Logistik Ridge adalah sebagai berikut.

$$\pi_i(x_i) = \frac{\exp(2,62665 - 0,06379x_1 - 0,10456x_4 + 0,04869x_5)}{1 + \exp(2,62665 - 0,06379x_1 - 0,10456x_4 + 0,04869x_5)}$$

dimana nilai *ridge parameter* (θ^\oplus) sebesar 0,04879089 dan terdapat tiga variabel prediktor yang dilibatkan dalam pemodelan, yaitu variabel angka kematian bayi (X_1), angka buta huruf (X_4) dan angka partisipasi sekolah (X_5).

3. Akurasi (ketepatan klasifikasi) model terbaik dari Indeks Pembangunan Manusia (IPM) untuk kabupaten/kota di Provinsi Jawa Timur dengan menggunakan metode

backward elimination dari Regresi Logistik Ridge sebesar 97,37%.

4. Prediksi dari model terbaik Regresi Logistik Ridge menghasilkan 5 kabupaten/kota tergolong dalam kategori IPM menengah bawah, yaitu Kabupaten Bangkalan, Sampang, Probolinggo, Situbondo dan Jember. Sedangkan 33 kabupaten/kota lainnya tergolong dalam IPM menengah atas. Kesalahan prediksi terjadi pada Kabupaten Bondowoso, dimana pada observasi, tergolong dalam IPM menengah bawah, sedangkan hasil prediksi tergolong IPM menengah atas.

5.2 Saran

Saran yang diberikan pada penelitian ini adalah :

1. Perlu adanya penelitian lebih lanjut mengenai metode lain terhadap kasus multikolinieritas yang terjadi pada faktor-faktor yang mempengaruhi indeks pembangunan manusia (IPM) untuk setiap kabupaten/kota di Provinsi Jawa Timur. Sehingga dapat dilakukan perbandingan dengan metode Regresi Logistik Ridge, sehingga dapat diketahui kelebihan dan kekurangan masing-masing metode.
2. Pemerintah Provinsi Jawa Timur harus memperhatikan proses peningkatan dan pemerataan pembangunan di setiap kabupaten/kota di Provinsi Jawa Timur untuk setiap sektor, baik dari segi kesehatan, pendidikan maupun ekonomi.

DAFTAR PUSTAKA

- Agresti, A. (2007). *An Introduction to Categorical Data Analysis Second Edition*. United State of America: A John Wiley & Sons, Inc.
- Astri, M. (2013). Pengaruh Pengeluaran Pemerintah Daerah Pada Sektor Pendidikan Dan Kesehatan Terhadap Indeks Pembangunan Manusia Di Indonesia. *Jurnal Pendidikan Ekonomi Dan Bisnis Vol. 1 No. 1*, 77-102.
- BPS, B. (2008). *Indeks Pembangunan Manusia 2006-2007*. Jakarta: Badan Pusat Statistik.
- BPS, B. (2010). *Penduduk Indonesia Menurut Provinsi 1971, 1980, 1990, 1995, 2000 dan 2010*. Diambil kembali dari Statistics Indonesia: http://bps.go.id/tab_sub/view.php?kat=1&tabel=1&daftar=1&id_subyek=12¬ab=1
- BPS, B. (2013). *Indeks Pembangunan Manusia dan Komponennya*. Diambil kembali dari Statistics Indonesia: http://bps.go.id/ipm.php?id_subyek=26¬ab=0
- BPS, B. (2014). *Konsep Indeks Pembangunan Manusia*. Diambil kembali dari Statistics Indonesia: http://bps.go.id/menutab.php?kat=1&tabel=1&id_subyek=26
- Cule, E., & De Iorio, M. (2013). Ridge Regression in Prediction Problem: Automatic Choice of the Ridge Parameter. *Genetic Epidemiology, Vol. 37, No. 7*, 704-714.
- Darwis, D. (2011). *Kependudukan Dalam Prespektif Pembangunan Ekonomi Guna Pembangunan Nasional*. Jawa Barat: BKKbN.
- Dickson. (2013, Desember 19). *10 Negara dengan Jumlah Penduduk Terbanyak di Dunia*. Diambil kembali dari Ilmu Pengetahuan Umum: <http://ilmupengetahuanumum.com/10-negara-dengan-jumlah-penduduk-populasi-terbanyak-di-dunia/>
- Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis : Third Edition*. Canada: John Wiley & Sons.

- HDR, H. (2014). *Sustaining Human Progress : Reducing Vulnerabilities and Building Resilience*. New York, United State of America: United Nations Development Programme (UNDP).
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression : Biased Estimation For Nonorthogonal Problems. *Technometrics*, Vol. 12, No. 1.
- Hosmer, D. W., & Lemeshow, S. (2000). *Applied Logistic Regression Second Edition*. New York, United State of America: John Wiley & Sons, Inc.
- Huang, H. (2014, Agustus 26). *Pengertian Statistik Deskriptif dan Statistik Inferensial*. Diambil kembali dari Globalstats Academic: <http://www.globalstatistik.com/pengertian-statistik-deskriptif-dan-statistik-inferensial/>
- Marcus, G. L., Wattimanela, H. J., & Lesnussa, Y. A. (2012). Analisis Regresi Komponen Utama Untuk Mengatasi Masalah Multikolinieritas Dalam Analisis Regresi Linier Berganda (Studi Kasus : Curah Hujan di Kota Ambon Tahun 2010). *Barekeng Vol. 6 No. 1*, 31-40.
- Masruroh, I. (2011). Pemilihan Model Regresi Linier Berganda Pada Kasus Multikolinieritas Dengan Metode Regresi Komponen Utama (Principal Component Regression) Dan Regresi Gulud (Ridge Regression). *FMIPA-Universitas Brawijaya*.
- Melliana, A. (2013). Analisis Statistika Faktor yang Mempengaruhi Indeks Pembangunan Manusia di Kabupaten/ Kota Provinsi Jawa Timur dengan Menggunakan Regresi Panel. *Jurnal Sains dan Seni POMITS Vol. 2*, No. 2.
- Nur, C. F. (2010). *Pemodelan IPM Provinsi Jawa Timur, Jawa Tengah, Jawa Barat dan Sumatera Utara dengan Metode Regresi Logistik Ordinal*. Surabaya: Jurusan Statistika ITS Surabaya.
- Pradita, N. P. (2011). *Geographically Weighted Logistic Regression dan Aplikasinya (Studi Kasus : Indeks*

- Pembangunan Manusia di Provinsi Jawa Timur*). Surabaya: Jurusan Statistika ITS Surabaya.
- Ryan, T. P. (1997). *Modern Regression Methods*. New York: John Wiley & Sons.
- Sunyoto. (2009). *Regresi Logistik Ridge : Pada Keberhasilan Siswa SMA Negeri 1 Kediri Diterima Di Perguruan Tinggi Negeri*. Surabaya: Institut Teknologi Sepuluh Nopember.
- Vago, H., & Kemény, S. (2006). Logistic Ridge For Clinical Data Analysis (A Case Study). *Applied Ecology And Environmental Research* 4(2), 171-179.
- Walpole, R. E. (1995). *Pengantar Statistika Edisi ke-3*. Jakarta: Gramedia Pustaka Utama.
- Yan, X., & Su, X. G. (2009). *Linear Regression Analysis : Theory and Computing*. Singapore: World Scientific.
- Yunitasari, M. (2007). *Analisis Hubungan Antara Pertumbuhan Ekonomi Dengan Pembangunan Manusia Jawa Timur*.
- Yurviany. (2007). *Analisis Regresi Logistik pada Data Indeks Pembangunan Manusia di Propinsi Jawa Timur*. Surabaya: Jurusan Statistika ITS Surabaya.

(Halaman ini sengaja dikosongkan)

DAFTAR LAMPIRAN

	Halaman
Lampiran A1	Data Penelitian Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur Tahun 201257
Lampiran A2	Data Penelitian Faktor-Faktor di Sektor Kesehatan dan Pendidikan58
Lampiran A3	Data Penelitian Faktor-Faktor di Sektor Ekonomi60
Lampiran B1	Estimasi Parameter Model Lengkap62
Lampiran B2	Estimasi Parameter <i>Backward Elimination</i> Ke-1.....63
Lampiran B3	Estimasi Parameter <i>Backward Elimination</i> Ke-2.....64
Lampiran B4	Estimasi Parameter <i>Backward Elimination</i> Ke-3.....65
Lampiran B5	Estimasi Parameter <i>Backward Elimination</i> Ke-4.....66
Lampiran B6	Estimasi Parameter <i>Backward Elimination</i> Ke-5.....67
Lampiran B7	Estimasi Parameter <i>Backward Elimination</i> Ke-6.....68
Lampiran B8	Estimasi Parameter <i>Backward Elimination</i> Ke-7.....69
Lampiran B9	Estimasi Parameter <i>Backward Elimination</i> Ke-8.....70
Lampiran B10	Estimasi Parameter <i>Backward Elimination</i> Ke-9.....71
Lampiran B11	Estimasi Parameter <i>Backward Elimination</i> Ke-10 (Model Terbaik)72

(Halaman ini sengaja dikosongkan)

LAMPIRAN

Lampiran A1 Data Penelitian Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur Tahun 2012

Kabupaten/Kota	IPM 2012	Kode IPM 2012
Pacitan	72.88	1
Ponorogo	71.91	1
Trenggalek	74.09	1
Tulung Agung	74.45	1
Blitar	74.43	1
Kediri	72.72	1
Malang	71.94	1
Lumajang	69.00	1
Jember	65.99	0
Banyuwangi	70.53	1
Bondowoso	64.98	0
Situbondo	65.06	0
Probolinggo	64.35	0
Pasuruan	69.17	1
Sidoarjo	77.36	1
Mojokerto	74.42	1
Jombang	73.86	1
Nganjuk	71.96	1
Madiun	70.88	1
Magetan	73.85	1
Ngawi	70.20	1
Bojonegoro	67.74	1
Tuban	69.18	1
Lamongan	71.05	1
Gresik	75.97	1
Bangkalan	65.69	0
Sampang	61.67	0
Pamekasan	66.51	1
Sumenep	66.41	1
Kota Kediri	77.20	1
Kota Blitar	78.31	1

Lanjutan **Lampiran A1**

Kabupaten/Kota	IPM 2012	Kode IPM 2012
Kota Malang	78.43	1
Kota Probolinggo	75.44	1
Kota Pasuruan	74.33	1
Kota Mojokerto	78.01	1
Kota Madiun	77.50	1
Kota Surabaya	78.33	1
Kota Batu	75.42	1
Jawa Timur	72.83	1

Lampiran A2 Data Penelitian Faktor-Faktor di Sektor Kesehatan dan Pendidikan

Kabupaten/Kota	X₁	X₂	X₃	X₄	X₅	X₆	X₇
Pacitan	22.63	32.94	970	11.12	61.05	913	41
Ponorogo	27.03	26.36	1376	8.28	73.77	936	26
Trenggalek	21.41	24.47	1088	6.62	64.14	636	29
Tulung Agung	22.02	30.49	1407	4.90	53.72	791	22
Blitar	23.71	29.63	1723	7.55	63.83	634	24
Kediri	27.79	28.01	2121	7.24	65.86	553	15
Malang	30.46	25.94	3249	8.37	49.48	815	11
Lumajang	37.89	16.89	1421	15.44	57.57	875	24
Jember	56.33	30.28	3184	15.35	50.03	873	12
Banyuwangi	34.81	32.55	2596	8.57	58.98	772	17
Bondowoso	53.93	37.18	1321	17.55	61.21	658	31
Situbondo	54.94	31.18	1050	20.44	47.59	881	30
Probolinggo	63.51	31.07	1586	17.46	47.42	1029	22
Pasuruan	51.07	24.18	2193	7.95	47.84	812	13
Sidoarjo	24.27	20.27	2026	2.27	78.73	585	15
Mojokerto	25.54	33.92	1610	5.54	64.17	857	18

Lanjutan Lampiran A2

Kabupaten/Kota	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
Jombang	27.56	34.76	1930	5.44	68.5	1058	18
Nganjuk	31.12	24.20	1617	8.48	68.85	724	23
Madiun	31.18	27.56	1095	11.17	79.14	635	31
Magetan	22.85	20.67	1086	8.07	77.78	932	40
Ngawi	27.06	25.18	1470	13.48	80.35	685	28
Bojonegoro	38.67	19.79	1888	13.84	51.41	1076	20
Tuban	34.41	22.37	1859	14.77	67.64	678	20
Lamongan	33.72	28.69	2320	10.22	67.65	1452	25
Gresik	23.27	20.92	1663	3.44	70.81	880	20
Bangkalan	54.56	17.82	1423	16.98	42.86	691	19
Sampang	54.48	32.44	1284	24.01	38.61	1047	21
Pamekasan	50.69	24.94	1137	13.84	62.07	1136	25
Sumenep	48.42	23.81	1827	19.52	65.71	1168	23
Kota Kediri	24.85	29.20	381	2.89	73.36	1778	66
Kota Blitar	19.5	27.19	188	2.85	70.52	1882	128
Kota Malang	24.74	39.70	763	1.50	74.15	933	24
Kota Probolinggo	25.12	28.80	249	7.20	68.51	1639	108
Kota Pasuruan	39.45	31.06	313	2.58	80.08	881	86
Kota Mojokerto	21.88	45.02	186	2.94	82.07	1696	151
Kota Madiun	23.24	25.26	302	2.94	80.15	2245	115
Kota Surabaya	23.18	22.89	2980	1.95	69.68	690	12
Kota Batu	28.87	26.93	217	3.95	71.97	1028	99
Jawa Timur	28.31	26.93	55099	9.65	61.68	855	22

Lampiran A3 Data Penelitian Faktor-Faktor di Sektor Ekonomi

Kabupaten/Kota	X₈	X₉	X₁₀	X₁₁	X₁₂	X₁₃
Pacitan	12.26	8.32	6.77	78.80	1.16	79.73
Ponorogo	12.65	11.31	6.67	71.01	3.26	73.41
Trenggalek	19.03	11.54	6.72	74.89	3.14	77.32
Tulung Agung	15.21	21.45	6.99	69.91	3.18	72.21
Blitar	6.16	14.65	6.44	71.51	2.86	73.61
Kediri	10.46	13.93	6.99	66.95	4.16	69.86
Malang	8.73	17.44	7.56	67.60	3.79	70.26
Lumajang	9.46	18.17	6.47	64.34	4.70	67.51
Jember	18.34	14.32	7.27	61.63	3.91	64.13
Banyuwangi	10.6	21.03	7.29	70.87	3.40	73.37
Bondowoso	21.28	12.45	6.47	67.89	3.75	70.53
Situbondo	9.94	16.57	6.62	67.08	3.31	69.37
Probolinggo	34.96	17.7	6.67	73.82	1.98	75.31
Pasuruan	25.03	13.8	7.29	65.88	6.43	70.40
Sidoarjo	4.3	39.13	7.23	63.22	5.21	66.70
Mojokerto	4.44	24.06	7.29	67.73	3.42	70.13
Jombang	14.02	15.33	6.99	62.08	6.69	66.54
Nganjuk	22.23	13.79	6.72	64.66	4.22	67.52
Madiun	9.96	13.4	6.58	67.08	4.16	69.99
Magetan	7.98	15.34	6.51	69.24	3.86	72.02
Ngawi	9.69	11.42	6.67	63.50	3.05	65.50
Bojonegoro	15.62	18.23	5.82	66.98	3.51	69.41
Tuban	10.77	21.81	6.19	63.72	4.25	66.55
Lamongan	6.3	13.24	7.22	64.89	4.98	68.29
Gresik	14.51	41.88	7.43	59.23	6.72	63.49
Bangkalan	36.03	10.76	6.45	66.51	5.32	70.25
Sampang	22.24	8.69	6.19	75.33	1.78	76.69

Lanjutan **Lampiran A3**

Kabupaten/Kota	X₈	X₉	X₁₀	X₁₁	X₁₂	X₁₃
Pamekasan	30.06	8.43	6.43	75.70	2.30	77.48
Sumenep	35.25	13.55	6.49	75.93	1.19	76.84
Kota Kediri	14.89	290.79	7.67	61.68	7.85	66.93
Kota Blitar	11.05	19.81	6.84	62.27	3.55	64.56
Kota Malang	7.37	48.94	7.71	59.32	7.68	64.26
Kota Probolinggo	52.96	26.94	6.96	64.18	5.12	67.65
Kota Pasuruan	10.92	17.91	6.59	65.02	4.34	67.97
Kota Mojokerto	4.43	30.09	7.19	65.84	7.32	71.04
Kota Madiun	9.36	37.05	7.88	58.33	6.71	62.53
Kota Surabaya	5.03	97.1	7.76	62.77	5.07	66.53
Kota Batu	12.51	21.25	8.26	67.70	3.41	70.09
Jawa Timur	14.42	26.44	7.27	66.75	4.12	69.62

Keterangan Variabel Penelitian :

Sektor	Simbol	Variabel
Kesehatan	X ₁	Angka kematian bayi
	X ₂	Prosentase keluhan kesehatan
	X ₃	Jumlah sarana kesehatan
Pendidikan	X ₄	Angka buta huruf (Usia 10 tahun ke atas)
	X ₅	Angka partisipasi sekolah (SMA)
	X ₆	Rasio guru-siswa (SMA)
	X ₇	Rasio sekolah-murid (SMA)
Ekonomi	X ₈	Prosentase penduduk miskin
	X ₉	PDRB perkapita
	X ₁₀	Pertumbuhan ekonomi
	X ₁₁	Prosentase penduduk usia 15 tahun ke atas yang bekerja
	X ₁₂	Tingkat pengangguran terbuka
	X ₁₃	Tingkat partisipasi angkatan kerja

Lampiran B1 Estimasi Parameter Model Lengkap

Pemilihan Nilai Ridge Parameter Model Lengkap

<i>R</i>	<i>df model</i>	<i>selisih df dan r</i>	θ_r^{\oplus}
1	3.0505	2.0505	0.1889
2	2.4116	0.4116	0.3693
3	3.6142	0.6142	0.1101
4*	3.7801	0.2199	0.0941
5	4.2121	0.7879	0.0623
6	4.3442	1.6558	0.0547
7	4.2326	2.7674	0.0610

Estimasi Parameter Model Lengkap

```
> data = read.table('D:/Tugas Akhir/Data/data.txt', header = TRUE)
> mod <- logisticRidge(Y ~ ., data = as.data.frame(data), nPCs=4)
> summary(mod)
```

Call:

```
logisticRidge(formula = Y ~ ., data = as.data.frame(data), nPCs=4)
Coefficients:
```

<i>Variables</i>	<i>Coef</i>	<i>Coef (scaled)</i>	<i>Std. error</i>	<i>t-value (scaled)</i>	<i>P-value</i>
(Intercept)	-0.1771	NA	NA	NA	NA
X1***	-0.0442	-3.4000	0.8601	-3.9530	0.0001
X2	-0.0412	-1.4890	1.0350	-1.4380	0.1504
X3	0.0000	-0.0527	0.9210	-0.0570	0.9544
X4***	-0.0754	-2.7360	0.7910	-3.4590	0.0005
X5**	0.0371	2.6080	0.9219	2.8290	0.0047
X6	0.0003	0.6892	0.7965	0.8650	0.3869
X7	0.0016	0.3440	0.6406	0.5370	0.5912
X8	-0.0162	-1.0450	0.9847	-1.0620	0.2885
X9	0.0009	0.2585	0.6053	0.4270	0.6694
X10	0.1653	0.5336	0.8760	0.6090	0.5424
X11	0.0108	0.3284	0.7000	0.4690	0.6389
X12	0.0340	0.3516	0.8021	0.4380	0.6612
X13	0.0191	0.4937	0.7613	0.6490	0.5166

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge paramter: 0.09413297

Degrees of freedom: model 3.78 , variance 2.844

Lampiran B2 Estimasi Parameter *Backward Elimination* Ke-1

Pemilihan Nilai Ridge Parameter *Backward Elimination* Ke-1

r	df model	selisih df dan r	θ_r^{\oplus}
1	2.9484	1.9484	0.1801
2	2.3631	0.3631	0.3476
3	3.4361	0.4361	0.1084
4*	3.7653	0.2347	0.0770
5	4.1007	0.8993	0.0538
6	4.0923	1.9077	0.0543
7	4.0554	2.9446	0.0565

Estimasi Parameter *Backward Elimination* Ke-1

```
> data = read.table('D:/Tugas Akhir/Data/backward1.txt', header =
  TRUE)
> mod <- logisticRidge(Y ~ ., data = as.data.frame (data), nPCs=4)
> summary(mod)
Call:
logisticRidge(formula = Y ~ ., data = as.data.frame(data), nPCs=4)
Coefficients:
```

<i>Variables</i>	<i>Coef</i>	<i>Coef (scaled)</i>	<i>Std. error</i>	<i>t-value (scaled)</i>	<i>P-value</i>
(Intercept)	-0.5671	NA	NA	NA	NA
X1***	-0.0485	-3.7237	0.9668	-3.8520	0.0001
X2	-0.0472	-1.7066	1.1718	-1.4560	0.1453
X4***	-0.0826	-2.9981	0.8931	-3.3570	0.0008
X5**	0.0406	2.8583	1.0551	2.7090	0.0067
X6	0.0003	0.7973	0.9087	0.8770	0.3803
X7	0.0016	0.3465	0.8024	0.4320	0.6659
X8	-0.0173	-1.1153	1.1122	-1.0030	0.3160
X9	0.0010	0.2696	0.6714	0.4020	0.6880
X10	0.1710	0.5520	1.0302	0.5360	0.5921
X11	0.0141	0.4284	0.7760	0.5520	0.5809
X12	0.0346	0.3575	0.9120	0.3920	0.6950
X13	0.0238	0.6154	0.8572	0.7180	0.4728

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Ridge paramter: 0.07696627
Degrees of freedom: model 3.765 , variance 3.06
```

Lampiran B3 Estimasi Parameter *Backward Elimination* Ke-2

Pemilihan Nilai Ridge Parameter *Backward Elimination* Ke-2

r	df model	selisih df dan r	θ_r^{\oplus}
1	3.1329	2.1329	0.1307
2	2.6796	0.6796	0.2167
3*	3.2206	0.2206	0.1187
4	3.6114	0.3886	0.0771
5	4.1111	0.8889	0.0431
6	4.1287	1.8713	0.0422

Estimasi Parameter *Backward Elimination* Ke-2

```
> data = read.table('D:/Tugas Akhir/Data/backward2.txt', header =
  TRUE)
> mod <- logisticRidge(Y ~ ., data = as.data.frame (data), nPCs=3)
> summary(mod)
Call:
logisticRidge(formula = Y ~ ., data = as.data.frame(data), nPCs=3)
Coefficients:
```

<i>Variables</i>	<i>Coef</i>	<i>Coef (scaled)</i>	<i>Std. error</i>	<i>t-value (scaled)</i>	<i>P-value</i>
(Intercept)	0.5147	NA	NA	NA	NA
X1***	-0.0397	-3.0521	0.7531	-4.0530	0.0001
X2	-0.0352	-1.2718	0.9009	-1.4120	0.1581
X4***	-0.0684	-2.4834	0.7101	-3.4970	0.0005
X5**	0.0332	2.3392	0.8009	2.9210	0.0035
X6	0.0002	0.5792	0.6929	0.8360	0.4033
X7	0.0016	0.3529	0.6253	0.5640	0.5725
X8	-0.0149	-0.9637	0.8577	-1.1240	0.2612
X9	0.0009	0.2580	0.5615	0.4590	0.6460
X10	0.1623	0.5240	0.7814	0.6710	0.5025
X11	0.0058	0.1768	0.6887	0.2570	0.7974
X13	0.0129	0.3324	0.7058	0.4710	0.6377

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge paramter: 0.1186818

Degrees of freedom: model 3.221 , variance 2.326

Lampiran B4 Estimasi Parameter *Backward Elimination* Ke-3

Pemilihan Nilai Ridge Parameter *Backward Elimination* Ke-3

r	df model	selisih df dan r	θ_r^{\oplus}
1	3.2981	2.2981	0.0919
2	3.2310	1.2310	0.0992
3*	2.9102	0.0898	0.1431
4	3.3514	0.6486	0.0864
5	3.9500	1.0500	0.0428
6	4.0407	1.9593	0.0382

Estimasi Parameter *Backward Elimination* Ke-3

```
> data = read.table('D:/Tugas Akhir/Data/backward3.txt', header =
  TRUE)
> mod <- logisticRidge(Y ~ ., data = as.data.frame (data), nPCs=3)
> summary(mod)
Call:
logisticRidge(formula = Y ~ ., data = as.data.frame(data), nPCs=3)
Coefficients:
```

<i>Variables</i>	<i>Coef</i>	<i>Coef (scaled)</i>	<i>Std. error</i>	<i>t-value (scaled)</i>	<i>P-value</i>
(Intercept)	0.9409	NA	NA	NA	NA
X1***	-0.0361	-2.7728	0.6732	-4.1190	0.0000
X2	-0.0304	-1.0997	0.7967	-1.3800	0.1675
X4***	-0.0621	-2.2537	0.6360	-3.5440	0.0004
X5**	0.0302	2.1290	0.7075	3.0090	0.0026
X6	0.0002	0.5068	0.6124	0.8280	0.4079
X7	0.0016	0.3450	0.5573	0.6190	0.5359
X8	-0.0138	-0.8945	0.7637	-1.1710	0.2415
X9	0.0009	0.2441	0.5080	0.4800	0.6309
X10	0.1532	0.4945	0.6944	0.7120	0.4764
X13	0.0107	0.2754	0.7789	0.3540	0.7237

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge paramter: 0.1431263

Degrees of freedom: model 2.91 , variance 1.977

Lampiran B5 Estimasi Parameter *Backward Elimination* Ke-4

Pemilihan Nilai Ridge Parameter *Backward Elimination* Ke-4

r	df model	selisih df dan r	θ_r^{\oplus}
1	3.1275	2.1275	0.0762
2	3.0740	1.0740	0.0821
3*	2.7839	0.2161	0.1223
4	3.3415	0.6585	0.0565
5	3.2978	1.7022	0.0601
6	3.2073	2.7927	0.0683

Estimasi Parameter *Backward Elimination* Ke-4

```
> data = read.table('D:/Tugas Akhir/Data/backward4.txt', header =
  TRUE)
> mod <- logisticRidge(Y ~ ., data = as.data.frame (data), nPCs=3)
> summary(mod)
Call:
logisticRidge(formula = Y ~ ., data = as.data.frame(data), nPCs=3)
Coefficients:
```

Variables	Coef	Coef (scaled)	Std. error	t-value (scaled)	P-value
(Intercept)	1.8486	NA	NA	NA	NA
X1***	-0.0390	-2.9996	0.7406	-4.0500	0.0001
X2	-0.0339	-1.2245	0.8869	-1.3810	0.1674
X4***	-0.0665	-2.4149	0.7070	-3.4150	0.0006
X5**	0.0326	2.2926	0.7856	2.9180	0.0035
X6	0.0002	0.5795	0.6766	0.8560	0.3917
X7	0.0016	0.3462	0.6092	0.5680	0.5699
X8	-0.0141	-0.9091	0.8582	-1.0590	0.2895
X9	0.0009	0.2412	0.5485	0.4400	0.6602
X10	0.1489	0.4805	0.7842	0.6130	0.5400

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge paramter: 0.1223143

Degrees of freedom: model 2.784 , variance 2.014

Lampiran B6 Estimasi Parameter *Backward Elimination* Ke-5

Pemilihan Nilai Ridge Parameter *Backward Elimination* Ke-5

r	df model	selisih df dan r	θ_r^{\oplus}
1	3.1492	2.1492	0.0669
2	3.0218	1.0218	0.0796
3*	3.4979	0.4979	0.0412
4	3.2890	0.7110	0.0552
5	3.2384	1.7616	0.0592

Estimasi Parameter *Backward Elimination* Ke-5

```
> data = read.table('D:/Tugas Akhir/Data/backward5.txt', header =
  TRUE)
> mod <- logisticRidge(Y ~ ., data = as.data.frame (data), nPCs=3)
> summary(mod)
Call:
logisticRidge(formula = Y ~ ., data = as.data.frame(data), nPCs=3)
Coefficients:
```

<i>Variables</i>	<i>Coef</i>	<i>Coef (scaled)</i>	<i>Std. Error</i>	<i>t-value (scaled)</i>	<i>P-value</i>
(Intercept)	3.0230	NA	NA	NA	NA
X1***	-0.0626	-4.8077	1.3710	-3.5070	0.0005
X2	-0.0671	-2.4271	1.6661	-1.4570	0.1452
X4**	-0.1054	-3.8268	1.3674	-2.7990	0.0051
X5*	0.0527	3.7068	1.5384	2.4100	0.0160
X6	0.0006	1.3520	1.3565	0.9970	0.3189
X7	0.0013	0.2855	1.1072	0.2580	0.7965
X8	-0.0165	-1.0671	1.6058	-0.6650	0.5064
X10	0.1438	0.4641	1.5407	0.3010	0.7632

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge paramter: 0.04116176

Degrees of freedom: model 3.498 , variance 3.736

Lampiran B7 Estimasi Parameter *Backward Elimination* Ke-6

Pemilihan Nilai Ridge Parameter *Backward Elimination* Ke-6

r	df model	selisih df dan r	θ_r^{\oplus}
1	3.2422	2.2422	0.0511
2	3.0888	1.0888	0.0634
3*	3.4289	0.4289	0.0390
4	3.2262	0.7738	0.0522
5	3.1271	1.8729	0.0601

Estimasi Parameter *Backward Elimination* Ke-6

```
> data = read.table('D:/Tugas Akhir/Data/backward6.txt', header =
  TRUE)
> mod <- logisticRidge(Y ~ ., data = as.data.frame (data), nPCs=3)
> summary(mod)
Call:
logisticRidge(formula = Y ~ ., data = as.data.frame(data), nPCs=3)
Coefficients:
```

<i>Variables</i>	<i>Coef</i>	<i>Coef (scaled)</i>	<i>Std. error</i>	<i>t-value (scaled)</i>	<i>P-value</i>
(Intercept)	3.1037	NA	NA	NA	NA
X1***	-0.0640	-4.9153	1.4192	-3.4630	0.0005
X2	-0.0686	-2.4808	1.7220	-1.4410	0.1497
X4**	-0.1077	-3.9094	1.4120	-2.7690	0.0056
X5*	0.0541	3.8068	1.5987	2.3810	0.0173
X6	0.0006	1.4238	1.4289	0.9960	0.3191
X8	-0.0165	-1.0631	1.6570	-0.6420	0.5212
X10	0.1409	0.4549	1.5917	0.2860	0.7750

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge paramter: 0.0390339

Degrees of freedom: model 3.429 , variance 3.881

Lampiran B8 Estimasi Parameter *Backward Elimination* Ke-7

Pemilihan Nilai Ridge Parameter *Backward Elimination* Ke-7

r	df model	selisih df dan r	θ_r^{\oplus}
1	3.1780	2.1780	0.0363
2	2.8646	0.8646	0.0610
3*	3.1474	0.1474	0.0382
4	3.0216	0.9784	0.0472

Estimasi Parameter *Backward Elimination* Ke-7

```
> data = read.table('D:/Tugas Akhir/Data/backward7.txt', header =
  TRUE)
> mod <- logisticRidge(Y ~ ., data = as.data.frame (data), nPCs=3)
> summary(mod)
Call:
logisticRidge(formula = Y ~ ., data = as.data.frame(data), nPCs=3)
Coefficients:
```

Variables	Coef	Coef (scaled)	Std. error	t-value (scaled)	P-value
(Intercept)	4.1055	NA	NA	NA	NA
X1***	-0.0645	-4.9566	1.4350	-3.4540	0.0006
X2	-0.0681	-2.4638	1.7605	-1.4000	0.1617
X4*	-0.1114	-4.0467	1.5838	-2.5550	0.0106
X5*	0.0543	3.8221	1.6275	2.3480	0.0189
X6	0.0006	1.4430	1.4520	0.9940	0.3203
X8	-0.0169	-1.0893	1.6760	-0.6500	0.5157

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge paramter: 0.03818807

Degrees of freedom: model 3.147 , variance 3.551

Lampiran B9 Estimasi Parameter *Backward Elimination* Ke-8

Pemilihan Nilai Ridge Parameter *Backward Elimination* Ke-8

r	df model	selisih df dan r	θ_r^{\oplus}
1	2.8868	1.8868	0.0315
2	2.9843	0.9843	0.0259
3*	2.8157	0.1843	0.0363
4	2.7043	1.2957	0.0452

Estimasi Parameter *Backward Elimination* Ke-8

```
> data = read.table('D:/Tugas Akhir/Data/backward8.txt', header =
TRUE)
> mod <- logisticRidge(Y ~ ., data = as.data.frame (data), nPCs=3)
> summary(mod)
Call:
logisticRidge(formula = Y ~ ., data = as.data.frame(data), nPCs=3)
Coefficients:
```

<i>Variables</i>	<i>Coef</i>	<i>Coef (scaled)</i>	<i>Std. error</i>	<i>t-value (scaled)</i>	<i>P-value</i>
(Intercept)	3.8818	NA	NA	NA	NA
X1***	-0.0691	-5.3132	1.5800	-3.3630	0.0008
X2	-0.0654	-2.3653	1.8438	-1.2830	0.1996
X4**	-0.1162	-4.2196	1.6279	-2.5920	0.0095
X5*	0.0559	3.9367	1.6736	2.3520	0.0187
X6	0.0006	1.3992	1.5242	0.9180	0.3586

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Ridge paramter: 0.03625792
Degrees of freedom: model 2.816 , variance 3.32
```

Lampiran B10 Estimasi Parameter *Backward Elimination* Ke-9

Pemilihan Nilai Ridge Parameter *Backward Elimination* Ke-9

r	df model	selisih df dan r	θ_r^{\oplus}
1	2.7294	1.7294	0.0275
2	2.7847	0.7847	0.0242
3*	2.6357	0.3643	0.0338

Estimasi Parameter *Backward Elimination* Ke-9

```
> data = read.table('D:/Tugas Akhir/Data/backward9.txt', header =
  TRUE)
> mod <- logisticRidge(Y ~ ., data = as.data.frame (data), nPCs=3)
> summary(mod)
Call:
logisticRidge(formula = Y ~ ., data = as.data.frame(data), nPCs=3)
Coefficients:
```

<i>Variables</i>	<i>Coef</i>	<i>Coef (scaled)</i>	<i>Std. error</i>	<i>t-value (scaled)</i>	<i>P-value</i>
(Intercept)	4.4503	NA	NA	NA	NA
X1***	-0.0710	-5.4570	1.6379	-3.3320	0.0009
X2	-0.0682	-2.4651	1.9064	-1.2930	0.1960
X4*	-0.1184	-4.3004	1.6952	-2.5370	0.0112
X5*	0.0587	4.1280	1.7473	2.3620	0.0182

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge paramter: 0.03380747

Degrees of freedom: model 2.636 , variance 3.184

Lampiran B11 Estimasi Parameter *Backward Elimination* Ke-10 (Model Terbaik)

Pemilihan Nilai Ridge Parameter *Backward Elimination* Ke-9
(Model Terbaik)

r	df model	selisih df dan r	θ_r^{\oplus}
1	2.2416	1.2416	0.0249
2*	2.0299	0.0299	0.0488

Estimasi Parameter *Backward Elimination* Ke-10 (Model Terbaik)

```
> data = read.table('D:/Tugas Akhir/Data/backward10.txt', header =
  TRUE)
> mod <- logisticRidge(Y ~ ., data = as.data.frame (data), nPCs=2)
> summary(mod)
Call:
logisticRidge(formula = Y ~ ., data = as.data.frame(data), nPCs=2)
Coefficients:
```

<i>Variables</i>	<i>Coef</i>	<i>Coef (scaled)</i>	<i>Std. error</i>	<i>t-value (scaled)</i>	<i>P-value</i>
(Intercept)	2.6267	NA	NA	NA	NA
X1***	-0.0638	-4.9017	1.3501	-3.6310	0.0003
X4***	-0.1046	-3.7968	1.3801	-2.7510	0.0059
X5*	0.0487	3.4268	1.4331	2.3910	0.0168

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Ridge paramter: 0.04879089

Degrees of freedom: model 2.03 , variance 2.291

BIODATA PENULIS



Penulis yang bernama lengkap Dwi Maumere Putra. Anak kedua dari tiga bersaudara yang lahir pada tanggal 14 Desember 1992 di Bangkalan. Penulis telah menempuh jenjang pendidikan formal di SDN Kedundung II Mojokerto, SMPN 16 Surabaya dan SMAN 1 Puri Mojokerto. Pada Tahun 2011 penulis diterima di Jurusan Statistika ITS Program Studi S1 melalui SNMPTN Jalur Tulis. Semasa kuliah Penulis sempat aktif dalam beberapa organisasi mahasiswa diantaranya

adalah sebagai Staff Departemen Pengembangan Sumber Daya Mahasiswa (PSDM) HIMASTA-ITS (2012/2013), Kabiro Pelatihan dan Kepemanduan Departemen Pengembangan Sumber Daya Mahasiswa (PSDM) HIMASTA-ITS (2013/2014), Staff Kementerian Pengembangan Sumber Daya Mahasiswa (PSDM) Badan Eksekutif Mahasiswa (BEM) ITS (2012/2013) dan Kepala Bidang Pengembangan Organisasi (PO) Ikatan Himpunan Mahasiswa Statistika Indonesia (IHMSI) (2012/2014). Kepanitiaan yang dijalani penulis selama masa kuliah diantaranya sebagai Ketua Character Capacity Building For Future Leader (CCBFL) 2013, Koor Perlengkapan CERITA 2013 dan lainnya. Bagi pembaca yang ingin berdiskusi, memberikan saran dan kritik tentang Tugas Akhir ini dapat disampaikan melalui nomor 087854044375 atau melalui email : mere.putra@gmail.com.

(Halaman ini sengaja dikosongkan)