ESTIMASI KOEFISIEN REGRESI LOGISTIK BINER DENGAN METODE *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR*

SKRIPSI

IMELDA WIDJAJA

H 121 11 262

PROGRAM STUDI STATISTIKA JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

JUNI 2015

ESTIMASI KOEFISIEN REGRESI LOGISTIK BINER DENGAN METODE *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR*

SKRIPSI

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin Makassar

IMELDA WIDJAJA

H 121 11 262

PROGRAM STUDI STATISTIKA JURUSAN MATEMATIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS HASANUDDIN

MAKASSAR

JUNI 2015

LEMBAR PERNYATAAN KEOTENTIKAN

Saya yang bertanda tangan di bawah ini menyatakan dengan sungguh-sungguh bahwa skripsi yang saya buat dengan judul:

Estimasi Koefisien Regresi Logistik Biner dengan Metode *Least Absolute Shrinkage and Selection Operator*

Adalah benar hasil karya saya sendiri, bukan hasil plagiat dan belum pernah dipublikasikan dalam bentuk apapun.

Makassar, Mei 2015

Imelda Widjaja
NIM. H 121 11 262

ESTIMASI KOEFISIEN REGRESI LOGISTIK BINER DENGAN METODE *LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR*

Disetujui Oleh :

Pembimbing Utama	Pembimbing Pertama
Anna Islamiyati, S.Si., M.Si	**Drs. Raupong, M.Si**
NIP.19770808 200501 2 002	**NIP. 19621015 198810 1 001**

Pada Tanggal : Mei 2015

HALAMAN PENGESAHAN

Skripsi ini diajukan oleh	:
Nama	: IMELDA WIDJAJA
NIM	: H 121 11 262
Program Studi	: STATISTIKA
Judul Skripsi	: Estimasi Koefisien Regresi Logistik Biner dengan Metode *Least Absolute Shrinkage and Selection Operator*

Telah berhasil dipertahankan dihadapan Dewan Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

DEWAN PENGUJI

		Tanda Tangan
1. Ketua	: **Prof.Dr.Hj.Aidawayati R.,MS**	(.................................)
2. Sekretaris	: **Andi Kresna Jaya,S.Si.,M.Si.,**	(.................................)
3. Anggota	: **Drs. Muhammad Zakir,M.Si.**	(.................................)
4. Anggota	: **Anna Islamiyati, S.Si., M.Si.**	(.................................)
5. Anggota	: **Drs. Raupong, M.Si.**	(.................................)

Ditetapkan di : Makassar

Tanggal : Mei 2015

KATA PENGANTAR

Puji dan syukur dipanjatkan hanya kepada Tuhan yang Maha Esa atas segala rahmat dan kuasaNya, yang telah menyertai penulis selama proses penyelesaian skripsi dengan judul **"Estimasi Koefisien Regresi Logistik Biner dengan Metode *Least Absolute Shrinkage and Selection Operator*"** sebagai salah satu syarat untuk memperoleh gelar Sarjana Sains pada Program Studi Statistika Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin Makassar.

Selesainya skripsi ini juga tidak terlepas dari dukungan dari banyak orang. Ucapan terima kasih dan penghargaan yang tak terhingga saya ucapkan kepada Ayahanda dan Ibunda tercinta ***Steve Widjaja*** dan ***Veronika*** atas doanya yang tak pernah putus, serta kasih sayang yang melimpah dalam mendidik dan membesarkan penulis dengan begitu banyak pengorbanan yang tak pernah ternilai harganya, juga untuk ***Zasmita, Mersi dan Vera*** atas motivasi dan semangat serta dukungannya kepada penulis yang begitu besar.

Terima kasih yang terhingga juga penulis pesembahkan kepada :

1. Ibu **Rektor Universitas Hasanuddin** beserta jajarannya, Bapak **Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam** beserta jajarannya atas ilmu dan kemudahan-kemudahan yang diberikan, baik di bidang akademik maupun di bidang kemahasiswaan.
2. Ibu **Prof.Dr.Hasmawati, M.Si.,** selaku Ketua Jurusan Matematika, atas ilmu dan nasehatnya. Bapak **Dr.Nurdin,M.Si** selaku Sekretaris Jurusan yang telah memberikan banyak bantuan selama penulis menjalani pendidikan. Terima kasih pula untuk jajaran Pegawai Akademik Jurusan Matematika(untuk Pak Nasir, Pak Said) untuk bantuannya selama ini dalam pengurusan akademik.
3. Ibu **Anna Islamiyati,S.Si.,M.Si.** selaku pembimbing utama dan Bapak **Drs. Raupong, M.Si.** selaku pembimbing pertama, yang selalu bersabar dalam membimbing penulis dengan sebaik-baiknya dan telah bersedia meluangkan waktu disela-sela rutinitas yang begitu padat demi membimbing penulis dari awal hingga menyesaikan skripsi ini.

4. Ibu **Prof. Dr. Hj. Aidawayati Rangkuti, MS** selaku ketua tim penguji penulis, Bapak **Andi Kresna Jaya,S.Si.,M.Si.** selaku sekretaris tim penguji, dan Bapak **Drs. Muhammad Zakir, M.Si.** selaku anggota tim penguji dan Penasihat Akademik dari penulis yang telah meluangkan waktunya untuk memberikan saran dan arahan kepada penulis khususnya dalam penyusunan skripsi ini.
5. Bapak dan Ibu **Dosen Jurusan Matematika FMIPA UNHAS** yang telah berbagi ilmu dengan penulis baik selama proses perkuliahan dan penyusunan skripsi ini.
6. **Teman-teman jurusan matematika angkatan 2011 dan POLINOM 2011** (terkhusus buat Indah, Riska, Yetti, Wahda, Besse, Rika, Ima, Ebas, Ashar, Rizal, Ansar, Chacha, dkk) atas segala bantuan yang selalu ada diwaktu yang tepat, kekompakan, kesenangan dan kebersamaannya selama menghadapi masa-masa terindah maupun tersulit dalam menuntut ilmu.
7. **Sahabat-sahabatku** (Wulan, Niluh, Agnes, Irianti dan Yunita) yang memberikan motivasi, semangat dan menemaniku selama ini.
8. **Senior-senior dari jurusan matematika** (terkhusus buat Kak Edi dan Kak Octa) atas segala masukan serta dukungan kepada penulis selama melaksanakan studi di Universitas Hasanuddin.
9. Seluruh warga **Himatika** sebagai keluarga yang memberi nasehat-nasehatnya sehingga penulis dapat lebih mengerti arti pentingnya kebersamaan dan semua pihak yang telah banyak membantu penulis dan tak sempat penulis sebutkan satu per satu. Semoga segala bantuan dan partisipasinya dibalas oleh Tuhan Yang Maha Esa dengan balasan yang terbaik disisiNya.

Akhir kata, hanya kepada Tuhan Yang Maha Esa kita bersandar, segala puji hanya milikNya. Mudah-mudahan tulisan ini memberikan manfaat kepada semua pihak yang membutuhkan dan terutama untuk penulis. *Salam Sejahtera.*

Makassar, Mei 2015

Penulis

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama	: Imelda Widjaja
NIM	: H121 11 262
Program Studi	: Statistika
Departemen	: Matematika
Fakultas	: Matematika dan Ilmu Pengetahuan Alam
Jenis karya	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Bebas Royalti Noneksklusif** (***Non-exclusive Royalty- Free Right***) atas karya ilmiah saya yang berjudul:

"Estimasi Koefisien Regresi Logistik Biner dengan Metode *Least Absolute Shrinkage and Selection Operator*"

Beserta perangkat yang ada (jika diperlukan). Terkait dengan hal di atas, maka pihak universitas berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar
Pada tanggal, Mei 2015

Yang menyatakan

(Imelda Widjaja)

ABSTRAK

Korelasi yang tinggi antar variabel bebas (multikolinearitas) merupakan sebuah masalah dalam regresi logistik. Ada beberapa metode yang dapat mengatasi masalah multikolinearitas, tetapi metode tersebut memiliki kelemahan, yaitu tidak dapat menyusutkan koefisien menuju 0 dan seleksi variabel bebas secara bersamaan. Metode *Least Absolute Shrinkage and Selection Operator* (*LASSO*) dengan *penalty Lagrangian* dapat menyusutkan koefisien menuju 0 dan seleksi variabel secara bersamaan. Skripsi ini bertujuan untuk mengestimasi koefisien regresi logistik biner dengan metode *LASSO*. Selanjutnya diterapkan pada data kemiskinan setiap desa di Kabupaten Jeneponto. Berdasarkan model regresi logistik *LASSO* diperoleh taksiran model yang menunjukkan jika jumlah rumah tangga yang menggunakan sember penerangan Pelita / Petromaks dan bukan PLN lebih banyak daripada jumlah rumah tangga yang menggunakan penerangan PLN dan bekerja di bidang industri maka desa tersebut dikategorikan desa miskin. Ketepatan pengklasifikasian dengan metode *LASSO* sebesar 80,53%.

Kata kunci : Multikolinearitas, Regresi logistik biner, Metode *LASSO*, Ketepatan pengklasifikasian , data kemiskinan setiap desa.

ABSTRACT

High correlation between predictor variables (multicollinearity) become a problem in logistic regression. There are some method to solve the problem, but the methods have drawbacks i.e. they can not shrink some coefficients to 0 and do variable-selection technique simultaneity. Least Absolute Shrinkage and Selection Operator (LASSO) method can shrink some coefficients to 0 and do variable-selection technique simultaneity. This paper proposed to estimate binary logistic regression coefficients with LASSO method. Then it will be applied on proverty data of all village in Jeneponto. According to the binary logistic regression model with LASSO method, the estimated model shows that if number of household that which using lighting source i.e. oil lamp and not State Electricity Enterprise more than number of household using which State Electricity Enterprise and work in industry sector then that village is classificated in proverty village. Accuracy of classification with LASSO method is 80,53%.

Keywords: Multicollinearity, Binary logistic regression, LASSO, Accuracy of Classification, Proverty data.

DAFTAR ISI

DAFTAR TABEL

DAFTAR LAMPIRAN

BAB I
PENDAHULUAN

1.1. Latar Belakang

Analisis regresi bertujuan untuk mengetahui hubungan antara dua variabel atau lebih sehingga sebuah variabel bisa diramalkan berdasarkan variabel lainnya. Pada analisis regresi terdapat dua jenis variabel yaitu variabel bebas (*independent variable*) dan variabel tak bebas (*dependent variable*). Variabel bebas adalah variabel yang nilainya dapat diamati namun tidak dapat dikendalikan, sedangkan variablel tak bebas adalah variabel yang nilainya bergantung pada variabel bebas.

Salah satu model regresi yang biasa digunakan ketika variabel responnya kualitatif adalah model regresi logistik. Regresi logistik terdiri atas model regresi logistik biner, multinomial dan ordinal. Analisis regresi logistik biner digunakan untuk menjelaskan hubungan antara variabel respon yang berupa data dikotomik/biner dengan variabel bebas yang berupa data interval dan atau kategorik (Hosmer & Lemeshow, 1989). Variabel dikotomik atau biner adalah variabel yang hanya mempunyai dua kategori saja, yaitu kategori yang menyatakan kejadian sukses $y = 1$ dan kategori yang menyatakan kejadian gagal $y = 0$.

Terdapat beberapa asumsi klasik dalam analisis regresi yang harus dipenuhi agar menghasilkan model regresi yang baik, antara lain galat berdistribusi normal (normalitas), antara galat-galat tidak berkorelasi atau bersifat saling bebas, dan ragam suku galat konstan dan homogen. Tidak terpenuhinya asumsi tersebut biasanya disebabkan oleh korelasi antar variabel bebas (multikolinearitas), masalah pada jumlah data dan jumlah variabel bebas yang diperoleh. Ketika diperoleh data yang jumlah variabel (variabel bebas) lebih banyak daripada ukuran sampel.

Masalah multikolinieritas dapat diatasi dengan harapan diperoleh model terbaik dengan ragam minimum. Untuk mengatasi masalah ini dapat digunakan beberapa metode regresi, yaitu regresi gulud (*ridge regression*) (Pusporini, 2012), penggunaan analisis komponen utama (Ruslan, 2013) dan metode PLS

(*Partial Least Square*) dan *LASSO* (*Least absolute shrinkage and selection operator*) (Dewi, 2010). Metode *LASSO* dapat juga diterapkan pada data *microarray* (Farmani, 2012). *LASSO* merupakan teknik regresi yang melakukan penaksiran dengan meminimumkan jumlah kuadrat error dengan suatu kendala. Karena kendala tersebut, maka *LASSO* menyusutkan sejumlah koefisien menuju 0 dan pada waktu yang sama melakukan seleksi variabel (Tibshirani,1996). Metode ini dapat diterapkan pada beberapa model diantaranya regresi cox (Tibshirani,1997) dan regresi logistik (Roth, 2004). Sehingga penulis akan mengkaji kembali penggunaan metode *LASSO* (*Least Absolute Shrinkage and Selection Operator*) dalam model regresi logistik biner yang diaplikasikan pada data kemiskinan setiap desa pada Kabupaten Jeneponto Tahun 2013 yang dipengaruhi oleh beberapa faktor. Pembahasan masalah tersebut akan dirampungkan dalam bentuk tugas akhir dengan judul **"Estimasi Koefisien Regresi Logistik Biner dengan Metode *Least Absolute Shrinkage and Selection Operator*".**

1.2. Rumusan Masalah

Masalah yang akan dibahas dalam tugas akhir ini adalah :

1. Bagaimana mengestimasi koefisien regresi logistik biner dengan metode *Least Absolute Shrinkage and Selection Operator* ?
2. Bagaimana penerapan model Regresi Logistik Biner dengan metode *Least Absolute Shrinkage and Selection Operator* pada data kemiskinan setiap desa di Kabupaten Jeneponto Tahun 2013 ?

1.3.Batasan Masalah

Batasan masalah sangat diperlukan untuk menjamin kesimpulan yang akan diperoleh dalam sebuah penelitian. Agar tidak terjadi penyimpangan dari tujuan semula dan pemecahan masalah lebih terkonsentrasi maka pembahasan ini dibatasi pada Algoritma Osborne's Dual (Lasso2) untuk regresi logistik *LASSO*.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah, maka tugas akhir ini memiliki tujuan yaitu :

1. Untuk memperoleh estimasi koefisien pada regresi logistik biner dengan metode *Least Absolute Shrinkage and Selection Operator*.
2. Untuk memperoleh model regresi logistik biner pada data kemiskinan setiap desa di Kabupaten Jeneponto Tahun 2013 dengan metode *Least Absolute Shrinkage and Selection Operator*.

1.5. Manfaat Penelitian

Manfaat dari penulisan tugas akhir ini yaitu kita dapat menyelesaikan masalah multikolinearitas pada regresi logistik menggunakan metode *LASSO* dan penerapan dalam data kemiskinan setiap desa di Kabupaten Jeneponto Tahun 2013.

BAB II
TINJAUAN PUSTAKA

2.1. Regresi Logistik Biner

Model untuk menganalisis data biner atau data binomial adalah model regresi logistik biner yang pertama kali diperkenalkan oleh Cox (1970). Regresi logistik biner, yaitu salah satu bentuk regresi logistik yang variabel responnya terdiri dari dua kategori.

Misalkan Y adalah sebuah variabel respon biner, maka Y dapat dicontohkan dalam pilihan suara pemilu di Amerika Serikat (demokrat atau republik), atau contoh lainnya yang banyak terdapat dalam kejadian riil. Hal tersebut menunjukkan bahwa nilai dari variabel Y hanya mengindikasikan 2 hal atau dapat ditulis sebagai 0 dan 1. Variabel Y yang demikian lebih tepat dikatakan sebagai variabel respon dan memenuhi distribusi Bernoulli. Fungsi distribusi peluang untuk Y dengan parameter π_i adalah :

$$f(y_i, \pi_i) = \begin{cases} \pi_i^{y_i}(1-\pi_i)^{1-y_i} \text{ , } y_i = 0,1 \\ 0 \quad \text{, untuk } y_i \text{ yang lainnya} \end{cases} \tag{2.1}$$

dengan $\pi_i = P(Y_i = 1)$ dan $(1-\pi_i) = P(Y_i = 0)$. Mean dari variabel acak Bernoulli adalah :

$$E(Y_i) = 1 \times P(Y_i = 1) + 0 \times P(Y_i = 0) = P(Y_i = 1)$$

Misalkan fungsi peluang yang dinyatakan sebagai $\pi_i(x)$ yang bergantung pada variabel bebas $X = (X_1, \ldots, X_n)$ dengan $E(Y_i) = \pi_i$ dan $0 \le \pi_i \le 1$, sehingga diperoleh

$$E(Y_i^2) = 1^2\pi_i(x) + 0^2[1-\pi_i(x)] = \pi_i(x)$$

Variansi dari Y adalah

$$V[Y_i] = E(Y_i^2) - [E(Y_i)]^2 = \pi_i(x)[1-\pi_i(x)]$$

Fungsi regresi logistiknya dapat dituliskan dengan:

$$f(z) = \frac{1}{1+e^{-z}} \text{ ekuivalen dengan } f(z) = \frac{e^z}{1+e^z} \tag{2.2}$$

dimana:

$z = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$, dengan p = banyaknya variabel bebas

Nilai z berada di antara $-\infty$ dan $+\infty$ sehingga nilai $f(z)$ terletak antara 0 dan 1 untuk setiap nilai z yang diberikan. Hal tersebut menunjukkan bahwa model logistik sebenarnya menggambarkan probabilitas atau risiko dari suatu objek. Model regresi logistiknya adalah:

$$\pi_i(x) = \frac{e^{(\beta_0+\beta_1 X_{i1}+\cdots+\beta_p X_{ip})}}{1+e^{(\beta_0+\beta_1 X_{i1}+\cdots+\beta_p X_{ip})}} \tag{2.3}$$

dimana β_p menyatakan parameter-parameter regresi,

X_{ip} adalah pengamatan ke-i dengan sejumlah p variabel bebas.

Untuk mempermudah pendugaan parameter regresi maka model regresi logistik pada persamaan (2.3) dapat diuraikan dengan menggunakan transformasi logit dari $\pi(x)$, sehingga diperoleh persamaan sebagai berikut:

$$g(x) = ln\left(\frac{\pi_i(x)}{1-\pi_i(x)}\right) = \beta_0+\beta_1 X_{i1}+\cdots+\beta_p X_{ip} \tag{2.4}$$

Model tersebut merupakan fungsi linier dari parameter-parameternya. Pada regresi logistik, variabel respon diekspresikan sebagai $y = \pi(x) + \varepsilon$, dimana ε mempunyai salah satu dari kemungkinan dua nilai yaitu $\varepsilon = 1 - \pi(x)$ dengan peluang π(x), jika $y = 1$, dan $\varepsilon = -\pi(x)$ dengan peluang $1 - \pi(x)$ jika $y = 0$ dan mengikuti distribusi binomial dengan rataan nol dan varians $(\pi(x))(1-\pi(x))$ (Agresti,1990).

2.2. Multikolinearitas

Multikolinearitas terjadi ketika terdapat korelasi antara dua atau lebih variabel bebas dalam regresi. Adanya multikolinearitas mengakibatkan penduga koefisien regresi yang diperoleh dari *OLS* akan menghasilkan variansi yang besar, meskipun tetap bias (Pasha & Shah (2004) dalam Pusporini (2012)).

Cara mendeteksi multikolinearitas dalam model regresi, sebagai berikut :

1. Korelasi yang tinggi antara pasangan variabel bebas dalam model.
2. *Variance Inflation Factor* (VIF) untuk sebuah parameter β lebih besar dari 10, dimana

$$(\text{VIF})_i = \frac{1}{1-R_i^2}, \quad i = 1,2,\ldots,k \tag{2.5}$$

R_i^2 merupakan koefisien determinasi regresi (Mendenhall&Sincich,2003).

2.3. *Least Absolute Shrinkage and Selection Operator*

Metode *Least Absolute Shrinkage and Selection Operator* (*LASSO*) diperkenalkan pertama kali oleh Tibshirani pada tahun 1996. Penduga koefisien *LASSO* diperoleh dengan cara pemrograman kuadratik. *LASSO* adalah salah satu teknik regresi penyusutan variabel bebas. *LASSO* menyusutkan koefisien regresi (β_k) yang berkorelasi menjadi nol atau mendekati nol. Sehingga menghasilkan penduga dengan varian yang lebih kecil dan model akhir yang lebih representatif (Tibshirani, 1996).

Misalkan $(X_i, Y_i), \ldots, (X_n, Y_n)$ adalah n pasangan variabel bebas atau variabel respon dengan $Y_i \in Y$ dan $X_i \in X$, dimana X dan Y adalah input dan output dengan Y variabel respon, p variabel bebas $X_1, X_2, \ldots, X_p$ dan n banyaknya pengamatan. Penduga parameter pada *LASSO* dalam model regresi linear adalah sebagai berikut (Farmani,2012):

$$R(\beta_0, \beta_1) = \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \sum_{k=1}^{p} X_{ik}\hat{\beta}_k\right)^2$$

$$(\beta)_{l_1} = \sum_{i=1}^{n}(Y_i - X_i^T\beta)^2 + \lambda\sum_{k=1}^{p}|\hat{\beta}_k|$$

$$= (\boldsymbol{Y} - \boldsymbol{X\beta})^T(\boldsymbol{Y} - \boldsymbol{X\beta}) + \lambda\|\hat{\boldsymbol{\beta}}\|_1 \quad (2.6)$$

λ adalah parameter yang mengontrol koefisien *LASSO* yang diatur dengan batasan $D = \sum_{k=1}^{p}|\hat{\beta}_k| \leq t$.

Dalam General Linear Model (GLM) , penduga *LASSO* didefinisikan oleh *penalizing the log-likelihood.* Penduga *LASSO* dalam model GLM sebagai berikut (Buhlmann *et al*,2010):

$$\hat{\mu}(\lambda), \hat{\beta}(\lambda) = \arg\min_{\mu,\beta}\left\{n^{-1}\sum_{i=1}^{n}\rho_{\mu,\beta}(X_i, Y_i) + \lambda\|\beta\|_1\right\} \quad (2.7)$$

dimana $n^{-1}\sum_{i=1}^{n}\rho_{\mu,\beta}(X_i, Y_i) = -\sum_{i=1}^{n}\log\left(p_{\mu,\beta}(Y_i|X_i)\right)$. Dalam regresi logistik, *log-likelihood* sama dengan

$$-\sum_{i=1}^{n}\log\left(p_{\mu,\beta}(Y_i|X_i)\right) = \sum_{i=1}^{n}\left(-Y_i f_{\mu,\beta}(X_i) + \log\left(1 + \exp\left(f_{\mu,\beta}(X_i)\right)\right)\right) \quad (2.8)$$

Kendala $\sum|\beta_k| \leq s$ sebagai $\frac{\sum \beta_k^2}{|\beta_k|} \leq s$. Kendala ini ekivalen dengan penambahan penalty Lagrangian $\frac{\lambda \sum \beta_k^2}{|\beta_k|}$ pada *log likelihood* dengan $\lambda \geq 0$ yang bergantung pada s. menggunakan matriks manipulasi standar , maka solusi kendala $\widetilde{\boldsymbol{\beta}}$ dapat ditulis (Tibshirani, 1997) :

$$\widetilde{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{D} \boldsymbol{X} + \lambda \boldsymbol{W})^{-1} \boldsymbol{X}^T \boldsymbol{D} \boldsymbol{z} \tag{2.9}$$

$\boldsymbol{W} = \mathrm{diag}(\boldsymbol{W}_j), \boldsymbol{W}_i = 1/|\tilde{\beta}_k|$ jika $|\tilde{\beta}_k| > 0$ dan 0 yang lainnya. Pendekatan banyaknya parameter efektif dalam kendala $\widetilde{\boldsymbol{\beta}}$ adalah

$$p(s) = \mathrm{tr}[\mathbf{X}(\mathbf{X}^T \mathbf{D} \mathbf{X} + \lambda \mathbf{W}^-)^{-1} \boldsymbol{X}^T \boldsymbol{D}] \tag{2.10}$$

dimana $\mathbf{W}^-$ adalah invers umum matriks

D adalah sebuah matriks manipulasi standar.

maka *generalized cross validation* untuk *LASSO* adalah

$$\mathrm{GCV}(s) = \frac{1}{N} \frac{-\ell_s}{N[1 - p(s)/N]^2} \tag{2.11}$$

Keterangan : $-\ell_s$ adalah *negative log-likelihood* dengan kendala s.

N adalah banyaknya data.

2.4. Algoritma Osborne's Dual

Osborne, dkk menyajikan sebuah algoritma yang lebih cepat konvergen dalam jurnalnya *LASSO and its Dual(2000)*. Osborne memperkenalkan algoritma umum yang berdasarkan pada konsep duality dalam jurnal aslinya. Dalam *LASSO and its Dual(2000)*, algoritma tersebut dapat digunakan untuk menghitung estimasi LASSO dalam berbagai model.

Misalkan $\sum_j |\hat{\beta}_j^0| = t_0$ dan $\sum_j |\hat{\beta}_j^L| = t$,

$$\begin{aligned} t_0 - t &= \sum_j |\hat{\beta}_j^0| - \sum_k |\hat{\beta}_j^L| = \sum_j \left\{ |\hat{\beta}_j^0| - \left(|\hat{\beta}_j^L| - \gamma\right)^+ \right\} \\ &= \sum_j |\hat{\beta}_j^0| \, I\left(|\hat{\beta}_j^0| \leq \gamma\right) + \gamma \sum_j I\left(|\hat{\beta}_j^0| \geq \gamma\right) \\ &= \sum_j b_j + \gamma(p - K) \end{aligned} \tag{2.12}$$

dimana p adalah banyak variabel: $b_1 \leq b_2 \leq \cdots \leq b_p$ adalah statistik terurut dari $|\hat{\beta}_1^0|, |\hat{\beta}_2^0| \ldots |\hat{\beta}_p^0|$. $K = \max\{j : b_j \leq \gamma\}$. Pada umumnya $t < t_0, K < p$ dan $b_K \leq \gamma \leq b_{K+1}$.

Langkah – langkah dari algoritma Osborne's Dual (Lasso2) sebagai berikut :

1. Mulai dengan $c_0 = 0$
2. Lakukan iterasi j dari 1 sampai p, $c_j = \sum_{i=1}^{j} b_i + b_j(p-j)$ sehingga $0 = c_0 \leq c_1 \leq \cdots \leq c_p = t_0$.
3. Misalkan $K = \max\{i: c_i \leq t_0 - t\}$ dimana akan lebih mudah dihitung setelah t dipilih oleh GCV
4. Berhubungan dengan $\gamma = \frac{[(t_0-t)-\sum_{i=1}^{K} b_i]}{(p-K)}$
5. Menyelesaikan semua estimasi LASSO oleh $\hat{\beta}_j^L = \text{sign}\,(\hat{\beta}_j^0)(\hat{\beta}_j^0 - \gamma)^+$. $j = 1,2,\ldots,p$.

   Berdasarkan algoritma Osborne. Justin Lokhorst, Bill Venables dan Berwin T mengembangkan sebuah package "lasso2" dalam R.

2.5. Uji Signifikansi Parameter

Uji yang dapat digunakan untuk menguji signifikansi koefisien β dari model dapat menggunakan uji secara parsial. Pengujian secara parsial dilakukan untuk mengetahui signifikansi parameter terhadap variabel respon. Pengujian signifikansi parameter menggunakan Uji Wald (Hosmer dan Lemeshow, 2000) dengan hipotesis sebagai berikut:

$H_0 : \beta_k = 0$, dengan $k = 1,2,\ldots,p$ (tidak ada pengaruh variabel bebas ke-k terhadap variabel respon)

$H_1 : \beta_k \neq 0$, dengan $k = 1,2,\ldots,p$ (ada pengaruh variabel bebas ke-k terhadap variabel respon)

Statistik Uji :

$$W = \left(\frac{\widehat{\beta_k}}{SE(\widehat{\beta_k})}\right)^2 \tag{2.13}$$

dengan

$$SE(\widehat{\beta_k}) = \left[\text{Var}(\widehat{\beta_k})\right]^{1/2} \tag{2.14}$$

dimana :

$\widehat{\beta_k}$: Penaksir dari β_k

$SE(\widehat{\beta_k})$: Standar error dari $\widehat{\beta_k}$

$\text{Var}(\widehat{\beta_k})$: variansi $\widehat{\beta_k}$

Daerah penolakan H_0 adalah jika $|W|Z_{\alpha/2}$atau $W^2 > \chi^2_{(v,\alpha)}$ dengan derajat bebas v (Hosmer dan Lemeshow,2000).

2.6. Uji Kelayakan Model

Menurut Hosmer dan Lemeshow (2000) salah satu ukuran kelayakan model dalam regresi logistik adalah jika memiliki peluang salah klasifikasi yang minimal. Ketepatan dan kesalahan klasifikasi dari model dapat diketahui dengan mengunakan tabel klasifikasi (*classification table*). Tabel klasifikasi untuk peubah respon dikotom terdiri atas dua kolom nilai prediksi dan dua baris nilai observasi(aktual). Untuk memperoleh ketepatan klasifikasi (*correct classification*) terhadap amatan harus menentukan nilai *cutpoint* (c) dan dibandingkan dengan peluang dugaan π(x). Jika π(x) lebih besar atau sama dengan c maka nilai dugaan termasuk pada respon y = 1 dan selain itu y = 0. *Total correct classification* yaitu ketepatan klasifikasi dalam menduga kejadian secara keseluruhan.

Tabel 2.1. Tabel Ketepatan Klasifikasi

Observasi	Prediksi		%
	0	1	Ketepatan
0	n_{11} (Benar)	n_{12} (Salah)	A
1	n_{21} (Salah)	n_{22} (Benar)	B
% Ketepatan Keseluruhan			

(Hosmer dan Lemeshow,2000)

di mana:

n_{11} : banyaknya kejadian gagal dari hasil amatan yang tepat diprediksikan sebagai kejadian gagal

n_{12} : banyaknya kejadian gagal dari hasil amatan yang tepat diprediksikan sebagai kejadian sukses

n_{21} : banyaknya kejadian sukses dari hasil amatan yang tepat diprediksikan sebagai kejadian gagal

n_{22} : banyaknya kejadian sukses dari hasil amatan yang tepat diprediksikan sebagai kejadian sukses

APER (*Apparent Error Rate)* merupakan suatu nilai yang digunakan untuk melihat peluang kesalahan dalam mengklasifikasi objek. Perhitungan nilai APER adalah (Hosmer dan Lemeshow,2000):

$$APER = \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}} \times 100\% \qquad (2.15)$$

2.7. Data Kemiskinan

Kemiskinan merupakan masalah serius yang dihadapi oleh negara berkembang. Jumlah penduduk miskin Indonesia pada Tahun 2012 mencapai 29,13 juta orang atau 11,31 persen dari jumlah penduduk yang terdistribusi di daerah perkotaan sebanyak 36,5 persen dan di daerah perdesaan sebanyak 63,5 persen (BPS, 2012). Data tersebut menunjukan bahwa daerah perdesaan dengan basis perekonomian pada sektor primer (perikanan, pertanian dan perkebunan) yang lebih baik justru memiliki jumlah masyarakat miskin yang lebih tinggi. Menurut Bappenas Miskin adalah kondisi dimana seseorang atau sekelompok orang, tidak mampu memenuhi hak-hak dasarnya untuk mempertahankan & mengembangkan kehidupan yang bermartabat. Hak-hak dasar tersebut antara lain: terpenuhinya kebutuhan pangan, kesehatan, pendidikan, perumahan, air bersih, pertanahan dan lain-lain. BPS mempunyai 14 indikator sehingga suatu rumah tangga dikatakan miskin, yaitu :

1. Luas lantai bangunan tempat tinggal kurang dari 8 m^2 per orang.
2. Jenis lantai tempat tinggal terbuat dari tanah/ bambu / kayu murahan.
3. Jenis dinding tempat tinggal dari bambu / rumbia / kayu berkualitas rendah / tembok tanpa diplester.
4. Tidak memiliki fasilitas buang air besar / bersama-sama dengan rumah tangga lain.
5. Sumber penerangan rumah tangga tidak menggunakan listrik.
6. Sumber air minum berasal dari sumur / mata air tidak terlindung / sungai /air hujan.
7. Bahan bakar untuk memasak sehari-hari adalah kayu bakar / arang / minyak tanah.
8. Hanya mengkonsumsi daging / susu / ayam satu kali dalam seminggu.

9. Hanya membeli satu stel pakaian baru dalam setahun.
10. Hanya sanggup makan sebanyak satu / dua kali dalam sehari.
11. Tidak sanggup membayar biaya pengobatan di puskesmas / poliklinik.
12. Sumber penghasilan kepala rumah tangga adalah : petani dengan luas lahan 500 m^2, buruh tani, nelayan, buruh bangunan, buruh perkebunan dan atau pekerjaan lainnya dengan pendapatan dibawah Rp. 600.000,- per bulan.
13. Pendidikan tertinggi kepala rumah tangga : tidak sekolah / tidak tamat SD/ hanya SD.
14. Tidak memiliki tabungan / barang yang mudah dijual dengan minimal Rp. 500.000,- seperti sepeda motor kredit / non kredit, emas, ternak, kapal motor, atau barang modal lainnya.

Badan Pusat Statistik (BPS) merilis jumlah penduduk miskin di Sulsel per Maret 2010 tercatat sebesar 11,6% dari jumlah penduduknya atau sebesar 913,4 ribu jiwa. Dari jumlah tersebut, 13,0% berada di daerah perkotaan sedangkan sisanya berada di daerah pedesaan. Mereka kebanyakan hidup dari buruh tani atau tak memiliki pekerjaan yang tak menentu.

Salah satu variabel yang mempengaruhi variabel indikator kemiskinan adalah sumber penerangan utama, yaitu :

1. Listrik PLN adalah sumber penerangan listrik yang dikelola oleh PLN. Rumah tangga dikatakan menggunakan listrik PLN baik menggunakan maupun tidak menggunakan meteran (volumetrik).
2. Listrik non-PLN adalah sumber penerangan listrik yang dikelola oleh instansi / pihak lain selain PLN termasuk yang menggunakan sumber penerangan dari *accu* (aki), generator, dan pembangkit listrik tenaga surya (yang tidak dikelola oleh PLN).
3. Lainnya adalah pelita, petromaks, lilin dan biji jarak.

BAB III
METODOLOGI PENELITIAN

3.1. Jenis dan Sumber Data

Sesuai dengan permasalahan yang akan dibahas dalam tugas akhir ini. Maka data yang digunakan dalam penelitian ini adalah data sekunder yaitu data kemiskinan Kabupaten Jeneponto setiap desa yang diperoleh dari publikasi BPS (Badan Pusat Statistik) Kabupaten Jeneponto Tahun 2014 (http://www.jenepontokab.bps.go.id/index.php?hal=publikasi).

3.2. Variabel Penelitian

Adapun variabel yang digunakan pada studi kasus untuk penulisan tugas akhir ini sebagai berikut :

Variabel respon (Y_i) yang digunakan adalah persentase rumah tangga klasifikasi kemiskinan setiap desa di Kabupaten Jeneponto Tahun 2013 yang terdiri dari dua kategorik, yaitu :

0 = desa tidak miskin (jika jumlah penduduk miskin dibawah 30%, maka desa tersebut dikategorikan desa tidak miskin)

1 = desa miskin (jika jumlah penduduk miskin diatas 30%, maka desa tersebut dikategorikan desa miskin).

Variabel bebas yang diduga berpengaruh terhadap respon, yaitu :

X_{i1} : Persentase jumlah rumah tangga menurut desa dan sumber penerangan yang digunakan berupa PLN

X_{i2} : Persentase jumlah rumah tangga menurut desa dan sumber penerangan yang digunakan berupa bukan PLN

X_{i3} : Persentase jumlah rumah tangga menurut desa dan sumber penerangan yang digunakan berupa petromaks atau pelita

X_{i4} : Persentase jumlah kepala Keluarga menurut desa dan sumber mata pencaharian sebagai petani atau nelayan

X_{i5} : Persentase jumlah kepala Keluarga menurut desa dan sumber mata pencaharian sebagai pedagang

X_{i6} : Persentase jumlah kepala Keluarga menurut desa dan sumber mata pencaharian dibidang industri

X_{i7} : Persentase jumlah kepala Keluarga menurut desa dan sumber mata pencaharian dibidang angkutan

3.3. Metode Penelitian

Metode analisis yang digunakan dalam penelitian ini adalah regresi logistik biner dengan metode *Least Absolute Shrinkage and Selection Operator*, untuk 7 variabel bebas model regresi logistik biner *LASSO* dapat dituliskan sebagai berikut :

$$Y_i = \frac{1}{1+\exp(-\beta_0-\beta_1X_{i1}-\beta_2X_{i2}-\beta_3X_{i3}-\beta_4X_{i4}-\beta_5X_{i5}-\beta_6X_{i6}-\beta_7X_{i7})}$$

Untuk mencapai tujuan penulisan tugas akhir ini, maka langkah – langkah yang ditempuh sebagai berikut :

1. Menaksir koefisien regresi logistik biner.
   a. Menentukan fungsi *likelihood* dari data yang berdistribusi bernoulli
   b. Menentukan fungsi *log-likelihood* bersama untuk vektor β
   c. Menambahkan penalty Lagrangian pada fungsi *log-likelihood* bersama untuk vektor β yang disebut *penalizing the log-likelihood.*
   d. Memaksimumkan fungsi *penalizing the log-likelihood* dengan menurunkan fungsi tersebut terhadap parameternya.
   e. Menyelesaikan turunan fungsi *penalizing the log-likelihood* dengan menggunakan iterasi Newton-Raphson.
   f. Sehingga diperoleh nilai estimasi koefisien $\hat{\beta}$.
2. Penerapan metode LASSO (*Least Absolute Shrinkage and Selection Operator*) pada data kemiskinan setiap desa Kabupaten Jeneponto
   a. Mengamsumsikan variabel respon Y berdistribusi Bernoulli dalam regresi logistik biner.
   b. Melakukan uji multikolinearitas dengan memeriksa matriks korelasi dan signifikansi dengan SPSS. Jika data terbukti mengandung multikolinearitas, maka analisis dapat dilanjutkan pada *LASSO.*
   c. Membakukan variabel bebas X sehingga memiliki nilai tengah nol dan ragam satu

d. Menentukan hasil estimasi parameter $\hat{\beta}$ awal .

e. Mencari nilai λ yang minimum, dengan $0 \leq \lambda$, di mana λ adalah parameter *tuning* yang nilainya dipilih dari *generalized cross validation* yang minimum.

f. Melakukan proses iterasi sehingga memperoleh nilai $\hat{\beta}$ yang konvergen. Sehingga memperoleh estimasi parameter $\hat{\beta}$

g. Melakukan uji kelayakan model dengan menggunakan tabel keakuratan klasifikasi.

h. Melakukan uji signifikansi parameter $\hat{\beta}$ dengan menggunakan uji Wald dengan hipotesis :

   $H_0 : \beta_k = 0$, dengan $k = 1,2,\dots,p$ (tidak ada pengaruh variabel bebas ke-k terhadap variabel respon)

   $H_1 : \beta_k \neq 0$, dengan $k = 1,2,\dots,p$ (ada pengaruh variabel bebas ke-k terhadap variabel respon)

BAB IV
HASIL DAN PEMBAHASAN

4.1. Estimasi Koefisien Regresi Logistik Biner dengan Metode *Least Absolute Shrinkage and Selection Operator*

Bagian ini menguraikan estimasi koefisien regresi logistik biner dengan metode *Least Absolute Shrinkage and Selection Operator* (*LASSO*) yang merupakan kajian ulang dari Roth (2004). Namun belum dijabarkan secara mendetail sehingga dalam tulisan ini akan diuraikan beberapa tahapan dalam mendapatkan estimasi koefisien tersebut.

Diketahui fungsi peluang untuk distribusi Bernoulli, yaitu :

$$f(y) = \pi_i^{y_i}(1-\pi_i)^{1-y_i} \quad , y = 0,1 \tag{4.1}$$

Jika $y_i = 1$ maka $f(y_i) = \pi_i^1(1-\pi_i)^0 = \pi_i$ sebagai peluang sukses dan jika $y_i = 0$ maka $f(y_i) = \pi_i^0(1-\pi_i)^1 = 1-\pi_i$ sebagai peluang gagal.

Model regresi logistik dari persamaan (2.2) adalah sebagai berikut :

$$\pi_i = P(Y_i = 1|X_i, \beta) = \frac{1}{1+\exp(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik})} \tag{4.2}$$

Berdasarkan persamaan (4.2) dapat diperoleh $1-\pi_i$ adalah :

$$\begin{aligned} 1-\pi_i &= 1 - \frac{1}{1+\exp(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik})} \\ &= \frac{1+\exp(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik}) - 1}{1+\exp(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik})} \\ &= \frac{\exp(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik})}{1+\exp(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik})} \end{aligned} \tag{4.3}$$

Fungsi likelihood dari distribusi bernoulli adalah

$$\begin{aligned} L(\boldsymbol{\beta}|y_1, \dots, y_n) &= \prod_{i=1}^{n} f(y_i|\boldsymbol{\beta}) \\ &= \left[\pi_1^{y_1}(1-\pi_1)^{1-y_1}\right] \dots \left[\pi_n^{y_n}(1-\pi_n)^{1-y_n}\right] \\ &= \prod_{i=1}^{n} \pi_i^{y_i}(1-\pi_i)^{1-y_i} \\ &= \prod_{i=1}^{n} \left(\frac{1}{1+\exp(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik})}\right)^{y_i} \left(\frac{\exp(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik})}{1+\exp(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik})}\right)^{1-y_i} \end{aligned} \tag{4.4}$$

Selanjutnya fungsi *log*-likelihood pada persamaan (4.4) adalah

$$l(\boldsymbol{\beta}) = \ln\left(L(\boldsymbol{\beta}|y_1, \dots, y_n)\right)$$

$$l(\boldsymbol{\beta}) = \ln\left[\prod_{i=1}^{n} \left(\frac{1}{1+\exp(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik})}\right)^{y_i} \left(\frac{\exp(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik})}{1+\exp(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik})}\right)^{1-y_i}\right]$$

$$= \sum_{i=1}^{n}\left[\ln\left(\frac{1}{1+\exp(-\beta_0-\sum_{k=1}^{p}\beta_k x_{ik})}\right)^{y_i} + \ln\left(\frac{\exp(-\beta_0-\sum_{k=1}^{p}\beta_k x_{ik})}{1+\exp(-\beta_0-\sum_{k=1}^{p}\beta_k x_{ik})}\right)^{1-y_i}\right]$$

$$= \sum_{i=1}^{n}\Big[y_i\big[\ln(1) - \ln\big(1+\exp\big(-\beta_0 - \textstyle\sum_{k=1}^{p}\beta_k x_{ik}\big)\big)\big] +$$
$$\ln\left(\frac{\exp(-\beta_0-\sum_{k=1}^{p}\beta_k x_{ik})}{1+\exp(-\beta_0-\sum_{k=1}^{p}\beta_k x_{ik})}\right)^{1}\left(\frac{\exp(-\beta_0-\sum_{k=1}^{p}\beta_k x_{ik})}{1+\exp(-\beta_0-\sum_{k=1}^{p}\beta_k x_{ik})}\right)^{-y_i}\Big]$$

$$= \sum_{i=1}^{n}\Big[y_i\ln(1) - y_i\ln\big(1+\exp\big(-\beta_0 - \textstyle\sum_{k=1}^{p}\beta_k x_{ik}\big)\big) +$$
$$\ln\exp\big(-\beta_0 - \textstyle\sum_{k=1}^{p}\beta_k x_{ik}\big) - \ln\big(1+\exp\big(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik}\big)\big) +$$
$$\big(-y_i\ln\big(\exp\big(-\beta_0 - \textstyle\sum_{k=1}^{p}\beta_k x_{ik}\big)\big)\big) - (-y_i)\ln\big(1+\exp\big(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik}\big)\big)\Big]$$

$$= \sum_{i=1}^{n}\Big[0 - y_i\ln\big(1+\exp\big(-\beta_0 - \textstyle\sum_{k=1}^{p}\beta_k x_{ik}\big)\big) + \ln\exp\big(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik}\big) -$$
$$\ln\big(1+\exp\big(-\beta_0 - \textstyle\sum_{k=1}^{p}\beta_k x_{ik}\big)\big) - y_i\ln\big(\exp\big(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik}\big)\big) +$$
$$y_i\ln\big(1+\exp\big(-\beta_0 - \textstyle\sum_{k=1}^{p}\beta_k x_{ik}\big)\big)\Big]$$

$$= \sum_{i=1}^{n}\Big[\ln\big(\exp\big(-\beta_0 - \textstyle\sum_{k=1}^{p}\beta_k x_{ik}\big)\big) - y_i\ln\big(\exp\big(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik}\big)\big) - \ln\big(1 +$$
$$\exp\big(-\beta_0 - \textstyle\sum_{k=1}^{p}\beta_k x_{ik}\big)\big)\Big]$$

$$= \sum_{i=1}^{n}\Big[(1-y_i)\ln\big(\exp\big(-\beta_0 - \textstyle\sum_{k=1}^{p}\beta_k x_{ik}\big)\big) - \ln\big(1+\exp\big(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik}\big)\big)\Big]$$

$$= \sum_{i=1}^{n}\Big[-(1-y_i)\big(\beta_0 + \textstyle\sum_{k=1}^{p}\beta_k x_{ik}\big) - \ln\big(1+\exp\big(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik}\big)\big)\Big]$$

$$= -\sum_{i=1}^{n}\Big[(1-y_i)\big(\beta_0 + \textstyle\sum_{k=1}^{p}\beta_k x_{ik}\big) + \ln\big(1+\exp\big(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik}\big)\big)\Big] \qquad (4.5)$$

dalam metode *LASSO* fungsi *log likelihood* pada persamaan (4.5) dengan kendala $\sum_{k=1}^{p}|\beta_k| \leq t$, t merupakan suatu besaran yang mengontrol besarnya penyusutan pada penduga koefisien *LASSO* dengan $t \geq 0$. Kendala tersebut ekivalen dengan menambahkan *penalty lagrangian. Penalty Lagrangian* berasal dari distribusi *Laplace* dengan mean 0. Fkp variabel acak distribusi *Laplace* dengan mean μ dan variansi $2/_{\lambda^2}$ adalah (Lee et al,2006) :

$$f(\beta_k) = \frac{\lambda}{2}\exp(-\lambda|\beta_k - \mu|) \qquad (4.6)$$

selanjutnya fungsi likelihood dari persamaan (4.6) adalah

$$L(\beta_1, \dots, \beta_p) = \left[\frac{\lambda}{2}\exp(-\lambda|\beta_1|)\right] \cdots \left[\frac{\lambda}{2}\exp(-\lambda|\beta_p|)\right]$$
$$= \prod_{k=1}^{p}\frac{\lambda}{2}\exp(-\lambda|\beta_k|) \qquad (4.7)$$

Persamaan (4.7) kemudian dibentuk fungsi *log-likelihood* sebagai berikut

$$\ln L(\beta_1, \dots, \beta_p) = \ln\left(\prod_{k=1}^{p}\frac{\lambda}{2}\exp(-\lambda|\beta_k|)\right)$$
$$= p\ln(\lambda) - p\ln(2) + \ln\big(\exp\big(-\lambda\textstyle\sum_{k=1}^{p}|\beta_k|\big)\big)$$
$$= p\ln(\lambda) - p\ln(2) - \lambda\sum_{k=1}^{p}|\beta_k| \qquad (4.8)$$

dimana fungsi *log-likelihood* pada persamaan (4.5) ditambahkan sebuah fungsi penalty yaitu *penalty Lagrangian* pada persamaan (4.8) sehingga diperoleh

$$l(\beta)_c = -\sum_{i=1}^{n}\left[(1-y_i)\left(\beta_0 + \sum_{k=1}^{p}\beta_k x_{ik}\right) + \ln\left(1 + \exp\left(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik}\right)\right)\right] + p\ln(\lambda) + p\ln(2) - \lambda\sum_{k=1}^{p}|\beta_k| \quad (4.9)$$

Persamaan (4.9) ekivalen dengan

$$L(\beta|Y_i) = \left[\prod_{i=1}^{n}\left(\frac{1}{1+\exp(-\beta_0-\sum_{k=1}^{p}\beta_k x_{ik})}\right)^{y_i}\left(\frac{\exp(-\beta_0-\sum_{k=1}^{p}\beta_k x_{ik})}{1+\exp(-\beta_0-\sum_{k=1}^{p}\beta_k x_{ik})}\right)^{1-y_i}\right]\cdot\left[\prod_{k=1}^{p}\frac{\lambda}{2}\exp(-\lambda|\beta_k|)\right]$$

Untuk mendapatkan estimasi $\hat{\beta}$ diperoleh dengan

$$\hat{\beta} = \arg\max_{\beta}[\ln L(Y,\beta) + \ln p(\beta)]$$

$$\hat{\beta} = \arg\max_{\beta}\left[-\sum_{i=1}^{n}(1-y_i)\left(\beta_0 + \sum_{k=1}^{p}\beta_k x_{ik}\right) + \ln\left(1 + \exp\left(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik}\right)\right) + p\ln(\lambda) + p\ln(2) - \lambda\sum_{k=1}^{p}|\beta_k|\right] \quad (4.10)$$

Selanjutnya persamaan (4.10) diturunkan terhadap parameter β dan hasil turunannya sama dengan nol seperti dijelaskan berikut :

$$l(\beta)_c = -\sum_{i=1}^{n}\left[(1-y_i)\left(\beta_0 + \sum_{k=1}^{p}\beta_k x_{ik}\right) + \ln\left(1 + \exp\left(-\beta_0 - \sum_{k=1}^{p}\beta_k x_{ik}\right)\right)\right] + p\ln(\lambda) + p\ln(2) - \lambda\sum_{k=1}^{p}|\beta_k|$$

Turunan $l(\beta)_c$ terhadap β_0

$$\frac{\partial l(\beta)_c}{\partial\beta_0} = -\sum_{i=1}^{n}(1-y_i)(1) + \left(\frac{1}{1+\exp(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip})}\right)(\exp(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip}))(-1) = 0$$

$$= -\sum_{i=1}^{n}(1-y_i) - \left(\frac{\exp\left(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip}\right)}{1+\left(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip}\right)}\right) = 0 \quad (4.11)$$

Turunan $l(\beta)_c$ terhadap β_k ; $k = 1,2,\dots,p$

$$\frac{\partial l(\beta)_c}{\partial\beta_1} = -\left[\sum_{i=1}^{n}(1-y_i)(x_{i1}) + \left(\frac{1}{1+\exp(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip})}\right)(\exp(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip}))(-x_{i1})\right] - \lambda\frac{|\beta_1|}{\beta_1} = 0$$

$$= -\left[\sum_{i=1}^{n}(1-y_i)(x_{i1}) + \left(\frac{\exp(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip})}{1+\exp(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip})}\right)(-x_{i1})\right] - \lambda\frac{|\beta_1|}{\beta_1} = 0$$

$$\frac{\partial l(\beta)_c}{\partial\beta_2} = -\left[\sum_{i=1}^{n}(1-y_i)(x_{i2}) + \left(\frac{1}{1+\exp(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip})}\right)(\exp(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip}))(-x_{i2})\right] - \lambda\frac{|\beta_2|}{\beta_2} = 0$$

$$= -\left[\sum_{i=1}^{n}(1-y_i)(x_{i2}) + \left(\frac{\exp\left(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip}\right)}{1+\exp\left(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip}\right)}\right)(-x_{i2})\right] - \lambda\frac{|\beta_2|}{\beta_2} = 0$$

$\vdots$

$$\frac{\partial l(\beta)_c}{\partial\beta_p} = -\left[\sum_{i=1}^{n}(1-y_i)(x_{ip}) + \left(\frac{1}{1+\exp(-\beta_0-\beta_1x_{i1}\cdots-\beta_px_{ip})}\right)\right.$$

$$(\exp(-\beta_0 - \beta_1 x_{i1} \ldots - \beta_p x_{ip}))(-x_{ip})] - \lambda \frac{|\beta_p|}{\beta_p} = 0$$

$$= -\sum_{i=1}^{n}(1 - y_i)(x_{ip}) - \left(\frac{\exp(-\beta_0 - \beta_1 x_{i1} \ldots - \beta_p x_{ip})}{1 + \exp(-\beta_0 - \beta_1 x_{i1} \ldots - \beta_p x_{ip})}\right)(-x_{ip}) - \lambda \frac{|\beta_p|}{\beta_p} = 0$$

Untuk β_k ; $k = 1,2, \ldots p$ diperoleh bentuk umum sebagai berikut :

$$\frac{\partial l(\beta)_c}{\partial \beta_k} =$$

$$-\left[\sum_{i=1}^{n}(1 - y_i)(x_{ik}) + \left(\frac{1}{1 + \exp(-\beta_0 - \sum_{k=1}^{p} \beta_k x_{ik})}\right)(\exp(-\beta_0 - \sum_{k=1}^{p} \beta_k x_{ik}))(-x_{ik})\right] -$$

$$\lambda \frac{\sum_{k=1}^{p}|\beta_k|}{\sum_{k=1}^{p} \beta_k} = 0$$

$$= -\sum_{i=1}^{n}(1 - y_i)(x_{ik}) + \left(\frac{\exp(-\beta_0 - \sum_{k=1}^{p} \beta_k x_{ik})}{1 + \exp(-\beta_0 - \sum_{k=1}^{p} \beta_k x_{ik})}\right)(-x_{ik}) - \lambda \frac{\sum_{k=1}^{p}|\beta_k|}{\sum_{k=1}^{p} \beta_k} = 0 \quad (4.12)$$

Persamaan (4.11) dan persamaan (4.12) dapat dilihat bahwa nilai turunan pertama dari fungsi *negative penalizing likelihood* tidak memberikan penyelesaian karena parameter-parameternya masih saling terkait satu sama lain. Sehingga pada estimasi parameter-paremeter ini akan dilakukan pendekatan lain untuk mendapatkan hasil estimasinya. Metode yang digunakan adalah metode iterasi Newton-Raphson.

Metode Newton-Raphson adalah metode numerik untuk menyelesaikan persamaan non-linier secara iterative seperti persamaan *likelihood* yang memaksimumkan suatu fungsi. Dasar dari metode ini adalah pendekatan deret Taylor. Dalam metode Newton-Raphson dibutuhkan turunan pertama dan kedua dari fungsi *log-likelihood*nya.

Taksiran parameter dapat ditunjukkan dalam bentuk :

$$\begin{bmatrix} \hat{\beta}_{0(t+1)} \\ \hat{\beta}_{1(t+1)} \\ \hat{\beta}_{2(t+1)} \\ \vdots \\ \hat{\beta}_{k(t+1)} \end{bmatrix} = \begin{bmatrix} \hat{\beta}_{0(t)} \\ \hat{\beta}_{1(t)} \\ \hat{\beta}_{2(t)} \\ \vdots \\ \hat{\beta}_{k(t)} \end{bmatrix} - \boldsymbol{D}^{-1}\boldsymbol{d} \quad (4.13)$$

dengan $\boldsymbol{d} = \frac{\partial l}{\partial(\boldsymbol{\beta}_t)}$ dan $\boldsymbol{D}(\boldsymbol{\beta}_t) = \frac{\partial^2 l}{\partial(\boldsymbol{\beta}_t)(\boldsymbol{\beta}_t)^t}$, dimana $\boldsymbol{\beta}_t = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}$, dan nilai $\hat{\beta}_t$ pada saat $t = 1$ adalah $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$.

Nilai entri-entri dari matriks $\boldsymbol{d}$ sudah ditunjukkan pada persamaan (4.11) dan persamaan (4.12). Sedangkan entri-entri dari matriks $\boldsymbol{D}$ dapat dicari dengan menurunkan persamaan (4.11) dan persamaan (4.12) terhadap β_0 dan $\beta_k = 1,2,..p$. Dengan menurunkan persamaan (4.11) terhadap β_0 diperoleh :

$$\frac{\partial^2 l(\beta)_c}{(\partial\beta_0)^2} = \frac{\begin{gathered}(1+\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip}))(\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip}))(-1)- \\ (\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip}))(-1)(\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip}))\end{gathered}}{\left(1+\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip})\right)^2}$$

Selanjutnya persamaan (4.11) diturunkan terhadap $\beta_k = 1,2,..p$, sehingga diperoleh :

$$\frac{\partial^2 l(\beta)_c}{\partial\beta_0\partial\beta_1} = \frac{\partial^2 l(\beta)_c}{\partial\beta_1\partial\beta_0} =$$

$$\frac{\begin{gathered}(1+\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip}))(\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip}))(-x_{i1})- \\ (\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip}))(-x_{i1})(\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip}))\end{gathered}}{\left(1+\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip})\right)^2}$$

$\vdots$

$$\frac{\partial^2 l(\beta)_c}{\partial\beta_0\partial\beta_p} = \frac{\partial^2 l(\beta)_c}{\partial\beta_p\partial\beta_0}$$

$$= \frac{\begin{gathered}(1+\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip}))(\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip}))(-x_{ip})- \\ (\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip}))(-x_{ip})(\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip}))\end{gathered}}{\left(1+\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip})\right)^2}$$

Selanjutnya dengan menurunkan persamaan (4.12) terhadap $\beta_k = 1,2,3,..p$, diperoleh :

$$\frac{\partial^2 l(\beta)_c}{(\partial\beta_1)^2}$$

$$= \frac{\begin{gathered}(-x_{i1})(1+\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip}))(\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip}))(-x_{i1})- \\ (\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip}))(-x_{i1})(\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip}))(-x_{i1})\end{gathered}}{\left(1+\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip})\right)^2}$$

$$\frac{\partial^2 l(\beta)_c}{\partial\beta_1\partial\beta_2} = \frac{\partial^2 l(\beta)_c}{\partial\beta_2\partial\beta_1}$$

$$= \frac{\begin{gathered}(-x_{i1})(1+\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip}))(\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip}))(-x_{i2})- \\ (\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip}))(-x_{i2})(\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip}))(-x_{i1})\end{gathered}}{\left(1+\exp(-\beta_0-\beta_1 x_{i1}-\cdots-\beta_p x_{ip})\right)^2}$$

$\vdots$

$$\frac{\partial^2 l(\beta)_c}{\partial\beta_1\partial\beta_p}=\frac{\partial^2 l(\beta)_c}{\partial\beta_p\partial\beta_1}$$

$$=\frac{\begin{array}{c}(-x_{i1})\left(1+\exp\left(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip}\right)\right)\left(\exp\left(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip}\right)\right)(-x_{ip})-\\ \left(\exp\left(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip}\right)\right)(-x_{ip})\left(\exp\left(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip}\right)\right)(-x_{i1})\end{array}}{\left(1+\exp\left(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip}\right)\right)^2}$$

$\vdots$

$$\frac{\partial^2 l(\beta)_c}{(\partial\beta_p)^2}$$

$$=\frac{\begin{array}{c}(-x_{ip})\left(1+\exp\left(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip}\right)\right)\left(\exp\left(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip}\right)\right)(-x_{ip})-\\ \left(\exp\left(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip}\right)\right)(-x_{ip})\left(\exp\left(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip}\right)\right)(-x_{ip})\end{array}}{\left(1+\exp\left(-\beta_0-\beta_1x_{i1}-\cdots-\beta_px_{ip}\right)\right)^2}$$

Berdasarkan persamaan diatas dapat diperoleh bentuk umum estimasi parameter dengan metode Newton-Raphson yaitu :

$$\begin{bmatrix}\hat\beta_{0(t+1)}\\ \hat\beta_{1(t+1)}\\ \hat\beta_{2(t+1)}\\ \vdots\\ \hat\beta_{p(t+1)}\end{bmatrix}=\begin{bmatrix}\hat\beta_{0(t)}\\ \hat\beta_{1(t)}\\ \hat\beta_{2(t)}\\ \vdots\\ \hat\beta_{p(t)}\end{bmatrix}-\begin{bmatrix}\frac{\partial^2 l(\beta)_c}{(\partial\beta_0)^2} & \frac{\partial^2 l(\beta)_c}{\partial\beta_0\partial\beta_1} & \frac{\partial^2 l(\beta)_c}{\partial\beta_0\partial\beta_2} & \cdots & \frac{\partial^2 l(\beta)_c}{\partial\beta_0\partial\beta_p}\\ \frac{\partial^2 l(\beta)_c}{\partial\beta_1\partial\beta_0} & \frac{\partial^2 l(\beta)_c}{(\partial\beta_1)^2} & \frac{\partial^2 l(\beta)_c}{\partial\beta_1\partial\beta_2} & \cdots & \frac{\partial^2 l(\beta)_c}{\partial\beta_1\partial\beta_p}\\ \frac{\partial^2 l(\beta)_c}{\partial\beta_2\partial\beta_0} & \frac{\partial^2 l(\beta)_c}{\partial\beta_2\partial\beta_1} & \frac{\partial^2 l(\beta)_c}{(\partial\beta_2)^2} & \cdots & \frac{\partial^2 l(\beta)_c}{\partial\beta_2\partial\beta_p}\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ \frac{\partial^2 l(\beta)_c}{\partial\beta_p\partial\beta_0} & \frac{\partial^2 l(\beta)_c}{\partial\beta_p\partial\beta_1} & \frac{\partial^2 l(\beta)_c}{\partial\beta_p\partial\beta_2} & \cdots & \frac{\partial^2 l(\beta)_c}{(\partial\beta_p)^2}\end{bmatrix}^{-1}\begin{bmatrix}\frac{\partial l(\beta)_c}{\partial\beta_0}\\ \frac{\partial l(\beta)_c}{\partial\beta_1}\\ \frac{\partial l(\beta)_c}{\partial\beta_2}\\ \vdots\\ \frac{\partial l(\beta)_c}{\partial\beta_p}\end{bmatrix} \quad (4.14)$$

dengan nilai entri-entri dari matriks, maktris turunan pertama **(d)** ditunjukkan pada turunan pertama, serta matriks turunan kedua **(D)** ditunjukkan pada turunan kedua. Perhitungan nilai-nilai dari $\beta_k = 1,2,3,\dots,p$, dengan metode Newton Raphson yang ditunjukkan pada persamaan (4.14) .

4.2. Penerapan Metode *Least Absolute Shrinkage and Selection Operator* pada data Kemiskinan

Contoh penerapan regresi logistik adalah data kemiskinan setiap desa Kabupaten Jeneponto Tahun 2013 dari publikasi Badan Pusat Statistika (BPS) Jeneponto dengan memperhitungkan variabel bebas berupa sumber penerangan yaitu PLN(X_1), bukan PLN(X_2) dan Pelita(X_3), serta sumber mata pencaharian yaitu petani(X_4), pedagang(X_5), dibidang industri(X_6) dan dibidang angkutan(X_7). Selengkapnya data yang digunakan ditunjukkan pada Lampiran 1 dan Lampiran 2.

Data kemiskinan menunjukkan bahwa data memenuhi asumsi distribusi Bernoulli karena variabel respon bernilai antara 0 atau 1.

Pemeriksaan hubungan antara variabel bebas melalui matriks korelasi adalah sebagai berikut :

Tabel 4.1. Nilai korelasi antara variabel bebas.

	Y	X1	X2	X3	X4	X5	X6	X7
Y	1	-0,079	0,153	-0,252	0,024	-0,006	0,109	0,012
X1	-0,079	1	-0,964	-0,334	-0,158	0	0,201	0,168
X2	0,153	-0,964	1	0,072	0,155	0,024	-0,162	-0,212
X3	-0,252	-0,334	0,072	1	0,047	-0,091	-0,179	0,121
X4	0,024	-0,158	0,155	0,047	1	-0,756	-0,444	-0,691
X5	-0,006	0	0,024	-0,091	-0,756	1	0,116	0,167
X6	0,109	0,201	-0,162	-0,179	-0,444	0,116	1	0,093
X7	0,012	0,168	-0,212	0,121	-0,691	0,167	0,093	1

Sumber : Data diolah, 2015.

Berdasarkan Tabel 4.1. terlihat ada beberapa pasang variabel bebas yang berkorelasi tinggi yaitu X_1-X_2, X_4-X_5 dan X_4-X_7 dikarenakan nilai korelasi lebih besar dari 0,5. Hal tersebut menunjukkan bahwa adanya multikolinearitas antara variabel bebas yaitu korelasi tinggi antara beberapa variabel bebas. Sehingga metode *LASSO* dapat diterapkan pada data ini. Namun pengujian awal dilakukan dengan menggunakan regresi logistik biner dan melalui uji kelayakan model.

Tabel 4.2. Hasil estimasi parameter pada data kemiskinan Kabupaten Jeneponto dengan regresi logistik biner.

Variabel Bebas	Koefisien Estimasi	SE	p-Value
1	2	3	4
Intersept	1,549	0,271	5,717
X_1	-32,415	30,052	-1,079
X_2	-30,081	28,404	-1,059
X_3	-9,390	8,047	-1,167
X_4	51,177	109,334	0,468
X_5	31,852	68,415	0,466
X_6	16,076	35,008	0,459
X_7	28,785	60,915	0,473
Uji deviance 8,850		Sig. 0,287	

Sumber : Data diolah, 2015.

Berdasarkan Tabel 4.2. diperoleh bahwa tidak ada variabel bebas yang signifikan mempengaruhi variabel respon hal ini ditunjukkan oleh tidak adanya nilai *p-value* yang lebih besar dari nilai $\alpha = 0{,}05$. Nilai signifikan uji deviance yang ditunjukkan pada Tabel 4.2. adalah $0{,}287 > \alpha = 0{,}05$, ini berarti bahwa model regresi logistik biner tidak layak untuk digunakan. Ini menunjukkan bahwa hasil estimasi regresi logistik biner kurang bagus akibat adanya multikolinearitas.

Selanjutnya data dianalisis menggunakan regresi logistik biner dengan metode *LASSO* melalui bantuan software R i386 3.1.2 package Lasso2.

LASSO memiliki kendala yaitu $\lambda \sum_{k=1}^{p} |\beta_k|$, dengan $\lambda \geq 0$ yang merupakan parameter *tuning* pada *LASSO*. Parameter λ yang digunakan merupakan nilai minimum dari GCV. Berdasarkan Lampiran 5 nilai λ yang diperoleh dari proses *LASSO* adalah 0,0513505. Kendala *LASSO* terpenuhi karena nilai λ yang diperoleh telah memenuhi syarat $\lambda \geq 0$. Selanjutnya dilakukan proses iterasi sampai konvergen, sehingga diperoleh hasil estimasi parameter dari regresi logistik biner dengan metode *LASSO*, yaitu :

Tabel 4.3. Hasil estimasi parameter dengan regresi logistik biner melalui metode *Least Absolute Shrinkage and Selection Operator*.

Parameter	Koefisien Estimasi	Standar Error
1	2	3
Intersept	1,5324	
X_1	-0,8307	0,0645
X_2	-0,2290	0,0608
X_3	-0,9354	0,0172
X_4	0,0000	0,3045
X_5	-0,1031	0,1905
X_6	-0,3258	0,0975
X_7	0,2810	0,1696

Sumber : Data diolah, 2015.

Berdasarkan hasil estimasi metode *LASSO* yang ditunjukkan pada Tabel 4.3. diperoleh nilai standar error koefisien regresi logistik *LASSO* lebih kecil dibandingkan standar error koefisien regresi logistik biner, dengan keakuratan pengklasifikasian sebesar 80,53% yang ditunjukkan pada Tabel 4.4.

4.2.1. Uji Kelayakan Model

Keakuratan klasifikasi dari suatu analisis regresi dapat diketahui melalui tabel klasifikasi. Tabel klasifikasi merupakan cara lain untuk menyatakan kelayakan suatu model yaitu seberapa besar persentase observasi secara tepat diklasifikasikan.

Tabel 4.4. Ketepatan Klasifikasi Kemiskinan di Kabupaten Jeneponto

Observasi	Prediksi		Total	% Ketepatan
	Tidak Miskin	Miskin		
Tidak Miskin	4	19	23	17,39%
Miskin	3	87	90	96,67%
Total	7	106	113	
% Ketepatan Keseluruhan				80,53%

Sumber : Data diolah, 2015.

Berdasarkan Tabel 4.4. diperoleh persentase ketepatan keseluruhan model dalam mengklasifikasikan observasi adalah 80,53%. Artinya dari 113 observasi, ada 91 yang tepat pengklasifikasiannya berdasarkan model regresi logistik *LASSO*. Jumlah observasi yang tepat pengklasifikasiannya dapat dilihat pada diagonal utama. Persentase ketepatan klasifikasi desa tidak miskin di Kabupaten Jeneponto adalah 17,39%, artinya dari 23 desa klasifikasi tidak miskin, ada 4 desa yang terklasifikasi dengan tepat. Sedangkan persentase ketepatan klasifikasi desa miskin di Kabupaten Jeneponto sebesar 96,67% yang berarti bahwa dari 90 desa klasifikasi miskin, ada 87 desa yang terklasifikasi dengan benar. Sehingga dapat dihitung besarnya nilai misklasifikasi berdasarkan Tabel 4.4. adalah :

$$APER = \left(\frac{3}{113} + \frac{19}{113}\right) \times 100\% = 19{,}47\%$$

Besarnya nilai misklasifikasi (APER) adalah 19,47%. Ukuran dari ketepatan klasifikasi yaitu jika nilai misklasifikasi lebih kecil dari nilai ketepatan klasifikasi maka dapat dikatakan pengklasifikasian sudah tepat.

Selanjutnya dilakukan uji signifikansi parameter, seperti uraian berikut :

4.2.2. Uji Signifikansi Parameter

Pengujian signifikansi parameter dilakukan melalui uji parsial. Uji parsial pada masing-masing koefisien regresi melalui uji Wald dengan hipotesis sebagai berikut:

$H_0 : \beta_k = 0$, dengan $k = 1,2,\dots,p$ (tidak ada pengaruh variabel bebas ke-k terhadap variabel respon)

$H_1 : \beta_k \neq 0$, dengan $k = 1,2,\dots,p$ (ada pengaruh variabel bebas ke-k terhadap variabel respon)

Berdasarkan Persamaan (2.13) yaitu $W = \left[\frac{\hat{\beta}_k}{se(\hat{\beta}_k)}\right]^2$ diperoleh hasil uji parameter secara parsial pada Tabel 4.5. sebagai berikut:

Tabel 4.5. Pengujian signifikansi parameter secara parsial

Parameter	Parameter	Wald	$\mathcal{X}^2_{0,05;1}$	Kesimpulan
1	2	3	4	5
X_1	-0,8307	166,0720	3,841	H_0 ditolak
X_2	-0,229	14,1402		H_0 ditolak
X_3	-0,9354	2970,685		H_0 ditolak
X_4	0,000	0,000		H_0 diterima
X_5	-0,1031	0,293		H_0 diterima
X_6	-0,3258	11,175		H_0 ditolak
X_7	0,281	2,745		H_0 diterima

Sumber : Data diolah, 2015.

Berdasarkan Tabel 4.5. diperoleh H_0 ditolak untuk beberapa koefisien regresi karena nilai statistik uji Wald $> \mathcal{X}^2_{0,05;1}$(nilai chikuadrat tabel) artinya terdapat 4 variabel bebas yang berpengaruh signifikan terhadap variabel respon yaitu sumber penerangan berupa PLN(X_1), sumber penerangan berupa bukan PLN(X_2), sumber penerangan berupa Pelita/Petromaks(X_3) dan sumber mata pencaharian dibidang industri(X_6) yang berpengaruh signifikan terhadap klasifikasi kemiskinan setiap desa di Kabupaten Jeneponto. Sehingga estimasi model regresi logistik pada data kemiskinan Kabupaten Jeneponto dengan variabel bebas yang signifikan diperoleh sebagai berikut :

$$\hat{Y}_i = \frac{1}{1+\exp(1{,}5324-(-0{,}971)x_1-(-0{,}425)x_2-(-0{,}921)x_3-(-0{,}321)x_6)} \quad (4.15)$$

Berdasarkan persamaan (4.15) maka dapat dijelaskan jika semakin sedikit rumah tangga yang menggunakan sumber penerangan berupa PLN dan semakin banyak rumah tangga yang menggunakan sumber penerangan berupa bukan PLN dan Pelita atau Petromaks serta jumlah kepala rumah tangga dengan sumber mata pencaharian dibidang industri semakin sedikit, maka desa tersebut diprediksi sebagai kategori desa miskin di Kabupaten Jeneponto. Sebaliknya jika rumah tangga yang menggunakan sumber penerangan berupa PLN semakin banyak dan rumah tangga yang menggunakan sumber penerangan berupa bukan PLN dan Pelita atau Petromaks semakin sedikit serta jumlah kepala rumah tangga dengan sumber mata pencaharian dibidang industri semakin banyak, maka desa tersebut diprediksi sebagai desa tidak miskin di Kabupaten Jeneponto. Misalkan jika suatu desa terdapat 100% jumlah rumah tangga yang menggunakan sumber penerangan berupa PLN dan 0% jumlah rumah tangga yang menggunakan sumber penerangan berupa bukan PLN dan Pelita/Petromaks serta 16% jumlah kepala keluarga yang bermata pencaharian dibidang industri, maka desa tersebut diprediksi sebagai kategori sebagai desa tidak miskin di Kabupaten Jeneponto. Sebaliknya jika suatu desa terdapat 26% jumlah rumah tangga yang menggunakan sumber penerangan berupa PLN, 57% jumlah rumah tangga yang menggunakan sumber penerangan berupa bukan PLN dan 18% jumlah rumah tangga yang menggunakan sumber penerangan berupa Pelita/Petromaks serta 3% jumlah kepala keluarga yang bermata pencaharian dibidang industri, maka desa tersebut diprediksi sebagai kategori sebagai desa miskin di Kabupaten Jeneponto.

BAB V
PENUTUP

5.1. Kesimpulan

Berdasarkan uraian pembahasan mengenai estimasi koefisien regresi logistik biner dengan metode *Least Absolute Shrinkage and Selection Operator*, maka dapat diambil kesimpulan sebagai berikut :

1. Taksiran model data kemiskinan setiap desa di Kabupaten Jeneponto melalui pendekatan regresi logistik *LASSO* adalah

$$\hat{Y}_i = \frac{1}{1+\exp(1{,}5324-(-0{,}971)x_1-(-0{,}425)x_2-(-0{,}921)x_3-(-0{,}321)x_6)}$$

   dengan menggunakan metode *Least Absolute Shrinkage and Selection Operator* diperoleh keakuratan pengklasifikasian yaitu sebesar 80,53%.
2. Berdasarkan penerapan regresi logistik biner dengan metode *Least Absolute Shrinkage and Selection Operator* data kemiskinan setiap desa di Kabupaten Jeneponto dipengaruhi oleh beberapa faktor yang signifikan yaitu sumber penerangan berupa PLN, bukan PLN dan Petromaks atau Pelita serta sumber mata pencaharian berupa pekerjaan dibidang industri.

5.2. Saran

Regresi logistik biner merupakan model regresi yang variabel responya berdistribusi Bernoulli dengan nilai yang terdiri dari dua kategori yaitu 0 dan 1. Penulis menyarankan untuk mengestimasi koefisien regresi dengan metode *Least Absolute Shrinkage and Selection Operator* pada regresi logistik multinomial dan ordinal serta dapat diterapkan pada model GLM (*Generalized Linear Model*) yang lain seperti regresi poisson.

Metode *Least Absolute Shrinkage and Selection Operator* ini dapat dikembangkan dalam bentuk kelompok LASSO.

DAFTAR PUSTAKA

Agresti,A. 1990. *Categorical Data Analysis*. John Wiley & Sons. New York.

[BPS]Badan Pusat Statistik. 2014.Publikasi Badan Pusat Statistik Kabupaten Jeneponto. (http://www.jenepontokab.bps.go.id/index.php?hal=publikasi). [18 Januari 2015].

Buhlmann P, Sara van de G. 2010. *Statistics for High Dimensional Data: Methods, Theory and Aplications*. New York: Springer.

Dewi YS. 2010. *OLS, LASSO dan PLS pada Data Mengandung Multikolinearitas*. Jurnal Ilmu Dasar 11(1); 83-91.

Farmani KD, Kencana IP, Sukarsa KG. 2012. *Perbandingan analisis least absolute shrinkage and selection operator dan partial least squares (Studi Kasus: Data Microarray)*. *e-Jurnal Matematika* 1(1): 75-80.

Hastie T, Tibshirani R, Friedman J. 2008. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction.* Ed ke-2. New York: Springer.

Hosmer, D.W. dan Lemeshow, S. 2000. *Applied Logistic Regression*. by John Wiley & Sons, Inc. USA.

Izenman AJ. 2008. *Modern Multivariate Statistical Techniques:Regression, Classification, and Manifold Learning*. New York: Springer.

Lee S., Lee H., Abeel P. and Ng A. (2006): Efficient $L1$ – regularized logistic regression,*Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-06)*.

Mendenhall W, Sincich T. 2003. *Regression Analysis. Sixth Edition*. Florida. Pearson Education Internasional.

Osborne, M., B. Presnell, and B. Turlach (2000). *On the LASSO and its dual.* Journal of Computational and Graphical Statistics 9, 319–337.

Pusporini Arum, Aunuddin, La Ode AR. 2012. *Penerapan Regresi Gulud dan Least Absolute Shrinkage and Selection Operator (LASSO) dalam Penyusutan Koefisien Regresi.* Jurnal Statistika. IPB.

Roth, V. 2004. *The Generalized LASSO. IEEE Trans. Neur. Networks*,15,16-28.

Ruslan. 2013. Skripsi. *Penggunaan Analisis Komponen Utama pada Regresi Logistik Biner*. FMIPA Universitas Hasanuddin; Makassar.

Susilo, Edi. 2014. Skripsi. *Model Regresi Logistik Biner dengan Metode Penalized Maximum Likelihood*. FMIPA Universitas Hasanuddin; Makassar.

Tibshirani R. 1996. *Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B* 58(1): 267-288.

Tibshirani R. 1997. *The Lasso Method for variable selection in the cox model. Statist. Med.*, 16; 385-395.