



**PENGEMBANGAN DAN APLIKASI GUI WEB INTERAKTIF
DENGAN R-SHINY UNTUK PEMODELAN LOGISTIK
MULTINOMIAL**

SKRIPSI

Oleh

**Dyah Kiki Langit Cahyani
NIM 111810101019**

**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2015**



**PENGEMBANGAN DAN APLIKASI GUI WEB INTERAKTIF
DENGAN R-SHINY UNTUK PEMODELAN LOGISTIK
MULTINOMIAL**

SKRIPSI

diajukan guna melengkapi tugas akhir dan memenuhi salah satu syarat
untuk menyelesaikan Program Studi Matematika (S1)
dan mencapai gelar Sarjana Sains

Oleh

**Dyah Kiki Langit Cahyani
NIM 1118101019**

**JURUSAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS JEMBER
2015**

PERSEMBAHAN

Skripsi ini saya persembahkan untuk :

1. Allah SWT yang telah memberikan semua kemudahan dan kesempurnaan dalam kehidupan ini;
2. Bapak Yon Misyono dan Ibu Dian Nur Rohaniah yang telah memberikan doa, cinta, kasih dan semangatnya;
3. adik tercinta Jagat Putra Pamungkas telah memberikan dukungan; serta Boni Muhardiyanto yang telah banyak memberikan bantuan dan semangat dengan penuh kesabaran;
4. seluruh guru dan dosen sejak taman kanak-kanak hingga perguruan tinggi yang telah memberikan ilmu dan membimbing dengan penuh perhatian dan kesabaran;
5. Almamater Jurusan Matematika FMIPA Universitas Jember, SMA N 1 Genteng, SMP N 1 Srono, SDN 2 Blambangan, TK Dharma Wanita.

MOTTO

“Orang yang menuntut ilmu berarti menuntut rahmat. Orang yang menuntut ilmu berarti menjalankan rukun Islam dan pahala yang diberikan sama dengan para Nabi”
(HR. Dailani dari Anas r.a)^{*)}

"Sesuatu yang belum dikerjakan, seringkali tampak mustahil. Kita baru yakin kalau kita telah berhasil melakukannya dengan baik."
(Evelyn Underhill)^{*)}

^{*)} <http://www.maribelajarbku.web.id/2015/03/contoh-motto-terbaru-dalam-skripsi.html>

PERNYATAAN

Saya yang bertanda tangan di bawah ini :

Nama : Dyah Kiki Langit Cahyani

NIM : 111810101019

menyatakan dengan sesungguhnya bahwa karya ilmiah yang berjudul “Pengembangan dan Aplikasi GUI Web Interaktif dengan R-shiny untuk Pemodelan Logistik Multinomial” adalah benar-benar hasil karya sendiri, kecuali kutipan yang sudah saya sebutkan sumbernya, belum pernah diajukan dalam institusi manapun dan juga bukan karya jiplakan. Saya bertanggung jawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenarnya tanpa ada tekanan dan paksaan dari pihak manapun serta bersedia mendapat sanksi akademik jika ternyata di kemudian hari pernyataan ini tidak benar.

Jember, November 2015

Yang menyatakan,

Dyah Kiki langit Cahyani
NIM 111810101019

SKRIPSI

**PENGEMBANGAN DAN APLIKASI GUI WEB INTERAKTIF
DENGAN R-SHINY UNTUK PEMODELAN LOGISTIK
MULTINOMIAL**

Oleh

**Dyah Kiki Langit Cahyani
NIM 1118101019**

Pembimbing

Dosen Pembimbing Utama : Prof. Drs. I Made Tirta M. Sc., Ph. D.

Dosen Pembimbing Anggota : Dian Anggraeni, S.Si., M.Si.

PENGESAHAN

Skripsi berjudul “Pengembangan dan Aplikasi GUI Web Interaktif dengan R-shiny untuk Pemodelan Logistik Multinomial” telah diuji dan disahkan pada :

Hari, tanggal :

Tempat : Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas
Jember

Tim penguji,

Ketua,

Sekretaris,

Prof. Drs. I Made Tirta M. Sc., Ph. D.
NIP. 195912201985031002

Dian Anggraeni, S. Si., M. Si.
NIP. 198202162006042002

Anggota I,

Anggota II,

Dr. Alfian Futuhul Hadi, S. Si., M. Si.
NIP. 197407192000121001

Kosala Dwidja Purnomo, S. Si., M. Si.
NIP. 196908281998021001

Mengesahkan

Dekan,

Prof. Drs. Kusno, DEA, Ph.D.
NIP 196101081986021001

RINGKASAN

Pengembangan dan Aplikasi GUI Web Interaktif dengan R-shiny untuk Pemodelan Logistik Multinomial; Dyah Kiki Langit Cahyani, 111810101019; 2015; 36 halaman; Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember.

Logistik multinomial merupakan regresi logistik yang variabel dependennya mempunyai skala yang bersifat *polychotomus* atau multinomial. Skala multinomial adalah suatu pengukuran yang dikategorikan menjadi lebih dari dua kategori. Logistik multinomial banyak digunakan dan bermanfaat karena banyak ditemui kasus yang menggunakan logistik multinomial untuk menganalisis data. Namun kendalanya akses untuk menggunakan analisis logistik multinomial terhalang karena menggunakan program yang susah bagi pengguna atau peneliti yang tidak mengerti program.

Program yang digunakan adalah R dan paket R yang digunakan untuk analisis data dengan logistik multinomial ini adalah paket *nnet* dengan fungsi `multinom()`. Salah satu program dalam R yang bisa membuat menu *user interface* yang berbasis web agar mudah diakses yaitu R-shiny. R-shiny merupakan *toolkit* dari program R yang dapat digunakan untuk membuat program secara *online*. Komponen dalam R-shiny ada 2 kelompok yaitu *user interface* dan *server*. *User interface* berupa file HTML yang digunakan untuk memasukkan nilai *input* dan penyajian *output*. Program dibentuk dalam bentuk tutorial yang meliputi ringkasan teori dan pengaplikasiannya. Analisis data menggunakan program R yang telah dibuat dalam bentuk web interaktif berbasis R-shiny. Program online ini dapat diakses di alamat <http://statslab-rshiny.fmipa.unej.ac.id/JORS/logmult/>. Program ini selain dapat digunakan untuk

data yang mempunyai variabel respon y lebih dari 2 kategori, juga dapat digunakan untuk data yang mempunyai variabel respon y yang mempunyai 2 kategori. Dalam program ini juga ada program untuk GLM apabila pengguna ingin membedakan hasil dari fungsi `multinom()` dengan fungsi `glm()` untuk data yang mempunyai variabel y 2 kategori.

Data yang digunakan adalah data simulasi yang mempunyai variabel prediktor x_1, x_2, x_3, x_4 dan variabel respon y yang memiliki lebih dari 2 kategori berupa A, B, C. Uji signifikansi menggunakan p - value dengan memilih nilai p - value ≤ 0.05 . Sedangkan uji kesesuaian model menggunakan nilai AIC dan BIC. Hasil analisis data diperoleh nilai AIC sebesar 121,0870 dan BIC sebesar 144,9073 untuk model formula $y \sim x_1 + x_2 + x_3 + x_4$ lebih kecil dibandingkan dengan nilai AIC sebesar 129,4233 dan BIC sebesar 148,4795 model formula $y \sim x_1 + x_2 + x_4$. Hasil tersebut menunjukkan bahwa model formula $y \sim x_1 + x_2 + x_3 + x_4$ lebih baik daripada model formula $y \sim x_1 + x_2 + x_4$. Jadi data simulasi dengan model terbaik yang didapat yaitu $y \sim x_1 + x_2 + x_3 + x_4$ dengan

$$g_1(x) = 3,1180 - 0,2205x_1 + 0,2078x_2 - 0,1231x_3 - 1,0677x_4 [P]$$

$$g_2(x) = -5,1752 - 0,1597x_1 + 0,2176x_2 + 0,1235x_3 - 0,6315x_4 [P]$$

PRAKATA

Puji syukur kehadirat Allah SWT atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Pengembangan dan Aplikasi GUI Web Interaktif dengan R-shiny untuk Pemodelan Logistik Multinomial Berbasis”. Skripsi ini disusun untuk memenuhi salah satu syarat untuk menyelesaikan pendidikan strata satu (S1) di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jember. Sholawat serta salam semoga selalu tercurahkan keharibaan beliau nabi Muhammad SAW yang telah menjadi pembawa rahmatan lil’alamin. Penyusunan skripsi ini tidak lepas dari bantuan berbagai pihak, secara langsung maupun tidak langsung. Penulis menyampaikan ucapan terima kasih kepada:

1. Prof. Drs. I Made Tirta, M.Sc., Ph.D. selaku Dosen Pembimbing Utama dan Dian Anggraeni, S.Si., M.Si. selaku Dosen Pembimbing Anggota yang telah meluangkan waktu, memberikan saran dan bimbingannya dalam penulisan skripsi ini;
2. Dr. Alfian Futuhul Hadi, S.Si., M.Si. selaku Dosen Penguji I dan Kosala Dwidja Purnomo, S. Si., M. Si. selaku Dosen Penguji II yang telah memberi masukan berupa saran dan kritiknya terhadap skripsi ini;
3. seluruh dosen dan karyawan Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam yang telah memberikan ilmu serta fasilitas yang membantu selama proses perkuliahan berlangsung;
4. sahabat kosan Kalimantan no 12 khususnya Mbak Amel, Mbak Denok, Indah, Alni, dan Juwita yang selalu memberikan doa, saran, dan motivasinya selama ini;
5. sahabat-sahabatku seperjuangan KRAMAT’11, khususnya Upik, Sella, Eka;
6. semua pihak yang tidak dapat disebutkan satu per satu.

Penulis juga menerima segala kritik dan saran dari semua pihak demi kesempurnaan skripsi ini. Akhirnya penulis berharap, semoga skripsi ini dapat bermanfaat bagi semua pihak.



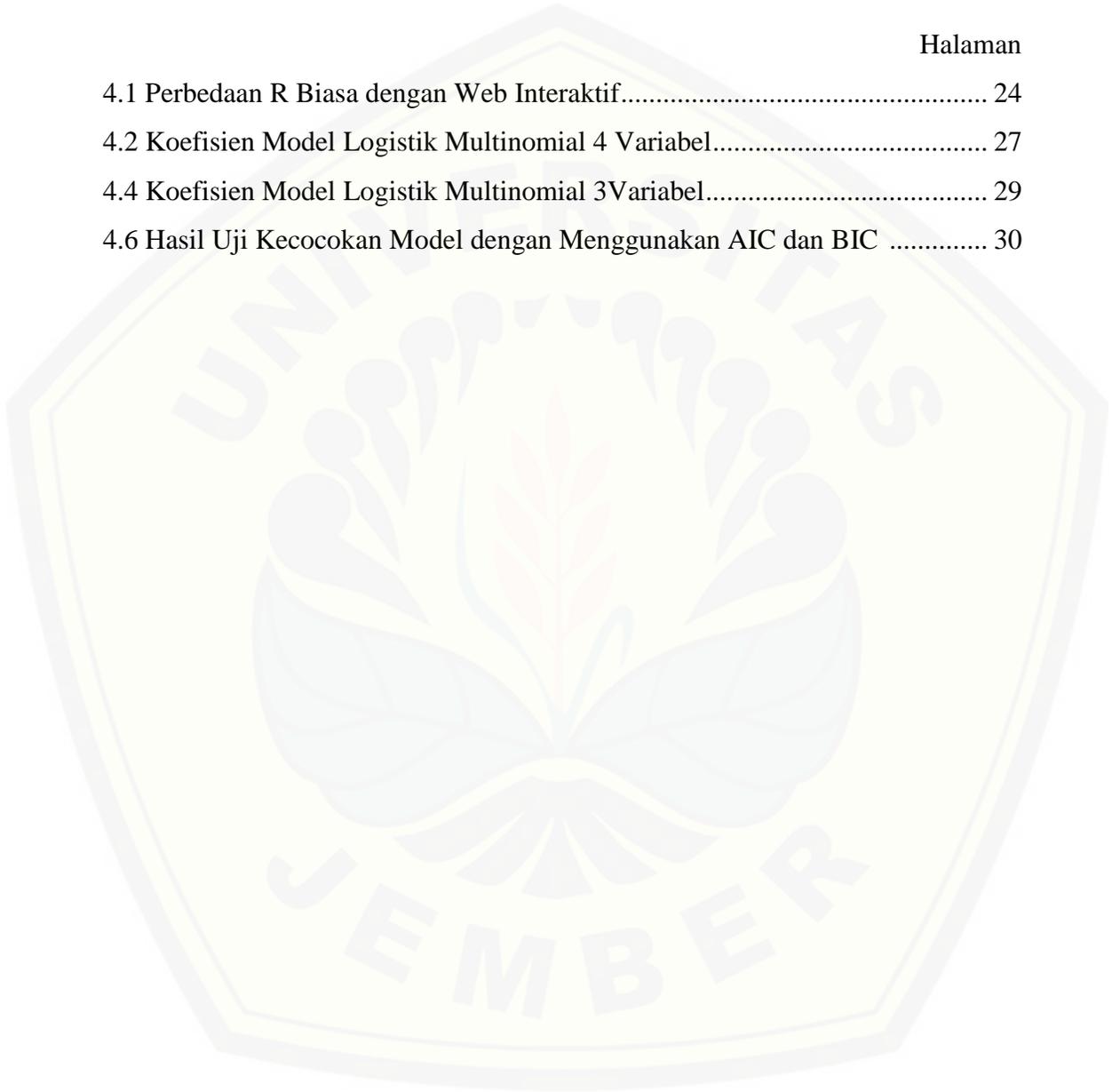
DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
PERSEMBAHAN	ii
MOTTO	iii
PERNYATAAN	iv
HALAMAN BIMBINGAN	v
PENGESAHAN	vi
RINGKASAN	vii
PRAKATA	ix
DAFTAR ISI	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xv
BAB 1. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.2 Batasan Masalah	3
1.3 Tujuan	3
1.4 Manfaat	3
BAB 2. TINJAUAN PUSTAKA	4
2.1 Model Linear	4
2.1.1 Model Linear Sederhana	4
2.1.2 Model Linear Tergeneralisir	4
2.2 Regresi Logistik	5
2.3 Distribusi Multinomial	7
2.4 Regresi Logistik Multinomial	7

2.5 Estimasi Parameter	9
2.6 Uji Kesesuaian Model	10
2.6.2 AIC (<i>Akaike's Information Criterion</i>)	10
2.6.1 BIC (<i>Bayesian Information Criterion</i>).....	10
2.7 R-Shiny.....	10
BAB 3. METODOLOGI PENELITIAN.....	14
BAB 4. HASIL DAN PEMBAHASAN	17
4.1 Hasil dan Pembahasan Fitur Web.....	17
4.1.1 Penyusunan Web.....	17
4.1.2 Tampilan Web.....	20
4.2 Ilustrasi Penggunaan Web Interaktif dengan R-shiny untuk Logistik Multinomial.....	25
4.2.1 Analisis dengan Model Formula $y \sim x_1 + x_2 + x_3 + x_4$	26
4.2.2 Analisis dengan Model Formula $y \sim x_1 + x_2 + x_4$	29
4.2.3 Uji Kesesuaian Model	28
4.2.4 Prediksi Peluang Model	30
BAB 5. PENUTUP.....	33
5.1 Kesimpulan	33
5.2 Saran	34
DAFTAR SUMBER RUJUKAN	35
LAMPIRAN.....	37

DAFTAR TABEL

	Halaman
4.1 Perbedaan R Biasa dengan Web Interaktif.....	24
4.2 Koefisien Model Logistik Multinomial 4 Variabel.....	27
4.4 Koefisien Model Logistik Multinomial 3 Variabel.....	29
4.6 Hasil Uji Kecocokan Model dengan Menggunakan AIC dan BIC	30

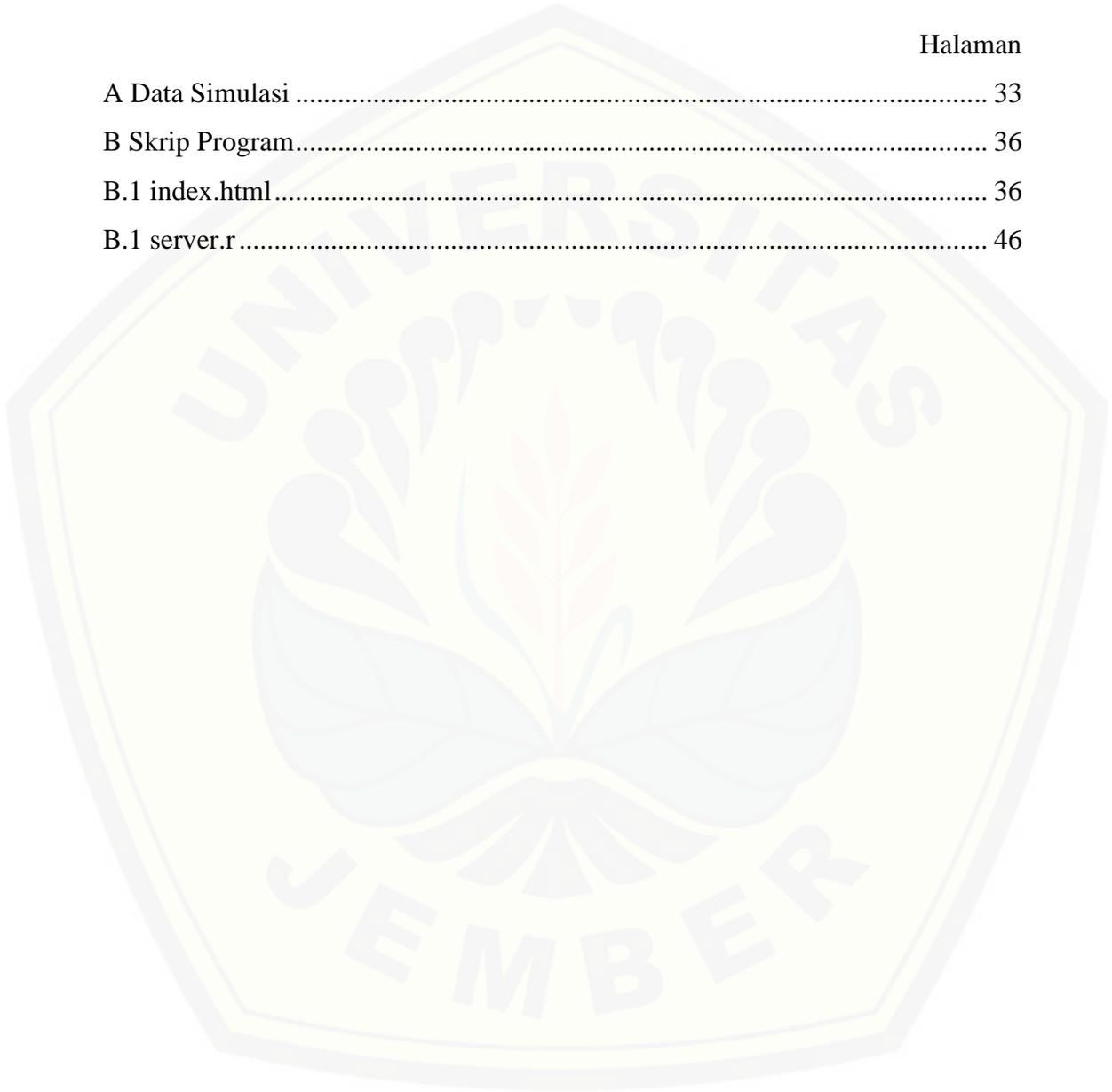


DAFTAR GAMBAR

	Halaman
2.1 Struktur Komponen R-Shiny.....	12
3.1 Diagram Alir Langkah Penelitian	16
4.1 Tampilan Awal.....	21
4.2 Pilih Data dan <i>Summary Data</i>	21
4.3 Pilih Variabel Prediktor dan Variabel Respon	21
4.4 Hasil <i>Output</i>	22
4.5 Hasil <i>Keseluruhan</i>	22
4.6 Cek GLM untuk Binomial untuk Respon Biner	23

DAFTAR LAMPIRAN

	Halaman
A Data Simulasi	33
B Skrip Program.....	36
B.1 index.html.....	36
B.1 server.r.....	46



BAB 1. PENDAHULUAN

1.1 Latar Belakang

Menurut Hosmer dan Lameshow (2000) regresi logistik merupakan bagian dari analisis regresi yang memprediksi probabilitas suatu peristiwa dengan mencocokkan data pada fungsi logit. Metode ini menggunakan beberapa variabel bebas, baik numerik maupun kategori (Sari *et al*, 2013), tetapi memiliki respon yang bersifat kategorik berskala nominal atau ordinal.

Logistik multinomial merupakan regresi logistik dimana variabel dependennya mempunyai skala yang bersifat *polychotomous* atau multinomial yang terdiri lebih dari dua kategori. Misalnya kasus penelitian untuk melihat bagaimana variabel independen nilai menulis (*writing*) dan status sosial ekonomi (*social economic status*) memengaruhi seorang memilih jenis kelas perkuliahan. Dalam hal ini jenis data untuk variabel independen nilai menulis adalah kuantitatif, jenis data untuk variabel status sosial ekonomi adalah kualitatif. Selanjutnya, jenis data untuk variabel dependen jenis kelas perkuliahan adalah kualitatif.

Logistik multinomial banyak digunakan dan bermanfaat karena banyak ditemui kasus yang menggunakan logistik multinomial untuk menganalisis data. Seperti kasus penelitian Shitarini (2007) menggunakan analisis regresi logistik multinomial untuk menentukan model regresi multinomial logistik terhadap indeks massa tubuh (IMT), kemudian Agustina (2010) menggunakan analisis regresi logistik multinomial untuk menganalisis pemilihan sekolah lanjutan atas pada siswa SMP kelas IX. Penelitian selanjutnya oleh Arini (2011) yang menggunakan analisis regresi logistik untuk mengoptimasi harga dengan model logit multinomial. Contoh kasus penelitian di atas menunjukkan bahwa logistik multinomial banyak dipakai dan bermanfaat. Namun kendalanya akses untuk menggunakan analisis logistik

multinomial terhalang karena menggunakan program yang susah bagi pengguna atau peneliti yang tidak mengerti program.

Beberapa program statistika yang biasa digunakan untuk program statistik diantaranya Stata, SAS dan mungkin yang paling sering R (Wallace, *et al*, 2012). Penelitian ini menggunakan program R untuk memodelkan analisis logistik multinomial dengan menggunakan paket `nnet` dan fungsi `multinom()`. Namun, dalam penggunaannya paket tersebut susah bagi pengguna atau peneliti yang tidak mengerti program terutama program R. Salah satu program dalam R yang bisa membuat menu *user interface* yang berbasis web adalah R-shiny. R-shiny merupakan *toolkit* dari program R yang dapat digunakan untuk membuat program secara *online* seperti GUI. Selain itu, program untuk logistik multinomial belum ada yang menggunakan sistem GUI. Sehingga, dalam tugas akhir ini paket `nnet` dalam mengembangkan analisis logistik multinomial ini akan dibuat *interface* web interaktif dengan R-shiny. Hal ini dimaksudkan untuk mempermudah seseorang yang kurang mengerti pemrograman menjadi lebih mudah untuk mengakses dan melakukan analisis logistik multinomial secara *online* hanya dengan memasukkan datanya. Serta dapat memilih menu tanpa harus menginstal program R dan bagi pengguna web tidak diwajibkan untuk menguasai program terlebih dahulu.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, disusun perumusan ,masalah masalah yaitu:

- a. bagaimana membuat *interface* web interaktif dengan R-shiny untuk analisis logistik multinomial yang memiliki kemampuan yang sama dengan pilihan menggunakan skrip?
- b. bagaimana menggunakan program web interaktif dengan R-shiny untuk menganalisis logistik multinomial pada data simulasi dan interpretasi hasilnya?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini adalah paket R yang digunakan dalam penelitian ini adalah `nnet` dengan fungsi `multinom()`.

1.4 Tujuan

Berdasarkan rumusan masalah yang sudah dibahas, dapat diketahui tujuan dari penelitian ini adalah:

- a. membuat program *interface* web interaktif dengan R-shiny untuk analisis logistik multinomial yang memiliki kemampuan yang sama dengan pilihan menggunakan skrip
- b. menggunakan program web interaktif untuk menganalisis logistik multinomial pada data simulasi dan menginterpretasi hasilnya.

1.5 Manfaat

Manfaat yang didapatkan dalam tugas akhir ini adalah dapat membantu orang lain yang tidak mengerti pemograman menjadi lebih mudah untuk mengakses dan melakukan analisis multinomial logistik secara *online* hanya dengan menginputkan datanya.

BAB 2. TINJAUAN PUSTAKA

2.1 Model Linier

Model linear telah digunakan untuk menganalisis data yang berdistribusi kontinyu. Teknik analisisnya didasarkan pada asumsi pada distribusi normal pada komponen acaknya dan adanya hubungan linear antara nilai tengah dengan peubah penjelas (Tirta, 2009).

2.1.1 Model Linear Sederhana

Analisis regresi yang melibatkan hubungan antara satu variabel respon (tidak bebas) dengan satu variabel prediktor (bebas) diistilahkan dengan regresi linear sederhana, dengan model persamaan:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

Dimana intercept β_0 dan slope β_1 merupakan parameter yang tidak diketahui nilainya, sedangkan ε_i adalah error random dengan rata-rata nol dan varians σ^2 .

Misalkan ada n pasangan observasi, katakan $(y_1, x_1), (y_2, x_2), \dots, (y_n, x_n)$ dengan y merupakan variabel tidak bebasnya atau variabel respon yang berhubungan dengan n variabel bebas diukur dengan error dapat diabaikan sehingga nilai harapan y untuk masing-masing x adalah:

$$E(y|x) = \beta_0 + \beta_1 X$$

2.1.2 Model Linear Tergeneralisir

Model linier klasik hanya berlaku untuk peubah respon yang berdistribusi normal. Namun dalam kenyataannya distribusi dari peubah respon tidak selalu normal. Untuk menangani kondisi dimana respon yang ada tidak berdistribusi normal, tetapi masih saling bebas, maka para statistikawan yang dipelopori oleh Nelder dan

Wedderburn (1972) telah mengembangkan model linier yang dikenal dengan *Generalized Linear Models (GLM)*. Model linier ini menggunakan asumsi bahwa respon memiliki distribusi keluarga eksponensial.

Menurut Tirta (2009) dikatakan model linier mempunyai beberapa hal yang sifatnya khas dan istimewa yaitu:

- 1) ada komponen tetap yang disebut prediktor linier
- 2) respon y_i berdistribusi normal dan saling independen dan
- 3) rerata y_i adalah $\mu_i = \sum_{j=0}^k x_{ij}\beta_j$.

Dalam model linier tergeneralisasi, hubungan di atas mengalami perluasan atau generalisasi, sebagaimana dalam definisi berikut:

Model linier tergeneralisasi adalah model yang mengandung tiga hal yaitu:

- 1) komponen tetap yang disebut prediktor linier $\eta_i = \sum_{j=0}^k x_{ij}\beta_j$ prediktor linier, dinotasikan dengan η_i dari bentuk model linier tergeneralisasi, yaitu: $\eta_i = x_i\beta$ dimana x_i adalah vektor regresi untuk unit sebanyak i dengan *fixed effect* β
- 2) respon y_i berdistribusi secara independen dalam keluarga eksponensial
- 3) hubungan antara mean dengan prediktor linier ditunjukkan oleh fungsi $g(\cdot)$ yang disebut fungsi 'link' sedemikian sehingga $g(\mu_i) = \eta_i$. Fungsi $g(\cdot)$ disebut fungsi hubungan (*link-function*).

2.2 Regresi Logistik

Regresi logistik merupakan salah satu metode yang dapat digunakan untuk mencari hubungan variabel respon yang bersifat berskala nominal atau ordinal dengan dua kategori atau mempunyai skala nominal atau ordinal dengan lebih dari dua kategori dengan satu atau lebih variabel prediktor dan variabel respon yang bersifat kontinyu atau kategorik. Regresi logistik juga merupakan hubungan antara regresi logit dan regresi probit. Yang termasuk dalam regresi ini adalah regresi biner (dengan respon Y hanya dua kategori) (Tirta, 2009).

Model dari regresi logistik ini, adalah:

$$\text{logit}(\pi_i) = \ln\left(\frac{\pi_i}{1-\pi_i}\right) = \alpha + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}, i = 1, \dots, n \quad (2.1)$$

dimana: $\pi_i = E(Y|X_i) = \Pr(Y_i = 1)$;

n adalah banyaknya variabel pengamatan; dan

X adalah variabel eksplanator.

Logaritma dari *odds* (π_i dibagi dengan $1 - \pi_i$) dari *outcome* dimodelkan sebagai fungsi linear variabel eksplanatorinya seperti (2.1) atau dinyatakan sebagai:

$$\pi_i = \Pr(Y_i = 1|X) = \frac{e^{\alpha + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}}}{1 + e^{\alpha + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}}} \quad (2.2)$$

Estimasi parameter β dapat diinterpretasikan sebagai pengaruh multiplikatif dari *odds ratio*, yakni pangkat eksponensialnya merupakan penjumlahan dari parameter-parameter model yang dapat dinyatakan sebagai multiplikasi (penggandaan) eksponensial dari parameter-parameter model. Suku dalam bilangan pangkat eksponensial $e^{[a+b]}$ dapat dinyatakan sebagai $e^a e^b$, sehingga dari persamaan (2.2) dapat dituliskan sebagai berikut:

$$\frac{\pi(Y|X=x_i)}{1-\pi(Y|X=x_i)} = e^{\beta_0} e^{(\beta_1 x_i)} \quad (2.3)$$

Pada sisi kanan persamaan (2.3) dihasilkan dua faktor eksponensial, dari nilai tersebut diperoleh nilai eksponensial ($e^{(a+b)}$) yang dapat dituliskan sebagai ($e^a e^b$) Kemudian sifat tersebut diterapkan kedalam persamaan (2.3) yang menghasilkan persamaan sebagai berikut:

$$\frac{\pi(x)}{1-\pi(x)} = e^{(\alpha + \beta x)} = e^\alpha (e^\beta)^x \quad (2.4)$$

Jadi persamaan (2.4) disebut juga sebagai versi multiplikatif dari *odds ratio*.

Suatu *odds ratio* adalah rasio antara probabilitas beberapa kejadian yang akan terjadi dengan probabilitas kejadian sama yang tidak akan terjadi. Formulanya adalah sebagai berikut:

$$\text{odds} = \frac{\pi}{1 - \pi}$$

Logaritma natural dari *odds ratio* atau disebut logit yang dinyatakan sebagai $\ln(\text{odds ratio})$. Untuk mendapatkan logit dari model regresi logistik dilakukan transformasi terhadap model tersebut. Model logit sederhana dinyatakan sebagai berikut:

$$\ln \left(\frac{\pi_i}{1 - \pi_i} \right) = \alpha + \beta x_i$$

dimana π_i adalah probabilitas sukses dari kejadian ke- i .

Logit dalam regresi logistik merupakan suatu kasus khusus dari fungsi *link* dalam model linier tergeneralisir (Agresti, 1990).

2.3 Distribusi Multinomial

Distribusi multinomial merupakan suatu distribusi yang sering digunakan dalam analisis data kategori. Apabila terdapat j kategori respon maka peluangnya dinotasikan dengan $\{\pi_1, \pi_2, \dots, \pi_j\}$ dengan $\sum_j \pi_j = 1$. Untuk n sampel, peluang multinomial bahwa n_1 termasuk kategori 1, n_2 termasuk kategori 2, ... , n_j termasuk kategori j dengan $\sum_j \pi_j = n$ adalah (Agresti, 1990):

$$P(n_1, n_2, \dots, n_j) = \left(\frac{n!}{n_1! n_2! \dots n_j!} \right) \pi_1^{n_1} \pi_2^{n_2} \dots \pi_j^{n_j}$$

2.4 Regresi Logistik Multinomial

Menurut Yudisasanta dan Ratna (2012) regresi logistik multinomial merupakan regresi logistik yang digunakan saat variabel dependen mempunyai skala yang bersifat *polychotomous* atau multinomial. Skala multinomial adalah suatu pengukuran yang dikategorikan menjadi lebih dari dua kategori. Metode yang digunakan dalam penelitian ini adalah regresi logistik multinomial dengan variabel dependen dengan tiga kategori.

Mengacu pada regresi logistik *trichotomous* untuk model regresi dengan variabel dependen berskala nominal tiga kategori digunakan kategori variabel hasil Y dikodekan 0, 1, dan 2. Variabel Y diparameterkan menjadi dua fungsi logit.

Sebelumnya perlu ditentukan kategori hasil mana yang digunakan untuk membandingkan. Pada umumnya digunakan $Y = 0$ sebagai pembanding. Untuk membentuk fungsi logit, akan dibandingkan $Y = 1$ dan $Y = 2$, terhadap $Y = 0$. Bentuk model regresi logistik dengan p variabel prediktor seperti pada persamaan (2.5).

$$\pi(x) = \frac{e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}} \quad (2.5)$$

Dengan menggunakan transformasi logit akan didapatkan dua fungsi logit,

$$\begin{aligned} g_1(x) &= \ln \left[\frac{P(Y = 1|x)}{P(Y = 0|x)} \right] \\ &= \beta_{10} + \beta_{11}x_1 + \dots + \beta_{1p}x_p \\ &= x' \beta_1 \end{aligned}$$

dan

$$\begin{aligned} g_2(x) &= \ln \left[\frac{P(Y = 2|x)}{P(Y = 0|x)} \right] \\ &= \beta_{20} + \beta_{21}x_1 + \dots + \beta_{2p}x_p \\ &= x' \beta_2 \end{aligned}$$

Berdasarkan kedua fungsi logit tersebut maka didapatkan model regresi logistik *trichotomous* sebagai berikut :

$$P(Y = 0|x) = \frac{1}{1 + e^{g_1(x)} + e^{g_2(x)}} \quad (2.6)$$

$$P(Y = 1|x) = \frac{e^{g_1(x)}}{1 + e^{g_1(x)} + e^{g_2(x)}} \quad (2.7)$$

dan

$$P(Y = 2|x) = \frac{e^{g_2(x)}}{1 + e^{g_1(x)} + e^{g_2(x)}} \quad (2.8)$$

Mengikuti aturan dari model logistik biner, misalkan $\pi_j(x) = P(Y = j|x)$ dengan $j = 0, 1, 2$ untuk setiap fungsi dari vektor $2(p+1)$ dengan parameter $\beta^T = (\beta_1^T \beta_2^T)$. Pernyataan umum untuk probabilitas bersyarat dalam model tiga kategori adalah:

$$P(Y = j|x) = \frac{e^{g_j(x)}}{\sum_{k=0}^2 e^{g_k(x)}}, j = 0, 1, 2. \quad (2.9)$$

dengan vektor $\beta_0 = 0$ sehingga $g_0(x) = 0$.

Model logistik untuk kategori variabel dependen lebih dari satu atau *polychotomous*, yaitu:

$$\ln \left[\frac{\pi_j}{\pi_q} \right] = \beta_0^{(j)} + \sum_{i=1}^k \beta_i^{(j)} x_i, j = 1, \dots, q - 1$$

pada persamaan (2.9) dapat dilihat bahwa salah satu dari kategori digunakan sebagai referensi dan disebut sebagai basis (*baseline*), yaitu kategori yang menjadi dasar pembandingan pengaruh kategori lainnya.

2.5 Estimasi Parameter

Nilai harapan antar variabel respon tidak linier serta memiliki varian-varian yang tidak sama dalam model regresi logistik, sehingga penduga parameter diperoleh melalui metode *Maximum Likelihood* (Hosmer dan Lemeshow, 2000). Untuk memecahkan masalah sistem persamaan nonlinier, solusi yang dilakukan adalah dengan mengestimasi melalui proses iterasi *Newton Raphson*. Karena variabel respon y_j diasumsikan saling bebas, maka diperoleh fungsi likelihood bersyarat untuk sampel sebanyak n observasi sebagai berikut:

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

Secara matematis, akan lebih mudah untuk mendapatkan nilai yang akan memaksimalkan fungsi likelihood di atas melalui log dari fungsi tersebut yaitu *log likelihood*. Dengan demikian maka fungsi *log likelihood*-nya adalah:

$$L(\beta) = \sum_{i=1}^n y_i g_1(x_i) + y_{2i} g_2(x_i) - \ln(1 + e^{g_1 x_i} + e^{g_2 x_i})$$

Untuk mendapatkan nilai yang memaksimumkan $L(\)$ maka dilakukan diferensiasi terhadap $L(\)$, dengan syarat $\frac{\partial L}{\partial \beta} = 0$ dan $\frac{\partial^2 L}{\partial^2 \beta} < 0$

2.6 Uji Kesesuaian Model (*Goodness of Fit*)

Untuk mengetahui apakah model dengan variabel dependen tersebut merupakan model yang sesuai, maka perlu dilakukan suatu uji kesesuaian model. Uji yang digunakan adalah uji AIC dan BIC.

2.6.1 AIC (*Akaike's Information Criterion*)

Pemilihan model dari sebuah data set yang terbaik salah satunya adalah dengan menggunakan metode AIC (*Akaike's Information Criterion*). AIC adalah suatu kriteria pemilihan model terbaik yang mempertimbangkan banyaknya parameter dalam model. Menurut Tirta (2009) besarnya AIC dihitung melalui rumus berikut :

$$AIC = -2l(\hat{\theta}) + 2q$$

dengan $l(\hat{\theta})$ adalah nilai likelihood dari model yang dihadapi dan q adalah banyaknya parameter dalam model. Pemilihan model terbaik dilihat dari nilai terkecil dari AIC.

2.6.2 BIC (*Bayesian Information Criterion*)

Selain AIC, pemilihan model terbaik juga bisa dilakukan dengan menggunakan BIC. Menurut Pardede (2013) BIC didefinisikan sebagai berikut :

$$BIC = -2l(\hat{\theta}) + q \ln N$$

dengan $-2l(\hat{\theta})$ adalah model log *likelihood*, adalah q jumlah parameter pada model, dan N adalah jumlah objek pengamatan. Model dengan nilai BIC lebih kecil dipilih sebagai model terbaik untuk data.

2.7 R-Shiny

Menurut Tirta (2014) R-shiny mempunyai komponen yang dibagi menjadi dua kelompok yaitu seperti berikut ini :

a. *User Interface*. Manfaat dari *user interface* adalah:

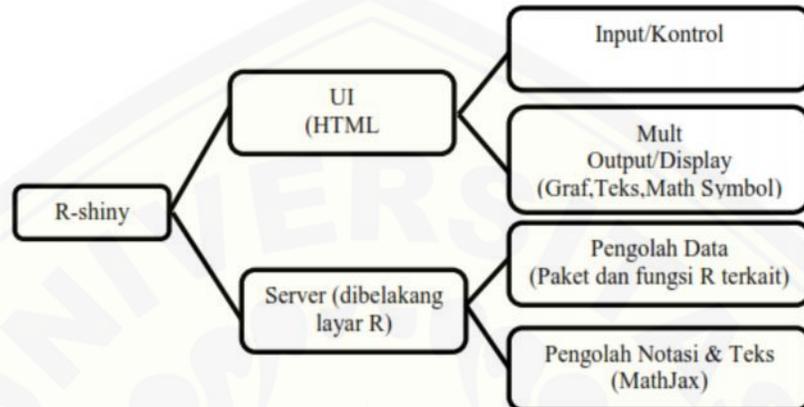
- 1) panel Kontrol adalah panel untuk mengontrol input berupa data, variabel, model, tergantung kompleksitas modul. Tampilan kontrol dapat berupa slider, radiobutton, check-box dan lain-lain
- 2) pemasukan permintaan nilai input (data dengan berbagai jenis variabel yang diperlukan, pemilihan model, jenis, dan kriteria uji statistika)
- 3) penyajian output terkait hasil analisis/uji. Hasil output dapat berupa
 - i. grafik (histogram, diagram pencar, dan lain-lain)
 - ii. bentuk angka/teks bisa berbentuk asli (verbatim) maupun dalam bentuk table
 - iii. teks khusus dengan notasi matematika dengan format LaTeX.

Untuk mengakomodasi berbagai jenis luaran tadi, shiny juga menyediakan berbagai format output seperti `plotOutput`, `textOutput`, `tableOutput`, `verbatimTextOutput`, dan lain-lainnya.

Bagian user interface ini dapat disaikan pada file khusus `ui.r`, dapat juga disajikan penuh melalui file HTML, misalnya `index.html`

b. *Server*. Bagian ini merupakan otak dari program yang bertugas melakukan simulasi, berbagai analisis data sesuai pilihan pengguna dan selanjutnya mengirim hasilnya ke bagian output. Bagian ini didukung oleh berbagai prosedur dan analisis data yang pada umumnya telah tersedia pada berbagai paket R. Bagian ini disimpan dalam file yang diberi nama `server.r`.

Struktur umum komponen pemrograman dengan R-shiny dapat dilihat pada Gambar 2.1



Gambar 2.1. Struktur Komponen R-Shiny

Menurut Tirta (2014) cara penyajian file ui dan file server ada dua macam yaitu

- membuat file server.r dan ui.r secara terpisah lalu diletakkan pada satu direktori yang diberi nama sesuai dengan topiknya. Selanjutnya yang dipanggil adalah nama direktorinya

Direktori: shiny1

File:

-server.r

-ui.r

Cara memanggil: `runApp("shiny1")`.

- menggabungkan file server dan ui dalam satu fungsi. Selanjutnya pemanggilan dilakukan dengan memanggil nama fungsinya.

```

fungsi.shiny1<-function(){
ui=,
server=
}
  
```

Cara memanggil: `fungsi.shiny1()`

- c. R-shiny juga dilengkapi kemampuan untuk menampilkan rumus-rumus matematika dengan memanfaatkan MathJax yang memiliki sintaks mirip LaTeX.

```
withMathJax(  
  helpText("Teks biasa $$simbol matematika $$"))
```



BAB 3. METODOLOGI PENELITIAN

Pembuatan program pengembangan dan aplikasi GUI web interaktif dengan R-shiny untuk pemodelan logistik multinomial mempunyai dua tahap.

Tahap I yaitu penyusunan web. Langkah-langkah ilustrasi penyusunan webnya sebagai berikut:

- a. mengidentifikasi *input* dan *output* logistik multinomial dengan paket `nnet`. *Input* yang akan dientrikan berupa data dan formula untuk dimodelkan dengan analisis logistik multinomial. *Output* yang akan dihasilkan berupa *summary*, hasil estimasi, hasil GOF. Paket yang digunakan yaitu paket `nnet` (Venables dan Ripley, 2002). Struktur fungsi multinom pada R dengan skrip sebagai berikut:

```
multinom(formula, data, trace, ...)
```

uraian lengkap mengenai perintah-perintah fungsi diatas sebagai berikut:

`formula` : sebuah formula dalam bentuk `respon~predictors`.
`data` : dataframe yang mengandung variabel yang disediakan formula.
`trace` : perubahan untuk melacak optimasi.
... : tambahan argument untuk multinom.

- b. menyusun dan mentransfer dokumen teori logistik multinomial kedalam format LaTeX berupa `index.html`. `index.html` merupakan file HTML yang digunakan untuk memasukkan nilai *input* dan penyajian *output* sehingga dapat ditampilkan secara online
- c. menyusun skrip R-shiny yang berupa `server.r`. `server.r` merupakan otak dari program sehingga pada bagian ini dilakukan pengaplikasian program analisis logistik multinomial pada R kedalam program R-shiny. Untuk menganalisis data analisis logistik multinomial menggunakan *software* R 3.1.3

- d. mengintegrasikan dan menguji coba dokumen teori dalam `index.html` dan `server.r` yang ada pada langkah b dan c untuk membuat program *online*

Tahap II yaitu ilustrasi penggunaan program web interaktif dengan R-shiny untuk analisis data dengan logistik multinomial. Langkah-langkah ilustrasi analisis datanya sebagai berikut:

- a. menginputkan data ke dalam program *online*. Data yang dianalisis yaitu data simulasi yang memiliki variabel respon berdistribusi multinomial yaitu kategori A, B, C. Variabel prediktor yang terdiri dari x_1, x_2, x_3 , dan x_4 berupa jenis kelamin yang terdiri dari L sebagai laki-laki dan P sebagai perempuan (Lampiran A). Data yang digunakan pada logistik multinomial menggunakan variabel yaitu:
 - 1) variabel respon
Untuk data multinomial variabel respon (y) lebih dari satu kategori untuk setiap responden (n).
 - 2) variabel prediktor
Variabel prediktor merupakan variabel yang mempengaruhi respon.
Variabel prediktor dinotasikan dengan (x).
- b. eksplorasi data yaitu melihat karakteristik data serta menentukan variabel respon dan variabel prediktor
- c. mengestimasi parameter dengan memilih menu program logistik multinomial. Dalam langkah ini program akan menampilkan hasil estimasi, standart *error*, *z-value* dan *p-value*
- d. melakukan uji signifikansi dengan menggunakan nilai *p-value*. Taraf signifikansi yang digunakan sebesar $\alpha = 0,05$. Variabel dikatakan berpengaruh terhadap variabel respon apabila variabel tersebut memiliki $p - value \leq 0.05$
- e. melakukan uji kesesuaian model dengan menggunakan AIC dan BIC. AIC dan BIC digunakan untuk memilih model dari sebuah data set yang terbaik yang dilihat dari nilai terkecil dari AIC dan BIC
- f. menginterpretasikan hasil.