



Pengujian akurasi model regresi logistik multinomial untuk memprediksi keberhasilan mahasiswa di perguruan tinggi menggunakan r

Sentot Eko Baskoro¹, Suhartono², Totok Chamidy³, Syahiduz Zaman⁴

¹ Sekolah Tinggi Ilmu Ekonomi GICI, Depok

^{2,3,4} Jurusan Teknik Informatika UIN Maulana Malik Ibrahim Malang

¹sentot@cbn.net.id, ²suhartono@ti.uin-malang.ac.id, ³totokchamidy@gmail.com, ⁴amansyahiduz@gmail.com

Info Artikel

Sejarah artikel:

Diterima 22 Agustus 2022

Disetujui 17 Oktober 2022

Diterbitkan 25 Oktober 2022

Kata kunci:

Kategorikal; Multinomial regresi logistik; Prediksi; Akurasi; Keberhasilan mahasiswa

Keywords :

Categorical; Multinomial logistic regression; Prediction; Accuracy; Student success

ABSTRAK

Lembaga pendidikan tinggi harus bisa menjamin keberhasilan mahasiswa dalam bidang akademik. Maka lembaga pendidikan tinggi harus dapat membuat model untuk memprediksi keberhasilan mahasiswa sedini mungkin. Untuk mendapatkan model dalam memprediksi keberhasilan mahasiswa perlu untuk menentukan faktor-faktor yang berpengaruh, dimana faktor-faktor tersebut berbentuk data kontinyu dan data kategori. Oleh karena itu tujuan penelitian ini adalah membangun model regresi logistik berdasarkan data campuran dari data kontinyu dan data kategori, kemudian melakukan pengujian akurasi model untuk memprediksi keberhasilan mahasiswa di perguruan tinggi. Data penelitian yang digunakan sebanyak 68 data dari mahasiswa. Enam tahapan penelitian yang telah dilakukan yaitu tahap pertama persiapan dan pengumpulan data, tahap kedua menganalisis data, tahap ketiga membangun model multinomial regresi logistik, tahap empat menguji koefisien model, tahap keempat pengujian dan validasi model, tahap ke lima mengukur akurasi model, tahap keenam menarik kesimpulan berdasarkan hasil analisis yang diperoleh. Hasil analisis dan akurasi model prediksi menggunakan metode multinomial regresi logistik, didapatkan model yang terbaik dengan faktor yang signifikan mempengaruhi lama studi adalah jenis kelamin, jurusan, dan jalur masuk, sedangkan akurasi model prediksi untuk masing-masing variabel respon lama studi adalah 96,4%.

ABSTRACT

Higher education institute should have maintained its student's success in their academic field. Therefore, higher education institute should make a model to predict its success as early as possible. To have such model, the impacting factors should be determined, which factor are in the forms of continuous dan categorized data. This research is aimed to build logistic regression model based on compounded data from continuous and categorized data, which then test the model accuracy to predict the student's success in such institute. Research data is using 68 student's data. There are 6 research steps, first, data preparation and collection, second, data analysis, third, building the logistic multinomial regression model, fourth, data testing and validation, fifth, measure the model accuracy, and the sixth step is drawing the conclusion based on the analysis output. The result of prediction analysis dan accuracy test using logistic multinomial regression is the best model with significant factors that influence the study time are gender, department and selection track, while prediction accuracy model for each study time response variable is 96.4%.



©2022 Penulis. Diterbitkan oleh Program Studi Akuntansi, Institut Koperasi Indonesia.

Ini adalah artikel akses terbuka di bawah lisensi CC BY NC

(<https://creativecommons.org/licenses/by/4.0/>)

PENDAHULUAN

Metode regresi adalah salah satu metode yang umum dan paling dasar digunakan dalam menentukan prediksi terhadap suatu data, dengan cara mengetahui pengaruh variabel respon dengan variabel *predictor* (Nafi'iyah, 2016; Rabbi et al., 2021). Tujuan dari metode regresi adalah memperoleh model prediksi yang baik yang ditunjukkan dengan pengujian akurasi dengan tingkat kesalahan yang rendah (Utama & Hajarisman, 2021).

Saat ini perkembangan akuisisi data terus dikembangkan, data penyimpanan dan analisis dan teknologi prediksi di era data besar pada bidang medis mencari lebih sederhana dan cepat (Lee & Han, 2020; Zhou et al., 2019). Metode regresi hanya dapat memproses variabel berbentuk data kontinyu dan variabel tersebut harus terdapat korelasi atau pengaruh, bila variabel tidak mempunyai sifat tersebut maka tidak bisa diselesaikan dengan metode regresi dan menjadi kelemahan metode tersebut karena pada saat proses pengambilan data mesti terdapat data yang berbentuk kategorikal (Nugraha, 2014; Purba & Ruslan, 2020; Sarwono, 2013).

Jika proses pengambilan data terdapat data yang berbentuk kategori saja, maka perlu metode linier yang digeneralisasi atau disebut metode regresi logistik. Dimana metode regresi logistik adalah metode regresi yang setiap variabelnya berbentuk kategori atau interval menurut Liang et al. (2020) dan tidak mengharuskan bahwa semua variabel responnya adalah berbentuk data kategori dan berskala dikotomi. Sedangkan tujuan dari metode ini adalah sama dengan regresi linear yaitu memperoleh model prediksi yang baik dengan ditunjukkan dengan tingkat akurasi yang rendah. Pemodelan prediksi keberhasilan mahasiswa dibutuhkan demi mengetahui apakah sistem pembelajaran yang dilakukan sudah cukup baik atau tidak (Guswandi et al., 2021; Heryana et al., 2020).

Berdasarkan penelitian sebelumnya yang dilakukan oleh Rizki et al. (2020) mengungkapkan bahwa model regresi logistik multinomial dipilih sebagai model terbaik yang menggambarkan kondisi penjurusan peserta didik, hal ini dikarenakan tingkat akurasinya lebih tinggi dibandingkan pada model diskriminan untuk setiap kemungkinan dalam pembagian proporsi data *training* dan *testing*. Kemudian penelitian Santosa & Chrismanto (2018) mengungkapkan bahwa rata-rata akurasi pemodelan lebih tinggi menggunakan Regresi Logistik. Tetapi untuk kestabilan model prediksi terlihat bahwa model Regresi Multinomial lebih stabil daripada model Regresi Logistik

Pada penelitian ini membahas tahapan membangun model multinomial regresi logistik dan mengukur akurasi model multinomial regresi logistic, penelitian ini menggunakan model multinomial regresi logistik karena data respon untuk prediksi adalah data dengan kategori lebih dari 2 kategori, data respon dan data *predictor* yang didapat adalah berbentuk data campuran, yang terdiri dari data kontinu dan data kategori. Pada multinomial regresi logistik untuk data kontinu harus dilakukan transformasi data terlebih dahulu, yaitu merubah bentuk data kontinu menjadi berbentuk data kategori. Penelitian ini juga menghitung akurasi model yang didapat, dengan ditunjukkan nilai *error* yang didapat.

METODE PENELITIAN

Penelitian ini dilaksanakan pada bulan Januari hingga Februari 2022. Penelitian ini bertempat di satu program studi pada lembaga pendidikan tinggi. Pengolahan data dilakukan di Laboratorium Data Sains Jurusan Teknik Informatika UIN Maulana Malik Ibrahim Malang. Penelitian ini menggunakan data primer dengan menjalankan kusioner kepada mahasiswa yang telah lulus pada tahun 2021 dalam satu program studi pada lembaga pendidikan tinggi di Jawa timur.

Pengambilan sampel data dalam penelitian ini berjumlah 90 mahasiswa, yaitu data mahasiswa yang telah lulus pada tahun 2021. Dari pengambilan sampel data tersebut terdapat *missing data* sebanyak 4 orang dimana dari 90 responden menjadi 86 responden yang diproses dengan metode multinomial regresi logistik.

Pada penelitian ini mengambil objek pada perkembangan mahasiswa di lembaga pendidikan tinggi, dimana salah satu untuk mengukur kesuksesan mahasiswa di lembaga Pendidikan tinggi adalah lama studi mahasiswa (Goodwin, 2011), maka pada penelitian ini mengambil variabel respon (*y*) yaitu Lama Studi sebagai *y1* dan sebagai variabel prediktor (*x*) adalah Jenis kelamin sebagai *x1*, Jalur masuk sebagai *x2*, Jenis SMA sebagai *x3*, pendapatan sebagai *x4*, asal daerah sebagai *x5*, IPM sebagai *x6* dan pekerjaan orang tua sebagai *x7*. Sebelum dilakukan pemrosesan data dengan metode multinomial regresi logistik perlu dilakukan transformasi data untuk mengubah data awal yang telah diperoleh ke dalam bentuk data yang dapat diolah dan sesuai dengan tujuan penelitian yaitu diskritisasi dan *concept hierarchy for nominal data* (Bi et al., 2020).

Tahapan penelitian dibagi menjadi enam tahapan penelitian yaitu tahap pertama persiapan dan pengumpulan data, tahap kedua menganalisis data, tahap ketiga membangun model multinomial regresi logistik, tahap empat menguji koefisien model, tahap keempat pengujian dan validasi model, tahap ke lima mengukur akurasi model, tahap keenam menarik kesimpulan berdasarkan hasil analisis yang diperoleh.

HASIL DAN PEMBAHASAN

Tahap pertama persiapan dan pengumpulan data

Sampel data dalam penelitian ini berjumlah 86 mahasiswa terdiri dari 80 mahasiswa sebagai data training dan 6 mahasiswa sebagai data test. Variabel respon (*y*) terdapat dua variabel yaitu Lama Studi (*y1*) sedangkan variabel prediktor terdapat tujuh variabel yaitu Jenis kelamin (*x1*), Jalur masuk (*x2*), Jenis SMA (*x3*), pendapatan (*x4*), asal daerah (*x5*), pekerjaan orang tua (*x6*) dan IPM (*x7*).

Setelah data terkumpul disimpan pada file excel "mahasiswaIPKup1.xlsx" . Kemudian dilakukan proses import dengan script kode program R pada gambar 1.

```
```\r}
Import the data
library(readxl)
seer_ordinal <- read_excel(path = "mahasiswaIPKupno.xlsx")
seer_ordinal
```\r}
```

Gambar 1 Kode program R membaca data file mahasiswaIPKup1.xlsx

Untuk melihat data dengan script kode program R pada gambar 5.

```
```\r}
head(seer_ordinal)
tail(seer_ordinal)
```\r}
```

Gambar 2 Kode program R menampilkan data set seer_multi

Hasil script kode program R untuk menampilkan data set seer_multi dapat dilihat pada gambar 3.

| | NIM | Lama_Studi | Lama_Studi_up1 | Jenis_Kelamin | Jenis_Kelamin_up1 | Jalur_Masuk | Jalur_Masuk_up1 | Jenis_SMA | Jenis_SMA_up1 | Pendapatan | Pendapatan_up1 | Asal_Daerah |
|----|-------|------------|----------------|---------------|-------------------|---------------|-----------------|------------|---------------|------------|----------------|--------------|
| | <dbl> | <dbl> | <dbl> | <fctr> | <fctr> | <fctr> | <fctr> | <fctr> | <fctr> | <dbl> | <dbl> | <fctr> |
| 23 | 23 | 9 | 3 | P | 2 | SNMPTN | 1 | SMA Negeri | 1 | 2,0-2,5 | 4 | Surabaya |
| 24 | 24 | 8 | 2 | P | 2 | SNMPTN | 1 | SMA Negeri | 1 | 1,0-1,5 | 2 | Non Surabaya |
| 25 | 25 | 8 | 2 | L | 1 | SNMPTN | 1 | SMA Negeri | 1 | 1,0-1,5 | 2 | Surabaya |
| 26 | 26 | 9 | 4 | L | 1 | SNMPTN | 1 | SMA Negeri | 1 | 2,0-2,5 | 4 | Non Surabaya |
| 27 | 27 | 8 | 2 | L | 1 | PKM Kemitraan | 3 | SMA Swasta | 2 | 1,5-2,0 | 3 | Non Surabaya |
| 28 | 28 | 7 | 1 | L | 1 | PKM Kemitraan | 3 | SMA Swasta | 2 | 2,0-2,5 | 4 | Non Surabaya |

6 rows | 1-13 of 18 columns

Gambar 3 Hasil Kode program R menampilkan data set seer_ordinal

Setelah mendapatkan data, maka dilakukan pembagian data menjadi dua bagian yaitu data training sebanyak 60% digunakan untuk membangun model prediksi dan data testing sebanyak 40% digunakan untuk mengukur akurasi terhadap model prediksi.

```
```\r}
set.seed(111)
ind <- sample(2, nrow(cmcData),
 replace = TRUE,
 prob = c(0.6, 0.4))
training <- cmcData[ind==1,]
testing <- cmcData[ind==2,]
training
testing
```

A tibble: 5 x 8

|  | Lama_Studi_up1 | Jenis_Kelamin_up1 | Jalur_Masuk_up1 | Jenis_SMA_up1 | Pendapatan_up1 |
|--|----------------|-------------------|-----------------|---------------|----------------|
|  | <dbl>          | <dbl>             | <dbl>           | <dbl>         | <dbl>          |
|  | 2              | 2                 | 1               | 1             | 1              |
|  | 2              | 1                 | 1               | 1             | 2              |
|  | 3              | 1                 | 1               | 1             | 7              |
|  | 4              | 1                 | 1               | 1             | 4              |
|  | 1              | 1                 | 3               | 2             | 4              |

5 rows | 1-5 of 8 columns

## Tahap kedua menganalisis data

Pada penelitian ini dilakukan analisis data dengan menggunakan analisis statistik deskriptif, analisis statistik deskriptif digunakan untuk menjelaskan karakteristik data yang dijadikan objek penelitian berdasarkan variabel predictor dan variabel respon. Hasil analisis statistik deskriptif variabel IPK mayoritas memiliki IP baik dengan prosentase sebesar 52,5%, variabel Lama Studi mayoritas lulus normal dengan prosentase sebesar 50,0%, variabel asal SMA mayoritas SMA Negeri, variabel jalur masuk mayoritas SNMPTN, variabel pendapatan orangtua mayoritas Rp 1.500.000,- sampai Rp 2.500.000,-, variabel pekerjaan orang tua mayoritas PNS Non Guru/Dosen, variabel asal daerah mayoritas Jawa Timur, variabel Indeks Pembangunan daerah (IPM) mayoritas menengah keatas. Untuk mendapatkan analisis statistik deskriptif variabel dengan script kode program R pada gambar 4.

```
```{r}
table(training$Lama_Studi_up1)
prop.table(table(training $Lama_Studi_up1))
```
```

**Gambar 4 Kode program R menampilkan data dari variabel Lama Studi**

Hasil Kode program R menampilkan data bentuk tabel dari variabel Lama Studi dapat dilihat pada gambar 5.

```
 1 2 3 4 5
 3 6 14 3 2

 1 2 3 4 5
0.10714286 0.21428571 0.50000000 0.10714286 0.07142857
```

**Gambar 5 Hasil Kode program R menampilkan data dari variabel Lama Studi**

Sedangkan analisis deskriptif variabel predictor terhadap variabel respon. Variabel lama studi dengan variabel jenis kelamin didapat bahwa terlambat lulus mayoritas laki-laki, variabel lama studi dengan variabel Asal SMA didapat bahwa lulus tepat waktu mayoritas SMA, variabel lama studi dengan variabel jalur masuk didapat bahwa lulus tepat waktu mayoritas jalur SNMPTN, variabel lama studi dengan variabel pendapatan orang tua didapat bahwa lulus tepat waktu mayoritas pendapatan orang tua Rp 1.000.000,- sampai Rp 1.500.000,-, variabel lama studi dengan variabel pekerjaan orang tua didapat bahwa lulus tepat waktu mayoritas pekerjaan orang tua PNS Guru/Dosen, variabel lama studi dengan variabel asal daerah didapat bahwa lulus tepat waktu mayoritas dari wilayah jawa timur, variabel lama studi dengan variabel IPM didapat bahwa lulus tepat waktu mayoritas tingkat IPM menengah keatas. Untuk mendapatkan analisis statistik deskriptif variabel lama studi dengan variabel jenis kelamin dengan script kode program R pada gambar 6.

```
```{r}
table(tarining$Lama_Studi_up1, tarining $Jenis_Kelamin_up1)
prop.table(table(tarining $Lama_Studi_up1, tarining $Jenis_Kelamin_up1))
```
```

**Gambar 6 Kode program R menampilkan probabilitas data dari variabel Jenis kelamin terhadap variabel Lama Studi**

Hasil Kode program R menampilkan probabilitas data dari variabel Jenis kelamin terhadap variabel Lama Studi dapat dilihat pada gambar 7.

|   | 1          | 2          |
|---|------------|------------|
| 1 | 0.03571429 | 0.07142857 |
| 2 | 0.10714286 | 0.10714286 |
| 3 | 0.35714286 | 0.14285714 |
| 4 | 0.10714286 | 0.00000000 |
| 5 | 0.07142857 | 0.00000000 |

**Gambar 7 Hasil Kode program R menampilkan probabilitas data dari variabel Jenis kelamin terhadap variabel Lama Studi**

Sampel data yang telah diperoleh kemudian dilakukan transformasi data untuk mentransformasikan atribut-atribut pada variable ke dalam bentuk yang dapat diolah dengan metode multinomial regresi logistic. Transformasi data yang dilakukan pada penelitian ini terbagi menjadi dua tahap, tahap pertama diskritisasi dan tahap kedua konsep hierarki untuk data nominal. Tahap pertama diskritisasi untuk mengunah variabel kontinu kedalam kategorikal. Tahap diskritisasi dilakukan pada variabel Lama Studi (y1) dan IPM (x7). Tahap kedua konsep hierarki untuk data nominal digunakan untuk me-generalisasikan atribut pada variabel ke tingkat yang lebih tinggi. Tahap konsep hierarki untuk data nominal dilakukan pada variabel Jalur masuk (x2), Jenis SMA (x3), pendapatan (x4), asal daerah (x5), IPM (x6) dan pekerjaan orang tua (x7). Untuk setiap variabel yang telah dilakukan transformasi data yaitu y1, x1, x2, x3, x4, x5,x6 dan x7 kemudian setiap variabel tersebut dianalisa karakteristik nya dengan fungsi histogram() dan density() dengan script kode program R pada gambar 8.

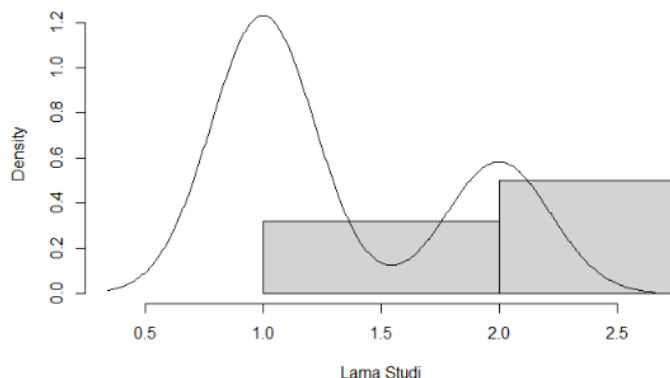
```

{r}
hist(training$Lama_Studi_up1, probability = TRUE, xlim = xlim, ylim = ylim, xlab="Lama Studi",
main="")
lines(dens)

```

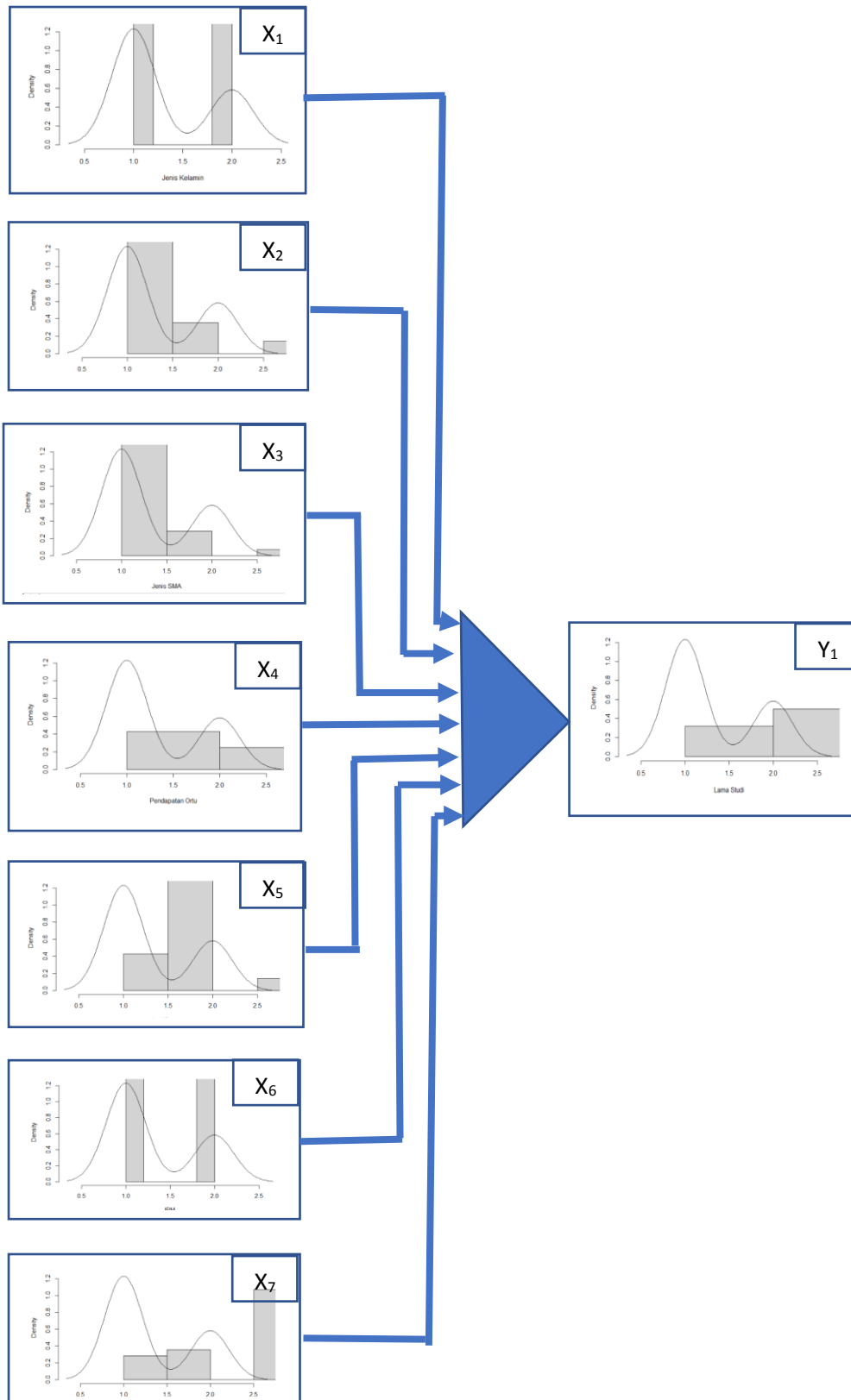
**Gambar 8 Kode program R karakteristik data Lama\_Studi\_up1**

Hasil Kode program R mengetahui karakteristik data Lama\_Studi\_up1 dapat dilihat pada gambar 9.



**Gambar 9 Hasil Kode program R mengetahui karakteristik data Lama\_Studi\_up1**

Setelah mendapatkan karakteristik data untuk setiap variabel maka dilakukan keterhubungan untuk setiap variabel yang akan menjadi input dan output pada model multinomial regresi logistik dengan variable multi kategori, keterhubungan variabel tersebut dapat dilihat pada gambar 10.



**Gambar 10 Keterhubungan variabel sebagai input dan output pada model multinomial regresi logistik**

Pada gambar 10 dapat dilihat bahwa untuk model multinomial regresi logistik sebagai variabel yang digunakan sebagai input sebanyak tujuh variabel yaitu variabel Jenis kelamin ( $x_1$ ), variabel Jalur masuk ( $x_2$ ), variabel Jenis SMA ( $x_3$ ), variabel pendapatan ( $x_4$ ), variabel asal daerah ( $x_5$ ), variabel pekerjaan orang tua ( $x_6$ ) dan variabel IPM ( $x_7$ ) sedangkan variabel yang digunakan sebagai output adalah variabel Lama Studi ( $y_1$ ). Dengan menggunakan density plot bahwa variabel-variabel tersebut berbentuk distribusi data yang mempunyai karakter yang mirip, dengan membentuk distribusi yang mirip akan didapatkan model multinomial regresi logistik yang optimum.

### Tahap ketiga membangun model multinomial regresi logistic

Multinomial regresi logistic dapat digunakan untuk memprediksi variabel dependen kategoris yang memiliki lebih dari dua level. Output model Multinomial regresi logistic dapat diprediksi dengan menggunakan satu atau lebih variabel bebas. Variabel bebas dapat bertipe nominal, ordinal, atau kontinu. Pada penelitian ini model Multinomial regresi logistic dapat memprediksi lama studi menurut semester. Dimana variabel lama studi sudah dilakukan diskritisasi yaitu melakukan perubahan data dari bentuk kontinu ke dalam bentuk kategorikal. Lama studi dikategorikan menjadi tiga kategori, kategori pertama adalah lama studi < 8 semester dengan keterangan lebih cepat, kategori kedua adalah lama studi = 8 semester dengan keterangan normal, kategori ketiga adalah lama studi > 8 semester dengan keterangan terlambat.

Pada penelitian ini membangun model multinomial regresi logistic dengan menggunakan fungsi multinorm pada package nnet, output model adalah lama studi dengan tiga level dengan script kode program R pada gambar 11.

```
```\r\nlibrary(nnet)\n## Run the multinomial model with the multinom function and summarize it\nmultimodel <- multinom(training$Lama_Studi_up1~ training$Jenis_Kelamin_up1+\ntraining$Jalur_Masuk_up1+ training$Jenis_SMA_up1+ training$Pendapatan_up1+\ntraining$Asal_Daerah_up1+ training$IPM_up1+ training$Pekerjaan_Ortu_up1)\n```\n
```

Gambar 11 Kode program R membuat model multinomial regresi logistic dengan fungsi multinorm()

Hasil Kode program R untuk melihat model multinomial regresi logistic pada variabel multi model dapat dilihat pada gambar 12.

```
\n# weights: 95 (72 variable)\ninitial value 45.064262\niter 10 value 4.298946\niter 20 value 1.391611\niter 30 value 1.386302\nfinal value 1.386294\nconverged\nNaNs producedCall:\n
```

Gambar 12 Hasil Kode program R melihat model multinomial regresi logistic

Pada gambar 12 terdapat beberapa output yang dihasilkan dengan menjalankan model, keluaran yang yang dihasilkan dari menjalankan multimodel terdapat tiga iterasi dan menyertakan log negatif pada iterasi nilai 30 didapat log negative akhir 1,386294.

Untuk melihat hasil model multinominal regresi logistic dengan fungsi summary() dapat menggunakan script kode program R pada gambar 11.

```
```{r}
summary(multimodel)
```
```

Hasil Kode program R untuk melihat hasil model multinominal regresi logistic dengan fungsi summary() dapat dilihat pada gambar 12.

| Coefficients: | | training \$Jenis_Kelamin_up12 | training \$Jalur_Masuk_up12 |
|---------------|--------------------------------|--------------------------------|--------------------------------|
| (Intercept) | | | |
| 2 | 38.95767 | 26.88719 | -39.6624333 |
| 3 | 16.92615 | -32.67641 | -83.6263768 |
| 4 | 20.27364 | -75.33936 | -24.5712549 |
| 5 | 21.82604 | -129.43179 | -0.3024562 |
| | training \$Jalur_Masuk_up13 | training \$Jenis_SMA_up12 | training \$Jenis_SMA_up13 |
| 2 | 58.41572 | 14.4495608 | 63.289803 |
| 3 | -83.72388 | -24.3513312 | -22.970539 |
| 4 | -54.07892 | 0.4658031 | -5.000226 |
| 5 | -31.45001 | -53.0548630 | -10.747203 |
| | training \$Pendapatan_up12 | training \$Pendapatan_up13 | training \$Pendapatan_up14 |
| 2 | -20.28240 | -18.69708 | -134.42363 |
| 3 | 42.32943 | 94.76927 | 50.30350 |
| 4 | -26.07885 | -12.95020 | 13.09094 |
| 5 | 43.21255 | 33.15658 | -34.79842 |
| | training \$Pendapatan_up15 | training \$Pendapatan_up16 | training \$Pendapatan_up17 |
| 2 | -23.42497 | -20.54189 | -9.917300 |
| 3 | 85.33528 | -25.962725 | 19.360907 |
| 4 | -17.78795 | 72.332674 | -2.029258 |
| 5 | -13.42889 | -8.175964 | -1.090844 |
| | training \$Asal_Daerah_up12 | training \$Asal_Daerah_up13 | training \$IPM_up12 |
| 2 | -49.220598 | -33.34227 | -65.45769 |
| 3 | 58.874941 | 104.69619 | -16.00727 |
| 4 | 69.945459 | -19.81721 | -104.74858 |
| 5 | -5.579176 | -14.51973 | 52.73015 |
| | training \$Pekerjaan_Ortu_up12 | training \$Pekerjaan_Ortu_up13 | training \$Pekerjaan_Ortu_up14 |
| 2 | -12.606417 | 40.74066 | 63.610386 |
| 3 | 34.536216 | -72.29543 | 3.005039 |
| 4 | -1.531388 | -18.30517 | -12.751695 |
| 5 | -2.461664 | -44.03001 | -14.799226 |

Gambar 13 Hasil Kode program R membuat model multinominal regresi logistic dengan fungsi summary()

Pada gambar 13 adalah ringkasan model multinominal regresi logistic dengan fungsi summary(), dari ringkasan model didapat residual deviance sebesar 2.7725 hasil ini didapat dari dua kali nilai log negative akhir pada iterasi nilai 30 sebesar 1,386294 dan dapat digunakan dalam perbandingan model bersarang.

Output dari ringkasan model memiliki blok coefisient dan blok standard error. Masing-masing blok ini memiliki satu baris nilai yang sesuai dengan persamaan model. Berfokus pada blok koefisien, kita dapat melihat baris pertama yang membandingkan Lama Studi = "2" dengan Lama Studi = "1" dan baris kedua membandingkan Lama Studi = "3" dengan Lama Studi dasar kami = "1". Jika kita menganggap koefisien dari baris pertama adalah b1, koefisien dari baris kedua adalah b2, koefisien dari baris ketiga adalah b3, koefisien dari baris keempat adalah b4, koefisien dari baris kelima adalah b5 , kita dapat menulis persamaan model kita:

$$\ln \left(\frac{P(\text{Lama Studi} = 2)}{P(\text{Lama Studi} = 1)} \right) = b_{10} + b_{11}\text{Jenis_Kelamin_up1} + b_{12}\text{Jalur_Masuk_up1} + b_{13}\text{Jenis_SMA_up1} + b_{14}\text{Pendapatan_up1} + b_{15}\text{Asal_Daerah_up1} + b_{16}\text{IPM_up1} + b_{17}\text{Pekerjaan_Ortu_up1}$$

b₁₁ adalah peningkatan satu unit dalam variabel Jenis_Kelamin_up1 dikaitkan dengan penurunan peluang log sebesar 26,88 .

b_{12} adalah peningkatan satu unit dalam variabel Jalur_Masuk_up1 dikaitkan dengan penurunan peluang log sebesar 39,66 .

Menghitung rasio probabilitas dalam memilih satu kategori hasil atas probabilitas memilih kategori dasar atau bisa disebut risiko relatif. Risiko relatif adalah persamaan linier sisi kanan yang dieksponekan, yang mengarah ke fakta bahwa koefisien regresi yang dieksponekan adalah rasio risiko relatif untuk perubahan unit dalam variabel prediktor. Untuk menghitung koefisien dari model digunakan untuk melihat rasio risiko dapat menggunakan script kode program R pada gambar 11.

```

```{r}
exp(coef(multimodel))
```
    
```

Hasil Kode program R untuk melihat rasio risiko model multinominal regresi logistic dapat dilihat pada gambar 14.

| | | | |
|--------------------------------|--------------------------------|--------------------------------|-----------------------------|
| (Intercept) | training \$jenis_kelamin_up12 | training \$jalur_Masuk_up12 | training \$jalur_Masuk_up13 |
| 2 8.300402e+16 | 4.752888e+11 | 5.954209e-18 | 2.342194e+25 |
| 3 2.243533e+07 | 6.438938e-15 | 4.803149e-37 | 4.356915e-37 |
| 4 6.378682e+08 | 1.907797e-33 | 2.132257e-11 | 3.264540e-24 |
| 5 3.012508e+09 | 6.144537e-57 | 7.390009e-01 | 2.195002e-14 |
| training \$jenis_SMA_up12 | training \$jenis_SMA_up13 | training \$Pendapatan_up12 | training \$Pendapatan_up13 |
| 2 1.885231e+06 | 3.064871e+27 | 1.554059e-09 | 7.585112e-09 |
| 3 2.656753e-11 | 1.056870e-10 | 2.417891e+18 | 1.438042e+41 |
| 4 1.593293e+00 | 6.736424e-03 | 4.721702e-12 | 2.375755e-06 |
| 5 9.090039e-24 | 2.150547e-05 | 5.847510e+18 | 2.510259e+14 |
| training \$Pendapatan_up14 | training \$Pendapatan_up15 | training \$Pendapatan_up16.0 | |
| 2 4.174077e-59 | 6.709138e-11 | 1.199706e-09 | |
| 3 7.023170e+21 | 1.149848e+37 | 5.303124e-12 | |
| 4 4.845333e+05 | 1.882750e-08 | 2.592275e+31 | |
| 5 7.713288e-16 | 1.471997e-06 | 2.813352e-04 | |
| training \$Pendapatan_up17 | training \$asal_Daerah_up12 | training \$asal_Daerah_up13 : | training \$IPM_up12 |
| 2 4.931414e-05 | 4.204998e-22 | 3.308554e-15 | 5.733243e-29 |
| 3 2.560561e+08 | 3.707335e+25 | 2.944255e+45 | 1.117202e-07 |
| 4 1.314330e-01 | 2.381917e+30 | 2.474554e-09 | 3.223058e-46 |
| 5 3.359328e-01 | 3.775675e-03 | 4.944919e-07 | 7.950876e+22 |
| training \$pekerjaan_Ortu_up12 | training \$pekerjaan_Ortu_up13 | training \$pekerjaan_Ortu_up14 | |
| 2 3.350446e-06 | 4.936794e+17 | 4.223179e+27 | |
| 3 9.974426e+14 | 4.003985e-32 | 2.018701e+01 | |
| 4 2.162354e-01 | 1.122451e-08 | 2.897404e-06 | |
| 5 8.529292e-02 | 7.551097e-20 | 3.739191e-07 | |

Gambar 14 Hasil Kode program R membuat rasio risiko model multinominal regresi logistic.

Rasio risiko relatif untuk peningkatan satu unit dalam variabel Jenis_Kelamin_up1 adalah 4,752 untuk Lama Studi '1' dengan Lama Studi '2'. Pembuatan model regresi logistik multinomial dilakukan berjenjang yaitu sebanyak 6 model, model regresi logistik multinomial pertama adalah multimodelfit1 dengan variabel respon lama studi dan variabel predictor jenis kelamin, model regresi logistik multinomial kedua adalah multimodelfit2 dengan variabel respon lama studi dan variabel predictor jenis kelamin dan jalur masuk, model regresi logistik multinomial ketiga adalah multimodelfit3 yaitu dengan variabel respon lama studi dan variabel predictor jenis kelamin, jalur masuk dan jenis masuk, model regresi logistik multinomial keempat adalah multimodelfit4 yaitu dengan variabel respon lama studi dan variabel predictor jenis kelamin, jalur masuk, jenis masuk dan Pendapatan, model regresi logistik multinomial kelima adalah multimodelfit5 yaitu dengan variabel respon lama studi dan variabel predictor jenis kelamin, jalur masuk, jenis masuk, Pendapatan dan asal daerah, model regresi logistik multinomial keenam adalah multimodelfit6 yaitu dengan variabel respon lama studi dan variabel predictor jenis kelamin, jalur masuk, jenis masuk, Pendapatan, asal daerah dan IPM, model regresi logistik multinomial ketujuh adalah multimodelfit7 yaitu dengan variabel respon lama studi dan variabel predictor jenis kelamin, jalur masuk, jenis masuk, Pendapatan, asal daerah, IPM dan pekerjaan orang tua dengan script kode program R pada gambar 15.

```
```\r}
library(nnet)
Run the multinomial model with the multinom function and summarize it
multimodelfit1 <- multinom (data = training, training $Lama_Studi_up1 ~ 1)
multimodelfit2 <- multinom(data = training, training $Lama_Studi_up1~ training $Jenis_Kelamin_up1)
multimodelfit3 <- multinom(data = training, training $Lama_Studi_up1~ training $Jenis_Kelamin_up1+
training $Jalur_Masuk_up1)
multimodelfit4 <- multinom(data = training, training $Lama_Studi_up1~ training $Jenis_Kelamin_up1+
training $Jalur_Masuk_up1+ training $Jenis_SMA_up1)
multimodelfit5 <- multinom(data = training, training $Lama_Studi_up1~ training $Jenis_Kelamin_up1+
training $Jalur_Masuk_up1+ training $Jenis_SMA_up1+ training $Pendapatan_up1)
multimodelfit6 <- multinom(data = training, training $Lama_Studi_up1~ training $Jenis_Kelamin_up1+
training $Jalur_Masuk_up1+ training $Jenis_SMA_up1+ training $Pendapatan_up1+ training
$Asal_Daerah_up1)
multimodelfit7 <- multinom(data = training, training $Lama_Studi_up1~ training $Jenis_Kelamin_up1+
training $Jalur_Masuk_up1+ training $Jenis_SMA_up1+ training $Pendapatan_up1+ training
$Asal_Daerah_up1+ training $IPM_up1)
multimodelfit8 <- multinom(data = training, training $Lama_Studi_up1~ training $Jenis_Kelamin_up1+
training $Jalur_Masuk_up1+ training $Jenis_SMA_up1+ training $Pendapatan_up1+ training
$Asal_Daerah_up1+ training $IPM_up1+ training $Pekerjaan_Ortu_up1)
```\r
```

Gambar 15 Kode program R membuat tujuh model multinomial regresi logistic dengan fungsi multinorm()

Hasil Kode program R membuat tujuh model multinomial regresi logistic dengan fungsi multinorm() dapat dilihat pada gambar 16.

```
# weights: 10 (4 variable)
initial value 45.064262
final value 37.626399
converged
# weights: 15 (8 variable)
initial value 45.064262
iter 10 value 34.586078
iter 20 value 34.488875
iter 30 value 34.488813
final value 34.488268
converged
# weights: 25 (16 variable)
initial value 45.064262
iter 10 value 29.171892
iter 20 value 28.524888
iter 30 value 28.520099
final value 28.520065
converged
# weights: 35 (24 variable)
initial value 45.064262
iter 10 value 23.917730
iter 20 value 23.241543
iter 30 value 23.240211
final value 23.240207
converged
# weights: 65 (48 variable)
initial value 45.064262
iter 10 value 15.303387
iter 20 value 14.090008
iter 30 value 14.074430
final value 14.074398
converged
# weights: 75 (56 variable)
initial value 45.064262
iter 10 value 11.685510
iter 20 value 9.378048
iter 30 value 9.364299
final value 9.364263
converged
# weights: 80 (60 variable)
initial value 45.064262
iter 10 value 10.077850
iter 20 value 6.947667
iter 30 value 6.931516
final value 6.931472
converged
# weights: 95 (72 variable)
initial value 45.064262
iter 10 value 4.298946
iter 20 value 1.391611
iter 30 value 1.386302
final value 1.386294
converged
```

Gambar 16 Hasil Kode program R membuat tujuh model multinomial regresi logistic dengan fungsi multinorm()

Pada penelitian ini menggunakan analisis varians (ANOVA) untuk menguji signifikansi model regresi, goodness of fit model dinilai secara tidak langsung dengan membandingkan model regresi dengan model regresi sebelum menambahkan variabel dengan script kode program R pada gambar 17.

```

```{r}
Melakukan test ANOVA pada tujuh model
anova (multimodelfit1, multimodelfit2, multimodelfit3, multimodelfit4, multimodelfit5,
multimodelfit6, multimodelfit7, multimodelfit8)
```
    
```

Gambar 17 Kode program R analisa anova pada tujuh model multinominal regresi logistic

Hasil Kode program R analisa anova pada tujuh model multinominal regresi logistic dapat dilihat pada gambar 18.

| | Resid. df | Resid. Dev | Test | Df | LR stat. | Pr(Chi) |
|--|-----------|------------|--------|----|-----------|------------|
| | 108 | 75.252798 | | NA | NA | NA |
| | 104 | 68.976537 | 1 vs 2 | 4 | 6.276261 | 0.17944503 |
| | 96 | 57.040130 | 2 vs 3 | 8 | 11.936406 | 0.15406391 |
| | 88 | 46.480415 | 3 vs 4 | 8 | 10.559716 | 0.22791520 |
| | 68 | 28.148796 | 4 vs 5 | 20 | 18.331619 | 0.56557294 |
| | 64 | 18.728525 | 5 vs 6 | 4 | 9.420271 | 0.05141154 |
| | 60 | 13.862944 | 6 vs 7 | 4 | 4.865581 | 0.30136983 |
| | 48 | 2.772589 | 7 vs 8 | 12 | 11.090355 | 0.52119155 |

Gambar 18 Hasil Kode program R keluaran dari model regresi logistik multinomial Hasil ANOVA dari perbandingan enam model.

Pada gambar 18, signifikansi model regresi dipertimbangkan dengan nilai $P < 0,10$ atau $P < 0,15$ saat membandingkan model regresi berganda. Nilai P pada multimodelfit6 menunjukkan bahwa model regresi logistik multinomial yang diusulkan adalah signifikan.

Tahap empat menguji koefisien model

Setelah membangun multinominal regresi logistic, tahap berikut adalah menghitung nilai P dari koefisien regresi dengan menggunakan uji Wald atau bisa disebut sebagai uji z. multinominal regresi logistic adalah mempunyai koefisien jika signifikasinya mempunyai nilai pada dua sisi sebesar $P < 0,05$ dengan script kode program R pada gambar 19.

```

```{r}
Wald tests/z tests of the coefficients
zvalues <- summary(multimodelfit6)$coefficients / summary(multimodelfit6)$standard.errors
pvalues <- pnorm(abs(zvalues), lower.tail = FALSE) * 2
pvalues
```
    
```

Gambar 19 Kode program R uji Wald pada model multinominal regresi logistic

Hasil Kode program R uji Wald pada model multinomial regresi logistic dapat dilihat pada gambar 20

```

R> naks producednaks produced (Intercept) training$jenis_kelamin_up12 training$jalur_masuk_up12 training$jalur_masuk_up13 training$jenis_SMA_up12
seer_ordinal$jenis_SMA_up13
2 0 0 0 0 0 0
3 0 0 0 0 0 0
4 0 0 0 0 0 0
5 n 0 0 0 0 0
training$pendapatan_up12 training$pendapatan_up13 training$pendapatan_up14 training$pendapatan_up15 training$pendapatan_up16 training$pendapatan_up17
2 0 0 0 0 0 0
3 0 0 0 0 0 0
4 0 0 0 0 0 0
5 0 0 0 0 0 0
training$sasal_Diberal_up12 training$sasal_Diberal_up13 training$SIPM_up12 training$Pekerjaan_Ortu_up12 training$Pekerjaan_Ortu_up13
2 0 0 0 0 0
3 0 0 0 0 0
4 0 0 0 0 0
5 0 0 0 0 0
training$Pekerjaan_Ortu_up14
2 0
3 0
4 0
5 NaN
    
```

Gambar 20 Hasil Kode program R uji Wald pada model multinomial regresi logistic

Pada gambar 20 didapat variabel dengan nilai P signifikan dari koefisien jika ditentukan sebagai faktor prognostik signifikan yang berkontribusi secara signifikan dengan lama studi terhadap jenis kelamin.

Tahap ke lima mengukur akurasi model

Selain itu, model regresi selanjutnya divalidasi dengan memprediksi hasil lama studi pada kumpulan data menggunakan fungsi prediksi script kode program R dapat dilihat pada gambar 21.

```

```{r}
Predict the category of outcomes for each patient
pred.multinom <- predict(multimodelfit6, testing)
pred.multinom
```
    
```

Gambar 21 Kode program R untuk prediksi dengan fungsi predict().

Hasil Kode program R untuk prediksi dengan fungsi predict() dapat dilihat pada gambar 22.

```

[1] 4 1 3 3 2 4 3 3 3 2 2 3 2 1
Levels: 1 2 3 4 5 6
    
```

Gambar 22 Hasil Kode program R untuk prediksi dengan fungsi predict().

Kemudian prediksi dari model terhadap lama studi pada kumpulan data diproses secara probabilitas menggunakan fungsi prediksi di R dapat dilihat pada gambar 23.

```

```{r}
Probabilities with each outcome for each patient
pprob <- predict(multimodelfit6, testing, type = "p")
head(pprob)
```
    
```

Gambar 23 Kode program R untuk prediksi dengan probabilitas.

Hasil Kode program R untuk prediksi dengan nilai probabilitas dapat dilihat pada gambar 24.

| | 1 | 2 | 3 | 4 | 5 |
|---|--------------|---------------|--------------|--------------|--------------|
| 1 | 7.502647e-83 | 2.484450e-129 | 1.000000e+00 | 1.887421e-95 | 1.452726e-39 |
| 2 | 1.427375e-35 | 1.000000e+00 | 1.543187e-15 | 1.198773e-34 | 3.730161e-91 |
| 3 | 6.209357e-21 | 1.662739e-16 | 5.000000e-01 | 5.000000e-01 | 3.118533e-14 |
| 4 | 4.305550e-27 | 1.000000e+00 | 1.133219e-67 | 9.500646e-53 | 2.111618e-81 |
| 5 | 7.483460e-68 | 7.396318e-109 | 1.000000e+00 | 8.706225e-80 | 1.698864e-23 |
| 6 | 1.869003e-36 | 9.119440e-60 | 1.000000e+00 | 2.440645e-56 | 3.203880e-11 |

Gambar 24 Hasil Kode program R untuk prediksi dengan nilai probabilitas

Untuk lebih memudahkan hasil model berupa prediksi dengan data real, maka dapat dibuat dengan satu data yaitu data_model script kode program R dapat dilihat pada gambar 25.

```

```{r}
data_model <- testing[c(1,3,18)]
data_model
```

```

Gambar 25 Kode program R untuk membuat data_model

Hasil Kode program R untuk membuat data_model yang berisi data prediksi data data real dapat dilihat pada gambar 26.

| | NIM <dbl> | Lama_Studi_up1 <fctr> | predict <fctr> |
|----|--------------|--------------------------|-------------------|
| 1 | 3 | 3 | 3 |
| 2 | 2 | 2 | 2 |
| 3 | 3 | 3 | 4 |
| 4 | 2 | 2 | 2 |
| 5 | 3 | 3 | 3 |
| 6 | 3 | 3 | 3 |
| 7 | 5 | 5 | 5 |
| 8 | 1 | 1 | 1 |
| 9 | 3 | 3 | 3 |
| 10 | 3 | 3 | 3 |

Gambar 26 Kode program R untuk membuat data_model

Tahap keenam menarik kesimpulan berdasarkan hasil analisis yang diperoleh.

Keakuratan model dalam memprediksi lama studi mahasiswa dengan fungsi mean(), keakuratan prediksi dapat diperoleh dengan menghitung rata-rata dari lama studi setiap data dengan lama studi predict. script kode program R dapat dilihat pada gambar 27.

```

```{r}
Calculate the prediction accuracy of the model
pred_accuracy <- mean(testing$predict == testing $Lama_Studi_up1)
pred_accuracy
```

```

Gambar 27 Kode program R untuk akurasi model

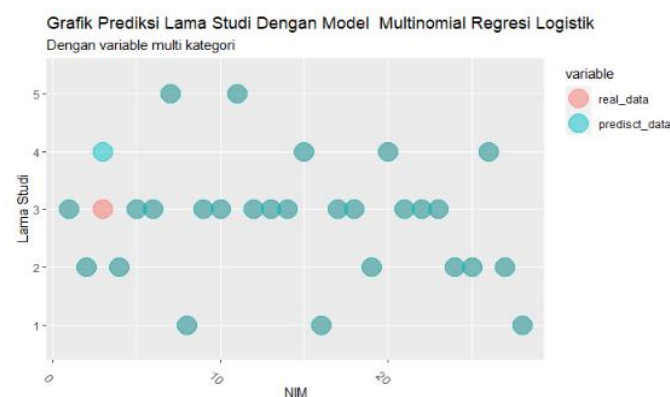
Hasil Kode program R untuk akurasi model dapat dilihat pada gambar 28.

[1] 0.9642857

Gambar 28 Hasil Kode program R untuk untuk akurasi model

Pada gambar 28 hasil keakuratan model menunjukkan bahwa 96,4% dari hasil lama studi dalam kumpulan data ini diverifikasi kebenarannya dengan menggunakan model regresi multinomial, dimana

hasil nya menunjukkan bahwa model regresi multinomial ini memiliki keandalan yang kuat. Dapat ditunjukkan plot data berupa grafik antara data real dengan data prediksi, plot warna merah adalah data real sedangkan plot warna biru adalah data prediksi, sedangkan plot warna biru gelap adalah irisan antara data real dengan data prediksi seperti pada gambar 27.



Gambar 29 Plot grafik data real dan data prediksi

KESIMPULAN

Dari model prediksi yang telah dibangun menggunakan metode multinomial regresi logistik dengan variable multi kategori untuk memprediksi keberhasilan mahasiswa di Perguruan Tinggi menggunakan perangkat lunak R. Model regresi logistic multinomial yang baik pada multimodelfit6 dengan variabel output adalah `training$Lama_Studi_up1` sedangkan variabel input adalah `training$Jenis_Kelamin_up1`, `training$Jalur_Masuk_up1`, `training$Jenis_SMA_up1`, `training $ Pendapatan_up1` dan `training $ Asal_Daerah_up1`. Hasil multimodelfit 6 menunjukkan akurasi sebesar 96,4% dari kumpulan data yang diverifikasi kebenarannya, sehingga dapat ditarik kesimpulan bahwa model regresi logistic multinomial yang dibangun memiliki keandalan yang kuat.

DAFTAR PUSTAKA

- Bi, G., Li, R., Liang, J., Hu, Z., & Zhan, C. (2020). A nomogram with enhanced function facilitated by nomogramEx and nomogramFormula. *Annals of Translational Medicine*, 8(4).
- Goodwin, B. (2011). *Simply better: Doing what matters most to change the odds for student success*. ASCD.
- Guswandi, D., Yanto, M., Hafizh, M., & Mayola, L. (2021). Analisis hybrid decision support system dalam penentuan status kelulusan mahasiswa. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(6), 1127–1136.
- Heryana, N., Ghassani, F. Z., Usrahmawan, M. A., Juardi, D., & Solehudin, A. (2020). Kajian model prediksi ketepatan pelulusan mahasiswa: Sebuah literatur review. *Syntax*, 9(1), 54–62.
- Lee, M., & Han, J. (2020). Statistical methods and models in the analysis of time to event data. *Annals of Translational Medicine*, 8(4).
- Liang, J., Bi, G., & Zhan, C. (2020). Multinomial and ordinal Logistic regression analyses with multi-categorical variables using R. *Annals of Translational Medicine*, 8(16).
- Nafi'iyah, N. (2016). Perbandingan regresi linear, backpropagation dan fuzzy mamdani dalam prediksi harga emas. *Prosiding SENIATI*, 291-B.
- Nugraha, J. (2014). *Pengantar analisis data kategorik: Metode dan aplikasi menggunakan program R*. Deepublish.

- Purba, M. Y., & Ruslan, S. (2020). Influence compensation, career development and job satisfaction to turnover intention. *Dinasti International Journal of Economics, Finance & Accounting*, 1(4), 642–658.
- Rabbi, F., Khan, S., Khalil, A., Mashwani, W. K., Shafiq, M., Göktaş, P., & Unvan, Y. A. (2021). Model selection in linear regression using paired bootstrap. *Communications in Statistics-Theory and Methods*, 50(7), 1629–1639.
- Rizki, N. A., Fendiyanto, P., & Jariah, A. (2020). Perbandingan klasifikasi penjurusan peserta didik pada model diskriminan dan regresi logistik multinomial. *METIK JURNAL*, 4(2), 49–54.
- Santosa, R. G., & Chrismanto, A. R. (2018). Perbandingan akurasi regresi logistik dengan regresi multinomial untuk prediksi kategori ip mahasiswa jalur prestasi. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, 4(2), 99–107.
- Sarwono, J. (2013). *12 jurus ampuh SPSS untuk riset skripsi*. Elex Media Komputindo.
- Utama, M. B. R., & Hajarisman, N. (2021). Metode pemilihan variabel pada model regresi poisson menggunakan metode nordberg. *Jurnal Riset Statistika*, 1(1), 35–42.
- Zhou, Z.-R., Wang, W.-W., Li, Y., Jin, K.-R., Wang, X.-Y., Wang, Z.-W., Chen, Y.-S., Wang, S.-J., Hu, J., & Zhang, H.-N. (2019). In-depth mining of clinical data: The construction of clinical prediction model with R. *Annals of Translational Medicine*, 7(23).