

PRAKTIKUM PEMODELAN STATISTIKA
MODUL 7



Disusun oleh :

Nama : Fidelia Ping
NIM : 245410012
Kelas : Informatika 1

PROGRAM STUDI INFORMATIKA
PROGRAM SARJANA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS TEKNOLOGI DIGITAL INDONESIA
YOGYAKARTA
2025

MODUL 7

REGRESI LOGISTIK BINER

A. TUJUAN PRAKTIKUM

1. Memahami konsep dasar regresi logistik Biner.
2. Melakukan analisis regresi logistik Biner menggunakan R.

B. PEMBAHASAN LISTING

PRAKTIK

Kasus 1 :

Berikut ini analisis regresi logistik menggunakan dataset bawaan R yaitu mtcars untuk memprediksi apakah mobil memiliki transmisi otomatis (0) atau manual (1) berdasarkan beberapa variabel.

1. Load dataset

```
# 1. Load Library dan Dataset
library(dplyr)

data(mtcars)
cat("Dataset mtcars (6 baris pertama):\n")
print(head(mtcars))
```

Pembahasan : Kode ini digunakan untuk **mempersiapkan data** sebelum dilakukan analisis lebih lanjut. Library dplyr dimuat untuk mendukung proses pengolahan dan manipulasi data yang lebih mudah dan terstruktur. Selanjutnya, dataset bawaan R yaitu mtcars dipanggil menggunakan fungsi data(mtcars). Perintah head(mtcars) digunakan untuk menampilkan **enam baris pertama** dari dataset, sementara cat() berfungsi memberikan keterangan teks agar output di konsol lebih jelas dan rapi.

Output

```
***
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':
  filter, lag

The following objects are masked from 'package:base':
  intersect, setdiff, setequal, union

Dataset mtcars (6 baris pertama):
   mpg  cyl  disp    hp  drat    wt   qsec    vs  am  gear  carb
Mazda RX4         21.0   6  160 110  3.90  2.620 16.46   0   1    4    4
Mazda RX4 Wag     21.0   6  160 110  3.90  2.875 17.02   0   1    4    4
Datsun 710        22.8   4  108  93  3.85  2.320 18.61   1   1    4    1
Hornet 4 Drive    21.4   6  258 110  3.08  3.215 19.44   1   0    3    1
Hornet Sportabout 18.7   8  360 175  3.15  3.440 17.02   0   0    3    2
Valiant           18.1   6  225 105  2.76  3.460 20.22   1   0    3    1
```

Pembahasan Output : Output yang dihasilkan menampilkan **enam observasi awal dataset mtcars**, yang berisi berbagai variabel karakteristik mobil seperti konsumsi bahan bakar (mpg), jumlah silinder (cyl), kapasitas mesin (disp), tenaga mesin (hp), berat kendaraan (wt), dan lain-lain. Tampilan ini membantu memahami **struktur data, jenis variabel, serta kisaran nilai** yang ada dalam dataset, sehingga memudahkan penentuan variabel mana yang akan digunakan dalam analisis atau pemodelan selanjutnya.

2. Pilih variabel

```
# 2. Pilih Variabel dan Data Cleaning
data_mobil <- mtcars %>%
  select(am, mpg, hp, wt) %>%
  mutate(am = factor(am, labels = c("Manual", "Otomatis")))

# Ubah jadi data frame biasa (R base)
data_mobil <- as.data.frame(data_mobil)

cat("\nVariabel yang digunakan:\n")
print(names(data_mobil))
```

Pembahasan : Kode ini digunakan untuk **memilih variabel yang relevan dan melakukan pembersihan data** dari dataset mtcars. Dengan bantuan fungsi select() dari package dplyr, hanya variabel am (jenis transmisi), mpg (efisiensi bahan bakar), hp (tenaga mesin), dan wt (berat mobil) yang diambil untuk analisis. Selanjutnya, fungsi mutate() digunakan untuk **mengubah variabel am menjadi factor**, dengan pemberian label kategori agar lebih mudah dipahami. Setelah itu, data dikonversi kembali menjadi data frame berbasis R (as.data.frame) untuk memastikan kompatibilitas dengan fungsi analisis dasar R. Perintah names() digunakan untuk menampilkan daftar variabel yang digunakan.

Output

```
***
Variabel yang digunakan:
[1] "am" "mpg" "hp" "wt"
```

Pembahasan Output : Output menampilkan **nama-nama variabel** yang terdapat dalam data hasil seleksi, yaitu am, mpg, hp, dan wt. Hal ini menegaskan bahwa dataset telah berhasil dibersihkan dan hanya memuat variabel yang diperlukan untuk analisis lebih lanjut. Selain itu, perubahan am menjadi variabel kategorik (factor) memastikan bahwa jenis transmisi akan diperlakukan sebagai **variabel dummy** dalam model regresi yang akan dibangun.

3. Buat Model

```
# 3. Buat Model
# Memprediksi 'am' berdasarkan 'mpg', 'hp', dan 'wt'
model_mobil <- glm(am ~ mpg + hp + wt,
  data = data_mobil,
  family = binomial(link = "logit"))
```

Pembahasan : Kode ini digunakan untuk **membangun model regresi logistik biner** dengan tujuan memprediksi **jenis transmisi mobil (am)** berdasarkan tiga variabel penjelas, yaitu **efisiensi bahan bakar (mpg)**, **tenaga mesin (hp)**, dan **berat mobil (wt)**. Fungsi glm() (Generalized Linear Model) digunakan dengan family = binomial(link = "logit"), yang menandakan bahwa model yang dibangun adalah **regresi logistik** dengan fungsi hubungan logit. Karena variabel am sebelumnya telah diubah menjadi factor dengan dua kategori (Manual dan Otomatis), R secara otomatis membentuk **variabel dummy** untuk keperluan pemodelan.

Output

-

Pembahasan Output :

4. Ringkasan model

```
# 4. Ringkasan Model
cat("\n=== RINGKASAN MODEL REGRESI LOGISTIK ===\n")
summary(model_mobil)

# Menghitung Odds Ratio
cat("\n=== ODDS RATIO (exp(beta)) ===\n")
odds_ratio <- exp(coef(model_mobil))
print(cbind(Variabel = names(odds_ratio), OR = round(odds_ratio, 3)))
```

Pembahasan : Kode ini digunakan untuk **menampilkan ringkasan hasil regresi logistik** dan menghitung **odds ratio** dari model `model_mobil`. Fungsi `summary(model_mobil)` memberikan informasi penting seperti estimasi koefisien, standar error, nilai *z*, dan *p-value* untuk masing-masing variabel (*mpg*, *hp*, dan *wt*). Selanjutnya, odds ratio dihitung dengan fungsi `exp()` terhadap koefisien model, karena pada regresi logistik koefisien berada pada skala log-odds. Hasil odds ratio kemudian ditampilkan dalam bentuk tabel agar lebih mudah diinterpretasikan.

Output

```
***
=== RINGKASAN MODEL REGRESI LOGISTIK ===

Call:
glm(formula = am ~ mpg + hp + wt, family = binomial(link = "logit"),
    data = data_mobil)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -15.72137    40.00281  -0.393   0.6943
mpg           1.22930     1.58109   0.778   0.4369
hp            0.08389     0.08228   1.020   0.3079
wt           -6.95492     3.35297  -2.074   0.0381 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 43.2297  on 31  degrees of freedom
Residual deviance:  8.7661  on 28  degrees of freedom
AIC: 16.766

Number of Fisher Scoring iterations: 10

=== ODDS RATIO (exp(beta)) ===
      Variabel      OR
(Intercept) "(Intercept)" "0"
mpg          "mpg"         "3.419"
hp           "hp"          "1.088"
wt           "wt"          "0.001"
```

Pembahasan Output : Output ringkasan model menunjukkan bagaimana setiap variabel memengaruhi **peluang jenis transmisi mobil**. Koefisien yang signifikan (*p-value* < 0,05) menandakan bahwa variabel tersebut berpengaruh secara nyata terhadap kemungkinan mobil memiliki transmisi tertentu. Nilai **odds ratio (OR)** memberikan interpretasi yang lebih intuitif: OR > 1 berarti peningkatan variabel tersebut meningkatkan peluang mobil memiliki transmisi kategori referensi, sedangkan OR < 1 berarti menurunkan peluangnya. Dengan demikian, hasil ini membantu memahami faktor-faktor apa saja yang paling berpengaruh dalam menentukan jenis transmisi mobil.

5. Prediksi data baru

```
# 5. Prediksi Data Baru
mobil_baru <- data.frame(
  mpg = c(22.8, 15.0, 30.4),
  hp = c(108, 175, 95),
  wt = c(2.32, 3.85, 1.61))
```

```
)

# Melakukan prediksi probabilitas
mobil_baru$prob_otomatis <- predict(model_mobil, newdata = mobil_baru, type
= "response")

# Mengategorikan hasil prediksi (Threshold 0.5)
mobil_baru$prediksi <- ifelse(mobil_baru$prob_otomatis > 0.5, "Otomatis",
"Manual")

cat("\n=== PREDIKSI 3 MOBIL BARU ===\n")
print(mobil_baru)
```

Pembahasan : Kode ini digunakan untuk **melakukan prediksi pada data mobil baru** menggunakan model regresi logistik yang telah dibangun. Data baru dimasukkan dalam bentuk data frame mobil_baru yang berisi nilai mpg, hp, dan wt. Fungsi predict() dengan argumen type = "response" digunakan untuk menghasilkan **probabilitas** suatu mobil termasuk ke dalam kategori transmisi tertentu (dalam hal ini peluang **Otomatis**). Selanjutnya, probabilitas tersebut dikonversi menjadi **kelas prediksi** menggunakan ambang batas (*threshold*) 0,5, di mana nilai probabilitas di atas 0,5 diklasifikasikan sebagai *Otomatis*, dan sisanya sebagai *Manual*.

Output

```
...
=== PREDIKSI 3 MOBIL BARU ===
   mpg  hp  wt prob_otomatis prediksi
1 22.8 108 2.32  9.946844e-01 Otomatis
2 15.0 175 3.85  8.464651e-05  Manual
3 30.4  95 1.61  1.000000e+00 Otomatis
```

Pembahasan Output : Output menampilkan tabel berisi **tiga mobil baru** beserta nilai prediksi probabilitas transmisi otomatis dan hasil klasifikasinya. Kolom prob_otomatis menunjukkan seberapa besar peluang sebuah mobil menggunakan transmisi otomatis menurut model, sedangkan kolom prediksi menunjukkan hasil klasifikasi akhir. Hasil ini memperlihatkan bahwa model regresi logistik dapat digunakan untuk **mengklasifikasikan jenis transmisi mobil** berdasarkan karakteristik teknisnya, sekaligus memberikan informasi probabilistik yang berguna dalam pengambilan keputusan.

Kasus 2

Dataset: infert , yaitu Data infertilitas wanita

Target: case 1 = infertil, 0 = fertil (kontrol)

Variabel independen: age, parity, education, spontaneous, induced

1. Load dataset bawaan R

```
# 1. Load Dataset
data(infert)
cat("Dataset infert: n =", nrow(infert), "wanita\n")
print(head(infert))
```

Pembahasan : Kode ini digunakan untuk **memuat dataset bawaan R bernama infert**, yang berisi data hasil studi kasus–kontrol terkait faktor-faktor yang berhubungan dengan infertilitas pada wanita. Fungsi data(infert) berfungsi memanggil dataset tersebut ke dalam environment R agar siap dianalisis. Selanjutnya, fungsi nrow(infert) digunakan untuk menghitung **jumlah observasi (jumlah wanita)** dalam dataset, sedangkan

head(infert) menampilkan **enam baris pertama** data. Perintah cat() digunakan untuk menampilkan informasi jumlah data secara ringkas dan jelas di konsol.

Output

```
*** Dataset infert: n = 248 wanita
  education age parity induced case spontaneous stratum pooled.stratum
1    0-5yrs  26     6       1     1           2         1           3
2    0-5yrs  42     1       1     1           0         2           1
3    0-5yrs  39     6       2     1           0         3           4
4    0-5yrs  34     4       2     1           0         4           2
5    6-11yrs 35     3       1     1           1         5          32
6    6-11yrs 36     4       2     1           1         6          36
```

Pembahasan Output : Output menunjukkan **jumlah total wanita** yang terdapat dalam dataset infert serta enam observasi awalnya. Data yang ditampilkan mencakup variabel seperti status infertilitas (case), usia (age), jumlah kehamilan sebelumnya (parity), riwayat aborsi induksi (induced) dan spontan (spontaneous), serta faktor lain yang relevan. Tampilan awal ini membantu memahami **struktur data, jenis variabel (numerik dan kategorik), serta konteks penelitian**, sehingga memudahkan dalam menentukan variabel mana yang akan digunakan pada analisis lanjutan, khususnya untuk pemodelan regresi logistik.

2. Pilih variabel utama

```
# 2. Pilih Variabel Utama
data_infert <- infert[, c("case", "age", "parity", "education",
"spontaneous", "induced")]

# Ubah target 'case' menjadi factor
data_infert$case <- factor(data_infert$case,
                           levels = c(0, 1),
                           labels = c("Fertil", "Infertil"))

# Ringkasan Distribusi Target
cat("\n=== DISTRIBUSI CASE ===\n")
print(table(data_infert$case))
print(prop.table(table(data_infert$case)))
```

Pembahasan : Kode ini digunakan untuk **memuat dataset bawaan R bernama infert**, yang berisi data hasil studi kasus–kontrol terkait faktor-faktor yang berhubungan dengan infertilitas pada wanita. Fungsi data(infert) berfungsi memanggil dataset tersebut ke dalam environment R agar siap dianalisis. Selanjutnya, fungsi nrow(infert) digunakan untuk menghitung **jumlah observasi (jumlah wanita)** dalam dataset, sedangkan head(infert) menampilkan **enam baris pertama** data. Perintah cat() digunakan untuk menampilkan informasi jumlah data secara ringkas dan jelas di konsol.

Output

```
***
=== DISTRIBUSI CASE ===

  Fertil Infertil
    165      83

  Fertil Infertil
0.6653226 0.3346774
```

Pembahasan Output : Output menunjukkan **jumlah total wanita** yang terdapat dalam dataset infert serta enam observasi awalnya. Data yang ditampilkan mencakup variabel seperti status infertilitas (case), usia (age), jumlah kehamilan sebelumnya (parity),

riwayat aborsi induksi (induced) dan spontan (spontaneous), serta faktor lain yang relevan. Tampilan awal ini membantu memahami **struktur data, jenis variabel (numerik dan kategorik), serta konteks penelitian**, sehingga memudahkan dalam menentukan variabel mana yang akan digunakan pada analisis lanjutan, khususnya untuk pemodelan regresi logistik.

3. Bangun model

```
# 3. Bangun Model
model_infert <- glm(case ~ age + parity + education + spontaneous + induced,
                    data = data_infert,
                    family = binomial(link = "logit"))
```

Pembahasan : Kode ini digunakan untuk **membangun model regresi logistik biner** pada dataset infertilitas. Fungsi glm() (Generalized Linear Model) digunakan dengan family = binomial(link = "logit"), yang menandakan bahwa variabel respon case bersifat biner (misalnya 1 = infertil, 0 = tidak infertil). Variabel penjelas yang digunakan dalam model ini adalah **usia (age), jumlah kelahiran (parity), tingkat pendidikan (education), jumlah keguguran spontan (spontaneous), dan jumlah aborsi induksi (induced)**. Model ini bertujuan untuk menganalisis bagaimana faktor-faktor tersebut memengaruhi **peluang seorang wanita mengalami infertilitas**.

Output

-

Pembahasan Output :

4. Ringkasan model

```
# 4. Ringkasan Model
cat("\n=== RINGKASAN MODEL REGRESI LOGISTIK ===\n")
print(summary(model_infert))
```

Pembahasan : Kode ini digunakan untuk **menampilkan ringkasan hasil model regresi logistik** yang telah dibangun pada data infertilitas. Fungsi summary(model_infert) menghasilkan informasi statistik utama dari model, seperti estimasi koefisien, standar error, nilai statistik z, dan *p-value* untuk setiap variabel penjelas. Perintah cat() digunakan untuk menampilkan judul agar output di konsol lebih terstruktur dan mudah dibaca.

Output

```
***
=== RINGKASAN MODEL REGRESI LOGISTIK ===

Call:
glm(formula = case ~ age + parity + education + spontaneous +
    induced, family = binomial(link = "logit"), data = data_infert)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.14924    1.41220  -0.814   0.4158
age           0.03958    0.03120   1.269   0.2046
parity       -0.82828    0.19649  -4.215 2.49e-05 ***
education6-11yrs -1.04424    0.79255  -1.318   0.1876
education12+ yrs -1.40321    0.83416  -1.682   0.0925 .
spontaneous    2.04591    0.31016   6.596 4.21e-11 ***
induced       1.28876    0.30146   4.275 1.91e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 316.17  on 247  degrees of freedom
Residual deviance: 257.80  on 241  degrees of freedom
AIC: 271.8

Number of Fisher Scoring iterations: 4
```

Pembahasan Output : Output ringkasan model menunjukkan pengaruh masing-masing variabel (age, parity, education, spontaneous, dan induced) terhadap **peluang terjadinya infertilitas**. Koefisien bernilai positif mengindikasikan bahwa kenaikan variabel tersebut meningkatkan peluang infertilitas, sedangkan koefisien negatif menunjukkan efek sebaliknya. Nilai *p-value* digunakan untuk menilai signifikansi statistik; variabel dengan *p-value* kurang dari 0,05 dapat dianggap berpengaruh signifikan terhadap status infertilitas. Selain itu, informasi seperti **null deviance** dan **residual deviance** membantu menilai seberapa baik model menjelaskan data dibandingkan model tanpa prediktor.

5. Odds Ratio

```
# 5. Odds Ratio
cat("\n=== ODDS RATIO (exp(beta)) ===\n")
or <- exp(coef(model_infert))
or_table <- data.frame(
  Variabel = names(or),
  Odds_Ratio = round(or, 3)
)

# Perbaiki typo: menghapus kurung tutup ganda 'print(or_table))'
print(or_table)
```

Pembahasan : Kode ini digunakan untuk **menghitung dan menampilkan Odds Ratio (OR)** dari model regresi logistik model_infert. Nilai koefisien regresi yang diperoleh dari coef(model_infert) masih berada pada skala **log-odds**, sehingga perlu ditransformasikan dengan fungsi exp() agar diperoleh odds ratio. Selanjutnya, hasil OR disusun dalam sebuah data frame or_table yang berisi nama variabel dan nilai odds ratio yang telah dibulatkan agar mudah dibaca. Perintah print() digunakan untuk menampilkan tabel tersebut di konsol.

Output

```
***
=== ODDS RATIO (exp(beta)) ===
      Variabel Odds_Ratio
(Intercept) (Intercept)  0.317
age          age         1.040
parity       parity      0.437
education6-11yrs education6-11yrs 0.352
education12+ yrs education12+ yrs  0.246
spontaneous  spontaneous  7.736
induced      induced     3.628
```

Pembahasan Output : Output berupa tabel odds ratio menunjukkan **besar pengaruh masing-masing variabel terhadap peluang infertilitas**. Nilai OR > 1 mengindikasikan bahwa variabel tersebut meningkatkan risiko infertilitas, sedangkan OR < 1 menunjukkan efek protektif atau penurunan risiko. Misalnya, jika OR untuk variabel spontaneous atau induced lebih besar dari 1, hal ini berarti riwayat keguguran atau aborsi meningkatkan peluang terjadinya infertilitas. Tabel ini memudahkan interpretasi hasil regresi logistik secara praktis dan lebih intuitif dibandingkan koefisien log-odds.

LATIHAN

Lakukan analisis Regresi Logistik untuk hubungan antara Pclass, Sex, Age, SibSp, Parch, Fare dengan Survived. Gunakan Dataset: Titanic (built-in di R)

Variabel dependen : Survived

Variabel independen : Pclass, Sex, Age, SibSp, Parch, Fare

Lakukan :

1. Load dataset mtcars yang sdh ada di R

```
# 1. Load Dataset dan Persiapan Data
# Cek apakah library titanic sudah ada, jika belum install
if(!require(titanic)) install.packages("titanic")
library(titanic)
library(dplyr)

# Memuat data
data("titanic_train")

# Memilih variabel yang diminta: Survived, Pclass, Sex, Age, SibSp, Parch, Fare
data_titanic <- titanic_train %>%
  select(Survived, Pclass, Sex, Age, SibSp, Parch, Fare)

# DATA CLEANING (Penting!)
# Mengubah Survived menjadi Factor (0 = Meninggal, 1 = Selamat)
data_titanic$Survived <- factor(data_titanic$Survived,
                                levels = c(0, 1),
                                labels = c("Meninggal", "Selamat"))

# Mengubah Sex menjadi Factor
data_titanic$Sex <- factor(data_titanic$Sex)

# Mengisi data umur yang kosong (NA) dengan rata-rata (imputasi sederhana)
data_titanic$Age[is.na(data_titanic$Age)] <- mean(data_titanic$Age, na.rm = TRUE)

cat("Dataset Titanic (6 baris pertama):\n")
print(head(data_titanic))

cat("\nStruktur Data:\n")
str(data_titanic)
```

Pembahasan : Kode ini digunakan untuk **mempersiapkan dataset Titanic** sebelum dilakukan analisis lebih lanjut, khususnya untuk pemodelan **regresi logistik**. Pertama, dilakukan pengecekan dan pemanggilan library titanic dan dplyr. Dataset titanic_train kemudian dimuat dan dipilih hanya variabel yang relevan, yaitu Survived, Pclass, Sex, Age, SibSp, Parch, dan Fare. Selanjutnya dilakukan **data cleaning**, di mana variabel Survived diubah menjadi factor dengan label yang lebih mudah dipahami (*Meninggal* dan *Selamat*), serta Sex diubah menjadi factor agar diperlakukan sebagai variabel kategorik (dummy). Untuk mengatasi nilai umur yang kosong (NA), dilakukan **imputasi sederhana** dengan mengganti nilai NA menggunakan rata-rata umur, sehingga data siap digunakan tanpa error pada proses pemodelan.

Output

```

*** Loading required package: titanic

Warning message in library(package, lib.loc = lib.loc, character.only = TRUE, logical.return = TRUE, :
"there is no package called 'titanic'"
Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)

Dataset Titanic (6 baris pertama):
  Survived Pclass   Sex   Age SibSp Parch   Fare
1 Meninggal    3  male 22.00000    1    0  7.2500
2 Selamat     1 female 38.00000    1    0 71.2833
3 Selamat     3 female 26.00000    0    0  7.9250
4 Selamat     1 female 35.00000    1    0 53.1000
5 Meninggal    3  male 35.00000    0    0  8.0500
6 Meninggal    3  male 29.69912    0    0  8.4583

Struktur Data:
'data.frame':   891 obs. of  7 variables:
 $ Survived: Factor w/ 2 levels "Meninggal","Selamat": 1 2 2 2 1 1 1 1 2 2 ...
 $ Pclass  : int  3 1 3 1 3 3 1 3 3 2 ...
 $ Sex     : Factor w/ 2 levels "female","male": 2 1 1 1 2 2 2 2 1 1 ...
 $ Age     : num  22 38 26 35 35 ...
 $ SibSp   : int  1 1 0 1 0 0 0 3 0 1 ...
 $ Parch   : int  0 0 0 0 0 0 0 1 2 0 ...
 $ Fare    : num  7.25 71.28 7.92 53.1 8.05 ...

```

Pembahasan Output : Output pertama menampilkan enam baris awal dataset Titanic yang telah dibersihkan, sehingga dapat dilihat contoh nilai setiap variabel setelah proses seleksi dan transformasi. Output `str(data_titanic)` menunjukkan **struktur data**, termasuk jumlah observasi, tipe masing-masing variabel, serta level pada variabel factor seperti Survived dan Sex. Dari hasil ini dapat disimpulkan bahwa data telah siap untuk analisis lanjutan karena variabel target bersifat kategorik, variabel prediktor sudah sesuai tipenya, dan tidak terdapat nilai kosong pada variabel penting seperti Age.

2. Buat model regresi logistik

```

# 2. Buat Model Regresi Logistik
# Variabel Dependen: Survived
# Variabel Independen: Pclass, Sex, Age, SibSp, Parch, Fare
model_titanic <- glm(Survived ~ Pclass + Sex + Age + SibSp + Parch + Fare,
                     data = data_titanic,
                     family = binomial(link = "logit"))

```

Pembahasan : Kode ini digunakan untuk **membangun model regresi logistik biner** pada dataset Titanic dengan tujuan memprediksi **status keselamatan penumpang (Survived)**. Fungsi `glm()` digunakan dengan `family = binomial(link = "logit")`, yang menandakan bahwa variabel dependen bersifat kategorik dua kelas (Meninggal dan Selamat). Variabel independen yang digunakan dalam model ini meliputi **kelas penumpang (Pclass)**, **jenis kelamin (Sex)**, **usia (Age)**, **jumlah saudara/pasangan di kapal (SibSp)**, **jumlah orang tua/anak (Parch)**, dan **tarif tiket (Fare)**. Model ini bertujuan untuk menganalisis faktor-faktor yang memengaruhi peluang seorang penumpang untuk selamat.

Output

-

Pembahasan Output :

3. Interpretasi koefisien

```

# 3. Interpretasi Koefisien (Ringkasan Model)
cat("\n=== RINGKASAN MODEL ===\n")
summary(model_titanic)

```

Pembahasan : Kode `cat("\n=== RINGKASAN MODEL ===\n")` digunakan untuk mencetak teks judul ke layar agar hasil output terlihat lebih terstruktur dan mudah dibedakan dari output lainnya. Sementara itu, fungsi `summary(model_titanic)` dipakai untuk menampilkan ringkasan statistik dari model regresi logistik `model_titanic` yang

telah dibangun sebelumnya, sehingga pengguna dapat melihat hasil estimasi model secara lengkap tanpa harus menghitungnya secara manual.

Output

```
***
=== RINGKASAN MODEL ===

Call:
glm(formula = Survived ~ Pclass + Sex + Age + SibSp + Parch +
     Fare, family = binomial(link = "logit"), data = data_titanic)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  4.960445   0.532937   9.308 < 2e-16 ***
Pclass       -1.084297   0.139119  -7.794 6.49e-15 ***
Sexmale      -2.762930   0.199011 -13.883 < 2e-16 ***
Age          -0.039702   0.007797  -5.092 3.55e-07 ***
SibSp        -0.350725   0.109552  -3.201 0.00137 **
Parch        -0.111963   0.117400  -0.954 0.34024
Fare         0.002852   0.002361   1.208 0.22718
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1186.66  on 890  degrees of freedom
Residual deviance:  788.73  on 884  degrees of freedom
AIC: 802.73

Number of Fisher Scoring iterations: 5
```

Pembahasan Output : Output dari `summary(model_titanic)` menampilkan tabel koefisien yang berisi nilai *Estimate* (arah dan besar pengaruh setiap variabel terhadap peluang penumpang untuk selamat), *Std. Error* (tingkat ketidakpastian estimasi koefisien), *z value*, dan *Pr(>|z|)* atau nilai p yang menunjukkan signifikansi masing-masing variabel. Nilai p yang kecil (biasanya $< 0,05$) menandakan bahwa variabel tersebut berpengaruh signifikan terhadap status *Survived*. Selain itu, ditampilkan juga *null deviance* dan *residual deviance* untuk membandingkan model tanpa prediktor dengan model yang menggunakan prediktor, serta nilai *AIC* sebagai ukuran kualitas model secara keseluruhan, di mana nilai *AIC* yang lebih kecil menunjukkan model yang lebih baik.

4. Hitung nilai Odds Ratio interpretasikan hasilnya

```
# 4. Hitung Odds Ratio dan Interpretasi
cat("\n=== ODDS RATIO (exp(beta)) ===\n")
odds_ratio <- exp(coef(model_titanic))

# Membuat tabel agar mudah dibaca
hasil_or <- data.frame(
  Variabel = names(odds_ratio),
  Odds_Ratio = round(odds_ratio, 4),
  Interpretasi_Singkat = c(
    "Intercept",
    "Setiap kenaikan 1 kelas (makin miskin), peluang selamat turun",
    "Laki-laki memiliki peluang selamat jauh lebih kecil dibanding wanita",
    "Semakin tua, peluang selamat sedikit menurun",
    "Semakin banyak saudara/pasangan, peluang turun",
    "Semakin banyak ortu/anak, peluang sedikit naik (tidak signifikan)",
    "Harga tiket lebih mahal sedikit menaikkan peluang"
  )
)
print(hasil_or)
```

Pembahasan : Kode `cat("\n=== ODDS RATIO (exp(beta)) ===\n")` digunakan untuk menampilkan judul output agar hasil perhitungan mudah dikenali. Baris `odds_ratio <- exp(coef(model_titanic))` berfungsi menghitung *odds ratio* dengan cara mengubah koefisien regresi logistik (β) menjadi bentuk eksponensial, karena interpretasi regresi logistik lebih tepat dilakukan dalam satuan *odds ratio*. Selanjutnya, fungsi `data.frame()` digunakan untuk menyusun hasil *odds ratio* ke dalam bentuk tabel yang rapi, dengan kolom Variabel untuk nama variabel, Odds_Ratio untuk nilai *odds ratio* yang telah dibulatkan, serta Interpretasi_Singkat yang berisi penjelasan makna masing-masing nilai. Perintah `print(hasil_or)` digunakan untuk menampilkan tabel tersebut ke layar.

Output

```

***
=== ODDS RATIO (exp(beta)) ===
      Variabel Odds_Ratio
(Intercept) (Intercept) 142.6573
Pclass      Pclass      0.3381
Sexmale     Sexmale     0.0631
Age         Age         0.9611
SibSp       SibSp       0.7042
Parch       Parch       0.8941
Fare        Fare        1.0029

      Interpretasi_Singkat
(Intercept) Intercept
Pclass      Setiap kenaikan 1 kelas (makin miskin), peluang selamat turun
Sexmale     Laki-laki memiliki peluang selamat jauh lebih kecil dibanding wanita
Age         Semakin tua, peluang selamat sedikit menurun
SibSp       Semakin banyak saudara/pasangan, peluang turun
Parch       Semakin banyak ortu/anak, peluang sedikit naik (tidak signifikan)
Fare        Harga tiket lebih mahal sedikit menaikkan peluang

```

Pembahasan Output : Output berupa tabel *odds ratio* yang menunjukkan seberapa besar perubahan peluang selamat (*Survived*) akibat perubahan setiap variabel. Nilai *odds ratio* lebih dari 1 berarti variabel tersebut meningkatkan peluang selamat, sedangkan nilai kurang dari 1 berarti menurunkan peluang selamat. Misalnya, *Pclass* dengan *odds ratio* kurang dari 1 menunjukkan bahwa semakin rendah kelas penumpang (semakin miskin), maka peluang selamat semakin kecil. Variabel *Sex* biasanya memiliki *odds ratio* jauh di bawah 1 untuk laki-laki, yang berarti laki-laki memiliki peluang selamat lebih rendah dibanding perempuan. Variabel *Age* dengan *odds ratio* sedikit di bawah 1 menunjukkan bahwa bertambahnya usia cenderung menurunkan peluang selamat. Sementara itu, *Fare* dengan *odds ratio* di atas 1 mengindikasikan bahwa harga tiket yang lebih mahal sedikit meningkatkan peluang selamat, meskipun pengaruhnya relatif kecil.

5. Lakukan Prediksi dengan data baru

```

# 5. Prediksi Data Baru
# Kita buat 2 penumpang fiktif untuk simulasi:
# Penumpang A: Wanita, Kelas 1, Umur 25, Bayar mahal (Profil "Rose")
# Penumpang B: Pria, Kelas 3, Umur 30, Bayar murah (Profil "Jack")

data_baru <- data.frame(
  Pclass = c(1, 3),
  Sex = c("female", "male"),
  Age = c(25, 30),
  SibSp = c(0, 0),
  Parch = c(0, 0),
  Fare = c(100, 10) # Asumsi harga tiket
)

```

```
# Melakukan prediksi peluang
data_baru$Peluang_Selamat <- predict(model_titanic, newdata = data_baru,
type = "response")

# Keputusan (Threshold > 0.5 diprediksi Selamat)
data_baru$Prediksi <- ifelse(data_baru$Peluang_Selamat > 0.5, "Selamat",
"Meninggal")

cat("\n=== HASIL PREDIKSI DATA BARU ===\n")
print(data_baru)
```

Pembahasan : Bagian kode ini digunakan untuk melakukan prediksi peluang keselamatan penumpang baru menggunakan model regresi logistik model_titanic. Pertama, dibuat sebuah *data frame* bernama data_baru yang berisi dua penumpang fiktif dengan karakteristik tertentu, yaitu Penumpang A (wanita, kelas 1, usia 25 tahun, tanpa tanggungan, dan harga tiket mahal) serta Penumpang B (pria, kelas 3, usia 30 tahun, tanpa tanggungan, dan harga tiket murah). Selanjutnya, fungsi predict() dengan argumen type = "response" digunakan untuk menghitung peluang selamat (nilai probabilitas antara 0 dan 1) berdasarkan model yang telah dilatih. Kemudian, fungsi ifelse() dipakai untuk menentukan label prediksi akhir, yaitu “Selamat” jika peluang selamat lebih dari 0,5 dan “Meninggal” jika sebaliknya. Terakhir, perintah cat() dan print() digunakan untuk menampilkan hasil prediksi ke layar.

Output

```
***
=== HASIL PREDIKSI DATA BARU ===
  Pclass Sex Age SibSp Parch Fare Peluang_Selamat Prediksi
1     1 female  25     0     0  100     0.9596427  Selamat
2     3  male  30     0     0   10     0.0981539  Meninggal
```

Pembahasan Output : Output yang dihasilkan berupa tabel berisi data penumpang fiktif beserta nilai Peluang_Selamat dan Prediksi. Kolom Peluang_Selamat menunjukkan probabilitas setiap penumpang untuk selamat berdasarkan model regresi logistik. Umumnya, penumpang A (wanita, kelas 1, dan tiket mahal) memiliki peluang selamat yang tinggi sehingga diprediksi “Selamat”, sedangkan penumpang B (pria, kelas 3, dan tiket murah) memiliki peluang selamat yang lebih rendah sehingga diprediksi “Meninggal”. Hasil ini konsisten dengan pola data Titanic, di mana faktor jenis kelamin, kelas penumpang, dan harga tiket sangat memengaruhi peluang keselamatan.

C. PEMBAHASAN TUGAS

KASUS : Analisis Berat Badan Lahir (birthwt)

Dataset: Faktor risiko bayi BBLR (Berat Badan Lahir Rendah)

Target: low (0 = Normal, 1 = BBLR/Low)

1. Load Dataset

```
# 1. Load Dataset
library(MASS) # Library bawaan yang memuat dataset birthwt
library(dplyr)

data(birthwt)

cat("Dataset birthwt (6 baris pertama):\n")
print(head(birthwt))
```

Pembahasan : Kode library(MASS) digunakan untuk memanggil paket MASS yang berisi dataset bawaan birthwt, sedangkan library(dplyr) dipanggil untuk mendukung pengolahan dan manipulasi data jika diperlukan pada tahap selanjutnya. Perintah

`data(birthwt)` berfungsi untuk memuat dataset `birthwt` ke dalam lingkungan kerja R. Selanjutnya, `cat("Dataset birthwt (6 baris pertama):\n")` digunakan untuk menampilkan judul atau keterangan pada console agar output lebih jelas, dan fungsi `head(birthwt)` dipakai untuk menampilkan enam baris pertama dari dataset sebagai gambaran awal struktur dan isi data.

Output

```
***
Attaching package: 'MASS'

The following object is masked from 'package:dplyr':

  select

Dataset birthwt (6 baris pertama):
  low age lwt race smoke ptl ht ui ftv bwt
85   0  19 182   2    0  0  0  1  0 2523
86   0  33 155   3    0  0  0  0  3 2551
87   0  20 105   1    1  0  0  0  1 2557
88   0  21 108   1    1  0  0  1  2 2594
89   0  18 107   1    1  0  0  1  0 2600
91   0  21 124   3    0  0  0  0  0 2622
```

Pembahasan Output : Output yang ditampilkan berupa enam baris pertama dari dataset `birthwt` yang berisi data berat badan bayi saat lahir beserta beberapa variabel penjelas, seperti usia ibu, status merokok, ras, riwayat hipertensi, dan faktor lainnya. Tampilan ini membantu pengguna memahami jenis variabel (numerik atau kategorik), kisaran nilai, serta memastikan bahwa dataset berhasil dimuat dengan benar sebelum dilakukan analisis atau pemodelan lebih lanjut.

2. Pilih Variabel & Data Cleaning

```
# 2. Pilih Variabel & Data Cleaning
# Kita akan gunakan variabel:
# - low    : Target (0 = >2.5kg, 1 = <2.5kg)
# - age    : Usia ibu
# - lwt    : Berat badan ibu sebelum hamil (pounds)
# - smoke  : Status merokok (1 = Ya, 0 = Tidak)
# - ht     : Riwayat hipertensi (1 = Ya, 0 = Tidak)

data_bayi <- birthwt %>%
  dplyr::select(low, age, lwt, smoke, ht) %>%
  mutate(
    # Ubah target jadi Label agar mudah dibaca
    low = factor(low, labels = c("Normal", "BBLR")),
    # Ubah variabel kategori menjadi factor
    smoke = factor(smoke, labels = c("Tidak Merokok", "Perokok")),
    ht = factor(ht, labels = c("Tidak Hipertensi", "Hipertensi"))
  )

cat("\nStruktur Data setelah cleaning:\n")
str(data_bayi)

cat("\nDISTRIBUSI KASUS BBLR:\n")
print(table(data_bayi$low))
```

Pembahasan : Kode ini digunakan untuk memilih variabel yang relevan dan melakukan *data cleaning* pada dataset `birthwt`. Dengan fungsi `select()`, hanya variabel `low` sebagai target serta `age`, `lwt`, `smoke`, dan `ht` sebagai variabel prediktor yang diambil. Selanjutnya, fungsi `mutate()` digunakan untuk mengubah variabel `low` menjadi tipe *factor* dengan label “Normal” dan “BBLR” agar target lebih mudah dipahami, serta mengonversi variabel

kategori smoke dan ht menjadi *factor* dengan label yang lebih deskriptif. Perintah `str(data_bayi)` digunakan untuk menampilkan struktur data setelah proses pembersihan, sedangkan `table(data_bayi$low)` dipakai untuk melihat distribusi jumlah kasus bayi normal dan BBLR.

Output

```
***
Struktur Data setelah cleaning:
'data.frame': 189 obs. of 5 variables:
 $ low : Factor w/ 2 levels "Normal","BBLR": 1 1 1 1 1 1 1 1 1 ...
 $ age : int 19 33 20 21 18 21 22 17 29 26 ...
 $ lwt : int 182 155 105 108 107 124 118 103 123 113 ...
 $ smoke: Factor w/ 2 levels "Tidak Merokok",...: 1 1 2 2 2 1 1 1 2 2 ...
 $ ht : Factor w/ 2 levels "Tidak Hipertensi",...: 1 1 1 1 1 1 1 1 1 1 ...

DISTRIBUSI KASUS BBLR:

Normal  BBLR
130     59
```

Pembahasan Output : Output `str(data_bayi)` menampilkan struktur dataset hasil *cleaning*, termasuk jumlah observasi, tipe data setiap variabel, serta level pada variabel bertipe *factor*. Dari output ini dapat dilihat bahwa variabel target `low` telah menjadi faktor dengan dua kategori, yaitu “Normal” dan “BBLR”, sementara `smoke` dan `ht` juga telah berhasil dikonversi menjadi variabel kategorik. Output `table(data_bayi$low)` menampilkan distribusi jumlah kasus bayi dengan berat lahir normal dan bayi dengan BBLR, sehingga pengguna dapat mengetahui apakah data seimbang atau tidak sebelum dilakukan analisis atau pemodelan lebih lanjut.

3. Buat Model Regresi Logistik

```
# 3. Buat Model Regresi Logistik
# Memprediksi 'low' berdasarkan usia, berat ibu, rokok, dan hipertensi
model_bblr <- glm(low ~ age + lwt + smoke + ht,
                  data = data_bayi,
                  family = binomial(link = "logit"))
```

Pembahasan : Kode ini digunakan untuk membangun model regresi logistik guna memprediksi status berat badan lahir bayi (`low`) berdasarkan beberapa faktor risiko, yaitu usia ibu (`age`), berat badan ibu sebelum hamil (`lwt`), status merokok (`smoke`), dan riwayat hipertensi (`ht`). Fungsi `glm()` digunakan dengan argumen `family = binomial(link = "logit")` karena variabel target bersifat biner (Normal vs BBLR). Model ini akan mengestimasi hubungan antara variabel-variabel tersebut dengan peluang terjadinya BBLR menggunakan fungsi `logit` sebagai penghubung.

Output

-

Pembahasan Output :

4. Ringkasan Model

```
# 4. Ringkasan Model
cat("\n=== RINGKASAN MODEL ===\n")
summary(model_bblr)

# Hitung Odds Ratio
cat("\n=== ODDS RATIO (exp(beta)) ===\n")
odds_ratio <- exp(coef(model_bblr))

# Tampilkan dalam bentuk tabel rapi
hasil_or <- data.frame(
  Variabel = names(odds_ratio),
```



```

Odds_Ratio = round(odds_ratio, 3),
Interpretasi = c(
  "Intercept",
  "Setiap kenaikan umur ibu, risiko BBLR berubah (lihat tanda +/-)",
  "Semakin berat ibu, risiko BBLR cenderung turun (OR < 1)",
  "Ibu perokok memiliki risiko BBLR jauh lebih tinggi (OR > 1)",
  "Ibu hipertensi memiliki risiko BBLR sangat tinggi (OR > 1)"
)
)
print(hasil_or)

```

Pembahasan : Kode `cat("\n=== RINGKASAN MODEL ===\n")` digunakan untuk menampilkan judul agar output ringkasan model mudah dikenali. Fungsi `summary(model_bblr)` menampilkan ringkasan lengkap hasil regresi logistik, termasuk koefisien, signifikansi, dan ukuran kecocokan model. Selanjutnya, `exp(coef(model_bblr))` digunakan untuk menghitung *odds ratio* dengan mengubah koefisien logit menjadi bentuk eksponensial agar lebih mudah diinterpretasikan. Hasil *odds ratio* kemudian disusun ke dalam sebuah *data frame* bernama `hasil_or` yang berisi nama variabel, nilai *odds ratio* yang telah dibulatkan, serta interpretasi singkat untuk setiap variabel, dan ditampilkan ke layar menggunakan `print()`.

Output

```

*** === RINGKASAN MODEL ===

Call:
glm(formula = low ~ age + lwt + smoke + ht, family = binomial(link = "logit"),
    data = data_bayi)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.766856   1.052266   1.679  0.09313 .
age          -0.035687   0.033365  -1.070  0.28480
lwt          -0.016955   0.006623  -2.560  0.01047 *
smokePerokok  0.679020   0.331939   2.046  0.04079 *
htHipertensi  1.788156   0.685882   2.607  0.00913 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 234.67  on 188  degrees of freedom
Residual deviance: 215.68  on 184  degrees of freedom
AIC: 225.68

Number of Fisher Scoring iterations: 4

=== ODDS RATIO (exp(beta)) ===
      Variabel Odds_Ratio
(Intercept) (Intercept)  5.852
age          age         0.965
lwt          lwt         0.983
smokePerokok smokePerokok 1.972
htHipertensi htHipertensi 5.978

                                Interpretasi
(Intercept) Intercept
age          Setiap kenaikan umur ibu, risiko BBLR berubah (lihat tanda +/-)
lwt          Semakin berat ibu, risiko BBLR cenderung turun (OR < 1)
smokePerokok Ibu perokok memiliki risiko BBLR jauh lebih tinggi (OR > 1)
htHipertensi Ibu hipertensi memiliki risiko BBLR sangat tinggi (OR > 1)

```

Pembahasan Output : Output pertama berupa ringkasan model regresi logistik yang menampilkan koefisien masing-masing variabel (*Estimate*), *Std. Error*, nilai *z*, dan *p-value*. Variabel dengan *p-value* kecil (misalnya $< 0,05$) dianggap berpengaruh signifikan terhadap risiko BBLR. Output kedua berupa tabel *odds ratio* yang menunjukkan besarnya perubahan risiko BBLR akibat setiap variabel. *Odds ratio* kurang dari 1, seperti pada berat badan ibu (*lwt*), menunjukkan bahwa semakin berat ibu sebelum hamil maka risiko BBLR cenderung menurun. Sebaliknya, *odds ratio* lebih dari 1 pada variabel *smoke* dan *ht* menandakan bahwa ibu perokok dan ibu dengan riwayat hipertensi memiliki risiko

melahirkan bayi BBLR yang jauh lebih tinggi dibandingkan ibu yang tidak merokok dan tidak hipertensi.

5. Prediksi Data Baru

```
# 5. Prediksi Data Baru
# Skenario: Membandingkan 2 Ibu dengan berat dan usia sama (25 thn, 120 lbs),
# tapi beda status merokok.

ibu_baru <- data.frame(
  age = c(25, 25),
  lwt = c(120, 120),
  smoke = factor(c("Tidak Merokok", "Perokok"), levels = c("Tidak Merokok",
"Perokok")),
  ht = factor(c("Tidak Hipertensi", "Tidak Hipertensi"), levels = c("Tidak
Hipertensi", "Hipertensi"))
)

# Prediksi Probabilitas
ibu_baru$Probabilitas_BBLR <- predict(model_bblr, newdata = ibu_baru, type =
"response")

# Klasifikasi (Threshold 0.3 karena kasus medis sering pakai threshold
rendah,
# tapi disini kita pakai standar 0.5)
ibu_baru$Prediksi_Status <- ifelse(ibu_baru$Probabilitas_BBLR > 0.5, "BBLR",
"Normal")

cat("\n=== PREDIKSI DATA BARU ===\n")
print(ibu_baru)
```

Pembahasan : Kode ini digunakan untuk melakukan prediksi risiko bayi lahir dengan berat badan rendah (BBLR) pada data ibu baru menggunakan model regresi logistik model_bblr. Pertama, dibuat *data frame* ibu_baru yang berisi dua ibu dengan usia dan berat badan sebelum hamil yang sama, tetapi berbeda pada status merokok, sementara status hipertensi dibuat sama. Penulisan factor() dengan argumen levels memastikan kategori variabel sesuai dengan level yang digunakan pada saat pelatihan model. Selanjutnya, fungsi predict() dengan type = "response" digunakan untuk menghitung probabilitas BBLR untuk masing-masing ibu. Setelah itu, fungsi ifelse() dipakai untuk mengklasifikasikan hasil prediksi menjadi “BBLR” atau “Normal” berdasarkan nilai ambang (threshold) 0,5. Terakhir, perintah cat() dan print() digunakan untuk menampilkan hasil prediksi ke layar.

Output

```
***
=== PREDIKSI DATA BARU ===
   age lwt   smoke      ht Probabilitas_BBLR Prediksi_Status
1  25 120 Tidak Merokok Tidak Hipertensi    0.2386785      Normal
2  25 120   Perokok Tidak Hipertensi    0.3820354      Normal
```

Pembahasan Output : Output yang dihasilkan berupa tabel berisi karakteristik dua ibu beserta kolom Probabilitas_BBLR dan Prediksi_Status. Nilai Probabilitas_BBLR menunjukkan peluang terjadinya BBLR untuk setiap skenario ibu. Umumnya, ibu dengan status “Perokok” memiliki probabilitas BBLR yang lebih tinggi dibandingkan ibu “Tidak Merokok”, meskipun usia dan berat badan sama. Berdasarkan threshold 0,5, ibu tidak merokok cenderung diprediksi “Normal”, sedangkan ibu perokok memiliki risiko lebih tinggi dan dapat diprediksi “BBLR”. Hasil ini menegaskan bahwa kebiasaan merokok

merupakan faktor penting yang meningkatkan risiko bayi lahir dengan berat badan rendah.

D. KESIMPULAN

Berdasarkan praktikum analisis regresi logistik biner yang telah saya lakukan menggunakan dataset *birthwt*, diperoleh hasil bahwa model regresi logistik mampu digunakan untuk memprediksi risiko bayi lahir dengan berat badan rendah (BBLR) berdasarkan faktor usia ibu, berat badan ibu sebelum hamil, status merokok, dan riwayat hipertensi. Dari ringkasan model dan nilai *odds ratio*, dapat disimpulkan bahwa berat badan ibu cenderung menurunkan risiko BBLR, sedangkan kebiasaan merokok dan riwayat hipertensi secara signifikan meningkatkan risiko terjadinya BBLR. Hasil prediksi pada data baru juga menunjukkan bahwa ibu perokok memiliki probabilitas BBLR yang lebih tinggi dibandingkan ibu yang tidak merokok meskipun usia dan berat badannya sama. Dengan demikian, regresi logistik biner efektif digunakan untuk menganalisis dan memprediksi kejadian BBLR serta membantu mengidentifikasi faktor-faktor risiko yang berpengaruh.