

# Projek SPL

Dinda Dwi Anugrah Pertiwi

## Contents

<b>PROJEK SUPERVISED LEARNING</b>	<b>1</b>
INPUT DATA . . . . .	1
REGRESI LOGISTIK . . . . .	2
Jawaban . . . . .	2

## PROJEK SUPERVISED LEARNING

Universitas ingin mengetahui faktor-faktor yang memengaruhi kelulusan tepat waktu mahasiswa (1=lulus tepat waktu).

Masukkan data ke R (`data2.txt`), fit modelnya, buat assesement model kemudian jawab pertanyaan berikut:

1. Apakah GPA berpengaruh signifikan terhadap peluang kelulusan?
2. Bagaimana tanda koefisien part\_time\_job diinterpretasikan?
3. Apa arti hasil uji Hosmer-Lemeshow yang signifikan/tidak signifikan?
4. Jika GPA naik 1 poin, berapa kali lipat peluang kelulusan meningkat?

## INPUT DATA

```
#INPUT DATA
data_mahasiswa <- read.table("C:/Users/dals/OneDrive - untirta.ac.id/KULIAH/Sem 5/SPL/DATA/data2.txt", 
head(data_mahasiswa, 10)

##          GPA attendance part_time_job graduate
## 1  3.548383    98.34913           1        0
## 2  2.774121    76.30728           1        1
## 3  3.145251    85.55487           1        1
## 4  3.253145    85.49067           0        1
## 5  3.161707    79.21644           0        1
## 6  2.957550    75.01261           1        0
## 7  3.604609    84.97567           0        1
## 8  2.962136    91.55512           1        0
## 9  3.807369    99.76842           0        1
## 10 2.974914    65.90847           0        1
```

```

str(data_mahasiswa)

## 'data.frame':   400 obs. of  4 variables:
## $ GPA        : num  3.55 2.77 3.15 3.25 3.16 ...
## $ attendance  : num  98.3 76.3 85.6 85.5 79.2 ...
## $ part_time_job: int  1 1 1 0 0 1 0 1 0 0 ...
## $ graduate    : int  0 1 1 1 1 0 1 0 1 1 ...

```

## REGRESI LOGISTIK

```

#PEMODELAN
model_logit <- glm(graduate ~ GPA + attendance + part_time_job,
                     data = data_mahasiswa,
                     family = binomial(link = "logit"))

summary(model_logit)

##
## Call:
## glm(formula = graduate ~ GPA + attendance + part_time_job, family = binomial(link = "logit"),
##      data = data_mahasiswa)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -9.93480  1.80086 -5.517 3.45e-08 ***
## GPA          3.16494  0.44524  7.108 1.17e-12 ***
## attendance   0.02898  0.01399  2.071  0.03837 *
## part_time_job -0.73463  0.27716 -2.651  0.00803 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
## Null deviance: 416.41 on 399 degrees of freedom
## Residual deviance: 340.14 on 396 degrees of freedom
## AIC: 348.14
##
## Number of Fisher Scoring iterations: 5

```

## Jawaban

Apakah GPA berpengaruh signifikan terhadap peluang kelulusan?

```

summary(model_logit)$coefficients["GPA", ]

##     Estimate Std. Error      z value      Pr(>|z|)
## 3.164937e+00 4.452374e-01 7.108425e+00 1.173744e-12

```

Interpretasi hasil:

- Dari hasil pemodelan dengan regresi logistik didapatkan nilai p-value sebesar 1.173744e-12 yang dimana nilai tersebut  $< 0.05$  yang berarti GPA berpengaruh signifikan terhadap peluang kelulusan tepat waktu.
- Koefisien bertanda positif menunjukkan bahwa semakin tinggi GPA, peluang mahasiswa untuk lulus tepat waktu meningkat.

Bagaimana tanda koefisien part\_time\_job diinterpretasikan?

```
summary(model_logit)$coefficients["part_time_job", ]  
  
##      Estimate   Std. Error     z value    Pr(>|z|)  
## -0.734634759  0.277155138 -2.650626521  0.008034263
```

Interperetasi hasil:

- Dari hasil pemodelan dengan regresi logistik didapatkan nilai p-value sebesar 0.008034263 yang dimana nilai tersebut  $< 0.05$  yang berarti part\_time\_job berpengaruh signifikan terhadap peluang kelulusan tepat waktu.
- Koefisien bertanda negatif (-0.734634759) menunjukkan bahwa mahasiswa yang memiliki part time job (pekerjaan paruh waktu) memiliki peluang lebih rendah untuk lulus tepat waktu dibandingkan mahasiswa yang tidak bekerja.

Apa arti hasil uji Hosmer-Lemeshow yang signifikan/tidak signifikan?

```
#Hosmer-Lemeshow  
library(ResourceSelection)  
  
## Warning: package 'ResourceSelection' was built under R version 4.3.3  
  
## ResourceSelection 0.3-6 2023-06-27  
  
hoslem.test(data_mahasiswa$graduate, fitted(model_logit))  
  
##  
## Hosmer and Lemeshow goodness of fit (GOF) test  
##  
## data: data_mahasiswa$graduate, fitted(model_logit)  
## X-squared = 14.771, df = 8, p-value = 0.06375
```

Interpretasi hasil:

- Dari hasil uji Hosmer-Lemeshow diperoleh nilai p-value sebesar 0.06375 yang dimana nilai tersebut  $> 0.05$  yang berarti tidak terdapat perbedaan yang signifikan antara nilai yang diprediksi oleh model dan nilai observasi pada data (model regresi logistik memiliki kecocokan yang baik dan sesuai dengan data).

Jika GPA naik 1 poin, berapa kali lipat peluang kelulusan meningkat?

```
#Menghitung odds ratio  
exp(coef(model_logit)[ "GPA" ])
```

```
##      GPA  
## 23.68725
```

Interpretasi hasil:

- Dari hasil odds ratio untuk GPA didapatkan nilai sebesar 23.68725, yang berarti jika GPA naik 1 poin maka peluang mahasiswa untuk lulus tepat waktu meningkat 23.68725 kali dengan asumsi attendance dan part time job tetap.