

# Final Project

123210098\_123210114

2023-12-04

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   :  2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean   : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.   :120.00
```

## Including Plots

You can also embed plots, for example:



```
#Library Yang digunakan#  
library(dplyr)
```

```
##  
## 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
```

```
library(tidyr)  
library(stringr)  
library(writexl)
```

```
## Warning: package 'writexl' was built under R version 4.3.2
```

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 4.3.2
```

```
## Loading required package: ggplot2
```

```
## Loading required package: lattice
```

```
library(here)
```

```
## Warning: package 'here' was built under R version 4.3.2
```

```
## here() starts at C:/Users/Lenovo/Downloads/Projek DS
```

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.3.2
```

```
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 4.3.2
```

```
##
```

```
## Attaching package: 'reshape2'
```

```
## The following object is masked from 'package:tidyr':
```

```
##
```

```
##      smiths
```

```
library(ggplot2)
```

```
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 4.3.2
```

```
##
```

```
## Attaching package: 'MASS'
```

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

```
##
```

```
##      select
```

```
library(scales)
```

```
# Membaca dataset
```

```
path <- "C:/Users/Lenovo/Downloads/Projek DS/JumlahPendudukBerdasarkanPendidikanPerKabBersihFinal.xlsx"
```

```
pendidikan <- read_excel(path) # Menyesuaikan skip jika diperlukan
```

```
#Menampilkan Struktur
```

```
str(pendidikan)
```

```
## tibble [700 x 5] (S3: tbl_df/tbl/data.frame)
## $ No : num [1:700] 1 2 3 4 5 6 7 8 9 10 ...
## $ Kabupaten/Kota : chr [1:700] "CILACAP" "CILACAP" "CILACAP" "CILACAP" ...
## $ Gender : chr [1:700] "Pria" "Pria" "Pria" "Pria" ...
## $ Tingkat Pendidikan: chr [1:700] "Tidak/Belum Sekolah" "Belum Tamat SD/Sederajat" "Tamat SD/Sedera
## $ Jumlah : num [1:700] 194237 83718 326975 158007 140179 ...
```

#Menampilkan dataset

```
print(pendidikan)
```

```
## # A tibble: 700 x 5
##       No 'Kabupaten/Kota' Gender 'Tingkat Pendidikan' Jumlah
##   <dbl> <chr>          <chr> <chr>          <dbl>
## 1     1 CILACAP      Pria  Tidak/Belum Sekolah 194237
## 2     2 CILACAP      Pria  Belum Tamat SD/Sederajat 83718
## 3     3 CILACAP      Pria  Tamat SD/Sederajat 326975
## 4     4 CILACAP      Pria  SLTP/Sederajat 158007
## 5     5 CILACAP      Pria  SLTA/Sederajat 140179
## 6     6 CILACAP      Pria  Diploma I/II 2083
## 7     7 CILACAP      Pria  Akademi/Diploma III/Sarjana Muda 6921
## 8     8 CILACAP      Pria  Diploma IV/Strata I 18833
## 9     9 CILACAP      Pria  Strata II 924
## 10    10 CILACAP      Pria  Strata III 35
## # i 690 more rows
```

#Menampilkan kolom dataset

```
names(pendidikan)
```

```
## [1] "No" "Kabupaten/Kota" "Gender"
## [4] "Tingkat Pendidikan" "Jumlah"
```

#Mengubah jenis kelamin menjadi factor

```
pendidikan$Gender <- as.factor(pendidikan$Gender)
pendidikan$Gender
```

```
## [1] Pria Pria Pria Pria Pria Pria Pria Pria Pria Pria
## [11] Pria Pria Pria Pria Pria Pria Pria Pria Pria Pria
## [21] Pria Pria Pria Pria Pria Pria Pria Pria Pria Pria
## [31] Pria Pria Pria Pria Pria Pria Pria Pria Pria Pria
## [41] Pria Pria Pria Pria Pria Pria Pria Pria Pria Pria
## [51] Pria Pria Pria Pria Pria Pria Pria Pria Pria Pria
## [61] Pria Pria Pria Pria Pria Pria Pria Pria Pria Pria
## [71] Pria Pria Pria Pria Pria Pria Pria Pria Pria Pria
## [81] Pria Pria Pria Pria Pria Pria Pria Pria Pria Pria
## [91] Pria Pria Pria Pria Pria Pria Pria Pria Pria Pria
## [101] Pria Pria Pria Pria Pria Pria Pria Pria Pria Pria
## [111] Pria Pria Pria Pria Pria Pria Pria Pria Pria Pria
## [121] Pria Pria Pria Pria Pria Pria Pria Pria Pria Pria
```

[illegible]

```
## [671] Wanita Wanita Wanita Wanita Wanita Wanita Wanita Wanita Wanita Wanita
## [681] Wanita Wanita Wanita Wanita Wanita Wanita Wanita Wanita Wanita Wanita
## [691] Wanita Wanita Wanita Wanita Wanita Wanita Wanita Wanita Wanita Wanita
## Levels: Pria Wanita
```

#Visualisasi

```
# Agregasi data untuk mendapatkan total jumlah untuk setiap TingkatPendidikan dan JenisKelamin
data_aggregated <- pendidikan %>%
  group_by(`Tingkat Pendidikan`, Gender) %>%
  summarize(Jumlah = sum(Jumlah, na.rm = TRUE))
```

```
## 'summarise()' has grouped output by 'Tingkat Pendidikan'. You can override
## using the '.groups' argument.
```

data\_aggregated

```
## # A tibble: 20 x 3
## # Groups:   Tingkat Pendidikan [10]
##   `Tingkat Pendidikan`      Gender  Jumlah
##   <chr>                  <fct>    <dbl>
## 1 Akademi/Diploma III/Sarjana Muda Pria    172498
## 2 Akademi/Diploma III/Sarjana Muda Wanita  214583
## 3 Belum Tamat SD/Sederajat    Pria    1944310
## 4 Belum Tamat SD/Sederajat    Wanita  1991685
## 5 Diploma I/II                Pria     54195
## 6 Diploma I/II                Wanita    72556
## 7 Diploma IV/Strata I         Pria    500193
## 8 Diploma IV/Strata I         Wanita   483691
## 9 SLTA/Sederajat              Pria    3066421
## 10 SLTA/Sederajat              Wanita   2543823
## 11 SLTP/Sederajat              Pria    3013293
## 12 SLTP/Sederajat              Wanita   2823242
## 13 Strata II                   Pria     37406
## 14 Strata II                   Wanita    21628
## 15 Strata III                  Pria      2116
## 16 Strata III                  Wanita    1279
## 17 Tamat SD/Sederajat          Pria   5665549
## 18 Tamat SD/Sederajat          Wanita  5950109
## 19 Tidak/Belum Sekolah        Pria   3594777
## 20 Tidak/Belum Sekolah        Wanita  3607705
```

```
# Visualisasi dengan Diagram Batang skala sumbu y disesuaikan
ggplot(data_aggregated, aes(x = `Tingkat Pendidikan`, y = Jumlah, fill = Gender)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  scale_y_continuous(labels = scales::comma) + # Menggunakan label dengan pemisah ribuan
  theme_minimal() +
  labs(title = "Distribusi Tingkat Pendidikan berdasarkan Jenis Kelamin",
       x = "Tingkat Pendidikan",
       y = "Jumlah (dalam ribuan)")
```

[illegible]

```
# Mengubah TingkatPendidikan menjadi faktor dengan urutan yang sesuai
data_aggregated$`Tingkat Pendidikan` <- factor(data_aggregated$`Tingkat Pendidikan`)
data_aggregated$Gender <- factor(data_aggregated$Gender)

# Membagi data menjadi set pelatihan dan pengujian
set.seed(123)
training_indices <- sample(1:nrow(data_aggregated), 0.8 * nrow(data_aggregated))
train_data <- data_aggregated[training_indices, ]
test_data <- data_aggregated[-training_indices, ]

# Membangun model dengan bobot
model <- polr(`Tingkat Pendidikan` ~ Gender, data = train_data, weights = train_data$Jumlah, Hess = TRUE)
```

```
# Memprediksi pada data pengujian
prediksi <- predict(model, newdata = test_data)

# Membuat confusion matrix
conf_matrix <- table(Predicted = prediksi, Actual = test_data$`Tingkat Pendidikan`)

# Menampilkan confusion matrix
print(conf_matrix)
```

##	Actual		
## Predicted	Akademi/Diploma III/Sarjana Muda		
## Akademi/Diploma III/Sarjana Muda		0	
## Belum Tamat SD/Sederajat		0	
## Diploma I/II		0	
## Diploma IV/Strata I		0	
## SLTA/Sederajat		0	
## SLTP/Sederajat		0	
## Strata II		0	
## Strata III		0	
## Tamat SD/Sederajat		1	
## Tidak/Belum Sekolah		0	
##	Actual		
## Predicted	Belum Tamat SD/Sederajat Diploma I/II		
## Akademi/Diploma III/Sarjana Muda	0	0	
## Belum Tamat SD/Sederajat	0	0	
## Diploma I/II	0	0	
## Diploma IV/Strata I	0	0	
## SLTA/Sederajat	0	0	
## SLTP/Sederajat	0	0	
## Strata II	0	0	
## Strata III	0	0	
## Tamat SD/Sederajat	0	0	
## Tidak/Belum Sekolah	0	0	
##	Actual		
## Predicted	Diploma IV/Strata I SLTA/Sederajat		
## Akademi/Diploma III/Sarjana Muda	0	0	
## Belum Tamat SD/Sederajat	0	0	
## Diploma I/II	0	0	
## Diploma IV/Strata I	0	0	
## SLTA/Sederajat	0	0	
## SLTP/Sederajat	0	0	
## Strata II	0	0	
## Strata III	0	0	
## Tamat SD/Sederajat	1	0	
## Tidak/Belum Sekolah	0	0	
##	Actual		
## Predicted	SLTP/Sederajat Strata II Strata III		
## Akademi/Diploma III/Sarjana Muda	0	0	0
## Belum Tamat SD/Sederajat	0	0	0
## Diploma I/II	0	0	0
## Diploma IV/Strata I	0	0	0
## SLTA/Sederajat	0	0	0
## SLTP/Sederajat	0	0	0
## Strata II	0	0	0
## Strata III	0	0	0
## Tamat SD/Sederajat	0	1	1
## Tidak/Belum Sekolah	0	0	0
##	Actual		
## Predicted	Tamat SD/Sederajat Tidak/Belum Sekolah		
## Akademi/Diploma III/Sarjana Muda	0	0	
## Belum Tamat SD/Sederajat	0	0	
## Diploma I/II	0	0	
## Diploma IV/Strata I	0	0	



##	SLTA/Sederajat	0	0
##	SLTP/Sederajat	0	0
##	Strata II	0	0
##	Strata III	0	0
##	Tamat SD/Sederajat	0	0
##	Tidak/Belum Sekolah	0	0

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.