Memuat paket-paket yang diperlukan

library(tidyverse)

## Warning: package 'ggplot2' was built under R version 4.3.3

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.3 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(VGAM)

## Loading required package: stats4  
## Loading required package: splines

library(nnet)  
library(caret)

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

## Loading required package: lattice  
##   
## Attaching package: 'caret'  
##   
## The following object is masked from 'package:VGAM':  
##   
## predictors  
##   
## The following object is masked from 'package:purrr':  
##   
## lift

library(readr)  
library(generalhoslem)

## Warning: package 'generalhoslem' was built under R version 4.3.3

## Loading required package: reshape  
##   
## Attaching package: 'reshape'  
##   
## The following object is masked from 'package:lubridate':  
##   
## stamp  
##   
## The following object is masked from 'package:dplyr':  
##   
## rename  
##   
## The following objects are masked from 'package:tidyr':  
##   
## expand, smiths  
##   
## Loading required package: MASS  
##   
## Attaching package: 'MASS'  
##   
## The following object is masked from 'package:dplyr':  
##   
## select

library(pscl)

## Warning: package 'pscl' was built under R version 4.3.3

## Classes and Methods for R originally developed in the  
## Political Science Computational Laboratory  
## Department of Political Science  
## Stanford University (2002-2015),  
## by and under the direction of Simon Jackman.  
## hurdle and zeroinfl functions by Achim Zeileis.

library(ROSE)

## Warning: package 'ROSE' was built under R version 4.3.2

## Loaded ROSE 0.0-4

Data

pkl <- read.csv("data\_pkl.csv")  
data <- subset(pkl, Status\_Akhir %in% c("SEMBUH", "MENINGGAL"))  
data = subset(data, select = -c(Tahun))

data <- data %>%  
 mutate\_all(as.factor)  
str(data)

## 'data.frame': 580 obs. of 9 variables:  
## $ Jenis\_Kelamin : Factor w/ 2 levels "L","P": 2 2 1 2 2 1 1 2 2 1 ...  
## $ Status\_Covid.19: Factor w/ 3 levels "KONFIRMASI","PROBABLE",..: 1 1 1 1 3 3 1 1 3 3 ...  
## $ Perawatan : Factor w/ 3 levels "IGD","RALAN",..: 3 2 3 3 2 2 3 3 2 2 ...  
## $ Sesak\_Nafas : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Gangguan\_Makan : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ DM : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ HT : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Umur : Factor w/ 2 levels "<=50",">50": 2 2 2 2 1 1 2 1 1 2 ...  
## $ Status\_Akhir : Factor w/ 2 levels "MENINGGAL","SEMBUH": 2 2 2 2 2 2 2 2 2 2 ...

Summary

summary(data)

## Jenis\_Kelamin Status\_Covid.19 Perawatan Sesak\_Nafas Gangguan\_Makan DM   
## L:285 KONFIRMASI:382 IGD : 99 0:479 0:569 0:577   
## P:295 PROBABLE : 7 RALAN:102 1:101 1: 11 1: 3   
## SUSPECT :191 RANAP:379   
## HT Umur Status\_Akhir  
## 0:571 <=50:290 MENINGGAL: 28   
## 1: 9 >50 :290 SEMBUH :552   
##

data$Sesak\_Nafas <- relevel(data$Sesak\_Nafas, ref="0")  
data$Gangguan\_Makan <- relevel(data$Gangguan\_Makan, ref="0")  
data$DM <- relevel(data$DM, ref="0")  
data$HT <- relevel(data$HT, ref="0")  
data$Umur <- factor(data$Umur, levels = c("<=50", ">50"))  
data$Status\_Covid.19 <- factor(data$Status\_Covid.19, levels = c("SUSPECT", "PROBABLE", "KONFIRMASI"))  
data$Perawatan <- factor(data$Perawatan, levels = c("RALAN", "RANAP", "IGD"))  
data$Status\_Akhir <- relevel(data$Status\_Akhir, ref="SEMBUH")

unique(data$Jenis\_Kelamin)

## [1] P L  
## Levels: L P

unique(data$Status\_Covid.19)

## [1] KONFIRMASI SUSPECT PROBABLE   
## Levels: SUSPECT PROBABLE KONFIRMASI

unique(data$Perawatan)

## [1] RANAP RALAN IGD   
## Levels: RALAN RANAP IGD

unique(data$Sesak\_Nafas)

## [1] 0 1  
## Levels: 0 1

unique(data$Gangguan\_Makan)

## [1] 0 1  
## Levels: 0 1

unique(data$DM)

## [1] 0 1  
## Levels: 0 1

unique(data$HT)

## [1] 0 1  
## Levels: 0 1

unique(data$Umur)

## [1] >50 <=50  
## Levels: <=50 >50

unique(data$Status\_Akhir)

## [1] SEMBUH MENINGGAL  
## Levels: SEMBUH MENINGGAL

Spliting Data

set.seed(2024)  
index <- sample(nrow(data), 0.7\*nrow(data))  
train <- data[index,]  
valid <- data[-index,]

table(valid$Status\_Akhir)

##   
## SEMBUH MENINGGAL   
## 163 11

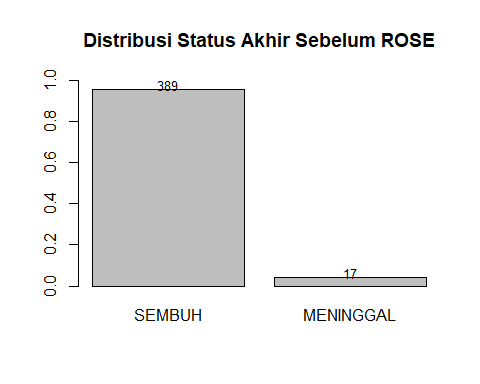
Handling Imbalance Data Using ROSE

table(train$Status\_Akhir)

##   
## SEMBUH MENINGGAL   
## 389 17

Data sebelum dilakukan ROSE

# Membuat barplot  
bar <- barplot(prop.table(table(train$Status\_Akhir)),   
 main = "Distribusi Status Akhir Sebelum ROSE",   
 ylim = c(0, max(prop.table(table(train$Status\_Akhir))) + 0.05))  
  
# Menghitung frekuensi absolut  
counts <- table(train$Status\_Akhir)  
  
# Menambahkan teks (jumlah) di atas tengah bar  
text(bar, prop.table(counts) + 0.02, labels = counts, cex = 0.8)



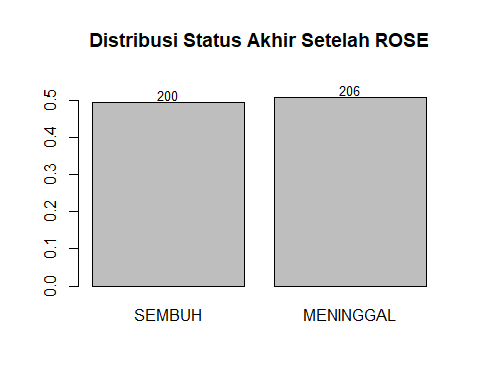
train\_rose <- ROSE(Status\_Akhir ~ ., data = train, seed = 2024)$data

table(train\_rose$Status\_Akhir)

##   
## SEMBUH MENINGGAL   
## 200 206

Data setelah dilakukan ROSE

# Membuat barplot  
bar <- barplot(prop.table(table(train\_rose$Status\_Akhir)),   
 main = "Distribusi Status Akhir Setelah ROSE",   
 ylim = c(0, max(prop.table(table(train\_rose$Status\_Akhir))) + 0.05))  
  
# Menghitung frekuensi absolut  
counts <- table(train\_rose$Status\_Akhir)  
  
# Menambahkan teks (jumlah) di atas tengah bar  
text(bar, prop.table(counts) + 0.02, labels = counts, cex = 0.8)



str(train\_rose)

## 'data.frame': 406 obs. of 9 variables:  
## $ Jenis\_Kelamin : Factor w/ 2 levels "L","P": 2 2 2 2 2 1 2 1 2 1 ...  
## $ Status\_Covid.19: Factor w/ 3 levels "SUSPECT","PROBABLE",..: 3 3 3 3 3 1 3 3 1 1 ...  
## $ Perawatan : Factor w/ 3 levels "RALAN","RANAP",..: 3 1 3 2 2 1 1 2 1 1 ...  
## $ Sesak\_Nafas : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Gangguan\_Makan : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ DM : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ HT : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Umur : Factor w/ 2 levels "<=50",">50": 1 1 1 1 1 2 1 1 1 1 ...  
## $ Status\_Akhir : Factor w/ 2 levels "SEMBUH","MENINGGAL": 1 1 1 1 1 1 1 1 1 1 ...

Uji independensi

1. Variabel Jenis Kelamin

tabel\_kontingensi1 <- table(train\_rose$Jenis\_Kelamin, train\_rose$Status\_Akhir)  
hasil\_chisq1 <- chisq.test(tabel\_kontingensi1)  
hasil\_chisq1

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: tabel\_kontingensi1  
## X-squared = 0.74697, df = 1, p-value = 0.3874

1. Variabel Umur

tabel\_kontingensi2 <- table(train\_rose$Umur, train\_rose$Status\_Akhir)  
hasil\_chisq2 <- chisq.test(tabel\_kontingensi2)  
hasil\_chisq2

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: tabel\_kontingensi2  
## X-squared = 138.22, df = 1, p-value < 2.2e-16

1. Variabel Status.Covid.19

tabel\_kontingensi3 <- table(train\_rose$Status\_Covid.19, train\_rose$Status\_Akhir)  
hasil\_chisq3 <- chisq.test(tabel\_kontingensi3)  
hasil\_chisq3

##   
## Pearson's Chi-squared test  
##   
## data: tabel\_kontingensi3  
## X-squared = 28.779, df = 2, p-value = 5.632e-07

1. Variabel Perawatan

tabel\_kontingensi4 <- table(train\_rose$Perawatan, train\_rose$Status\_Akhir)  
hasil\_chisq4 <- chisq.test(tabel\_kontingensi4)  
hasil\_chisq4

##   
## Pearson's Chi-squared test  
##   
## data: tabel\_kontingensi4  
## X-squared = 72.135, df = 2, p-value < 2.2e-16

1. Variabel Sesak\_Nafas

tabel\_kontingensi5 <- table(train\_rose$Sesak\_Nafas, train\_rose$Status\_Akhir)  
hasil\_chisq5 <- chisq.test(tabel\_kontingensi5)  
hasil\_chisq5

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: tabel\_kontingensi5  
## X-squared = 9.9085, df = 1, p-value = 0.001645

1. Variabel Gangguan Makan

tabel\_kontingensi6 <- table(train\_rose$Gangguan\_Makan, train\_rose$Status\_Akhir)  
hasil\_chisq6 <- chisq.test(tabel\_kontingensi6)  
hasil\_chisq6

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: tabel\_kontingensi6  
## X-squared = 26.498, df = 1, p-value = 2.638e-07

1. Variabel DM

tabel\_kontingensi7 <- table(train\_rose$DM, train\_rose$Status\_Akhir)  
hasil\_chisq7 <- chisq.test(tabel\_kontingensi7)

## Warning in chisq.test(tabel\_kontingensi7): Chi-squared approximation may be  
## incorrect

hasil\_chisq7

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: tabel\_kontingensi7  
## X-squared = 7.0341, df = 1, p-value = 0.007997

1. Variabel HT

tabel\_kontingensi8 <- table(train\_rose$HT, train\_rose$Status\_Akhir)  
hasil\_chisq8 <- chisq.test(tabel\_kontingensi8)  
hasil\_chisq8

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: tabel\_kontingensi8  
## X-squared = 32.761, df = 1, p-value = 1.042e-08

Hasil uji independensi menunjukan bahwa terdapat varibel jenis kelamin tidak berpengaruh signifikan terhadap status akhir pasien covid-19, sehingga variabel tersebut tidak dimasukan dalam pembentukan model

train\_baru = subset(train\_rose, select = -(Jenis\_Kelamin))  
str(train\_baru)

## 'data.frame': 406 obs. of 8 variables:  
## $ Status\_Covid.19: Factor w/ 3 levels "SUSPECT","PROBABLE",..: 3 3 3 3 3 1 3 3 1 1 ...  
## $ Perawatan : Factor w/ 3 levels "RALAN","RANAP",..: 3 1 3 2 2 1 1 2 1 1 ...  
## $ Sesak\_Nafas : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Gangguan\_Makan : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ DM : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ HT : Factor w/ 2 levels "0","1": 1 1 1 1 1 1 1 1 1 1 ...  
## $ Umur : Factor w/ 2 levels "<=50",">50": 1 1 1 1 1 2 1 1 1 1 ...  
## $ Status\_Akhir : Factor w/ 2 levels "SEMBUH","MENINGGAL": 1 1 1 1 1 1 1 1 1 1 ...

Regresi Logistik (Model Semua Variabel)

mlog <- glm(Status\_Akhir~., data = train\_baru, family = binomial)  
summary(mlog)

##   
## Call:  
## glm(formula = Status\_Akhir ~ ., family = binomial, data = train\_baru)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -6.3614 1.1189 -5.685 1.31e-08 \*\*\*  
## Status\_Covid.19PROBABLE 38.9102 3193.4648 0.012 0.990279   
## Status\_Covid.19KONFIRMASI 1.0826 0.4389 2.467 0.013640 \*   
## PerawatanRANAP 3.9445 1.0458 3.772 0.000162 \*\*\*  
## PerawatanIGD -15.6627 1391.9735 -0.011 0.991022   
## Sesak\_Nafas1 -0.9818 0.4808 -2.042 0.041137 \*   
## Gangguan\_Makan1 6.4838 1.5279 4.244 2.20e-05 \*\*\*  
## DM1 -0.6685 4173.6931 0.000 0.999872   
## HT1 19.8707 2137.7203 0.009 0.992584   
## Umur>50 2.6800 0.4001 6.698 2.11e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 562.75 on 405 degrees of freedom  
## Residual deviance: 245.71 on 396 degrees of freedom  
## AIC: 265.71  
##   
## Number of Fisher Scoring iterations: 18

Uji Signifikan Serentak

pR2(mlog)

## fitting null model for pseudo-r2

## llh llhNull G2 McFadden r2ML r2CU   
## -122.8561579 -281.3734187 317.0345216 0.5633697 0.5419941 0.7227114

qchisq(0.95,6)

## [1] 12.59159

Uji Signifikan Parsial

summary(mlog)

##   
## Call:  
## glm(formula = Status\_Akhir ~ ., family = binomial, data = train\_baru)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -6.3614 1.1189 -5.685 1.31e-08 \*\*\*  
## Status\_Covid.19PROBABLE 38.9102 3193.4648 0.012 0.990279   
## Status\_Covid.19KONFIRMASI 1.0826 0.4389 2.467 0.013640 \*   
## PerawatanRANAP 3.9445 1.0458 3.772 0.000162 \*\*\*  
## PerawatanIGD -15.6627 1391.9735 -0.011 0.991022   
## Sesak\_Nafas1 -0.9818 0.4808 -2.042 0.041137 \*   
## Gangguan\_Makan1 6.4838 1.5279 4.244 2.20e-05 \*\*\*  
## DM1 -0.6685 4173.6931 0.000 0.999872   
## HT1 19.8707 2137.7203 0.009 0.992584   
## Umur>50 2.6800 0.4001 6.698 2.11e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 562.75 on 405 degrees of freedom  
## Residual deviance: 245.71 on 396 degrees of freedom  
## AIC: 265.71  
##   
## Number of Fisher Scoring iterations: 18

mlog2 <- glm(Status\_Akhir~ Umur+Status\_Covid.19+Perawatan+Sesak\_Nafas+Gangguan\_Makan, data = train\_baru, family = binomial)  
summary(mlog2)

##   
## Call:  
## glm(formula = Status\_Akhir ~ Umur + Status\_Covid.19 + Perawatan +   
## Sesak\_Nafas + Gangguan\_Makan, family = binomial, data = train\_baru)  
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -6.5485 1.1168 -5.863 4.53e-09 \*\*\*  
## Umur>50 2.8232 0.3922 7.199 6.06e-13 \*\*\*  
## Status\_Covid.19PROBABLE 36.9213 1932.4585 0.019 0.984757   
## Status\_Covid.19KONFIRMASI 1.1364 0.3958 2.871 0.004091 \*\*   
## PerawatanRANAP 3.9774 1.0452 3.805 0.000142 \*\*\*  
## PerawatanIGD -14.6299 833.9554 -0.018 0.986004   
## Sesak\_Nafas1 0.2320 0.3940 0.589 0.555927   
## Gangguan\_Makan1 5.3365 1.4852 3.593 0.000327 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 562.75 on 405 degrees of freedom  
## Residual deviance: 275.61 on 398 degrees of freedom  
## AIC: 291.61  
##   
## Number of Fisher Scoring iterations: 17

Rasio odds

exp(coef((mlog)))

## (Intercept) Status\_Covid.19PROBABLE Status\_Covid.19KONFIRMASI   
## 1.726873e-03 7.915967e+16 2.952281e+00   
## PerawatanRANAP PerawatanIGD Sesak\_Nafas1   
## 5.165153e+01 1.576742e-07 3.746240e-01   
## Gangguan\_Makan1 DM1 HT1   
## 6.544482e+02 5.124970e-01 4.263211e+08   
## Umur>50   
## 1.458513e+01

Uji Kesesuaian Model

logitgof(train\_baru$Status\_Akhir,fitted(mlog2))

## Warning in logitgof(train\_baru$Status\_Akhir, fitted(mlog2)): At least one cell  
## in the expected frequencies table is < 1. Chi-square approximation may be  
## incorrect.

## Warning in logitgof(train\_baru$Status\_Akhir, fitted(mlog2)): Not possible to  
## compute 10 rows. There might be too few observations.

##   
## Hosmer and Lemeshow test (binary model)  
##   
## data: train\_baru$Status\_Akhir, fitted(mlog2)  
## X-squared = 30.146, df = 6, p-value = 3.688e-05

qchisq(0.95,5)

## [1] 11.0705

Prediksi

# Prediksi probabilitas pada data validasi  
problog <- predict(mlog2, valid, type = "response")  
  
# Prediksi biner berdasarkan ambang batas 0.5  
predlog <- rep("SEMBUH", nrow(valid)) # Default prediksi adalah "SEMBUH"  
predlog[problog > 0.5] <- "MENINGGAL" # Jika probabilitas > 0.5, prediksi "MENINGGAL"  
  
# Pastikan kelas dalam predlog sama dengan kelas dalam valid$Status\_Akhir  
predlog <- factor(predlog, levels = c("SEMBUH", "MENINGGAL"))  
valid$Status\_Akhir <- factor(valid$Status\_Akhir, levels = c("SEMBUH", "MENINGGAL"))

Confusion Matrix Model Regresi Logistik

contingency\_table <- table(predlog, valid$Status\_Akhir)  
contingency\_table

##   
## predlog SEMBUH MENINGGAL  
## SEMBUH 99 5  
## MENINGGAL 64 6

conf\_matrix <- confusionMatrix(contingency\_table)  
conf\_matrix

## Confusion Matrix and Statistics  
##   
##   
## predlog SEMBUH MENINGGAL  
## SEMBUH 99 5  
## MENINGGAL 64 6  
##   
## Accuracy : 0.6034   
## 95% CI : (0.5266, 0.6767)  
## No Information Rate : 0.9368   
## P-Value [Acc > NIR] : 1   
##   
## Kappa : 0.0437   
##   
## Mcnemar's Test P-Value : 2.902e-12   
##   
## Sensitivity : 0.60736   
## Specificity : 0.54545   
## Pos Pred Value : 0.95192   
## Neg Pred Value : 0.08571   
## Prevalence : 0.93678   
## Detection Rate : 0.56897   
## Detection Prevalence : 0.59770   
## Balanced Accuracy : 0.57641   
##   
## 'Positive' Class : SEMBUH   
##

#citation("ROSE")

#RStudio.Version()