|  |  |  |
| --- | --- | --- |
| **Nama :** Ibnu Fajar Setiawan  **NIM :** 065002000006 | c9824-logouniversitastrisakti  **Praktikum Data Analitik** | **Modul 7**  **Nama Dosen:** **Syandra Sari, S.Kom, M.Kom** |
| **Hari/Tanggal** :  Kamis, 10 November 2022 | **Nama Aslab :**   1. **Azzahra Nuranisa (065001900044)** 2. **Ida Jubaidah (065001900037)** |

**Praktikum 7 – REGRESI LOGISTIK**

**DESKRIPSI MODUL** : Melakukan pengujian Regresi Logistik

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | Elemen Kompetensi | Indikator Kinerja | Jml  Jam | hlm |
| 1 | Mampu melakukan pengujian regresi Logistik | Mampu melakukan pengujian regresi Logistik | 2 |  |

**TEORI SINGKAT**

Berbeda dengan regresi lininer yang digunakan untuk memprediksi variabel Y yang bersifat kontinu, regresi logistic digunakan untuk memprediksi variabel Y yang bersifat kategorik. Kasus regresi logistic dengan Y yang terdiri dari hanya dua kelas dinamakan *binary classification problems (****binomial logistic regression)****.* Prediktor dapat bersifat kontinu, kategorik maupun gabungan keduanya.

# ELEMEN KOMPETENSI I

|  |  |
| --- | --- |
| Deskripsi : | Dapat melakukan pengujian regresi Logistik |
| Kompetensi Dasar : | Mampu melakukan pengujian regresi Logistik |

**Kasus 1 : Titanic Dataset**

Gunakan data berikut ini untuk membangun model prediktif. Berikan interpretasi atas setiap output yang dihasilkan. Mulailah analisis dengan membuat tabulasi silang setiap predictor yang bersifat kategorik dengan respon (Y).

### Data Dictionary

|  |  |  |
| --- | --- | --- |
| **Variable** | **Definition** | **Key** |
| survival | Survival | 0 = No, 1 = Yes |
| pclass | Ticket class | 1 = 1st, 2 = 2nd, 3 = 3rd |
| sex | Sex |  |
| Age | Age in years |  |
| sibsp | # of siblings / spouses aboard the Titanic |  |
| parch | # of parents / children aboard the Titanic |  |
| ticket | Ticket number |  |
| fare | Passenger fare |  |
| cabin | Cabin number |  |
| embarked | Port of Embarkation | C = Cherbourg, Q = Queenstown, S = Southampton |

**REGRESI LOGISTIK**

|  |
| --- |
| > databaru=read.delim("clipboard")  > str(databaru)  > sampel1<-sample(1:nrow(databaru),0.75\*nrow(databaru))  > traininglogistik<-data.frame(databaru)[sampel1,]  > testinglogistik<-data.frame(databaru)[-sampel1,]  > modellogistik=glm(Survived~.,data=traininglogistik,family = binomial)  > summary(modellogistik) |

**#MELAKUKAN PREDIKSI**

|  |
| --- |
| > prediksilogistik=predict(modellogistik,testinglogistik)  > pred\_logreg<-as.numeric(prediksilogistik>.5)  > tabel\_logreg<-table(pred\_logreg,testinglogistik$Survived)  > tabel\_logreg |

**ELEMEN KOMPETENSI II**

|  |  |
| --- | --- |
| Deskripsi : | Dapat melakukan pengujian regresi Logistik dengan dataset Iris |
| Kompetensi Dasar : | Mampu melakukan pengujian regresi Logistik dengan Dataset Iris |

# Kasus 2. The iris data set (*species virginica and versicolor only)*

Gunakan data berikut ini untuk membangun model prediktif. Berikan interpretasi atas setiap output yang dihasilkan. Mulailah analisis dengan membuat tabulasi silang setiap predictor yang bersifat kategorik dengan respon (Y).

# make a reduced iris data set that only contains virginica and versicolor species

> library(dplyr)

> iris.small <- filter(iris, Species %in% c("virginica", "versicolor"))

# logistic regression

> glm.out <- glm(Species ~ Sepal.Width + Sepal.Length + Petal.Width + Petal.Length,

+ data = iris.small,

+ family = binomial) # family = binomial required for logistic regression

> summary(glm.out)

> exp(coef(glm.out))

> glm.out <- glm(Species ~ Sepal.Width + Petal.Width + Petal.Length,

+ data = iris.small,

+ family = binomial)

> exp(coef(glm.out))

*Text

Description automatically generated*

*Text

Description automatically generated*

**TUGAS**

*Make a plot of the fitted probability as a function of the linear predictor, colored by species identity. Hint: you will have to make a new data frame combining data from the fitted model with data from the iris.small data frame.*

> lr\_data <- data.frame(predictor=glm.out$linear.predictors, prob=glm.out$fitted.values, Species=iris.small$Species)

> ggplot(lr\_data, aes(x=predictor, y=prob, color=Species)) + geom\_point()

Chart, scatter chart

Description automatically generated

*Make a density plot that shows how the two species are separated by the linear predictor.*

ggplot(lr\_data, aes(x=predictor, fill=Species)) + geom\_density(alpha=.5)

Chart, histogram

Description automatically generated

*Assume you have obtained samples from three plants, with measurements as listed below. Predict the likelihood that each of these plants belongs to the species virginica.*

> plant1 <- data.frame(Sepal.Length=6.4, Sepal.Width=2.8, Petal.Length=4.6, Petal.Width=1.8)

> plant2 <- data.frame(Sepal.Length=6.3, Sepal.Width=2.5, Petal.Length=4.1, Petal.Width=1.7)

> plant3 <- data.frame(Sepal.Length=6.7, Sepal.Width=3.3, Petal.Length=5.2, Petal.Width=2.3)

> predict(glm.out, plant1, type="response")

> predict(glm.out, plant2, type="response")

> predict(glm.out, plant3, type="response")

Graphical user interface, text, application

Description automatically generated

*Pick a cutoff predictor value at which you would decide that a specimen belongs to virginica rather than versicolor. Calculate how many virginicas you call correctly and how many incorrectly given that choice.*

> cutoff <- 0

> virg\_true <- sum(lr\_data$predictor > cutoff & lr\_data$Species=="virginica")

> virg\_false <- sum(lr\_data$predictor <= cutoff & lr\_data$Species=="virginica")

> virg\_true

> virg\_false

A picture containing text

Description automatically generated

*Now do the same calculation for versicolor.*

> vers\_true <- sum(lr\_data$predictor <= cutoff & lr\_data$Species=="versicolor")

> vers\_false <- sum(lr\_data$predictor > cutoff & lr\_data$Species=="versicolor")

> vers\_true

> vers\_false

Text

Description automatically generated with medium confidence

*If we define a call of virginica as a positive and a call of versicolor as a negative, what are the true positive rate (sensitivity, true positives divided by all possible positives) and the true negative rate (specificity, true negatives divided by all possible negatives) in your analysis?*

> tp <- virg\_true/(virg\_true + virg\_false)

> tn <- vers\_true/(vers\_true + vers\_false)

> tp

> tn

Text

Description automatically generated with medium confidence

Sumber : [**http://wilkelab.org/classes/SDS348/2015\_spring\_worksheets/class11\_solutions.html**](http://wilkelab.org/classes/SDS348/2015_spring_worksheets/class11_solutions.html)

**1.** **Cek List**

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Elemen Kompetensi** | **Penyelesaian** | |
| **Selesai** | **Tidak** |
| **1** | **Elemen Kompetensi I**  Dapat melakukan pengujian regresi Logistik. | Checkmark with solid fill |  |
| **2** | **Elemen Kompetensi II**  Dapat melakukan pengujian regresi Logistik dengan dataset Iris. | Checkmark with solid fill |  |

**2.** **Form Umpan Balik**

|  |  |  |
| --- | --- | --- |
| **Elemen Kompetensi** | **Waktu Pengerjaan** | **Kriteria** |
| **Elemen Kompetensi I**  Dapat melakukan pengujian regresi Logistik. | 30 | 1 |
| **Elemen Kompetensi II**  Dapat melakukan pengujian regresi Logistik dengan dataset Iris | 30 | 1 |

Kriteria

1.Sangat Menarik

2.Cukup Menarik

3.Kurang Menarik

4.Sangat Kurang Menarik