Nama Kelompok :

* Mario Andreas 15.01.53.0001
* Muchammad Teguh S 15.01.53.0034
* Irfan Ardhiyanto 15.01.53.0061

**Program Naive Bayes Menggunakan Bahasa R**

install.packages("naivebayes")

install.packages("gdata")

install.packages("e1071")

install.packages("mlr")

install.packages("dplyr")

install.packages("psych")

install.packages("ggplot2")

install.packages("magrittr")

library(naivebayes)

library(gdata)

library(e1071)

library(mlr)

library(dplyr)

library(psych)

library(ggplot2)

library(magrittr)

# melihat direktori data sedang aktif

getwd()

# menentukan folder lokasi data

#setwd("D:/datamining2/data/winequality-red.csv")

# membuka data

data <- read.csv(file = 'D:/datamining2/data/winequality-red.csv', header = TRUE, sep = ";", as.is = 1)

# melihat data

head(data)

# melihat array data ke-1

data[1, ]

str(data)

xtabs(~alcohol+quality, data = data)

#ubah factor

data$alcohol <- as.factor(data$alcohol)

data$quality <- as.factor(data$quality)

# Visualization

pairs.panels(data[-1])

data %>%

ggplot(aes(x=alcohol, y=quality, fill = quality)) +

geom\_boxplot() +

ggtitle("Box Plot")

data %>% ggplot(aes(x=quality, fill = quality)) +

geom\_density(alpha=0.5, color= 'red') +

ggtitle("Density Plot")

# Data Partition

set.seed(12)

ind <- sample(2, nrow(data), replace = T, prob = c(0.8, 0.2))

wine <- data[ind == 1,]

wine2 <- data[ind == 2,]

# Naive Bayes Model

model <- naive\_bayes(quality ~ ., data = wine, usekernel = T)

model

wine %>%

filter(quality == "1") %>%

summarise(mean(pH), sd(pH))

plot(model)

# Predict

p <- predict(model, wine, type = 'prob')

head(cbind(p, wine))

# Confusion Matrix - wine data

p1 <- predict(model, wine)

(tab1 <- table(p1, wine$quality))

1 - sum(diag(tab1)) / sum(tab1)

# Confusion Matrix - wine2 data

p2 <- predict(model, wine2)

(tab2 <- table(p2, wine2$quality))

1 - sum(diag(tab2)) / sum(tab2)

#countsToCases <- function(x,countcol="Freq")

#{

# Get the row indices to pull from x

#idx <- rep.int(seq\_len(nrow(x)),x[[countcol]])

# Drop count column

# x[[countcol]]<-NULL

# Get the rows from x

# x[idx,]

#}

#caseTita<-countsToCases(as.data.frame(dataku))

#nrow(caseTita)

#data$fixed acidity <- as.factor(data$V1)

#data$citric acid <- as.factor(data$V3)

#data$residual.sugar <- as.factor(data$V4)

#data$alcohol <- as.factor(data$V11)

# Visualization

#pairs.panels(data[-1])

#data %>%

# ggplot(aes(x=fixed acidity, y=citric.acid, fill = pH)) +

# geom\_boxplot() +

# ggtitle("Box Plot")

**Hasil**



