

Minusha Attygala

27305

Programming in R – Lab Report

- Without Preprocessing

The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help, and Addins. Below the menu is a toolbar with various icons. The main workspace shows two tabs: "Without Preprocessing.R" and "With Preprocessing.R". The "Without Preprocessing.R" tab contains the following R code:

```
1 # Load Libraries
2 library(naivebayes)
3 library(caret)
4 library(e1071)
5
6 # Load the dataset
7 data <- read.csv("C:\\\\Users\\\\Minusha Attygala\\\\OneDrive\\\\Documents\\\\R\\\\Lab Report\\\\student_portuguese_clean.csv")
8
9 # Split the dataset into training and testing sets
10 set.seed(600)
11 splitIndex <- createDataPartition(data$final_grade, p = 0.7, list = FALSE)
12 train_data <- data[splitIndex, ]
13 test_data <- data[-splitIndex, ]
14
15 # Train the Naive Bayes model
16 model <- naiveBayes(final_grade ~ ., data = train_data)
17
18 # Make predictions on the test set
19 predictions <- predict(model, test_data)
20
21 # Evaluate the model
22 accuracy <- sum(predictions == test_data$final_grade) / nrow(test_data)
23 print(paste("Accuracy without preprocessing:", round(accuracy, 4)))
24
```

The "Console" tab at the bottom shows the output of running this script in R version 4.3.1:

```
R 4.3.1 --> 
> library(naivebayes)
> model <- naiveBayes(final_grade ~ ., data = train_data)
> # Make predictions on the test set
> predictions <- predict(model, test_data)
> # Evaluate the model
> accuracy <- sum(predictions == test_data$final_grade) / nrow(test_data)
> print(paste("Accuracy without preprocessing:", round(accuracy, 4)))
[1] "Accuracy without preprocessing: 0.2694"
>
```

Accuracy without preprocessing: 0.2694

This accuracy value represents the proportion of correctly classified instances by the Naive Bayes model when no preprocessing is applied to the data. Here, approximately 26.94% of the instances are correctly classified.

- With Preprocessing

The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. Below the menu is a toolbar with various icons. The main workspace shows two tabs: 'Without Preprocessing.R' and 'With Preprocessing.R'. The 'With Preprocessing.R' tab contains the following R code:

```

6 # Load the dataset
7 data <- read.csv("C:/Users/Minusha Attygala/OneDrive/Documents/R/Lab Report/student_portuguese_clean.csv")
8
9 # Preprocess the data
10 data_preprocessed <- data %>% mutate(
11   age = as.factor(age),
12   study_time = as.factor(study_time),
13   final_grade = as.factor(final_grade)
14 )
15
16 # Split the preprocessed data into training and testing sets
17 train_index <- createDataPartition(data_preprocessed$final_grade, p = 0.7, list = FALSE)
18 train_data <- data[train_index, ]
19 test_data <- data[-train_index, ]
20
21 # Set seed for reproducibility
22 resample <- createResample(data_preprocessed$final_grade, times = 1, list = FALSE)
23
24 # Train the Naive Bayes model with preprocessing
25 model <- naiveBayes(final_grade ~ ., data = train_data)
26
27 # Predict the final grade for the test data
28 predictions <- predict(model, test_data)
29
30 # Evaluate the model performance with preprocessing
31 confusion_matrix <- table(test_data$final_grade, predictions)
32 accuracy <- mean(diag(confusion_matrix))
33 print(paste("Accuracy with preprocessing:", accuracy))

```

The bottom panel shows the R Console with the following output:

```

> accuracy <- mean(diag(confusion_matrix))
> print(paste("Accuracy with preprocessing:", accuracy))
[1] "Accuracy with preprocessing: 0.538461538461538"
>

```

Accuracy with preprocessing: 0.538461538461538

This accuracy value, ranging between 0 and 1, indicates the proportion of correctly classified instances by the Naive Bayes model when preprocessing is applied to the data. In this case, about 53.85% of the instances are correctly classified.

Comparing the two accuracies:

The accuracy with preprocessing is higher than the accuracy without preprocessing. This suggests that the preprocessing steps applied to the data have a positive impact on the model's performance, as it leads to a higher proportion of correct predictions.