
Title: Business Intelligence Project Author: Natasha Gichira, Emmanuel Agre, Ryan Gitonga, Fredrick Koech, Dennis Muriuki

Date: (23/10/23)

Student Details

Student ID Number	124255,120415,124562,1182
Student Name	Natasha Gichira, Em- manuel Agre, Ryan Gi- tonga, Fredrick Koech, Dennis Muriuki
BBIT 4.2 Group	C
BI Project Group Name/ID (if applicable)	Challengers

EVALUATION METRICS

Loading the Dataset

```
library(readr)
```

```
machine_readable_business_employment_data_jun_2023_quarter <- read_csv("machine-readable-business-employment-data-jun-2023-quarter.csv")
```

Source:

The dataset that was used can be downloaded here: <https://www.stats.govt.nz/large-datasets/csv-files-for-download/>

Reference:

*<## StatsNZ & Tatauranga, A. (2023). Business Employment Data June Quater . <https://www.stats.govt.nz/large-datasets/csv-files-for-download/> # nolint -->

Executable code

```
set.seed(7)
class_model_glm <- train(class ~ ., data = machine_readable_business_employment_data_jun_2023_quarter_train,
method = "glm", metric = "Accuracy", trControl = train_control)
```

```

print(class_model_glm)

set.seed(7) class_model_knn <- train(class ~ ., data = machine_readable_business_employment_data_jun_2023_quarter,
method = "knn", metric = "ROC", trControl = train_control)

print(class_model_knn)

predictions <- predict(class_model_knn, machine_readable_business_employment_data_jun_2023_quarter_test[,
1:8])

print(predictions) confusion_matrix <- caret::confusionMatrix(predictions, machine_readable_business_employment_data_jun_2023_quarter_test[,
1:9]$class)

print(confusion_matrix)

predictions <- predict(class_model_knn, machine_readable_business_employment_data_jun_2023_quarter_test[,
1:8], type = "prob")

print(predictions)

roc_curve <- roc(machine_readable_business_employment_data_jun_2023_quarter_test$class, predictions$neg)

```

Plot the ROC curve

```

plot(roc_curve, main = "ROC Curve for KNN Model", print.auc = TRUE, print.auc.x = 0.6, print.auc.y =
0.6, col = "blue", lwd = 2.5)

train_control <- trainControl(method = "repeatedcv", number = 5, repeats = 3, classProbs = TRUE,
summaryFunction = mnLogLoss) set.seed(7)

Employment_model_cart <- train(class ~ ., data = machine_readable_business_employment_data_jun_2023_quarter,
method = "rpart", metric = "logLoss", trControl = train_control)

print(Employment_model_cart)

```