



MACHINE LEARNING WITH TREE-BASED MODELS IN R

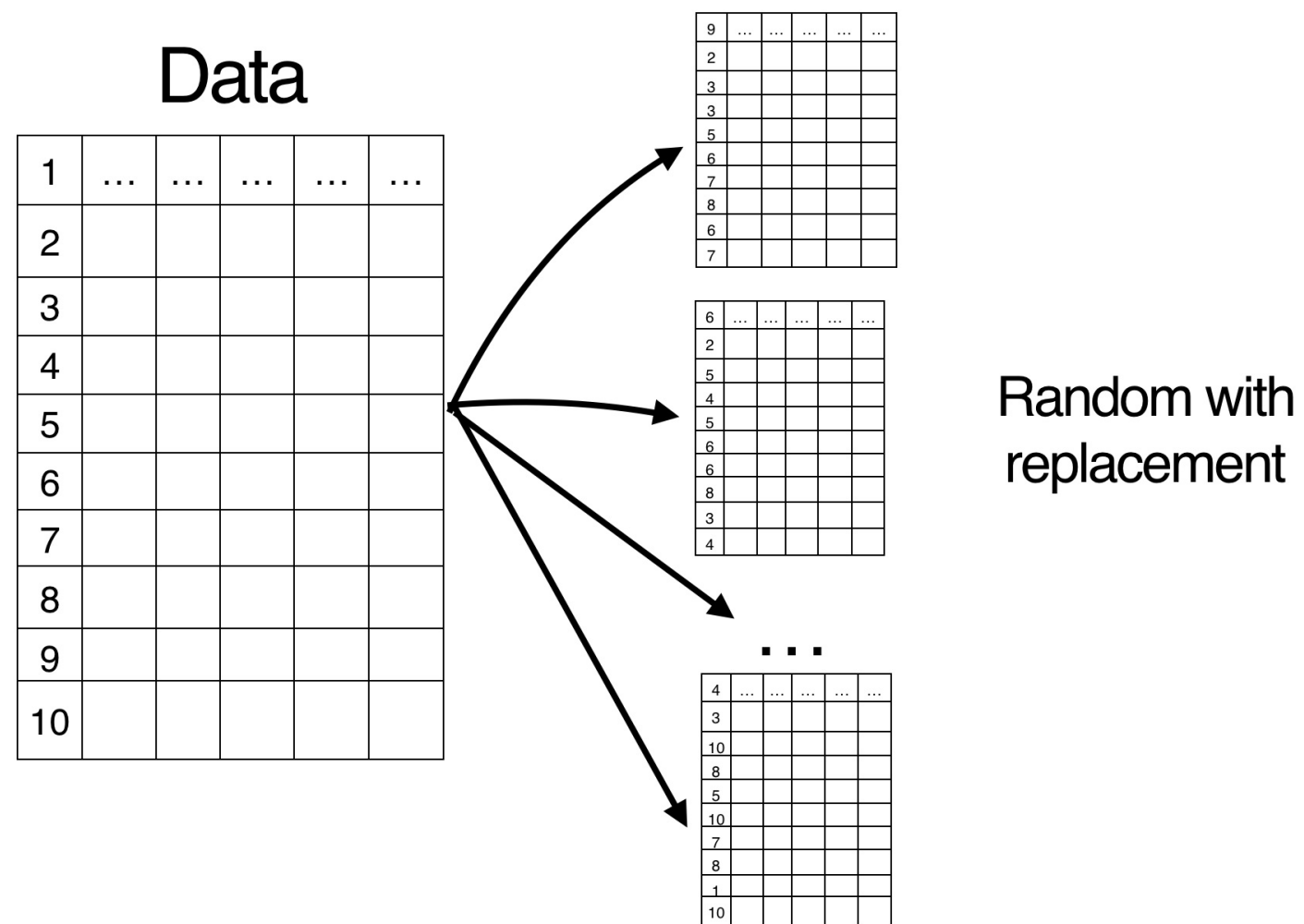
# Introduction to bagged trees

Gabriela de Queiroz  
Instructor



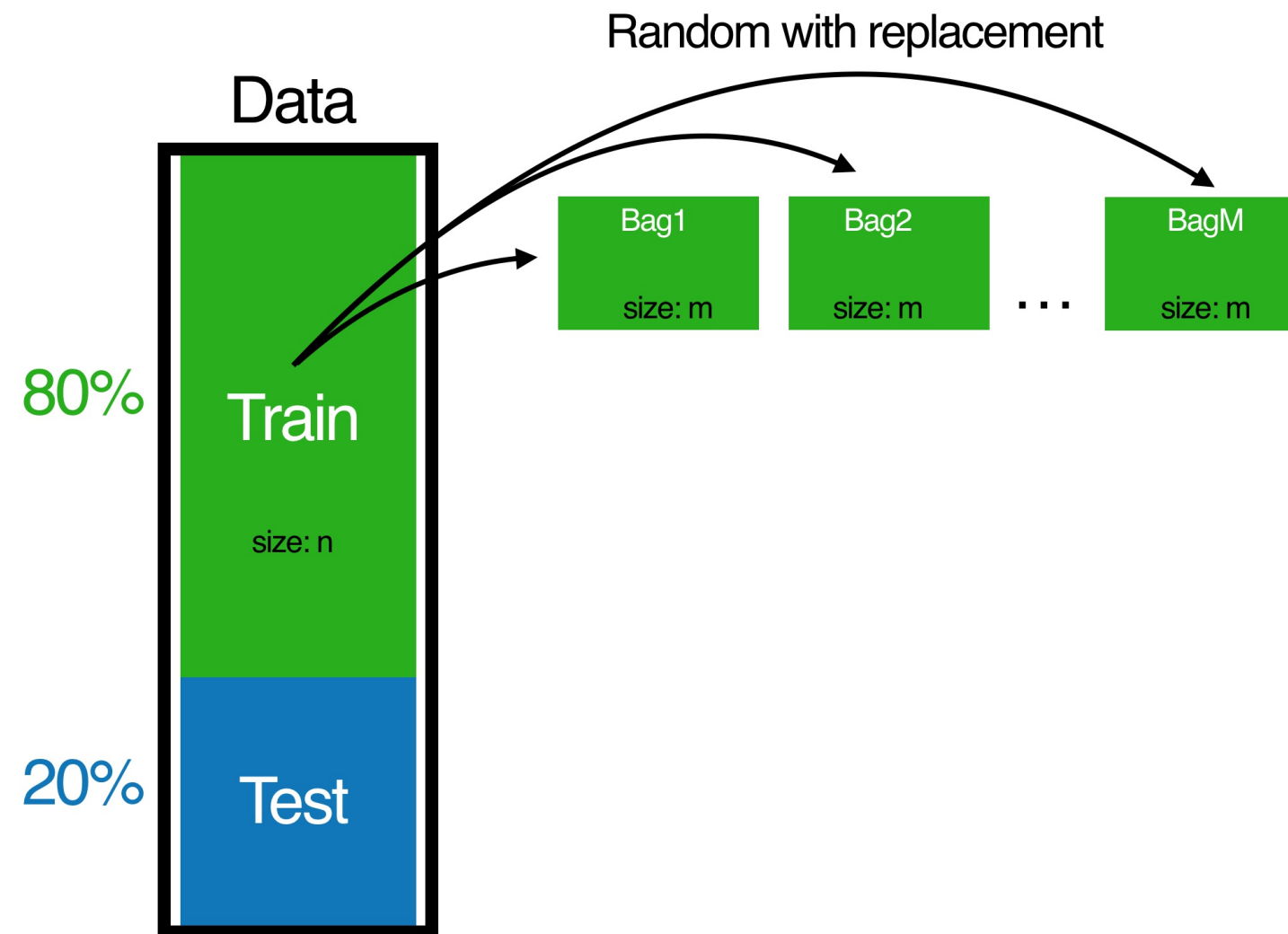
# Bagging

## Bootstrap **AGG**regat**ING**



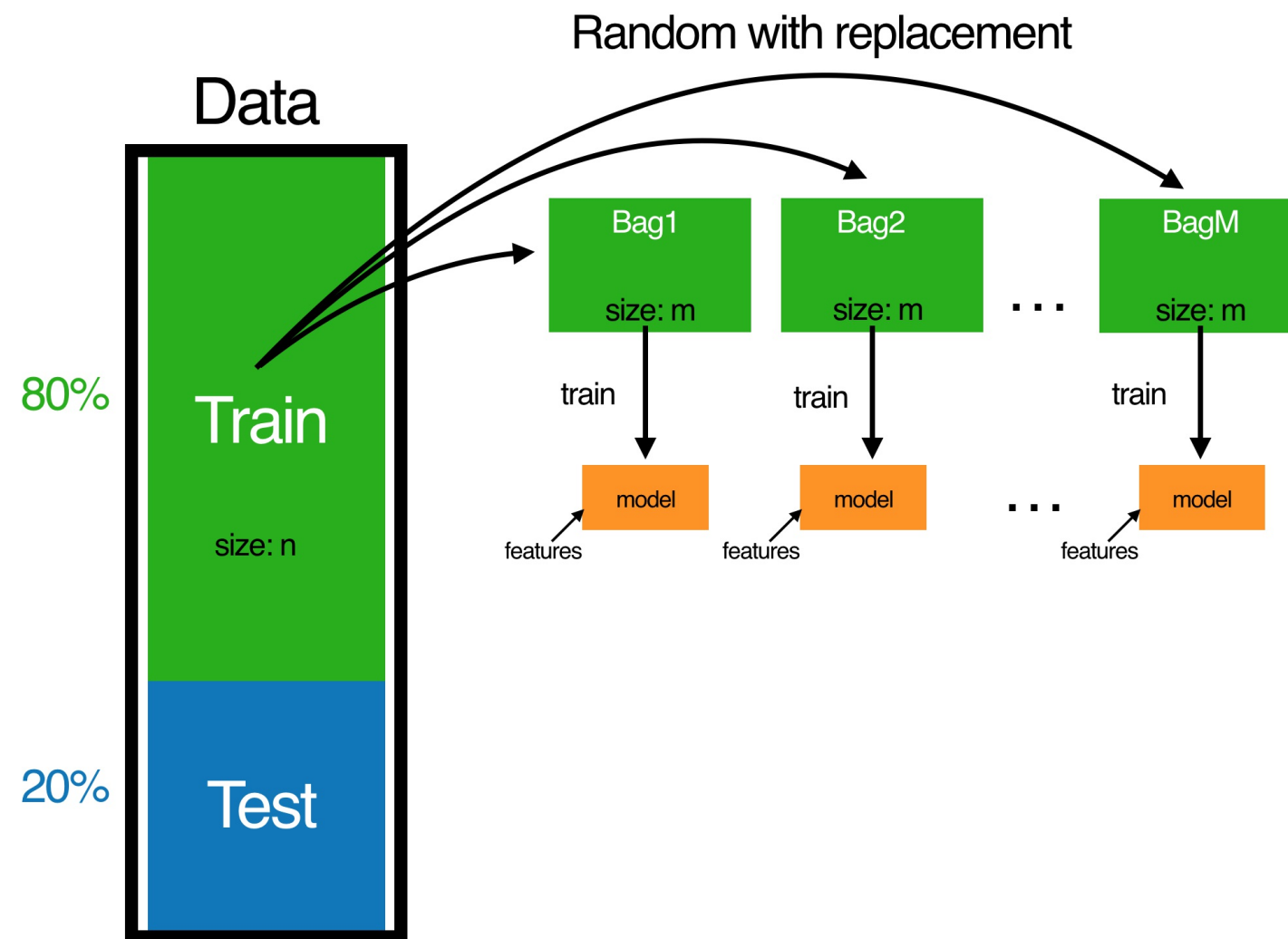


# Step 1



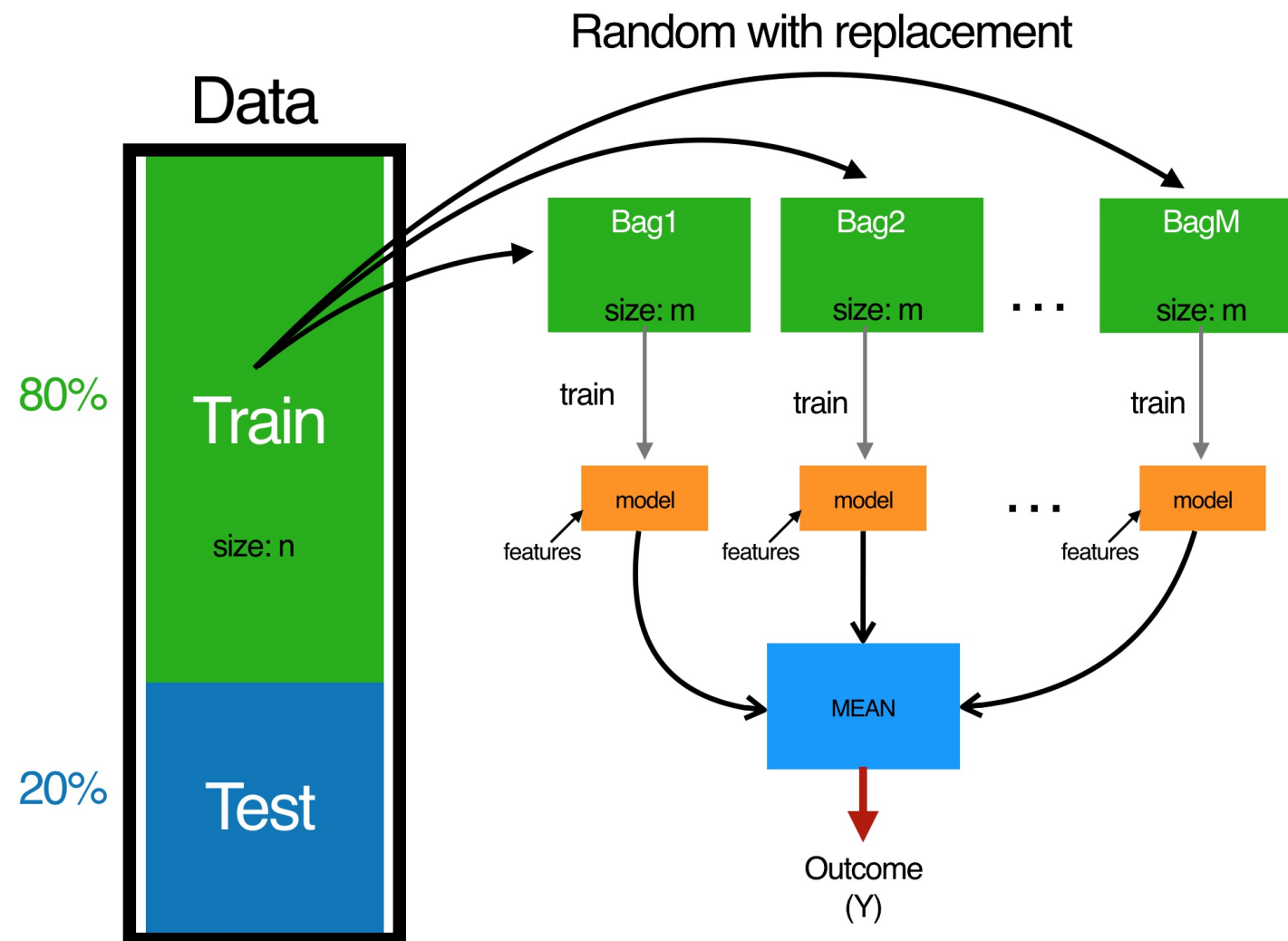


# Step 2





# Bagging





# Bagging in R

```
> library(ipred)
> bagging(formula = response ~ ., data = dat)
```



## MACHINE LEARNING WITH TREE-BASED MODELS IN R

**Let's practice!**



MACHINE LEARNING WITH TREE-BASED MODELS IN R

# Evaluating the performance of bagged tree models

Gabriela de Queiroz  
Instructor





# Generate Predictions

```
> class_predictions <- predict(object = rest_model_bag, # model object
                               newdata = restaurant_test, # test dataset
                               type = "class") # return classification labels

> print(class_predictions)
[1] Yes Yes Yes Yes No No Yes No Yes Yes Yes Yes No No No Yes No Yes Yes I
Levels: No Yes
```



# Confusion Matrix

```
> confusionMatrix(data = class_predictions,           # predicted classes
                  reference = restaurant_test$will_wait) # actual classes
```

## Confusion Matrix and Statistics

	Reference	
Prediction	No	Yes
No	5	3
Yes	1	12

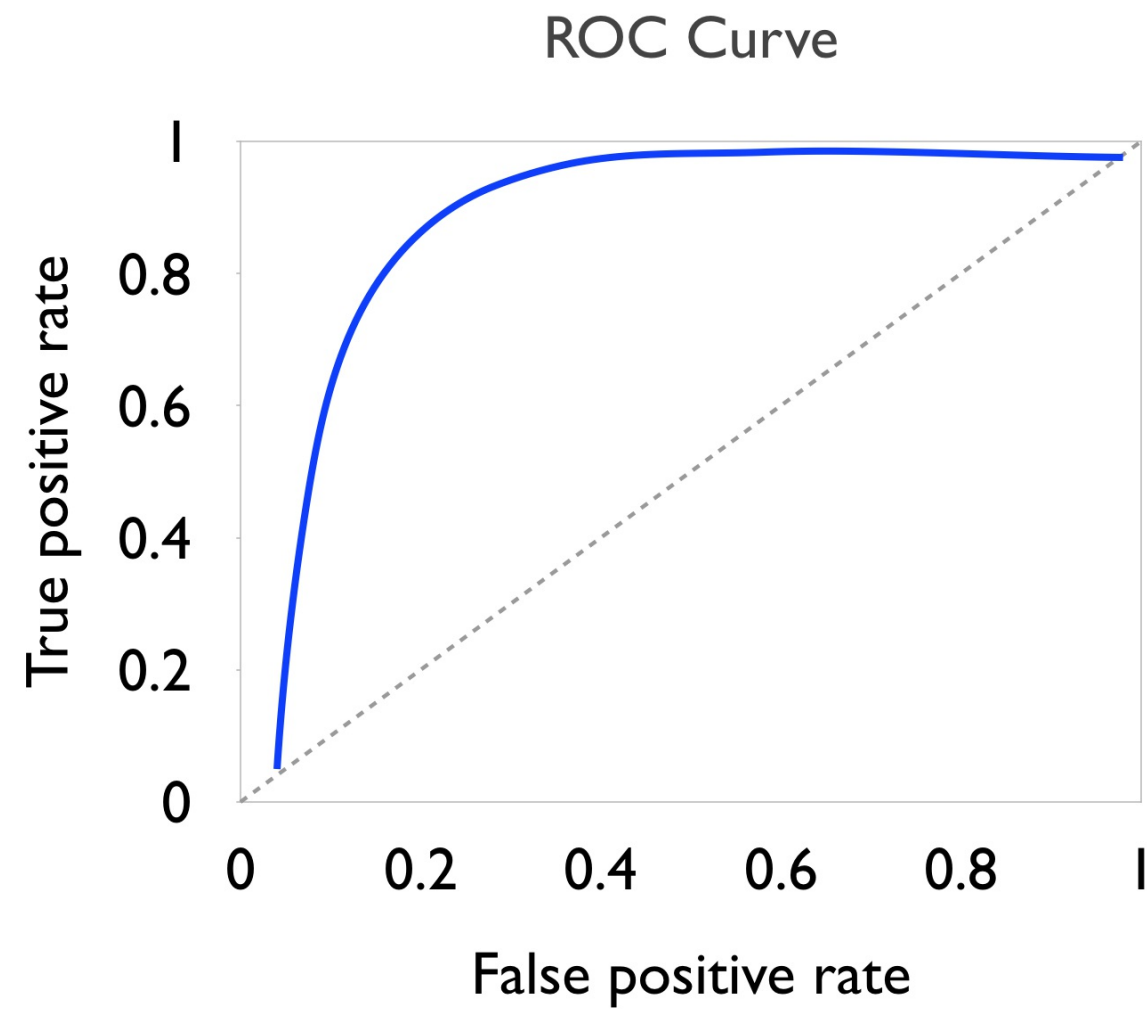
Accuracy : 0.8095  
95% CI : (0.5809, 0.9455)  
No Information Rate : 0.7143  
P-Value [Acc > NIR] : 0.2402

Kappa : 0.5758  
Mcnemar's Test P-Value : 0.6171

Sensitivity : 0.8333  
Specificity : 0.8000  
Pos Pred Value : 0.6250  
Neg Pred Value : 0.9231  
...



# ROC Curve





# AUC

```
> library(Metrics)
> auc(actual, predicted)
[1] .76765
```



## MACHINE LEARNING WITH TREE-BASED MODELS IN R

**Let's practice!**



MACHINE LEARNING WITH TREE-BASED MODELS IN R

# Cross-validation

Gabriela de Queiroz  
Instructor



# K-fold Cross-validation

- dataset size = 200 rows
- $k = 10$  (number of cross validation folds)

20	20	20	20	20
20	20	20	20	20



# K-fold Cross-validation

20	20	20	20	20
20	20	20	20	20

- 10 estimates of test set AUC
- the average is the cross-validated estimate of AUC





# Using caret for cross-validating models

```
> library(caret)
```

- train()
- trainControl()



# Training configuration

```
# Specify the training configuration
ctrl <- trainControl(method = "cv",           # Cross-validation
                     number = 5,            # 5 folds
                     classProbs = TRUE,     # For AUC
                     summaryFunction = twoClassSummary) # For AUC
```



# Training configuration

```
set.seed(1) #for reproducibility
credit_model <- train(default ~ .,
                      data = credit_train,
                      method = "treebag",
                      metric = "ROC",
                      trControl = ctrl)
```



## MACHINE LEARNING WITH TREE-BASED MODELS IN R

**Let's practice!**