**STT 811**

**In-Class Assignment 15**

This problem will use the Weekly dataset, with Direction as the target.

1. Take a look at the data.

View(weekly)

1. Split the data into training and test datasets (with a 75/25 split).

split\_pct <- 0.75

n <- length(weekly$Direction)\*split\_pct

row\_samp <- sample(1:length(weekly$Direction), n, replace = FALSE)

train <- weekly[row\_samp,]

test <- weekly[-row\_samp,]

1. Build a xgboost based on the lagged values and volume, using nrounds = 50, eta = 0.1, and max\_depth = 3. Calculate the accuracy from the confusion matrix.

weekly\_xgb <- xgboost(data = data.matrix(train[,c(2:7)]), nrounds = 50, max\_depth = 3, eta = 0.1, label = train$Direction, objective = "binary:logistic")

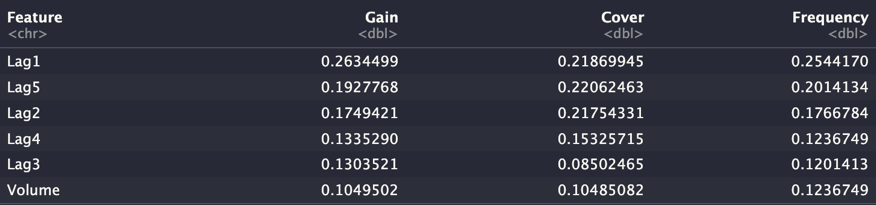
pred <- predict(weekly\_xgb, data.matrix(test[,c(2:7)]))

confusionMatrix(as.factor(as.integer(2\*pred)), as.factor(test$Direction))

Text

Description automatically generated

1. Calculate the variable importance from this model

xgb.importance(colnames(train[,c(2:7)]), model = weekly\_xgb)

1. Try different values for the hyperparameters, calculate the confusion matrix for each, and compare the results. Let the class know which set works particularly well.

Using: nrounds = 75, max\_depth = 5, eta = 0.8, the accuracy was boosted to 0.5531136.