Martin\_Alonso\_HW7

Martin Alonso

November 25, 2018

# HW7 Instructions

In this homework, you will use SVMs, kNN, and Random Forest algorithms for handwriting recognition, and compare their performance with the naïve Bayes and decision tree models you built in previous week.

## Loading the packages and data

We’ll start by loading the caret, e1071, and dplyr packages that will be used. Though the original exercise involved submitting to Kaggle, this is no longer necessary, so we’ll just work with the training set. Due to the size of the dataset, we will only explore part of it so as to not saturate the output.

# Load the required packages.  
require(caret)

## Loading required package: caret

## Loading required package: lattice

## Loading required package: ggplot2

require(e1071)

## Loading required package: e1071

require(rpart)

## Loading required package: rpart

require(dplyr)

## Loading required package: dplyr

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

require(stringr)

## Loading required package: stringr

require(randomForest)

## Loading required package: randomForest

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

##   
## Attaching package: 'randomForest'

## The following object is masked from 'package:dplyr':  
##   
## combine

## The following object is masked from 'package:ggplot2':  
##   
## margin

# Load the train dataset.   
dataSet <- read.csv('Kaggle-digit-train.csv')  
  
# Explore the dimension of the dataset, and the structure and summary of the first ten rows.   
dim(dataSet)

## [1] 42000 785

str(dataSet[, 1:10])

## 'data.frame': 42000 obs. of 10 variables:  
## $ label : int 1 0 1 4 0 0 7 3 5 3 ...  
## $ pixel0: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ pixel1: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ pixel2: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ pixel3: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ pixel4: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ pixel5: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ pixel6: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ pixel7: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ pixel8: int 0 0 0 0 0 0 0 0 0 0 ...

summary(dataSet[, 1:10])

## label pixel0 pixel1 pixel2 pixel3   
## Min. :0.000 Min. :0 Min. :0 Min. :0 Min. :0   
## 1st Qu.:2.000 1st Qu.:0 1st Qu.:0 1st Qu.:0 1st Qu.:0   
## Median :4.000 Median :0 Median :0 Median :0 Median :0   
## Mean :4.457 Mean :0 Mean :0 Mean :0 Mean :0   
## 3rd Qu.:7.000 3rd Qu.:0 3rd Qu.:0 3rd Qu.:0 3rd Qu.:0   
## Max. :9.000 Max. :0 Max. :0 Max. :0 Max. :0   
## pixel4 pixel5 pixel6 pixel7 pixel8   
## Min. :0 Min. :0 Min. :0 Min. :0 Min. :0   
## 1st Qu.:0 1st Qu.:0 1st Qu.:0 1st Qu.:0 1st Qu.:0   
## Median :0 Median :0 Median :0 Median :0 Median :0   
## Mean :0 Mean :0 Mean :0 Mean :0 Mean :0   
## 3rd Qu.:0 3rd Qu.:0 3rd Qu.:0 3rd Qu.:0 3rd Qu.:0   
## Max. :0 Max. :0 Max. :0 Max. :0 Max. :0

# Let's also check what the first 5 rows of the dataset look like.  
head(dataSet[, 1:10], n = 5)

## label pixel0 pixel1 pixel2 pixel3 pixel4 pixel5 pixel6 pixel7 pixel8  
## 1 1 0 0 0 0 0 0 0 0 0  
## 2 0 0 0 0 0 0 0 0 0 0  
## 3 1 0 0 0 0 0 0 0 0 0  
## 4 4 0 0 0 0 0 0 0 0 0  
## 5 0 0 0 0 0 0 0 0 0 0

Now that we have checked what the dataset look like, it’s time to do some additional processing and then train the models.

# The Datasets.

For this exercise, we’ll work on two new models: k-Nearest Neighbours (kNN) and Support Vector Machines (SVM). We’ll also work on a third model - Random Forest - which is an expansion of the decision tree algorithm we have previously worked on. Also, due to the size of the training set, we’ll only work with 30 percent of the dataset, setting the seeds and selecting two-thirds of this subset to serve as our training set and the remaining third to serve as our testing set. All these subsets will be chosen at random.

## Data preprocessing

Furthermore, because there are so many variables, we’ll make sure that we work with a select number of objects. Just to make sure that we don’t eliminate unneccesary columns, only columns where there is no variance will be eliminated.

# Replace NAs with 0, count number of columns, and create a list that will populate the columns that sum 0.   
dataSet[is.na(dataSet)] <- 0  
cols <- ncol(dataSet)  
vars <- list()  
  
# Loop over the variables to find which present no variance.   
for(i in 2:cols){  
 colVar <- var(dataSet[[i]])  
 if(colVar == 0){  
 vars <- append(vars, i)  
 }  
}  
  
# Drop the columns that have no variance in both the test and training set.   
colsToDrop <- unlist(vars)  
  
dataSet <- dataSet[, -colsToDrop]  
  
# With the columns now selected, we'll now choose 20 percent of both the training and testing datasets.   
set.seed(1024)  
  
split <- sample(nrow(dataSet), nrow(dataSet) \* .3)  
dataSubset <- dataSet[split, ]  
  
# Now to select the training and testing sets from this subset.   
set.seed(760)  
  
trainSplit <- sample(nrow(dataSubset), nrow(dataSubset) \* 2/3)  
trainSet <- dataSubset[trainSplit, ]  
testSet <- dataSubset[-trainSplit, ]

After selecting the columns, we’ll end up working with 8,400 records to train the model and 4,200 records to test the model.

# Training and testing the models.

## k-Nearest Neighbours

Starting off with kNN, we’ll build the model using two-thirds of the training set and then use the remaining third of the same data to test the results. Just to avoid redoing this process, we’ll set the seed for the split and work from there.

# We'll first start by converting the label variable into a factor. Because we are building a classifier that will determine the probability that a observation belongs to a certain number, we need to convert the labels into factors that can then be converted to variable names  
trainSet$label <- factor(paste0('X', trainSet$label), levels = c('X0', 'X1', 'X2', 'X3', 'X4', 'X5', 'X6', 'X7', 'X8', 'X9'))  
testSet$label <- factor(paste0('X', testSet$label), levels = c('X0', 'X1', 'X2', 'X3', 'X4', 'X5', 'X6', 'X7', 'X8', 'X9'))  
  
# Now to start creating the kNN. We'll do three-fold CV as in the previous exercise.   
set.seed(239)  
x <- trainControl(method = 'repeatedcv', repeats = 3, classProbs = TRUE)  
knnTrain <- train(label ~ ., data = trainSet, method = 'knn', preProcess = c('center', 'scale'), trControl = x, metric = 'ROC', tuneLength = 3)

## Warning in train.default(x, y, weights = w, ...): The metric "ROC" was not  
## in the result set. Accuracy will be used instead.

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel393, pixel587, pixel588, pixel615, pixel616, pixel617, pixel643,  
## pixel698, pixel702, pixel724, pixel725, pixel726, pixel752, pixel753,  
## pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel393, pixel587, pixel588, pixel615, pixel616, pixel617, pixel643,  
## pixel698, pixel702, pixel724, pixel725, pixel726, pixel752, pixel753,  
## pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel393, pixel587, pixel588, pixel615, pixel616, pixel617, pixel643,  
## pixel698, pixel702, pixel724, pixel725, pixel726, pixel752, pixel753,  
## pixel777, pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel533, pixel559, pixel587, pixel588, pixel615, pixel616, pixel617,  
## pixel642, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel533, pixel559, pixel587, pixel588, pixel615, pixel616, pixel617,  
## pixel642, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel533, pixel559, pixel587, pixel588, pixel615, pixel616, pixel617,  
## pixel642, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel777, pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel33, pixel34, pixel35, pixel50, pixel51, pixel59,  
## pixel62, pixel110, pixel113, pixel167, pixel169, pixel224, pixel252,  
## pixel280, pixel308, pixel336, pixel364, pixel504, pixel587, pixel588,  
## pixel615, pixel616, pixel617, pixel643, pixel698, pixel724, pixel725,  
## pixel726, pixel752, pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel33, pixel34, pixel35, pixel50, pixel51, pixel59,  
## pixel62, pixel110, pixel113, pixel167, pixel169, pixel224, pixel252,  
## pixel280, pixel308, pixel336, pixel364, pixel504, pixel587, pixel588,  
## pixel615, pixel616, pixel617, pixel643, pixel698, pixel724, pixel725,  
## pixel726, pixel752, pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel33, pixel34, pixel35, pixel50, pixel51, pixel59,  
## pixel62, pixel110, pixel113, pixel167, pixel169, pixel224, pixel252,  
## pixel280, pixel308, pixel336, pixel364, pixel504, pixel587, pixel588,  
## pixel615, pixel616, pixel617, pixel643, pixel698, pixel724, pixel725,  
## pixel726, pixel752, pixel753, pixel777, pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel49, pixel50, pixel51, pixel59, pixel60, pixel110,  
## pixel113, pixel167, pixel169, pixel224, pixel252, pixel280, pixel308,  
## pixel335, pixel336, pixel364, pixel587, pixel588, pixel615, pixel616,  
## pixel617, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel49, pixel50, pixel51, pixel59, pixel60, pixel110,  
## pixel113, pixel167, pixel169, pixel224, pixel252, pixel280, pixel308,  
## pixel335, pixel336, pixel364, pixel587, pixel588, pixel615, pixel616,  
## pixel617, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel49, pixel50, pixel51, pixel59, pixel60, pixel110,  
## pixel113, pixel167, pixel169, pixel224, pixel252, pixel280, pixel308,  
## pixel335, pixel336, pixel364, pixel587, pixel588, pixel615, pixel616,  
## pixel617, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel777, pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel363,  
## pixel364, pixel391, pixel419, pixel587, pixel588, pixel589, pixel615,  
## pixel616, pixel617, pixel643, pixel698, pixel724, pixel725, pixel726,  
## pixel752, pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel363,  
## pixel364, pixel391, pixel419, pixel587, pixel588, pixel589, pixel615,  
## pixel616, pixel617, pixel643, pixel698, pixel724, pixel725, pixel726,  
## pixel752, pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel363,  
## pixel364, pixel391, pixel419, pixel587, pixel588, pixel589, pixel615,  
## pixel616, pixel617, pixel643, pixel698, pixel724, pixel725, pixel726,  
## pixel752, pixel753, pixel777, pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel477, pixel587, pixel588, pixel615, pixel616, pixel617, pixel643,  
## pixel698, pixel724, pixel725, pixel726, pixel752, pixel753, pixel761,  
## pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel477, pixel587, pixel588, pixel615, pixel616, pixel617, pixel643,  
## pixel698, pixel724, pixel725, pixel726, pixel752, pixel753, pixel761,  
## pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel477, pixel587, pixel588, pixel615, pixel616, pixel617, pixel643,  
## pixel698, pixel724, pixel725, pixel726, pixel752, pixel753, pixel761,  
## pixel777, pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel58, pixel59, pixel110, pixel113,  
## pixel167, pixel169, pixel224, pixel252, pixel280, pixel308, pixel336,  
## pixel364, pixel587, pixel588, pixel615, pixel616, pixel617, pixel643,  
## pixel698, pixel724, pixel725, pixel726, pixel752, pixel753, pixel777,  
## pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel58, pixel59, pixel110, pixel113,  
## pixel167, pixel169, pixel224, pixel252, pixel280, pixel308, pixel336,  
## pixel364, pixel587, pixel588, pixel615, pixel616, pixel617, pixel643,  
## pixel698, pixel724, pixel725, pixel726, pixel752, pixel753, pixel777,  
## pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel58, pixel59, pixel110, pixel113,  
## pixel167, pixel169, pixel224, pixel252, pixel280, pixel308, pixel336,  
## pixel364, pixel587, pixel588, pixel615, pixel616, pixel617, pixel643,  
## pixel698, pixel724, pixel725, pixel726, pixel752, pixel753, pixel777,  
## pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel80, pixel81, pixel110,  
## pixel113, pixel167, pixel169, pixel224, pixel252, pixel280, pixel308,  
## pixel336, pixel364, pixel447, pixel587, pixel588, pixel615, pixel616,  
## pixel617, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel80, pixel81, pixel110,  
## pixel113, pixel167, pixel169, pixel224, pixel252, pixel280, pixel308,  
## pixel336, pixel364, pixel447, pixel587, pixel588, pixel615, pixel616,  
## pixel617, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel80, pixel81, pixel110,  
## pixel113, pixel167, pixel169, pixel224, pixel252, pixel280, pixel308,  
## pixel336, pixel364, pixel447, pixel587, pixel588, pixel615, pixel616,  
## pixel617, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel777, pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel307, pixel308, pixel335,  
## pixel336, pixel364, pixel447, pixel531, pixel559, pixel587, pixel588,  
## pixel615, pixel616, pixel617, pixel643, pixel698, pixel724, pixel725,  
## pixel726, pixel752, pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel307, pixel308, pixel335,  
## pixel336, pixel364, pixel447, pixel531, pixel559, pixel587, pixel588,  
## pixel615, pixel616, pixel617, pixel643, pixel698, pixel724, pixel725,  
## pixel726, pixel752, pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel307, pixel308, pixel335,  
## pixel336, pixel364, pixel447, pixel531, pixel559, pixel587, pixel588,  
## pixel615, pixel616, pixel617, pixel643, pixel698, pixel724, pixel725,  
## pixel726, pixel752, pixel753, pixel777, pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel49, pixel50, pixel51, pixel59, pixel110, pixel113,  
## pixel167, pixel169, pixel224, pixel252, pixel280, pixel308, pixel336,  
## pixel363, pixel364, pixel391, pixel419, pixel587, pixel588, pixel615,  
## pixel616, pixel617, pixel642, pixel643, pixel698, pixel702, pixel724,  
## pixel725, pixel726, pixel752, pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel49, pixel50, pixel51, pixel59, pixel110, pixel113,  
## pixel167, pixel169, pixel224, pixel252, pixel280, pixel308, pixel336,  
## pixel363, pixel364, pixel391, pixel419, pixel587, pixel588, pixel615,  
## pixel616, pixel617, pixel642, pixel643, pixel698, pixel702, pixel724,  
## pixel725, pixel726, pixel752, pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel49, pixel50, pixel51, pixel59, pixel110, pixel113,  
## pixel167, pixel169, pixel224, pixel252, pixel280, pixel308, pixel336,  
## pixel363, pixel364, pixel391, pixel419, pixel587, pixel588, pixel615,  
## pixel616, pixel617, pixel642, pixel643, pixel698, pixel702, pixel724,  
## pixel725, pixel726, pixel752, pixel753, pixel777, pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel33, pixel34, pixel35, pixel36, pixel50, pixel51,  
## pixel59, pixel110, pixel113, pixel167, pixel169, pixel224, pixel252,  
## pixel280, pixel308, pixel336, pixel364, pixel533, pixel587, pixel588,  
## pixel615, pixel616, pixel617, pixel643, pixel698, pixel724, pixel725,  
## pixel726, pixel752, pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel33, pixel34, pixel35, pixel36, pixel50, pixel51,  
## pixel59, pixel110, pixel113, pixel167, pixel169, pixel224, pixel252,  
## pixel280, pixel308, pixel336, pixel364, pixel533, pixel587, pixel588,  
## pixel615, pixel616, pixel617, pixel643, pixel698, pixel724, pixel725,  
## pixel726, pixel752, pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel33, pixel34, pixel35, pixel36, pixel50, pixel51,  
## pixel59, pixel110, pixel113, pixel167, pixel169, pixel224, pixel252,  
## pixel280, pixel308, pixel336, pixel364, pixel533, pixel587, pixel588,  
## pixel615, pixel616, pixel617, pixel643, pixel698, pixel724, pixel725,  
## pixel726, pixel752, pixel753, pixel777, pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel58, pixel59, pixel60, pixel110,  
## pixel113, pixel167, pixel169, pixel224, pixel252, pixel280, pixel308,  
## pixel336, pixel364, pixel504, pixel587, pixel588, pixel615, pixel616,  
## pixel617, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel58, pixel59, pixel60, pixel110,  
## pixel113, pixel167, pixel169, pixel224, pixel252, pixel280, pixel308,  
## pixel336, pixel364, pixel504, pixel587, pixel588, pixel615, pixel616,  
## pixel617, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel58, pixel59, pixel60, pixel110,  
## pixel113, pixel167, pixel169, pixel224, pixel252, pixel280, pixel308,  
## pixel336, pixel364, pixel504, pixel587, pixel588, pixel615, pixel616,  
## pixel617, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel777, pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel80, pixel81, pixel110,  
## pixel113, pixel167, pixel169, pixel224, pixel252, pixel280, pixel308,  
## pixel336, pixel364, pixel477, pixel587, pixel588, pixel615, pixel616,  
## pixel617, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel80, pixel81, pixel110,  
## pixel113, pixel167, pixel169, pixel224, pixel252, pixel280, pixel308,  
## pixel336, pixel364, pixel477, pixel587, pixel588, pixel615, pixel616,  
## pixel617, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel80, pixel81, pixel110,  
## pixel113, pixel167, pixel169, pixel224, pixel252, pixel280, pixel308,  
## pixel336, pixel364, pixel477, pixel587, pixel588, pixel615, pixel616,  
## pixel617, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel777, pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel589, pixel615, pixel616, pixel617, pixel643,  
## pixel698, pixel724, pixel725, pixel726, pixel752, pixel753, pixel761,  
## pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel589, pixel615, pixel616, pixel617, pixel643,  
## pixel698, pixel724, pixel725, pixel726, pixel752, pixel753, pixel761,  
## pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel589, pixel615, pixel616, pixel617, pixel643,  
## pixel698, pixel724, pixel725, pixel726, pixel752, pixel753, pixel761,  
## pixel777, pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel393, pixel587, pixel588, pixel615, pixel616, pixel617, pixel643,  
## pixel697, pixel698, pixel724, pixel725, pixel726, pixel752, pixel753,  
## pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel393, pixel587, pixel588, pixel615, pixel616, pixel617, pixel643,  
## pixel697, pixel698, pixel724, pixel725, pixel726, pixel752, pixel753,  
## pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel393, pixel587, pixel588, pixel615, pixel616, pixel617, pixel643,  
## pixel697, pixel698, pixel724, pixel725, pixel726, pixel752, pixel753,  
## pixel777, pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel33, pixel34, pixel35, pixel50, pixel51, pixel59,  
## pixel110, pixel113, pixel167, pixel169, pixel224, pixel252, pixel280,  
## pixel308, pixel336, pixel364, pixel587, pixel588, pixel615, pixel616,  
## pixel617, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel33, pixel34, pixel35, pixel50, pixel51, pixel59,  
## pixel110, pixel113, pixel167, pixel169, pixel224, pixel252, pixel280,  
## pixel308, pixel336, pixel364, pixel587, pixel588, pixel615, pixel616,  
## pixel617, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel33, pixel34, pixel35, pixel50, pixel51, pixel59,  
## pixel110, pixel113, pixel167, pixel169, pixel224, pixel252, pixel280,  
## pixel308, pixel336, pixel364, pixel587, pixel588, pixel615, pixel616,  
## pixel617, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel777, pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel363,  
## pixel364, pixel365, pixel391, pixel393, pixel419, pixel587, pixel588,  
## pixel615, pixel616, pixel617, pixel643, pixel698, pixel724, pixel725,  
## pixel726, pixel752, pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel363,  
## pixel364, pixel365, pixel391, pixel393, pixel419, pixel587, pixel588,  
## pixel615, pixel616, pixel617, pixel643, pixel698, pixel724, pixel725,  
## pixel726, pixel752, pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel363,  
## pixel364, pixel365, pixel391, pixel393, pixel419, pixel587, pixel588,  
## pixel615, pixel616, pixel617, pixel643, pixel698, pixel724, pixel725,  
## pixel726, pixel752, pixel753, pixel777, pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel58, pixel59, pixel86, pixel110,  
## pixel113, pixel167, pixel169, pixel224, pixel252, pixel280, pixel308,  
## pixel336, pixel364, pixel477, pixel587, pixel588, pixel589, pixel615,  
## pixel616, pixel617, pixel643, pixel698, pixel724, pixel725, pixel726,  
## pixel752, pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel58, pixel59, pixel86, pixel110,  
## pixel113, pixel167, pixel169, pixel224, pixel252, pixel280, pixel308,  
## pixel336, pixel364, pixel477, pixel587, pixel588, pixel589, pixel615,  
## pixel616, pixel617, pixel643, pixel698, pixel724, pixel725, pixel726,  
## pixel752, pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel58, pixel59, pixel86, pixel110,  
## pixel113, pixel167, pixel169, pixel224, pixel252, pixel280, pixel308,  
## pixel336, pixel364, pixel477, pixel587, pixel588, pixel589, pixel615,  
## pixel616, pixel617, pixel643, pixel698, pixel724, pixel725, pixel726,  
## pixel752, pixel753, pixel777, pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel702, pixel724, pixel725, pixel726, pixel752, pixel753, pixel777,  
## pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel702, pixel724, pixel725, pixel726, pixel752, pixel753, pixel777,  
## pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel702, pixel724, pixel725, pixel726, pixel752, pixel753, pixel777,  
## pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel335, pixel336,  
## pixel364, pixel587, pixel588, pixel615, pixel616, pixel617, pixel643,  
## pixel698, pixel724, pixel725, pixel726, pixel752, pixel753, pixel777,  
## pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel335, pixel336,  
## pixel364, pixel587, pixel588, pixel615, pixel616, pixel617, pixel643,  
## pixel698, pixel724, pixel725, pixel726, pixel752, pixel753, pixel777,  
## pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel335, pixel336,  
## pixel364, pixel587, pixel588, pixel615, pixel616, pixel617, pixel643,  
## pixel698, pixel724, pixel725, pixel726, pixel752, pixel753, pixel777,  
## pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel49, pixel50, pixel51, pixel59, pixel110, pixel113,  
## pixel167, pixel169, pixel224, pixel252, pixel280, pixel308, pixel336,  
## pixel364, pixel447, pixel475, pixel504, pixel559, pixel587, pixel588,  
## pixel615, pixel616, pixel617, pixel642, pixel643, pixel698, pixel724,  
## pixel725, pixel726, pixel752, pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel49, pixel50, pixel51, pixel59, pixel110, pixel113,  
## pixel167, pixel169, pixel224, pixel252, pixel280, pixel308, pixel336,  
## pixel364, pixel447, pixel475, pixel504, pixel559, pixel587, pixel588,  
## pixel615, pixel616, pixel617, pixel642, pixel643, pixel698, pixel724,  
## pixel725, pixel726, pixel752, pixel753, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel49, pixel50, pixel51, pixel59, pixel110, pixel113,  
## pixel167, pixel169, pixel224, pixel252, pixel280, pixel308, pixel336,  
## pixel364, pixel447, pixel475, pixel504, pixel559, pixel587, pixel588,  
## pixel615, pixel616, pixel617, pixel642, pixel643, pixel698, pixel724,  
## pixel725, pixel726, pixel752, pixel753, pixel777, pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel60, pixel80, pixel81,  
## pixel110, pixel113, pixel167, pixel169, pixel224, pixel252, pixel280,  
## pixel308, pixel336, pixel364, pixel587, pixel588, pixel615, pixel616,  
## pixel617, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel761, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel60, pixel80, pixel81,  
## pixel110, pixel113, pixel167, pixel169, pixel224, pixel252, pixel280,  
## pixel308, pixel336, pixel364, pixel587, pixel588, pixel615, pixel616,  
## pixel617, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel761, pixel777, pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel60, pixel80, pixel81,  
## pixel110, pixel113, pixel167, pixel169, pixel224, pixel252, pixel280,  
## pixel308, pixel336, pixel364, pixel587, pixel588, pixel615, pixel616,  
## pixel617, pixel643, pixel698, pixel724, pixel725, pixel726, pixel752,  
## pixel753, pixel761, pixel777, pixel778, pixel779

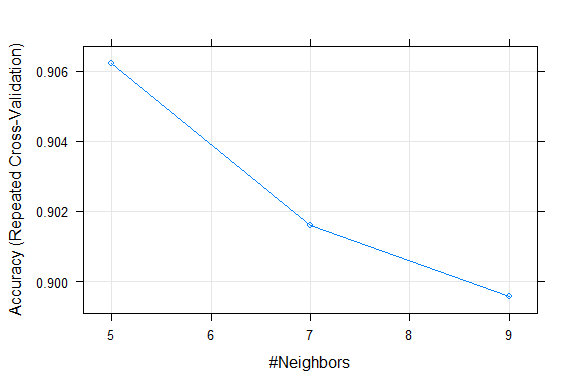
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel533, pixel587, pixel588, pixel615, pixel616, pixel617, pixel643,  
## pixel698, pixel724, pixel725, pixel726, pixel752, pixel753, pixel777,  
## pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel533, pixel587, pixel588, pixel615, pixel616, pixel617, pixel643,  
## pixel698, pixel724, pixel725, pixel726, pixel752, pixel753, pixel777,  
## pixel778, pixel779  
  
## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel533, pixel587, pixel588, pixel615, pixel616, pixel617, pixel643,  
## pixel698, pixel724, pixel725, pixel726, pixel752, pixel753, pixel777,  
## pixel778, pixel779

## Warning in preProcess.default(thresh = 0.95, k = 5, freqCut = 19, uniqueCut  
## = 10, : These variables have zero variances: pixel12, pixel13, pixel14,  
## pixel15, pixel32, pixel50, pixel51, pixel59, pixel110, pixel113, pixel167,  
## pixel169, pixel224, pixel252, pixel280, pixel308, pixel336, pixel364,  
## pixel587, pixel588, pixel615, pixel616, pixel617, pixel643, pixel698,  
## pixel724, pixel725, pixel726, pixel752, pixel753, pixel777, pixel778,  
## pixel779

# Model summary  
knnTrain

## k-Nearest Neighbors   
##   
## 8400 samples  
## 708 predictor  
## 10 classes: 'X0', 'X1', 'X2', 'X3', 'X4', 'X5', 'X6', 'X7', 'X8', 'X9'   
##   
## Pre-processing: centered (708), scaled (708)   
## Resampling: Cross-Validated (10 fold, repeated 3 times)   
## Summary of sample sizes: 7561, 7558, 7562, 7559, 7562, 7560, ...   
## Resampling results across tuning parameters:  
##   
## k Accuracy Kappa   
## 5 0.9062305 0.8957703  
## 7 0.9015868 0.8906054  
## 9 0.8995636 0.8883551  
##   
## Accuracy was used to select the optimal model using the largest value.  
## The final value used for the model was k = 5.

plot(knnTrain)



# Given these numbers, it appears that the model is extremely good at predicting the class label. Let's run the model on the test set.   
knnPred <- predict(knnTrain, testSet, type = 'prob')  
knnPred <- as.data.frame(knnPred)  
  
# Select maximum prediction.  
# We subtract 1 because we are showing the class label (i.e 0 is 1, 1 is 2, and so on.)  
knnPredictedValues <- data.frame(apply(knnPred, 1, which.max) - 1)   
colnames(knnPredictedValues) <- 'prediction'  
  
# Compare to actual values  
knnResults <- testSet %>% select(label) %>% mutate(real = str\_remove(label, 'X')) %>% bind\_cols(knnPredictedValues) %>% mutate(real = as.factor(real), prediction = as.factor(prediction))  
  
confusionMatrix(knnResults$real, knnResults$prediction)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1 2 3 4 5 6 7 8 9  
## 0 396 0 0 3 0 2 2 0 1 1  
## 1 0 476 1 0 1 0 0 0 0 0  
## 2 5 9 412 5 5 1 3 2 3 0  
## 3 1 7 8 368 0 12 1 7 4 2  
## 4 0 11 3 0 358 0 1 4 0 18  
## 5 2 6 5 17 1 319 11 2 6 12  
## 6 8 0 6 0 2 3 425 0 0 0  
## 7 1 11 2 1 9 0 0 380 2 18  
## 8 3 15 4 18 5 31 4 4 307 3  
## 9 1 5 2 7 19 1 0 29 2 358  
##   
## Overall Statistics  
##   
## Accuracy : 0.9045   
## 95% CI : (0.8952, 0.9132)  
## No Information Rate : 0.1286   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 0.8938   
## Mcnemar's Test P-Value : NA   
##   
## Statistics by Class:  
##   
## Class: 0 Class: 1 Class: 2 Class: 3 Class: 4 Class: 5  
## Sensitivity 0.94964 0.8815 0.9300 0.87828 0.89500 0.86450  
## Specificity 0.99762 0.9995 0.9912 0.98889 0.99026 0.98382  
## Pos Pred Value 0.97778 0.9958 0.9258 0.89756 0.90633 0.83727  
## Neg Pred Value 0.99447 0.9828 0.9917 0.98654 0.98896 0.98691  
## Prevalence 0.09929 0.1286 0.1055 0.09976 0.09524 0.08786  
## Detection Rate 0.09429 0.1133 0.0981 0.08762 0.08524 0.07595  
## Detection Prevalence 0.09643 0.1138 0.1060 0.09762 0.09405 0.09071  
## Balanced Accuracy 0.97363 0.9405 0.9606 0.93359 0.94263 0.92416  
## Class: 6 Class: 7 Class: 8 Class: 9  
## Sensitivity 0.9508 0.88785 0.94462 0.86893  
## Specificity 0.9949 0.98834 0.97755 0.98258  
## Pos Pred Value 0.9572 0.89623 0.77919 0.84434  
## Neg Pred Value 0.9941 0.98729 0.99527 0.98570  
## Prevalence 0.1064 0.10190 0.07738 0.09810  
## Detection Rate 0.1012 0.09048 0.07310 0.08524  
## Detection Prevalence 0.1057 0.10095 0.09381 0.10095  
## Balanced Accuracy 0.9729 0.93809 0.96108 0.92575

The model takes a long time to run - training this model took almost two hours! - but the end results show that it is highly accurate when predicting class labels. We can see that for both the training and test sets, the accuracy is slightly over 90 percent which, so far, are the best results we have seen.

## Support Vector Machines

Now, our attention will turn to SVMs. Since the data has already been processed, we will train the model.

# Run the model using a linear kernel to classify the data. We'll do 3 k-fold cross validatiion.   
svmTrain <- svm(label ~ ., data = trainSet, type = 'C', kernel = 'linear', cross = 3, probability = TRUE)

## Warning in svm.default(x, y, scale = scale, ..., na.action = na.action):  
## Variable(s) 'pixel12' and 'pixel13' and 'pixel14' and 'pixel15' and  
## 'pixel32' and 'pixel50' and 'pixel51' and 'pixel59' and 'pixel110' and  
## 'pixel113' and 'pixel167' and 'pixel169' and 'pixel224' and 'pixel252' and  
## 'pixel280' and 'pixel308' and 'pixel336' and 'pixel364' and 'pixel587' and  
## 'pixel588' and 'pixel615' and 'pixel616' and 'pixel617' and 'pixel643' and  
## 'pixel698' and 'pixel724' and 'pixel725' and 'pixel726' and 'pixel752' and  
## 'pixel753' and 'pixel777' and 'pixel778' and 'pixel779' constant. Cannot  
## scale data.

summary(svmTrain)

##   
## Call:  
## svm(formula = label ~ ., data = trainSet, type = "C", kernel = "linear",   
## cross = 3, probability = TRUE)  
##   
##   
## Parameters:  
## SVM-Type: C-classification   
## SVM-Kernel: linear   
## cost: 1   
## gamma: 0.001412429   
##   
## Number of Support Vectors: 2365  
##   
## ( 286 221 128 235 184 283 249 287 320 172 )  
##   
##   
## Number of Classes: 10   
##   
## Levels:   
## X0 X1 X2 X3 X4 X5 X6 X7 X8 X9  
##   
## 3-fold cross-validation on training data:  
##   
## Total Accuracy: 91.33333   
## Single Accuracies:  
## 91.32143 91.28571 91.39286

These results are really amazing! SVM has a higher degree of accuracy within the training set than what kNN showed originally. Now to run the test set on the model.

svmPred <- predict(svmTrain, testSet, type = 'prob')  
svmPred <- as.data.frame(svmPred)  
colnames(svmPred) <- 'results'  
  
# Build the data frame to compare the results.   
svmResults <- testSet %>% select(label) %>% bind\_cols(svmPred) %>% mutate(real = factor(as.character(str\_remove(label, 'X'))), prediction = factor(as.character(str\_remove(results, 'X'))))  
  
confusionMatrix(svmResults$real, svmResults$prediction)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1 2 3 4 5 6 7 8 9  
## 0 392 0 2 0 1 4 4 1 1 0  
## 1 0 469 0 1 1 1 0 1 5 0  
## 2 1 3 408 3 8 1 5 4 12 0  
## 3 1 5 10 351 1 22 2 4 10 4  
## 4 1 3 1 0 369 0 2 3 1 15  
## 5 5 2 4 21 5 319 8 1 12 4  
## 6 3 0 4 0 8 4 419 0 6 0  
## 7 2 3 6 2 6 1 0 387 2 15  
## 8 4 11 8 9 1 16 4 3 333 5  
## 9 0 3 2 6 16 2 0 23 6 366  
##   
## Overall Statistics  
##   
## Accuracy : 0.9079   
## 95% CI : (0.8987, 0.9164)  
## No Information Rate : 0.1188   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 0.8976   
## Mcnemar's Test P-Value : NA   
##   
## Statistics by Class:  
##   
## Class: 0 Class: 1 Class: 2 Class: 3 Class: 4 Class: 5  
## Sensitivity 0.95844 0.9399 0.91685 0.89313 0.88702 0.86216  
## Specificity 0.99657 0.9976 0.99015 0.98450 0.99313 0.98381  
## Pos Pred Value 0.96790 0.9812 0.91685 0.85610 0.93418 0.83727  
## Neg Pred Value 0.99552 0.9919 0.99015 0.98892 0.98765 0.98665  
## Prevalence 0.09738 0.1188 0.10595 0.09357 0.09905 0.08810  
## Detection Rate 0.09333 0.1117 0.09714 0.08357 0.08786 0.07595  
## Detection Prevalence 0.09643 0.1138 0.10595 0.09762 0.09405 0.09071  
## Balanced Accuracy 0.97750 0.9687 0.95350 0.93882 0.94007 0.92299  
## Class: 6 Class: 7 Class: 8 Class: 9  
## Sensitivity 0.94369 0.90632 0.85825 0.89487  
## Specificity 0.99334 0.99019 0.98400 0.98470  
## Pos Pred Value 0.94369 0.91274 0.84518 0.86321  
## Neg Pred Value 0.99334 0.98941 0.98555 0.98861  
## Prevalence 0.10571 0.10167 0.09238 0.09738  
## Detection Rate 0.09976 0.09214 0.07929 0.08714  
## Detection Prevalence 0.10571 0.10095 0.09381 0.10095  
## Balanced Accuracy 0.96852 0.94826 0.92112 0.93978

The model’s consistency is very impressive. It manages to also predict at a high accuracy.

## Random Forest

Let’s turn to the final model that we will work on: random forest.

x <- trainControl(method = 'repeatedcv', number = 3, repeats = 3)  
rfTrain <- train(label ~ ., data = trainSet, method = 'rf', metric = 'Accuracy', trControl = x, type = 'C')  
  
rfTrain

## Random Forest   
##   
## 8400 samples  
## 708 predictor  
## 10 classes: 'X0', 'X1', 'X2', 'X3', 'X4', 'X5', 'X6', 'X7', 'X8', 'X9'   
##   
## No pre-processing  
## Resampling: Cross-Validated (3 fold, repeated 3 times)   
## Summary of sample sizes: 5601, 5599, 5600, 5600, 5601, 5599, ...   
## Resampling results across tuning parameters:  
##   
## mtry Accuracy Kappa   
## 2 0.8767850 0.8629762  
## 37 0.9452385 0.9391393  
## 707 0.9253181 0.9170037  
##   
## Accuracy was used to select the optimal model using the largest value.  
## The final value used for the model was mtry = 37.

# With these results, we see that the train has a minimum accuracy on two variables sampled, but this increases once we reach 37.

With 94.6 percent accuracy, this is by far the superior model when comparing the three models we’ve built and the results according to the training set. But the most important thing is to compare the results of the testing set.

# Test the random forest model.   
rfPred <- predict(rfTrain, testSet, type = 'prob')  
rfPred <- as.data.frame(rfPred)  
  
rfPredictedValues <- data.frame(apply(rfPred, 1, which.max) - 1)   
colnames(rfPredictedValues) <- 'results'  
  
# Random forest actual and predicted labels, and confusion matrix to comparte the model's prediciton capabilities.  
rfResults <- testSet %>% select(label) %>% bind\_cols(rfPredictedValues) %>% mutate(real = factor(as.character(str\_remove(label, 'X'))), prediction = factor(results))  
  
confusionMatrix(rfResults$real, rfResults$prediction)

## Confusion Matrix and Statistics  
##   
## Reference  
## Prediction 0 1 2 3 4 5 6 7 8 9  
## 0 398 0 1 0 0 1 2 0 3 0  
## 1 0 470 2 1 1 2 1 0 0 1  
## 2 1 1 424 1 6 0 3 3 4 2  
## 3 2 6 12 368 1 8 0 4 6 3  
## 4 0 1 2 0 379 0 3 0 0 10  
## 5 4 1 1 9 0 359 6 0 1 0  
## 6 4 1 0 0 1 2 432 1 3 0  
## 7 0 2 6 2 2 0 1 403 2 6  
## 8 2 7 0 8 2 5 5 1 359 5  
## 9 3 2 0 9 7 4 1 12 5 381  
##   
## Overall Statistics  
##   
## Accuracy : 0.946   
## 95% CI : (0.9387, 0.9526)  
## No Information Rate : 0.1169   
## P-Value [Acc > NIR] : < 2.2e-16   
##   
## Kappa : 0.9399   
## Mcnemar's Test P-Value : NA   
##   
## Statistics by Class:  
##   
## Class: 0 Class: 1 Class: 2 Class: 3 Class: 4 Class: 5  
## Sensitivity 0.96135 0.9572 0.9464 0.92462 0.94987 0.94226  
## Specificity 0.99815 0.9978 0.9944 0.98895 0.99579 0.99424  
## Pos Pred Value 0.98272 0.9833 0.9528 0.89756 0.95949 0.94226  
## Neg Pred Value 0.99578 0.9944 0.9936 0.99208 0.99474 0.99424  
## Prevalence 0.09857 0.1169 0.1067 0.09476 0.09500 0.09071  
## Detection Rate 0.09476 0.1119 0.1010 0.08762 0.09024 0.08548  
## Detection Prevalence 0.09643 0.1138 0.1060 0.09762 0.09405 0.09071  
## Balanced Accuracy 0.97975 0.9775 0.9704 0.95679 0.97283 0.96825  
## Class: 6 Class: 7 Class: 8 Class: 9  
## Sensitivity 0.9515 0.95047 0.93734 0.93382  
## Specificity 0.9968 0.99444 0.99083 0.98866  
## Pos Pred Value 0.9730 0.95047 0.91117 0.89858  
## Neg Pred Value 0.9941 0.99444 0.99369 0.99285  
## Prevalence 0.1081 0.10095 0.09119 0.09714  
## Detection Rate 0.1029 0.09595 0.08548 0.09071  
## Detection Prevalence 0.1057 0.10095 0.09381 0.10095  
## Balanced Accuracy 0.9742 0.97246 0.96408 0.96124

So results are on par with what was expected: 94.7 percent accuracy, by far the best out of these three models.

# Final comments

All three models tested this week present high degrees of accuracy: all of them are accurate over 90 percent of the time. However, there are some caveats to bear in mind. First, the size of the data is immense, and so, we had to cut corners. We dismissed using the original test set as it did not have class labels, which we would’ve only been able to actually test in Kaggle. Because of this, we worked only on the training dataset. However, because this dataset was over 40,000 records long, we also limited working with 30 percent of this - with all observations chosen at random. Since this was done prior to any splitting into a new training and testing set, there is no chance for observations to appear in both sets. What is concerning is that maybe this would probably cause the models to overfit because of the limited size of the training sample. This could be ammended by using a larger set and adding more computing power.

Despite all this, these initial results are satisfying and merit additional investigation towards finer parameter tuning and slowly add more observations to the training set to produce a better model.