

```

# Implementing ML techniques in R

library(caret)

dataset <- iris

sapply(dataset, class)

levels(dataset$Species)

# Can we build a model with ML algorithms and use that model to do
# predictions? From validation data, using model and first 4 columns can
# we predict the factor/species?

# 80% split for training data and 20% split for validation data
validation_index <- createDataPartition(dataset$Species, p=0.80,
list=FALSE)

validation <- dataset[-validation_index,]
dataset <- dataset[validation_index,]

# cross validations: 3, 5, 10-fold cross validations

control <- trainControl(method="cv", number=10)
metric <- "Accuracy"

# Linear Discriminant Analysis (LDA), Classification and regression
# trees (CART), k-Nearest Neighbors (kNN), Support Vector Machines
# (SVM), Random Forest (RF).

set.seed(7)
fit.lda <- train(Species~., data=dataset, method="lda", metric=metric,
trControl= control)

set.seed(7)
fit.cart <- train(Species~., data=dataset, method="rpart",
metric=metric, trControl= control)

set.seed(7)
fit.knn <- train(Species~., data=dataset, method="knn", metric=metric,
trControl= control)

set.seed(7)
fit.SVM <- train(Species~., data=dataset, method="svmRadial",
metric=metric, trControl= control)

set.seed(7)
fit.rf <- train(Species~., data=dataset, method="rf", metric=metric,
trControl= control)

```

```
# Summarizing the accuracy of these 5 models
results <- resamples(list(lda=fit.lda, cart=fit.cart, knn=fit.knn,
svm=fit.SVM, rf=fit.rf))

summary(results)

# Based upon the accuracy values we choose lda model to do our
predictions on the validation dataset

predictions <- predict(fit.lda, validation)

confusionMatrix(predictions, validation$Species)
```