

```

> head(titanic_clean)
  Survived   Sex Pclass Age  Fare SibSp Parch FamilySize Embarked
1         0  male     3  22  7.25     1     0         2         S
2         1 female     1  38 71.28     1     0         2         C
3         1 female     3  26  7.92     0     0         1         S
4         1 female     1  35 53.10     1     0         2         S
5         0  male     3  35  8.05     0     0         1         S
6         0  male     3  28  8.46     0     0         1         Q
>
>
>
> train_lda <- train(Survived ~ Fare, method = "lda", data = train_set)
> lda_preds <- predict(train_lda, test_set)
> mean(lda_preds == test_set$Survived)
[1] 0.693
> train_qda <- train(Survived ~ Fare, method = "qda", data = train_set)
> qda_preds <- predict(train_qda, test_set)
> mean(qda_preds == test_set$Survived)
[1] 0.693
>
>
>
> train_glm <- train(Survived ~ Age, method = "glm", data = train_set)
> glm_preds <- predict(train_glm, test_set)
> mean(glm_preds == test_set$Survived)
[1] 0.615
> train_glm <- train(Survived ~ Sex + Pclass + Fare + Age , method = "glm", data = train_set)
> glm_preds <- predict(train_glm, test_set)
> mean(glm_preds == test_set$Survived)
[1] 0.849
> train_glm <- train(Survived ~ ., method = "glm", data = train_set)
There were 25 warnings (use warnings() to see them)
> glm_preds <- predict(train_glm, test_set)
Warning message:
In predict.lm(object, newdata, se.fit, scale = 1, type = if (type == :
prediction from a rank-deficient fit may be misleading
> mean(glm_preds == test_set$Survived)
[1] 0.849
>
>
>
> set.seed(6, sample.kind = 'Rounding')
Warning message:
In set.seed(6, sample.kind = "Rounding") :
non-uniform 'Rounding' sampler used
> train_knn <- train(Survived ~ ., method = "knn", data = train_set, tuneGrid = data.frame(k = seq(3, 51,
2)))
> train_knn$bestTune
      k
5 11
>
>
>
> plot(train_knn)
>
>
>
> knn_preds <- predict(train_knn, test_set)
> mean(knn_preds == test_set$Survived)
[1] 0.709
>
>
>
> set.seed(8, sample.kind = 'Rounding')
Warning message:

```

```

In set.seed(8, sample.kind = "Rounding") :
  non-uniform 'Rounding' sampler used
> control <- trainControl(method = "cv", number = 10, p = .9)
> train_knn_cv <- train(Survived ~ ., method = 'knn', data = train_set, tuneGrid = data.frame(k = seq(3, 5
1, 2)), trControl = control)
> train_knn_cv$bestTune
  k
2 5
> knn_cv_preds <- predict(train_knn_cv, test_set)
> mean(knn_cv_preds == test_set$Survived)
[1] 0.648
>
>
>
> suppressWarnings(set.seed(10, sample.kind = 'Rounding'))
> train_rpart <- train(Survived ~ ., method = 'rpart', tuneGrid = data.frame(cp = seq(0, 0.05, 0.002)), da
ta = train_set)
> train_rpart$bestTune
  cp
9 0.016
> rpart_preds <- predict(train_rpart, test_set)
> mean(rpart_preds == test_set$Survived)
[1] 0.838
>
>
>
> train_rpart$finalModel
n= 712

```

```

node), split, n, loss, yval, (yprob)
  * denotes terminal node

```

```

1) root 712 273 0 (0.6166 0.3834)
 2) Sexmale>=0.5 463 91 0 (0.8035 0.1965)
   4) Age>=3.5 449 80 0 (0.8218 0.1782) *
   5) Age< 3.5 14 3 1 (0.2143 0.7857) *
 3) Sexmale< 0.5 249 67 1 (0.2691 0.7309)
   6) Pclass>=2.5 118 59 0 (0.5000 0.5000)
      12) Fare>=23.4 24 3 0 (0.8750 0.1250) *
      13) Fare< 23.4 94 38 1 (0.4043 0.5957) *
      7) Pclass< 2.5 131 8 1 (0.0611 0.9389) *
> plot(train_rpart$finalModel, margin = 0.1)
> text(train_rpart$finalModel)
>
>
>
> suppressWarnings(set.seed(14, sample.kind = "Rounding"))
> train_rf <- train(Survived ~ ., data = train_set, method = "rf", ntree = 100, tuneGrid = data.frame(mtry
= seq(1:7)))
> train_rf$bestTune
  mtry
2 2
> rf_preds <- predict(train_rf, test_set)
> mean(rf_preds == test_set$Survived)
[1] 0.844
> varImp(train_rf)
rf variable importance

```

	Overall
Sexmale	100.000
Fare	65.091
Age	45.533
Pclass	32.529
FamilySize	18.275
SibSp	7.881

Parch	7.150
EmbarkedS	2.839
EmbarkedQ	0.122
EmbarkedC	0.000

>