```
> getwd()
[1] "/Users/icaptain/Desktop/Rlang"
> iris_data<-iris
> View(iris_data)
> head(iris_data)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
> summary(iris_data)
  Sepal.Length    Sepal.Width     Petal.Length    Petal.Width          Species
 Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100   setosa    :50
 1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300   versicolor:50
 Median :5.800   Median :3.000   Median :4.350   Median :1.300   virginica :50
 Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
 3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
 Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
> dim(iris_data)
[1] 150   5
> split=sample.split(iris_data,SplitRatio = 0.80)
> train_data=subset(iris_data,split==TRUE)
> test_data=subset(iris_data,split==FALSE)
> train_scale=scale(train_data[,1:4])
> test_scale=scale(test_data[,1:4])
> dim(train_data)
[1] 120   5
> dim(train_scale)
[1] 120   4
> dim(test_scale)
[1] 30  4
> set.seed(1)
> library(naivebayes)
> classifier_naive=naive_bayes(Species~.,data = train_data)
-------------------------------------------------------------------------------
:: Petal.Width (Gaussian)
-------------------------------------------------------------------------------

Petal.Width     setosa versicolor  virginica
       mean 0.24000000 1.32750000 2.04000000
       sd   0.09001424 0.20253205 0.27623847

-------------------------------------------------------------------------------

>
```

```
> classifier_naive

========================================= Naive Bayes =========================

Call:
naive_bayes.formula(formula = Species ~ ., data = train_data)

-------------------------------------------------------------------------------

Laplace smoothing: 0

-------------------------------------------------------------------------------

A priori probabilities:

    setosa versicolor  virginica
 0.3333333  0.3333333  0.3333333

-------------------------------------------------------------------------------

Tables:

-------------------------------------------------------------------------------
:: Sepal.Length (Gaussian)
-------------------------------------------------------------------------------

Sepal.Length    setosa versicolor virginica
        mean 5.0200000  5.9650000 6.6250000
        sd   0.3131314  0.5206259 0.6581832

-------------------------------------------------------------------------------
:: Sepal.Width (Gaussian)
-------------------------------------------------------------------------------

Sepal.Width    setosa versicolor virginica
       mean 3.4375000  2.8000000 3.0075000
       sd   0.3739515  0.3137858 0.3237501

-------------------------------------------------------------------------------
:: Petal.Length (Gaussian)
-------------------------------------------------------------------------------

Petal.Length    setosa versicolor virginica
        mean 1.4625000  4.2625000 5.5650000
        sd   0.1734824  0.4204317 0.5404414
```

```
> model <- randomForest(Survived ~ ., data = trainingset, na.action = na.exclude)
Warning message:
In randomForest.default(m, y, ...) :
  The response has five or fewer unique values.  Are you sure you want to do regression?
> unique(titanic$Survived)
[1] 0 1
> summary(model)
                Length Class  Mode
call              4    -none- call
type              1    -none- character
predicted       713    -none- numeric
mse             500    -none- numeric
rsq             500    -none- numeric
oob.times       713    -none- numeric
importance        6    -none- numeric
importanceSD      0    -none- NULL
localImportance   0    -none- NULL
proximity         0    -none- NULL
ntree             1    -none- numeric
mtry              1    -none- numeric
forest           11    -none- list
coefs             0    -none- NULL
y               713    -none- numeric
test              0    -none- NULL
inbag             0    -none- NULL
terms             3    terms  call
> View(model)
> testset$prediction=predict(model,testset)
> View(testset)
> testset$binary=ifelse(testset$prediction>0.5,1,0)
> cor(testset$Survived,testset$binary)
[1] 0.6128629
> table(testset$Survived,testset$binary)

    0  1
  0 99 11
  1 21 47
```

```
> getwd()
[1] "/Users/icaptain/Desktop/Rlang/modeling_"
Warning message:
R graphics engine version 15 is not supported by this version of RStudio. The Plots tab will be disabled until a newer
 version of RStudio is installed.
> titanic=read.csv("titanic.csv",na.strings = "")
> titanic=titanic[-c(1,3,4,9,11)]
> table(titanic$Survived)

  0   1
549 342
> table(titanic$Sex)

female   male
   314    577
> table(titanic$Embarked)

  C   Q   S
168  77 644
> titanic$Age[is.na(titanic$Age)]=mean(titanic$Age,na.rm = TRUE)
> titanic$Embarked[is.na(titanic$Embarked)]="S"
> library(caTools)
> sample=sample.split(titanic$Survived,SplitRatio = 0.80)
> trainingset=subset(titanic,sample==TRUE)
> testset=subset(titanic,sample==FALSE)
> install.packages("randomForest")
trying URL 'https://cran.rstudio.com/bin/macosx/contrib/4.2/randomForest_4.7-1.1.tgz'
Content type 'application/x-gzip' length 256701 bytes (250 KB)
==================================================
downloaded 250 KB


The downloaded binary packages are in
        /var/folders/h8/qvvtq7xd2qq19lpjk1nf_sw40000gn/T//Rtmpt87zaC/downloaded_packages
> library(randomForest)
randomForest 4.7-1.1
Type rfNews() to see new features/changes/bug fixes.
```