

Previendo a Ocorrência de Câncer - Usando algoritmo Knn

Evanil Tiengo Junior

Descrição

Este projeto é parte integrante do curso Big Data Analytics com R e Microsoft Azure da Formação Cientista de Dados. O objetivo é analisar dados reais sobre exames de câncer de mama realizado com mulheres nos EUA e então prever a ocorrência de novos casos. Os dados do câncer da mama incluem 569 observações de biópsias de câncer, cada um com 32 características (variáveis). O diagnóstico é codificado como "M" para indicar maligno ou "B" para indicar benigno. Todo o projeto será descrito de acordo com suas etapas.

Objetivos

O objetivo será prever a ocorrência de câncer. O modelo preditivo deve ser bastante preciso, pois estamos tratando de doença.

Local de armazenamento e pacotes instalados

```
# Local armazenamento
setwd("~/Mini-Projeto02")
getwd()

# Pacotes utilizados.
#install.packages("class")
library(class)
#install.packages("caTools")
library(caTools)
# install.packages("gmodels")
library(gmodels)
#install.packages("ggplot2")
library(ggplot2)
#install.packages("caret")
library(caret)

## Loading required package: lattice
```

Etapa 1 - Coletando os Dados

Os dados foram fornecidos pela DSA. Sendo assim preciso realizar a carga. Antes de realizar a carga do arquivo é necessário saber o formato do mesmo, que neste caso é .csv!

```
Dados <- read.csv("C:/Users/evanil.tiengo/Desktop/Cursos/DSA/Big Data Analytics com R e Microsoft Azure Machine Learning/Mini-Projeto02/bc_data.csv", stringsAsFactors = FALSE)
```

Etapa 2 - EDA (Exploratory Data Analysis)

Com o comando abaixo é possível identificar as classes de cada variável
str(Dados)

```
## 'data.frame':    569 obs. of  32 variables:
## $ id           : int  87139402 8910251 905520 868871 9012568 9065
39 925291 87880 862989 89827 ...
## $ diagnosis    : chr  "B" "B" "B" "B" ...
## $ radius_mean  : num  12.3 10.6 11 11.3 15.2 ...
## $ texture_mean : num  12.4 18.9 16.8 13.4 13.2 ...
## $ perimeter_mean : num  78.8 69.3 70.9 73 97.7 ...
## $ area_mean    : num  464 346 373 385 712 ...
## $ smoothness_mean : num  0.1028 0.0969 0.1077 0.1164 0.0796 ...
## $ compactness_mean : num  0.0698 0.1147 0.078 0.1136 0.0693 ...
## $ concavity_mean : num  0.0399 0.0639 0.0305 0.0464 0.0339 ...
## $ points_mean   : num  0.037 0.0264 0.0248 0.048 0.0266 ...
## $ symmetry_mean  : num  0.196 0.192 0.171 0.177 0.172 ...
## $ dimension_mean : num  0.0595 0.0649 0.0634 0.0607 0.0554 ...
## $ radius_se     : num  0.236 0.451 0.197 0.338 0.178 ...
## $ texture_se    : num  0.666 1.197 1.387 1.343 0.412 ...
## $ perimeter_se  : num  1.67 3.43 1.34 1.85 1.34 ...
## $ area_se       : num  17.4 27.1 13.5 26.3 17.7 ...
## $ smoothness_se : num  0.00805 0.00747 0.00516 0.01127 0.00501 ...
## $ compactness_se : num  0.0118 0.03581 0.00936 0.03498 0.01485 ...
## $ concavity_se  : num  0.0168 0.0335 0.0106 0.0219 0.0155 ...
## $ points_se     : num  0.01241 0.01365 0.00748 0.01965 0.00915 ...
## $ symmetry_se   : num  0.0192 0.035 0.0172 0.0158 0.0165 ...
## $ dimension_se  : num  0.00225 0.00332 0.0022 0.00344 0.00177 ...
## $ radius_worst  : num  13.5 11.9 12.4 11.9 16.2 ...
## $ texture_worst : num  15.6 22.9 26.4 15.8 15.7 ...
## $ perimeter_worst : num  87 78.3 79.9 76.5 104.5 ...
## $ area_worst    : num  549 425 471 434 819 ...
## $ smoothness_worst : num  0.139 0.121 0.137 0.137 0.113 ...
## $ compactness_worst : num  0.127 0.252 0.148 0.182 0.174 ...
## $ concavity_worst : num  0.1242 0.1916 0.1067 0.0867 0.1362 ...
## $ points_worst  : num  0.0939 0.0793 0.0743 0.0861 0.0818 ...
## $ symmetry_worst : num  0.283 0.294 0.3 0.21 0.249 ...
## $ dimension_worst : num  0.0677 0.0759 0.0788 0.0678 0.0677 ...
```

Temos 569 observações (linhas) de biópsias de câncer, cada um com 32 variáveis (colunas). Uma característica é um número de identificação (ID), outro é o diagnóstico de câncer, e 30 são medidas laboratoriais numéricas. O diagnóstico é codificado como “M” para indicar maligno ou “B” para indicar benigno.

Exploração do dataset:

Independentemente do método de aprendizagem de máquina, deve sempre ser excluídas variáveis de ID. Caso contrário, isso pode levar a resultados errados porque o ID pode ser usado para “prever” cada exemplo. Por conseguinte, um modelo que inclui um identificador que sofre de superajuste, e é improvável que generalize bem a outros dados.

```
# Excluindo a coluna ID
```

```
Dados <- Dados[-1]
```

```
str(Dados)
```

```
## 'data.frame': 569 obs. of 31 variables:
## $ diagnosis : chr "B" "B" "B" "B" ...
## $ radius_mean : num 12.3 10.6 11 11.3 15.2 ...
## $ texture_mean : num 12.4 18.9 16.8 13.4 13.2 ...
## $ perimeter_mean : num 78.8 69.3 70.9 73 97.7 ...
## $ area_mean : num 464 346 373 385 712 ...
## $ smoothness_mean : num 0.1028 0.0969 0.1077 0.1164 0.0796 ...
## $ compactness_mean : num 0.0698 0.1147 0.078 0.1136 0.0693 ...
## $ concavity_mean : num 0.0399 0.0639 0.0305 0.0464 0.0339 ...
## $ points_mean : num 0.037 0.0264 0.0248 0.048 0.0266 ...
## $ symmetry_mean : num 0.196 0.192 0.171 0.177 0.172 ...
## $ dimension_mean : num 0.0595 0.0649 0.0634 0.0607 0.0554 ...
## $ radius_se : num 0.236 0.451 0.197 0.338 0.178 ...
## $ texture_se : num 0.666 1.197 1.387 1.343 0.412 ...
## $ perimeter_se : num 1.67 3.43 1.34 1.85 1.34 ...
## $ area_se : num 17.4 27.1 13.5 26.3 17.7 ...
## $ smoothness_se : num 0.00805 0.00747 0.00516 0.01127 0.00501 ...
## $ compactness_se : num 0.0118 0.03581 0.00936 0.03498 0.01485 ...
## $ concavity_se : num 0.0168 0.0335 0.0106 0.0219 0.0155 ...
## $ points_se : num 0.01241 0.01365 0.00748 0.01965 0.00915 ...
## $ symmetry_se : num 0.0192 0.035 0.0172 0.0158 0.0165 ...
## $ dimension_se : num 0.00225 0.00332 0.0022 0.00344 0.00177 ...
## $ radius_worst : num 13.5 11.9 12.4 11.9 16.2 ...
## $ texture_worst : num 15.6 22.9 26.4 15.8 15.7 ...
## $ perimeter_worst : num 87 78.3 79.9 76.5 104.5 ...
## $ area_worst : num 549 425 471 434 819 ...
## $ smoothness_worst : num 0.139 0.121 0.137 0.137 0.113 ...
## $ compactness_worst : num 0.127 0.252 0.148 0.182 0.174 ...
## $ concavity_worst : num 0.1242 0.1916 0.1067 0.0867 0.1362 ...
## $ points_worst : num 0.0939 0.0793 0.0743 0.0861 0.0818 ...
## $ symmetry_worst : num 0.283 0.294 0.3 0.21 0.249 ...
## $ dimension_worst : num 0.0677 0.0759 0.0788 0.0678 0.0677 ...
```

```
# Identificação de NA's e Vazios!
```

```
any(is.na(Dados))
```

```
## [1] FALSE
```

```
any(Dados == "")
```

```
## [1] FALSE
```

False significa que não existe nenhum campo com NA ou Vazio!

Analizando as variáveis:

#\$diagnosis: Muitos classificadores requerem que as variáveis sejam do tipo Fator.

No nosso dataset a variável está como caracter. Temos que converter-la

```
table(Dados$diagnosis)
```

```
##
```

```
## B M
```

```
## 357 212
```

```
Dados$diagnosis <- factor(Dados$diagnosis, levels = c("B", "M"), labels = c("Benigno", "Maligno"))
```

```
str(Dados$diagnosis)
```

```
## Factor w/ 2 levels "Benigno","Maligno": 1 1 1 1 1 1 1 2 1 1 ...
```

Verificando a proporção

```
round(prop.table(table(Dados$diagnosis)) * 100, digits = 1)
```

```
##
```

```
## Benigno Maligno
```

```
## 62.7 37.3
```

Pelo dataset temos que 62.7% dos casos analisados são do tipo Benigno e 37.3% são do tipo Maligno.

As demais medidas são numéricas. Elas são os resultados dos exames da biopsia, como por exemplo o raio, a textura, o perímetro, a área, a concavidade, etc...

Como temos várias medidas numéricas e de diferentes exames, podemos ter um problema de normalização. O propósito da normalização é minimizar os problemas oriundos do uso de unidades e dispersões distintas entre as variáveis. No nosso dataset podemos detectar esse problema de escala entre os dados. Então precisam ser normalizados!

Segue abaixo 6 variáveis que demonstrarão a necessidade da normalização:

```
summary(Dados[c("texture_mean", "perimeter_mean", "area_mean", "concavity_mean", "smoothness_se", "perimeter_worst")])
```

##	texture_mean	perimeter_mean	area_mean	concavity_mean
## Min.	: 9.71	Min. : 43.79	Min. : 143.5	Min. :0.00000
## 1st Qu.:	:16.17	1st Qu.: 75.17	1st Qu.: 420.3	1st Qu.:0.02956
## Median	:18.84	Median : 86.24	Median : 551.1	Median :0.06154
## Mean	:19.29	Mean : 91.97	Mean : 654.9	Mean :0.08880
## 3rd Qu.:	:21.80	3rd Qu.:104.10	3rd Qu.: 782.7	3rd Qu.:0.13070
## Max.	:39.28	Max. :188.50	Max. :2501.0	Max. :0.42680

```
## smoothness_se      perimeter_worst
## Min.      :0.001713   Min.       : 50.41
## 1st Qu.:0.005169   1st Qu.: 84.11
## Median :0.006380   Median : 97.66
## Mean      :0.007041   Mean      :107.26
## 3rd Qu.:0.008146   3rd Qu.:125.40
## Max.      :0.031130   Max.      :251.20
```

Iremos utilizar a normalização segundo a amplitude. O cálculo de distância pelo Knn é dependente das medidas de escala nos dados de entrada!

```
# Criando um função de normalização
Normalizar <- function(x) {
  return ((x - min(x)) / (max(x) - min(x)))
}

# Testando a função de normalização - os resultados devem ser idênticos
Normalizar(c(1, 2, 3, 4, 5))

## [1] 0.00 0.25 0.50 0.75 1.00

Normalizar(c(10, 20, 30, 40, 50))

## [1] 0.00 0.25 0.50 0.75 1.00
```

Normalização ok!

```
# Normalizando os dados
Dados_Norm <- as.data.frame(lapply(Dados[2:31], Normalizar))

# Confirmando que a normalização funcionou
summary(Dados[c("texture_mean", "perimeter_mean", "area_mean", "concavity_mean", "smoothness_se", "perimeter_worst")])

## texture_mean      perimeter_mean      area_mean      concavity_mean
## Min.      : 9.71      Min.      : 43.79      Min.      : 143.5      Min.      :0.00000
## 1st Qu.:16.17      1st Qu.: 75.17      1st Qu.: 420.3      1st Qu.:0.02956
## Median :18.84      Median : 86.24      Median : 551.1      Median :0.06154
## Mean      :19.29      Mean      : 91.97      Mean      : 654.9      Mean      :0.08880
## 3rd Qu.:21.80      3rd Qu.:104.10      3rd Qu.: 782.7      3rd Qu.:0.13070
## Max.      :39.28      Max.      :188.50      Max.      :2501.0      Max.      :0.42680
## smoothness_se      perimeter_worst
## Min.      :0.001713   Min.       : 50.41
## 1st Qu.:0.005169   1st Qu.: 84.11
## Median :0.006380   Median : 97.66
## Mean      :0.007041   Mean      :107.26
## 3rd Qu.:0.008146   3rd Qu.:125.40
## Max.      :0.031130   Max.      :251.20

summary(Dados_Norm[c("texture_mean", "perimeter_mean", "area_mean", "concavity_mean", "smoothness_se", "perimeter_worst")])
```

```
## texture_mean perimeter_mean area_mean concavity_mean
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.00000
## 1st Qu.:0.2185 1st Qu.:0.2168 1st Qu.:0.1174 1st Qu.:0.06926
## Median :0.3088 Median :0.2933 Median :0.1729 Median :0.14419
## Mean :0.3240 Mean :0.3329 Mean :0.2169 Mean :0.20806
## 3rd Qu.:0.4089 3rd Qu.:0.4168 3rd Qu.:0.2711 3rd Qu.:0.30623
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.00000
## smoothness_se perimeter_worst
## Min. :0.0000 Min. :0.0000
## 1st Qu.:0.1175 1st Qu.:0.1678
## Median :0.1586 Median :0.2353
## Mean :0.1811 Mean :0.2831
## 3rd Qu.:0.2187 3rd Qu.:0.3735
## Max. :1.0000 Max. :1.0000
```

Os dados foram normalizados!

Etapa 3 - Modelagem

Treinando o Modelo. Iremos usar o algoritmo Knn (K nearest neighbors) para o modelo. O Knn é o K vizinhos mais próximos. É um dos classificadores mais simples de ser implementado, de fácil compreensão e ainda hoje pode obter bons resultados dependendo de sua aplicação. A ideia principal do Knn é determinar o rótulo de classificação de uma amostra baseado nas amostras vizinhas advindas de um conjunto de treinamento.

```
# Criando amostras randômicas
str(Dados_Norm)
```

```
## 'data.frame': 569 obs. of 30 variables:
## $ radius_mean : num 0.253 0.171 0.192 0.203 0.389 ...
## $ texture_mean : num 0.0906 0.3125 0.2408 0.1245 0.1184 ...
## $ perimeter_mean : num 0.242 0.176 0.187 0.202 0.372 ...
## $ area_mean : num 0.136 0.0861 0.0974 0.1024 0.2411 ...
## $ smoothness_mean : num 0.453 0.399 0.497 0.576 0.244 ...
## $ compactness_mean : num 0.155 0.292 0.18 0.289 0.153 ...
## $ concavity_mean : num 0.0934 0.1496 0.0714 0.1086 0.0795 ...
## $ points_mean : num 0.184 0.131 0.123 0.238 0.132 ...
## $ symmetry_mean : num 0.454 0.435 0.33 0.359 0.334 ...
## $ dimension_mean : num 0.202 0.315 0.283 0.227 0.115 ...
## $ radius_se : num 0.0451 0.1228 0.0309 0.0822 0.0242 ...
## $ texture_se : num 0.0675 0.1849 0.2269 0.2172 0.0116 ...
## $ perimeter_se : num 0.043 0.1259 0.0276 0.0515 0.0274 ...
## $ area_se : num 0.0199 0.0379 0.0126 0.0365 0.0204 ...
## $ smoothness_se : num 0.215 0.196 0.117 0.325 0.112 ...
## $ compactness_se : num 0.0717 0.252 0.0533 0.2458 0.0946 ...
## $ concavity_se : num 0.0425 0.0847 0.0267 0.0552 0.0392 ...
## $ points_se : num 0.235 0.259 0.142 0.372 0.173 ...
## $ symmetry_se : num 0.16 0.382 0.131 0.111 0.121 ...
## $ dimension_se : num 0.0468 0.0837 0.045 0.088 0.0301 ...
```

```
## $ radius_worst      : num  0.198 0.141 0.159 0.142 0.294 ...
## $ texture_worst     : num  0.0965 0.291 0.3843 0.0999 0.0989 ...
## $ perimeter_worst  : num  0.182 0.139 0.147 0.13 0.269 ...
## $ area_worst       : num  0.0894 0.0589 0.0703 0.0611 0.1558 ...
## $ smoothness_worst : num  0.445 0.331 0.434 0.433 0.274 ...
## $ compactness_worst: num  0.0964 0.2175 0.1173 0.1503 0.142 ...
## $ concavity_worst  : num  0.0992 0.153 0.0852 0.0692 0.1088 ...
## $ points_worst     : num  0.323 0.272 0.255 0.296 0.281 ...
## $ symmetry_worst   : num  0.249 0.271 0.282 0.106 0.182 ...
## $ dimension_worst  : num  0.0831 0.1366 0.1559 0.084 0.0828 ...

set.seed(1)
Amostra <- sample.split(Dados_Norm, SplitRatio = 0.70)

# Treinamos o nosso modelo nos dados de treino
# Dados_Treino
Dados_Treino <- subset(Dados_Norm, Amostra == TRUE)
# Dados_Testes
Dados_Testes <- subset(Dados_Norm, Amostra == FALSE)

# Criando os labels para os dados de treino e de teste
Dados_Treino_Labels <- subset(Dados[1:569,1], Amostra == TRUE)
Dados_Testes_Labels <- subset(Dados[1:569,1], Amostra == FALSE)

# Criando o modelo
Modelo_v1 <- knn(train = Dados_Treino, test = Dados_Testes, cl = Dados_Treino_Labels,
                 k = 10)

# A função knn() retorna um objeto do tipo fator com as previsões para cada exemplo no dataset de teste
class(Modelo_v1)

## [1] "factor"
```

Etapa 4 - Interpretando o Modelo

Nesta etapa iremos analisar a performance do Modelo_v1.

```
# Criando uma tabela cruzada dos dados previstos x dados atuais
confusionMatrix(Dados_Testes_Labels, Modelo_v1)

## Confusion Matrix and Statistics
##
##              Reference
## Prediction Benigno Maligno
##    Benigno    114      1
##    Maligno      4     52
##
##              Accuracy : 0.9708
##              95% CI : (0.9331, 0.9904)
```

```
##      No Information Rate : 0.6901
##      P-Value [Acc > NIR] : <2e-16
##
##              Kappa : 0.9327
##  Mcnemar's Test P-Value : 0.3711
##
##      Sensitivity : 0.9661
##      Specificity : 0.9811
##      Pos Pred Value : 0.9913
##      Neg Pred Value : 0.9286
##      Prevalence : 0.6901
##      Detection Rate : 0.6667
##      Detection Prevalence : 0.6725
##      Balanced Accuracy : 0.9736
##
##      'Positive' Class : Benigno
##
```

Interpretando os Resultados A tabela cruzada mostra 4 possíveis valores:

A primeira coluna lista os labels originais nos dados observados As duas colunas do modelo (Benigno e Maligno) do modelo, mostram os resultados da previsão Temos:

Cenário 1: Célula Benigno (label) x Benigno (Modelo) - 114 casos - true negative

Cenário 2: Célula Benigno (label) x Maligno (Modelo) - 001 casos - false positive (o modelo errou)

Cenário 3: Célula Maligno (label) x Benigno (Modelo) - 004 casos - false negative (o modelo errou)

Cenário 4: Célula Maligno (label) x Maligno (Modelo) - 052 casos - true positive

Lendo a Confusion Matrix (Perspectva de ter ou não a doença):

True Negative = nosso modelo previu que a pessoa NÃO tinha a doença e os dados mostraram que realmente a pessoa NÃO tinha a doença

False Positive = nosso modelo previu que a pessoa tinha a doença e os dados mostraram que NÃO, a pessoa não tinha a doença

False Negative = nosso modelo previu que a pessoa NÃO tinha a doença e os dados mostraram que SIM, a pessoa tinha a doença

True Positive = nosso modelo previu que a pessoa tinha a doença e os dados mostraram que SIM, a pessoa tinha a doença

Falso Positivo - Erro Tipo I Falso Negativo - Erro Tipo II

Etapa 5 - Otimização do Modelo

Testando diferentes valores para k

#k=1

```
Dados_Test_Pred <- knn(train = Dados_Treino, test = Dados_Testes,  
                       cl = Dados_Treino_Labels, k=1)  
confusionMatrix(Dados_Testes_Labels, Dados_Test_Pred)
```

```
## Confusion Matrix and Statistics
```

```
##
```

```
##           Reference
```

```
## Prediction Benigno Maligno
```

```
##   Benigno    110      5
```

```
##   Maligno      3    53
```

```
##
```

```
##           Accuracy : 0.9532
```

```
##           95% CI : (0.9099, 0.9796)
```

```
##   No Information Rate : 0.6608
```

```
##   P-Value [Acc > NIR] : <2e-16
```

```
##
```

```
##           Kappa : 0.8948
```

```
## Mcnemar's Test P-Value : 0.7237
```

```
##
```

```
##           Sensitivity : 0.9735
```

```
##           Specificity : 0.9138
```

```
##           Pos Pred Value : 0.9565
```

```
##           Neg Pred Value : 0.9464
```

```
##           Prevalence : 0.6608
```

```
##           Detection Rate : 0.6433
```

```
##   Detection Prevalence : 0.6725
```

```
##           Balanced Accuracy : 0.9436
```

```
##
```

```
##           'Positive' Class : Benigno
```

```
##
```

Taxa Acerto = 95,32

#k=12

```
Dados_Test_Pred <- knn(train = Dados_Treino, test = Dados_Testes,  
                       cl = Dados_Treino_Labels, k=12)  
confusionMatrix(Dados_Testes_Labels, Dados_Test_Pred)
```

```
## Confusion Matrix and Statistics
```

```
##
```

```
##           Reference
```

```
## Prediction Benigno Maligno
```

```
##   Benigno    114      1
```

```
##   Maligno      5    51
```

```
##
```

```
##           Accuracy : 0.9649
```

```

##          95% CI : (0.9252, 0.987)
##      No Information Rate : 0.6959
##      P-Value [Acc > NIR] : <2e-16
##
##          Kappa : 0.9189
##  McNemar's Test P-Value : 0.2207
##
##          Sensitivity : 0.9580
##          Specificity : 0.9808
##          Pos Pred Value : 0.9913
##          Neg Pred Value : 0.9107
##          Prevalence : 0.6959
##          Detection Rate : 0.6667
##      Detection Prevalence : 0.6725
##          Balanced Accuracy : 0.9694
##
##      'Positive' Class : Benigno
##
# Taxa Acerto = 97,08

#k=20
Dados_Test_Pred <- knn(train = Dados_Treino, test = Dados_Testes,
                       cl = Dados_Treino_Labels, k=20)
confusionMatrix(Dados_Testes_Labels, Dados_Test_Pred)

## Confusion Matrix and Statistics
##
##          Reference
## Prediction Benigno Maligno
##   Benigno      115         0
##   Maligno         6        50
##
##          Accuracy : 0.9649
##          95% CI : (0.9252, 0.987)
##      No Information Rate : 0.7076
##      P-Value [Acc > NIR] : < 2e-16
##
##          Kappa : 0.9181
##  McNemar's Test P-Value : 0.04123
##
##          Sensitivity : 0.9504
##          Specificity : 1.0000
##          Pos Pred Value : 1.0000
##          Neg Pred Value : 0.8929
##          Prevalence : 0.7076
##          Detection Rate : 0.6725
##      Detection Prevalence : 0.6725
##          Balanced Accuracy : 0.9752
##

```

```

##          'Positive' Class : Benigno
##

# Taxa Acerto = 97,08

#k=29
Dados_Test_Pred <- knn(train = Dados_Treino, test = Dados_Testes,
                       cl = Dados_Treino_Labels, k=29)
confusionMatrix(Dados_Testes_Labels, Dados_Test_Pred)

## Confusion Matrix and Statistics
##
##          Reference
## Prediction Benigno Maligno
##   Benigno      115         0
##   Maligno         7        49
##
##          Accuracy : 0.9591
##          95% CI : (0.9175, 0.9834)
##   No Information Rate : 0.7135
##   P-Value [Acc > NIR] : < 2e-16
##
##          Kappa : 0.904
##  Mcnemar's Test P-Value : 0.02334
##
##          Sensitivity : 0.9426
##          Specificity : 1.0000
##   Pos Pred Value : 1.0000
##   Neg Pred Value : 0.8750
##   Prevalence : 0.7135
##   Detection Rate : 0.6725
##   Detection Prevalence : 0.6725
##   Balanced Accuracy : 0.9713
##
##          'Positive' Class : Benigno
##

# Taxa Acerto = 95,91

#k=36
Dados_Test_Pred <- knn(train = Dados_Treino, test = Dados_Testes,
                       cl = Dados_Treino_Labels, k=36)
confusionMatrix(Dados_Testes_Labels, Dados_Test_Pred)

## Confusion Matrix and Statistics
##
##          Reference
## Prediction Benigno Maligno
##   Benigno      115         0
##   Maligno         7        49
##

```

```

##              Accuracy : 0.9591
##              95% CI : (0.9175, 0.9834)
##      No Information Rate : 0.7135
##      P-Value [Acc > NIR] : < 2e-16
##
##              Kappa : 0.904
##  McNemar's Test P-Value : 0.02334
##
##              Sensitivity : 0.9426
##              Specificity : 1.0000
##              Pos Pred Value : 1.0000
##              Neg Pred Value : 0.8750
##              Prevalence : 0.7135
##              Detection Rate : 0.6725
##      Detection Prevalence : 0.6725
##              Balanced Accuracy : 0.9713
##
##      'Positive' Class : Benigno
##

# Taxa Acerto = 95,91

#k=50
Dados_Test_Pred <- knn(train = Dados_Treino, test = Dados_Testes,
                       cl = Dados_Treino_Labels, k=50)
confusionMatrix(Dados_Testes_Labels, Dados_Test_Pred)

## Confusion Matrix and Statistics
##
##              Reference
## Prediction Benigno Maligno
##   Benigno      115         0
##   Maligno         7        49
##
##              Accuracy : 0.9591
##              95% CI : (0.9175, 0.9834)
##      No Information Rate : 0.7135
##      P-Value [Acc > NIR] : < 2e-16
##
##              Kappa : 0.904
##  McNemar's Test P-Value : 0.02334
##
##              Sensitivity : 0.9426
##              Specificity : 1.0000
##              Pos Pred Value : 1.0000
##              Neg Pred Value : 0.8750
##              Prevalence : 0.7135
##              Detection Rate : 0.6725
##      Detection Prevalence : 0.6725
##              Balanced Accuracy : 0.9713

```

```
##
##      'Positive' Class : Benigno
##
# Taxa Acerto = 95,91
```

Etapa 6 - Calculando a Taxa de Acerto

```
## Calculando a taxa de acerto
Prev = NULL
Taxa_Acerto = NULL

suppressWarnings(for(i in 1:50)
{
  set.seed(1)
  Prev = knn(train = Dados_Treino, test = Dados_Testes,
             cl = Dados_Treino_Labels, k = i)
  CT = CrossTable(x = Dados_Testes_Labels, y = Prev,
                 prop.chisq=FALSE)
  Taxa_Acerto[i] <- ((CT$t[1] + CT$t[4]) / sum(CT$t))*
100
})
```

```
##
##
##      Cell Contents
## |-----|
## |                      N |
## |      N / Row Total |
## |      N / Col Total |
## |      N / Table Total |
## |-----|
##
##
## Total Observations in Table:  171
##
```

	Prev		
Dados_Testes_Labels	Benigno	Maligno	Row Total
Benigno	110	5	115
	0.957	0.043	0.673
	0.973	0.086	
	0.643	0.029	
Maligno	3	53	56
	0.054	0.946	0.327
	0.027	0.914	
	0.018	0.310	
Column Total	113	58	171

##		0.661	0.339	
##	-----	-----	-----	-----

##

Cell Contents

##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	-----

##

##

Total Observations in Table: 171

##

##

##		Prev		
##	Dados_Testes_Labels	Benigno	Maligno	Row Total
##	-----	-----	-----	-----
##	Benigno	110	5	115
##		0.957	0.043	0.673
##		0.957	0.089	
##		0.643	0.029	
##	-----	-----	-----	-----
##	Maligno	5	51	56
##		0.089	0.911	0.327
##		0.043	0.911	
##		0.029	0.298	
##	-----	-----	-----	-----
##	Column Total	115	56	171
##		0.673	0.327	
##	-----	-----	-----	-----

##

##

##

##

Cell Contents

##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	-----

##

##

Total Observations in Table: 171

##

##

##	Prev		
## Dados_Testes_Labels	Benigno	Maligno	Row Total
##			
## Benigno	113	2	115
##	0.983	0.017	0.673
##	0.966	0.037	
##	0.661	0.012	
##			
## Maligno	4	52	56
##	0.071	0.929	0.327
##	0.034	0.963	
##	0.023	0.304	
##			
## Column Total	117	54	171
##	0.684	0.316	
##			

##	
##	
##	
##	
##	Cell Contents
##	
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	

##

##

##

Total Observations in Table: 171

##

##	Prev		
## Dados_Testes_Labels	Benigno	Maligno	Row Total
##			
## Benigno	113	2	115
##	0.983	0.017	0.673
##	0.966	0.037	
##	0.661	0.012	
##			
## Maligno	4	52	56
##	0.071	0.929	0.327
##	0.034	0.963	
##	0.023	0.304	
##			
## Column Total	117	54	171
##	0.684	0.316	
##			

##

##

```

##
##
##      Cell Contents
## |-----|
## |                      N |
## |      N / Row Total |
## |      N / Col Total |
## |      N / Table Total |
## |-----|
##
##
## Total Observations in Table:  171
##
##
##      Dados_Testes_Labels | Prev
##      Benigno | Maligno | Row Total |
## -----|-----|-----|
##      Benigno |      115 |         0 |      115 |
##              |      1.000 |      0.000 |      0.673 |
##              |      0.966 |      0.000 |
##              |      0.673 |      0.000 |
## -----|-----|-----|
##      Maligno |         4 |        52 |        56 |
##              |      0.071 |      0.929 |      0.327 |
##              |      0.034 |      1.000 |
##              |      0.023 |      0.304 |
## -----|-----|-----|
##      Column Total |      119 |        52 |      171 |
##              |      0.696 |      0.304 |
## -----|-----|-----|
##
##
##
##      Cell Contents
## |-----|
## |                      N |
## |      N / Row Total |
## |      N / Col Total |
## |      N / Table Total |
## |-----|
##
##
## Total Observations in Table:  171
##
##
##      Dados_Testes_Labels | Prev
##      Benigno | Maligno | Row Total |
## -----|-----|-----|
##      Benigno |      115 |         0 |      115 |

```


##		1.000	0.000	0.673
##		0.966	0.000	
##		0.673	0.000	
##	-----	-----	-----	-----
##	Maligno	4	52	56
##		0.071	0.929	0.327
##		0.034	1.000	
##		0.023	0.304	
##	-----	-----	-----	-----
##	Column Total	119	52	171
##		0.696	0.304	
##	-----	-----	-----	-----

##

Cell Contents

##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	-----

##

Total Observations in Table: 171

##

##		Prev		
##	Dados_Testes_Labels	Benigno	Maligno	Row Total
##	-----	-----	-----	-----
##	Benigno	115	0	115
##		1.000	0.000	0.673
##		0.975	0.000	
##		0.673	0.000	
##	-----	-----	-----	-----
##	Maligno	3	53	56
##		0.054	0.946	0.327
##		0.025	1.000	
##		0.018	0.310	
##	-----	-----	-----	-----
##	Column Total	118	53	171
##		0.690	0.310	
##	-----	-----	-----	-----

##

Cell Contents

##	-----
----	-------

```
## |          N |
## |      N / Row Total |
## |      N / Col Total |
## |      N / Table Total |
## |-----|
##
##
## Total Observations in Table:  171
##
##
```

	Prev		
Dados_Testes_Labels	Benigno	Maligno	Row Total
Benigno	114	1	115
	0.991	0.009	0.673
	0.974	0.019	
	0.667	0.006	
Maligno	3	53	56
	0.054	0.946	0.327
	0.026	0.981	
	0.018	0.310	
Column Total	117	54	171
	0.684	0.316	

```
##
##
##
##
## Cell Contents
```

```
## |          N |
## |      N / Row Total |
## |      N / Col Total |
## |      N / Table Total |
## |-----|
##
##
```

```
## Total Observations in Table:  171
##
##
```

	Prev		
Dados_Testes_Labels	Benigno	Maligno	Row Total
Benigno	114	1	115
	0.991	0.009	0.673
	0.966	0.019	
	0.667	0.006	

##	Maligno	4	52	56
##		0.071	0.929	0.327
##		0.034	0.981	
##		0.023	0.304	
##	-----	-----	-----	-----
##	Column Total	118	53	171
##		0.690	0.310	
##	-----	-----	-----	-----

##

Cell Contents

##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	-----

##

##

Total Observations in Table: 171

##

##

##		Prev		
##	Dados_Testes_Labels	Benigno	Maligno	Row Total
##	-----	-----	-----	-----
##	Benigno	114	1	115
##		0.991	0.009	0.673
##		0.966	0.019	
##		0.667	0.006	
##	-----	-----	-----	-----
##	Maligno	4	52	56
##		0.071	0.929	0.327
##		0.034	0.981	
##		0.023	0.304	
##	-----	-----	-----	-----
##	Column Total	118	53	171
##		0.690	0.310	
##	-----	-----	-----	-----

##

##

##

##

Cell Contents

##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total

```
## |-----|
##
##
## Total Observations in Table: 171
##
##
```

	Prev		
Dados_Testes_Labels	Benigno	Maligno	Row Total
Benigno	114	1	115
	0.991	0.009	0.673
	0.966	0.019	
	0.667	0.006	
Maligno	4	52	56
	0.071	0.929	0.327
	0.034	0.981	
	0.023	0.304	
Column Total	118	53	171
	0.690	0.310	

```
## |-----|
##
##
##
##
```

```
##
##
##
##
## Cell Contents
## |-----|
## | N |
## | N / Row Total |
## | N / Col Total |
## | N / Table Total |
## |-----|
##
##
## Total Observations in Table: 171
##
##
```

	Prev		
Dados_Testes_Labels	Benigno	Maligno	Row Total
Benigno	114	1	115
	0.991	0.009	0.673
	0.966	0.019	
	0.667	0.006	
Maligno	4	52	56
	0.071	0.929	0.327
	0.034	0.981	
	0.023	0.304	

##	-----	-----	-----	-----
##	Column Total	118	53	171
##		0.690	0.310	
##	-----	-----	-----	-----

##

##

##

##

Cell Contents

##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	-----

##

##

Total Observations in Table: 171

##

##

##		Prev		
##	Dados_Testes_Labels	Benigno	Maligno	Row Total
##	-----	-----	-----	-----
##	Benigno	114	1	115
##		0.991	0.009	0.673
##		0.958	0.019	
##		0.667	0.006	
##	-----	-----	-----	-----
##	Maligno	5	51	56
##		0.089	0.911	0.327
##		0.042	0.981	
##		0.029	0.298	
##	-----	-----	-----	-----
##	Column Total	119	52	171
##		0.696	0.304	
##	-----	-----	-----	-----

##

##

##

##

Cell Contents

##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	-----

##

##

Total Observations in Table: 171

```
##
##
##
## Dados_Testes_Labels | Prev
```

	Benigno	Maligno	Row Total
-----	-----	-----	-----
Benigno	114	1	115
	0.991	0.009	0.673
	0.958	0.019	
	0.667	0.006	
-----	-----	-----	-----
Maligno	5	51	56
	0.089	0.911	0.327
	0.042	0.981	
	0.029	0.298	
-----	-----	-----	-----
Column Total	119	52	171
	0.696	0.304	
-----	-----	-----	-----

```
##
##
##
##
## Cell Contents
```

N
N / Row Total
N / Col Total
N / Table Total

```
##
##
## Total Observations in Table: 171
```

```
##
##
##
## Dados_Testes_Labels | Prev
```

	Benigno	Maligno	Row Total
-----	-----	-----	-----
Benigno	115	0	115
	1.000	0.000	0.673
	0.966	0.000	
	0.673	0.000	
-----	-----	-----	-----
Maligno	4	52	56
	0.071	0.929	0.327
	0.034	1.000	
	0.023	0.304	
-----	-----	-----	-----
Column Total	119	52	171
	0.696	0.304	
-----	-----	-----	-----

##	-----	-----	-----	-----
##	Benigno	115	0	115
##		1.000	0.000	0.673
##		0.958	0.000	
##		0.673	0.000	
##	-----	-----	-----	-----
##	Maligno	5	51	56
##		0.089	0.911	0.327
##		0.042	1.000	
##		0.029	0.298	
##	-----	-----	-----	-----
##	Column Total	120	51	171
##		0.702	0.298	
##	-----	-----	-----	-----

##

Cell Contents

##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	-----

##

Total Observations in Table: 171

##

##		Prev		
##	Dados_Testes_Labels	Benigno	Maligno	Row Total
##	-----	-----	-----	-----
##	Benigno	115	0	115
##		1.000	0.000	0.673
##		0.958	0.000	
##		0.673	0.000	
##	-----	-----	-----	-----
##	Maligno	5	51	56
##		0.089	0.911	0.327
##		0.042	1.000	
##		0.029	0.298	
##	-----	-----	-----	-----
##	Column Total	120	51	171
##		0.702	0.298	
##	-----	-----	-----	-----

##

Cell Contents

##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	-----

##

##

Total Observations in Table: 171

##

##

##		Prev		
##	Dados_Testes_Labels	Benigno	Maligno	Row Total
##	-----	-----	-----	-----
##	Benigno	115	0	115
##		1.000	0.000	0.673
##		0.950	0.000	
##		0.673	0.000	
##	-----	-----	-----	-----
##	Maligno	6	50	56
##		0.107	0.893	0.327
##		0.050	1.000	
##		0.035	0.292	
##	-----	-----	-----	-----
##	Column Total	121	50	171
##		0.708	0.292	
##	-----	-----	-----	-----

##

##

##

##

Cell Contents

##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	-----

##

##

Total Observations in Table: 171

##

##

##		Prev		
##	Dados_Testes_Labels	Benigno	Maligno	Row Total
##	-----	-----	-----	-----
##	Benigno	115	0	115
##		1.000	0.000	0.673
##		0.958	0.000	

##		0.673	0.000	
##				
##	Maligno	5	51	56
##		0.089	0.911	0.327
##		0.042	1.000	
##		0.029	0.298	
##				
##	Column Total	120	51	171
##		0.702	0.298	
##				

##

##

##

##

Cell Contents

##	
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	

##

##

Total Observations in Table: 171

##

##

##		Prev		
##	Dados_Testes_Labels	Benigno	Maligno	Row Total
##				
##	Benigno	115	0	115
##		1.000	0.000	0.673
##		0.958	0.000	
##		0.673	0.000	
##				
##	Maligno	5	51	56
##		0.089	0.911	0.327
##		0.042	1.000	
##		0.029	0.298	
##				
##	Column Total	120	51	171
##		0.702	0.298	
##				

##

##

##

##

Cell Contents

##	
##	N
##	N / Row Total

```
## |          N / Col Total |
## |          N / Table Total |
## |-----|
##
##
## Total Observations in Table: 171
##
```

	Prev		
Dados_Testes_Labels	Benigno	Maligno	Row Total
Benigno	115	0	115
	1.000	0.000	0.673
	0.958	0.000	
	0.673	0.000	
Maligno	5	51	56
	0.089	0.911	0.327
	0.042	1.000	
	0.029	0.298	
Column Total	120	51	171
	0.702	0.298	

```
##
##
##
## Cell Contents
```

	N
N / Row Total	
N / Col Total	
N / Table Total	

```
##
##
## Total Observations in Table: 171
##
```

	Prev		
Dados_Testes_Labels	Benigno	Maligno	Row Total
Benigno	115	0	115
	1.000	0.000	0.673
	0.958	0.000	
	0.673	0.000	
Maligno	5	51	56
	0.089	0.911	0.327

##		0.042	1.000	
##		0.029	0.298	
##	-----	-----	-----	-----
##	Column Total	120	51	171
##		0.702	0.298	
##	-----	-----	-----	-----

##

##	Cell Contents
##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	-----

##

Total Observations in Table: 171

##

##		Prev		
##	Dados_Testes_Labels	Benigno	Maligno	Row Total
##	-----	-----	-----	-----
##	Benigno	115	0	115
##		1.000	0.000	0.673
##		0.950	0.000	
##		0.673	0.000	
##	-----	-----	-----	-----
##	Maligno	6	50	56
##		0.107	0.893	0.327
##		0.050	1.000	
##		0.035	0.292	
##	-----	-----	-----	-----
##	Column Total	121	50	171
##		0.708	0.292	
##	-----	-----	-----	-----

##

##	Cell Contents
##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	-----

##

```
##
## Total Observations in Table: 171
##
##
##
```

Dados_Testes_Labels	Prev Benigno	Maligno	Row Total
Benigno	115	0	115
	1.000	0.000	0.673
	0.950	0.000	
	0.673	0.000	
Maligno	6	50	56
	0.107	0.893	0.327
	0.050	1.000	
	0.035	0.292	
Column Total	121	50	171
	0.708	0.292	

```
##
##
##
##
## Cell Contents
```

N
N / Row Total
N / Col Total
N / Table Total

```
##
##
## Total Observations in Table: 171
##
##
##
```

Dados_Testes_Labels	Prev Benigno	Maligno	Row Total
Benigno	115	0	115
	1.000	0.000	0.673
	0.950	0.000	
	0.673	0.000	
Maligno	6	50	56
	0.107	0.893	0.327
	0.050	1.000	
	0.035	0.292	
Column Total	121	50	171

##		0.708	0.292	
##	-----	-----	-----	-----

##

##	Cell Contents
##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	-----

##

Total Observations in Table: 171

##

##		Prev		
##	Dados_Testes_Labels	Benigno	Maligno	Row Total
##	-----	-----	-----	-----
##	Benigno	115	0	115
##		1.000	0.000	0.673
##		0.950	0.000	
##		0.673	0.000	
##	-----	-----	-----	-----
##	Maligno	6	50	56
##		0.107	0.893	0.327
##		0.050	1.000	
##		0.035	0.292	
##	-----	-----	-----	-----
##	Column Total	121	50	171
##		0.708	0.292	
##	-----	-----	-----	-----

##

##

##	Cell Contents
##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	-----

##

Total Observations in Table: 171

##

##	Prev		
## Dados_Testes_Labels	Benigno	Maligno	Row Total
##			
## Benigno	115	0	115
##	1.000	0.000	0.673
##	0.950	0.000	
##	0.673	0.000	
##			
## Maligno	6	50	56
##	0.107	0.893	0.327
##	0.050	1.000	
##	0.035	0.292	
##			
## Column Total	121	50	171
##	0.708	0.292	
##			

##	
##	
##	
##	
##	Cell Contents
##	
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	

##

##

##

Total Observations in Table: 171

##

##	Prev		
## Dados_Testes_Labels	Benigno	Maligno	Row Total
##			
## Benigno	115	0	115
##	1.000	0.000	0.673
##	0.943	0.000	
##	0.673	0.000	
##			
## Maligno	7	49	56
##	0.125	0.875	0.327
##	0.057	1.000	
##	0.041	0.287	
##			
## Column Total	122	49	171
##	0.713	0.287	
##			

##

##

```

##
##
##      Cell Contents
## |-----|
## |                      N |
## |      N / Row Total |
## |      N / Col Total |
## |      N / Table Total |
## |-----|
##
##
## Total Observations in Table:  171
##
##
##      Dados_Testes_Labels | Prev
##      Benigno | Maligno | Row Total |
## -----|-----|-----|
##      Benigno |      115 |          0 |      115 |
##              |      1.000 |      0.000 |      0.673 |
##              |      0.950 |      0.000 |
##              |      0.673 |      0.000 |
## -----|-----|-----|
##      Maligno |         6 |         50 |         56 |
##              |      0.107 |      0.893 |      0.327 |
##              |      0.050 |      1.000 |
##              |      0.035 |      0.292 |
## -----|-----|-----|
##      Column Total |      121 |         50 |      171 |
##              |      0.708 |      0.292 |
## -----|-----|-----|
##
##
##
##      Cell Contents
## |-----|
## |                      N |
## |      N / Row Total |
## |      N / Col Total |
## |      N / Table Total |
## |-----|
##
##
## Total Observations in Table:  171
##
##
##      Dados_Testes_Labels | Prev
##      Benigno | Maligno | Row Total |
## -----|-----|-----|
##      Benigno |      115 |          0 |      115 |

```


##		1.000	0.000	0.673
##		0.950	0.000	
##		0.673	0.000	
##	-----	-----	-----	-----
##	Maligno	6	50	56
##		0.107	0.893	0.327
##		0.050	1.000	
##		0.035	0.292	
##	-----	-----	-----	-----
##	Column Total	121	50	171
##		0.708	0.292	
##	-----	-----	-----	-----

##

Cell Contents

##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	-----

##

Total Observations in Table: 171

##

##		Prev		
##	Dados_Testes_Labels	Benigno	Maligno	Row Total
##	-----	-----	-----	-----
##	Benigno	115	0	115
##		1.000	0.000	0.673
##		0.950	0.000	
##		0.673	0.000	
##	-----	-----	-----	-----
##	Maligno	6	50	56
##		0.107	0.893	0.327
##		0.050	1.000	
##		0.035	0.292	
##	-----	-----	-----	-----
##	Column Total	121	50	171
##		0.708	0.292	
##	-----	-----	-----	-----

##

Cell Contents

##	-----
----	-------

```
## |          N |
## |      N / Row Total |
## |      N / Col Total |
## |      N / Table Total |
## |-----|
##
##
## Total Observations in Table:  171
##
##
```

	Prev		
Dados_Testes_Labels	Benigno	Maligno	Row Total
Benigno	115	0	115
	1.000	0.000	0.673
	0.943	0.000	
	0.673	0.000	
Maligno	7	49	56
	0.125	0.875	0.327
	0.057	1.000	
	0.041	0.287	
Column Total	122	49	171
	0.713	0.287	

```
##
##
##
##
## Cell Contents
## |-----|
## |          N |
## |      N / Row Total |
## |      N / Col Total |
## |      N / Table Total |
## |-----|
##
##
## Total Observations in Table:  171
##
##
```

	Prev		
Dados_Testes_Labels	Benigno	Maligno	Row Total
Benigno	115	0	115
	1.000	0.000	0.673
	0.943	0.000	
	0.673	0.000	

##	Maligno	7	49	56
##		0.125	0.875	0.327
##		0.057	1.000	
##		0.041	0.287	
##	-----	-----	-----	-----
##	Column Total	122	49	171
##		0.713	0.287	
##	-----	-----	-----	-----

##

Cell Contents

##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	-----

##

##

Total Observations in Table: 171

##

##

##		Prev		
##	Dados_Testes_Labels	Benigno	Maligno	Row Total
##	-----	-----	-----	-----
##	Benigno	115	0	115
##		1.000	0.000	0.673
##		0.943	0.000	
##		0.673	0.000	
##	-----	-----	-----	-----
##	Maligno	7	49	56
##		0.125	0.875	0.327
##		0.057	1.000	
##		0.041	0.287	
##	-----	-----	-----	-----
##	Column Total	122	49	171
##		0.713	0.287	
##	-----	-----	-----	-----

##

##

##

##

Cell Contents

##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total

```
## |-----|
##
##
## Total Observations in Table: 171
##
##
```

	Prev		
Dados_Testes_Labels	Benigno	Maligno	Row Total
Benigno	115	0	115
	1.000	0.000	0.673
	0.943	0.000	
	0.673	0.000	
Maligno	7	49	56
	0.125	0.875	0.327
	0.057	1.000	
	0.041	0.287	
Column Total	122	49	171
	0.713	0.287	

```
## |-----|
##
##
##
##
```

```
##
##
##
##
## Cell Contents
## |-----|
## | N |
## | N / Row Total |
## | N / Col Total |
## | N / Table Total |
## |-----|
##
##
## Total Observations in Table: 171
##
##
```

	Prev		
Dados_Testes_Labels	Benigno	Maligno	Row Total
Benigno	115	0	115
	1.000	0.000	0.673
	0.943	0.000	
	0.673	0.000	
Maligno	7	49	56
	0.125	0.875	0.327
	0.057	1.000	
	0.041	0.287	

##	-----	-----	-----	-----
##	Column Total	122	49	171
##		0.713	0.287	
##	-----	-----	-----	-----

##

##

##

##

Cell Contents

##	-----	
##		N
##	N / Row Total	
##	N / Col Total	
##	N / Table Total	
##	-----	

##

##

Total Observations in Table: 171

##

##

##		Prev		
##	Dados_Testes_Labels	Benigno	Maligno	Row Total
##	-----	-----	-----	-----
##	Benigno	115	0	115
##		1.000	0.000	0.673
##		0.943	0.000	
##		0.673	0.000	
##	-----	-----	-----	-----
##	Maligno	7	49	56
##		0.125	0.875	0.327
##		0.057	1.000	
##		0.041	0.287	
##	-----	-----	-----	-----
##	Column Total	122	49	171
##		0.713	0.287	
##	-----	-----	-----	-----

##

##

##

##

Cell Contents

##	-----	
##		N
##	N / Row Total	
##	N / Col Total	
##	N / Table Total	
##	-----	

##

##

Total Observations in Table: 171

```
##
##
##
## Dados_Testes_Labels | Prev
```

	Benigno	Maligno	Row Total
-----	-----	-----	-----
Benigno	115	0	115
	1.000	0.000	0.673
	0.943	0.000	
	0.673	0.000	
-----	-----	-----	-----
Maligno	7	49	56
	0.125	0.875	0.327
	0.057	1.000	
	0.041	0.287	
-----	-----	-----	-----
Column Total	122	49	171
	0.713	0.287	
-----	-----	-----	-----

```
##
##
##
##
## Cell Contents
```

N
N / Row Total
N / Col Total
N / Table Total

```
##
##
## Total Observations in Table: 171
##
```

```
##
##
## Dados_Testes_Labels | Prev
```

	Benigno	Maligno	Row Total
-----	-----	-----	-----
Benigno	115	0	115
	1.000	0.000	0.673
	0.943	0.000	
	0.673	0.000	
-----	-----	-----	-----
Maligno	7	49	56
	0.125	0.875	0.327
	0.057	1.000	
	0.041	0.287	
-----	-----	-----	-----
Column Total	122	49	171
	0.713	0.287	
-----	-----	-----	-----

```

##
##
##
##
## Cell Contents
## |-----|
## |                N |
## |      N / Row Total |
## |      N / Col Total |
## |      N / Table Total |
## |-----|
##
##
## Total Observations in Table:  171
##
##
##
## DADOS TESTE
## Prev
## DADOS TESTE LABELS | Benigno | Maligno | Row Total |
## |-----|-----|-----|-----|
## | Benigno |      115 |         0 |      115 |
## |          |      1.000 |      0.000 |      0.673 |
## |          |      0.943 |      0.000 |          |
## |          |      0.673 |      0.000 |          |
## |-----|-----|-----|-----|
## | Maligno |         7 |        49 |        56 |
## |          |      0.125 |      0.875 |      0.327 |
## |          |      0.057 |      1.000 |          |
## |          |      0.041 |      0.287 |          |
## |-----|-----|-----|-----|
## | Column Total |      122 |        49 |      171 |
## |          |      0.713 |      0.287 |          |
## |-----|-----|-----|-----|
##
##
##
## Cell Contents
## |-----|
## |                N |
## |      N / Row Total |
## |      N / Col Total |
## |      N / Table Total |
## |-----|
##
##
## Total Observations in Table:  171
##
##
##
## DADOS TESTE
## Prev
## DADOS TESTE LABELS | Benigno | Maligno | Row Total |

```

##	-----	-----	-----	-----
##	Benigno	115	0	115
##		1.000	0.000	0.673
##		0.943	0.000	
##		0.673	0.000	
##	-----	-----	-----	-----
##	Maligno	7	49	56
##		0.125	0.875	0.327
##		0.057	1.000	
##		0.041	0.287	
##	-----	-----	-----	-----
##	Column Total	122	49	171
##		0.713	0.287	
##	-----	-----	-----	-----

##

##

##

##

Cell Contents

##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	-----

##

##

Total Observations in Table: 171

##

##

##		Prev		
##	Dados_Testes_Labels	Benigno	Maligno	Row Total
##	-----	-----	-----	-----
##	Benigno	115	0	115
##		1.000	0.000	0.673
##		0.943	0.000	
##		0.673	0.000	
##	-----	-----	-----	-----
##	Maligno	7	49	56
##		0.125	0.875	0.327
##		0.057	1.000	
##		0.041	0.287	
##	-----	-----	-----	-----
##	Column Total	122	49	171
##		0.713	0.287	
##	-----	-----	-----	-----

##

##

##

##

Cell Contents

##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	-----

##

##

Total Observations in Table: 171

##

##

##		Prev		
##	Dados_Testes_Labels	Benigno	Maligno	Row Total
##	-----	-----	-----	-----
##	Benigno	115	0	115
##		1.000	0.000	0.673
##		0.943	0.000	
##		0.673	0.000	
##	-----	-----	-----	-----
##	Maligno	7	49	56
##		0.125	0.875	0.327
##		0.057	1.000	
##		0.041	0.287	
##	-----	-----	-----	-----
##	Column Total	122	49	171
##		0.713	0.287	
##	-----	-----	-----	-----

##

##

##

##

Cell Contents

##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	-----

##

##

Total Observations in Table: 171

##

##

##		Prev		
##	Dados_Testes_Labels	Benigno	Maligno	Row Total
##	-----	-----	-----	-----
##	Benigno	115	0	115
##		1.000	0.000	0.673
##		0.943	0.000	

##		0.673	0.000	
##				
##	Maligno	7	49	56
##		0.125	0.875	0.327
##		0.057	1.000	
##		0.041	0.287	
##				
##	Column Total	122	49	171
##		0.713	0.287	
##				

##

##

##

##

Cell Contents

##	
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	

##

##

Total Observations in Table: 171

##

##

##		Prev		
##	Dados_Testes_Labels	Benigno	Maligno	Row Total
##				
##	Benigno	115	0	115
##		1.000	0.000	0.673
##		0.943	0.000	
##		0.673	0.000	
##				
##	Maligno	7	49	56
##		0.125	0.875	0.327
##		0.057	1.000	
##		0.041	0.287	
##				
##	Column Total	122	49	171
##		0.713	0.287	
##				

##

##

##

##

Cell Contents

##	
##	N
##	N / Row Total

```
## |          N / Col Total |
## |          N / Table Total |
## |-----|
##
##
## Total Observations in Table: 171
##
```

	Prev		
Dados_Testes_Labels	Benigno	Maligno	Row Total
Benigno	115	0	115
	1.000	0.000	0.673
	0.943	0.000	
	0.673	0.000	
Maligno	7	49	56
	0.125	0.875	0.327
	0.057	1.000	
	0.041	0.287	
Column Total	122	49	171
	0.713	0.287	

```
##
##
##
## Cell Contents
```

	N
N / Row Total	
N / Col Total	
N / Table Total	

```
##
## Total Observations in Table: 171
##
```

	Prev		
Dados_Testes_Labels	Benigno	Maligno	Row Total
Benigno	115	0	115
	1.000	0.000	0.673
	0.935	0.000	
	0.673	0.000	
Maligno	8	48	56
	0.143	0.857	0.327

##		0.065	1.000	
##		0.047	0.281	
##	-----	-----	-----	-----
##	Column Total	123	48	171
##		0.719	0.281	
##	-----	-----	-----	-----

##

##	Cell Contents
##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	-----

##

Total Observations in Table: 171

##

##		Prev		
##	Dados_Testes_Labels	Benigno	Maligno	Row Total
##	-----	-----	-----	-----
##	Benigno	115	0	115
##		1.000	0.000	0.673
##		0.943	0.000	
##		0.673	0.000	
##	-----	-----	-----	-----
##	Maligno	7	49	56
##		0.125	0.875	0.327
##		0.057	1.000	
##		0.041	0.287	
##	-----	-----	-----	-----
##	Column Total	122	49	171
##		0.713	0.287	
##	-----	-----	-----	-----

##

##	Cell Contents
##	-----
##	N
##	N / Row Total
##	N / Col Total
##	N / Table Total
##	-----

##

```
##
## Total Observations in Table: 171
##
##
##      Dados_Testes_Labels | Prev
##      Benigno | Maligno | Row Total |
## -----|-----|-----|
##      Benigno | 115 | 0 | 115 |
##      | 1.000 | 0.000 | 0.673 |
##      | 0.943 | 0.000 | |
##      | 0.673 | 0.000 | |
## -----|-----|-----|
##      Maligno | 7 | 49 | 56 |
##      | 0.125 | 0.875 | 0.327 |
##      | 0.057 | 1.000 | |
##      | 0.041 | 0.287 | |
## -----|-----|-----|
##      Column Total | 122 | 49 | 171 |
##      | 0.713 | 0.287 | |
## -----|-----|-----|
##
##
```

Obtendo os valores de k e das taxas de erro

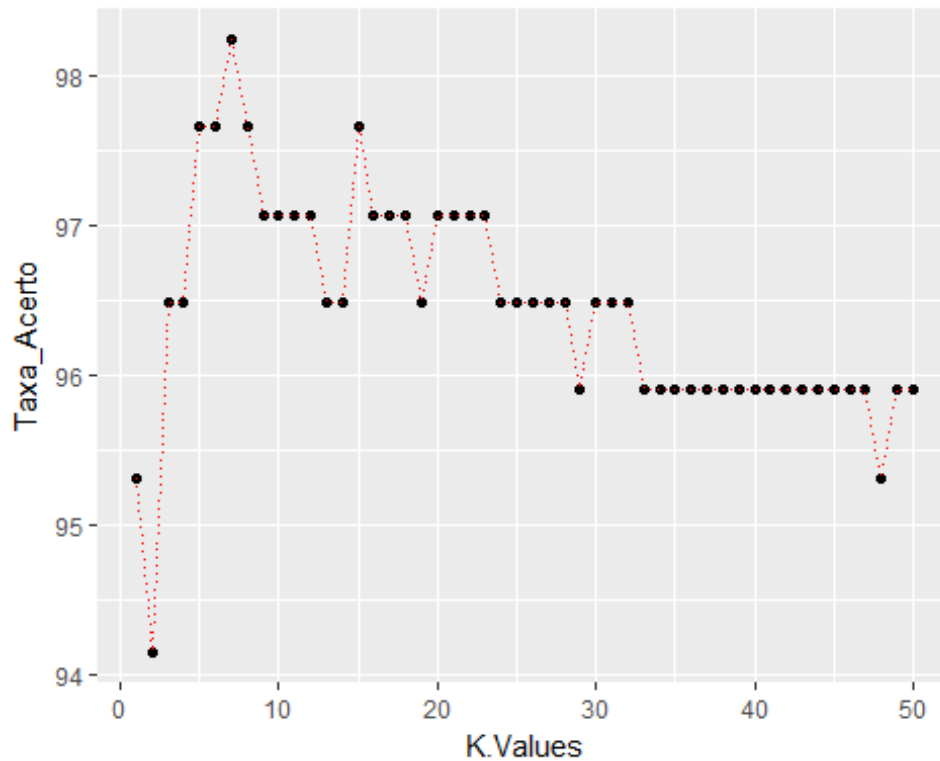
```
K.Values <- 1:50
Df_Acerto <- data.frame(Taxa_Acerto, K.Values)
Df_Acerto
```

```
##      Taxa_Acerto K.Values
## 1      95.32164      1
## 2      94.15205      2
## 3      96.49123      3
## 4      96.49123      4
## 5      97.66082      5
## 6      97.66082      6
## 7      98.24561      7
## 8      97.66082      8
## 9      97.07602      9
## 10     97.07602     10
## 11     97.07602     11
## 12     97.07602     12
## 13     96.49123     13
## 14     96.49123     14
## 15     97.66082     15
## 16     97.07602     16
## 17     97.07602     17
## 18     97.07602     18
## 19     96.49123     19
## 20     97.07602     20
## 21     97.07602     21
```

```
## 22    97.07602    22
## 23    97.07602    23
## 24    96.49123    24
## 25    96.49123    25
## 26    96.49123    26
## 27    96.49123    27
## 28    96.49123    28
## 29    95.90643    29
## 30    96.49123    30
## 31    96.49123    31
## 32    96.49123    32
## 33    95.90643    33
## 34    95.90643    34
## 35    95.90643    35
## 36    95.90643    36
## 37    95.90643    37
## 38    95.90643    38
## 39    95.90643    39
## 40    95.90643    40
## 41    95.90643    41
## 42    95.90643    42
## 43    95.90643    43
## 44    95.90643    44
## 45    95.90643    45
## 46    95.90643    46
## 47    95.90643    47
## 48    95.32164    48
## 49    95.90643    49
## 50    95.90643    50
```

```
# Grafico
```

```
ggplot(Df_Acerto, aes(x = K.Values, y = Taxa_Acerto)) + geom_point() + geom_line(lty = "dotted", color = 'red')
```



O k=7 foi o que apresentou melhor taxa de acerto

```
max(Df_Acerto)

## [1] 98.24561

Dados_Test_Pred <- knn(train = Dados_Treino, test = Dados_Testes,
                       cl = Dados_Treino_Labels, k=7)
confusionMatrix(Dados_Testes_Labels, Dados_Test_Pred)

## Confusion Matrix and Statistics
##
##              Reference
## Prediction Benigno Maligno
##   Benigno    115         0
##   Maligno      3         53
##
##               Accuracy : 0.9825
##               95% CI : (0.9496, 0.9964)
##   No Information Rate : 0.6901
##   P-Value [Acc > NIR] : <2e-16
##
##               Kappa : 0.9596
##   Mcnemar's Test P-Value : 0.2482
##
##               Sensitivity : 0.9746
##               Specificity : 1.0000
##               Pos Pred Value : 1.0000
```

```
##          Neg Pred Value : 0.9464
##          Prevalence : 0.6901
##          Detection Rate : 0.6725
##    Detection Prevalence : 0.6725
##          Balanced Accuracy : 0.9873
##
##          'Positive' Class : Benigno
##
```

Conclusão

Concluimos que com K=7 temos uma Acuracia de 98,25%, ou seja, erramos 1,75% das previsões. Como na área da saúde não podemos ter erros o ideal é atingir a menor taxa de erro possível. Analisando a nossa matriz temos:

Cenário 1: Célula Benigno (label) x Benigno (Modelo) - 115 casos - true negative

Cenário 2: Célula Benigno (label) x Maligno (Modelo) - 000 casos - false positive (o modelo errou)

Cenário 3: Célula Maligno (label) x Benigno (Modelo) - 003 casos - false negative (o modelo errou)

Cenário 4: Célula Maligno (label) x Maligno (Modelo) - 053 casos - true positive

Lendo a Confusion Matrix (Perspectva de ter ou não a doença):

True Negative = nosso modelo previu que a pessoa NÃO tinha a doença e os dados mostraram que realmente a pessoa NÃO tinha a doença

False Positive = nosso modelo previu que a pessoa tinha a doença e os dados mostraram que NÃO, a pessoa não tinha a doença

False Negative = nosso modelo previu que a pessoa NÃO tinha a doença e os dados mostraram que SIM, a pessoa tinha a doença

True Positive = nosso modelo previu que a pessoa tinha a doença e os dados mostraram que SIM, a pessoa tinha a doença

Ficamos com 3 casos que nosso modelo previu que a pessoa não tinha a doença e na verdade ela tinha. Para melhorar o resultado precisamos treinar o modelo com mais casos para assim melhorar ainda mais a taxa de acerto.

Evanil Tiengo Junior

Belo Horizonte, 07 de novembro de 2018