

# PARCIAL3\_SARMIENTO\_POLO

KLEYDERMSARMIENTO,LUISPOLO

28/11/2021

## UNIVERSIDAD AUTONOMA DEL CARIBE

### CIENCIA DE DATOS EN EL SECTOR SALUD

#### PARCIAL FINAL

#### INTEGRANTES:

KLEYDERM ANDRES SARMIENTO (252010055)

LUIS JOSE POLO (251210277)

#### INSTRUCTOR :

JAIR VILLANUEVA

## RECONOCIMIENTO.

### LIBRERIAS NECESARIAS :

```
library(caret)
library(dplyr)
library(ggplot2)
library(tidyverse)
library(readr)
library(skimr)
library(stats)
library(corrplot)
library(standardize)
library(PerformanceAnalytics)
```

### DATA SET UTILIZADO. \*KLEYDERM

Estoy usando el conjunto de datos cardíacos “SPECTF” de los datos disponibles en el sitio web de ‘UCI Machine Learning Repository’ de 2004.

Link de descarga: <https://archive.ics.uci.edu/ml/datasets/SPECTF+Heart>

El conjunto de datos describe el diagnóstico de imágenes cardíacas de tomografía computarizada por emisión de protón único (SPECT). Cada uno de los pacientes se clasifica en dos categorías: normales y anormales. La base de datos de 267 conjuntos de imágenes SPECT (pacientes) se procesó para extraer características que resumen las imágenes SPECT originales. Como resultado, se crearon 44 patrones de características continuas para cada paciente.

## CARGAMOS EL DATA SET DE ENTRENAMIENTO. \*LUIS

```
library(readr)
SPECTF_train <- read_csv("SPECTF_train.CSV")
```

## CARGAMOS EL DATA SET DE PRUEBA. \*KLEYDERM

```
library(readr)
SPECTF_test <- read_csv("SPECTF_test.CSV")
```

## PREGUNTA DE INVESTIGACIÓN.

Se realizaron exámenes cardiacos a pacientes por medio de tomografía computarizada por emisión de protón único (SPECT), este es un estudio de diagnóstico por imágenes de medicina nuclear que utiliza sustancias radioactivas y una cámara especial con la que se crean imágenes tridimensionales de la parte estudiada en el cuerpo del paciente, las cuales después deben ser visualizadas por un médico quien indica si el paciente tiene o no una anomalía cardiaca.

A cada imagen tomada, se le agregaron 22 regiones de interés (ROI), los cuales se encargan de buscar un similar con una muestra ya preestablecida (diferente para cada ROI), en este caso la muestra que el programa de visión artificial busca es de un paciente sano, por lo que el resultado que arroja cada ROI es de 0 a 100, donde 0 quiere decir que hay 0% de coincidencia y la región de interés no está sana y el 100 quiere decir que la región está sana y exacta a la muestra.

Este data set muestra los resultados de ROI para cada paciente donde se le tomó una tomografía en reposo y una en estrés, por lo tanto, cada paciente tiene 22 ROIS en reposo y 22 ROIS en estrés.

Nuestra pregunta de investigación es: ¿Se puede realizar una integración entre visión artificial y Machine learning para lograr establecer si un paciente tiene o no una anomalía cardiaca solo con hacer una tomografía en reposo y en estrés, luego pasarla por un procesamiento de Rois y que el programa de Machine Learning estudie estos 44 resultados y de un diagnóstico?

## LIMPIEZA.

## ESTADO DEL DATA SET DE ENTRENAMIENTO. \*LUIS

```
summary (SPECTF_train)
```

##	OVERALL_DIAGNOSIS	F1R	F1S	F2R
##	Min. :0.0	Min. :57.00	Min. :51.00	Min. :52.00
##	1st Qu.:0.0	1st Qu.:64.75	1st Qu.:64.00	1st Qu.:66.00
##	Median :0.5	Median :69.00	Median :67.00	Median :70.00
##	Mean :0.5	Mean :68.65	Mean :67.55	Mean :69.55
##	3rd Qu.:1.0	3rd Qu.:72.00	3rd Qu.:73.00	3rd Qu.:74.00
##	Max. :1.0	Max. :80.00	Max. :81.00	Max. :82.00
##				
##	F2S	F3R	F3S	F4R
##	Min. :51.00	Min. :47.00	Min. :46.00	Min. :63.00
##	1st Qu.:67.00	1st Qu.:63.75	1st Qu.:63.75	1st Qu.:68.75
##	Median :72.00	Median :68.00	Median :68.50	Median :72.50
##	Mean :70.85	Mean :67.10	Mean :67.69	Mean :72.08
##	3rd Qu.:75.25	3rd Qu.:71.00	3rd Qu.:71.25	3rd Qu.:75.00
##	Max. :84.00	Max. :78.00	Max. :89.00	Max. :79.00
##				

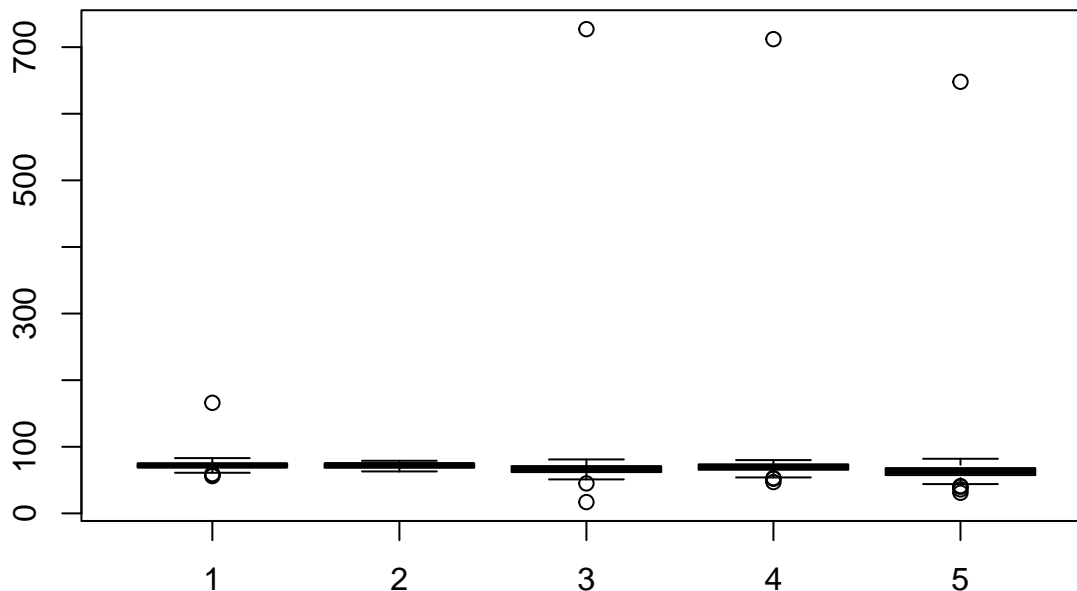
##	F4S	F5R	F5S	F6R	
##	Min. : 56.00	Min. : 26.00	Min. :30.00	Min. :11.00	
##	1st Qu.: 69.00	1st Qu.: 61.00	1st Qu.:59.75	1st Qu.:68.00	
##	Median : 72.00	Median : 65.00	Median :65.50	Median :72.00	
##	Mean : 72.64	Mean : 72.36	Mean :63.95	Mean :69.91	
##	3rd Qu.: 75.00	3rd Qu.: 68.25	3rd Qu.:69.00	3rd Qu.:75.00	
##	Max. :166.00	Max. :713.00	Max. :78.00	Max. :81.00	
##				NA's :1	
##	F6S	F7R	F7S	F8R	
##	Min. :12.00	Min. :43.00	Min. :48.00	Min. :25.00	
##	1st Qu.:68.00	1st Qu.:67.00	1st Qu.:66.75	1st Qu.:62.00	
##	Median :72.00	Median :70.00	Median :71.00	Median :65.00	
##	Mean :70.68	Mean :68.84	Mean :69.56	Mean :64.36	
##	3rd Qu.:76.00	3rd Qu.:73.00	3rd Qu.:74.00	3rd Qu.:69.25	
##	Max. :81.00	Max. :79.00	Max. :80.00	Max. :76.00	
##	NA's :1	NA's :1			
##	F8S	F9R	F9S	F10R	
##	Min. :19.00	Min. :39.00	Min. :37.00	Min. :35.00	
##	1st Qu.:59.00	1st Qu.:66.00	1st Qu.:65.00	1st Qu.:63.00	
##	Median :64.00	Median :69.00	Median :68.00	Median :67.00	
##	Mean :62.33	Mean :68.51	Mean :67.61	Mean :66.65	
##	3rd Qu.:68.00	3rd Qu.:73.00	3rd Qu.:73.00	3rd Qu.:72.00	
##	Max. :74.00	Max. :77.00	Max. :77.00	Max. :81.00	
##		NA's :2			
##	F10S	F11R	F11S	F12R	
##	Min. : 17.00	Min. :56.00	Min. :52.00	Min. : 47.00	
##	1st Qu.: 62.00	1st Qu.:72.00	1st Qu.:71.00	1st Qu.: 65.75	
##	Median : 66.00	Median :75.00	Median :75.00	Median : 70.00	
##	Mean : 73.54	Mean :74.23	Mean :73.91	Mean : 76.81	
##	3rd Qu.: 71.00	3rd Qu.:78.50	3rd Qu.:78.00	3rd Qu.: 73.25	
##	Max. :727.00	Max. :83.00	Max. :84.00	Max. :712.00	
##	NA's :1	NA's :1	NA's :2		
##	F12S	F13R	F13S	F14R	
##	Min. :47.00	Min. :19.00	Min. :14.00	Min. : 8.00	
##	1st Qu.:67.00	1st Qu.:56.50	1st Qu.:55.75	1st Qu.:58.75	
##	Median :71.00	Median :63.00	Median :63.50	Median :64.00	
##	Mean :69.36	Mean :60.52	Mean :59.83	Mean :62.79	
##	3rd Qu.:74.00	3rd Qu.:67.00	3rd Qu.:68.25	3rd Qu.:68.25	
##	Max. :81.00	Max. :77.00	Max. :79.00	Max. :79.00	
##		NA's :1			
##	F14S	F15R	F15S	F16R	
##	Min. : 31.00	Min. :17.00	Min. :20.00	Min. :49.00	
##	1st Qu.: 57.75	1st Qu.:62.00	1st Qu.:61.00	1st Qu.:71.00	
##	Median : 63.50	Median :67.00	Median :67.00	Median :73.00	
##	Mean : 69.17	Mean :64.55	Mean :64.24	Mean :72.06	
##	3rd Qu.: 68.00	3rd Qu.:69.00	3rd Qu.:69.00	3rd Qu.:75.00	
##	Max. :648.00	Max. :76.00	Max. :77.00	Max. :79.00	
##				NA's :1	
##	F16S	F17R	F17S	F18R	F18S
##	Min. :52.00	Min. :49.00	Min. :48.00	Min. :31.0	Min. :13.00
##	1st Qu.:73.00	1st Qu.:59.00	1st Qu.:61.00	1st Qu.:63.0	1st Qu.:63.00
##	Median :74.00	Median :62.00	Median :64.00	Median :66.0	Median :65.50
##	Mean :73.56	Mean :62.39	Mean :63.71	Mean :65.2	Mean :64.66
##	3rd Qu.:76.25	3rd Qu.:65.00	3rd Qu.:66.25	3rd Qu.:68.0	3rd Qu.:69.00

```
## Max. :81.00 Max. :75.00 Max. :80.00 Max. :73.0 Max. :73.00
## NA's :1 NA's :1
## F19R F19S F20R F20S
## Min. :33.00 Min. :23.00 Min. :31.00 Min. :30.00
## 1st Qu.:62.75 1st Qu.:62.00 1st Qu.:65.00 1st Qu.:65.00
## Median :66.00 Median :66.50 Median :70.50 Median :71.00
## Mean :65.61 Mean :65.01 Mean :69.44 Mean :68.41
## 3rd Qu.:71.00 3rd Qu.:71.00 3rd Qu.:74.00 3rd Qu.:74.00
## Max. :77.00 Max. :77.00 Max. :83.00 Max. :80.00
##
## F21R F21S F22R F22S
## Min. :31.00 Min. :30.0 Min. :11.00 Min. :11.00
## 1st Qu.:64.00 1st Qu.:63.0 1st Qu.:53.00 1st Qu.:52.75
## Median :69.00 Median :68.0 Median :59.00 Median :58.00
## Mean :66.34 Mean :65.1 Mean :55.98 Mean :55.08
## 3rd Qu.:73.25 3rd Qu.:71.0 3rd Qu.:63.00 3rd Qu.:63.00
## Max. :79.00 Max. :83.0 Max. :71.00 Max. :73.00
##
```

Dado que en la explicación nos dicen que los valores de los rois se entregan entre 0 y 100, podemos observar mediante la función “summary” que hay variables con valores mayores de 100, por lo tanto, el data set cuenta con Out liers y Missing datas.

## VISUALIZAMOS LOS OUTLIERS \* KLEYDERM

```
boxplot (SPECTF_train$F4S,SPECTF_train$F4R,SPECTF_train$F10S,SPECTF_train$F12R,SPECTF_train$F14S )
```



## CONVERTIMOS LOS OUT LIERS EN NA. \*LUIS

```
SPECTF_train$F4S[SPECTF_train$F4S > 100 ] <- NA
SPECTF_train$F5R[SPECTF_train$F5R > 100 ] <- NA
SPECTF_train$F10S[SPECTF_train$F10S > 100 ] <- NA
SPECTF_train$F12R[SPECTF_train$F12R > 100 ] <- NA
SPECTF_train$F14S[SPECTF_train$F14S > 100 ] <- NA
```

```
SPECTF_train[SPECTF_train > 100]
```

```
## <unspecified> [0]
```

## REEMPLAZAMOS LOS NA POR EL PROMEDIO. \*KLEYDERM

```
library(mice)
imp <- mice(SPECTF_train, m = 5, method = "mean")
imp$imp$F4S
imp$imp$F5R
imp$imp$F6R
imp$imp$F6S
imp$imp$F7R
imp$imp$F9R
imp$imp$F10S
imp$imp$F11R
imp$imp$F11S
imp$imp$F12R
imp$imp$F13R
imp$imp$F14S
imp$imp$F16R
imp$imp$F17R
imp$imp$F18R
SPECTF_train_IMP <- complete(imp,1)
```

```
sum(is.na(SPECTF_train_IMP))
```

```
## [1] 0
```

## ESTUDIAMOS EL DATASET YA LIMPIO. \*LUIS

```
skim(SPECTF_train_IMP)
```

## TRANSFORMACIÓN DE DATOS

### CONVERTIMOS LA VARIABLE CATEGORICA EN OUTCOME

\*LUIS

```
SPECTF_train_IMP <-
  mutate(SPECTF_train_IMP , outcome = (as.factor(
    SPECTF_train_IMP$OVERALL_DIAGNOSIS)))
SPECTF_train_IMP <- SPECTF_train_IMP[,-1]
c.entrenamiento <- SPECTF_train_IMP
```

### CONVERTIMOS EL OUTCOME EN NORMAL Y ANORMAL \*KLEYDERM

```
SPECTF_train_IMP$outcome <- as.character.factor(SPECTF_train_IMP$outcome)
SPECTF_train_IMP$outcome <-
  replace(SPECTF_train_IMP$outcome,
    SPECTF_train_IMP$outcome == "1" , "ANORMAL")
SPECTF_train_IMP$outcome <-
```

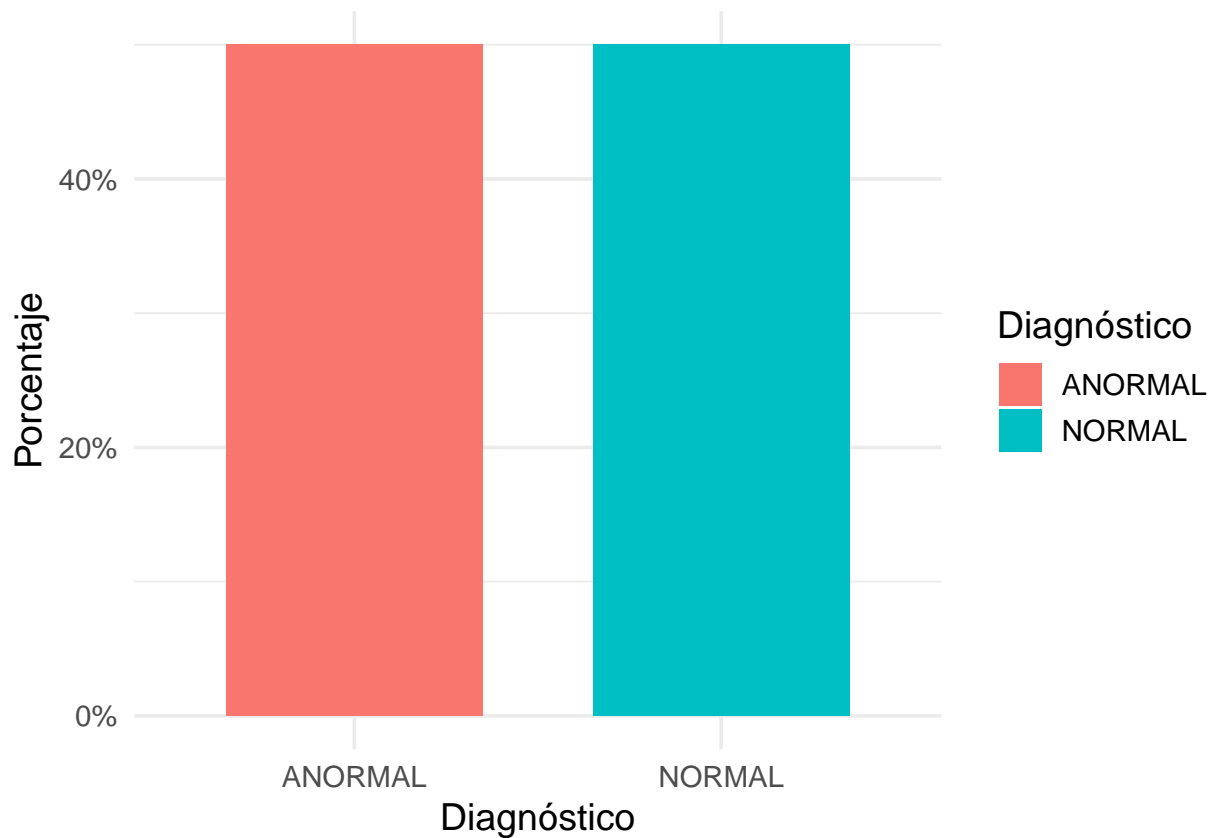
```
replace(SPECTF_train_IMP$outcome,
        SPECTF_train_IMP$outcome == "0" , "NORMAL")
SPECTF_train_IMP$outcome[30:50]
```

```
## [1] "ANORMAL" "ANORMAL" "ANORMAL" "ANORMAL" "ANORMAL" "ANORMAL" "ANORMAL"
## [8] "ANORMAL" "ANORMAL" "ANORMAL" "ANORMAL" "NORMAL" "NORMAL" "NORMAL"
## [15] "NORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL"
```

## VISUALIZACIÓN DE LOS DATOS

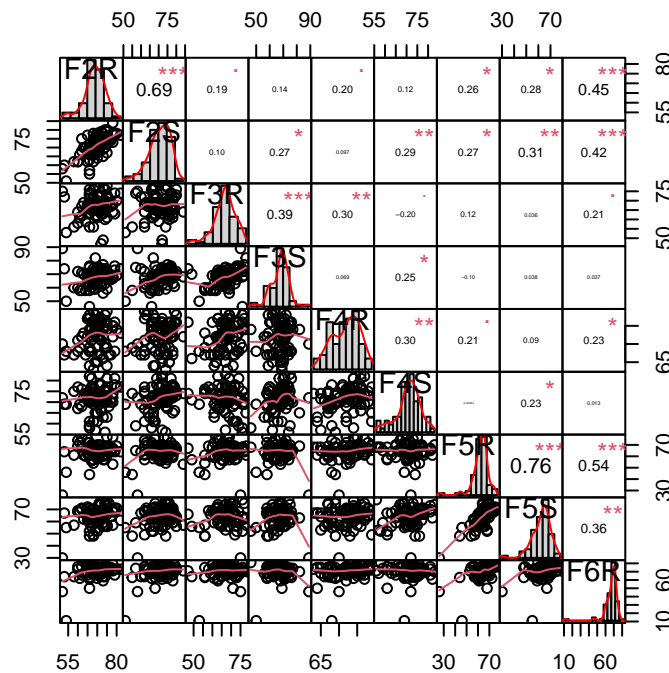
### PACIENTES CON TOMOGRAFÍAS NORMALES Y ANORMALES \*LUIS

```
pacientes <- SPECTF_train_IMP %>%
group_by(outcome) %>%
summarise(count = n()) %>%
mutate(perc = count/sum(count))
brks <- c(0, 0.2, 0.4, 0.6, 0.8, 1)
ggplot(pacientes, aes(x = outcome, y = perc, fill = factor(outcome))) +
geom_bar(stat="identity", width = 0.7) +
scale_y_continuous(breaks = brks, labels = scales::percent(brks)) +
labs(x = "Diagnóstico", y = "Porcentaje", fill = "Diagnóstico") +
theme_minimal(base_size = 14)
```



## CORRELACIÓN DE VARIABLES \*KLEYDERM

```
chart.Correlation(SPECTF_train_IMP[, c(3:11)], histogram = TRUE, col="grey10", pch=1, main="Cancer Mean".
```



## CREAMOS MODELO DE REGRESIÓN #1 \*LUIS

```
model1 <- glm(outcome ~.,family = binomial, data=
c.entrenamiento )
```

## VALORES ESTADISTICOS DEL MODEL1 \*KLEYDERM

```
summary(model1)
```

```
##
## Call:
## glm(formula = outcome ~ ., family = binomial, data = c.entrenamiento)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -9.907e-06 -2.802e-06  0.000e+00  8.575e-07  1.009e-05
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  9.164e+02  8.186e+06      0      1
## F1R          8.954e-01  5.842e+04      0      1
## F1S         -2.185e+00  5.548e+04      0      1
## F2R          1.253e+00  4.498e+04      0      1
## F2S         -2.888e+00  5.378e+04      0      1
## F3R         -3.092e+00  3.807e+04      0      1
## F3S          1.957e+00  4.676e+04      0      1
## F4R         -4.012e+00  7.040e+04      0      1
## F4S          2.289e+00  4.565e+04      0      1
```

```
## F5R      8.735e+00  6.357e+04      0      1
## F5S     -4.607e-01  9.009e+04      0      1
## F6R      3.223e+00  6.287e+04      0      1
## F6S     -2.416e+00  2.051e+04      0      1
## F7R      3.847e+00  3.998e+04      0      1
## F7S     -5.155e-01  2.460e+04      0      1
## F8R     -2.147e+00  5.355e+04      0      1
## F8S      6.710e+00  7.257e+04      0      1
## F9R     -7.334e+00  5.770e+04      0      1
## F9S     -4.773e+00  3.283e+04      0      1
## F10R     -4.170e+00  6.328e+04      0      1
## F10S     -3.990e-01  5.459e+04      0      1
## F11R      2.030e+00  2.985e+04      0      1
## F11S     -1.802e+00  3.320e+04      0      1
## F12R      3.235e+00  5.133e+04      0      1
## F12S      1.972e+00  2.311e+04      0      1
## F13R      1.092e+00  5.224e+04      0      1
## F13S     -7.317e+00  1.003e+05      0      1
## F14R      3.692e+00  2.842e+04      0      1
## F14S     -1.285e+00  4.854e+04      0      1
## F15R     -2.927e+00  4.647e+04      0      1
## F15S     -6.162e-01  3.761e+04      0      1
## F16R     -5.179e+00  5.344e+04      0      1
## F16S     -2.910e+00  3.562e+04      0      1
## F17R     -3.933e+00  6.468e+04      0      1
## F17S      1.255e+00  3.109e+04      0      1
## F18R      4.469e+00  1.421e+05      0      1
## F18S      4.215e+00  2.169e+04      0      1
## F19R     -2.914e+00  3.747e+04      0      1
## F19S      2.238e+00  3.562e+04      0      1
## F20R      1.635e+00  3.242e+04      0      1
## F20S     -4.338e+00  6.164e+04      0      1
## F21R     -5.096e-01  3.273e+04      0      1
## F21S     -1.489e+00  3.709e+04      0      1
## F22R     -2.442e+00  5.687e+04      0      1
## F22S      5.051e+00  2.499e+04      0      1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1.1090e+02  on 79  degrees of freedom
## Residual deviance: 1.8485e-09  on 35  degrees of freedom
## AIC: 90
##
## Number of Fisher Scoring iterations: 25
```

## INGRESAMOS DATA SET DE VALIDACIÓN \*LUIS

```
c.validacion <- SPECTF_test[,-1]
```

## REALIZAMOS PRUEBAS AL MODEL1 #KLEYDERM

```
pred1 <- predict(model1, newdata = c.validacion , type = "response")
pred1
```



##	1	2	3	4	5	6
##	1.180441e-05	9.999999e-01	9.999376e-01	1.000000e+00	1.000000e+00	1.000000e+00
##	7	8	9	10	11	12
##	1.000000e+00	7.576507e-02	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
##	13	14	15	16	17	18
##	1.000000e+00	1.000000e+00	2.220446e-16	1.000000e+00	1.000000e+00	1.000000e+00
##	19	20	21	22	23	24
##	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
##	25	26	27	28	29	30
##	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	2.220446e-16	1.000000e+00
##	31	32	33	34	35	36
##	3.488692e-06	2.220446e-16	3.046406e-11	1.000000e+00	2.220446e-16	1.000000e+00
##	37	38	39	40	41	42
##	1.000000e+00	2.220446e-16	1.000000e+00	1.000000e+00	1.000000e+00	4.190128e-09
##	43	44	45	46	47	48
##	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	2.220446e-16	1.000000e+00
##	49	50	51	52	53	54
##	7.263778e-01	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
##	55	56	57	58	59	60
##	2.220446e-16	1.000000e+00	1.000000e+00	1.000000e+00	1.279208e-06	1.000000e+00
##	61	62	63	64	65	66
##	1.000000e+00	1.000000e+00	1.129737e-05	1.000000e+00	4.328336e-08	1.000000e+00
##	67	68	69	70	71	72
##	1.000000e+00	1.000000e+00	1.000000e+00	2.220446e-16	1.000000e+00	1.000000e+00
##	73	74	75	76	77	78
##	4.599124e-13	2.220446e-16	9.999524e-01	2.220446e-16	1.000000e+00	9.999961e-01
##	79	80	81	82	83	84
##	1.000000e+00	1.000000e+00	2.220446e-16	1.000000e+00	1.000000e+00	1.000000e+00
##	85	86	87	88	89	90
##	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00	1.000000e+00
##	91	92	93	94	95	96
##	1.000000e+00	1.000000e+00	2.220446e-16	2.220446e-16	2.220446e-16	1.000000e+00
##	97	98	99	100	101	102
##	1.000000e+00	2.220446e-16	1.000000e+00	2.220446e-16	9.998474e-01	4.896335e-08
##	103	104	105	106	107	108
##	1.000000e+00	9.999070e-01	2.220446e-16	9.999998e-01	1.000000e+00	2.220446e-16
##	109	110	111	112	113	114
##	1.000000e+00	3.660979e-07	9.999813e-01	1.000000e+00	1.000000e+00	1.000000e+00
##	115	116	117	118	119	120
##	2.576465e-10	5.847889e-12	1.000000e+00	1.222587e-13	1.000000e+00	7.777679e-03
##	121	122	123	124	125	126
##	1.000000e+00	1.000000e+00	2.220446e-16	2.220446e-16	2.220446e-16	2.220446e-16
##	127	128	129	130	131	132
##	1.000000e+00	1.000000e+00	1.000000e+00	2.220446e-16	1.000000e+00	1.000000e+00
##	133	134	135	136	137	138
##	1.000000e+00	1.000000e+00	1.000000e+00	4.759546e-04	1.000000e+00	1.000000e+00
##	139	140	141	142	143	144
##	1.000000e+00	2.220446e-16	2.220446e-16	1.000000e+00	1.000000e+00	1.000000e+00
##	145	146	147	148	149	150
##	1.000000e+00	2.220446e-16	1.000000e+00	1.000000e+00	2.220446e-16	1.000000e+00
##	151	152	153	154	155	156
##	2.220446e-16	1.000000e+00	8.168194e-01	1.000000e+00	7.124894e-10	2.220446e-16
##	157	158	159	160	161	162
##	2.220446e-16	1.000000e+00	1.000000e+00	2.220446e-16	5.040419e-01	1.000000e+00

```
##          163          164          165          166          167          168
## 2.220446e-16 1.000000e+00 2.220446e-16 1.000000e+00 2.220446e-16 1.000000e+00
##          169          170          171          172          173          174
## 6.511919e-11 1.000000e+00 1.000000e+00 1.000000e+00 9.999998e-01 2.220446e-16
##          175          176          177          178          179          180
## 1.000000e+00 3.577422e-05 1.000000e+00 1.000000e+00 2.220446e-16 1.000000e+00
##          181          182          183          184          185          186
## 1.000000e+00 2.220446e-16 1.000000e+00 3.140278e-08 3.072105e-05 9.847697e-01
##          187
## 2.478453e-11
```

## REALIZAMOS EL PUNTO DE CORTE Y CONVERTIMOS LA PREDICCIÓN A FACTOR

\*LUIS

```
predic1<- ifelse(pred1 > 0.5, "NORMAL", "ANORMAL")
predic1
```

```
##          1          2          3          4          5          6          7          8
## "ANORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "ANORMAL"
##          9         10         11         12         13         14         15         16
## "NORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "ANORMAL" "NORMAL"
##         17         18         19         20         21         22         23         24
## "NORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL"
##         25         26         27         28         29         30         31         32
## "NORMAL" "NORMAL" "NORMAL" "NORMAL" "ANORMAL" "NORMAL" "ANORMAL" "ANORMAL"
##         33         34         35         36         37         38         39         40
## "ANORMAL" "NORMAL" "ANORMAL" "NORMAL" "NORMAL" "ANORMAL" "NORMAL" "NORMAL"
##         41         42         43         44         45         46         47         48
## "NORMAL" "ANORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "ANORMAL" "NORMAL"
##         49         50         51         52         53         54         55         56
## "NORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "ANORMAL" "NORMAL"
##         57         58         59         60         61         62         63         64
## "NORMAL" "NORMAL" "ANORMAL" "NORMAL" "NORMAL" "NORMAL" "ANORMAL" "NORMAL"
##         65         66         67         68         69         70         71         72
## "ANORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "ANORMAL" "NORMAL" "NORMAL"
##         73         74         75         76         77         78         79         80
## "ANORMAL" "ANORMAL" "NORMAL" "ANORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL"
##         81         82         83         84         85         86         87         88
## "ANORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL"
##         89         90         91         92         93         94         95         96
## "NORMAL" "NORMAL" "NORMAL" "NORMAL" "ANORMAL" "ANORMAL" "ANORMAL" "NORMAL"
##         97         98         99        100        101        102        103        104
## "NORMAL" "ANORMAL" "NORMAL" "ANORMAL" "NORMAL" "ANORMAL" "NORMAL" "NORMAL"
##        105        106        107        108        109        110        111        112
## "ANORMAL" "NORMAL" "NORMAL" "ANORMAL" "NORMAL" "ANORMAL" "NORMAL" "NORMAL"
##        113        114        115        116        117        118        119        120
## "NORMAL" "NORMAL" "ANORMAL" "ANORMAL" "NORMAL" "ANORMAL" "NORMAL" "ANORMAL"
##        121        122        123        124        125        126        127        128
## "NORMAL" "NORMAL" "ANORMAL" "ANORMAL" "ANORMAL" "ANORMAL" "NORMAL" "NORMAL"
##        129        130        131        132        133        134        135        136
## "NORMAL" "ANORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "ANORMAL"
##        137        138        139        140        141        142        143        144
```

```
## "NORMAL" "NORMAL" "NORMAL" "ANORMAL" "ANORMAL" "NORMAL" "NORMAL" "NORMAL"
##      145      146      147      148      149      150      151      152
## "NORMAL" "ANORMAL" "NORMAL" "NORMAL" "ANORMAL" "NORMAL" "ANORMAL" "NORMAL"
##      153      154      155      156      157      158      159      160
## "NORMAL" "NORMAL" "ANORMAL" "ANORMAL" "ANORMAL" "NORMAL" "NORMAL" "ANORMAL"
##      161      162      163      164      165      166      167      168
## "NORMAL" "NORMAL" "ANORMAL" "NORMAL" "ANORMAL" "NORMAL" "ANORMAL" "NORMAL"
##      169      170      171      172      173      174      175      176
## "ANORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "ANORMAL" "NORMAL" "ANORMAL"
##      177      178      179      180      181      182      183      184
## "NORMAL" "NORMAL" "ANORMAL" "NORMAL" "NORMAL" "ANORMAL" "NORMAL" "ANORMAL"
##      185      186      187
## "ANORMAL" "NORMAL" "ANORMAL"
```

**CAMBIAMOS EL OVERALL DEL DATA SET TEST POR OUTCOME Y LO CONVERTIMOS A FACTOR DE CARACTER. \*KLEYDERM**

```
SPECTF_test <-
  mutate(SPECTF_test , outcome = (as.factor(
    SPECTF_test$OVERALL_DIAGNOSIS)))
SPECTF_test <- SPECTF_test[,-1]
SPECTF_test$outcome<- as.character.factor(SPECTF_test$outcome)
SPECTF_test$outcome <-
  replace(SPECTF_test$outcome,
    SPECTF_test$outcome == "1" , "ANORMAL")
SPECTF_test$outcome <-
  replace(SPECTF_test$outcome,
    SPECTF_test$outcome == "0" , "NORMAL")
SPECTF_test$outcome[170:186]
```

```
## [1] "ANORMAL" "ANORMAL" "ANORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL"
## [8] "NORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL" "NORMAL"
## [15] "NORMAL" "NORMAL" "NORMAL"
```

**IGUALAMOS AMBOS OUTCOMES, ANTES Y DESPUÉS DE LA PREDICCIÓN \*LUIS**

```
real <- factor(SPECTF_test$outcome)
prediction1 <- data.frame(predic1,real)
prediction1[1:10,]
```

```
##      predic1    real
## 1  ANORMAL ANORMAL
## 2   NORMAL ANORMAL
## 3   NORMAL ANORMAL
## 4   NORMAL ANORMAL
## 5   NORMAL ANORMAL
## 6   NORMAL ANORMAL
## 7   NORMAL ANORMAL
## 8  ANORMAL ANORMAL
## 9   NORMAL ANORMAL
## 10  NORMAL ANORMAL
```

## EVALUAMOS LA CANTIDAD DE ACIERTOS QUE TUVO LA PREDIC- TION1 \*KLEYDERM

```
mean(predic1 == real)
```

```
## [1] 0.3208556
```

NUESTRO PRIMER MODELO OBTIENE UN PORCENTAJE DE ACERTACIÓN DEL 32%

## CREAMOS UN NUEVA REGRESIÓN (MODEL2) CON LAS VARIBALES MÁS RELEVANTES \*LUIS

```
model2 <- glm(outcome ~ F2R + F4S #40%  
              + F20R,  
              family = binomial,data=c.entrenamiento )
```

## VALORES ESTADISTICOS DEL MODEL2 \*KLEYDERM

```
summary(model2)
```

```
##  
## Call:  
## glm(formula = outcome ~ F2R + F4S + F20R, family = binomial,  
##      data = c.entrenamiento)  
##  
## Deviance Residuals:  
##      Min       1Q   Median       3Q      Max  
## -2.3887  -0.9351  -0.1862   0.8344   2.2399  
##  
## Coefficients:  
##              Estimate Std. Error z value Pr(>|z|)  
## (Intercept)  26.47541    6.95378   3.807  0.00014 ***  
## F2R          -0.08981    0.05327  -1.686  0.09179 .  
## F4S          -0.12354    0.05138  -2.404  0.01620 *  
## F20R         -0.16246    0.05029  -3.231  0.00123 **  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## (Dispersion parameter for binomial family taken to be 1)  
##  
##      Null deviance: 110.904  on 79  degrees of freedom  
## Residual deviance:  86.093  on 76  degrees of freedom  
## AIC: 94.093  
##  
## Number of Fisher Scoring iterations: 5
```

## REALIZAMOS PRUEBAS AL MODEL2 \*LUIS

```
pred2 <- predict(model2, newdata = c.validacion , type = "response")  
pred2
```

```
##           1           2           3           4           5           6           7  
## 0.61886724 0.72760238 0.64694710 0.36977373 0.54501561 0.27962720 0.64238745
```

##	8	9	10	11	12	13	14
##	0.22469163	0.99996347	0.24576714	0.81279421	0.98070749	0.74125584	0.91844348
##	15	16	17	18	19	20	21
##	0.80794505	0.99998466	0.41833071	0.18362985	0.88887567	0.91882252	0.81167154
##	22	23	24	25	26	27	28
##	0.98948417	0.90237800	0.44877391	0.81129957	0.62374481	0.68061452	0.40447654
##	29	30	31	32	33	34	35
##	0.64351913	0.56129422	0.32507857	0.57797346	0.99999741	0.76962729	0.32167661
##	36	37	38	39	40	41	42
##	0.65367505	0.49010430	0.86495694	0.99999972	0.99977234	0.99999414	0.28160406
##	43	44	45	46	47	48	49
##	0.33048016	0.94097705	0.44991460	0.48765732	0.93279726	0.98217460	0.21461047
##	50	51	52	53	54	55	56
##	0.73614359	0.19567678	0.99998068	0.83224263	0.71263830	0.12623613	0.13255162
##	57	58	59	60	61	62	63
##	0.99969798	0.88931923	0.83683704	0.27142430	0.52824002	0.33531340	0.14971306
##	64	65	66	67	68	69	70
##	0.72490309	0.15398388	0.99999764	0.20660810	0.83246673	0.49036256	0.90152185
##	71	72	73	74	75	76	77
##	0.70342699	0.33762754	0.88634784	0.99998241	0.39365333	0.82276117	0.81303845
##	78	79	80	81	82	83	84
##	0.50268220	0.71520024	0.29425585	0.75755285	0.19947253	0.99486989	0.72593541
##	85	86	87	88	89	90	91
##	0.92535984	0.99998649	0.81263691	0.98266767	0.20407108	0.46371280	0.29284803
##	92	93	94	95	96	97	98
##	0.99994046	0.65072512	0.30097198	0.25970201	0.37740031	0.23236522	0.71301884
##	99	100	101	102	103	104	105
##	0.51655107	0.94245599	0.62921341	0.13012125	0.47911788	0.72435413	0.86544087
##	106	107	108	109	110	111	112
##	0.48052647	0.07859944	0.22567629	0.26612842	0.32543102	0.20011804	0.77517470
##	113	114	115	116	117	118	119
##	0.83783965	0.97321619	0.34170683	0.10787191	0.47366118	0.39996594	0.93201372
##	120	121	122	123	124	125	126
##	0.39938253	0.84767755	0.16997690	0.14885092	0.32293370	0.74076776	0.81888234
##	127	128	129	130	131	132	133
##	0.80535897	0.26147858	0.17545554	0.85450669	0.99862702	0.98658109	0.77447026
##	134	135	136	137	138	139	140
##	0.99364676	0.35287770	0.94268060	0.99993274	0.56129422	0.56692731	0.97785192
##	141	142	143	144	145	146	147
##	0.47584543	0.99999939	0.29318072	0.88805334	0.78881205	0.49983254	0.99839207
##	148	149	150	151	152	153	154
##	0.98670407	0.96055037	0.30252296	0.90161356	0.53380264	0.35088004	0.42377561
##	155	156	157	158	159	160	161
##	0.55015130	0.69665710	0.27131108	0.99974330	0.99963817	0.94646393	0.83648948
##	162	163	164	165	166	167	168
##	0.48895228	0.65786289	0.79625099	0.49943104	0.25003841	0.45930744	0.33160315
##	169	170	171	172	173	174	175
##	0.41388475	0.35810092	0.32205153	0.89531372	0.72086502	0.34417148	0.51255227
##	176	177	178	179	180	181	182
##	0.32555673	0.13849984	0.77745237	0.45194158	0.86861347	0.19921620	0.46396980
##	183	184	185	186	187		
##	0.21286853	0.79641859	0.13660073	0.13129901	0.30264378		

## REALIZAMOS EL PUNTO DE CORTE Y CONVERTIMOS LA PREDICCIÓN A FACTOR

\*KLEYDERM

```
predic2<- ifelse(pred2 > 0.5, "NORMAL", "ANORMAL")
predic2
```

##	1	2	3	4	5	6	7	8
##	"NORMAL"	"NORMAL"	"NORMAL"	"ANORMAL"	"NORMAL"	"ANORMAL"	"NORMAL"	"ANORMAL"
##	9	10	11	12	13	14	15	16
##	"NORMAL"	"ANORMAL"	"NORMAL"	"NORMAL"	"NORMAL"	"NORMAL"	"NORMAL"	"NORMAL"
##	17	18	19	20	21	22	23	24
##	"ANORMAL"	"ANORMAL"	"NORMAL"	"NORMAL"	"NORMAL"	"NORMAL"	"NORMAL"	"ANORMAL"
##	25	26	27	28	29	30	31	32
##	"NORMAL"	"NORMAL"	"NORMAL"	"ANORMAL"	"NORMAL"	"NORMAL"	"ANORMAL"	"NORMAL"
##	33	34	35	36	37	38	39	40
##	"NORMAL"	"NORMAL"	"ANORMAL"	"NORMAL"	"ANORMAL"	"NORMAL"	"NORMAL"	"NORMAL"
##	41	42	43	44	45	46	47	48
##	"NORMAL"	"ANORMAL"	"ANORMAL"	"NORMAL"	"ANORMAL"	"ANORMAL"	"NORMAL"	"NORMAL"
##	49	50	51	52	53	54	55	56
##	"ANORMAL"	"NORMAL"	"ANORMAL"	"NORMAL"	"NORMAL"	"NORMAL"	"ANORMAL"	"ANORMAL"
##	57	58	59	60	61	62	63	64
##	"NORMAL"	"NORMAL"	"NORMAL"	"ANORMAL"	"NORMAL"	"ANORMAL"	"ANORMAL"	"NORMAL"
##	65	66	67	68	69	70	71	72
##	"ANORMAL"	"NORMAL"	"ANORMAL"	"NORMAL"	"ANORMAL"	"NORMAL"	"NORMAL"	"ANORMAL"
##	73	74	75	76	77	78	79	80
##	"NORMAL"	"NORMAL"	"ANORMAL"	"NORMAL"	"NORMAL"	"NORMAL"	"NORMAL"	"ANORMAL"
##	81	82	83	84	85	86	87	88
##	"NORMAL"	"ANORMAL"	"NORMAL"	"NORMAL"	"NORMAL"	"NORMAL"	"NORMAL"	"NORMAL"
##	89	90	91	92	93	94	95	96
##	"ANORMAL"	"ANORMAL"	"ANORMAL"	"NORMAL"	"NORMAL"	"ANORMAL"	"ANORMAL"	"ANORMAL"
##	97	98	99	100	101	102	103	104
##	"ANORMAL"	"NORMAL"	"NORMAL"	"NORMAL"	"NORMAL"	"ANORMAL"	"ANORMAL"	"NORMAL"
##	105	106	107	108	109	110	111	112
##	"NORMAL"	"ANORMAL"	"ANORMAL"	"ANORMAL"	"ANORMAL"	"ANORMAL"	"ANORMAL"	"NORMAL"
##	113	114	115	116	117	118	119	120
##	"NORMAL"	"NORMAL"	"ANORMAL"	"ANORMAL"	"ANORMAL"	"ANORMAL"	"NORMAL"	"ANORMAL"
##	121	122	123	124	125	126	127	128
##	"NORMAL"	"ANORMAL"	"ANORMAL"	"ANORMAL"	"NORMAL"	"NORMAL"	"NORMAL"	"ANORMAL"
##	129	130	131	132	133	134	135	136
##	"ANORMAL"	"NORMAL"	"NORMAL"	"NORMAL"	"NORMAL"	"NORMAL"	"ANORMAL"	"NORMAL"
##	137	138	139	140	141	142	143	144
##	"NORMAL"	"NORMAL"	"NORMAL"	"NORMAL"	"ANORMAL"	"NORMAL"	"ANORMAL"	"NORMAL"
##	145	146	147	148	149	150	151	152
##	"NORMAL"	"ANORMAL"	"NORMAL"	"NORMAL"	"NORMAL"	"ANORMAL"	"NORMAL"	"NORMAL"
##	153	154	155	156	157	158	159	160
##	"ANORMAL"	"ANORMAL"	"NORMAL"	"NORMAL"	"ANORMAL"	"NORMAL"	"NORMAL"	"NORMAL"
##	161	162	163	164	165	166	167	168
##	"NORMAL"	"ANORMAL"	"NORMAL"	"NORMAL"	"ANORMAL"	"ANORMAL"	"ANORMAL"	"ANORMAL"
##	169	170	171	172	173	174	175	176
##	"ANORMAL"	"ANORMAL"	"ANORMAL"	"NORMAL"	"NORMAL"	"ANORMAL"	"NORMAL"	"ANORMAL"
##	177	178	179	180	181	182	183	184
##	"ANORMAL"	"NORMAL"	"ANORMAL"	"NORMAL"	"ANORMAL"	"ANORMAL"	"ANORMAL"	"NORMAL"

```
##          185          186          187
## "ANORMAL" "ANORMAL" "ANORMAL"
```

## IGUALAMOS AMBOS OUTCOMES, ANTES Y DESPUÉS DE LA PREDICCIÓN2 \*LUIS

```
prediction2 <- data.frame(predic2,real)
prediction2[1:10,]
```

```
##   predic2    real
## 1  NORMAL ANORMAL
## 2  NORMAL ANORMAL
## 3  NORMAL ANORMAL
## 4 ANORMAL ANORMAL
## 5  NORMAL ANORMAL
## 6 ANORMAL ANORMAL
## 7  NORMAL ANORMAL
## 8 ANORMAL ANORMAL
## 9  NORMAL ANORMAL
## 10 ANORMAL ANORMAL
```

## EVALUAMOS LA CANTIDAD DE ACIERTOS QUE TUVO LA PREDICCIÓN2 \*KLEYDERM

```
mean(predic2 == real)
```

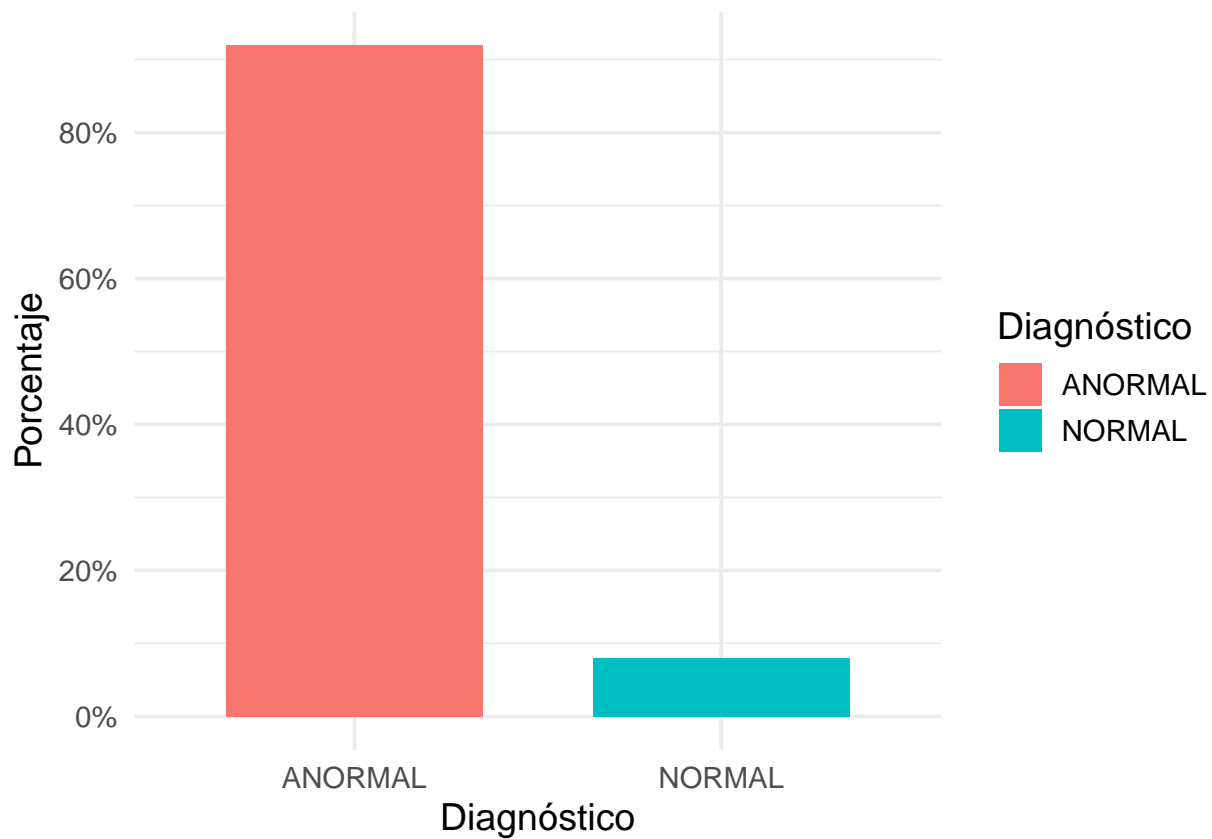
```
## [1] 0.4010695
```

NUESTRO SEGUNDO MODELO OBTIENE UN PORCENTAJE DE ACERTACIÓN DEL 40,15%

#CONCLUSIONES

## DIAGRAMA DE BARRAS DE DIAGNOSTICO GRUPO DE TEST \*LUIS

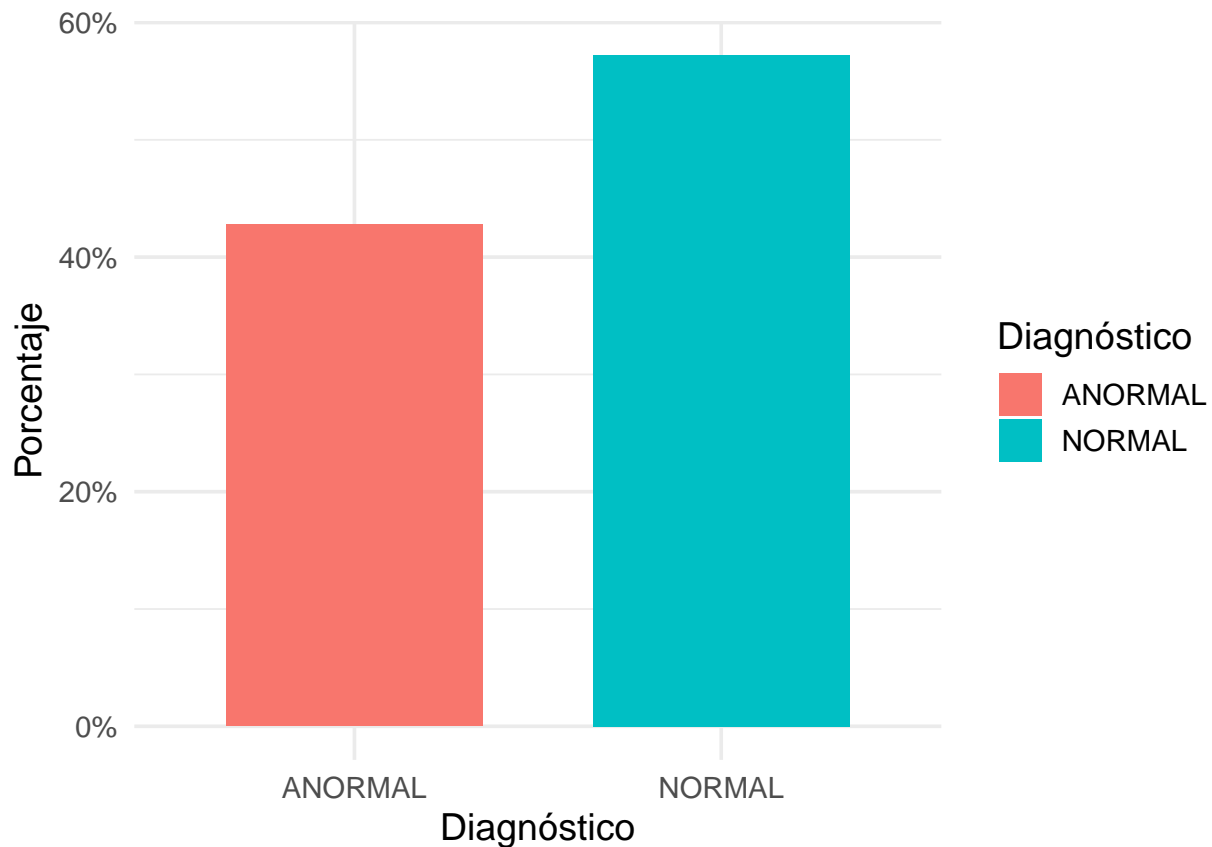
```
pacientes <- SPECTF_test %>%
group_by(outcome) %>%
summarise(count = n()) %>%
mutate(perc = count/sum(count))
brks <- c(0, 0.2, 0.4, 0.6, 0.8, 1)
ggplot(pacientes, aes(x = outcome, y = perc, fill = factor(outcome))) +
geom_bar(stat="identity", width = 0.7) +
scale_y_continuous(breaks = brks, labels = scales::percent(brks)) +
labs(x = "Diagnóstico", y = "Porcentaje", fill = "Diagnóstico") +
theme_minimal(base_size = 14)
```



**DIAGRAMA DE BARRAS DE DIAGNOSTICO MODELO DE REGRESIÓN CON ACERTACIÓN MÁS ALTO. \*KLEYDERM**

```
pacientes <- prediction2 %>%
group_by(predic2) %>%
summarise(count = n()) %>%
mutate(perc = count/sum(count))
brks <- c(0, 0.2, 0.4, 0.6, 0.8, 1)
ggplot(pacientes, aes(x = predic2, y = perc, fill = factor(predic2))) +
geom_bar(stat="identity", width = 0.7) +
scale_y_continuous(breaks = brks, labels = scales::percent(brks)) +
labs(x = "Diagnóstico", y = "Porcentaje", fill = "Diagnóstico") +
theme_minimal(base_size = 14)
```





## PREGUNTA DE INVESTIGACIÓN \*LUIS

¿Se puede realizar una integración entre visión artificial y Machine learning para lograr establecer si un paciente tiene o no una anomalía cardíaca solo con hacer una tomografía en reposo y en estrés, luego pasarla por un procesamiento de Rois y que el programa de Machine Learning estudie estos 44 resultados y de un diagnóstico?

## DATOS OBTENIDOS EN EL PROCESO

### DATOS OBTENIDOS PARA MODELO #1 \*KLEYDERM

```
predic1 <- factor(predic1)
real <- factor(real)
confusionMatrix(predic1,real, positive = "NORMAL")
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction ANORMAL NORMAL
##   ANORMAL      52      7
##   NORMAL      120      8
##
##           Accuracy : 0.3209
##           95% CI : (0.2546, 0.3929)
##   No Information Rate : 0.9198
##   P-Value [Acc > NIR] : 1
##
```

```
##           Kappa : -0.037
##
## Mcnemar's Test P-Value : <2e-16
##
##           Sensitivity : 0.53333
##           Specificity : 0.30233
##           Pos Pred Value : 0.06250
##           Neg Pred Value : 0.88136
##           Prevalence : 0.08021
##           Detection Rate : 0.04278
##           Detection Prevalence : 0.68449
##           Balanced Accuracy : 0.41783
##
##           'Positive' Class : NORMAL
##
```

## DATOS OBTENIDOS PARA MODELO #2 \*LUIS

```
predic2 <- factor(predic2)
real <- factor(real)
confusionMatrix(predic2,real, positive = "NORMAL")
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction ANORMAL NORMAL
##   ANORMAL      70      10
##   NORMAL      102       5
##
##           Accuracy : 0.4011
##           95% CI : (0.3302, 0.4751)
##   No Information Rate : 0.9198
##   P-Value [Acc > NIR] : 1
##
##           Kappa : -0.0684
##
## Mcnemar's Test P-Value : <2e-16
##
##           Sensitivity : 0.33333
##           Specificity : 0.40698
##           Pos Pred Value : 0.04673
##           Neg Pred Value : 0.87500
##           Prevalence : 0.08021
##           Detection Rate : 0.02674
##           Detection Prevalence : 0.57219
##           Balanced Accuracy : 0.37016
##
##           'Positive' Class : NORMAL
##
```

## ¿SON SUS HALAZGOS LO QUE ESPERABA? ¿POR QUÉ O POR QUÉ NO? \*KLEYDERM

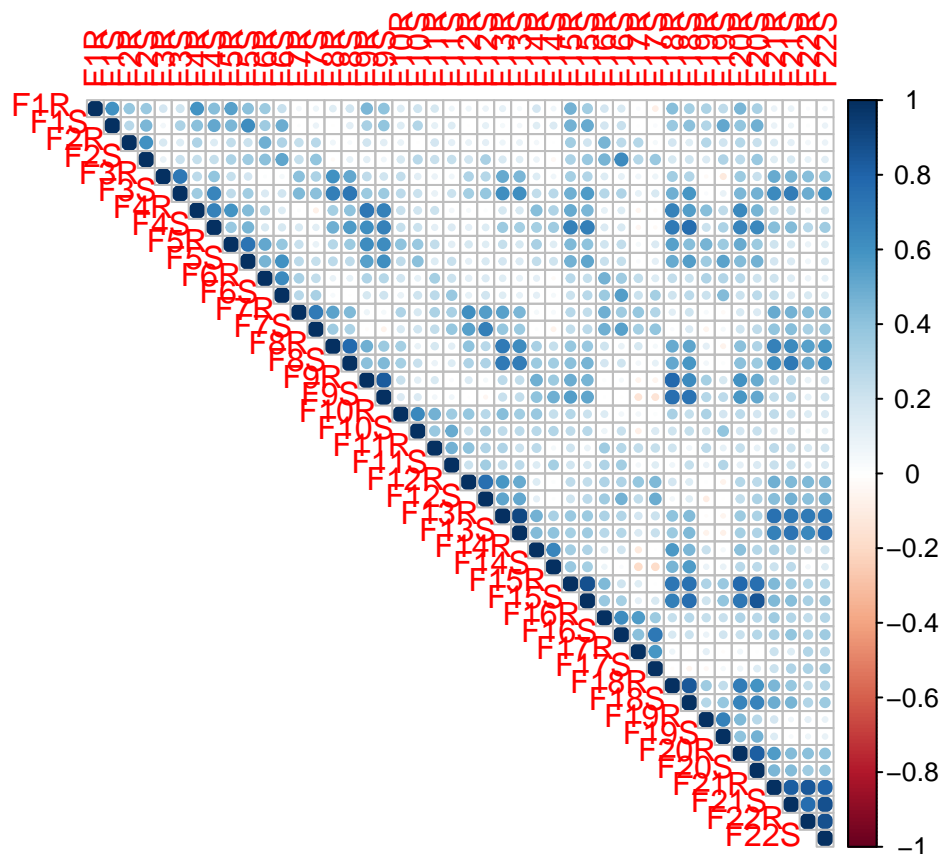
Los hallazgos no fueron los esperados, debido a que el porcentaje de acierto de nuestro modelo no superó el 50%.

Al ser muchas variables, éstas se relacionan entre ellas de manera muy aleatoria, y al momento de decidirnos por cual mantener en la regresión se complica, lo que nos permite determinar que a partir de un modelo predictivo que reciba resultados de una cámara de visión artificial NO se puede determinar si la tomografía de un paciente es anormal o no sin necesidad de la presencia de un especialista.

Las variables muy pocas veces dependerán de alguna otra, pues recordemos que cada variable contiene el porcentaje de igualación de una imagen en condiciones sanas con la imagen tomada al paciente en cierta zona del cuerpo, es por ello que entre variables es difícil que haya relación pues el resultado de un ROI es totalmente diferente al de otro órgano.

## CORRELACIÓN DE LAS 44 VARIABLES. \*LUIS

```
corrplot(cor(c.validacion), type="upper")
```



## GIT HUB

ENLACE PÚBLICO: [https://github.com/luisp03/DataScience\\_UAC.git](https://github.com/luisp03/DataScience_UAC.git)