MODELADO FASE REM DEL SUEÑO

ANALÍTICA DE DATOS EN SALUD

GUILLEM MIRALLES GADEA

CONTEXTO

El **sueño se divide en 5 ciclos.**

- NOREM: N1, N2, N3 y N4
- **REM**

Nuestro trabajo trata de detectar la fase **REM** (rapid eye movements):

- Movimiento rápido de los ojos, frecuencias respiratorias y cardíacas fluctuantes, fluctuación de la tensión arterial
- Donde soñamos y captamos gran cantidad de información de nuestro entorno
- 25% del ciclo del sueño.

3 pacientes:

- Paciente **9 y 10** -> **Train**
- Paciente **8** -> **Test**

DIVISIÓN DEL TRABAJO:

1. Lectura y transformación de características
2. Análisis Exploratorio
3. Modelos
4. Interpretabilidad
5. Conclusiones

1. LECTURA Y TRANSFORMACIÓN DE CARACTERÍSTICAS

1. LECTURA Y TRANSFORMACIÓN DE LOS DATOS

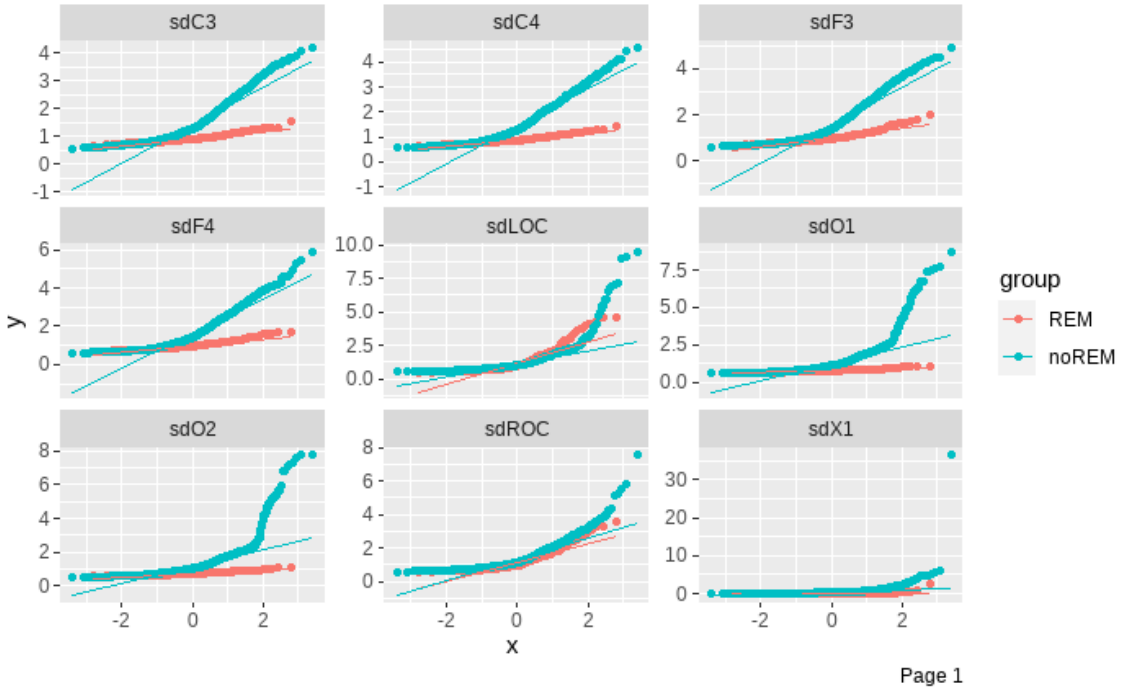
- Datos disponibles de **11 canales** (6 EEG, 2 EOG, 3EMG).
- De cada 30s de canal obtenemos **3 estadísticos** que estimen esta fase
  - **Mean** – Media.
  - **Sd** – Desviación Estándar.
  - **dfreq** - Frecuencia dominante.

meanF3	sdF3	dfreqF3	meanC3	sdC3	dfreqC3	meanO1	sdO1	dfreqO1	meanF4	sdF4	dfreqF4
-3.680505e-03	1.7943498	0.0004666667	-4.128918e-03	1.3382350	0.0012333333	-0.0043696121	0.9566125	0.0104333333	0.0091383686	1.8725329	0.0004666667
9.208585e-04	1.5704560	0.0004666667	2.980247e-04	0.9636163	0.0004666667	-0.0034295512	0.7050007	0.0006000000	0.0106487931	1.7833156	0.0004666667
6.043920e-03	1.1775754	0.0099666667	4.631620e-03	0.8965963	0.0099666667	-0.0046627210	0.7856079	0.0099666667	0.0053711422	1.1495238	0.0099666667
-1.935051e-03	1.0275330	0.0007333333	-1.590550e-03	0.8197547	0.0007333333	0.0124859202	0.6453957	0.0007666667	-0.0160158457	1.2482475	0.0005333333
9.853435e-03	1.0929176	0.0011666667	5.039361e-03	0.7970588	0.0099000000	-0.0091899388	0.7377766	0.0011333333	0.0200409434	0.9521512	0.0012000000
4.429656e-03	1.0302643	0.0009000000	-9.897055e-04	0.7857381	0.0099333333	-0.0060678283	0.7004962	0.0100333333	-0.0014294591	0.9120877	0.0010666667

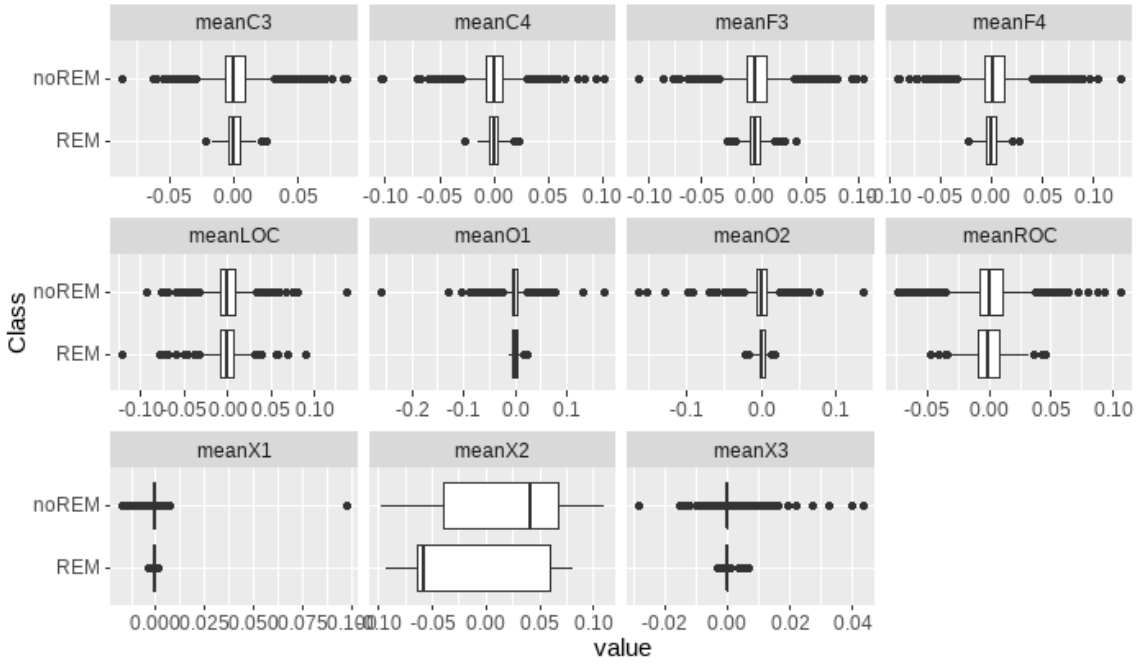
2. ANÁLISIS EXPLORATORIO

2. ANÁLISIS EXPLORATORIO

QQ – Plot Desviación Típica

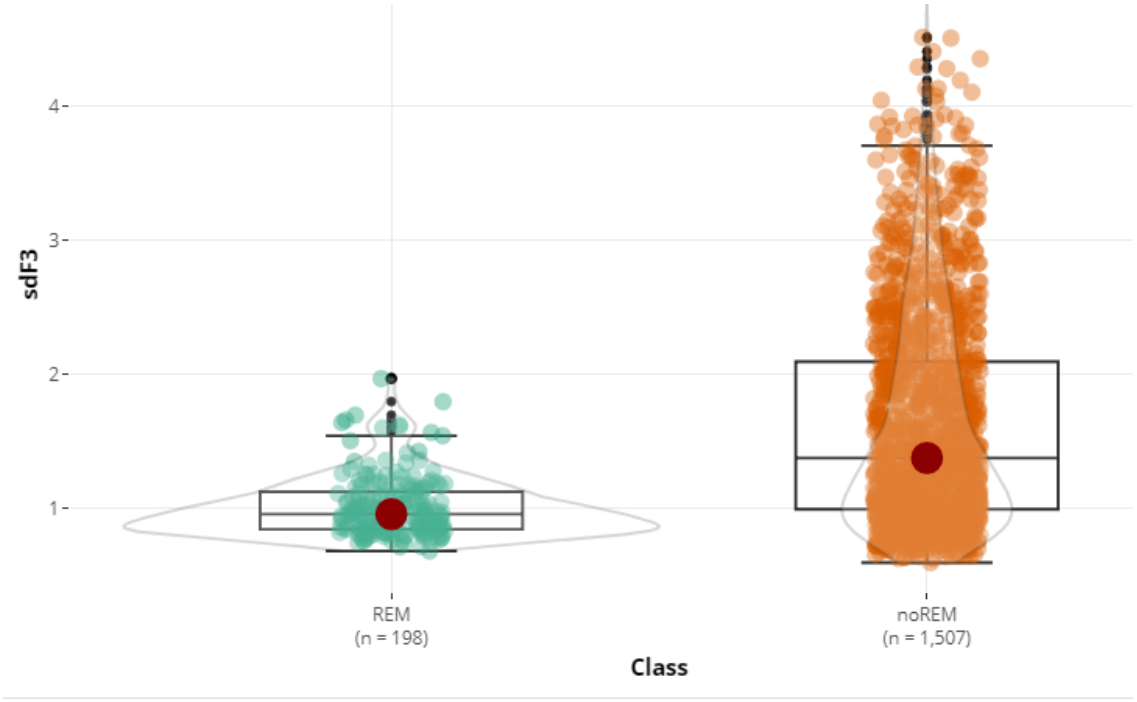


Boxplot de la Media

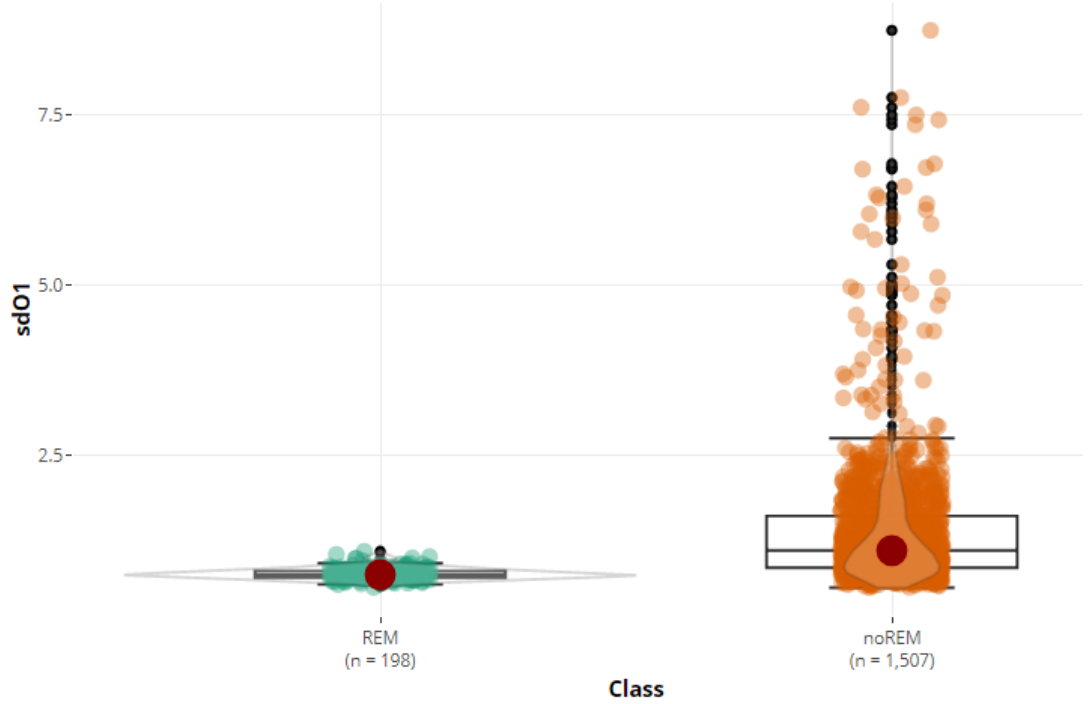


2. ANÁLISIS EXPLORATORIO

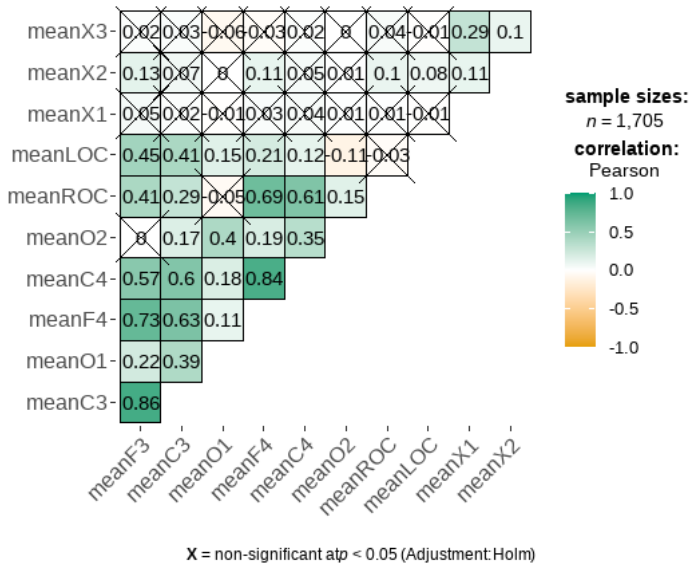
Violin plot- sdF3

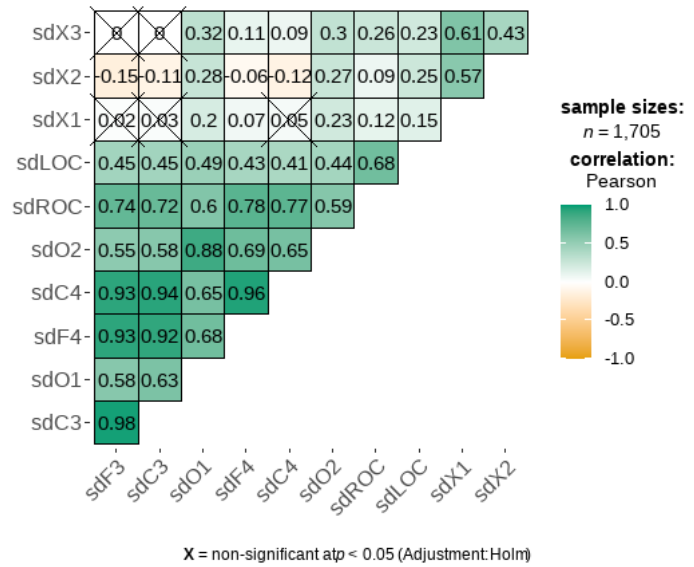


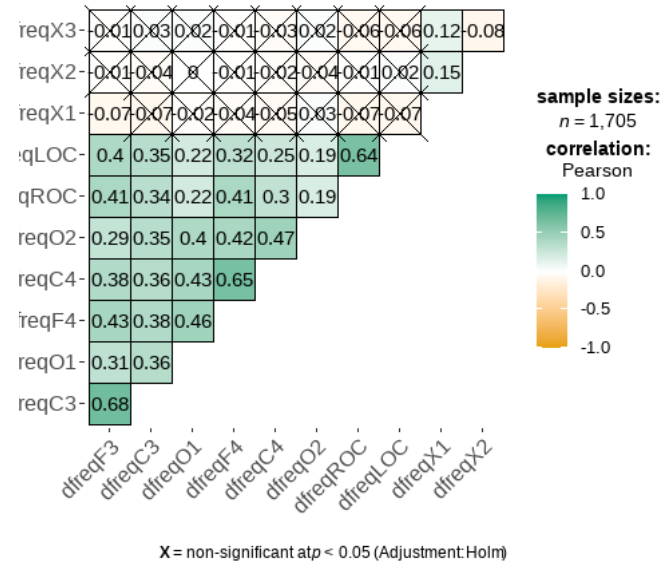
Violin plot- sdO1

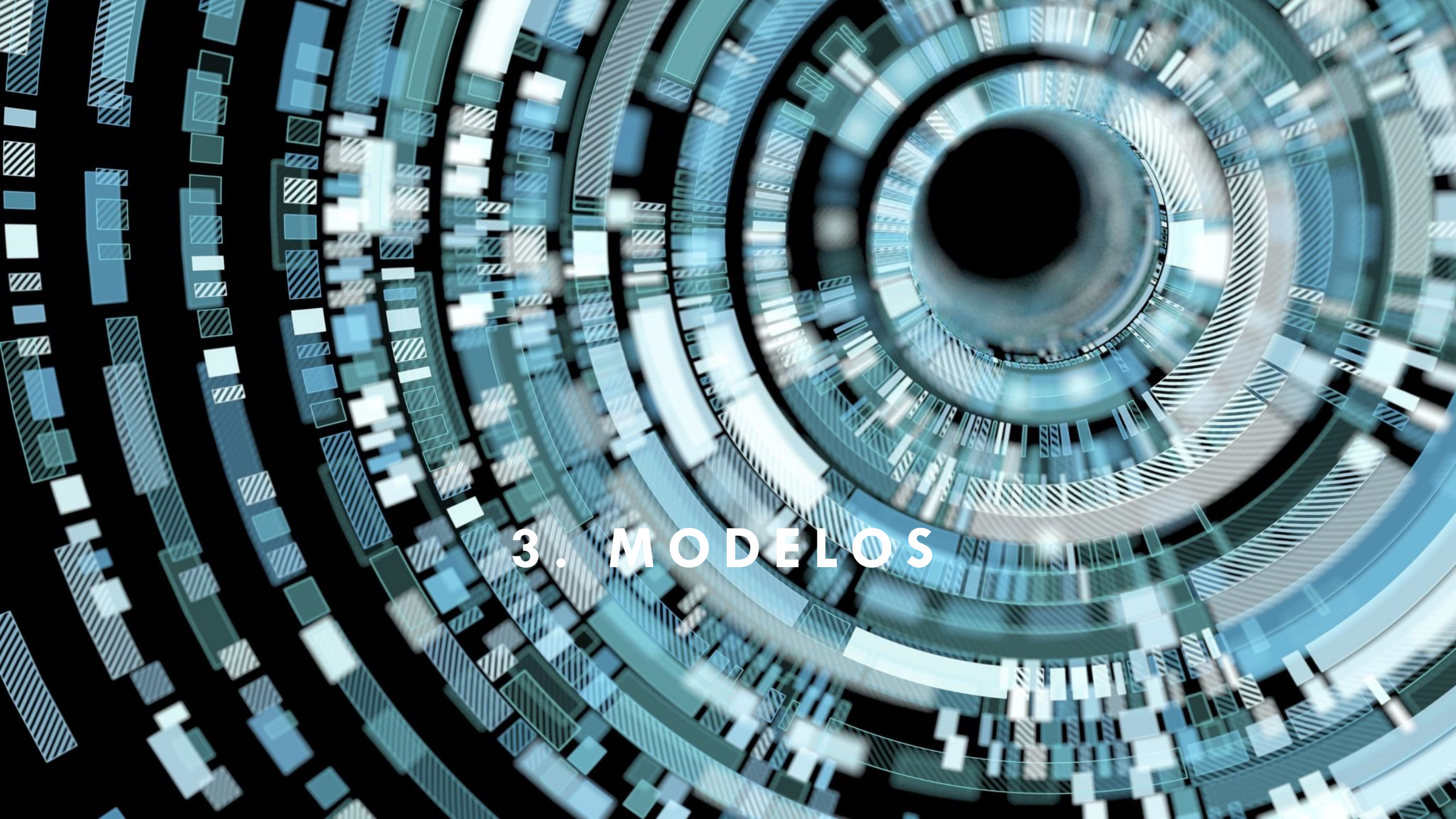


2. ANÁLISIS EXPLORATORIO



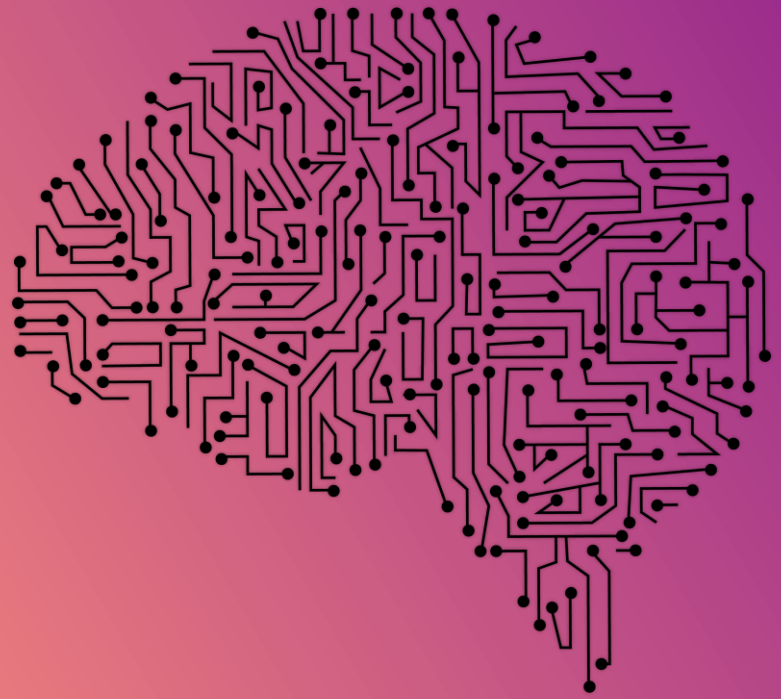





3. MODELOS

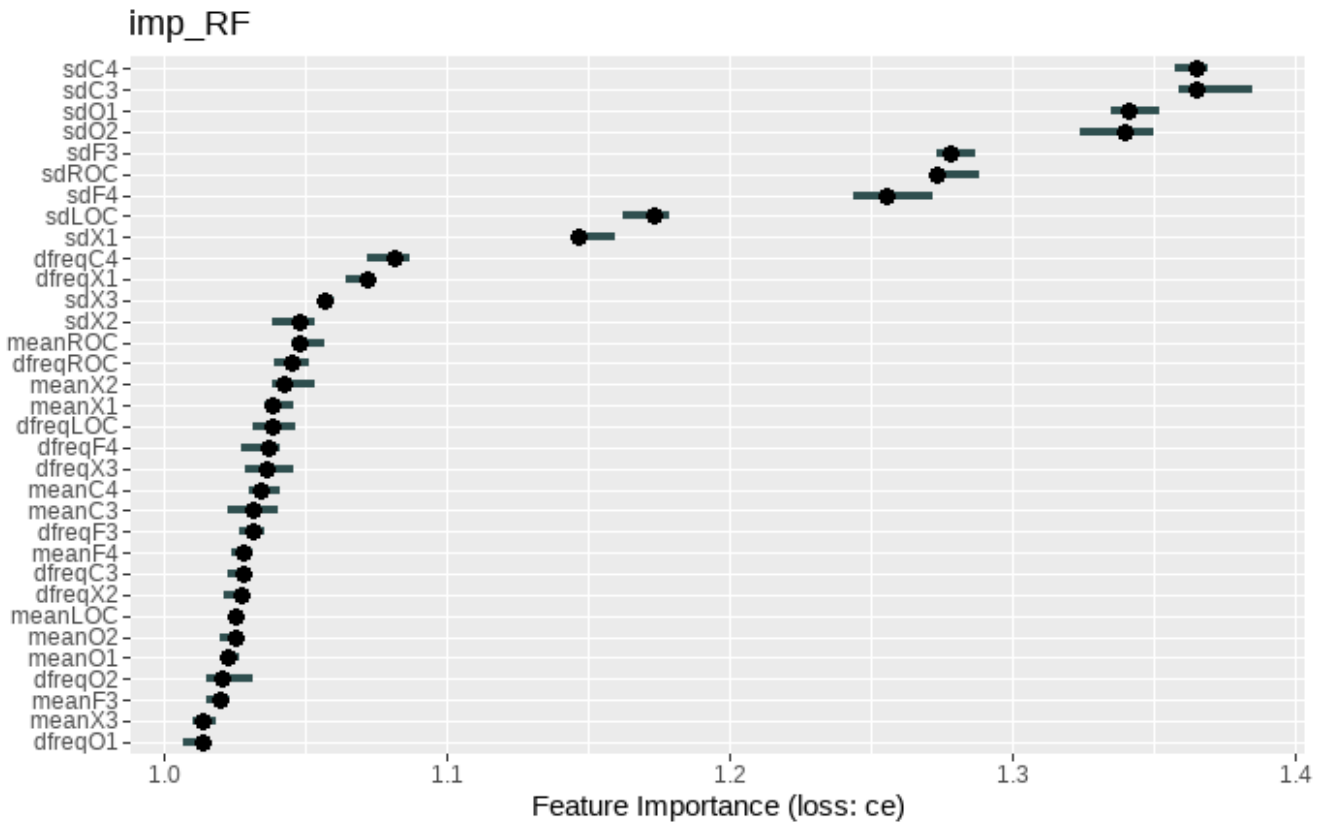
3. MODELOS

- Construimos los siguientes modelos:
  - **GLM**, con los parámetros que mejor estiman el modelo.
  - **Random Forest**, encontrando los hiperparámetros.
  - **SVM**, ajustando los hiperparámetros.
  - **Modelo TAN** , con discretización y clases balanceadas.
  - **CBA Classifier**, con datos discretizados.



ESTRATÉGIA RANDOM FOREST

- **Construimos el RF .**
- Ajustamos los parámetros y vemos la importancia de las variables.
- Volvemos a construir dos modelos:
  - **RF_1**: Con las 14 primeras variabes más importantes.
  - **RF_2**: Con las variables que representan la desviación típica.



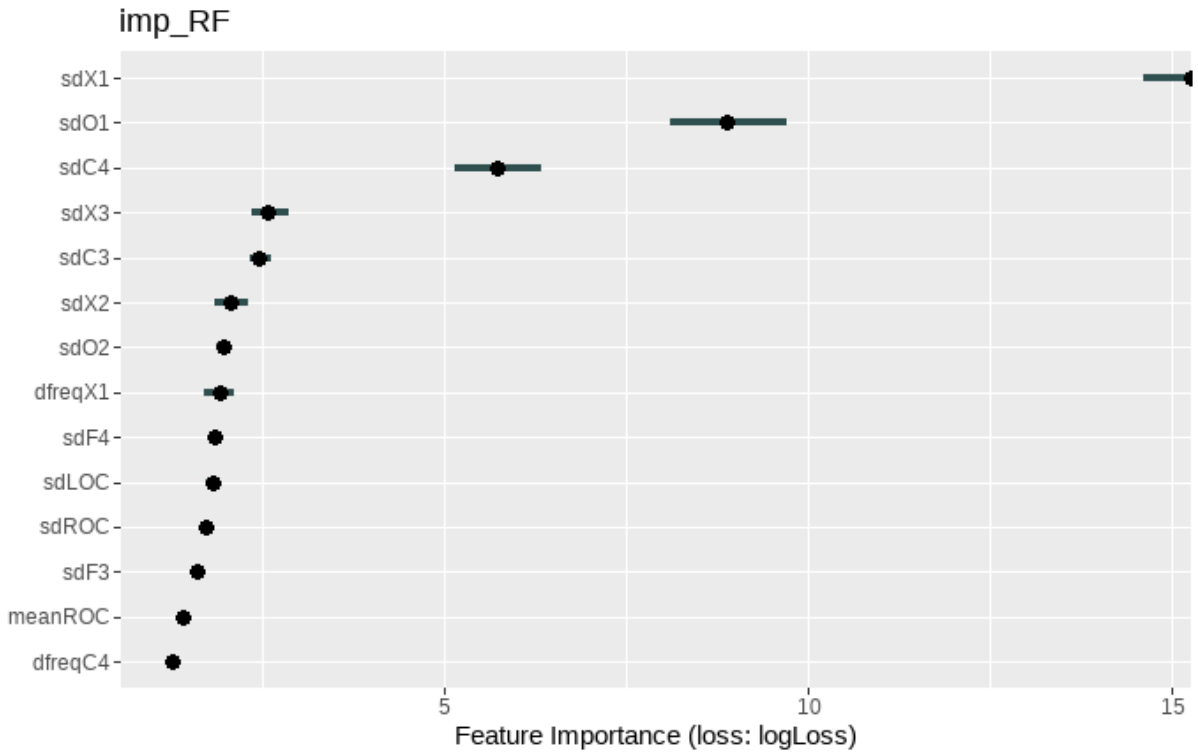
RESULTADOS DE LOS MODELOS

	Accuracy	*PPV*	*Sensitividad*	*Sensibilidad*	*Balanced - Acc*
GLM	0.9041	0.6333	0.8926	0.9062	0.8994
Random Forest 1	0.9268	0.7010	0.9128	0.9294	0.9211
Random Forest 2	0.9247	0.6959	0.9060	0.9281	0.9171
SVM	0.9268	0.85455	0.63087	0.98051	0.80569
TAN	0.8649	0.5326	0.8658	0.8648	0.8653
CBA	0.9392	0.9167	0.6644	0.9890	0.8267

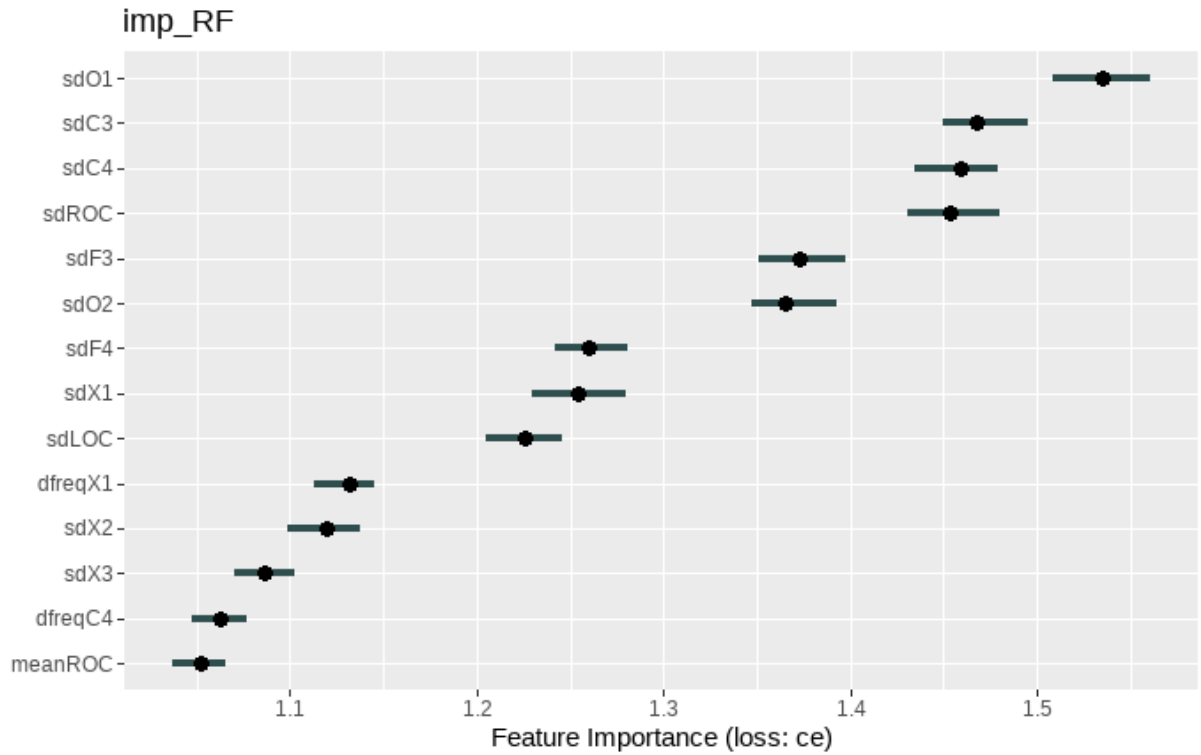
4. INTERPRETABILIDAD

INTERPRETABILIDAD

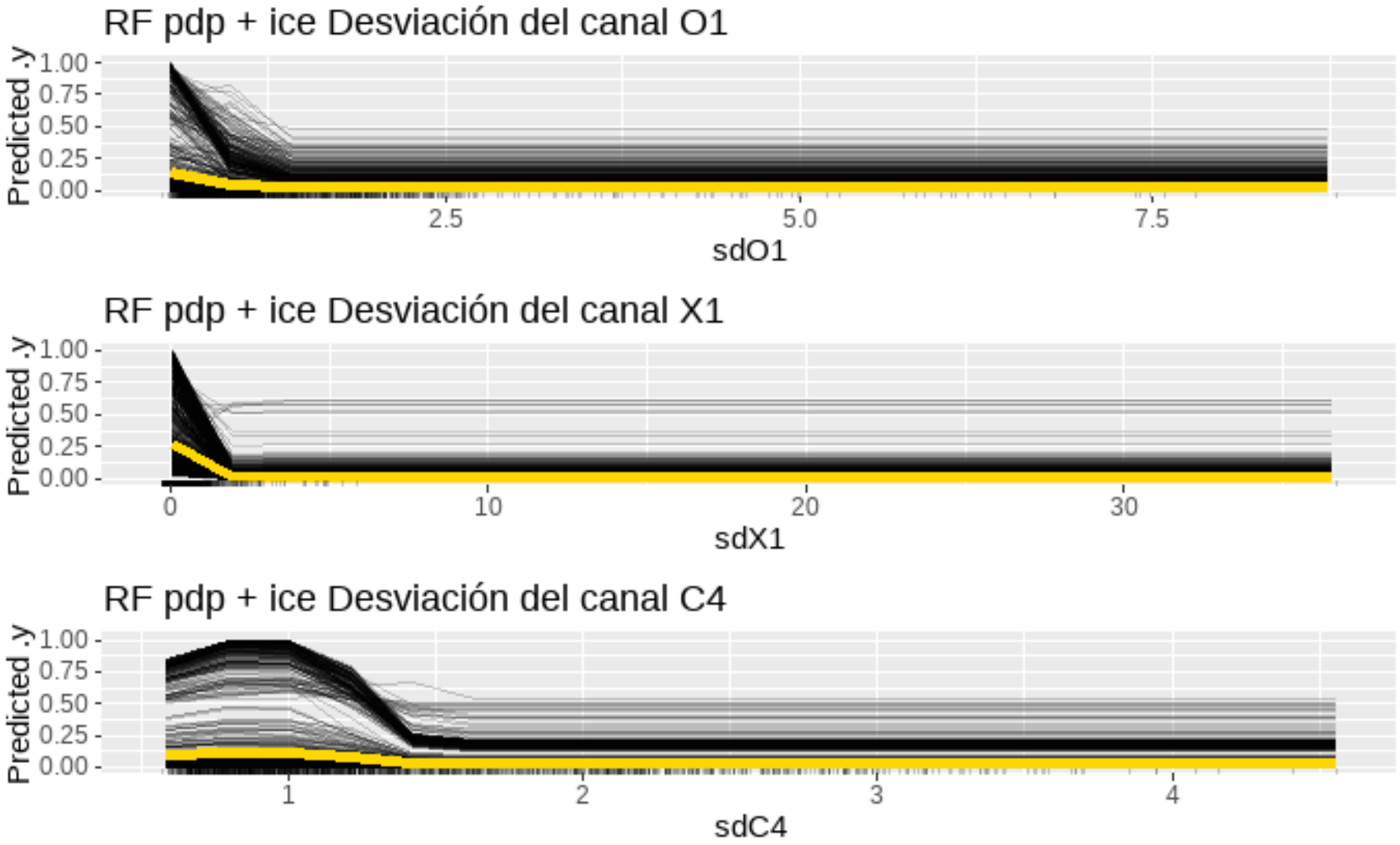
- loss = "logLoss”



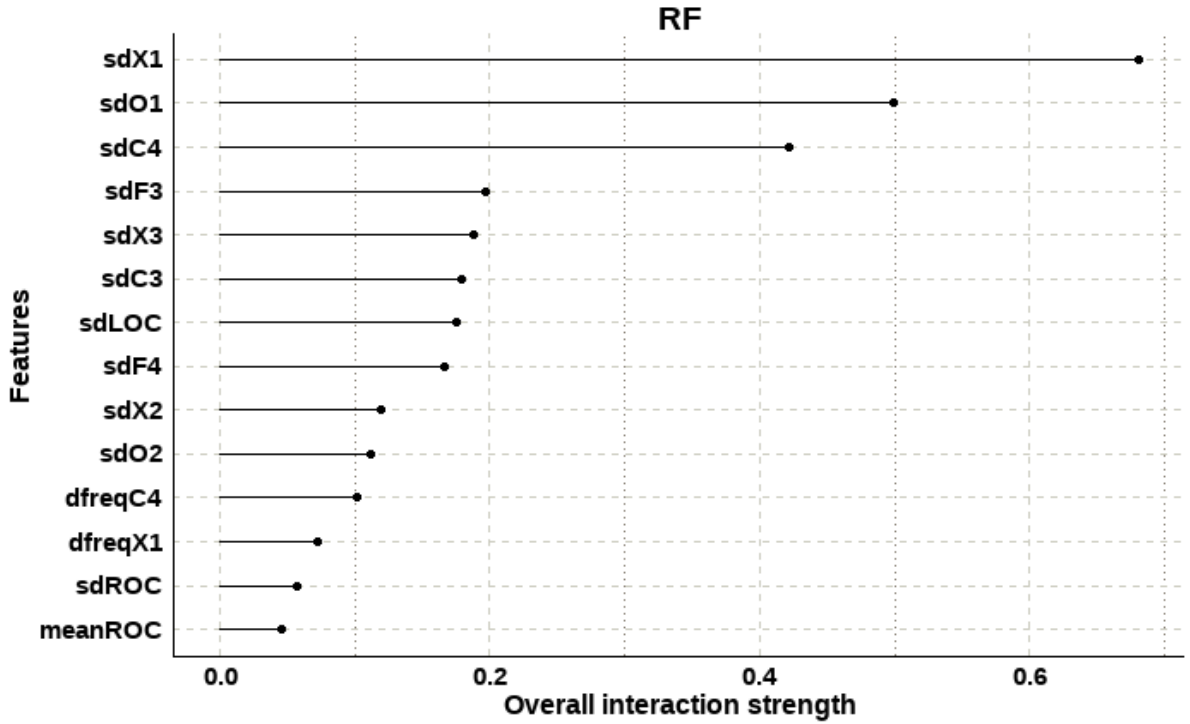
- loss = “ce”

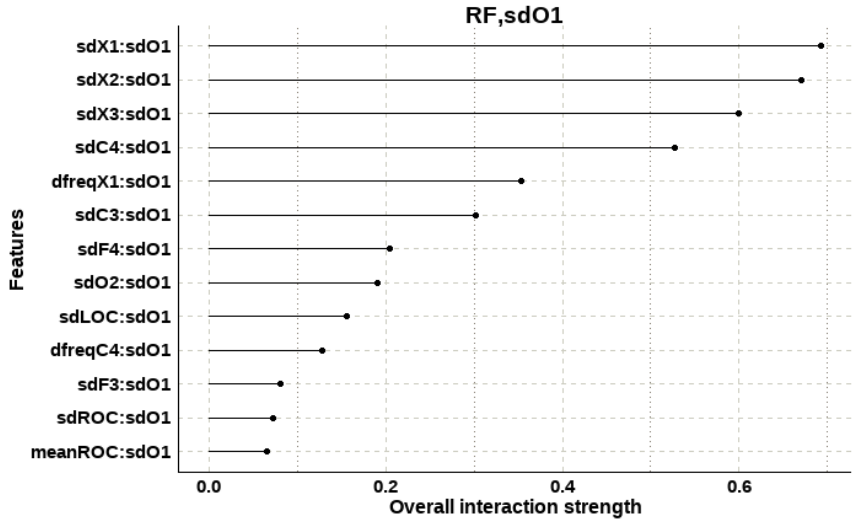


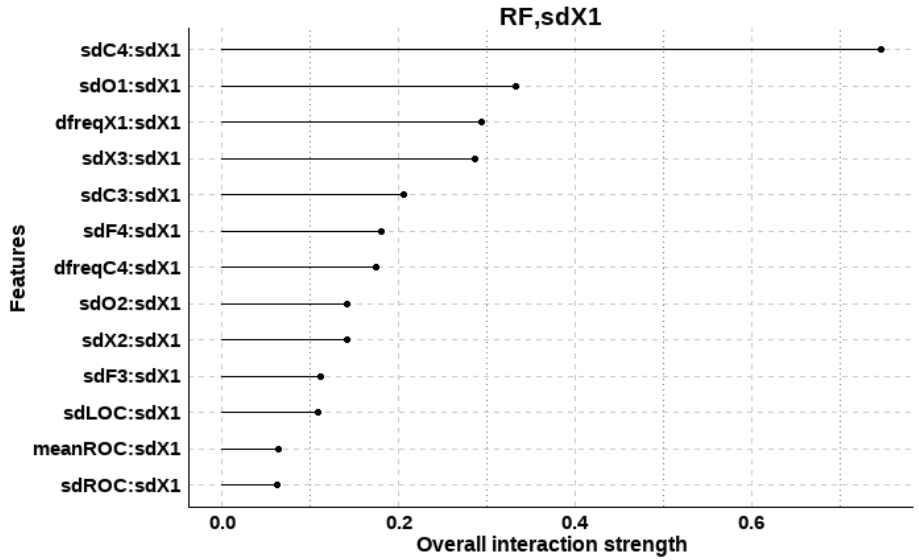
INTERPRETABILIDAD



INTERACCIONES

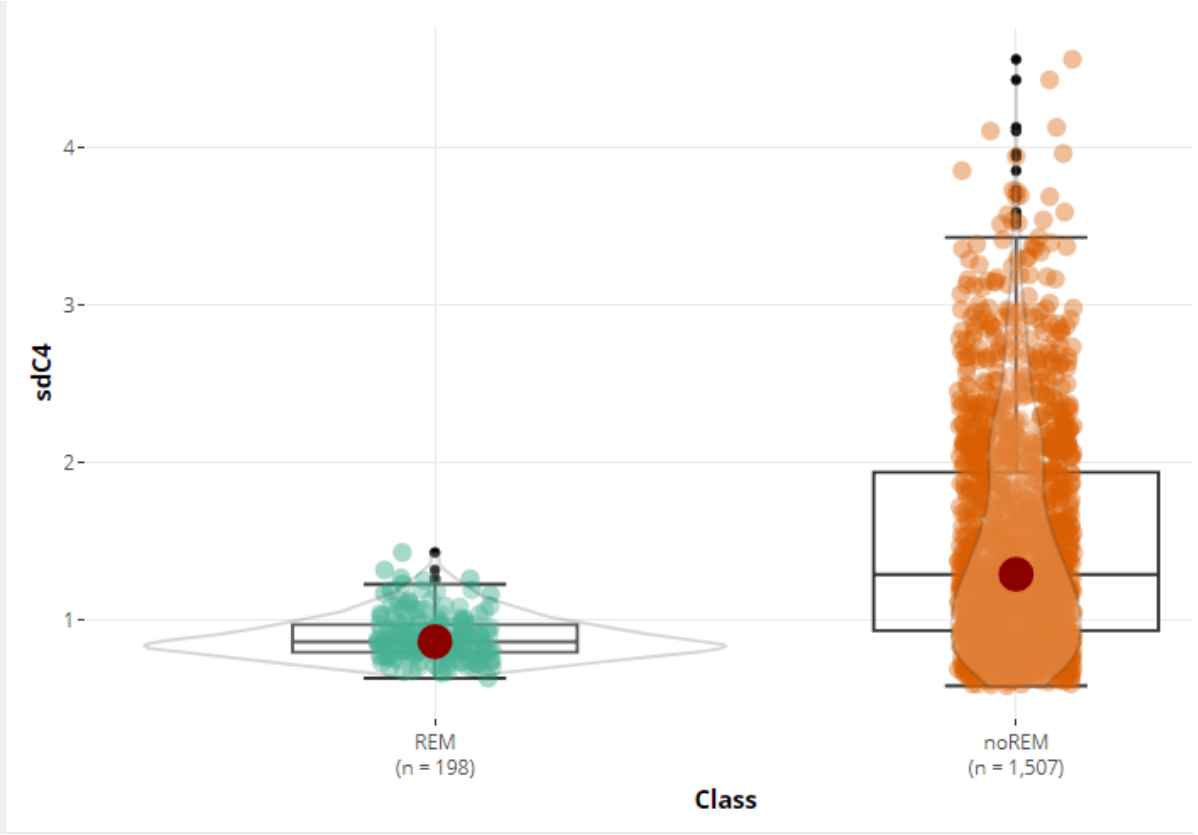




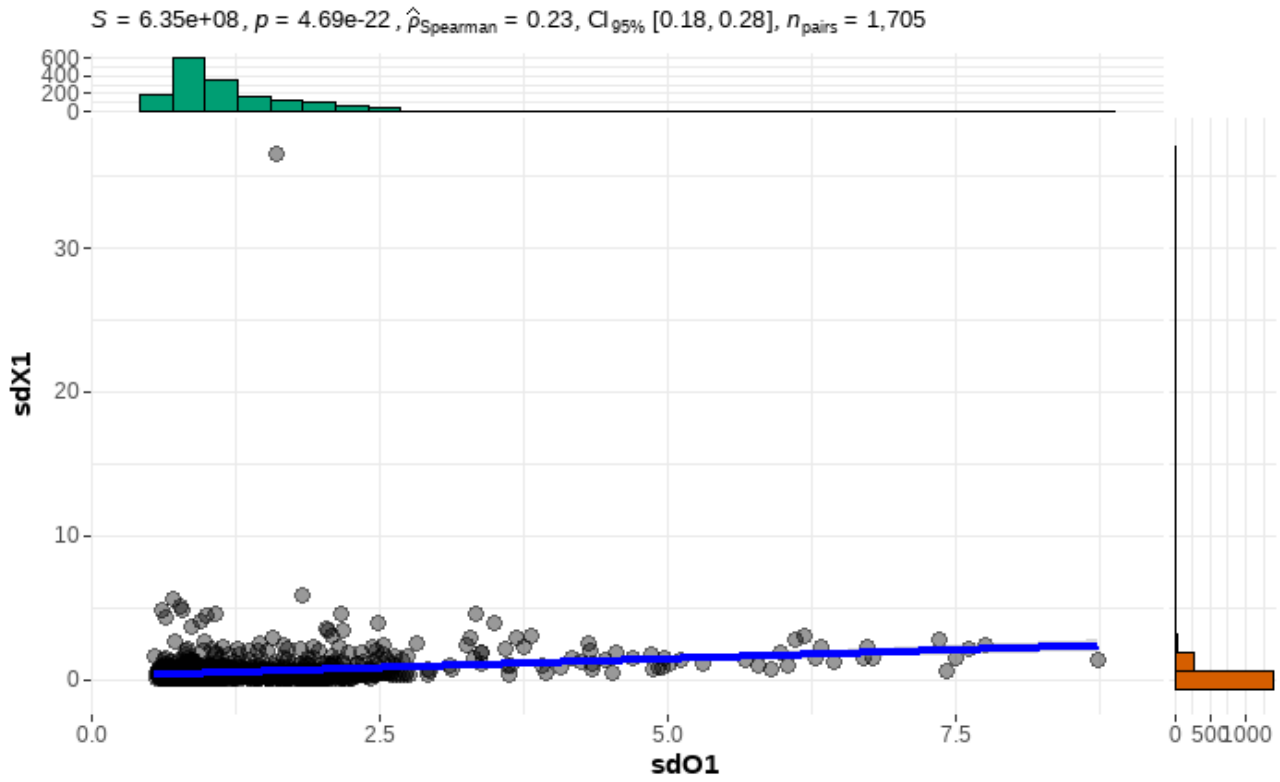


R E A L C I O N E S

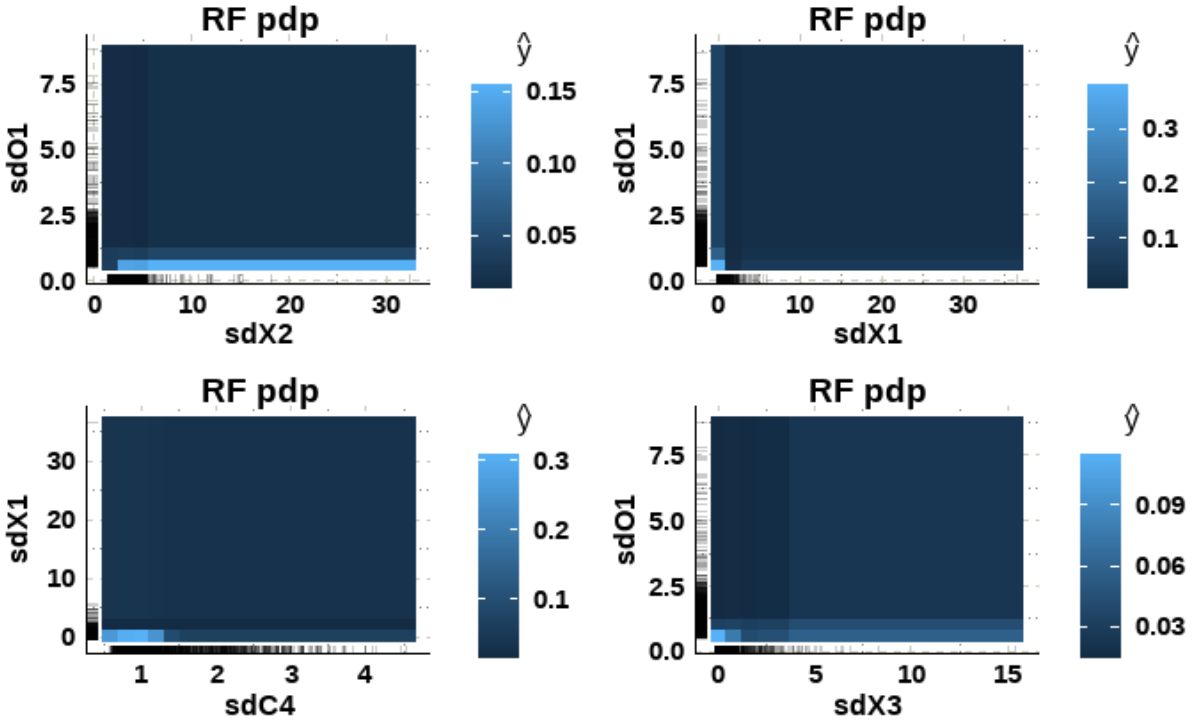
- **Violin Plot: sdC4**

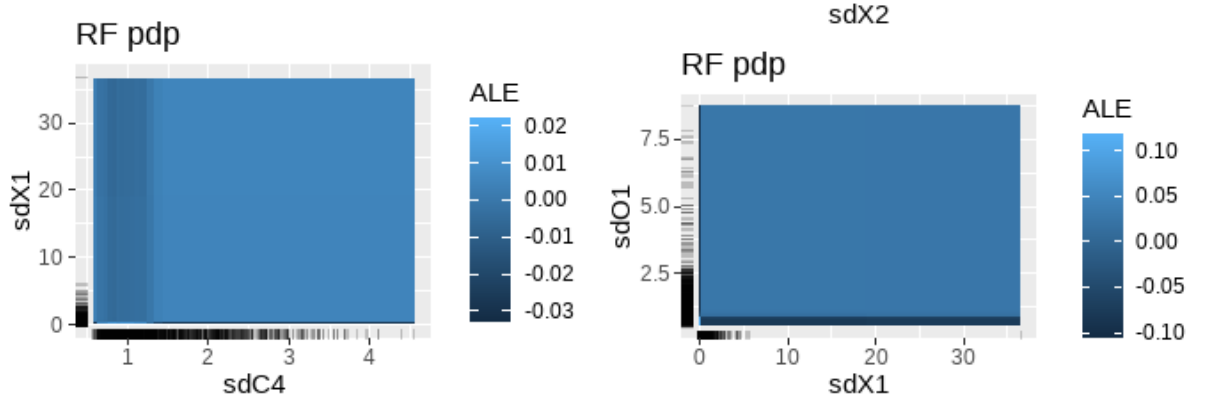


- **Correlación sdO1 y sdX1**

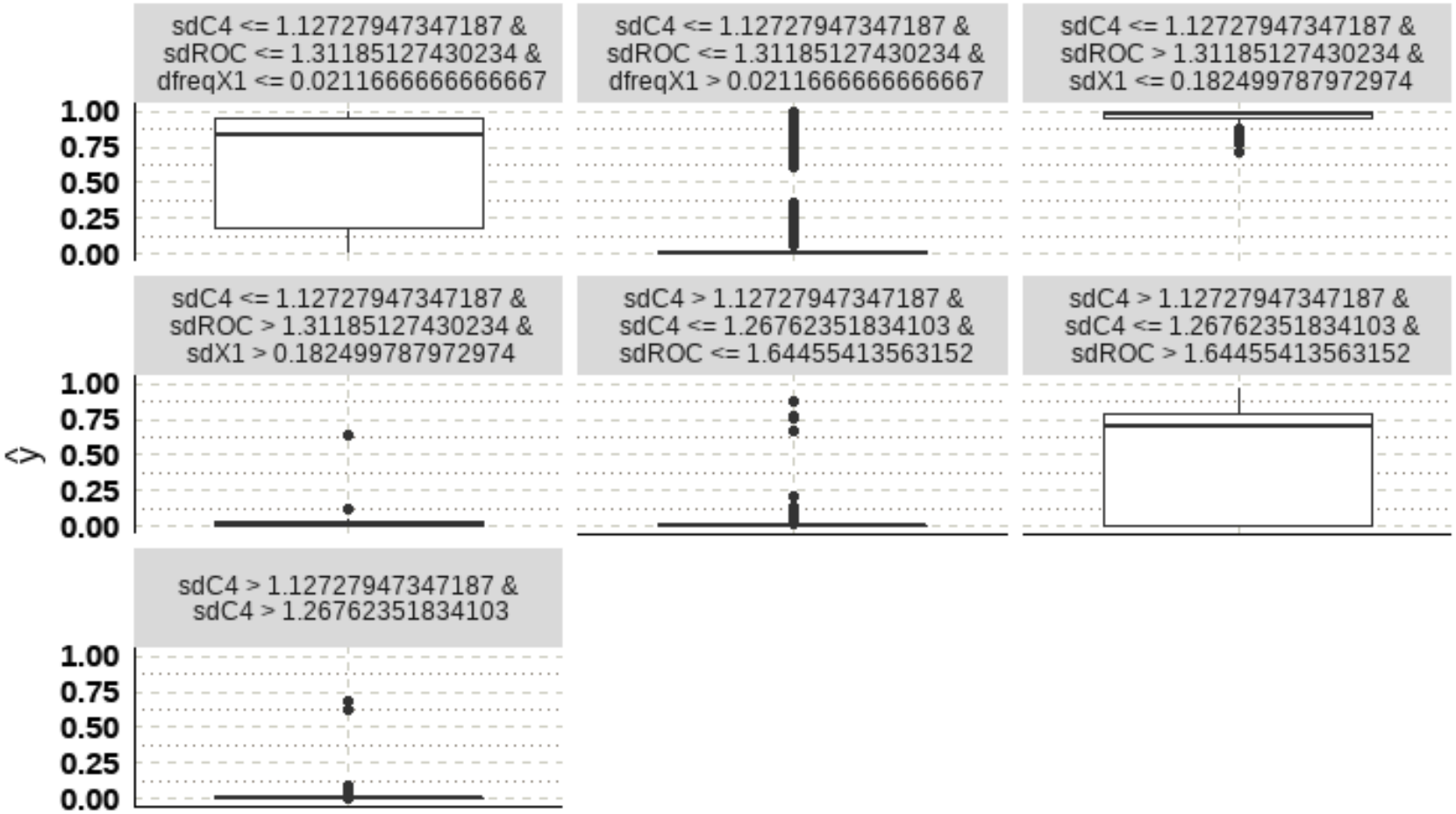


INTERPRETABILIDAD
ALE



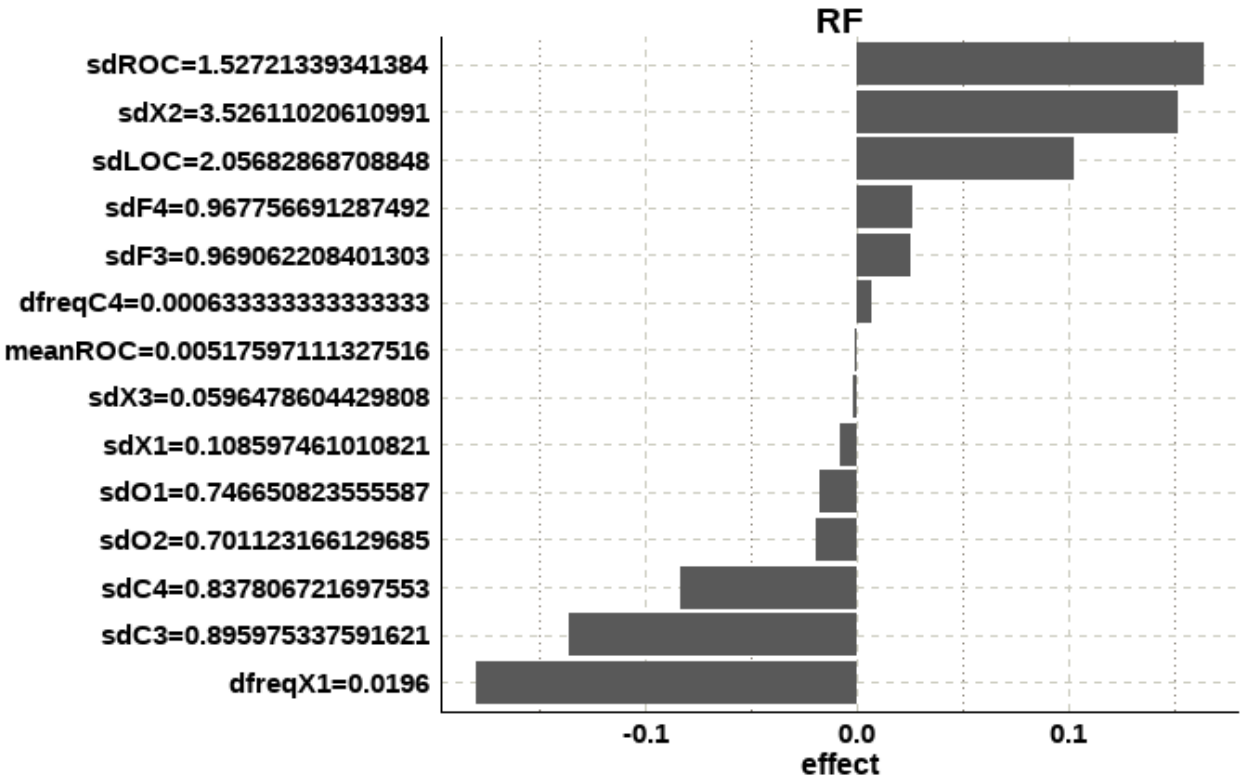


MODEL WHITE BOX

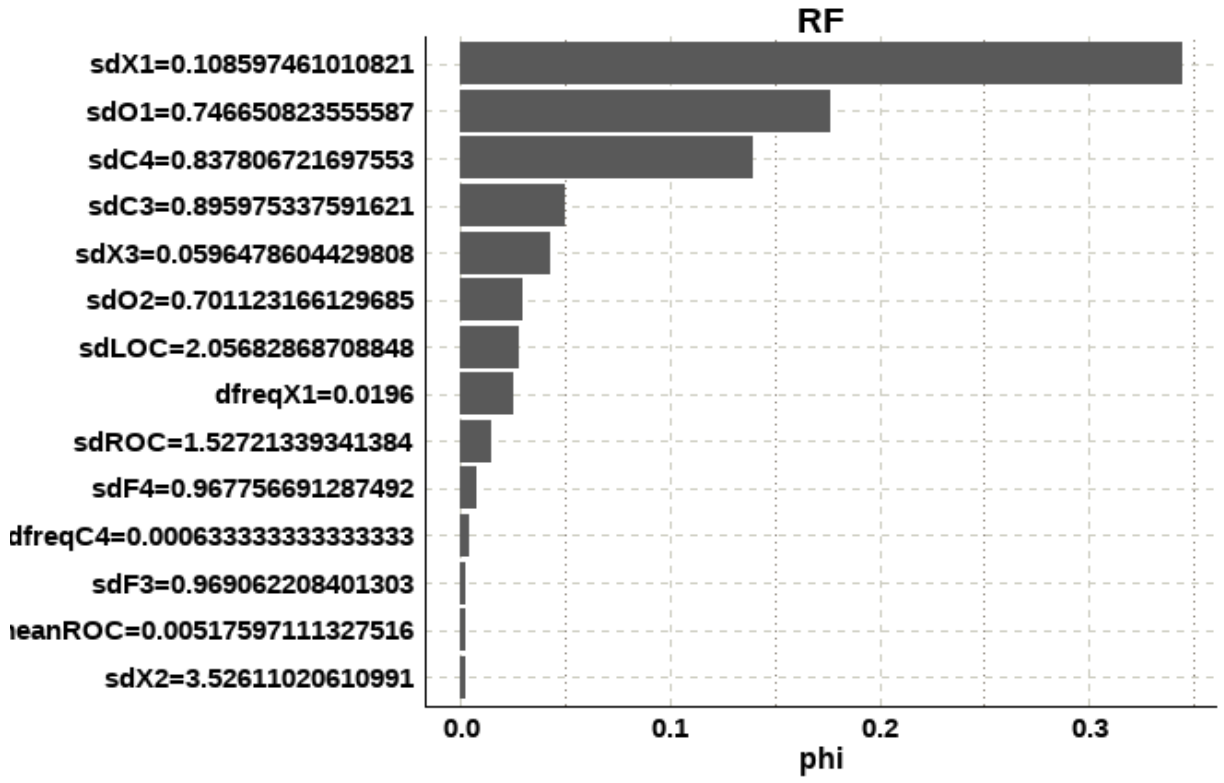


EFECTOS PARA LA CLASIFICACIÓN CON MAYOR PROBABILIDAD

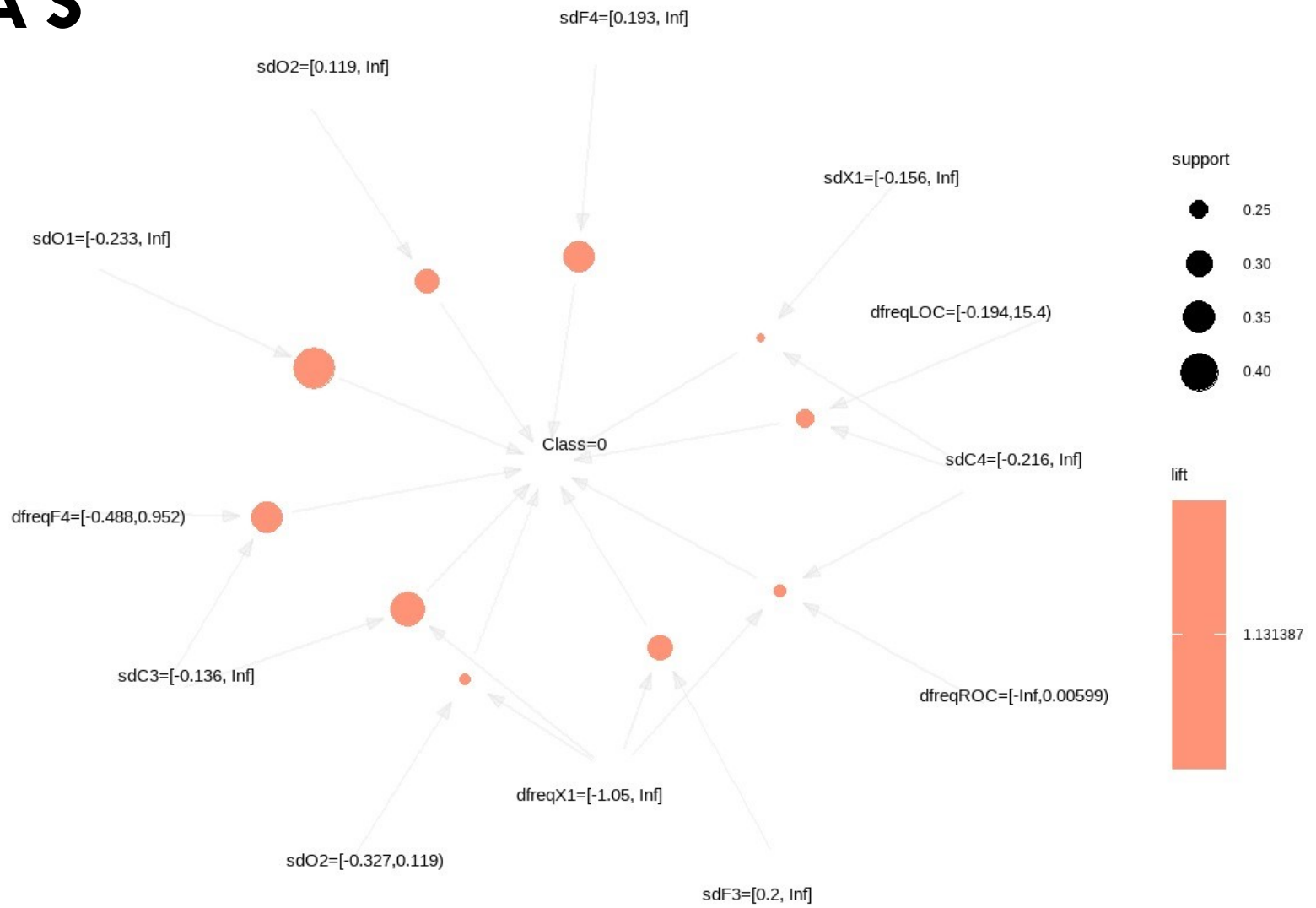
- **LIME**



- **SHAP VALUES**



REGLAS


sdF4=[0.193, Inf]
sdO2=[0.119, Inf]
sdX1=[-0.156, Inf]
sdO1=[-0.233, Inf]
dfreqLOC=[-0.194,15.4)
Class=0
sdC4=[-0.216, Inf]
dfreqF4=[-0.488,0.952)
sdC3=[-0.136, Inf]
dfreqROC=[-Inf,0.00599)
dfreqX1=[-1.05, Inf]
sdO2=[-0.327,0.119)
sdF3=[0.2, Inf]
support
0.25
0.30
0.35
0.40
lift
1.131387

5. CONCLUSIONES


FINAL