EVALUACIÓN DE LOS MODELOS

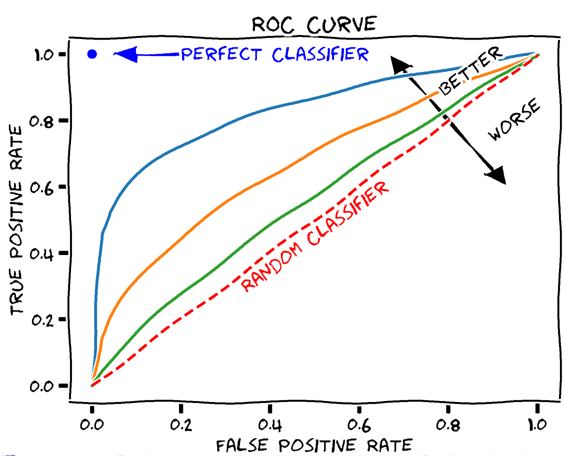
$$Sensitivity = \frac{TP}{TP + FN}$$

[same as recall; aka **true positive rate**]

$$Specificity = \frac{TN}{TN + FP}$$

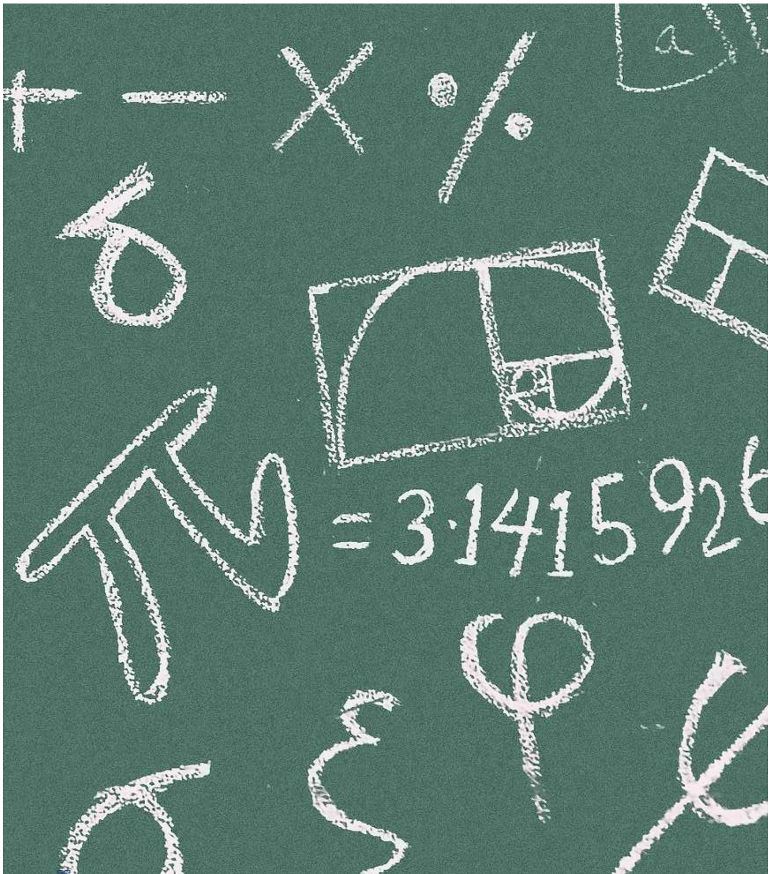
[aka **true negative rate**]

		Actual: Has Heart Disease	Actual: Does Not Have Heart Disease
Predicted	Has Heart Disease	True Positives	False Positives
	Does Not Have Heart Disease	False Negatives	True Negatives



La curva ROC nos permite ver todos los umbrales de clasificación sin leer +1000 matrices de confusión

METODOLOGIA -PASOS

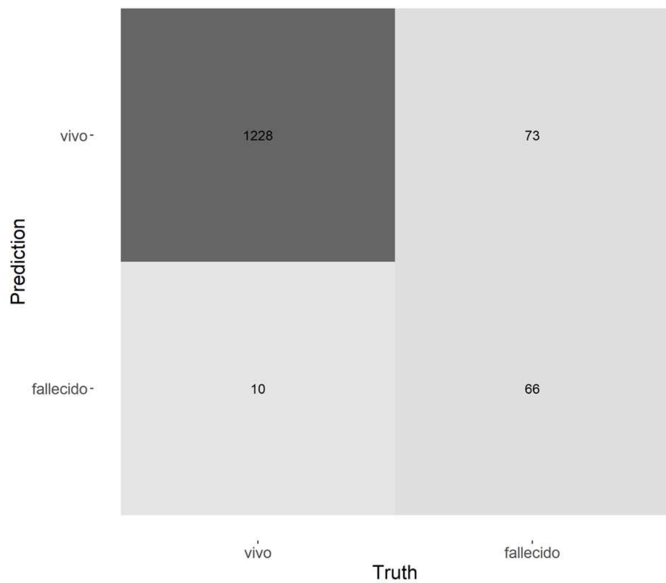


IPS UNIVERSITARIA
Servicios de Salud
Universidad de Antioquia

1. Seleccionar un desenlace a modelar (hospitalización, mortalidad, cirugía, urgencias)
2. Pre-procesar los datos en el formato correcto.
3. Fraccionar el conjunto de datos en partes para construir el modelo y para probarlo (10 Fold CV)
4. Definir una métrica de base para comparar (ROC / AUC)
5. Escoger varios modelos, evaluar sus métricas y seleccionar el mejor.
6. Usar el modelo por cada paciente en una medición.
7. Validar la metodología con investigación (epidemiología, dirección de salud) y publicar.

NUESTRO MODELO IPS DE MORTALIDAD

Modelo de predicción de mortalidad datos del 2020 del PGP



n=5511 pacientes

4134 datos de entrenamiento
1377 para prueba

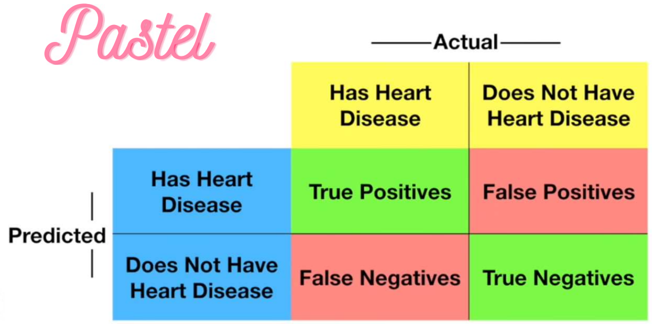
$$Sensitivity = \frac{TP}{TP + FN}$$

[same as recall; aka **true positive rate**]

$$Specificity = \frac{TN}{TN + FP}$$

[aka **true negative rate**]

Pastel



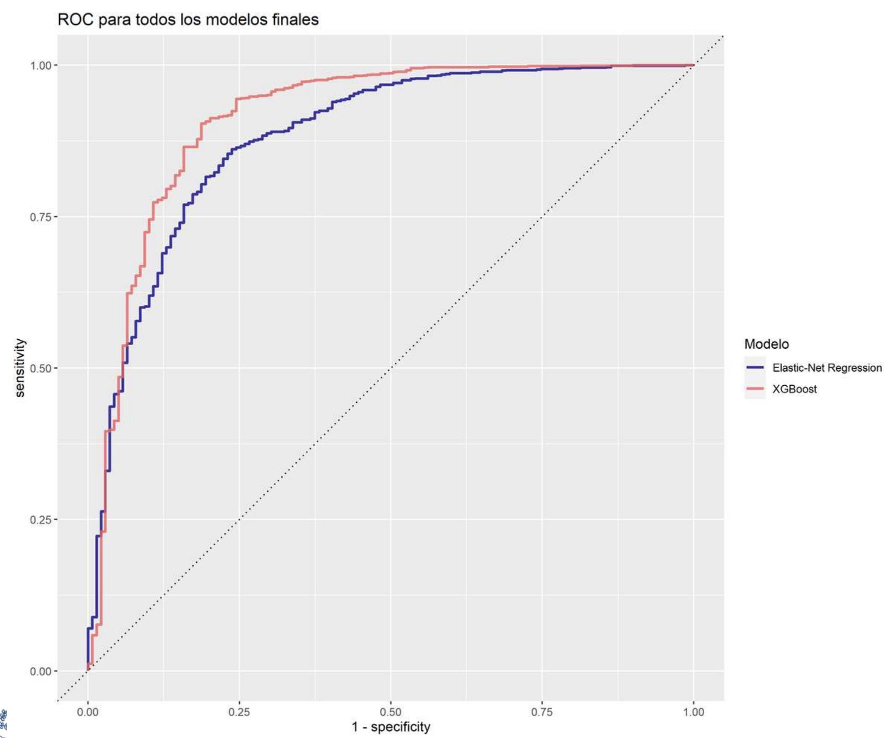
Modelo Mortalidad XGBoost con grid de 100 modelos y 10-Fold CV

.metric	.estimate
sens	0.9919225
spec	0.4748201
accuracy	0.9397240
roc_auc	0.9119431

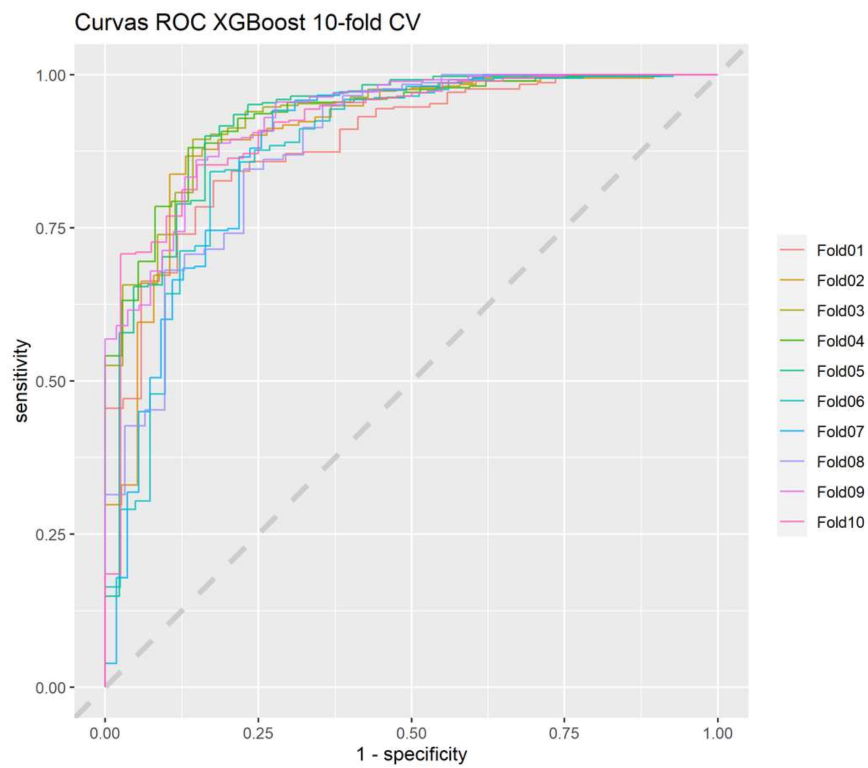
Universidad de Antioquia

NUESTRO MODELO IPS DE MORTALIDAD

Modelo de predicción de mortalidad datos del 2020 del PGP

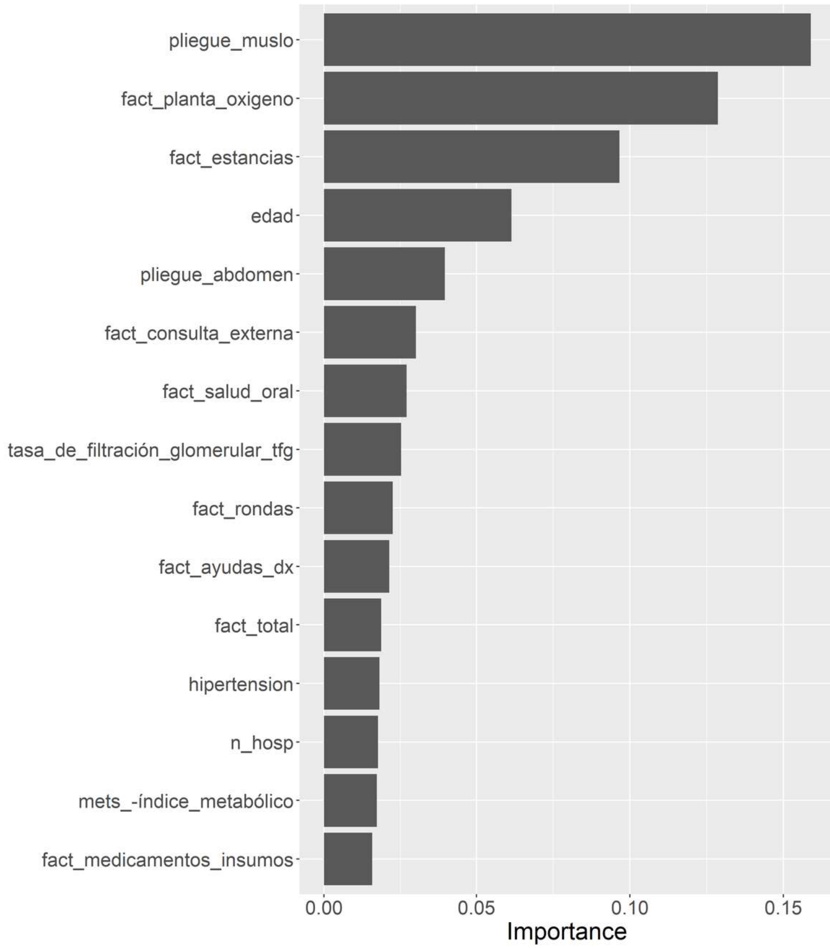


ROC-AUC= 0.91
Accuracy= 0.94



Universidad de Antioquia

Model-agnostic variable importance (XGBoost Classifier)



Pasos en el algoritmo de Gradient Boosting

Input: Data $\{(x_i, y_i)\}_{i=1}^{n}$, and a differentiable **Loss Function** $L(y_i, F(x))$

Step 1: Initialize model with a constant value: $F_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^{n} L(y_i, \gamma)$

Step 2: for $m = 1$ to M:

(A) Compute $r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}(x)}$ for $i = 1,\ldots,n$

(B) Fit a regression tree to the r_{im} values and create terminal regions R_{jm}, for $j = 1 \ldots J_m$

(C) For $j = 1 \ldots J_m$ compute $\gamma_{jm} = \underset{\gamma}{\operatorname{argmin}} \sum_{x_i \in R_{ij}} L(y_i, F_{m-1}(x_i) + \gamma)$

(D) Update $F_m(x) = F_{m-1}(x) + \nu \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm})$

Step 3: Output $F_M(x)$

Universidad de Antioquia