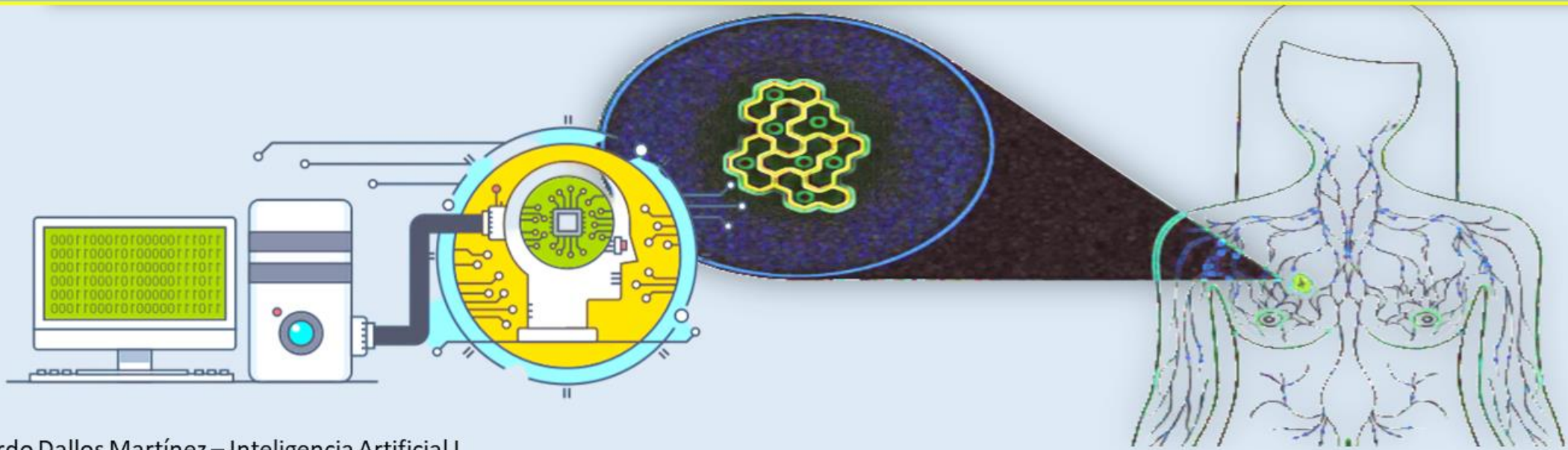


IDENTIFICACIÓN DE LAS CARACTERÍSTICAS MAS INFLUYENTES EN LA PREDICCIÓN DE CÁNCER DE MAMAA PARTIR DE MEDICIONES A MUESTRAS OBTENIDAS MEDIANTE LA TÉCNICA FNA



Leonardo Dallos Martínez – Inteligencia Artificial I

Identificación de los características mas influyentes en la predicción de cáncer de mama a partir de mediciones a muestras obtenidas mediante la técnica FNA

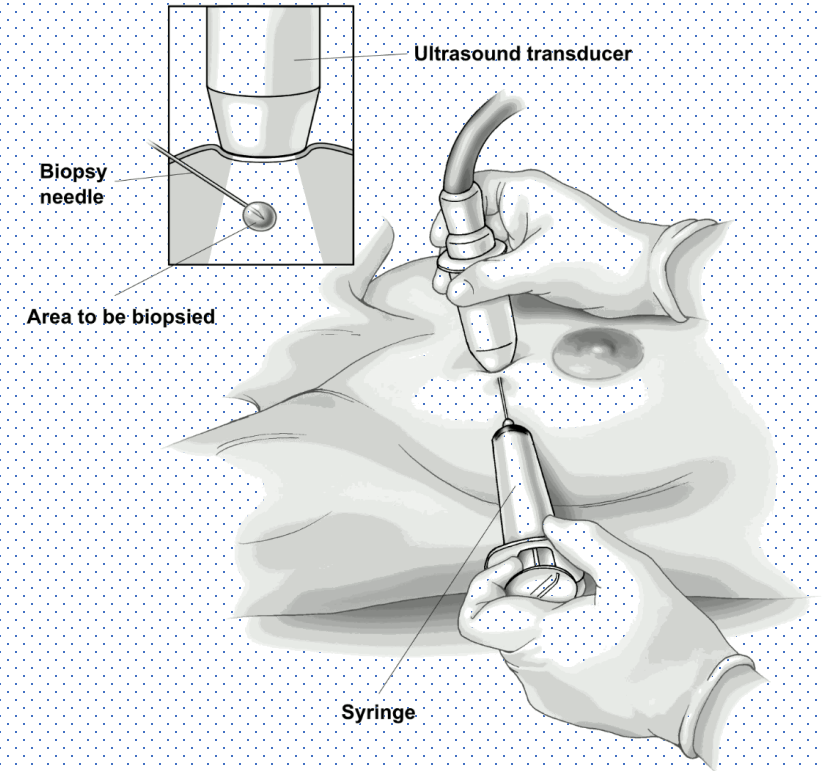
OBJETIVOS

- Realizar un modelo de predicción para identificación de cáncer de mama a partir de datos de muestras de tejido sospechoso obtenidas con la técnica FNA.
- Identificar cuales son los factores que mas influyen en la identificación de cáncer de seno a partir de modelos matemáticos.

Identificación de los características mas influyentes en la predicción de cáncer de mama a partir de mediciones a muestras obtenidas mediante la técnica FNA

DATASET

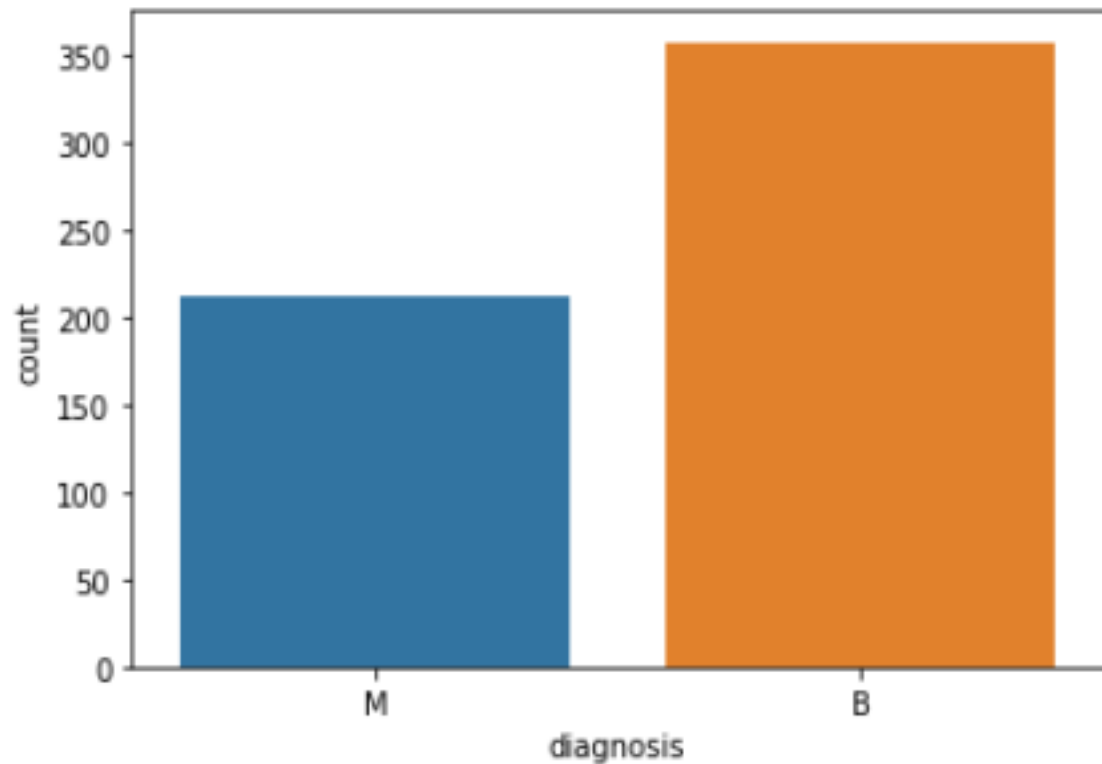
- Muestras para estudio obtenidas de tejido sospechosos mediante la técnica FNA (biopsia por aspiración con aguja fina).
- 569 Muestras
- 30 características para cada una de las muestras
- 1 clase clasificadora (benigno, maligno)



Fine needle aspiration using ultrasound

Identificación de las características más influyentes en la predicción de cáncer de mama a partir de mediciones a muestras obtenidas mediante la técnica FNA

DATASET



A cada muestra se le tomaron diversas mediciones para cada una de sus características y se agruparon en 3 grandes grupos así:

- id
- radio
- textura
- perímetro
- área
- suavidad
- compactividad
- concavidad
- puntos concavos
- simetría
- dimensión fractal

- Datos de media.
- Datos de desviación estándar.
- Datos de peor medida (worst).

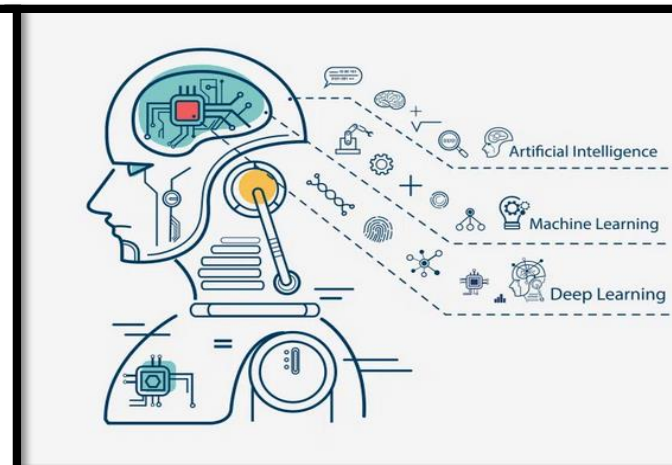
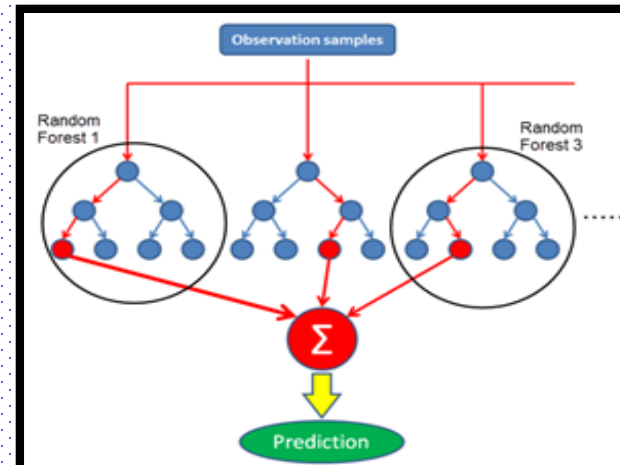
Identificación de los características mas influyentes en la predicción de cáncer de mama a partir de mediciones a muestras obtenidas mediante la técnica FNA

DISEÑO:

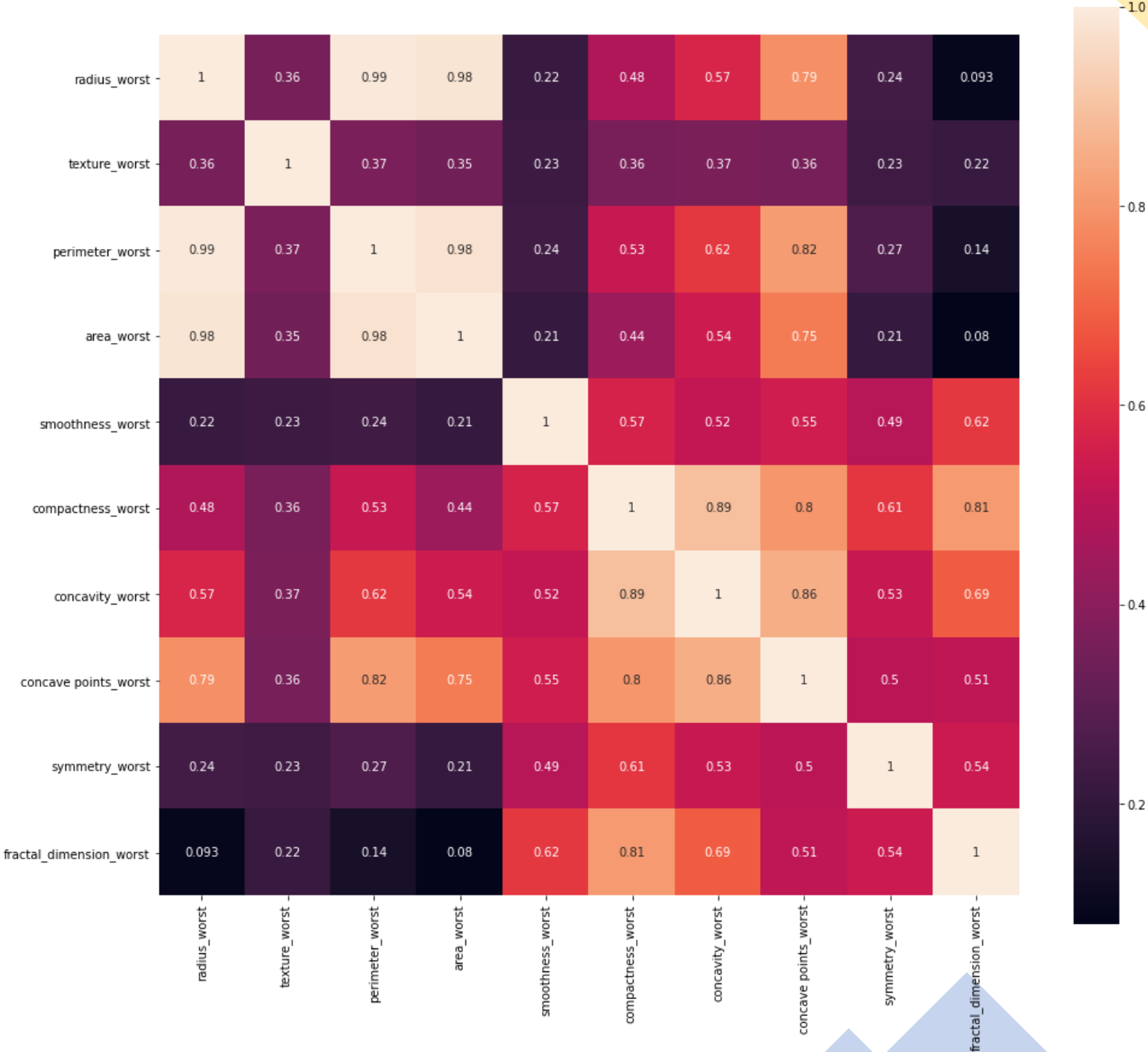
- Análisis estadístico :
 1. Preprocesamiento de datos.
 2. Testeos con los modelos de clasificación con el uso de herramientas tales como: PCA, cross-validation.
 3. Testeos con DNN.
 4. Selección del modelo con el mejor comportamiento.
 5. Determinación de las características mas influyentes para el modelo seleccionado.

MODELOS USADOS:

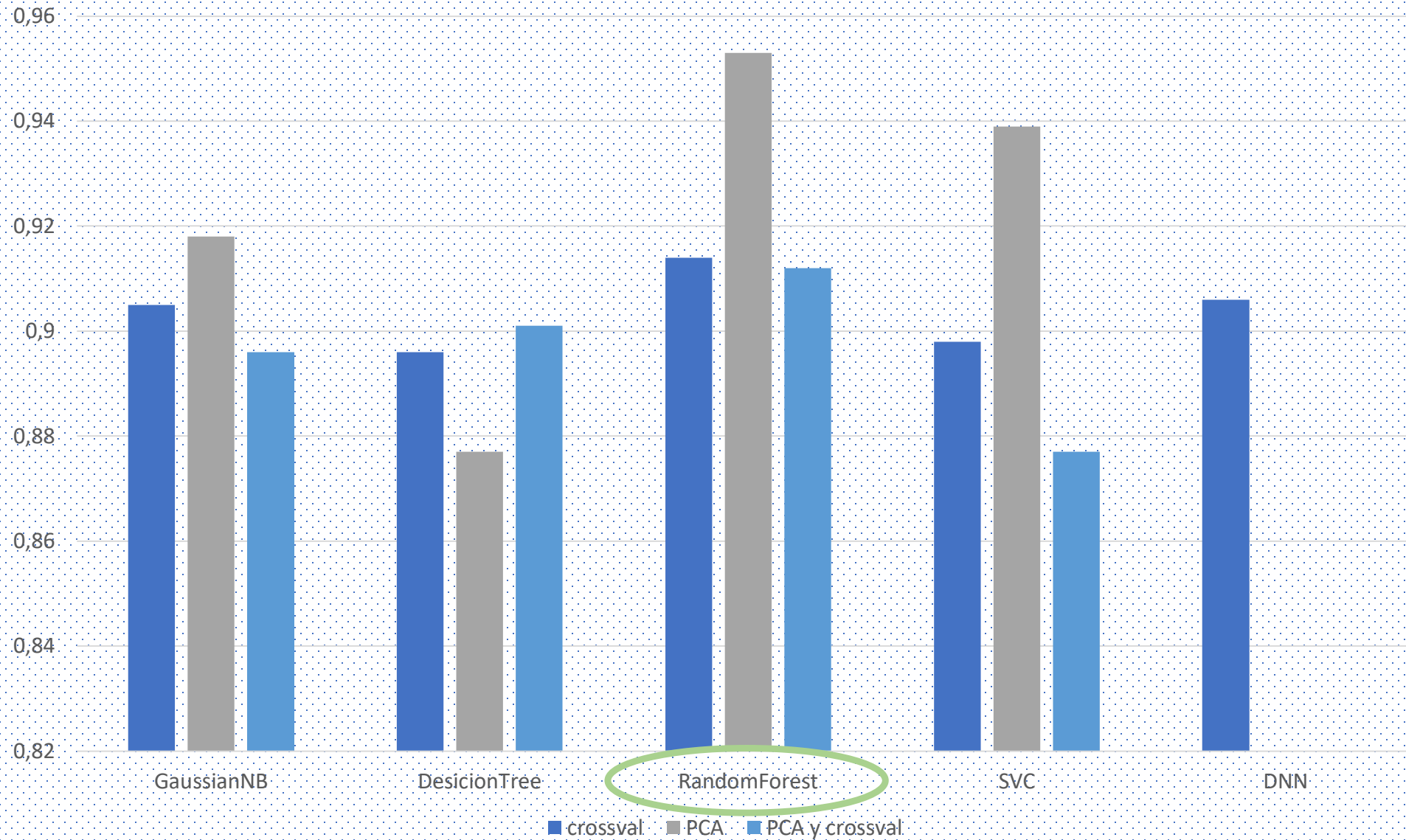
- GaussianNB.
- Decisión Tree clasifier.
- Random forest clasifier.
- SVC.
- Redes neuronales profundas (DNN)



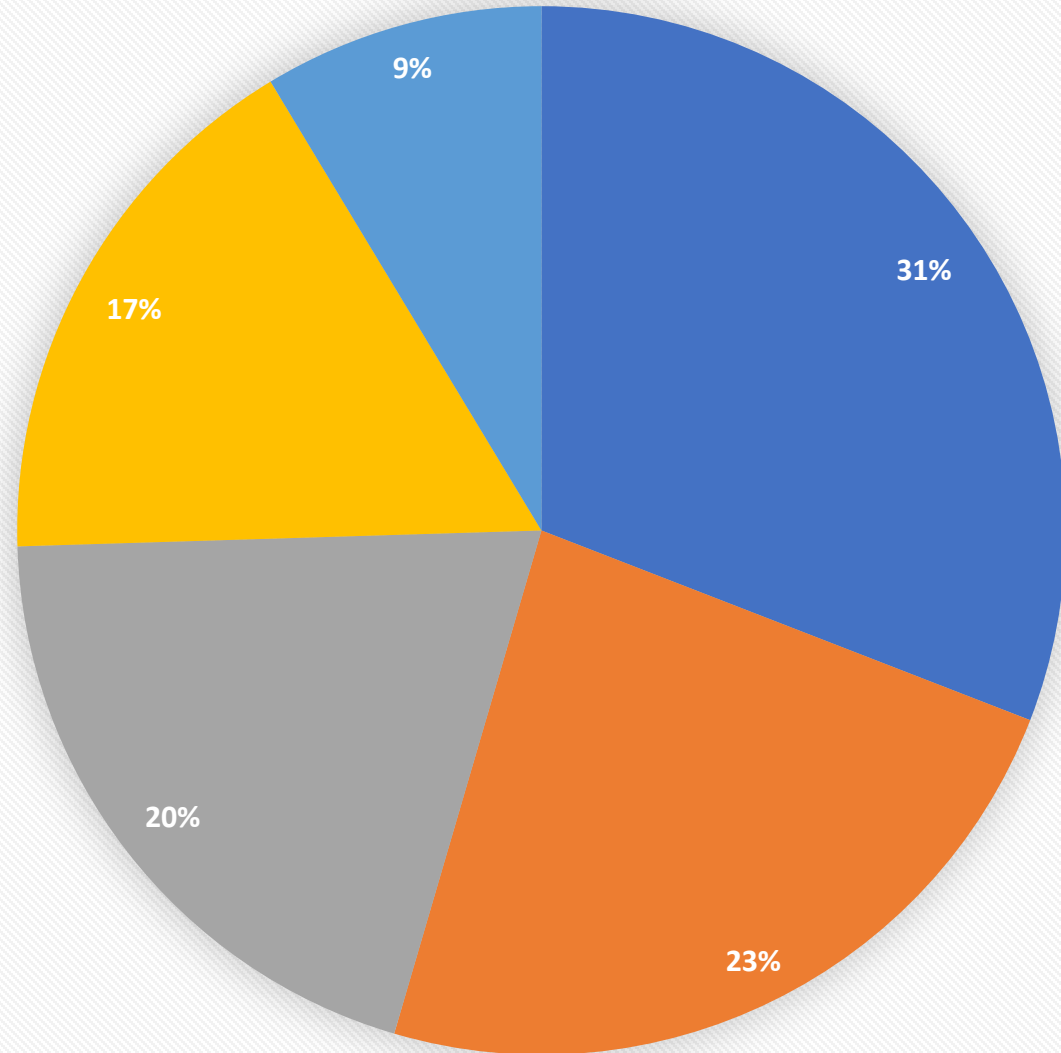
CORRELACIÓN DE DATOS



PRECISION DE LOS MODELOS DE ESTUDIO:



Características mas influyentes



■ concave points mean

■ perimeter mean

■ area mean

■ radio mean

■ compactness mean

CONCLUSIONES

:

1. Para el testeo de los distintos modelos matemáticos se debe modificar correctamente el dataset para quitarle información que no aporte y no corresponda para el estudio y así tener un resultados mas acertados.

2. El algoritmo seleccionado RandomForest no es un algoritmo perfecto pero bien muestra un error bajo en sus predicciones.

3. El modelo de redes neuronales profundas (DNN) mostro un comportamiento similar al modelo de clasificación seleccionado (RandomForest) y bien podría también ser seleccionado como modelo para el análisis con un mayor estudio de su configuración para su uso.

4. Escalar la información en el modelo por red neuronal profunda permite generar una mejor versión del modelo, resultando en un menor error y mejor puntuación.

5. Las características mas relevantes identificadas (perímetro, área, radio, compactividad) según el modelo seleccionado vemos que son las mediciones mas notorias en una muestra FNA.



Universidad
Industrial de
Santander

#LaUISqueQueremos

¡Gracias!

