

## **1. Selección de conjuntos de datos:**

Los conjuntos de datos utilizados son:

- banknote
- diabetes
- liver-disorders
- ilpd
- mammographic

Se eligieron estos conjuntos tratando de tener la menor cantidad de columnas posibles y la mayor cantidad de filas.

## **2. Aplicar análisis de correlación**

Se realizó un análisis de correlación y se borraron las columnas que no eran necesarias.

## **3. Clasificadores binarios**

Se usaron los siguientes clasificadores:

- KNN
- MLP

Se puede ver la ejecución de ambos clasificadores desde el archivo "mainProgram.py". Cada clasificador tiene un archivo dedicado, donde existe una clase específica que se manda a llamar por el método "Execute("Ruta del archivo")" el cual recibe un dato tipo String con la ruta del archivo que se va a analizar. Los archivos dedicados se pueden encontrar con los nombres: "knn.py" y "MLP.py".

Para cada clasificación, se hace un 5-fold cross-validation. Así mismo se calcula el AUC promedio, tiempo promedio de entrenamiento y tiempo promedio de clasificación.

Los resultados de estos clasificadores fueron los siguientes:

-----  
Resultados de KNN:

AUC = 0.721294739785  
Tiempo promedio de entrenamiento = 0.00620002746582  
Tiempo promedio de clasificacion = 0.0  
-----

-----  
Resultados de MLP:

AUC = 0.775248837228  
Tiempo promedio de entrenamiento = 0.00320000648499  
Tiempo promedio de clasificacion = 0.302999973297  
-----

#### 4. Clasificadores de una clase

Se usaron los siguientes clasificadores:

- KMeans
- OCKRA

Se puede ver la ejecución de ambos clasificadores desde el archivo "mainProgram.py". Cada clasificador tiene un archivo dedicado, donde existe una clase específica que se manda a llamar por el método "Execute("Ruta del archivo")" el cual recibe un dato tipo String con la ruta del archivo que se va a analizar. Los archivos dedicados se pueden encontrar con los nombres: "KM.py" y "OCKRA.py".

Para cada clasificación, se hace un 5-fold cross-validation. Así mismo se calcula el AUC promedio, tiempo promedio de entrenamiento y tiempo promedio de clasificación.

Los resultados de estos clasificadores fueron los siguientes:

-----  
Resultados de KMeans:

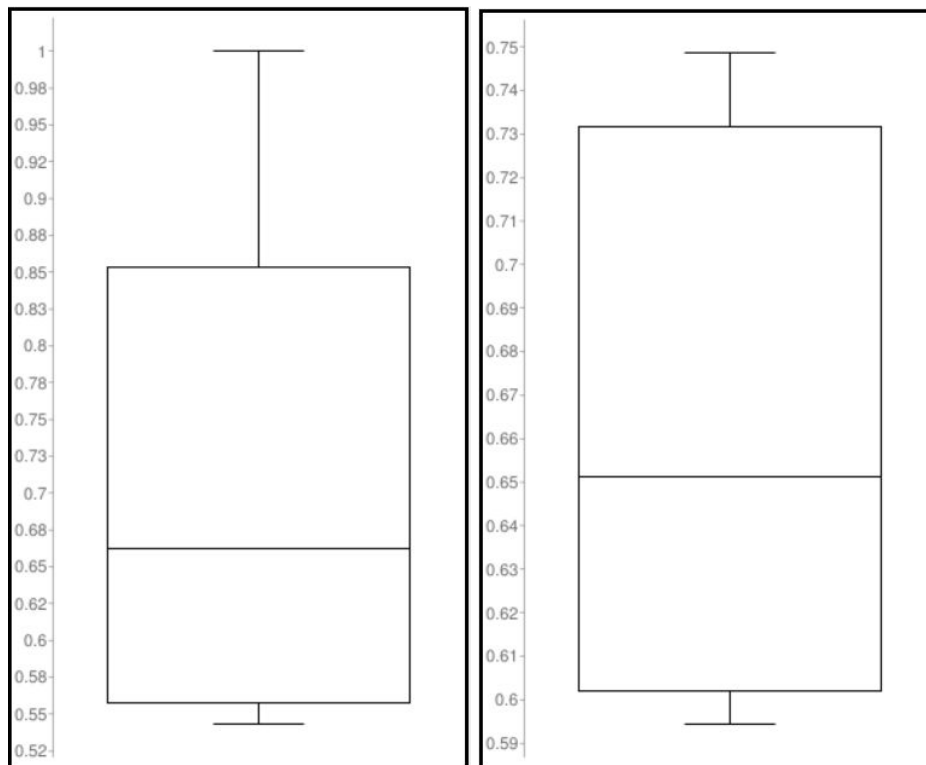
AUC = 0.675700899736  
Tiempo promedio de entrenamiento = 0.0843999862671  
Tiempo promedio de clasificacion = 0.0  
-----

-----  
Resultados de OCKRA:

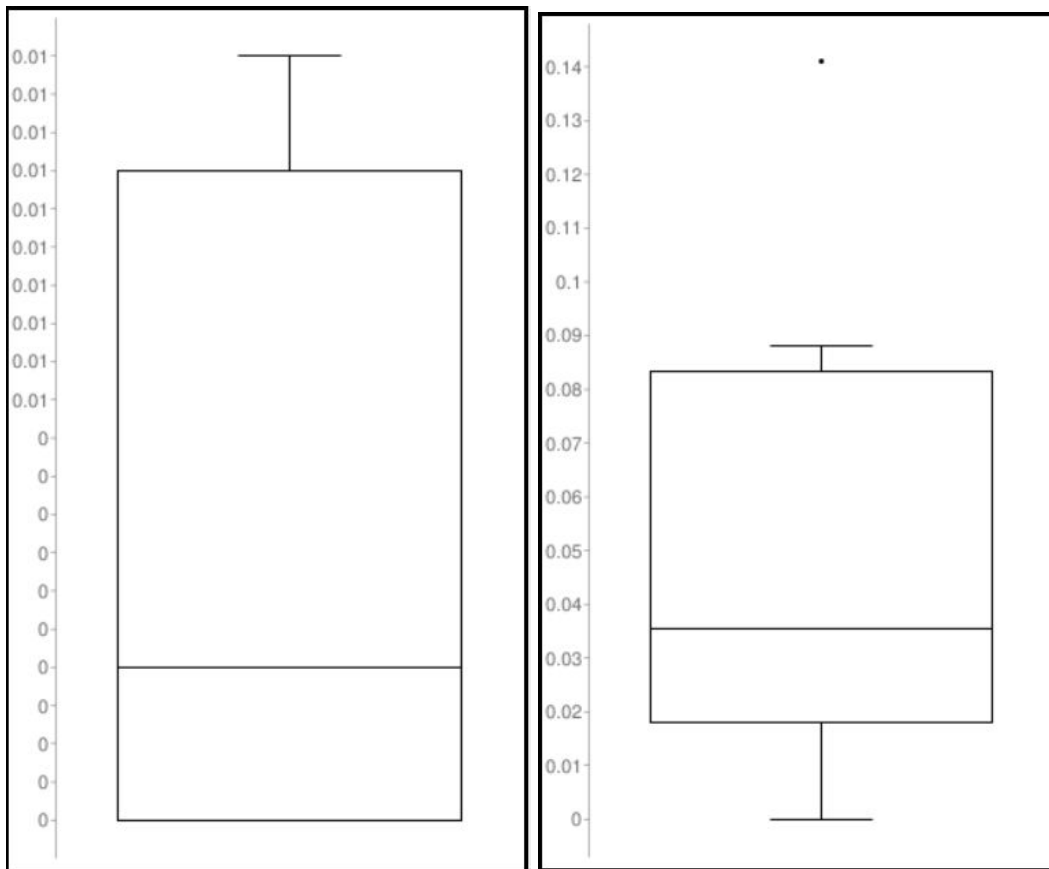
AUC = 0.653629183162  
Tiempo promedio de entrenamiento = 0.0124000549316  
Tiempo promedio de clasificacion = 2.2259999752  
-----

### 5. Determinar qué clasificadores son mejores:

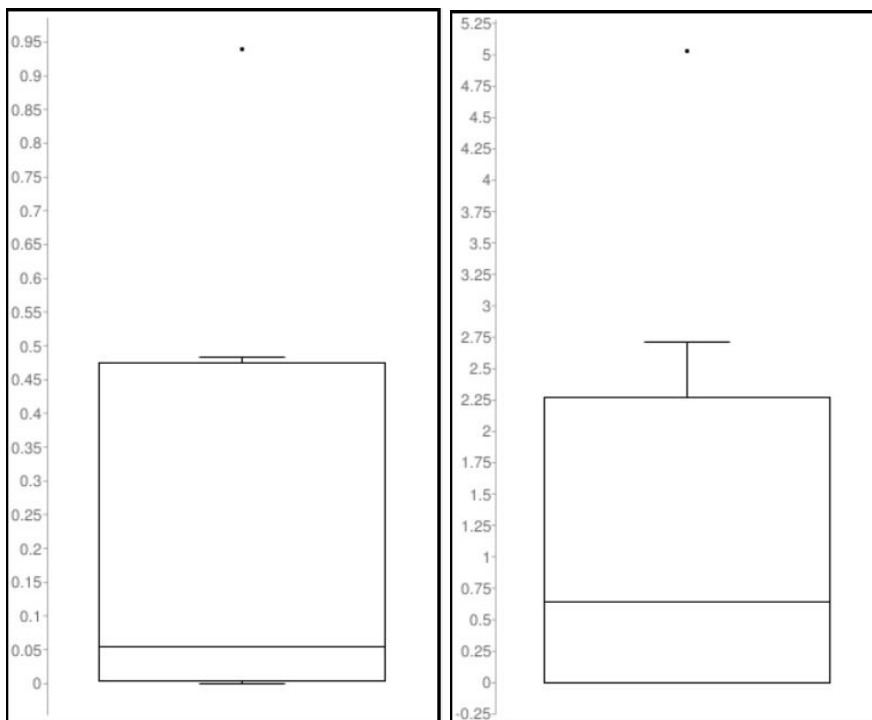
Comparación de AUC de clasificadores binarios (izquierda) vs clasificadores de una clase (derecha)



Comparación de tiempo de Entrenamiento de clasificadores binarios (izquierda) vs clasificadores de una clase (derecha)



Comparación de tiempo de Clasificación de clasificadores binarios (izquierda) vs clasificadores de una clase (derecha)



A partir de los resultados obtenidos de las comparaciones entre ambos tipos de clasificadores con los diagramas de cajas y bigotes, se puede decir que los clasificadores binarios tienen un AUC promedio superior al de los clasificadores de una clase.

Así mismo, los clasificadores binarios tardan menos tiempo en ejecutar las clasificaciones que los clasificadores de una clase.

\*Los datos usados para los diagramas de cajas y bigotes fueron los obtenidos por cada vez que se ejecutó un fold para cada tipo de clasificador

## **6. Análisis estadístico**

Tomando en cuenta el AUC de ambos tipos de clasificadores si existe una diferencia en el desempeño entre los clasificadores de una clase y los binarios. Sin embargo no es una diferencia contundente.

por otra parte hay que destacar que los clasificadores binarios tienen un AUC promedio superior al de los clasificadores de una clase.