

mÉTODOS DE APRENDIZAJE ESTADÍSTICO, ÁRBOLES DE DECISIÓN Y DISEÑO EXPERIMENTAL

PRÁCTICA 2 – INTELIGENCIA COMPUTACIONAL



antía FRAGA LOURIDO

VERÓNICA ARANDA JARRETA

1. Introducción

En esta práctica se ha realizado el entrenamiento de varios modelos para dos conjuntos de datos: Iris y Breast Cancer Wisconsin. Para ello se han empleado diferentes métodos estadísticos para la estimación del error para cada uno de los conjuntos de datos.

2. Materiales

* Dos conjuntos de datos:
  + **Iris**: 150 muestras, 4 variables y 3 clases.
  + **Breast Cancer Wisconsin (original)**: 699 muestras, 10 variables y 2 clases.
* MATLAB

3. Métodos

En primer lugar, se ha realizado un análisis de los dos conjuntos de datos para verificar si es necesario hacer un preprocesado de los mismo. En el caso del conjunto de datos de Iris, no ha sido necesario, pero, este procedimiento si se ha tenido en el caso del conjunto de datos de Breast Cancer Wisconsin, al tener ‘missing values’.

En este caso, lo que se ha hecho ha sido sustituir los ‘missing values por la media del atributo en el que se encontraban.

Para los dos conjuntos de datos se ha empleado el método de la validación cruzada, siendo para el Iris el caso extremo del Leaving one-out y para el Breast Cancer Wisconsin el K-fold (k = 10).

Por otro lado, los modelos entrenados para ambos conjuntos de datos han sido:

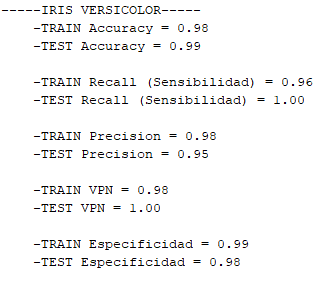
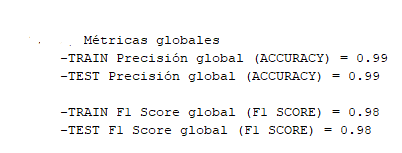
* Modelo empleando el discriminante lineal de Fisher
* Modelo empleando el discriminante cuadrático
* Tres modelos empleando Árboles de Decisión, variando las métricas empleadas en cada uno

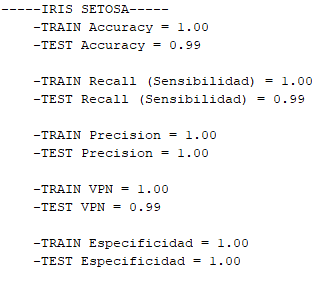
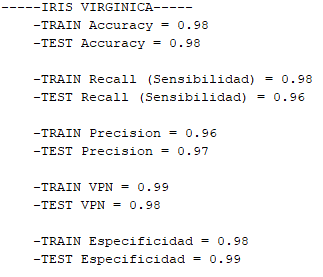
Además, para cada modelo entrenado y también para el conjunto de entrenamiento, se calcularon una serie de métricas estadísticas:

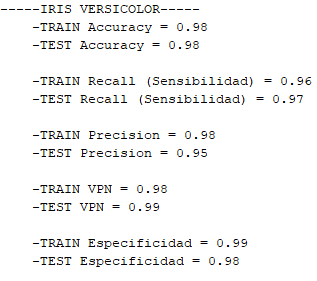
* Accuracy (Exactitud)
* Recall (Sensibilidad)
* PPV (Precisión)
* VPN (Valor predictivo negativo)
* Especificidad
* F1 Score

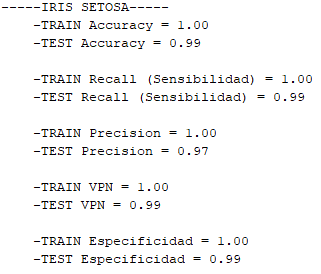
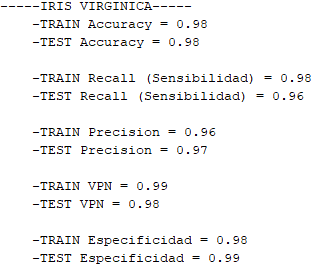
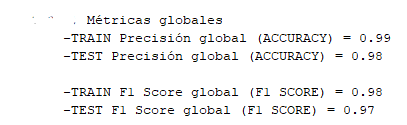
4. Resultados

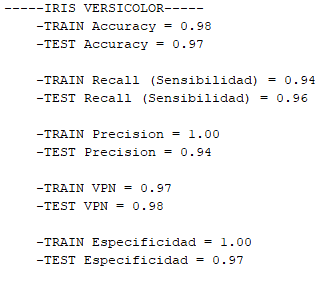
4.1 Resultados Iris

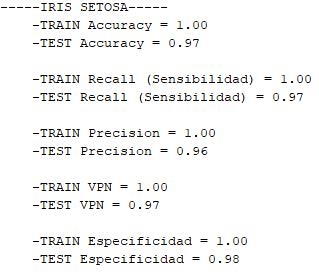
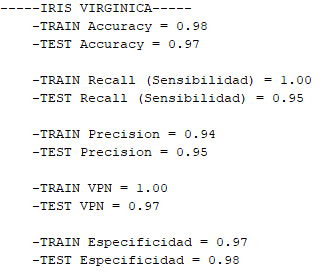
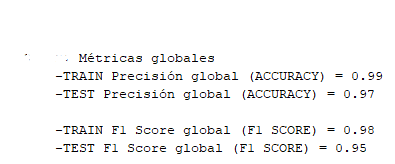
4.1.1 Resultados del modelo entrenado con Discriminante Lineal de Fisher

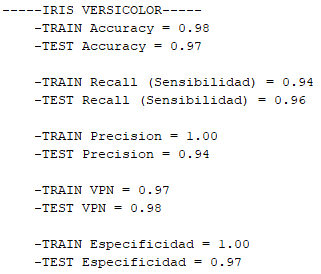


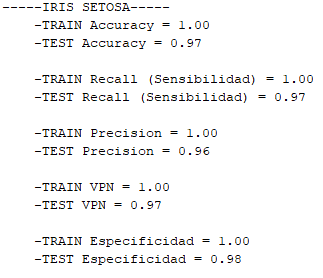
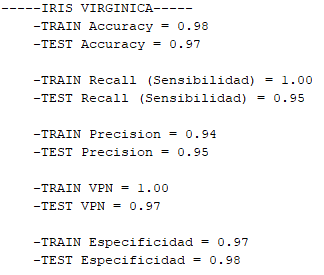
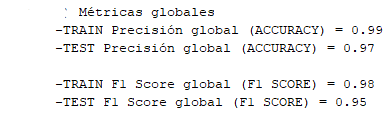
4.1.2 Resultados del modelo entrenado con Discriminante Cuadrático

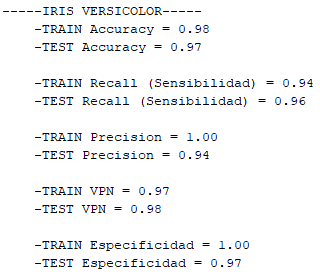


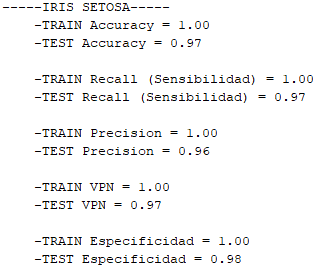
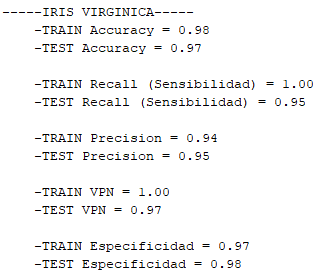
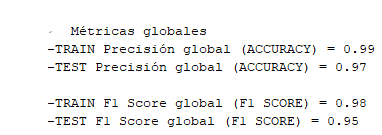
4.1.3 Resultados del modelo entrenado con Árboles de Decisión (modelo I)



4.1.4 Resultados del modelo entrenado con Árboles de Decisión (modelo II)



4.1.5 Resultados del modelo entrenado con Árboles de Decisión (modelo III)



4.1.6 Resultados generales **¿? 🡪 ¿Cómo llamarlo?**

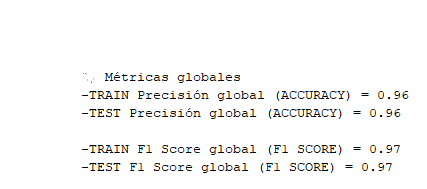
Imagen que contiene captura de pantalla

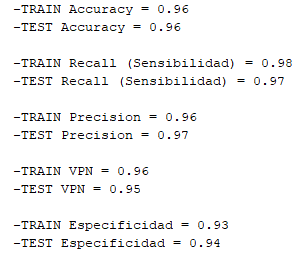
Descripción generada automáticamente

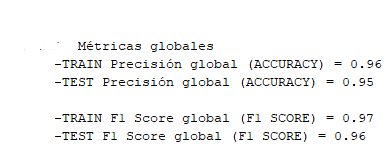
Imagen que contiene texto

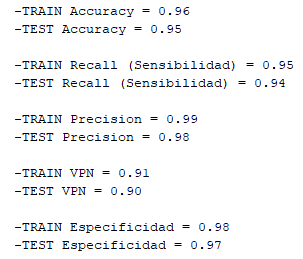
Descripción generada automáticamente

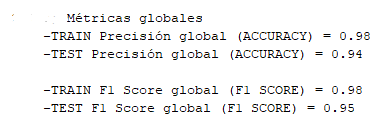
4.2 Resultados Breast Cancer Wisconsin

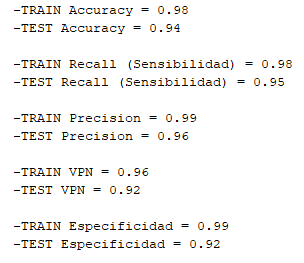
4.2.1 Resultados del modelo entrenado con Discriminante Lineal de Fisher

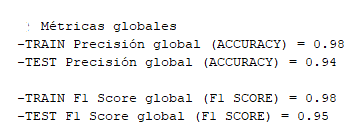


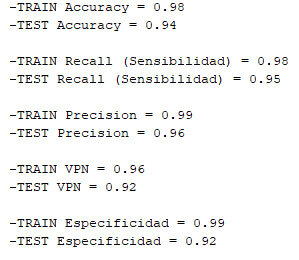
4.2.2 Resultados del modelo entrenado con Discriminante Cuadrático

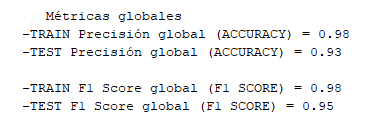


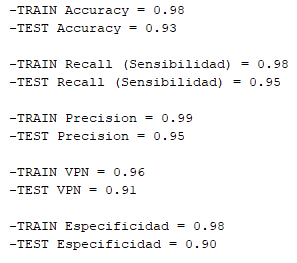
4.2.3 Resultados del modelo entrenado con Árboles de Decisión (modelo I)



4.2.4 Resultados del modelo entrenado con Árboles de Decisión (modelo II)



4.2.5 Resultados del modelo entrenado con Árboles de Decisión (modelo III)



4.2.6 Resultados generales **¿? Como llamarlo**

Imagen que contiene captura de pantalla

Descripción generada automáticamente

Imagen que contiene texto

Descripción generada automáticamente

5. Conclusiones