2020 - I semester
INFO257 - Inteligencia Artificial
Unidad 1 - Aprendizaje Supervisada

Tarea práctica

Dataset

El dataset *Sloan Digital Sky Survey DR14* (*https://www.sdss.org/dr14/*) contiene 10.000 observaciones del espacio tomadas por el SDSS (https://www.sdss.org/). Cada observación es descrita por 17 columnas de características y una columna de clase que la identifica como una *estrella*, *galaxia* o *quasar*. Los datos del SDSS están descritas por características obtenidos por varias mediciones de espectros ópticos y infrarrojos.

Descargar el dataset:
https://github.com/matthieuvernier/INFO257_2020/blob/master/unidad1/datos/SDSS-DR14.csv

Tarea

El ejercicio consiste en aprender distintos modelos de clasificación supervisada capaz de clasificar nuevos datos del SDSS en tres clases: *estrella*, *galaxia* o *quasar*.

Podrán utilizar los algoritmos de Machine Learning implementado en la libreria Python Scikit-Learn: https://scikit-learn.org/stable/

- Cuidarán el preprocesamiento inicial de los datos brutos para eliminar cualquier inconsistencia.
- Analizarán la distribución de las características para evaluar eventuales desequilibros en el dataset.
- Entrenarán y evaluarán el rendimiento de al menos tres modelos distintos para resolver esta tarea, utilizando métricas de evaluación apropiadas.
- Compararan el rendimiento de los modelos obtenidos y conversarán las ventajas y desventajas potenciales de cada modelo.
- Analizarán los modelos obtenidos para identificar cuáles son las características más relevantes para la clasificación de estrellas, galaxia y quasar.

Entrega de la tarea:

- Notebook Python en su cuenta GitHub
- Plazo: **jueves 28 de mayo** (Enviarme una notificación por Slack con el enlace de su cuenta)