

UNIDAD TEMÁTICA 2: Procesamiento previo de los datos

PRACTICOS DOMICILIARIOS INDIVIDUALES #1

Ejercicio 1

En RapidMiner Studio, ve a la sección de Help / Tutorials, y de la parte “Prepare Data” completa los tutoriales:

1. “Handling Missing Values” y
2. “Normalization and Outlier detection”

En ambos casos, desarrolla un breve documento (no más de media carilla) con los hallazgos más importantes, salva los procesos y modelos que generes y remite todo en un archivo comprimido a la tarea correspondiente en la webasignatura.

Ejercicio 2

El objetivo de este ejercicio es practicar el uso de los bloques disponibles en RapidMiner para manejar datos faltantes, outliers, estandarización y normalización de atributos.

1. Descarga el dataset “wine” del repositorio UCI e impórtalo al repositorio de RAPIDMINER.
2. Analiza y documenta el problema a resolver, los atributos, sus estadísticas, la existencia de valores faltantes y outliers.
3. Crea un nuevo proceso en RapidMiner, que tendrá dos canales: en uno utiliza el dataset tal cual está, y en el otro aplica bloques para normalizar y estandarizar los datos.
4. En ambos canales utiliza un bloque para particionar los datos en entrenamiento y test.
5. Utiliza un modelo sencillo de Naive Bayes para realizar la clasificación, y bloques de evaluación de la performance pertinentes (utilizando el dataset de test) para obtener los resultados.
6. ¿Cómo se comparan los resultados de ambos canales?