

2021

**Sistemas de Soporte de Toma de Decisiones / Soporte a
Decisiones de Negocios/ Minería de Datos**

Jueves Mañana

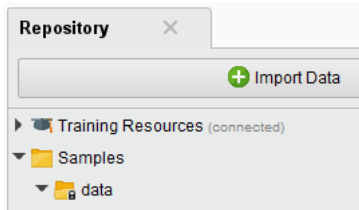


INGENIERÍA EN INFORMÁTICA

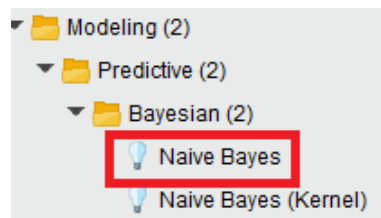
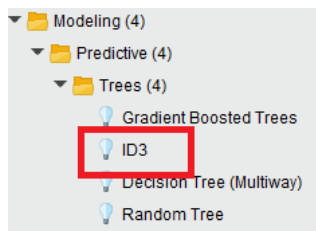
Trabajo Práctico Obligatorio – Data Mining

Explicación

El objetivo del presente trabajo es llevar adelante algunas pruebas de los algoritmos de data mining vistos en la materia. Para tal fin se utilizará el *dataset* “Titanic Training” y “Titanic Unlabeled”, ambos disponibles en el repositorio de RapidMiner *Samples->Data*



Para este trabajo práctico se utilizarán dos algoritmos: Árboles de decisión (ID3) y Naive Bayes. Tendrán que aplicar ambos algoritmos para predecir que pasajeros sobreviven y cuáles no.



Para entrenar ambos algoritmos utilizarán el *dataset* “Titanic Training”. Tienen que entrenar ambos algoritmos y comparar los resultados para ver cual predice mejor. Para esto tendrán que mostrar los resultados de la matriz de confusión. Si creen que algún gráfico complementario puede ayudar al entendimiento, lo pueden agregar.

Una vez obtenido el algoritmo que mayor predice ejecutarlo con el *dataset* “Titanic Unlabeled”. Describir los resultados. ¿Se puede comparar con el *dataset* real “Titanic” para ver qué tan cerca estuvimos de la realidad?

Entrega

- Se deberá entregar un informe con print screen de ambos procesos (en Rapidminer u otra herramienta o lenguaje).
- Se deberán mostrar los resultados de ambas predicciones y explicar cuál predice mejor y por qué llegaron a esa conclusión.
- Explicar la comparación con el *dataset* real.