

Trabajo Practico N° 2

Minería de datos

Objetivo

El objetivo de este trabajo practico es integrar los conocimientos adquiridos en las unidades 4 y 5 en dos problemas reales asociados uno al comportamiento financiero de 1000 empresas y otro a un juego de cartas.

Actividades

1. Descargar el conjunto de datos, 1000_Companies.csv¹, para realizar el trabajo práctico. Analizar los atributos del conjunto de datos (distribuciones, valores, outliers, tipos de datos, etc.) y elegir un método de estandarización.
2. Realizar la estimación del atributo Profit utilizando árboles de decisión (Regresión) analizando los parámetros máximo profundidad, número mínimo de observaciones, número mínimo de observaciones por separación y criterio de separación. Graficar el árbol obtenido en el proceso de entrenamiento y mostrar los resultados sobre dos conjuntos de test (Precisión, Exhaustividad y Exactitud).
3. Descargar el conjunto de datos, PokemonDB.csv², para realizar el trabajo práctico. Analizar los atributos del conjunto de datos (distribuciones, valores, outliers, tipos de datos, etc.)
4. Realizar la estimación del atributo Type utilizando árboles de decisión (Clasificación) analizando los parámetros máximo profundidad, número mínimo de observaciones, número mínimo de observaciones por separación y criterio de separación. Graficar el árbol obtenido en el proceso de entrenamiento y mostrar los resultados sobre dos conjuntos de test (Precisión, Exhaustividad y Exactitud).
5. Realizar la estimación del atributo Type utilizando Bayes Ingenuo. Aquí deberá considerar un criterio de división de los atributos para discretizarlos. Mostrar los resultados sobre dos conjuntos de test (Precisión, Exhaustividad y Exactitud).
6. Realizar la estimación del atributo Type utilizando k-NN analizando los parámetros cantidad de vecinos, métrica y valor de p. Mostrar los resultados sobre un conjunto de test (Precisión, Exhaustividad y Exactitud).

Presentación

La entrega es por grupos de dos estudiantes y se entregan dos archivos por grupo. Cualquier integrante del grupo puede hacer la entrega mediante el campus de la materia.

¹ <https://www.kaggle.com/datasets/subhamp7/company-profit-and-expenditures>

² <https://www.kaggle.com/datasets/rounakbanik/pokemon>

Trabajo Practico N° 2

Minería de datos

El informe deberá tener una carátula en la que se indique: año, materia, integrantes. Además, deberá contar con una sección de conclusiones al final del mismo.

El formato del informe deberá ser pdf con el código en formato py. También podrá enviarse en formato ipynb mientras que se cumplan las especificaciones previamente definidas.

Las entregas fuera del plazo establecido no serán consideradas salvo excepciones previamente justificadas por el grupo.