

Trabajo Practico N° 3

Minería de datos

Objetivo

El objetivo de este trabajo practico es integrar los conocimientos adquiridos en las unidades 5 y 6 en un problema real asociado a la determinación de la estación del año mediante la medición atributos característicos.

Actividades

1. Descargar el conjunto de dxWeather.csv¹, para realizar el trabajo práctico. Realizar un análisis exploratorio de datos: visualizar distribuciones, valores faltantes, correlaciones, etc. Limpiar el conjunto de datos (manejar valores faltantes, eliminar outliers) si es necesario. Codificar variables categóricas (si es necesario). Normalizar o estandarizar las características.
2. Realizar la predicción del atributo Estación utilizando máquinas de vectores con kernel lineal analizando el parámetro costo. Mostrar los resultados sobre los conjuntos de test (Precisión, Exhaustividad y Exactitud) utilizando validación cruzada con $k = 5$.
3. Realizar la predicción del atributo Estación utilizando máquinas de vectores con kernel gaussiano analizando los parámetros costo y gama. Mostrar los resultados sobre los conjuntos de test (Precisión, Exhaustividad y Exactitud) utilizando validación cruzada con $k = 5$.
4. Realizar la predicción del atributo Estación utilizando Random Forest analizando los parámetros cantidad de estimadores y la máxima profundidad de los árboles. Mostrar los resultados sobre los conjuntos de test (Precisión, Exhaustividad y Exactitud) utilizando validación cruzada con $k = 5$.

Nota: Para optimizar los parámetros debe elegir una partición de las cinco obtenidas.

Presentación

La entrega es por grupos de dos estudiantes y se entrega un archivo por grupo.

Cualquier integrante del grupo puede hacer la entrega mediante el campus de la materia.

El informe deberá tener una cabecera 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 en formato ipynb y no debe contener las definiciones teóricas ni el significado de los parámetros de los métodos dados en clase.

¹ <https://www.kaggle.com/datasets/nikhil7280/weather-type-classification/>

Trabajo Practico N° 3

Minería de datos

Las gráficas mostradas en el informe deben contener una explicación de lo observado y si es coherente que su hipótesis previa, por tanto, la cantidad de graficas debe estar acotadas y ser representativas.

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