

Instrucciones Miniproyecto 2:

Aplicación algoritmo Random Forest y KNN para un caso propuesto.

Descripción

El caso que se presentará a continuación consiste en analizar y generar una predicción de una base de datos de ventas de tiendas de *retail*. Se entrega información del ID de la tienda, fecha por semana, la temperatura promedio en la región que está ubicada, precio del combustible en la región, índice del precio al consumidor, tasa de desempleo, si es feriado o no y ventas semanales.

Lo que se pide en este miniproyecto es que aplique dos algoritmos a la base de datos propuesta: Random Forest y KNN. Deberá realizar predicción, analizar los resultados obtenidos aplicando distintos sets de parámetros y mostrar las diferencias en la predicción obtenida con cada algoritmo

Se espera como entregables un archivo .ipynb

Trabajo a realizar

Usando Python en jupyter notebook o Google Colab, los pasos para realizar este miniproyecto son:

1. Cargar y revisar la base de datos, ver su contenido y características, filtrar, limpiar, depurar, homogeneizar y normalizar. Debe justificar la toma de decisión.

2. Aplicar el algoritmo KNN al set de datos propuestos variando los parámetros. Probar con 3 set de parámetros diferentes.
3. Realizar visualización.
4. Realizar una predicción de los datos entregados en la tabla 1 con la configuración que resulta mejor según lo observado y justifique.
5. Aplicar el algoritmo Random Forest al set de datos propuestos variando los parámetros. Probar con 3 set de parámetros diferentes.
6. Realizar visualización de un árbol, y también entregar un gráfico que muestre la importancia de los atributos.
7. Realizar una predicción de los datos entregados en la tabla 1 con la configuración que resulta mejor según lo observado y justifique.
8. Analizar las diferencias entre las predicciones obtenidas por cada algoritmo.