

Práctica curso 24-25

Métodos Clásicos para Predicción

Grado en Ciencia de Datos e Inteligencia Artificial

Esta práctica consiste en dos apartados, uno en el que se aplicarán técnicas de reducción de dimensionalidad y otro en el que se aplicarán técnicas de predicción y análisis de series temporales.

Descarga de datos:

<https://drive.upm.es/s/Hu7ifQpXMSYwY8x>

<https://drive.upm.es/s/xpR82iBgLhJlWVX>

PROBLEMA 1: PROTEGIENDO LA ETSISI DE ATAQUES DE RED

En este apartado se proporciona un dataset “trazas_red.csv” de trazas de ataques de red categorizadas. Para cada traza se indica el tipo de ataque de red detectado con una etiqueta en la columna “label”.

Se pide lo siguiente:

1. Realiza un estudio cualitativo del dataset, describiendo lo máximo posible qué son las diferentes características y tipos de ataque de red. Es importante que, dado que existen muchos tipos de ataque, puedas agruparlos para lograr mejores explicaciones de los datos
2. Realiza un estudio cuantitativo de las características del dataset, indicando si existen variables con alta correlación
3. Haz uso de técnicas de reducción de dimensionalidad para lograr visualizaciones que permitan entender los datos. El objetivo fundamental es lograr visualizaciones que mejor consigan separar los diferentes tipos de ataque de red entre si.
4. Haz uso de técnicas de reducción de dimensionalidad para lograr una mejor clasificación de las trazas de red. Realiza una comparación exhaustiva contra diferentes clasificadores de scikit-learn, evaluando si permiten lograr mejores resultados

PROBLEMA 2: MEJORANDO LA EFICIENCIA ENERGÉTICA EN LA UPM

En este apartado se proporciona un dataset “consumo_UPM.csv”, con datos de consumo eléctrico detallados de Rectorado, el Campus de Montegancedo y el Campus Sur. Se pide lo siguiente:

1. Realiza un estudio detallado del dataset, cualitativo y cuantitativo
2. Realiza visualizaciones que nos permitan entender el consumo medio por bloque y campus en función de diferentes medidas de tiempo (días, semanas y meses). Describe los momentos temporales en los que se han producido eventos que han producido un importe efecto en la tendencia de consumo.

3. Desarrolla modelos de predicción de consumo en los 6 meses posteriores a la fecha de último dato recogido en el dataset. Haz un análisis detallado de los resultados obtenidos, y describe los problemas (si los hay) para conseguir cifras de acierto más altas.
4. Haz un análisis individualizado por cada uno de los bloques. Extrae la serie temporal y agrupa los datos por día. Aplica todas las técnicas de análisis de series temporales vistas en clase y realiza una comparativa y análisis exhaustivo de cada bloque.

Entrega

Se debe entregar un fichero comprimido .zip con el siguiente nombre:

APELLIDO1_APELLIDO2_GRUPO_XX.zip

donde APELLIDO1 y APELLIDO2 sean los primeros apellidos de los dos integrantes de la práctica, y XX el número asignado al grupo de prácticas en Moodle.

Este fichero contendrá:

- Memoria.pdf: una memoria detallada de todo el trabajo realizado, sin código
- ataques_red.ipynb: un notebook con todos los pasos realizados en el ejercicio 1
- consumo_upm.ipynb: un notebook con todos los pasos realizados en el ejercicio 2

La práctica se realizará en parejas, que deberán estar dadas de alta en Moodle.

Importante

Todos los notebooks y código deberán estar correctamente documentados utilizando celdas de markdown, explicando todos los pasos realizados, los resultados obtenidos y toda decisión tomada. Una explicación escasa o no suficientemente detallada conllevará el suspenso en la práctica.