2024-06-15

JOSE ANASTACIO HERNANDEZ SALDAÑA

Especificaciones de la práctica 1 Análisis exploratorio

Utilizando el dataset de los pagos de la UANL (<https://github.com/ppGodel/data_mining/blob/main/csv/typed_uanl.csv>) van a realizar un analisis exploratorio. Hay que crear una sript donde realices lo siguiente:

1. Identificar las entidades (que es lo que puede ser sujeto de estudio) del conjunto. agregalo como un comentario en tu script.
2. Obtener las estadisticas (descriptiva) de cada entidad (min, max, avg, std)
3. Hacer agrupaciones por las entidades y sacar estadisticas de las agrupaciones.
4. Crear imagenes de estas estadisticas, ya sean histogramas, graficas de pastel, etc, al menos 3 graficas diferentes, Hay que crear muchas imagenes, asi que creen ciclos y recorran las agrupaciones.
5. De alguna de las agrupaciones, hacer una pureba ANOVA, para ver si hay diferencias entre los elementos. recordando, hay que probar que las muestras son o no normales. si son normales anova para la prueba y t-student para saber quien es el diferente. Si la muestra no es normal, prueba con kruskall-wallis y tukey para saber quien es el diferente.

Realiza un reporte escrito de tus hallazgos. No quiero una descripcion de que hiciste, sino de que encontraste, cosas que no sabias y supiste una vez que comenzaste a analizar los datos, cuenta que cosas guiaron tu atencion para encontrar informacion que consideras significativa, y muestra esos hallazgos con imagenes, asi que prepara tus imagenes para que alguien las vea.

Hay que crear un repositorio de github/gitlab con 2 carpetas en la raiz del repositorio: Reportes y Practicas. En Reporte estara el reporte en formato PDF con el nombre Practica#.pdf donde # es el numero de la practica al que corresponde. En  la carpeta de practicas por cada practica habra una carpeta de nombre Practica# donde estara todo el codigo, imagenes, conjuntos de datos que usaste para la practica.

Link para pasarme tu repositorio:

<https://forms.office.com/r/YuvnZHMif0>

La fecha limite para subir el reporte y el codigo a git es el domingo 30 de junio.

like 6

typed\_uanl.csv

ppGodel/data\_mining