

Prueba Data Science (cerca de feina)

Descripción de la prueba

Bienvenidos a la prueba de evaluación para los egresados del curso de Data Science. En esta prueba, se espera que demuestres tus habilidades en el manejo de datos utilizando el lenguaje de programación Python. Tendrás a tu disposición tres conjuntos de datos diferentes, de los cuales deberás seleccionar uno para realizar el análisis correspondiente.

- Conjunto de datos 1: *Encuesta sobre la brecha digital en la ciudad de Barcelona*
Este dataset proporciona una amplia gama de información sobre las percepciones y características socioeconómicas de los habitantes de Barcelona en relacionados a la brecha digital.

El dataset se puede descargar en el siguiente enlace: <https://opendata-ajuntament.barcelona.cat/data/es/dataset/enquesta-escletxa-digital>

- Conjunto de datos 2: *Censo locales en planta baja destinados a actividad económica de la ciudad de Barcelona*
En este dataset se encuentra información del censo de locales en planta baja de la ciudad de Barcelona destinados a actividad económica y que se encuentran en activo o pendientes de actividad

El dataset se puede descargar en el siguiente enlace: <https://opendata-ajuntament.barcelona.cat/data/es/dataset/cens-locales-planta-baixa-act-economica>

- Conjunto de datos 3: *Accidentes según causa conductor gestionados por la Guàrdia Urbana a la ciutat de Barcelona*
En este dataset se encuentra información sobre los accidentes según causa conductor gestionados por la Guàrdia Urbana a la ciutat de Barcelona

El dataset se puede descargar en el siguiente enlace: https://opendata-ajuntament.barcelona.cat/data/es/dataset/accidents_causa_conductor_gu_bcn

En "Descripción complementaria y definición de campos" podrás obtener mayor información de cada conjunto de datos.

Tu tarea consistirá en seleccionar uno de estos conjuntos y aplicar uno de los siguientes modelos de aprendizaje automático:

- Algoritmos de Aprendizaje Supervisado: Regresión
- Algoritmos de Aprendizaje Supervisado: Clasificación
- Algoritmos de Aprendizaje NO Supervisado: Agrupación

En esta tarea se espera que demuestres tu capacidad para seleccionar el modelo de aprendizaje automático más adecuado y aplicarlo de manera efectiva al conjunto de datos seleccionado.

Criterios de Evaluación:

- Documentación del proceso: se evaluará su legibilidad, organización y buenas prácticas de programación.
- Correcta descripción del objetivo del estudio.
- Calidad del análisis exploratorio de los datos.
- Selección del modelo de aprendizaje automático en función del objetivo del estudio.
- Proceso realizado en la depuración de datos.
- Correcta interpretación de los resultados.

Recuerda que en la ciencia de datos es importante justificar las decisiones tomadas, interpretar los resultados y comunicarlos de manera efectiva.

Presentación del Trabajo:

Una vez completado el análisis, debes presentar tus resultados en un enlace a tu repositorio GitHub.

En el repositorio, se debe incluir:

- Un archivo Jupyter Notebook con todas las líneas de códigos ejecutadas y las interpretaciones que consideres necesarias.
- Un archivo README con la siguiente información:
 1. Introducción: Presentación del conjunto de datos y de las variables seleccionadas.
 2. Depuración de datos: Descripción detallada de las técnicas de preprocesado aplicadas y los criterios de evaluación utilizados.
 3. Resultados: Presentación de los resultados obtenidos.
 4. Conclusiones: Principales inferencias derivadas de los resultados obtenidos.

Recuerda que la claridad, organización y el procedimiento realizado en la tarea serán aspectos clave en la evaluación de tu desempeño. ¡Buena suerte!