

Proyecto Final: Asteroids Orbits Data Elective-I

Instrucciones:

- Fecha de entrega #1: 22 de Marzo de 2025
- Fecha de entrega #2: 29 de Mayo 2025
- Medio de entrega: Sustentación del proyecto final plataforma y video.

Conjunto de datos: Contiene datos sobre posibles impactos de asteroides con la Tierra.

<https://www.kaggle.com/datasets/nasa/asteroid-impacts>

Ejercicio:

- ° El proyecto final estará enfocado en entrenar un modelo de Machine Learning. Por ahora, para este avance se pide lo siguiente: Se debe encontrar un conjunto de datos de mínimo 5000 registros, plantear un problema, realizar un análisis exploratorio y usar el algoritmo Kmeans para encontrar patrones en los datos.
- ° Documento de máximo 2 páginas con formato IEEE, donde se indique: problema, link del conjunto de datos, análisis exploratorio, explicación de lo realizado en el avance y cuáles serán los siguientes pasos a realizar.
- ° Video de máximo 3 minutos, explicando el problema y la posible solución.

Entregables:

- Análisis exploratorio en notebook (10 puntos).
- Preprocesamiento de los datos (15 puntos).
- Entrenamiento del Modelo de Machine Learning con redes neuronales (10 puntos).
- Evaluación del modelo 75% Accuracy (10 puntos).

Proyecto Final: Asteroids Orbits Data Elective I.

Asteroids Orbits Data Elective-I: <https://github.com/dg2c4/Asteroids-Orbits-Data-Elective-I>

Problema:

¿Cuáles son los asteroides con mayor probabilidad de impacto con la Tierra?

El problema central es identificar y analizar los asteroides que representan un riesgo potencial de impacto con la Tierra, basándose en sus características orbitales y físicas. Esto incluye determinar cuáles son los asteroides más peligrosos, cuáles tienen órbitas más cercanas a la Tierra y cuáles podrían representar una amenaza en el futuro.

Objetivo:

Desarrollar un análisis completo que permita extraer datos sobre posibles impactos de asteroides con la Tierra, utilizando herramientas de visualización de datos para presentar los resultados de manera efectiva. El trabajo se divide en tres componentes principales:

Análisis Exploratorio:

- Formulación de preguntas de investigación.
- Creación de hipótesis basadas en un problema específico.
- Análisis inicial del conjunto de datos seleccionado.

Preprocesamiento de Datos:

- Limpieza de datos.
- Transformación de variables.
- Reducción de datos cuando sea necesario.
- Discretización de datos según corresponda.

Modelado y Evaluación:

- Implementación de modelo de regresión lineal o logística.
- Selección y cálculo de métricas de evaluación.
- Análisis del rendimiento del modelo.

Systems engineering:

User: David Gutierrez Chaves

Code: 506222728

Subject: Big Data Electiva-I

Institution: Fundación Universitaria Konrad Lorenz

Institutional email: david.gutierrec@konradlorenz.edu.co

Created By: <https://github.com/dg2c4>