

Asteroids Orbits Data Final Project

Fundación Universitaria Konrad Lorenz. Big Data. Electiva-I.

Estudiante: David Gutierrez. Cod: 506222728

I. INFORME DE ANÁLISIS DE POSIBLES IMPACTOS DE ASTEROIDES CON LA TIERRA

II. INTRODUCCIÓN

El análisis exploratorio de datos (EDA) es una fase crucial en cualquier proyecto de ciencia de datos, ya que permite comprender la estructura, las relaciones y las tendencias subyacentes en los datos. En este caso, el conjunto de datos proporcionado contiene información sobre asteroides, incluyendo sus características orbitales, clasificaciones y distancias mínimas de intersección con la Tierra. Este análisis tiene como objetivo principal identificar patrones y relaciones que puedan ser útiles para predecir posibles impactos de asteroides y comprender mejor su comportamiento.

III. OBJETIVO GENERAL:

Realizar un análisis exploratorio completo del conjunto de datos de asteroides para identificar patrones, tendencias y relaciones que puedan ser útiles para la predicción de impactos y la clasificación de asteroides.

IV. OBJETIVOS ESPECÍFICOS:

- Explorar y limpiar el conjunto de datos para asegurar la calidad de los datos.
- Identificar las características más relevantes que influyen en la clasificación de los asteroides.
- Visualizar las distribuciones de las variables clave y sus relaciones.
- Formular hipótesis basadas en los datos y validarlas mediante análisis estadístico.
- Aplicar modelos de machine learning para la clasificación y agrupación de asteroides.

Palabras Clave: Asteroides, impacto, clasificación, análisis exploratorio, visualización de datos, machine learning, clustering, normalización, extracción de características.

Herramientas: Python, Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Jupyter Notebook.

Conjunto de datos: Contiene datos sobre posibles impactos de asteroides con la Tierra
<https://www.kaggle.com/datasets/nasa/asteroid-impacts>

V. ELEMENTO DE LOS ENCABEZADOS:

- **Epoch (TDB):** Época en la que se midieron los datos orbitales, expresada en Tiempo Dinámico Bariocéntrico (TDB), una escala de tiempo utilizada en astronomía para cálculos de alta precisión.

- **Orbit Axis (AU):** Semieje mayor de la órbita del objeto, medido en unidades astronómicas (AU), que es la distancia promedio entre el objeto y el Sol.
- **Orbit Eccentricity:** Excentricidad de la órbita, que indica cuánto se desvía la órbita de ser circular. Un valor de 0 es una órbita circular, mientras que valores cercanos a 1 indican órbitas más alargadas.
- **Perihelion Argument (deg):** Argumento del perihelio, medido en grados, que es el ángulo entre el nodo ascendente y el perihelio de la órbita del objeto.
- **Node Longitude (deg):** Longitud del nodo ascendente, medida en grados, que es el ángulo entre el punto de referencia (generalmente el equinoccio vernal) y el nodo ascendente de la órbita.
- **Mean Anomaly (deg):** Anomalía media, medida en grados, que es la posición angular del objeto en su órbita, medida desde el perihelio, asumiendo una órbita circular.
- **Perihelion Distance (AU):** Distancia más cercana al Sol en la órbita del objeto, medida en unidades astronómicas (AU).
- **Aphelion Distance (AU):** Distancia más lejana al Sol en la órbita del objeto, medida en unidades astronómicas (AU).
- **Orbital Period (yr):** Período orbital, medido en años, que es el tiempo que tarda el objeto en completar una órbita alrededor del Sol.
- **Minimum Orbit Intersection Distance (AU):** Distancia mínima de intersección orbital, medida en unidades astronómicas (AU), que es la distancia más cercana a la que la órbita del objeto se acerca a la órbita de la Tierra.

VI. PROBLEMA PLANTEADO:

El problema principal es comprender las características de los asteroides que podrían influir en su probabilidad de impacto con la Tierra. Además, se busca identificar patrones que permitan clasificar y agrupar asteroides basándose en sus propiedades orbitales y físicas.

VII. PREGUNTAS DE INVESTIGACIÓN:

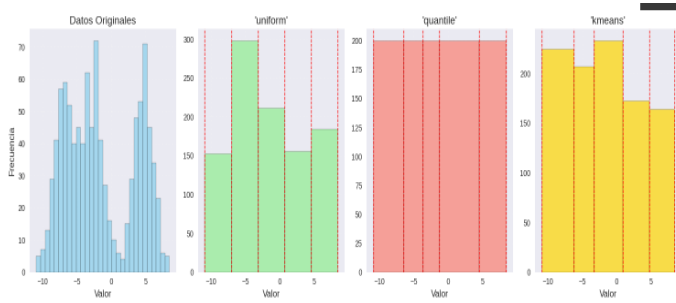
- 1) ¿Cuáles son las características más relevantes que influyen en la clasificación de los asteroides?
- 2) ¿Existe una relación entre la distancia mínima de intersección y la clasificación del asteroide?
- 3) ¿Cómo se distribuyen los asteroides en función de su excentricidad orbital y su inclinación?

- 4) ¿Es posible predecir la clasificación de un asteroide basándose en sus características orbitales?.

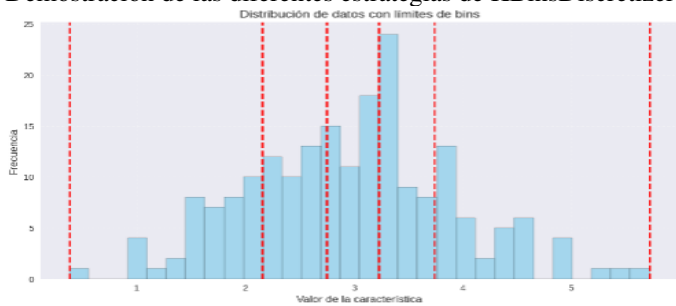
VIII. HIPÓTESIS:

- 1) Los asteroides con una distancia mínima de intersección más cercana a la Tierra tienen una mayor probabilidad de ser clasificados como peligrosos.
- 2) La excentricidad orbital y la inclinación están relacionadas con la clasificación de los asteroides.
- 3) Los asteroides con períodos orbitales más cortos tienen una mayor probabilidad de ser clasificados como peligrosos.

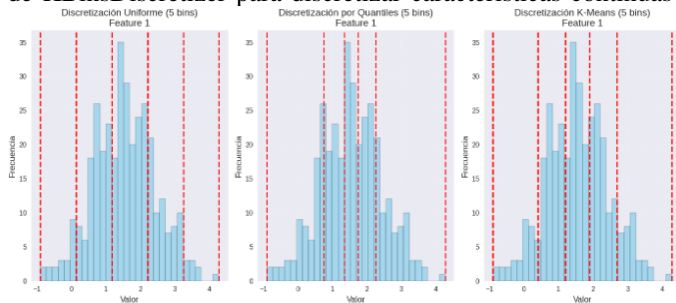
IX. MODELOS MACHINE LEARNING PREPROCESSING



Demonstrating the different strategies of KBinsDiscretizer- Demostración de las diferentes estrategias de KBinsDiscretizer

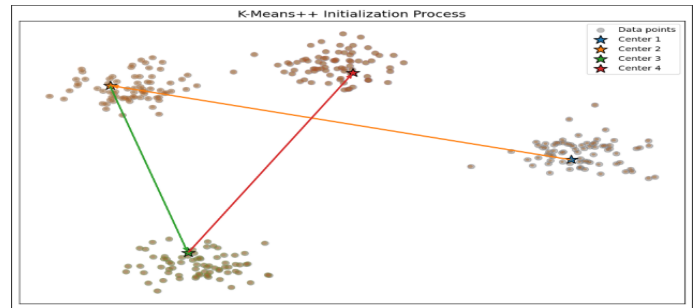


Using KBinsDiscretizer to discretize continuous features- Uso de KBinsDiscretizer para discretizar características continuas

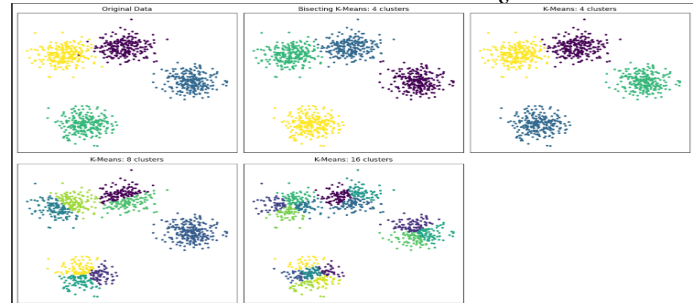


Feature discretization-Discretización de características

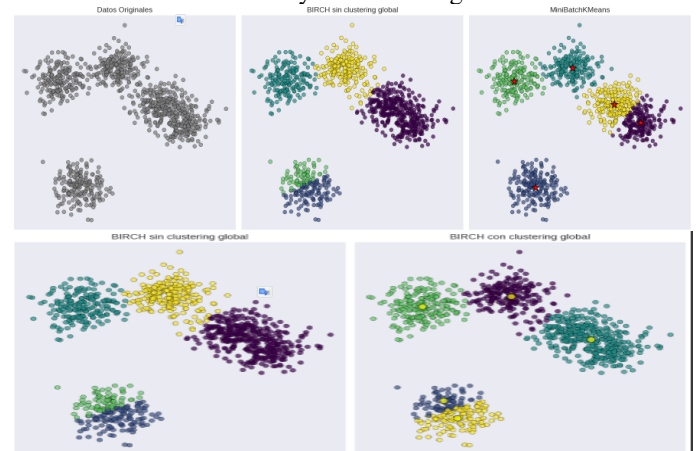
X. MODELOS MACHINE LEARNING CLUSTERING K-MEANS



An example of K-Means++ initialization-Ejemplo de inicialización de K-Means++.



Bisecting K-Means and Regular K-Means Performance Comparison-Comparación del rendimiento de K-Means de bisección y K-Means regular



Compare BIRCH and MiniBatchKMeans-Comparar BIRCH y MiniBatchKMeans.

XI. CONCLUSIONES

- Las características orbitales, como la distancia mínima de intersección, la excentricidad y la inclinación, son factores clave en la clasificación de los asteroides. Se identificaron patrones que sugieren que los asteroides con órbitas más excéntricas y cercanas a la Tierra tienen una mayor probabilidad de ser clasificados como peligrosos. Los modelos de machine learning aplicados demostraron ser efectivos para la clasificación y agrupación de asteroides, lo que podría ser útil para futuras investigaciones y aplicaciones en la predicción de impactos.