

Dataset Meteorite Landings

Wilmer Santiago Soto Vidal

Kaggle: <https://www.kaggle.com/datasets/nasa/meteorite-landings>

Contexto de aplicación.

El dataset que escogí es el de Meteorite Landings el cual es un dataset de la Nasa acerca de meteoritos registrados.

Se encuentra información como nombre, masa, ubicación, composición, año y si fue observada su caída o no.

Objetivo de machine learning (queremos predecir X, dada tal información)

Objetivo General: Analizar y predecir los patrones espaciales y temporales de las caídas de meteoritos.

Objetivos específicos.

1. **Predicción de probabilidad de caída por región**
 - 1.1. **Predicción espacial - Clasificación**

Dado las características geográficas y espaciales de una región, quiero predecir la probabilidad de que esa región sea clasificada como Alto o Bajo Riesgo de caídas de meteoritos.
 - 1.2. **Predicción espacial - Regresión**

Dado las características geográficas y espaciales de una región, quiero predecir el número esperado de caídas de meteoritos dentro de esa región.
2. **Predicción de tendencias históricas.**

Dado el historial de la frecuencia de caídas de meteoritos a lo largo de los años, quiero predecir la frecuencia esperada (número de caídas) en futuros períodos de tiempo.

Dataset: tipo de datos, tamaño (número de datos y tamaño en disco), distribución de las clases

Se compone de 45 716 datos con 10 columnas. Con un peso de .csv de 4.21 mb

El dataset posee las siguientes columnas,:

- **name:** el nombre del meteorito que suele ser típicamente una ubicación modificada con un número, año, su composición, etc..
- **id:** identificador único del meteorito

- **nametype:** uno de estos dos:
 - **valid:** un meteorito típico
 - **relict:** un meteorito que se ha degradado fuertemente debido al clima en la tierra
- **recclass:** la clase del meteorito, es una del gran número de clases basadas en características físicas, químicas y otras. referencia wikipedia: [meteorite classification](#)
- **mass:** la masa del meteorito en gramos
- **fall:** si el meteorito fue visto cayendo o si fue descubierto después de su impacto, es uno de estos:
 - **Fell:** se observó la caída del meteorito
 - **Found:** no se observó la caída del meteorito
- **year:** el año cuando impactó o el año cuando se encontró (depende del valor de fall)
- **reclat:** la latitud donde impactó
- **reclong:** la longitud donde impactó
- **GeoLocation:** una tupla que junta la latitud y longitud registrada.

Métricas de desempeño (de machine learning y negocio)

Para cada predicción entonces se manejan diferentes métricas de desempeño.

1. Predicción de probabilidad de caída por región

1. Predicción espacial - Clasificación

- ROC AUC
- Recall
- Precision
- F1-Score
- Matriz de confusión

2. Predicción espacial - Regresión

- MAE. Error absoluto medio.
- RMSE. Raíz del error cuadrático medio
- SMAPE. Error Porcentual Absoluto Medio Simétrico

2. Predicción de tendencias históricas.

- MAE. Error absoluto medio.
- RMSE. Raíz del error cuadrático medio
- SMAPE. Error Porcentual Absoluto Medio Simétrico

Para temas de negocio, estas predicciones ayudan a diferentes cosas.

- **Optimizar la búsqueda de meteoritos**

Como se ve en el dataset, gran parte (98%) de los meteoritos son encontrados y no directamente evidenciados. Un modelo que use el espacio puede ayudar a identificar zonas donde son más frecuentes las caídas de meteoritos.

- **Previsión global de caídas.**

Ayuda a predecir el # de caídas por año.

- **Identificar y analizar posibles patrones en los datos.**

El análisis del dataset mediante ML puede causar que encontremos patrones de los datos que posiblemente sean de ayuda. Por ejemplo, donde suelen caer más los meteoritos o si la altura/elevación influye.

Referencias y resultados previos

[Dataset Meteorite Landings](#)

[Notebook previo realizado por alguien más en Kaggle \(Meteorite-EDA Analysis & Predictive Modeling\)](#)