



Exploración de datos en repositorios públicos

Javier Díaz Machado

Paso 1: Accede a los siguientes sitios web:

- Instituto Nacional de Estadística (INE)
- Eurostat
- Organización Mundial de la Salud (OMS) - Estadísticas de Salud Mundial
- UCI Machine Learning Repository
- Kaggle

Plataforma seleccionada: **Kaggle**

Paso 2: Explora las secciones de datos disponibles en estos sitios y selecciona una categoría o conjunto de datos que te interese, como estadísticas de salud, datos de la Unión Europea o conjuntos de datos de aprendizaje automático.

DataSet seleccionado: **Road Traffic Accidents**

Link al DataSet: <https://www.kaggle.com/datasets/saurabhshahane/road-traffic-accidents>

Paso 3: Examina los atributos de los conjuntos de datos que elijas, prestando especial atención a la fuente de los datos, fechas de obtención, actualización y validez, así como la metodología de recogida o muestreo.

Fuente de los datos:

Registros manuales de datos de accidentes de tráfico de la ciudad de **Addis Abeba**, proporcionado por **Tarikwa Tesfa** y publicado por **SAURABH SHAHANE**.

Fechas de obtención de los datos: **2017-20:**

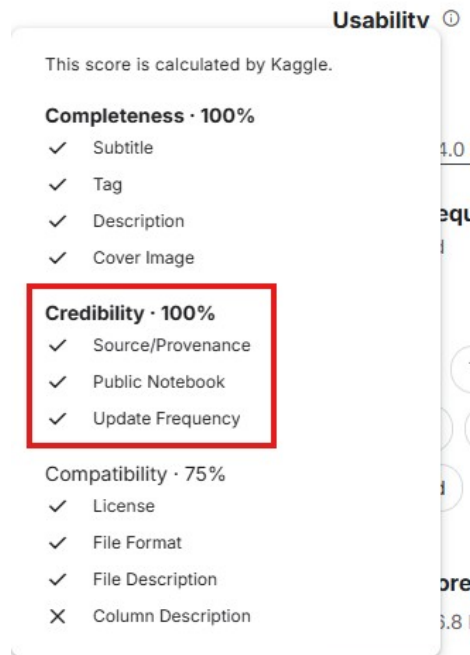
Content

The data set has been prepared from manual records of **road traffic accident of the year 2017-20**. All the sensitive information have been excluded during data encoding and finally it has 32 features and 12316 instances of the accident. Then it is preprocessed and for identification of major causes of the accident by analyzing it using different machine learning classification algorithms. RTA Dataset.csv is the dataset before preprocessing and cleaned.csv is the preprocessed dataset.

Frecuencia de Actualización: **Sin especificar (Última vez, hace 3 años)**

Validez de los datos:

Tiene una puntuación del 100% en credibilidad:



Paso 4: Anota la información relevante sobre el conjunto de datos que hayas seleccionado.

Datos proporcionados por el dataSet:

- Hora
- Day_of_week
- Age_band_of_driver
- Sex_of_driver
- Educational_level
- Vehicle_driver_relation
- Driving_experience
- Type_of_vehicle
- Owner_of_vehicle
- Service_year_of_vehicle

Este dataSet se podría utilizar para saber bajo qué situaciones hay más accidentes de tráfico, para así poder emplear campañas de la Dirección General de Tráfico de forma localizada.

Por ejemplo, si los viernes hay más accidentes, se pueden emplear más esfuerzos en reducir estas cifras con campañas todos los viernes.