

Ejercicio Modelado de redes

En este ejercicio tendréis que reflexionar sobre las distintas redes que se pueden crear a partir de un conjunto de datos. Es importante hacer énfasis en qué información codificada cada red y qué análisis nos permite realizar. A continuación, os proponemos los siguientes casos de uso.

Casos de uso

Reviews de cerveza:

<https://www.kaggle.com/datasets/thedevastator/1-5-million-beer-reviews-from-beer-advocate/>

¿Te gusta la cerveza? ¿Te gusta probar cervezas nuevas y calificarlas? ¡Bueno, este conjunto de datos es perfecto para ti! Contiene 1.5 millones de reseñas de cervezas de Beer Advocate, lo que lo hace ideal para poner a prueba tus habilidades con los datos. Los datos incluyen calificaciones en cinco aspectos: apariencia, aroma, paladar, sabor e impresión general. Las reseñas también incluyen información del producto y del usuario. Entonces, ¿a qué estás esperando? ¡Comienza hoy mismo!

Papers publicados en arxiv:

<https://www.kaggle.com/datasets/Cornell-University/arxiv>

Este conjunto de datos es una submuestra del extenso repositorio de ArXiv y presenta metadatos de artículos en formato JSON. Incluye detalles de los artículos como el ID de ArXiv, el remitente, los autores, el título, comentarios, información sobre la revista, DOI, resumen, categorías y el historial de versiones, proporcionando una referencia concisa para investigadores.

Datos de transporte:

<https://www.kaggle.com/datasets/carlosborrajo/madrid-bike-sharing-dataset>

Este conjunto de datos proporciona información sobre el uso del servicio de bicicletas compartidas BiciMad durante 2022. Cada fila representa una interacción entre un usuario y el sistema (un viaje entre estaciones). Los metadatos incluyen detalles como la fecha del viaje, el ID de la bicicleta, la flota, la duración del viaje, la geolocalización al inicio y final del viaje, direcciones, hora de inicio y finalización del viaje, estado de las bicicletas, información de las estaciones y detalles de anclaje.

Diálogos de películas:

<https://www.kaggle.com/datasets/kornflex/harry-potter-movies-dataset>

Contiene transcripciones de los diálogos entre los personajes de las películas de Harry Potter. Para cada película se conoce su título de la película, la taquilla que hizo, su año de estreno ... etc. Además, se cuenta con la transcripción de la película, cada película está dividida en capítulo, cada capítulo tiene el nombre del personaje que hable y el texto de lo que dice.

Foros de reddit:

<https://www.kaggle.com/datasets/kaggle/reddit-comments-may-2015>

El conjunto de datos de comentarios de Reddit de mayo de 2015 ofrece una instantánea de las interacciones en Reddit con 1,7 mil millones de comentarios. Incluye varios campos como marcas de tiempo, votos positivos, información sobre subreddits, detalles del autor y contenido del comentario. Este conjunto de datos es un recurso valioso para el procesamiento del lenguaje natural y el análisis de datos, brindando información sobre las discusiones de primavera en Reddit.

Dataset medico:

<https://www.kaggle.com/datasets/prasad22/healthcare-dataset>

Este conjunto de datos contiene información detallada sobre pacientes, sus ingresos y servicios de atención médica. Incluye nombres de pacientes, edades, géneros, tipos de sangre, condiciones médicas, fechas de ingreso, médicos a cargo, nombres de hospitales, proveedores de seguros, montos de facturación, números de habitación, tipos de ingreso, fechas de alta, medicamentos recetados y resultados de pruebas. Este conjunto de datos es valioso para diversas tareas de análisis y modelado en el ámbito de la salud.