API de Kaggle

Kaggle es una plataforma web para ciencia de datos. En Kagle se pueden encontrar bases de datos, cuadernos de Python con análisis de datos, concursos de analítica de datos y muchas herramientas para ejecutar proyectos de ciencia de datos, aprender e incluso trabajar. Kaggle se puede utilizar para buscar bases de datos, explorar modelos y construirlos en una interfaz web, compartir datos y análisis con otras personas y participar en competencias sobre diferentes temas.

Utilizar la API de Kaggle proporciona varias ventajas para los usuarios que desean interactuar con la plataforma de Kaggle de manera programática, como por ejemplo:

Acceso a conjuntos de datos:

Los usuarios pueden acceder a una amplia variedad de conjuntos de datos alojados en Kaggle mediante la API. Esto facilita la obtención de datos para análisis, modelado, investigación y desarrollo de proyectos.

Integración con flujos de trabajo:

La API de Kaggle se puede integrar con flujos de trabajo existentes en Python u otros lenguajes de programación, lo que permite a los usuarios incorporar datos y funcionalidades de Kaggle en sus procesos de análisis de datos y aprendizaje automático.

Facilidad de acceso a competiciones:

Los participantes en competiciones de Kaggle pueden usar la API para descargar datos de entrenamiento y prueba, enviar predicciones y realizar un seguimiento del rendimiento de sus modelos de forma programática, lo que facilita la gestión y el seguimiento de sus participaciones.

Para ingresar a Kaggle se debe visitar el link www.kaggle.com y luego iniciar sesión con una cuenta de correos. Es necesario que los estudiantes creen una cuenta ya que por medio de esta plataforma descargarán los datos necesarios en diferentes actividades del curso.

Una vez creada la cuenta, vamos a crear un entorno virtual de Python y allí instalaremos la librería de kaggle mediante el comando pip install kaggle.

Para poder realizar descargas, necesitamos un token. Este se puede obtener en la página de kaggle haciendo click en el perfil de usuario (https://www.kaggle.com/settings/account) y luego en “create new token”. Se descargará un archivo .json que, si estamos en sistemas Windows, se debe almacenar en la ruta: C:\Users\<Windows-username>\.kaggle\kaggle.json

Posteriormente, abrir una consola en la carpeta en la que queramos guardar el conjunto de datos, para obtener datos de kaggle, basta con copiar los comandos desde la página web, como se aprecia en la figura 1, empleando el comando “copy API command” Para este ejemplo descargaremos el dataset de aerolíneas, disponible en el link: https://www.kaggle.com/datasets/saadharoon27/airlines-dataset el comando que permite la descarga desde consola será:

kaggle datasets download -d saadharoon27/airlines-dataset

Como ejercicio se propone localizar el dataset de iris y el de Titanic y descargarlos mediante la línea de comandos. Luego validar que los datos descargados si pertenezcan al conjunto de datos esperado.