

Programa de formación Machine Learning and Data Science MLDS



Módulo 5

Deep learning: Introducción al Aprendizaje Profundo con Python

Proyecto Aplicado
Conjunto de Datos

Conjunto de datos

El planteamiento y desarrollo del proyecto no es restrictivo en la elección del conjunto de datos. Si lo desea puede realizar un proyecto personal/laboral que aplique los conceptos y herramientas discutidas en el transcurso del curso. Si tiene dificultad para encontrar un tema para su proyecto, dejamos a su disposición algunos temas propuestos de los que se puede guiar para llevar a cabo la actividad.

A continuación, se presentan algunas plataformas con distintos *dataset* disponibles:

Kaggle:

[Kaggle](#) es una plataforma en línea que ofrece competencias de aprendizaje automático y una comunidad de data scientists y analistas de datos. Kaggle es una excelente fuente de datos para proyectos de *Deep Learning* por varias razones:

1. **Gran variedad de conjuntos de datos:** Kaggle cuenta con una amplia variedad de conjuntos de datos en diferentes áreas, desde la visión por computadora hasta el procesamiento del lenguaje natural. Esto significa que puedes encontrar conjuntos de datos que se adapten a tu proyecto específico.
2. **Conjuntos de datos etiquetados:** Muchos conjuntos de datos en Kaggle están etiquetados, lo que significa que los datos están clasificados y organizados de tal manera que los algoritmos de *Deep Learning* pueden aprender a partir de ellos. Esto es particularmente útil para proyectos de clasificación o detección de objetos.
3. **Acceso a la comunidad de Kaggle:** Kaggle tiene una comunidad activa de científicos de datos y entusiastas de *Machine Learning* que pueden ayudarte en tu proyecto. Puedes hacer preguntas en el foro, discutir ideas y colaborar en proyectos en equipo.
4. **Desafíos y competencias:** Kaggle organiza regularmente desafíos y competencias de *Machine Learning* en los que puedes participar para medir tus habilidades y aprender de otros expertos. Estos desafíos son una excelente manera de aprender nuevas técnicas de *Deep Learning* y aplicarlas a problemas reales.

Google Dataset Search:

[Google Dataset Search](#) es una buena fuente de datos para proyectos de *Deep Learning* por varias razones:

1. **Gran variedad de conjuntos de datos:** Brinda una gran variedad de conjuntos de datos gracias a su capacidad para buscar y organizar datos de manera efectiva, así como por la amplia gama de fuentes que utiliza para recopilar los datos. De esta manera, se asegura de ofrecer una amplia gama de conjuntos de datos de diferentes campos y áreas de interés.
2. **Acceso a conjuntos de datos académicos:** Google Dataset Search incluye conjuntos de datos de instituciones académicas y de investigación de todo el mundo. Estos conjuntos de datos suelen ser de alta calidad y están bien documentados.

Conjunto de datos

3. **Documentación detallada:** Muchos conjuntos de datos en Google Dataset Search están bien documentados, lo que significa que puedes entender fácilmente la estructura y contenido del conjunto de datos antes de descargarlo.
4. **Gratis:** Muchos de los conjuntos de datos en Google Dataset Search son gratuitos, lo que significa que no tienes que preocuparte por los costos asociados con el acceso a los datos

Visual Data:

[Visual Data](#) es una plataforma que se especializa en proporcionar conjuntos de datos de alta calidad para proyectos de visión por computadora. La plataforma ofrece una amplia variedad de conjuntos de datos etiquetados y diseñados específicamente para aplicaciones de visión por computadora, como el reconocimiento facial, la detección de objetos y el seguimiento de objetos en movimiento.

Visual Data es una buena fuente de datos para proyectos de *Deep Learning* por varias razones:

1. **Alta calidad de los datos:** Los conjuntos de datos en Visual Data son de alta calidad y están diseñados específicamente para aplicaciones de visión por computadora. Esto significa que los datos son útiles y relevantes para la tarea que deseas abordar.
2. **Amplia variedad de conjuntos de datos:** Visual Data ofrece una amplia variedad de conjuntos de datos en diferentes áreas, desde el reconocimiento facial hasta la detección de objetos en imágenes satelitales.
3. **Apoyo al usuario:** Visual Data ofrece soporte al usuario, lo que significa que puedes obtener ayuda y asistencia para seleccionar y utilizar los conjuntos de datos adecuados para tu proyecto.

Papers With Code:

[PaperswithCode](#) es una excelente fuente de datos para proyectos de *Deep Learning* por varias razones:

1. **Acceso a una amplia variedad de algoritmos y modelos de Deep Learning:** PaperswithCode es una plataforma en línea que recopila y organiza una gran cantidad de artículos científicos en el campo del aprendizaje automático. La plataforma también proporciona código abierto asociado con estos artículos, lo que significa que puedes acceder a una amplia variedad de modelos y algoritmos de *Deep Learning*.
2. **Validación de resultados:** PaperswithCode también ofrece una validación de resultados asociada con los modelos y algoritmos de *Deep Learning* que presenta. Esto significa que puedes comparar y evaluar el rendimiento de diferentes modelos en diferentes conjuntos de datos y métricas.

Conjunto de datos

3. **Referencias y citas:** Los artículos científicos en PaperswithCode también tienen referencias y citas que pueden ser útiles para investigar más a fondo un problema o una técnica de aprendizaje automático específica.
4. **Comunidad activa:** PaperswithCode tiene una comunidad activa de científicos de datos y entusiastas de *Machine Learning* que pueden ayudarte en tu proyecto. Puedes hacer preguntas en el foro, discutir ideas y colaborar en proyectos en equipo.
5. **Actualizaciones regulares:** La plataforma se actualiza regularmente con nuevos artículos, códigos y resultados de modelos de *Deep Learning*.

Programa de formación Machine Learning and Data Science MLDS

Facultad de
I N G E N I E R Í A
Sede Bogotá



UNIVERSIDAD
NACIONAL
DE COLOMBIA