

Day-18: Hugging Face and Pre-Trained Models

What is Hugging Face?

- It is an emoji 🤗
- It is the GitHub of the ML world
- A collaborative platform - build, train and deploy using open-source code
- That means Hugging Face comes with pre-trained models
- So developers don't have to start from scratch
- Just load and fine-tune
- The AI community for building the future

Key Components

Transformers Library

Consists of an extensive collection of pre-trained models

- text classification
- language generation
- translation
- summarization

``pipeline()`` method

allows users to easily apply complex models

models are easily accessible and customizable

Datasets Library

a vast repository of NLP datasets

diverse collection of datasets that can be used to train, test, and benchmark

Tokenizers

Tokenizers are crucial in NLP

responsible for converting text into a format that machine learning models can understand

transformation of tokens into vector representations for model input

Hugging Face provides a range of user-friendly tokenizers

Let's Explore

Models

Spaces

Datasets

Hands-on: Text classification with BERT