Sentiment-analysis-case-study

Author: Santhi Maddipudi

GitHub: https://github.com/santhimaddipudi/Sentiment-analysis-case-study

Sentiment analysis is the automated process of tagging data according to their sentiment, such as positive, negative and neutral. Sentiment analysis allows companies to analyze data at scale, detect insights and automate processes.

we will use the IMDB dataset to fine-tune a DistilBERT model for sentiment analysis.

The IMDB dataset contains 25,000 movie reviews labeled by sentiment for training a model and 25,000 movie reviews for testing it. DistilBERT is a smaller, faster and cheaper version of BERT. It has 40% smaller than BERT and runs 60% faster while preserving over 95% of BERT's performance. You'll use the IMDB dataset to fine-tune a DistilBERT model that is able to classify whether a movie review is positive or negative. Once you train the model, you will use it to analyze new data!

Tech stack:

Python

PyTorch (or TensorFlow – depending on what you used)

Hugging Face Datasets

Scikit-learn (metrics, evaluation)