Sentiment Analysis with VADER and DistilRoBERTa

This repository contains code for performing Sentiment Analysis on Twitter data using both VADER sentiment analysis and a fine-tuned DistilRoBERTa model.

Sentiment analysis, also known as opinion mining, is the process of determining the sentiment or emotional (Positive, Negative or Neutral) tone expressed in a piece of text.

 # Overview
- Data loading and preprocessing.
- VADER sentiment analysis.
- Definition of a custom PyTorch dataset class.
- Initialization of the DistilRoBERTa model and tokenizer.
- Model training and evaluation.
- Visualization of confusion matrices for both VADER and DistilRoBERTa.
- Accuracy comparison between VADER and DistilRoBERTa.

 # Dependencies
Ensure you have the required libraries installed.

 # Technologies Used:
This project is built using the following technologies:

Python: An interpreted high-level general-purpose programming language. Transformers: A powerful library for natural language processing (NLP) tasks.DistilRoberta model is used.

 # Data (Available in Repository)
Training data: twitter_training.csv
Testing data: twitter_validation.csv

 # Model Details
The code uses the "cardiffnlp/twitter-roberta-base-sentiment-latest" pre-trained DistilRoBERTa model fine-tuned for sentiment analysis on Twitter data.

 # Results
VADER Accuracy: 0.4819
DistilRoBERTa Accuracy: 0.7548