**Tweet Emotion Detection using Transformer Models**

**Overview**

This project focuses on **tweet sentiment classification** using **Hugging Face Transformers**. Three different models are fine-tuned and compared for performance:

* **Albert Base**
* **DistilBERT**
* **RoBERTa**

Each model is trained to classify tweets into multiple emotion categories, such as *joy, sadness, anger, optimism*, and more.

**Models Used**

This project fine-tunes three different transformer models for tweet sentiment classification:

* **ALBERT (A Lite BERT)**: A lightweight version of BERT with reduced memory usage and faster performance.
* **DistilBERT**: A smaller, faster version of BERT that retains 97% of its accuracy while being 60% faster.
* **RoBERTa**: An optimized BERT model trained with dynamic masking and larger batch sizes for improved performance.

**Features**

* **Multi-Label Classification**: A tweet can express multiple emotions at the same time.
* **Pre-trained Transformers**: Fine-tuned **Albert Base, DistilBERT, and RoBERTa** for NLP tasks.
* **Tokenization**: Used Hugging Face tokenizers to preprocess text efficiently.
* **Training & Evaluation**: Models trained with Hugging Face's Trainer API and evaluated using custom metrics.
* **Experiment Tracking**: Logged results with **Weights & Biases (WandB)** for performance analysis.

**Installation**

Ensure all necessary libraries are installed:

pip install torch transformers evaluate wandb datasets accelerate

**Dataset**

This project uses a **tweet-based sentiment analysis dataset** that classifies tweets into multiple emotions:

* **Columns:** Tweet, anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, trust
* **Type:** Multi-label classification
* **Purpose:** To detect emotions from tweets

**Model Training**

Each notebook fine-tunes a different model. To train a specific model, run:

python train\_model.py --model\_name albert-base-v2

python train\_model.py --model\_name distilbert-base-uncased

python train\_model.py --model\_name roberta-base

**Results**

Each model's performance is evaluated using accuracy, F1-score, and other relevant metrics. The results are logged using **Weights & Biases (WandB)**.

**Model Comparison**

To evaluate the effectiveness of the three transformer models, we compared their performance on tweet sentiment classification. Below is a brief summary of their strengths:

* **ALBERT (A Lite BERT)**: Optimized for memory efficiency and speed but slightly lower accuracy compared to the others.
* **DistilBERT**: A compressed version of BERT, offering a balance between speed and performance.
* **RoBERTa**: Achieved the highest accuracy, making it the best-performing model for this task.

Based on evaluation metrics, **RoBERTa outperformed ALBERT and DistilBERT**, making it the most suitable choice for tweet sentiment classification.

**Conclusion**

After evaluating the models on tweet sentiment classification, **RoBERTa** emerged as the best-performing model. It achieved superior accuracy in capturing emotions like trust, optimism, and overall sentiment classification, making it the most effective model for this task. After evaluating the models on tweet sentiment classification, **RoBERTa** emerged as the best-performing model. It achieved superior accuracy in capturing emotions like trust, optimism, and overall sentiment classification, making it the most effective model for this task.

**Contributing**

Feel free to fork this repository and contribute improvements or additional experiments.

**License**

This project is open-source and available under the MIT License.