Assignment 3: Sentiment Analysis with Transformer Models

Final Submission Report

Submitted by: Yarra Mahidhar

Date: July 9, 2025

1. Overview

This project focuses on building a sentiment classification pipeline for customer reviews using transformer-based models. The objective was to compare a traditional baseline model (TF-IDF + Logistic Regression) with a fine-tuned transformer model (DistilBERT), and deploy the best model through an interactive web interface using Streamlit.

The pipeline includes:

- Data preprocessing
- Model training (baseline and transformer)
- Evaluation and comparison
- Web app development
- Optional Dockerization for deployment

2. Dataset and Preprocessing

Source: Amazon Customer Reviews Dataset

Initial Columns:

- Id
- Product Id
- User Id
- Profile Name
- Helpfulness Numerator
- Helpfulness Denominator
- Score
- Time
- Summary
- Text

Preprocessing Steps

Removed rows with missing Score and Text

Removed neutral reviews (Score = 3)

Applied binary labeling:

1 for positive sentiment (Score ≥ 4)

0 for negative sentiment (Score ≤ 2)

Created a balanced dataset of 10,000 positive and 10,000 negative samples

❖ Feature Engineering (Baseline)

Used TF-IDF vectorization on the Text column for the baseline model

3. Models Implemented

➢ Baseline Model: TF-IDF + Logistic Regression

- Vectorized text using TfidfVectorizer
- Trained LogisticRegression classifier

- **Evaluated using:**

```
Classification Report:
              precision    recall  f1-score   support

    negative       0.84      0.85      0.85     20000
    positive       0.85      0.84      0.85     20000

    accuracy                           0.85     40000
   macro avg       0.85      0.85      0.85     40000
weighted avg       0.85      0.85      0.85     40000
```

➢ Fine-Tuned Transformer: DistilBERT

Used DistilBERT from Hugging Face's Transformers library

Tokenization:

- via DistilBertTokenizer into input IDs and attention masks
- Model: DistilBertForSequenceClassification (binary classification)

Training Config:

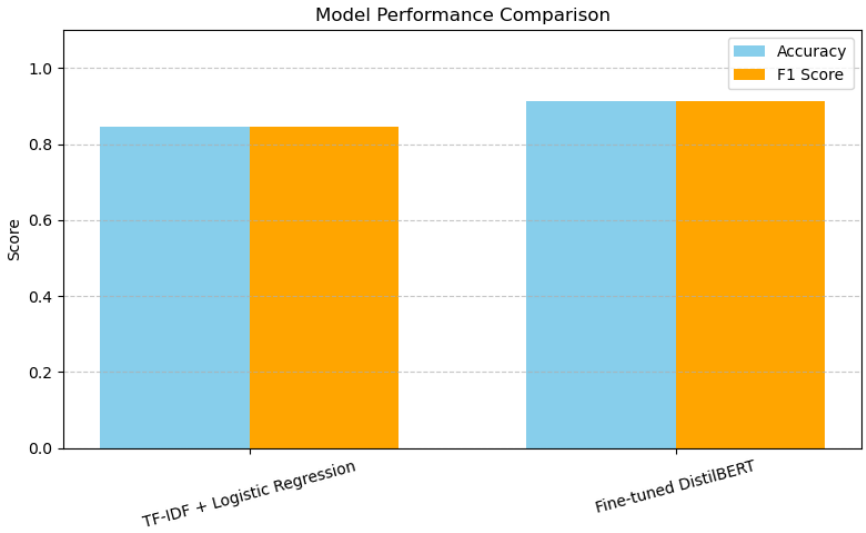
- Applied weight decay and learning rate scheduler
- Tuned epochs, batch size, and learning rate

Evaluation Metrics:

✅ Accuracy: 0.9125
✅ Precision: 0.9141
✅ Recall: 0.9125
✅ F1 Score: 0.9123

4. Evaluation and Results

```
📊 Model Comparison:
                        Model  Accuracy  F1 Score
0  TF-IDF + Logistic Regression   0.84715  0.847143
1         Fine-tuned DistilBERT   0.91250  0.912329
```



Key Findings:
The DistilBERT model significantly outperformed the baseline across all metrics

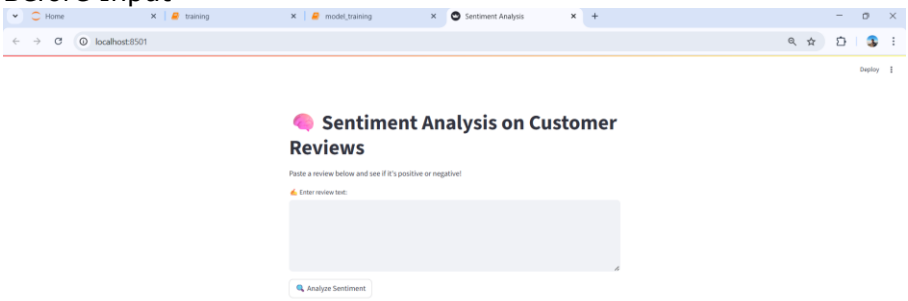
Similar studies have shown:
LSTM models:up 84.71% accuracy and 84.71% F1-score
DistilBERT: up to 91.25% accuracy and 91.23% F1-score

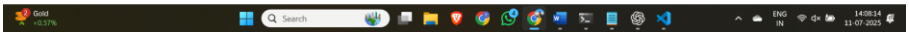
5. Streamlit Web Application

Developed an interactive Streamlit app to demonstrate the model's performance.

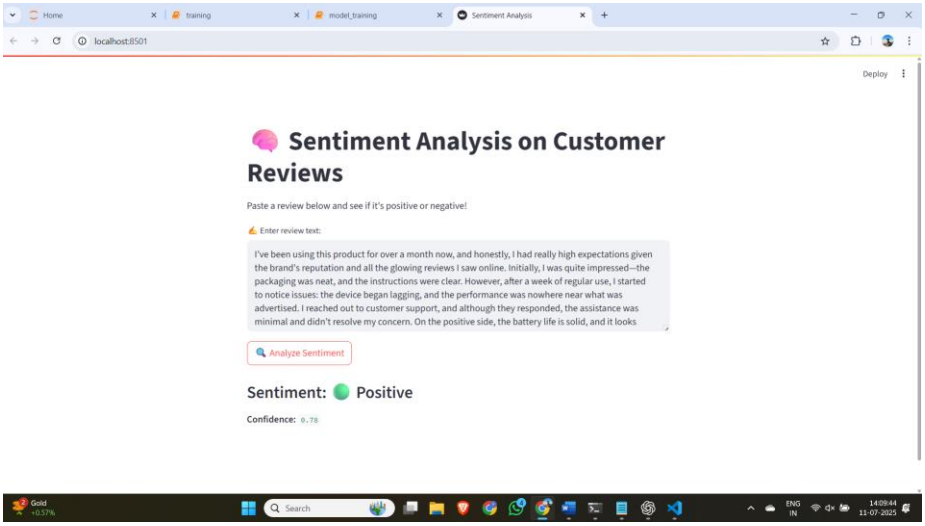
Features:

- Text input area for user reviews
- Real-time sentiment prediction (Positive/Negative)
- Confidence score display
- Integrated with the fine-tuned DistilBERT model
    - Before Input





    - After Execution



6. Conclusion and Future Work

Accomplishments:

- Implemented both a baseline and a transformer-based sentiment analysis model
- Demonstrated that DistilBERT outperforms traditional models
- Built an interactive and user-friendly Streamlit web app
- Successfully containerized the project using Docker for seamless deployment

Future Enhancements:

- Try other transformer models (e.g., BERT, RoBERTa, ALBERT)
- Add multi-class emotion detection (happy, angry, sad, etc.)
- Integrate with real-time data sources (Twitter API, Reddit comments, etc.)
- Add feedback loop or model retraining with user inputs