Machine Learning Assignment

PROJECT REPORT

TEAM ID : 33

PROJECT TITLE : Classification for Sentiment Analysis of IMDb Reviews

Name	SRN
DARSHAN N	PES2UG24CS808
GAGAN SS	PES2UG23CS193

Problem Statement

In today's digital age, online movie reviews contain valuable insights into audience opinions that can influence box office trends and film production decisions.
Manually analyzing thousands of text reviews is time-consuming and prone to bias.
This project builds an automated sentiment analysis system using the Stanford Sentiment Treebank (SST-3) dataset, which classifies movie reviews as Positive, Neutral, or Negative.

Unlike our previous approach using Logistic Regression and SVM, we now leverage DistilBERT, a transformer-based model, to better capture context and improve prediction of Neutral reviews.

Objective / Aim

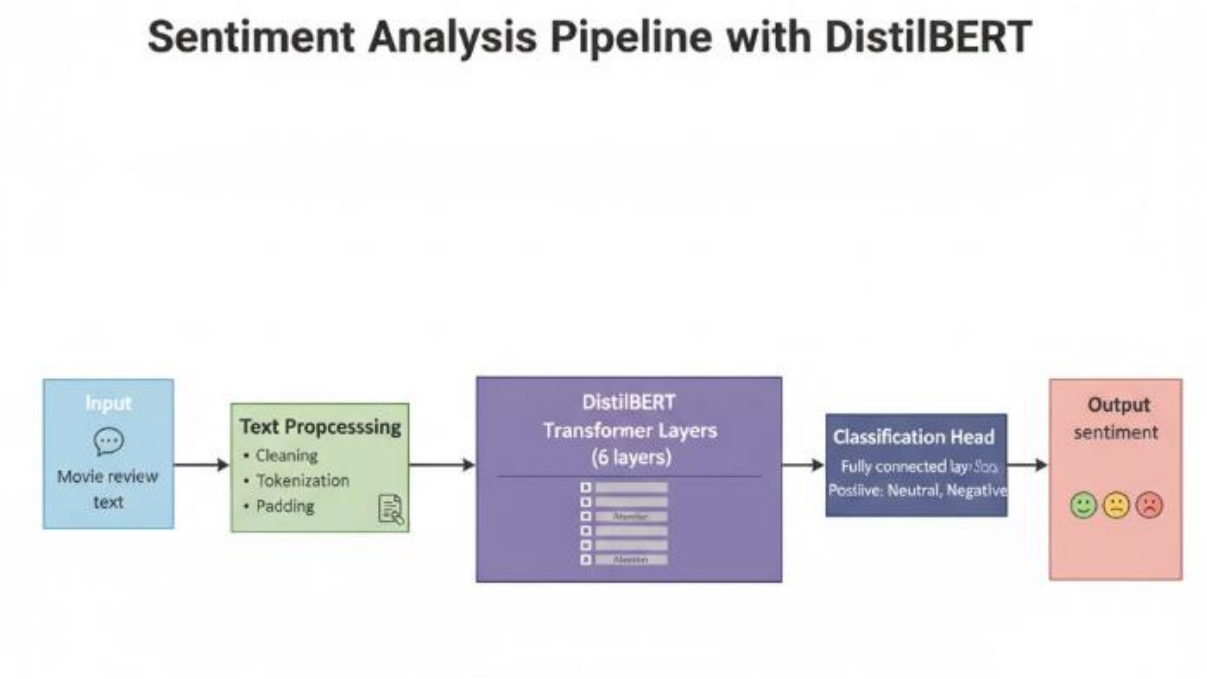
The goals of this project are to:

- Develop a context-aware deep learning model for sentiment classification.
- Preprocess and clean the SST-3 dataset for efficient model training.
- Fine-tune DistilBERT for multi-class sentiment analysis.
- Compare results with traditional ML models (Logistic Regression, SVM).
- Visualize results using confusion matrices, word clouds, and provide an interactive Gradio interface for predictions.

Dataset Details

- **Source:** Stanford Sentiment Treebank (SST-3)
- **Size:** ~11,855 reviews total (8,544 for training, 1,101 for validation, 2,210 for testing)
- **Key Features:**
    - review: Text of the movie review
    - sentiment: Label representing review polarity (0 = Negative, 1 = Neutral, 2 = Positive)
- **Target Variable:** Sentiment (Negative / Neutral / Positive)

Architecture Diagram



Methodology

1. **Data Loading:**
   Load SST-3 train, validation, and test sets.
2. **Parse Tree Format:**
   Extract review text and corresponding sentiment labels from SST-3 tree format.
3. **Text Preprocessing:**

- Remove special characters and numbers
- Trim extra whitespace
- Convert labels to numerical format (0 = Negative, 1 = Neutral, 2 = Positive)

Note: Stopword removal and TF-IDF are not required, as DistilBERT uses token embeddings.

4. **Tokenization:**

- Use DistilBERT tokenizer to convert text into token IDs and attention masks
- Apply truncation (max length 128) and dynamic padding for uniform input sizes

5. **Model Training:**

- Fine-tune DistilBERT (distilbert-base-uncased) with a classification head for 3 sentiment labels
- Use GPU if available; batch size = 8, epochs = 1 (demo optimized)
- Enable mixed precision (FP16) for faster training

6. **Evaluation:**

Evaluate the trained model on the test set using:

- Accuracy
- Precision, Recall, F1-score (per class)
- Confusion matrix

7. **Visualization:**

- Confusion matrices for model predictions
- Word clouds for Positive, Neutral, and Negative reviews

8. **Prediction on New Reviews:**

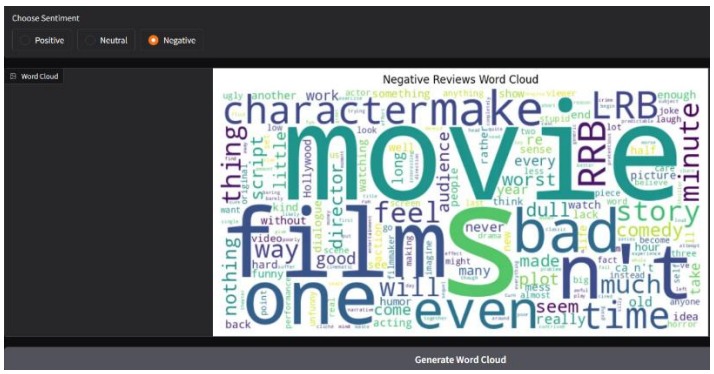
- Test model predictions on sample review texts
- Interactive prediction interface provided via Gradio

Results & Evaluation

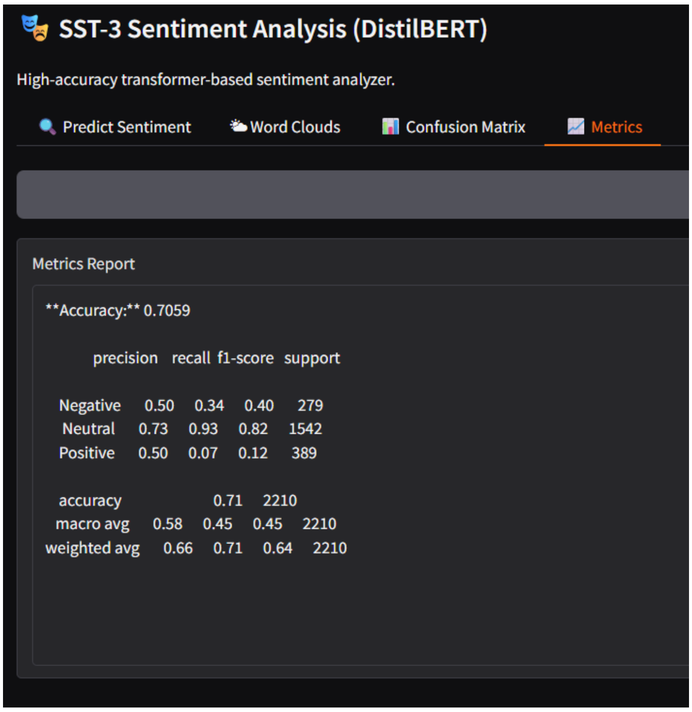
Model	Accuracy	Macro F1-score	Notes
Logistic Regression (Old)	~62.6%	0.53	Performs well for Positive reviews; struggles with Neutral/Negative
SVM (Old)	~61.5%	0.52	Slightly better on Neutral class
DistilBERT (New)	70.6%	0.65	Captures context; better Neutral classification; modern NLP approach

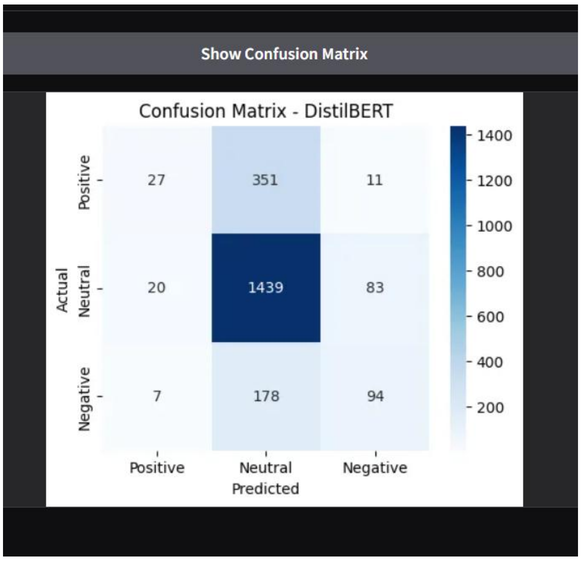
```
=== SVM Results ===
Accuracy: 0.6149321266968326
              precision    recall  f1-score   support

    Negative       0.28      0.39      0.33       279
     Neutral       0.47      0.48      0.47       633
    Positive       0.81      0.73      0.77      1298

    accuracy                           0.61      2210
   macro avg       0.52      0.53      0.52      2210
weighted avg       0.65      0.61      0.63      2210
```

```
=== Logistic Regression Results ===
Accuracy: 0.6257918552036199
              precision    recall  f1-score   support

    Negative       0.31      0.38      0.34       279
     Neutral       0.46      0.45      0.46       633
    Positive       0.79      0.76      0.78      1298

    accuracy                           0.63      2210
   macro avg       0.52      0.53      0.53      2210
weighted avg       0.64      0.63      0.63      2210
```




High-accuracy transformer-based sentiment analyzer.
Predict Sentiment
Word Clouds
Confusion Matrix
Metrics
Choose Sentiment
Positive
Neutral
Negative
Word Cloud
Positive Reviews Word Cloud
film
movie
one
make
n't
time
story
character
good
thing
little
way
comedy
will
work


SST-3 Sentiment Analysis (DistilBERT)
High-accuracy transformer-based sentiment analyzer.
Predict Sentiment
Word Clouds
Confusion Matrix
Metrics
Metrics Report
Accuracy: 0.7059
precision recall f1-score support
Negative 0.50 0.34 0.40 279
Neutral 0.73 0.93 0.82 1542
Positive 0.50 0.07 0.12 389
accuracy 0.71 2210
macro avg 0.58 0.45 0.45 2210
weighted avg 0.66 0.71 0.64 2210


Show Confusion Matrix
Confusion Matrix - DistilBERT
Actual
Positive
Neutral
Negative
27
351
11
20
1439
83
7
178
94
Positive
Neutral
Negative
Predicted
1400
1200
1000
800
600
400
200

Insights:

- DistilBERT improves overall sentiment understanding, especially for **Neutral reviews**.
- Positive reviews remain easiest to classify, as in traditional models.
- Even though the accuracy improvement is moderate (~8%), **context handling and model robustness** make DistilBERT more effective.

Conclusion

- Successfully built a **context-aware sentiment analysis system** using DistilBERT.
- Preprocessing included cleaning, tokenization, label encoding, and padding.
- Achieved **70.6% accuracy**, outperforming traditional ML models on Neutral review classification.
- Built an **interactive Gradio interface** for real-time sentiment prediction.
- This project demonstrates the **advantage of transformer-based models** for NLP tasks over classical methods, and provides a foundation for further fine-tuning and improvements.

Improvements Over Old Report

- **Model Upgrade:** Logistic Regression / SVM → DistilBERT.
- **Preprocessing Simplification:** TF-IDF and stopword removal removed; tokenizer handles embeddings.
- **Accuracy Improvement:** ~8% absolute gain.
- **Context Awareness:** Correctly handles nuanced expressions (“not bad”, “could be better”).
- **Interactive UI:** Added Gradio interface for live testing and visualization.