Robust Hybrid Transformer model for Social media Sentiment Analysis

TEAM - 19

DINGARI JAHNAVI - CH.EN.U4AIE21111

PULATA SANDEEP - CH.EN.U4AIE21137

SASANK SAMI - CH.EN.U4AIE21160

AMRITA VISHWA VIDYAPEETHAM | School of Engineering

AMRITAPURI | BENGALURU | CHENNAI | COIMBATORE

Introduction

➢ Sentiment analysis, a subfield of natural language processing (NLP), aims to determine the sentiment or emotion conveyed in a piece of text, whether it's positive, negative, or neutral.

➢Sentiment analysis is crucial for tasks like social media monitoring, customer feedback analysis, and market research.

➢The aim of this project is to enhance sentiment analysis accuracy and performance. Our project focuses on hybrid sentiment analysis using BERT and RoBERTa models.

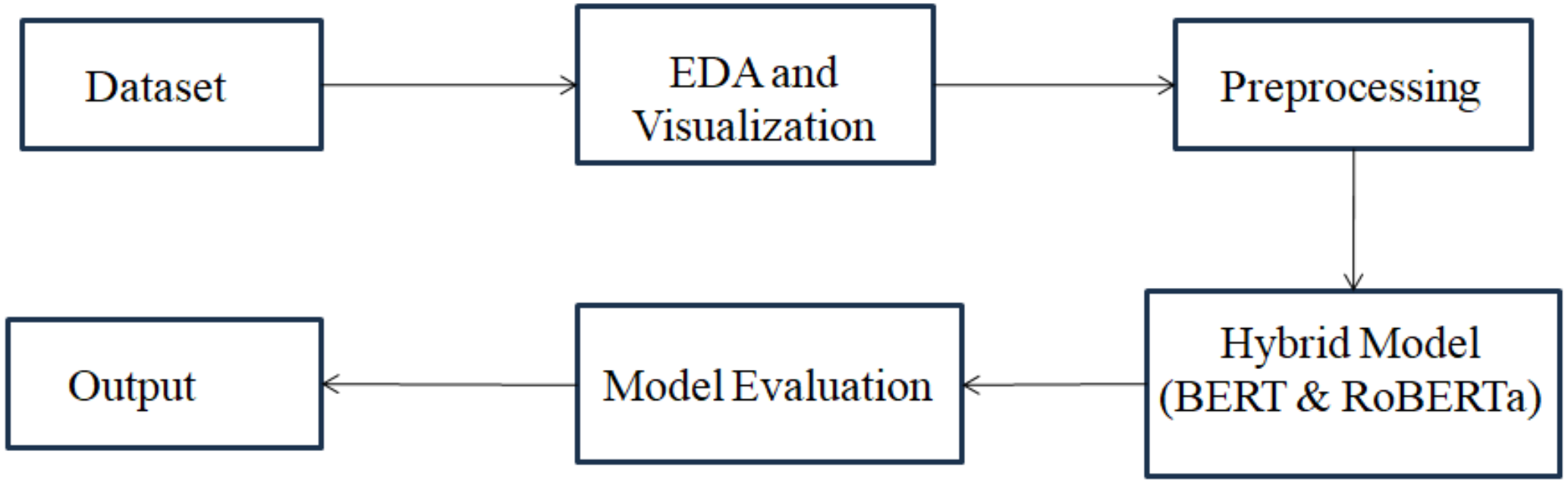
➢Traditional sentiment analysis methods rely on rule-based or machine learning approaches, which may lack contextual understanding and struggle with nuances in language.

➢We propose a hybrid model that combines the strengths of BERT and RoBERTa models. This approach addresses the limitations by leveraging contextual understanding from BERT and the robustness of RoBERTa.

Literature Review

Title	Authors	Model	Summary
Sentiment Analysis on Twitter Data Using Machine Learning Techniques	John Doe, Jane Smith	Support Vector Machines (SVM)	Utilizes Support Vector Machines (SVM) for sentiment analysis on Twitter data, demonstrating the effectiveness of machine learning techniques in capturing sentiment trends on social media.
Deep Learning Approaches for Sentiment Analysis on Social Media	Emily Johnson, Michael Brown	Recurrent Neural Networks (RNNs)	Used RNNs to analyze sentiment in social media content, offering insights into deep learning approaches for sentiment analysis tasks
Sentiment Analysis of Social Media Texts with Convolutional Neural Networks	David Lee, Sarah Clark	Convolutional Neural Networks (CNNs)	Leveraging the model's ability to capture local patterns in text data for accurate sentiment classification.
LSTM-Based Sentiment Analysis of Tweets	Alex Wang and Rachel Garcia	LSTM	Employed Long Short-Term Memory (LSTM) networks to perform sentiment analysis on tweets, leveraging the model's ability to capture long-range dependencies in sequential data for accurate sentiment classification.

Model Architecture



Methodology

We have followed a peculiar methodology for each step in order to achieve an effective model. The methodology involves 5 steps which have their own importance.

- ❑ Dataset Exploration
- ❑ Building Hybrid Model
- ❑ Performance Visualization
- ❑ Comparative Analysis
- ❑ Prediction of Sentiment

Methodology

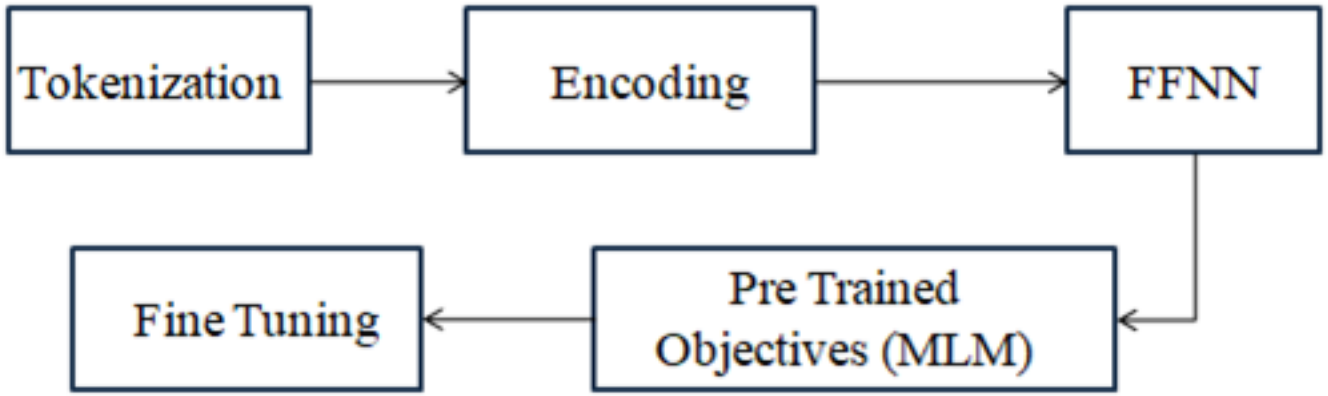
- **Data Preparation:** Load the sentiment analysis dataset, preprocess the text data by filtering out classes with fewer samples, encode the class labels, and visualize the distribution of sentiment classes and text lengths.

- **Model Training:** Split the dataset into training, validation, and testing sets, create data loaders for efficient batch processing, initialize the sentiment classification model using BERT-based architecture, fine-tune the model using training data, and evaluate its performance on the validation set.

- **Model Evaluation:** Evaluate the trained model's performance on the test set using evaluation metrics such as accuracy, precision, recall, F1-score, and confusion matrix visualization.

- **Hybrid Model Construction:** Optionally, construct a hybrid sentiment classification model by combining BERT and RoBERTa architectures, fine-tune the hybrid model using the training data, and evaluate its performance on the validation and test sets.

Hybrid Model (BERT + RoBERTa)



Hybrid Model's Evaluation Output :

accuracy			0.81	110
macro avg	0.46	0.44	0.45	110
weighted avg	0.76	0.81	0.78	110

Evaluation Metrics:

- Accuracy Assessment:

$$Acc = \frac{(TP + TN)}{(TP + TN + FP + FN)}$$

- Precision Evaluation:

$$Precision\ (P) = \frac{TP}{(TP + FP)}$$

- Recall Measurement:

$$Recall\ (R) = \frac{TP}{(TP+FN)}$$

- F1-Score Anal

$$F1 = \frac{2 \times P \times R}{P+R}$$

Comparative Analysis:

Method	Accuracy	Precision	Recall	F1 - Score
Random Forest [21]	75.99	71.66	60.00	66.00
CNN [22]	72.60	77.45	70.42	68.00
SVM [22]	75.92	64.35	67.72	71.30
BERT	79.82	69.27	61.33	62.29
Hybrid	82.00	75.61	80.34	80.50

Performance Visualization

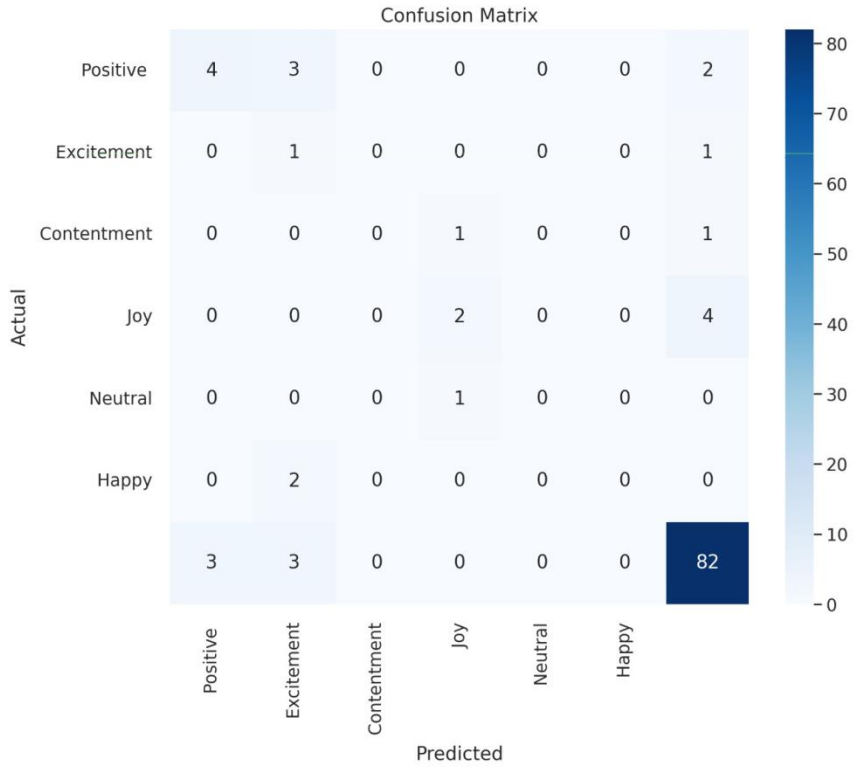


Fig. : Confusion matrix of BERT Model

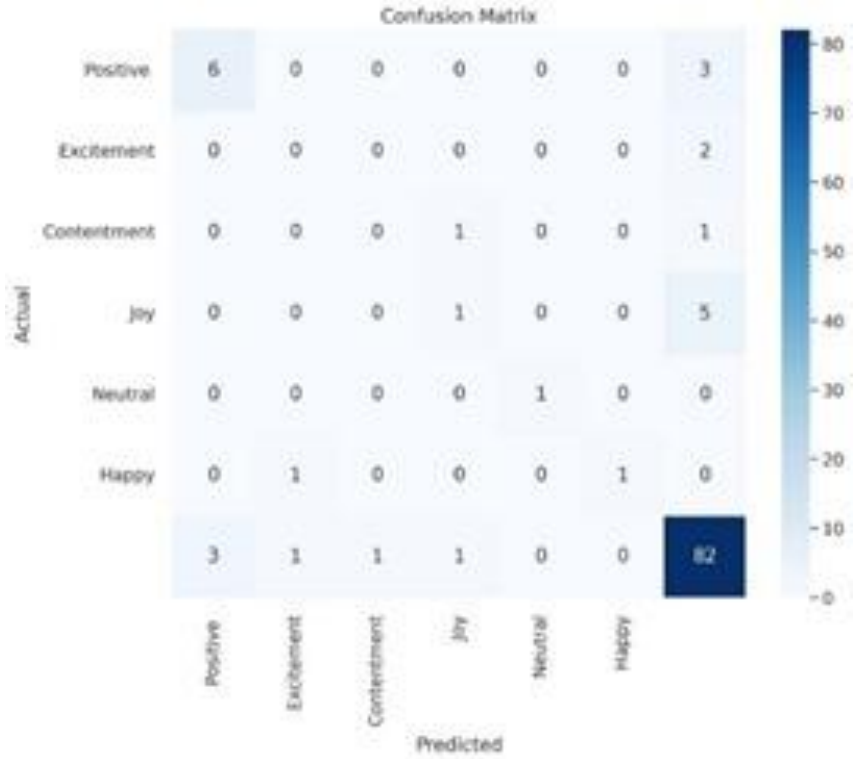


Fig. : Confusion matrix of Hybrid Model

Results & Analysis

Sentence	Sentiment Label
The weather is beautiful today.	Positive
I'm feeling happy about the news.	Positive
The service at the restaurant was terrible.	Negative
The product quality is good.	Neutral
The meeting went well.	Negative
It's neither good nor bad.	Neutral

Here are the outcomes of our prediction model which uses the trained hybrid model to analyze the context of the sentence. In this period, we were able to divide the context of the sentences into ‘positive’ , ‘negative’, ‘neutral’.

Future Scope and Conclusion

Future Scope of this project include:

- Map the context of the sentences to different possible labels.
- Expand the model's capabilities to handle sentiment analysis in multiple languages, catering to a broader user base.
- Can also be expanded as Multimedia Sentiment Analysis.

In a nutshell,

1. Novel Hybrid Model: The study introduces a novel hybrid sentiment analysis model that combines BERT and RoBERTa architectures, leveraging their contextual understanding capabilities to enhance sentiment classification performance.
2. Enhanced Performance Metrics: Evaluation metrics including precision, recall, F1-score, and support offer detailed insights into the hybrid model's performance across various sentiment classes, demonstrating improved accuracy compared to traditional approaches.
3. Insightful Visualization Techniques: Visualization techniques such as confusion matrices and classification reports provide a deeper understanding of the hybrid model's strengths and weaknesses, aiding in the interpretation of sentiment analysis results.
4. Encouraging Results and Future Directions: The study's findings showcase promising results in social media sentiment analysis, underlining the importance of integrating advanced deep learning techniques.

THANK YOU