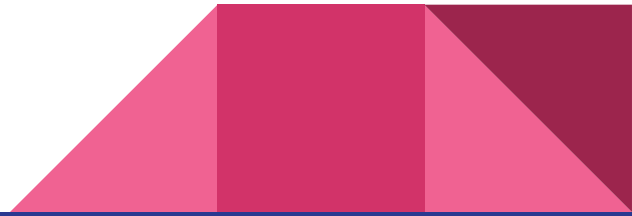
Sentiment Analysis Using RoBERTa Model

Team Members:
Naveen Preetham Kaveti
Nagabhushan Reddy Kadiri

Project Objectives

Statement of Project Objectives:

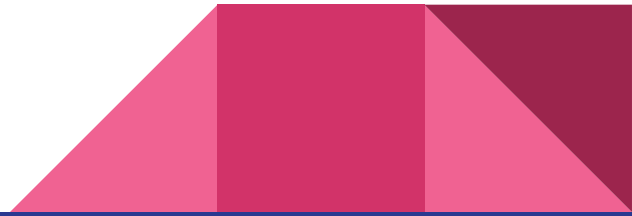
- Implement sentiment analysis on social media data(Twitter).
- Utilize advanced RoBERTa models for improved accuracy.
- Classify sentiments into categories like happy, angry, sad, etc.



Statement of Value

Importance of Sentiment Analysis:

- Understanding public sentiment aids businesses and policymakers.
- RoBERTa models enhance accuracy, leading to more informed decisions.
- Real-time analysis enables rapid response to changing sentiments.



Review of the State of the Art

Brief Overview of Sentiment Analysis:

Traditional Methods vs. Deep Learning Approaches:

"Traditional sentiment analysis methods often rely on predefined rules, while deep learning approaches leverage neural networks to grasp complex patterns in language, enabling more accurate and nuanced sentiment classification."

Advantages of RoBERTa over Traditional Models:

"RoBERTa, a robustly optimized transformer, outperforms traditional models by capturing intricate language structures, leading to superior sentiment analysis accuracy and adaptability to diverse datasets."

Approach

Data Collection and Preprocessing:

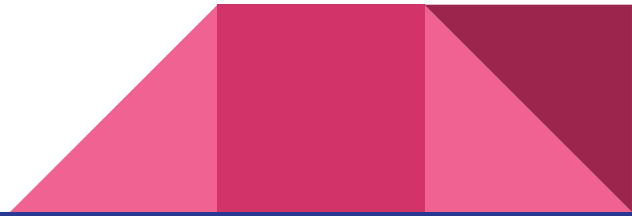
- Gather social media data from reliable sources.
- Preprocess data to remove noise and irrelevant information.

RoBERTa Model Implementation:

- Utilize pre-trained RoBERTa models for fine-tuning.
- Tokenize and encode data for model input.

Ensemble Learning (Optional):

- Utilize multiple RoBERTa models for ensemble predictions.
- Discuss the benefits of model ensembles.



Deliverables

- Dataset- SMILE Twitter Emotion Dataset, Trained RoBERTa model and code on GitHub for sentiment analysis
- Detailed report on model performance and insights.
- Documentation

Evaluation Methodology

Performance Metrics:

- Accuracy, F1 Score, and Confusion Matrix.
- Comparison of RoBERTa models with traditional methods.

Validation Process:

- Split dataset into training and validation sets.
- Evaluate models using unseen data to ensure generalization.

References

https://arxiv.org/abs/1810.04805

https://www.kaggle.com/code/chayan8/sentiment-analysis-using-bert-pytorch/notebook

https://www.kaggle.com/code/soham27/roberta-sentiment-analysis

Analyzing the Performance of Sentiment Analysis using BERT, DistilBERT, and RoBERTa(https://ieeexplore.ieee.org/document/10059542)


THANK YOU!