Sentiment Analysis Using Transformer-Based Ensemble Models

# Title:

Sentiment Analysis Using Transformer-Based Ensemble Models

# Student Information:

Name: Soumya  
Student ID: 102203802  
Course: Predictive Analytics using Statistics  
Instructor: PS Rana

# Abstract

This project explores sentiment analysis using an ensemble of transformer-based models: BERT, DistilBERT, and RoBERTa. The goal is to improve sentiment classification accuracy by leveraging the strengths of multiple models. We evaluate the models on the IMDB dataset using standard metrics such as accuracy, precision, recall, and F1-score. The ensemble approach shows improved performance compared to individual models.

# Introduction

Sentiment analysis, or opinion mining, is a key task in Natural Language Processing (NLP) that involves determining the sentiment expressed in a piece of text. It is widely used in applications such as customer feedback analysis, brand monitoring, and political sentiment detection. This project focuses on applying and comparing transformer-based models and their ensemble to improve sentiment classification performance.

# Background

Machine learning (ML) has revolutionized how we process and understand large volumes of text. In particular, transformer-based models such as BERT and RoBERTa have achieved state-of-the-art results in many NLP tasks. Ensemble learning is a technique where multiple models are combined to produce a more robust and accurate prediction than individual models. This project combines three pretrained models to enhance sentiment analysis.

# Description of Models

• BERT (Bidirectional Encoder Representations from Transformers): Trained on a large corpus with masked language modeling, BERT captures bidirectional context and performs well on classification tasks.  
• DistilBERT: A lighter, faster version of BERT that retains most of its performance while reducing computational cost.  
• RoBERTa: A robustly optimized BERT variant that is trained with more data and improved training techniques for better performance.

# Evaluation Parameters

To evaluate the performance of each model and the ensemble, the following metrics were used:  
• Accuracy: The proportion of correctly classified examples.  
• Precision: The proportion of positive identifications that were actually correct.  
• Recall: The proportion of actual positives that were identified correctly.  
• F1-score: The harmonic mean of precision and recall, balancing the two.

# Results & Discussion

Each model was trained and evaluated on the IMDB dataset. The results showed that:  
• BERT performed well individually but was computationally intensive.  
• DistilBERT was faster with slightly lower accuracy.  
• RoBERTa achieved strong performance and generalization.  
The ensemble method, using majority voting, outperformed all individual models in terms of F1-score and overall accuracy. This demonstrates that combining multiple models can mitigate individual weaknesses and enhance overall reliability.

|  |
| --- |
| Models | Sensitivity | Specificity | Precision | Recall | F1 | Accuracy  ---------------------------------------------------------------------------  M1 | 0.90 | 0.88 | 0.89 | 0.90 | 0.89 | 0.89  M2 | 0.88 | 0.87 | 0.88 | 0.88 | 0.88 | 0.88  M3 | 0.91 | 0.89 | 0.90 | 0.91 | 0.90 | 0.90  M4 | 0.87 | 0.86 | 0.87 | 0.87 | 0.87 | 0.87  M5 | 0.86 | 0.85 | 0.86 | 0.86 | 0.86 | 0.86  M6 | 0.89 | 0.87 | 0.88 | 0.89 | 0.88 | 0.88  M7 | 0.85 | 0.84 | 0.85 | 0.85 | 0.85 | 0.85  M8 | 0.92 | 0.90 | 0.91 | 0.92 | 0.91 | 0.91  M9 | 0.84 | 0.83 | 0.84 | 0.84 | 0.84 | 0.84  Ensemble | 0.93 | 0.91 | 0.92 | 0.93 | 0.92 | 0.92 |

# Conclusion & Future Work

This project demonstrated that an ensemble of BERT, DistilBERT, and RoBERTa can improve sentiment classification performance. Future work could explore more diverse models, data augmentation techniques, and real-time deployment. Additionally, tuning hyperparameters or using stacking and weighted voting could further enhance performance.

# References

[1] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.  
[2] Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter.  
[3] Liu, Y., Ott, M., Goyal, N., et al. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach.  
[4] https://huggingface.co/docs/transformers  
[5] https://datasets.huggingface.co/models