

# Sentiment Analysis Project Report

## Introduction

Sentiment Analysis is a Natural Language Processing (NLP) task that aims to determine the sentiment expressed in a piece of text, such as reviews, social media posts, or any other form of textual data. In this project, we developed a Sentiment Analysis application using the BERT (Bidirectional Encoder Representations from Transformers) model, a state-of-the-art transformer-based language model developed by Google.

## Abstract

This project focuses on building a Sentiment Analysis model using the BERT model pre-trained on the English language. The application can analyze both individual text reviews and entire datasets, providing sentiment classification (positive or negative). The model was fine-tuned on a large dataset, and an optimized training approach was used to achieve efficient performance on limited hardware.

## Tools Used

- Programming Language: Python
- Libraries: PyTorch, Transformers (Hugging Face), Pandas, NumPy, Streamlit
- Model: BERT (Bidirectional Encoder Representations from Transformers)
- Development Environment: Jupyter Notebook, Streamlit
- Visualization: Matplotlib, Seaborn
- Hardware: NVIDIA GPU (if available), CPU

## Steps Involved in Building the Project

1. **Data Collection:** The IMDB Dataset containing text reviews and sentiment labels was used.
2. **Data Preprocessing:** The text data was pre-processed, including text cleaning, tokenization using BERT tokenizer, and conversion of labels to numeric values (positive = 1, negative = 2).
3. **Sentiment Analysis:** VADER (Valence Aware Dictionary for Sentiment Reasoning) was used to perform sentiment analysis on the text reviews, providing a sentiment score for each review.
4. **Model Selection:** The BERT (bert-base-uncased) model from Hugging Face's Transformers library was used.
5. **Fine-Tuning:** The model was fine-tuned using the IMDB dataset with an optimized training loop for efficiency, using a batch size of 4, learning rate of  $2e-5$ , and AdamW optimizer.
6. **Application Development:** A user-friendly application was developed using Streamlit, allowing users to input text or upload datasets for sentiment analysis.

7. **Evaluation:** The model's performance was evaluated using training and validation loss curves and sentiment distribution visuals.

### **Predictive Model Summary**

- Model: BERT (Bidirectional Encoder Representations from Transformers)
- Fine-Tuned on: IMDB Dataset
- Training Time: Approximately 30-45 minutes (varies based on hardware)
- Optimizer: AdamW with learning rate of  $2e-5$
- Evaluation Metric: Loss and Accuracy
- Application: Streamlit-based app for single review or dataset analysis

### **Conclusion**

The Sentiment Analysis project successfully demonstrated the use of a transformer-based model (BERT) for sentiment classification. The application can efficiently analyze both single reviews and datasets. The use of an optimized training approach ensured faster training and reduced memory usage, making the model suitable for use even on limited hardware. The model achieved reliable performance in sentiment analysis, but further improvements can be made by exploring larger datasets, using more advanced transformer models, or optimizing the app for deployment.

**By-**

**Prerana V**