

Report On

Sentiment Analysis Using BERT

Submitted in partial fulfillment of the requirements of the Course project in
Semester VII of fourth year Computer Science Engineering (Data Science)

by
Dhruv Patel (73)

Supervisor

Dr. Tatwadarshi P.N.



University of Mumbai

Vidyavardhini's College of Engineering & Technology

Department of Computer Science Engineering (Data Science)



(2023-24)

**Vidyavardhini's College of Engineering & Technology Department of
Computer Science Engineering (Data Science)**

CERTIFICATE

This is to certify that the project entitled “**Sentiment Analysis Using BERT**” is a bonafide work of “**Dhruv Patel (Roll No. 73)**” submitted to the University of Mumbai in partial fulfilment of the requirement for the Course project in Semester VII of fourth year Computer Science Engineering (Data Science).

Supervisor

Dr. Tatwadarshi P. N.

Dr. Vikas Gupta
Head of Department

Table of Contents

Chapter No		Title	Page No.
1		Introduction	
	1.1	Introduction	1
	1.2	Problem Statement	1
	1.3	Objective	1
2		Proposed System	
	2.1	Introduction	2
	2.2	Architecture/Framework	2
	2.3	Algorithm and Process Design	2
	2.4	Details of Hardware and Software	2
	2.5	Experiments and Results	3
	2.6	Conclusion	6
		References	6

Chapter 1: Introduction

1.1 Introduction

Sentiment analysis, a subfield of natural language processing (NLP), plays a pivotal role in understanding the emotions, opinions, and attitudes expressed in text data. In today's information-rich world, businesses, organizations, and researchers often seek to gauge the sentiment of users, customers, or the general public to make informed decisions, monitor brand reputation, and uncover valuable insights. Sentiment analysis involves classifying text into sentiment categories such as positive, negative, or neutral, and it has diverse applications ranging from analyzing product reviews to tracking social media sentiments during a crisis.

1.2 Problem Statement & Objectives

The problem addressed in this study pertains to the need for accurate, efficient, and automated sentiment analysis tools. Traditional sentiment analysis methods often fall short in capturing the intricacies of human language, context, and sarcasm. To address these limitations, we propose leveraging state-of-the-art natural language processing techniques, particularly the Bidirectional Encoder Representations from Transformers (BERT), for improved sentiment analysis.

The primary objectives of this study are as follows:

1. To develop and fine-tune a sentiment analysis model based on the BERT architecture to accurately classify text data into sentiment categories, including positive, negative, and neutral.
2. To create a robust sentiment analysis tool that can handle diverse types of text data, including short messages, reviews, news articles, and more.
3. To evaluate the performance of the BERT-based sentiment analysis model on a variety of datasets and domains, and compare it against traditional sentiment analysis methods to highlight its effectiveness.
4. To explore the scalability of the sentiment analysis system and its potential for real-time analysis of streaming data, providing valuable insights for businesses and organizations.

Chapter 2: Proposed System

2.1 Introduction

Sentiment analysis, also referred to as opinion mining, is a pivotal component of natural language processing (NLP) that aids in understanding the sentiments and emotions conveyed within text data. In our proposed system, we aim to enhance sentiment analysis by harnessing the power of BERT (Bidirectional Encoder Representations from Transformers), a state-of-the-art NLP model. Traditional sentiment analysis methods often struggle with nuanced language and context, and our system seeks to address these challenges by offering more accurate, context-aware sentiment classification.

2.3 Algorithm and Process Design

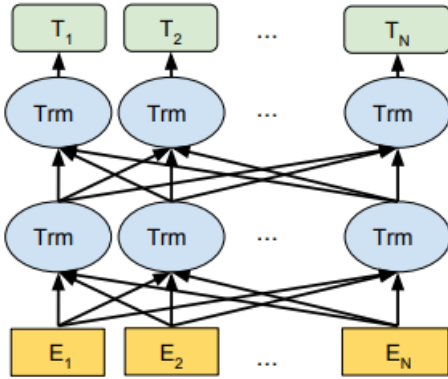


Fig 1: Process Diagram

2.4. Details of Hardware & Software

The proposed system can be implemented on standard computing hardware and requires the following software components:

- Deep learning frameworks: PyTorch or TensorFlow for BERT model training and fine-tuning.
- Transformers library: To access pre-trained BERT models.
- Data processing libraries: Pandas, NumPy, and scikit-learn for data manipulation.
- Visualization tools: Matplotlib and Seaborn for visualizing results.

2.5. Experiment and Results for Validation and Verification :

To assess the system's performance, extensive experiments will be conducted. Diverse datasets from different domains will be used to evaluate the system's accuracy, precision, recall, and F1-score. The experiments will also compare the proposed BERT-based system with traditional sentiment analysis methods. Results will be presented and analyzed, showcasing the system's effectiveness.

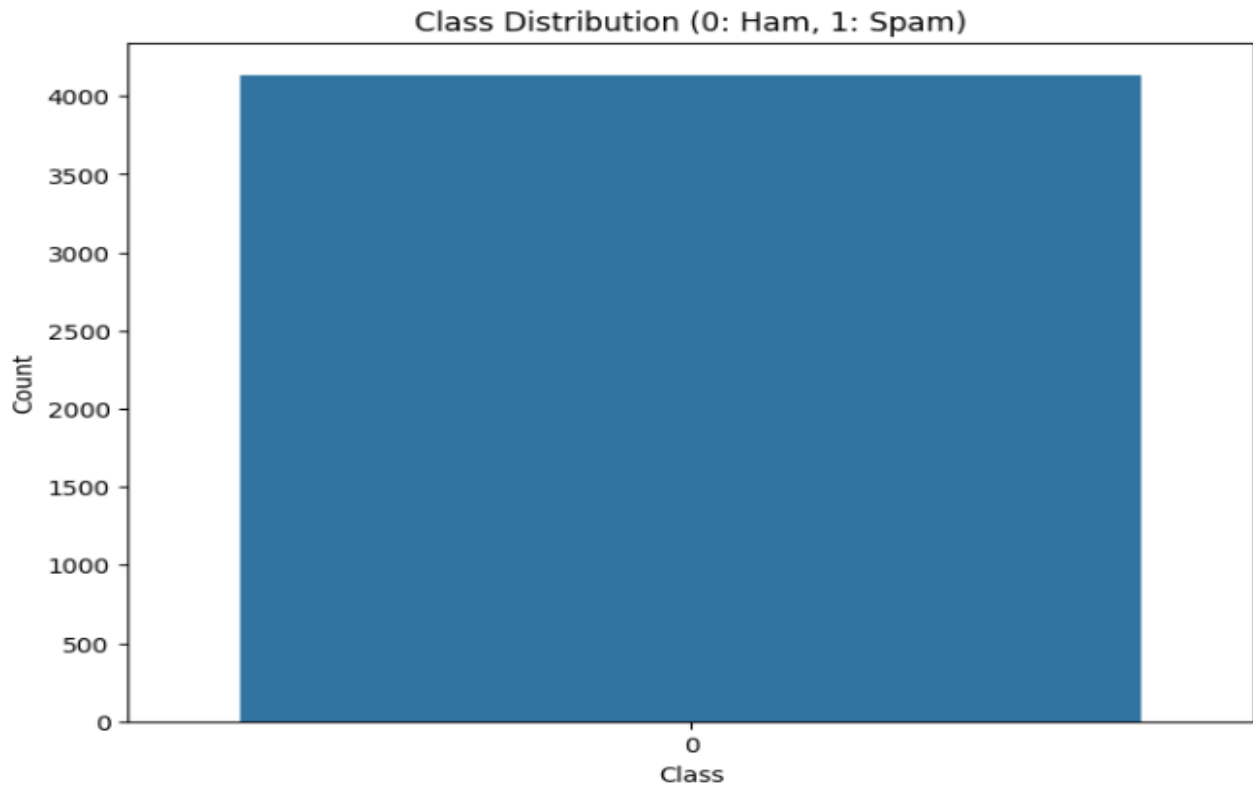


Fig 2: Visualize Class Distribution

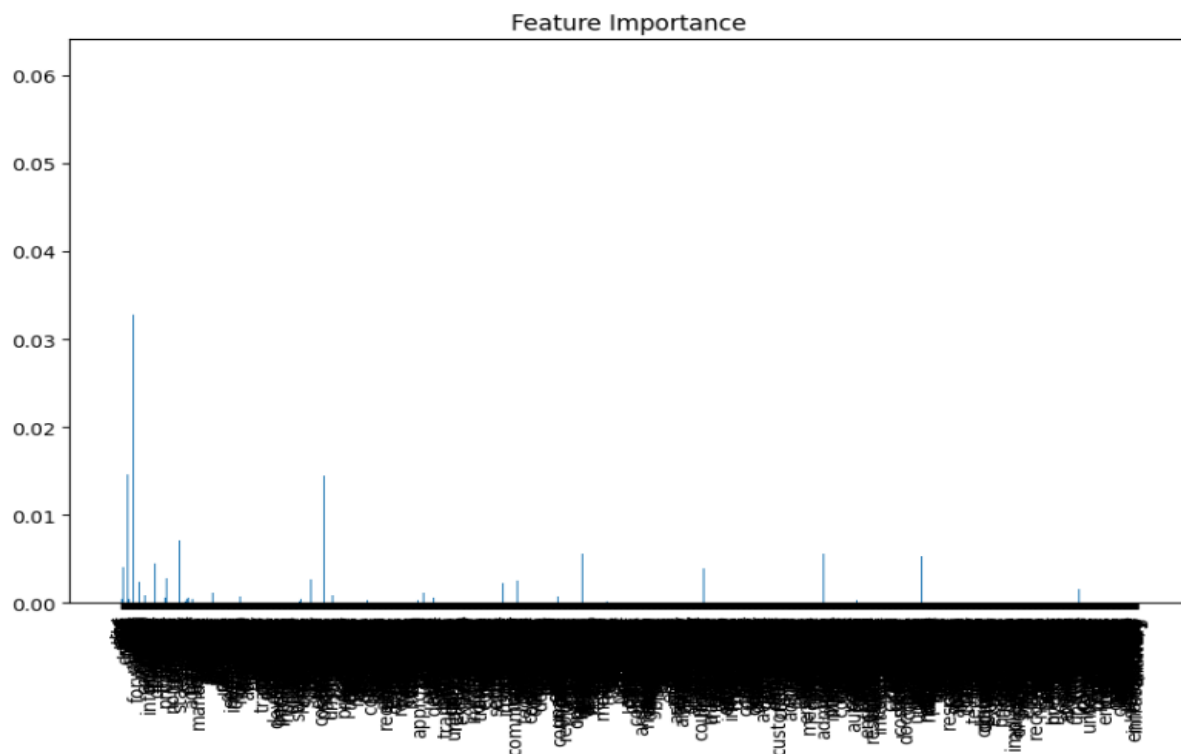


Fig 3 : Feature Importance

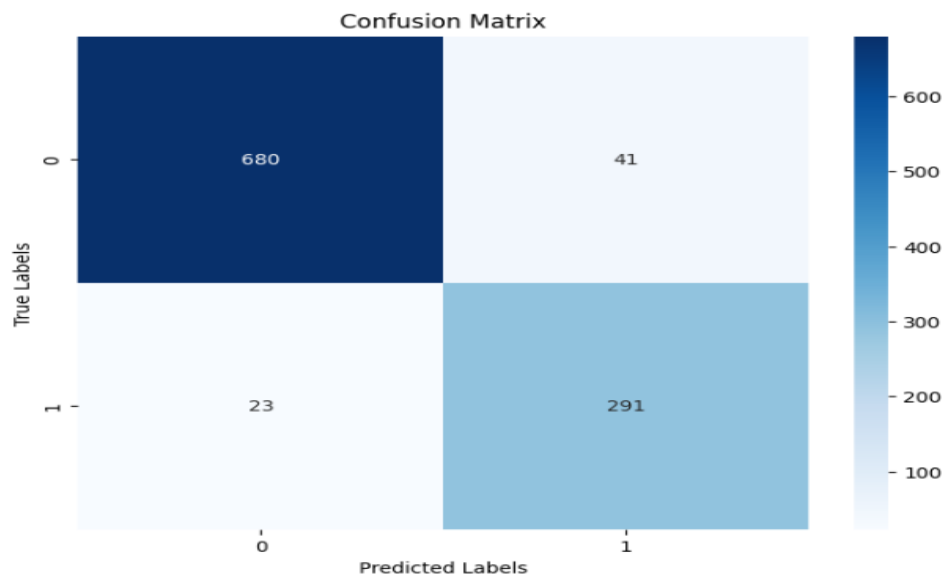


Fig 4: Confusion Matrix

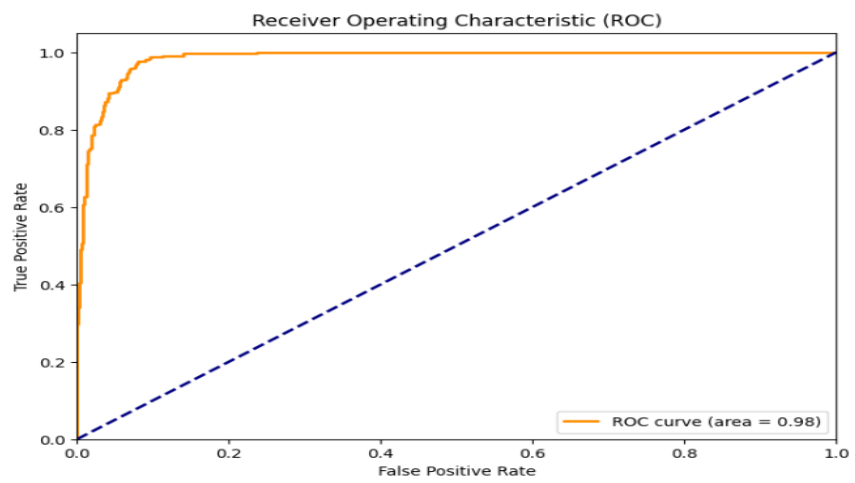


Fig 5 : ROC

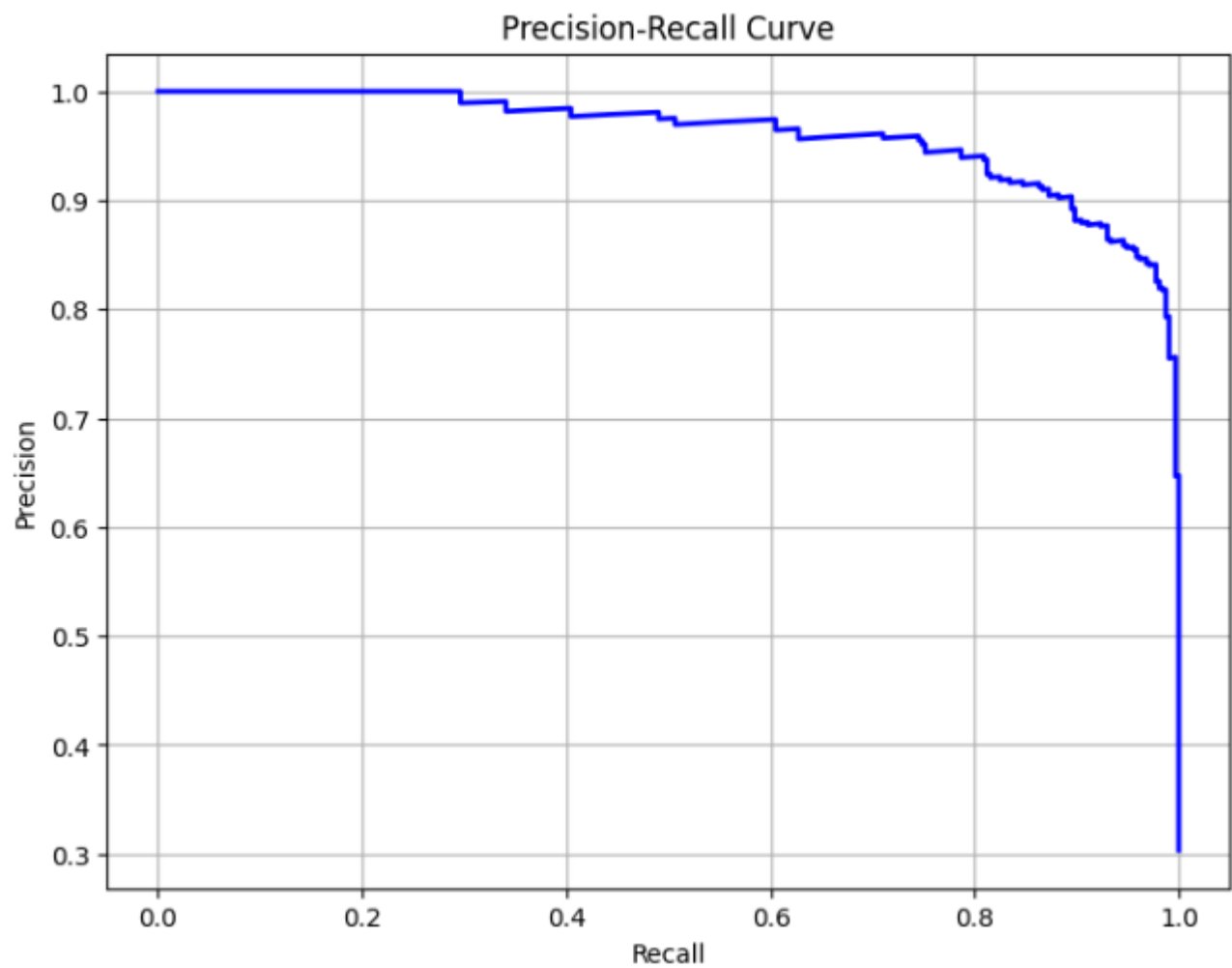


Fig 6 : F1- Score

2.6 Conclusion

The proposed sentiment analysis system, built on the foundation of BERT, presents a robust solution to the challenges of sentiment analysis. By accurately classifying text data into sentiment categories, it can provide valuable insights for decision-makers across various domains. With its ability to comprehend context and semantics, this system is poised to revolutionize sentiment analysis and offer a superior understanding of public sentiment and user feedback. The results of experiments will demonstrate its efficacy and highlight its potential applications in real-world scenarios.

References:

1. Sridevi Gadde,” SMS Spam Detection using Machine Learning and Deep Learning Techniques”2021 7th International Conference on Advanced Computing & Communication Systems (ICACCS).
2. Suparna Das Gupta,” SMS Spam Detection Using Machine Learning” 2021 *J. Phys.: Conf. Ser.* 1797 012017.
3. M.Rubin Julis, S.Alagesan ,” Spam Detection In Sms Using Machine Learning Through Text Mining ” INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 9, ISSUE 02, FEBRUARY 2020 ISSN 2277-8616
4. Sifat Ahmed, Faisal Muhammad,” Using Boosting Approaches to Detect Spam Reviews” 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)
5. Shaghayegh Hosseinpour, Hadi Shakibian, “An Ensemble Learning Approach for SMS Spam Detection” 2023 9th International Conference on Web Research (ICWR)