BERT

BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS

SUPERVISOR DETAILS:
Dr.B.Abirami, AP/CSE
Dr.C.Shanmuganathan, AP/CSE

PROJECT MEMBERS:
Ruchitraa Rajagopal(396)
Aaradhya Joshi(398)
Harshitha Pulluru(421)

OBJECTIVE

The objective for BERT (Bidirectional Encoder Representations from Transformers) is to pre-train a deep neural network model on a large corpus of text data in an unsupervised manner and then fine-tune it for specific downstream natural language processing (NLP) tasks.

Unlike previous models that relied on unidirectional context, BERT aims to understand both the left and right contexts of a word simultaneously.

SCOPE

Versatility: BERT can be used for a wide range of NLP tasks, making it suitable for various language-related challenges.

Multilingual Power: BERT's ability to work with multiple languages and cultures promotes global communication and accessibility.

Cross-Domain Relevance: BERT's deep language understanding has applications in many fields like healthcare, finance, law, and social sciences, where accurate language comprehension is vital for gaining insights and making informed decisions.

ABSTRACT

BERT presents a revolutionary approach to language representation.

BERT pretrains deep bidirectional representations using unlabeled text, effectively capturing contextual information from both left and right context in all layers.

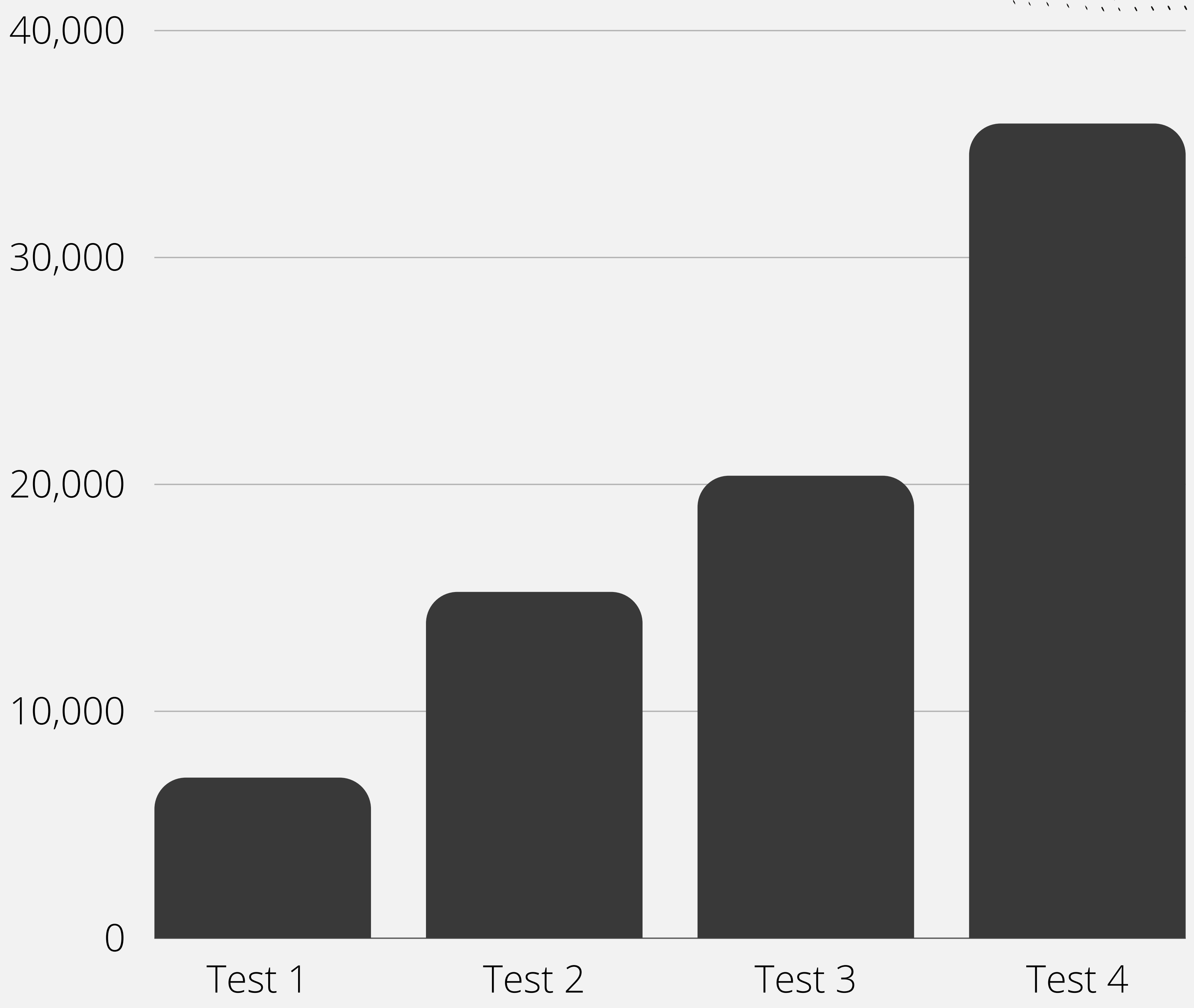
This unique design enables BERT's pretrained model to achieve remarkable performance across diverse tasks, requiring only minor adjustments to the output layer during fine-tuning.

INTRODUCTION

It helps to understand the context of words in a sentence by considering both left and right contexts. It achieves this through a "masked language model" pre-training approach, where some words in a sentence are hidden, and BERT learns to predict those words.

Sentiment Analysis

Keyword Extraction Tool



SYSTEM REQUIREMENTS

HARDWARE REQUIREMENTS

High Performance GPUs or Cloud based GPUS like AWS, Azure.
Atleast 16-32GB RAM
Storage for training data (SSDs)

SOFTWARE REQUIREMENTS

Deep Learning Framework such as PyTorch.
Python
Libraries for Data and Text Preprocessing Tools
Virtual Environment

LITERATURE REVIEW SURVEY TABLE

Title : BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Authors: Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova

Year : 2019

Inference: Using the pre-trained model to process input text and generate meaningful predictions or representations.

Advantage: BERT's bidirectional context enables it to capture complex linguistic relationships effectively.

Disadvantage: BERT's large size and computational demands can be resource-intensive.

EXISTING SYSTEM

Unidirectional models, such as the original, only consider context from one direction (either left-to-right or right-to-left) when pretraining.

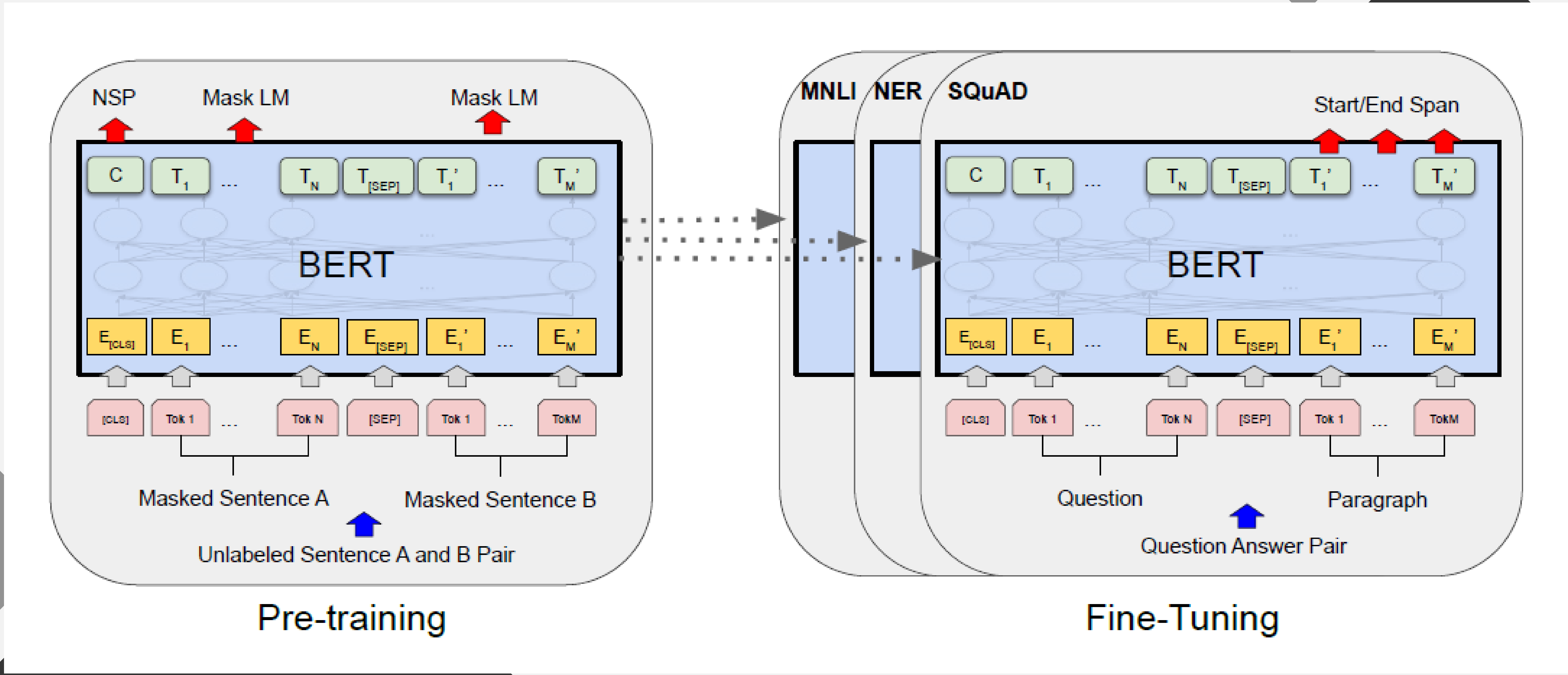
1. **Limited Contextual Understanding**: Traditional NLP models struggle to grasp nuanced language context.
2. **Task-Specific Complexity:** They often require intricate task-specific designs and manual efforts.

PROPOSED SYSTEM

Bidirectional models, such as BERT, considers context from both the directions (left-to-right and right-to-left) when pretraining.

1. **Deep Contextual Understanding**: BERT excels at capturing rich language context.
2. **Efficient Transfer Learning**: It offers efficient transfer learning, reducing task-specific complexities

ARCHITECTURE DIAGRAM



MODULE DESCRIPTION & ALGORITHMS USED

- Pre-Training BERT
- Fine-Tuning
- Unsupervised Feature-Based Approaches
- Unsupervised Fine-Tuning Approaches
- Transfer Learning from Supervised Data

PRE-TRAINING BERT

TASK 1 : MASKED LANGUAGE MODEL

TASK 2 : NEXT SENTENCE PREDICTION

TASK 1 : MASKED LANGUAGE MODEL (MLM)

- UNDERSTANDING CONTEXT
- MASKED LANGUAGE MODEL
- REPLACE THE TOKEN
  - ACTUAL [MASK] TOKEN (80%)
  - RANDOM WORD (10%)
  - ORIGINAL WORD (10%)
- TRAINING THE MODEL
- PROBLEMS DURING FINE-TUNING

TASK 2 : NEXT SENTENCE PREDICTION (NSP)

- WHAT IS NEXT SENTENCE PREDICTION ?
- TASKS SUCH AS :
  - QUESTION ANSWERING (Q&A)
  - NATURAL LANGUAGE INFERENCE (NLI)
- NEXT SENTENCE PREDICTION
  - TAKE A PAIR OF SENTENCES (A & B)
  - B IsNext of A (50%)
  - B NotNext of A (50%)
- TRAINING THE MODEL

PRE -TRAINING APPROACHES FOR GENERAL LANGUAGES -

UNSUPERVISED FEATURE-BASED APPROACHES

- Text-Based Learning
- Enhancing Language Comprehension
- Language Pattern Recognition
- Learning Word Meanings from Context
- Improving Language Understanding
- Language-Driven Computer Tasks

Example: Pre-training a model on a vast collection of books to teach it the general meanings of words and sentence structures.

UNSUPERVISED FEATURE-BASED APPROACHES

- **Language Learning**: Teach computers language by exposing them to diverse text.
- **Linguistic Foundation**: Build strong language basics for word meanings and grammar.
- **Versatility**: Models generalize this knowledge for various language tasks.
- **Pre-training**: Train models on big text datasets for language understanding.
- **Applications**: Used in tasks like text classification and sentiment analysis.

UNSUPERVISED FINE-TUNING APPROACHES

- Techniques for Specialized Training
- Conceptualizing Language-Savvy Computers
- Unsupervised Fine-Tuning Example
- Transformation into a Domain-Specific Expert
- Leveraging General Language Comprehension
- Generalist to Specialist Analogy

Example:
Think of BERT like a student who becomes a history expert after intense history study.

TRANSFER LEARNING FROM SUPERVISED DATA

- Transfer Learning from Supervised Data
- Leveraging Knowledge
- Adapting Knowledge to Limited Data
- Analogous to Cross-Subject Expertise

Example:

Source Task: BERT is trained on movie reviews to understand sentiment.
Target Task: Fine-tune BERT for a customer support chatbot.

FINE-TUNING

Fine-tuning BERT is a process used to adapt the pre-trained BERT model for specific natural language processing tasks, such as text classification or sentiment analysis.

Instead of training a model from scratch, fine-tuning leverages BERT's deep contextual understanding of language, making it efficient and effective.

It allows for impressive performance across a range of NLP tasks, as it benefits from the rich linguistic knowledge acquired during pre-training while adapting to specific tasks.

FINE-TUNING

Data Preparation: Task-specific datasets are created, usually including labeled examples for the target task.

Architecture Modification: BERT's pre-trained layers are retained, but a new classification layer is added on top. This layer is customized to match the specific task.

Training: The modified BERT model is fine-tuned using the task-specific data. During training, the model's weights are updated to better predict the task's output.

Evaluation: The fine-tuned model is tested on a separate validation dataset to assess its performance.

Deployment: Once the model meets the desired accuracy, it can be deployed for real-world applications.

BERT FINE-TUNING RESULTS ON 11 NLP TASKS

- Text Classification
- Named Entity Recognition (NER)
- Part-of-Speech Tagging
- Dependency Parsing
- Question Answering
- Text Summarization
- Machine Translation
- Language Modeling
- Text Similarity
- Sentiment Analysis
- Named Entity Linking (NEL)

NLP BENCHMARK DATASETS

GLUE (General Language Understanding Evaluation) (80.5%) :

- GLUE is a collection of NLP benchmark tasks that assess a model's general language understanding, including tasks like text classification, sentence pair classification, and more.
- It was designed to evaluate and compare the performance of various models, including BERT, on a diverse set of natural language understanding tasks.

SQuAD v1 (Stanford Question Answering Dataset, Version 1) (93.2%) :

- SQuAD v1 is a popular reading comprehension dataset that consists of questions posed on a set of Wikipedia articles, with the goal of extracting precise answers from the text.
- It is widely used to evaluate the ability of NLP models to answer questions by highlighting relevant passages in the provided context.

NLP BENCHMARK DATASETS

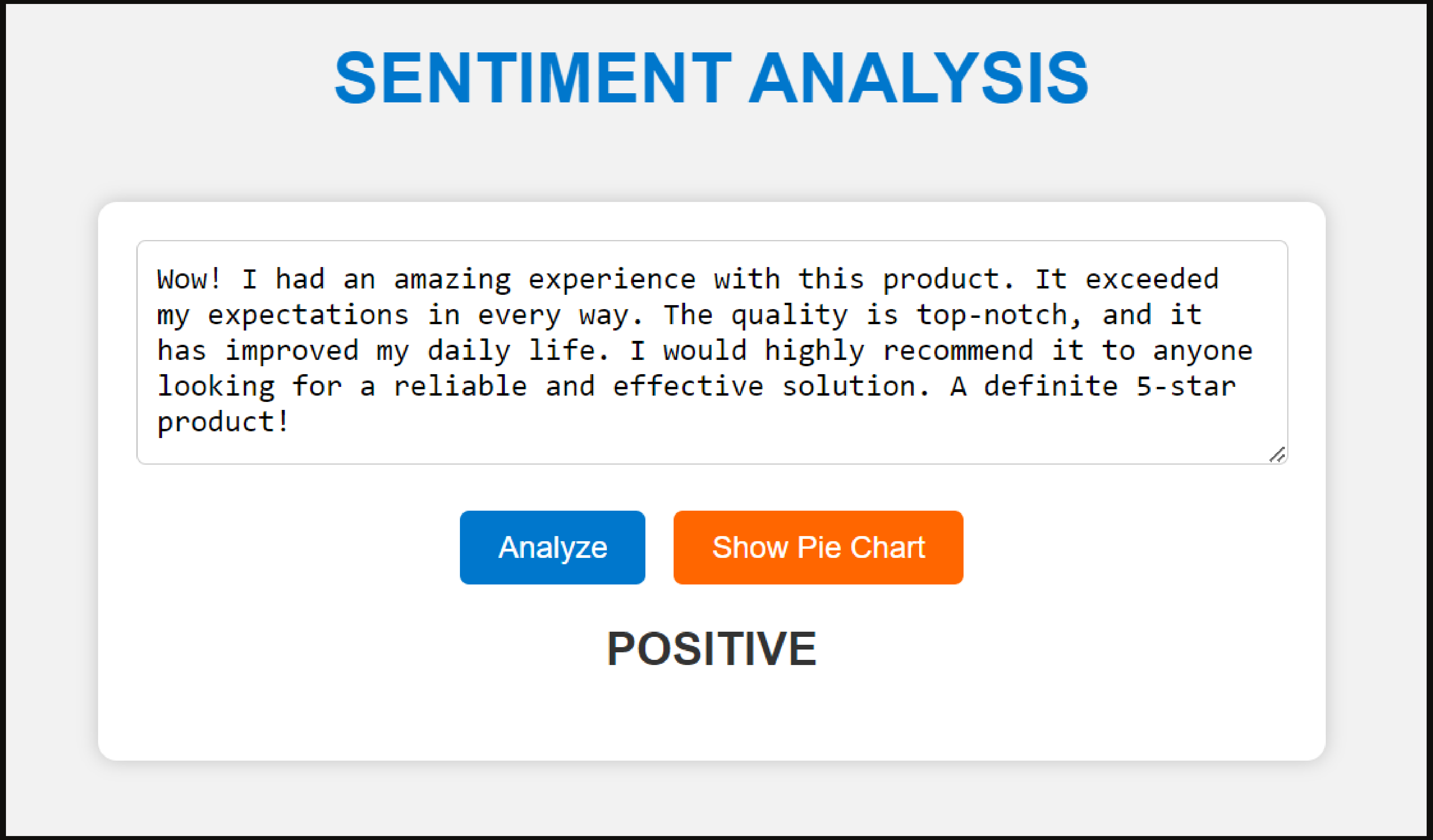
SQuAD v2 (Stanford Question Answering Dataset, Version 2) (83.1%) :

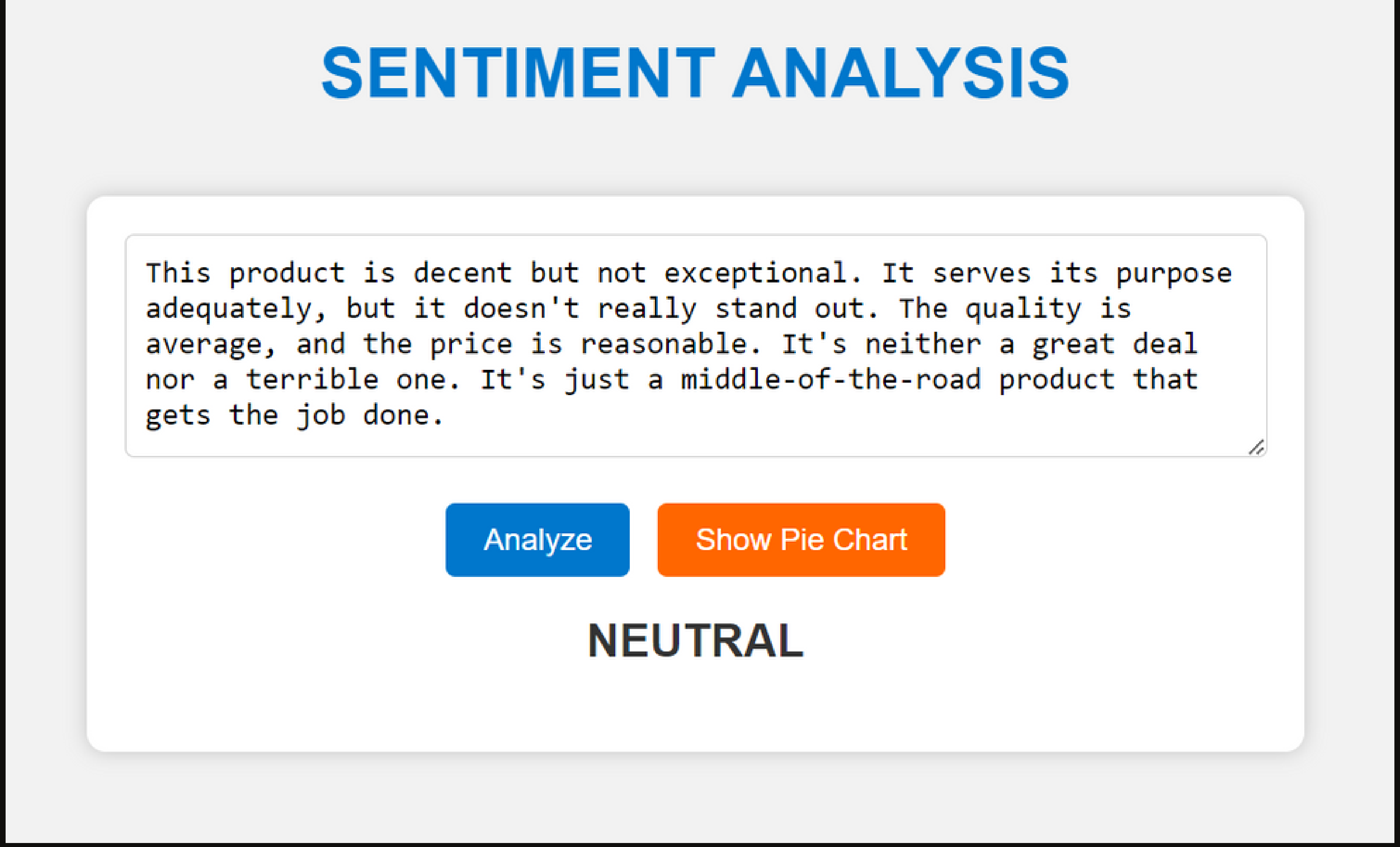
- SQuAD v2 is an extension of SQuAD v1, with the addition of unanswerable questions, making it more challenging for models to determine when an answer is not present in the passage.
- This dataset encourages models to provide a "no answer" response when a question cannot be answered from the given context.

SWAG (Situations With Adversarial Generations) (86.3%) :

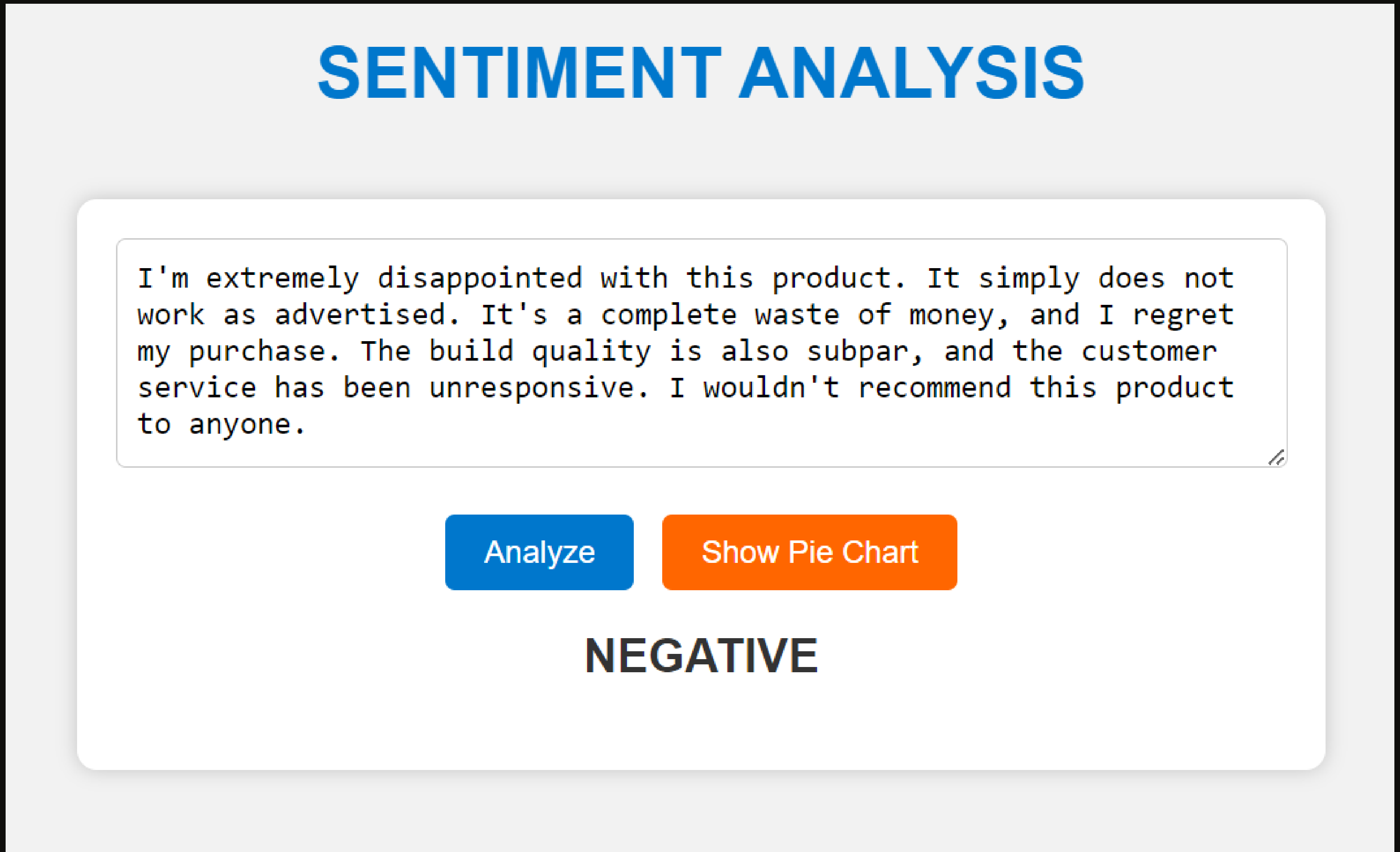
- SWAG is a dataset designed to assess a model's commonsense reasoning capabilities by presenting a sentence and asking the model to choose the most plausible continuation from multiple choices.
- It includes situations that require models to make inferences and choose contextually appropriate completions, challenging their world knowledge and reasoning abilities.

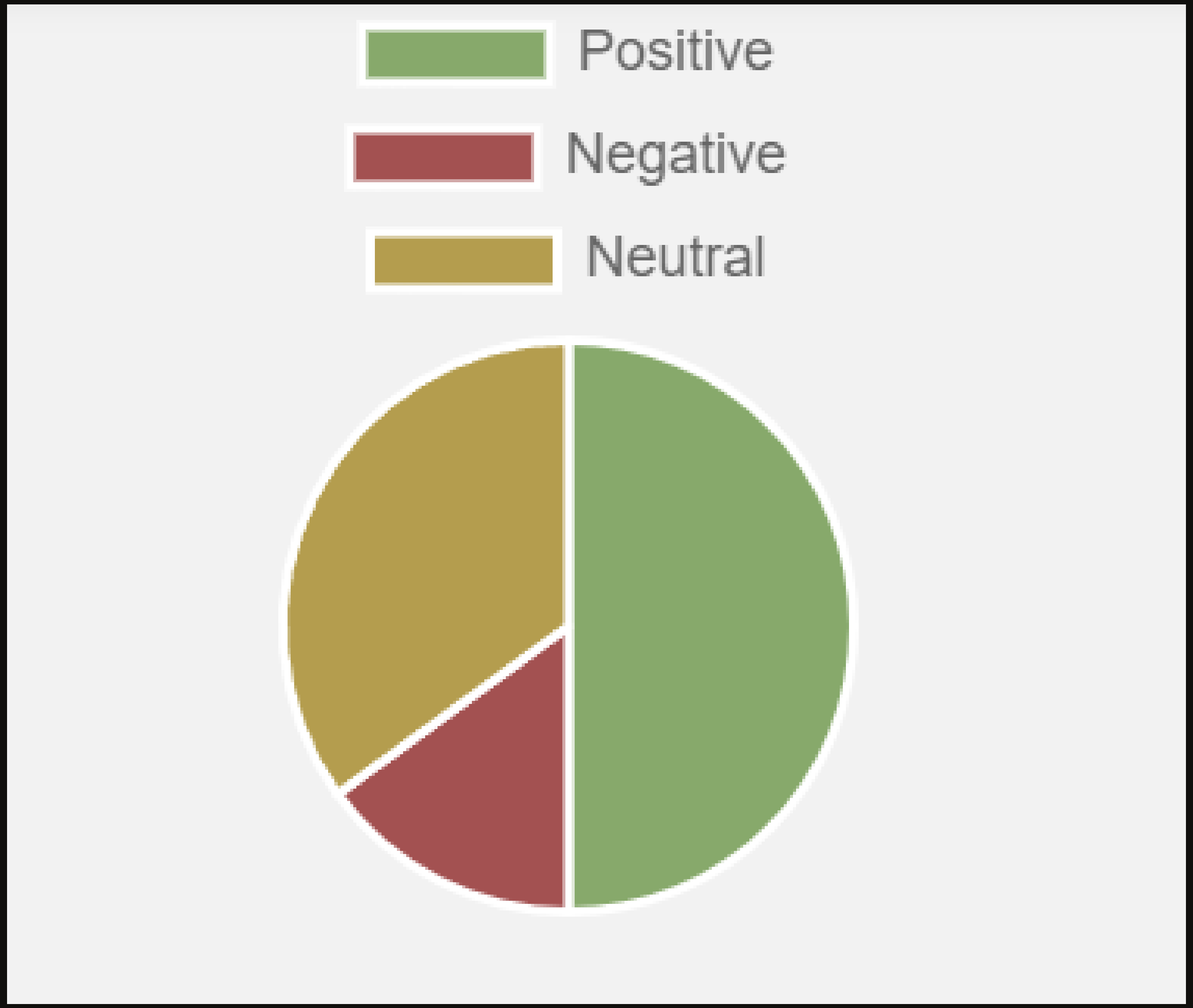
RESULT


SENTIMENT ANALYSIS
Wow! I had an amazing experience with this product. It exceeded my expectations in every way. The quality is top-notch, and it has improved my daily life. I would highly recommend it to anyone looking for a reliable and effective solution. A definite 5-star product!
Analyze
Show Pie Chart
POSITIVE


SENTIMENT ANALYSIS
This product is decent but not exceptional. It serves its purpose adequately, but it doesn't really stand out. The quality is average, and the price is reasonable. It's neither a great deal nor a terrible one. It's just a middle-of-the-road product that gets the job done.
Analyze
Show Pie Chart
NEUTRAL

RESULT


SENTIMENT ANALYSIS
I'm extremely disappointed with this product. It simply does not work as advertised. It's a complete waste of money, and I regret my purchase. The build quality is also subpar, and the customer service has been unresponsive. I wouldn't recommend this product to anyone.
Analyze
Show Pie Chart
NEGATIVE


Positive
Negative
Neutral

CONCLUSION

In conclusion, BERT has made a significant impact in the field of natural language processing. Its ability to pre-train on a massive amount of text data and capture contextual information has led to substantial advancements in various NLP tasks. BERT serves as a versatile foundation for fine-tuning on specific NLP tasks, consistently achieving state-of-the-art performance in areas such as text classification, question answering, and language understanding.

REFERENCES

- Daniel Cer, Mona Diab, Eneko Agirre, Inigo Lopez Gazpio, and Lucia Specia. 2017. Semeval-2017 task 1: Semantic textual similarity multilingual and crosslingual focused evaluation. In Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017), pages 1–14, Vancou ver, Canada. Association for Computational Lin guistics
- Alexis Conneau, Douwe Kiela, Holger Schwenk, Lo¨ ıc Barrault, and Antoine Bordes. 2017. Supervised learning of universal sentence representations from natural language inference data. In Proceedings of the 2017 Conference on Empirical Methods in Nat ural Language Processing, pages 670–680, Copen hagen, Denmark. Association for Computational Linguistics.
- Dan Hendrycks and Kevin Gimpel. 2016. Bridging nonlinearities and stochastic regularizers with gaus sian error linear units. CoRR, abs/1606.08415.
- Jeremy Howard and Sebastian Ruder. 2018. Universal language model fine-tuning for text classification. In ACL. Association for Computational Linguistics.

THANK YOU