**<FORKSCORE>**

**Submitted for**

**Statistical Machine Learning CSET211**

Submitted by:

**(<E23CSEU2103>) AVANTIKA SINGH**

**(<E23CSEU2127>) SHIVANI SHUKLA**

Submitted to

**DR. SUSMITA DAS**

**July-Dec 2024**

**SCHOOL OF COMPUTER SCIENCE AND ENGINEERING**

A close-up of a logo

Description automatically generated

**INDEX**

|  |  |  |
| --- | --- | --- |
| Sr.No | Content | Page No |
| 1 | ABSTRACT | 3 |
| 2 | INTRODUCTION | 3 |
| 3 | RELATED WORK | 4 |
| 4 | METHODOLOGY | 5 |
| 5 | SOFTWARE REQUIRED | 6 |
| 6 | EXPERIMENTAL RESULTS | 6 |
| 7 | CONCLUSION | 7 |
| 8 | FUTURE SCOPE | 8 |
| 9 | GITHUB LINK | 8 |

# 1. Abstract

This project focuses on analyzing customer reviews using machine learning techniques, specifically employing the RoBERTa model for sentiment analysis. The application provides insights into customer sentiment and associated emotions, helping businesses understand feedback effectively. It classifies reviews as Positive, Neutral, or Negative, offering confidence scores and visual distributions for better interpretability.

# 2. Introduction

In today’s digital age, customer reviews significantly influence the perception of products and services. Businesses can leverage customer feedback to enhance offerings and address concerns. However, manual analysis of extensive review data can be challenging. This project automates sentiment analysis using Natural Language Processing (NLP) and machine learning, enabling businesses to extract actionable insights efficiently.

The application uses RoBERTa, a transformer-based architecture, for sentiment classification, enabling accurate understanding and classification of review sentiments. By integrating the model with a Streamlit application, users can analyze reviews in real-time, making it a practical tool for businesses.

# 3. Related Work

Several studies and advancements have contributed to sentiment analysis using machine learning and NLP:

1. **BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding**
   * Authors: Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova
   * Summary: Introduced a robust transformer-based model for NLP tasks.
2. **RoBERTa: A Robustly Optimized BERT Pretraining Approach**
   * Authors: Yinhan Liu et al.
   * Summary: Optimized BERT architecture for better performance on downstream NLP tasks, including sentiment analysis.
3. **A Survey on Sentiment Analysis: Techniques and Applications**
   * Authors: A. P. Singh, B. K. Verma
   * Summary: Explores sentiment analysis techniques, highlighting machine learning approaches for improving accuracy and efficiency.
4. **Natural Language Processing with Transformers**
   * Authors: Lewis Tunstall, Leandro von Werra, Thomas Wolf
   * Summary: Discusses advanced NLP applications using transformer models like RoBERTa and Hugging Face tools.

# 4. Methodology

**Data Collection**

* Gathered customer reviews from platforms like Amazon and Yelp.
* Preprocessed the text data to remove noise, standardize formats, and prepare the dataset for modeling.

**Model Selection**

* Utilized the RoBERTa model, known for its robust pretraining and performance in NLP tasks.
* Fine-tuned the model on a labeled dataset of reviews to ensure optimal accuracy.

**Implementation**

* Developed a user-friendly **Streamlit application** for real-time review input and sentiment analysis.
* Integrated the fine-tuned RoBERTa model to classify sentiments and provide confidence scores.

**Sentiment Analysis**

* Classified reviews into three categories: Positive, Neutral, and Negative.
* Generated confidence scores and identified associated emotions for better interpretability.
* Visualized results through confidence distributions, enhancing user understanding.

## Software Required

* Programming Language: Python 3.8+
* Libraries: Hugging Face Transformers, PyTorch, Streamlit, NumPy, Pandas, Matplotlib
* Development Environment: Jupyter Notebook or VS Code

# 6. Experimental Results

* Successfully classified reviews with a high degree of accuracy using the fine-tuned RoBERTa model.
* The application provides:
  + **Primary Sentiment**: Positive, Neutral, Negative.
  + **Confidence Scores**: Distribution visualized for better insights.
  + **Associated Emotions**: Highlighting emotional tone of feedback.
* Case studies with sample data demonstrated effective sentiment categorization and user-friendly visualizations.

# 7. Conclusion

This project demonstrates the potential of machine learning techniques, specifically transformer-based models like RoBERTa, in automating sentiment analysis. The developed application serves as a practical tool for businesses, enabling efficient extraction of customer feedback insights.

# 8. Future Scope

1. **Aspect-Based Sentiment Analysis**:

* Identify specific features (e.g., product quality, delivery service) influencing customer sentiment.

2. **Model Enhancement**:

* Incorporate additional data from diverse sources to improve model robustness and accuracy.

3. **Multilingual Support**:

* Extend sentiment analysis capabilities to multiple languages to cater to a global audience.

4. **Advanced Visualizations**:

* Include dynamic dashboards for better representation of insights and trends over time.

**References:**

1. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv.
2. Liu, Y., Ott, M., Goyal, N., et al. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach*. arXiv.
3. Singh, A. P., & Verma, B. K. (2020). *A Survey on Sentiment Analysis: Techniques and Applications*. International Journal of Computer Applications.
4. Tunstall, L., von Werra, L., & Wolf, T. (2022). *Natural Language Processing with Transformers: Building Language Applications with Hugging Face*. O'Reilly Media.

# 9. GitHub link