SentimentScope: A Specialized Sentiment Analysis Web Platform

1st Hanuma Sashank Samudrala
Dept. of Computer Science
University of Maryland Baltimore County
Baltimore, USA
hanumas1@umbc.edu

2nd Ashish Athimamula
Dept. of Computer Science
University of Maryland Baltimore County
Baltimore, USA
ka42455@umbc.edu

3nd Varun G
Dept. of Computer Science
University of Maryland Baltimore County
Baltimore, USA
varung@umbc.edu

I. Abstract

SentimentScope is an advanced AI-driven web platform designed for precise sentiment and emotion analysis across diverse text sources, including tweets, product reviews, and general user texts. The platform offers an interactive user experience, allowing users to input text and receive sentiment predictions from three purpose-trained models. Each model is dedicated to a specific type of user input and is accessible through dedicated buttons, enabling seamless model experimentation. By leveraging specialized models trained on specific types of text data, SentimentScope achieves higher accuracy in sentiment analysis compared to generic models. The platform is ideal for businesses, marketers, and researchers seeking to understand customer sentiment, brand perception, and public opinion trends. It provides valuable insights that can be used to inform marketing strategies, product development, and social science research.

II. Project Description

SentimentScope exhibits an advanced AI-driven web platform tailored for precise sentiment and emotion analysis across diverse text sources, including tweets, product reviews, and general user texts. we've enhanced our web app's UI to offer a more interactive experience where users can input their text and sentiment is predicted for the input text by all the three purpose-trained models and based on the model's scores achieved the output from the particular model with highest score is displayed to the users as sentiment for the user text.SentimentScope uses special models to improve sentiment analysis accuracy. We achieve this by using three different models, each designed for specific types of text(trained RoBERTa Model on Tweets data, DistilBERT model on user-text-emotion data and DistilBert model on amazon product reviews data).This focused strategy helps us understand different data sources better, considering their unique details, which a generic model might miss.

This tool is ideal for businesses looking to gauge customer sentiment from online reviews or social media, marketers analyzing brand perception, or researchers studying public opinion trends. For instance, a marketing team could use SentimentScope to analyze customer feedback on a new product launch across social media platforms and review sites, enabling them to quickly adjust strategies based on real-time sentiment and emotion insights. Similarly, social scientists might employ the platform to understand public sentiment on societal issues, providing valuable data for studies and reports.

III. Related Work

SentimentScope's development has been guided by significant research in natural language processing (NLP) and sentiment analysis. We draw inspiration from several key studies that have advanced the understanding and application of sentiment analysis techniques.

A. DistilBERT Model for User-Text-Emotion Data

The DistilBERT model is a distilled version of the BERT model, designed for faster and more efficient performance while maintaining high accuracy. In SentimentScope, the DistilBERT model is trained on a diverse dataset of user-generated text containing emotions[1], model excels in capturing subtle emotional nuances and expressions in user text, providing valuable insights into user sentiments and emotions.

B. RoBERTa Model for Tweets

The Roberta model is a variant of the BERT (Bidirectional Encoder Representations from Transformers) model, specifically fine-tuned for understanding tweets. Tweets often contain informal language, abbreviations, and hashtags, making sentiment analysis challenging. The RoBERTa model is trained to capture these nuances and provide accurate sentiment analysis results for tweets. Its deep learning architecture allows it to understand context and sentiment expressions in short, informal text snippets[2], study's findings guided us to develop dedicated models trained on Twitter data, enhancing SentimentScope's ability to capture sentiment nuances within the social media context.

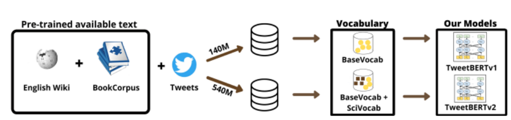

Fig. 1. Overview of the pre-training TweetBERTs [2]

C. DistilBERT Model for Amazon Poduct Reviews

Similar to the DistilBERT model, the DistilBERT model for Amazon reviews on mobile electronics is fine-tuned specifically for analyzing sentiments and opinions in product reviews related to mobile electronics on Amazon. This model is trained on a large corpus of Amazon reviews, enabling SentimentScope to accurately assess customer sentiments, preferences, and feedback regarding mobile electronic products[3], showed that DistilBERT is efficient in sentiment analysis. While we agree, we use multiple advanced models like RoBERTa and DistilBERT tailored for different text types, making SentimentScope more accurate.

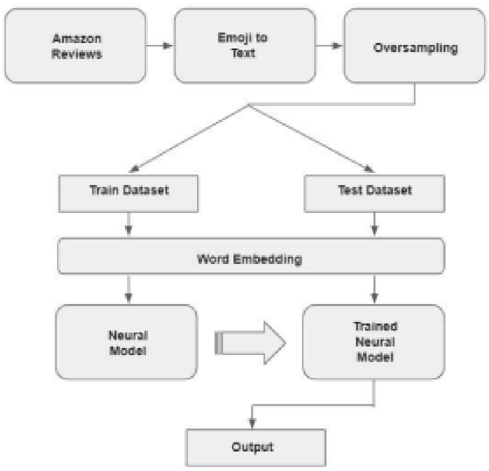

Fig. 2. The Proposed Algorithm of our work to train and test the models[3]

These studies have not only informed but also inspired SentimentScope's methodology, emphasizing the importance of context and advanced NLP models in enhancing sentiment analysis capabilities.

IV. Methods

A. Data Collection and Preprocessing

We collected three distinct datasets for SentimentScope: real-time Twitter tweets, Amazon product reviews, and user-generated content reflecting emotions from various sources. Prior to analysis, the datasets underwent preprocessing to enhance quality and usability. This preprocessing involved the removal of noise, such as treatment of missing values, handling special characters, and tokenization of text from tweets, product reviews, and general user-generated texts. These steps were crucial to ensure the data was cleaned and formatted appropriately for subsequent analysis.

B. Model Selection

For SentimentScope, we implemented a multi-model approach, selecting three models tailored to specific text sources: a RoBERTa model for tweets, a DistilBERT model for user-text-emotion data, and another DistilBERT model for Amazon product reviews. These models were chosen based on their demonstrated performance in handling the nuances of sentiment analysis within their respective text domains. The selection process focused on models known for their ability to effectively capture sentiment nuances in each dataset.

C. Fine-tuning and Training

Each selected model underwent a rigorous fine-tuning and training(on training set of source dataset) process to optimize its performance for sentiment and emotion analysis tasks. Fine-tuning involved adjusting model parameters and hyperparameters to suit the characteristics of the target data, which was performed on a validation split of the source dataset. The fine-tuned models were then integrated into the SentimentScope platform, which features a user-friendly interface for sentiment analysis. This platform allows users to input text, and sentiment is predicted by all three purpose-trained models. Based on the model's scores achieved, the output from the model with the highest score is displayed to the users as the sentiment for the input text.

D. Analysis and Interpretation

SentimentScope provides detailed analysis results, including sentiment scores, emotion classifications, and sentiment trends over time. The platform also offers visualizations and insights to aid in interpreting the sentiment and emotional content of the analyzed text data.

We conducted extensive evaluations to assess the performance and accuracy of SentimentScope in comparison to existing sentiment analysis tools. Evaluation metrics such as precision, recall, F1 score and accuracy were used to measure the backend model's performance on sentiment and emotion analysis tasks. User feedback and iterative testing were instrumental in refining SentimentScope's functionality and user experience. Continuous improvements were made based on user's suggestions and emerging trends in sentiment analysis research.

V. Data

In developing the SentimentScope project, we meticulously curated three datasets, each meticulously aligned with specific analysis objectives: Sentiment Analysis of Product Reviews, Text Emotion Detection, and Twitter Sentiment Analysis. Our selection criteria prioritized relevance, diversity, and the capacity to finely train our models for maximum accuracy.

A. Sentiment Analysis for Product Reviews

We are utilizing the 'amazon_us_reviews/Mobile_Electronics_v1_00/0.1.0.csv' dataset from TensorFlow Datasets[1], which contains 131,000 product reviews. Each review is accompanied by a star rating that we use as an indicator of sentiment—categorized as positive, negative, or neutral. Example reviews range from positive feedback like "This is a great wiring kit I used it to set up..." to negative experiences such as "Does not work." We've structured our data into 80% for training, 10% for validation, and 20% for testing purposes. The preprocessing

[1] Product reviews Tensorflow Dataset(Amazon US Reviews)

stage focuses on cleaning the dataset by eliminating irrelevant columns to ensure only significant features are used for model training. A significant challenge we face is addressing the subjectivity in sentiment expression and the variability in user ratings, which requires nuanced handling to accurately train our models.

B. Text-Emotion Detection

Our project incorporates the ‘emotion.csv‘ dataset, sourced from Hugging Face Datasets[2], which features 20,000 entries, each tagged with specific emotions like joy, anger, and sadness. Examples from the dataset include expressions such as ”I didn’t feel humiliated,” indicating joy, and ”I am feeling grouchy,” denoting anger. We’ve allocated 80% of the data for training, 10% for validation, and the remaining 10% for testing. During the preprocessing phase, we’ve concentrated on examining the dataset to gain insights into the distribution of emotion labels and the length of texts, both vital for effective model training and achieving high accuracy. One anticipated challenge is the subjective nature of emotion interpretation and the necessity for our model to be finely attuned to the nuances of emotional expression.

C. Twitter Data Sentiment Analysis

We are working with the ‘twitterentitysentiment-analysis.csv‘ dataset obtained from Kaggle[3], which comprises 79,089 tweets, each annotated with a sentiment label: positive, negative, or neutral. Examples of the dataset range from negative sentiments like ”@Kenichan I dived many times for the ball. Managed to save 50% The rest go out of bounds” to positive ones such as ”@britishmuseum @TudorHistory What a beautiful jewel/portrait. Is the ’R’ for Rex?”. For our project, we’ve divided the data into 69,491 entries for training, 998 for validation, and 8,600 for testing. A key challenge we face with this dataset is its inherent nature of tweets being brief and laden with informal language, slang, and abbreviations, all of which demand sophisticated processing techniques to accurately discern the underlying sentiment.

Each dataset presents unique challenges, from the variability and subjectivity of sentiments and emotions to the specific linguistic features of tweets. Our preprocessing efforts are geared towards mitigating these challenges by cleaning and structuring the data for optimal model performance. The diversity of our data sources enriches our models’ training, equipping them to handle real-world applications effectively.

VI. Evaluation

In evaluating the performance of SentimentScope, we focus on two key aspects: accuracy metrics and human evaluation.

[2]User Text Emotions Huggingface Dataset
[3]Twitter Tweets Kaggle Dataset

A. Accuracy Metrics

We assess the performance of our models using standard evaluation metrics such as precision, recall, F1 score and accuracy. These metrics provide quantitative measures of the models’ performance in sentiment analysis and emotion detection tasks. For sentiment analysis tasks using RoBERTa and DistilBERT, we measure the accuracy of sentiment classification into positive, negative, Neutral and irrelevant categories. High precision and recall indicate the models’ ability to correctly classify sentiments. In emotion detection, we measure the accuracy of identifying specific emotions like joy, sadness, anger, love, fear, and surprise. A high F1 score signifies the models’ effectiveness in multi-label emotion detection.

B. Human Evaluation

For human evaluation, we provided 15 real-time examples from each category, including tweets, Amazon product reviews, and user-generated texts. We manually checked whether the web app delivered the expected output for each input. Specifically, we verified if the final sentiment prediction was based on the highest confidence score achieved by the models for the given input, as per our web app design. This involved assessing whether the web app correctly implemented the intended functionality of selecting the sentiment prediction from the model with the highest confidence score. By doing this, we ensured that the system was functioning as designed and providing accurate sentiment analysis results for diverse text inputs.

VII. Results

A. Text-Emotion Detection

The DistilBERT model trained on the ”emotions” dataset demonstrates strong and improving performance over two epochs in sentiment analysis. Initially, the model starts with a training loss of 0.860600 and a validation loss of 0.343724, but by the second epoch, these figures significantly drop to 0.263400 and 0.224668 respectively, indicating effective learning and good generalization capabilities. The accuracy and F1 scores start at a commendable 89.40% and 88.90% and rise to 92.15% and 92.13% by the end of the second epoch, showing enhanced predictive accuracy and a balanced precision-recall relationship. Overall, the model’s performance is robust, making it well-suited for practical applications in NLP sentiment analysis.

[500/500 04:10, Epoch 2/2]

Epoch	Training Loss	Validation Loss	Accuracy	F1
1	0.860600	0.343724	0.894000	0.889023
2	0.263400	0.224668	0.921500	0.921336

Fig. 3. DistlBERT emotion detection evaluation results

B. Twitter Data Sentiment Analysis

The RoBERTa model trained for sentiment classification on Twitter tweets exhibits consistent improvement and robust performance across three epochs. The model's training loss decreases steadily from 0.5002 in Epoch 1 to 0.5379 in Epoch 3, indicating effective learning. Performance metrics for sentiment categories show high precision and recall, particularly for the 'Positive' class with a precision of 0.8623 and an impressive recall of 0.9495, leading to a high F1-score of 0.9038. Other classes—'Irrelevant', 'Negative', and 'Neutral'—also demonstrate good model performance with F1-scores above 0.7892. The overall accuracy of the model reaches 0.87, with the macro and weighted averages for precision, recall, and F1-score above 0.86, suggesting that the model is well-calibrated and performs strongly across different tweet sentiments. This model showcases its capability to effectively classify sentiments in Twitter data, making it a valuable tool for real-world applications in NLP.

```
Some weights of RobertaForSequenceClassification were not initializ
You should probably TRAIN this model on a down-stream task to be ab
/usr/local/lib/python3.10/dist-packages/transformers/optimization.p
  warnings.warn(
Epoch 1, Loss: 0.500424861907959
Epoch 2, Loss: 0.7420348525047302
Epoch 3, Loss: 0.5379140377044678
              precision    recall  f1-score   support

  Irrelevant     0.8187    0.7616    0.7892       172
    Negative     0.8938    0.9173    0.9054       266
     Neutral     0.8855    0.8140    0.8483       285
    Positive     0.8623    0.9495    0.9038       277

    accuracy                         0.8700      1000
   macro avg     0.8651    0.8606    0.8616      1000
weighted avg     0.8698    0.8700    0.8687      1000

Accuracy: 0.87
```

Fig. 4. RoBERTa model for Tweets evaluation results

C. Amazon Product Reviews

The DistilBERT model trained on Amazon product reviews shows a promising trend across three epochs, demonstrating substantial improvements in performance metrics. Initially, the model presents a training loss of 0.141000, which interestingly increases in the second epoch to 0.242700, before decreasing significantly to 0.089600 in the third epoch, suggesting some adjustments in learning. The validation loss starts relatively high at 0.718505 but improves consistently, reaching 0.868370 by the third epoch. Accuracy and F1 scores show consistent growth, starting at 83.1187% and 82.5322%, respectively, and increasing to 83.3572% and 83.4624% by the third epoch. Precision and recall metrics remain balanced, hovering around 84%, which indicates a reliable model in accurately predicting sentiment classifications. The overall evaluation scores, with an accuracy of 86.97%, precision of 84.21%, and recall of 83.69%, underscore the model's robustness in handling diverse sentiment analysis tasks for e-commerce reviews, making it an effective tool for real-world applications in sentiment analysis.

For human evaluation, we tested 45 real-time examples, with 15 examples from each category: tweets, Amazon product reviews, and user-generated texts. Our goal was to verify whether the web app provided the expected outputs based on

Epoch	Training Loss	Validation Loss	Accuracy	F1	Precision	Recall
1	0.141000	0.718505	0.831187	0.825322	0.825993	0.831187
2	0.242700	0.734169	0.829757	0.835645	0.843065	0.829757
3	0.089600	0.868370	0.833572	0.834624	0.835858	0.833572

```
[132/132 00:08]
{'eval_loss': 0.8235502243041992, 'eval_accuracy': 0.8369876072449952, 'eval_f1': 0.8389!
```

Fig. 5. DistilBERT model for Amazon Product reviews evaluation results

the highest scoring model prediction. Out of the 45 examples tested, the web app correctly predicted the sentiment for 44 examples, demonstrating its effectiveness and accuracy in delivering the intended results. This evaluation confirmed that the web app functions as designed, successfully selecting the sentiment prediction from the model with the highest score.

VIII. Discussion

The SentimentScope project showcased the challenges and opportunities of building a comprehensive sentiment analysis tool using advanced AI models. One challenge we encountered was the complexity of fine-tuning multiple models (RoBERTa and DistilBERT) across diverse data types. Early attempts to unify model predictions proved less effective due to varying text sources and nuances in sentiment expressions. In hindsight, establishing dedicated models for different text sources from the start was a more efficient approach. If we were to restart the project, more rigorous cross-validation and diverse hyperparameter tuning would be prioritized to enhance model generalization. Integrating the optimized models into our web platform, ensuring a seamless user experience, is a also little challenging we're currently facing.

For future work, enhancing data preprocessing to address informal text patterns more comprehensively, especially on Twitter data, could yield improved predictions. Moreover, incorporating external lexicons or real-time sentiment adjustments could refine sentiment interpretation, while integrating additional language support may expand the tool's applicability. A good next step involves developing user-friendly visualization tools that summarize sentiment trends effectively, empowering businesses and researchers to uncover nuanced insights for strategic decision-making.

IX. Source Code Git Repo

Here is the link to our source code git repo : Source Code

X. References

[1]Acheampong, Nunoo-Mensah, & Chen. (2021, February 8). Transformer models for text-based emotion detection: A review of BERT-based approaches. SpringerLink.https://link.springer.com/article/10.1007/s10462-021-09958-2

[2]Qudar,& Mago. (2020, October). TweetBERT: A Pretrained Language Representation Model for Twitter Text Analysis. arXiv.org e-Print archive. https://arxiv.org/pdf/2010.11091

[3]Pramanik, & Maliha. (2022). Analyzing sentiment towards a product using DistilBERT and LSTM. IEEE Xplore. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10037634