CSE 573

Project Demo Presentation

Stance Detection

Group 9

Group Members

- Avish Khosla
- Baibhav Phukan
- Gautham Maraswami
- Sai Rathnam Pallayam Ramanarasaiah
- Sai Vikhyath Kudhroli
- Tanuja Renu Sudha

Problem Statement

- In our project, have determined the stance of given individuals towards certain topics by evaluating their tweets.
- We perform this analysis using deep learning models such as Bi-LSTMs and BERT along with Support Vector Machines.
- Later, we have and compared the findings from the above techniques used in this study.

System Architecture and algorithm

BERT with SVM

- With its ability to analyze word and phrase relationships in a sentence, BERT is an ideal tool for capturing the context and meaning of the text.
- We used pre-trained BERT-Base model to generate a set of features, which were then used as input to an SVM classifier for predicting stance. SVM is a suitable choice because it can establish clear decision boundaries for the three stances (Favor, Against, and None).

System Architecture and algorithm

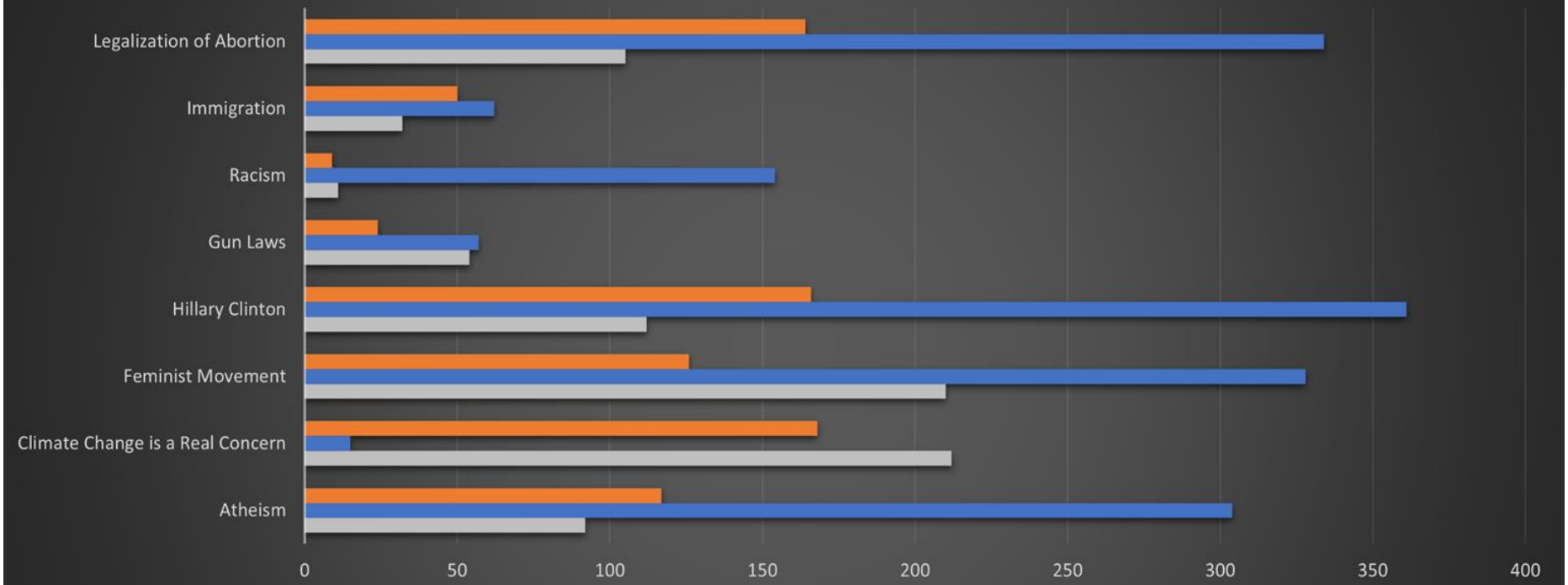
Bidirectional Long Short-Term Memory Networks (Bi-LSTMs)

- Bi-LSTMs are suitable because of their ability to process sequential text in both forward and backward directions, allowing them to capture contextual information from the entire input sequence.
- Furthermore, Bi-LSTMs can overcome the issue of vanishing gradients in deep neural networks by utilizing LSTM cells, which maintain information over time.

Datasets

- We are training our models on SemEval-2016 Task 6A dataset. This dataset contains 2,914 Tweets related to five different topics:
  - Atheism
  - Climate change
  - The feminist movement
  - Hillary Clinton
  - Legalization of abortion.
- Each Tweet has been manually annotated by humans and assigned a stance label of positive, negative, or none towards the topic.
- We are also further augmenting our training set by scraping more tweets related to the above topics plus recently trending topics as well. We aim to assign labels to these tweets by running a clustering algorithm on them.

Dataset Description



	Atheism	Climate Change is a Real Concern	Feminist Movement	Hillary Clinton	Gun Laws	Racism	Immigration	Legalization of Abortion
Number of Tweets in Neither	117	168	126	166	24	9	50	164
Number of Tweets in Against	304	15	328	361	57	154	62	334
Number of Tweets in Favor	92	212	210	112	54	11	32	105

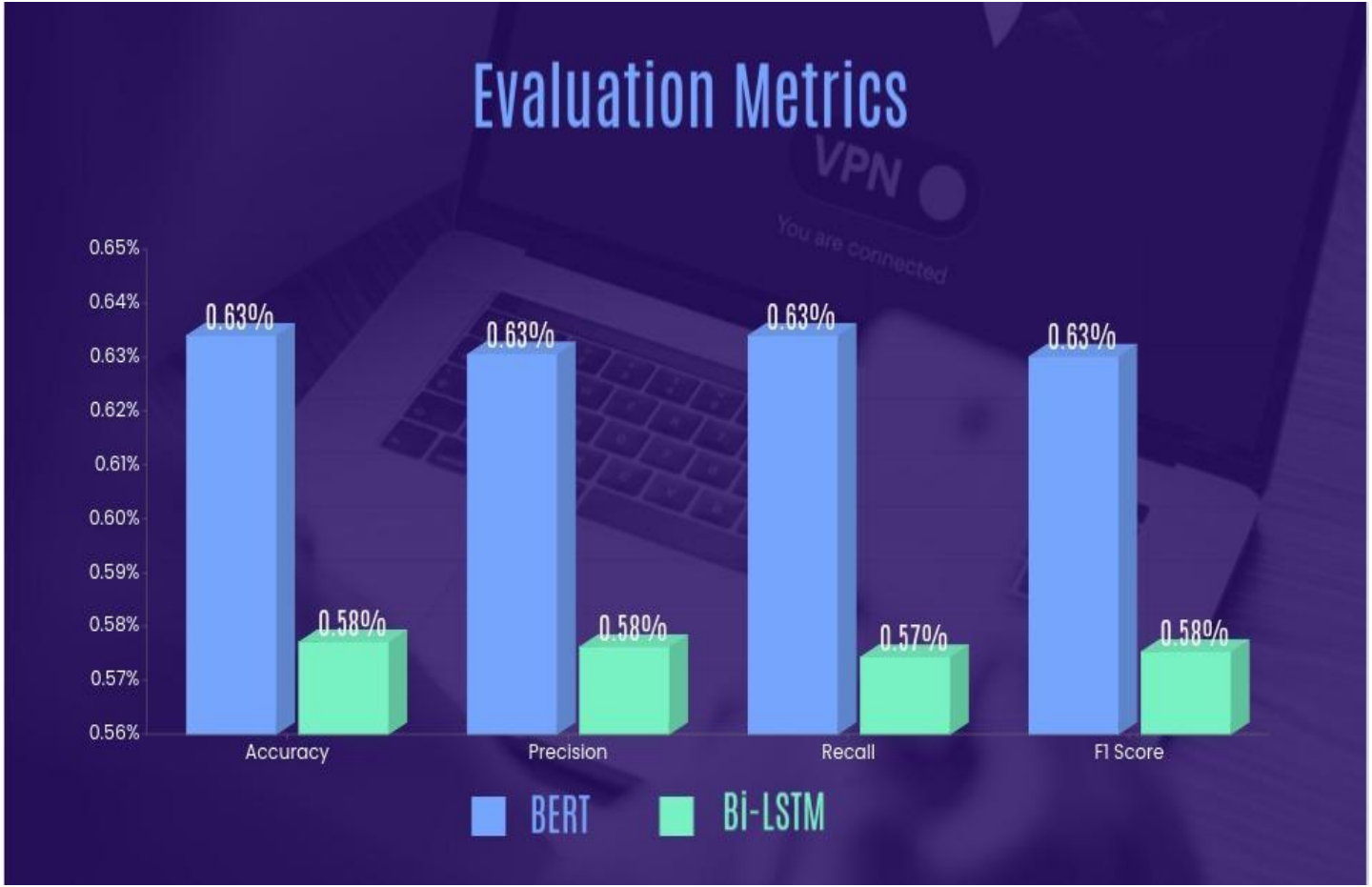
Data Processing

Data Preprocessing

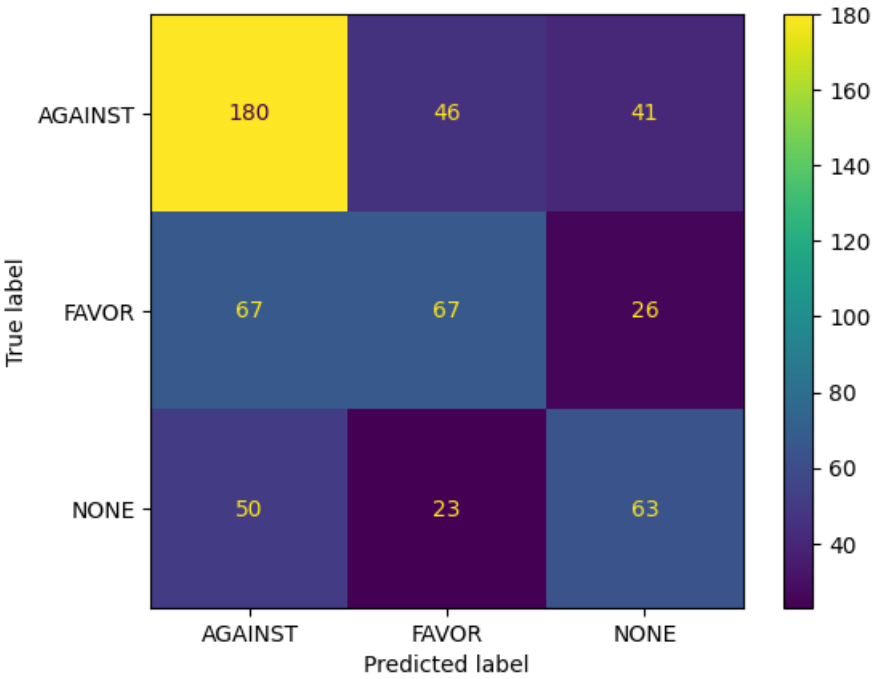
1. **Data Cleaning** - Removing URLs and mentions, replacing emoticons with their corresponding sentiment, correcting any spelling errors in the text.
2. **Tokenization** - Splitting the text into individual words to enable us to represent the text as numerical vectors.
3. **Removal of stop words** - Removing common words (like *and, the, a, an*, etc.) to focus more on the important and informative parts of the text.
4. **Stemming and Lemmatization** - Clustering similar words together for pinpointing pertinent keywords and extracting the most significant features.

Data Augmentation

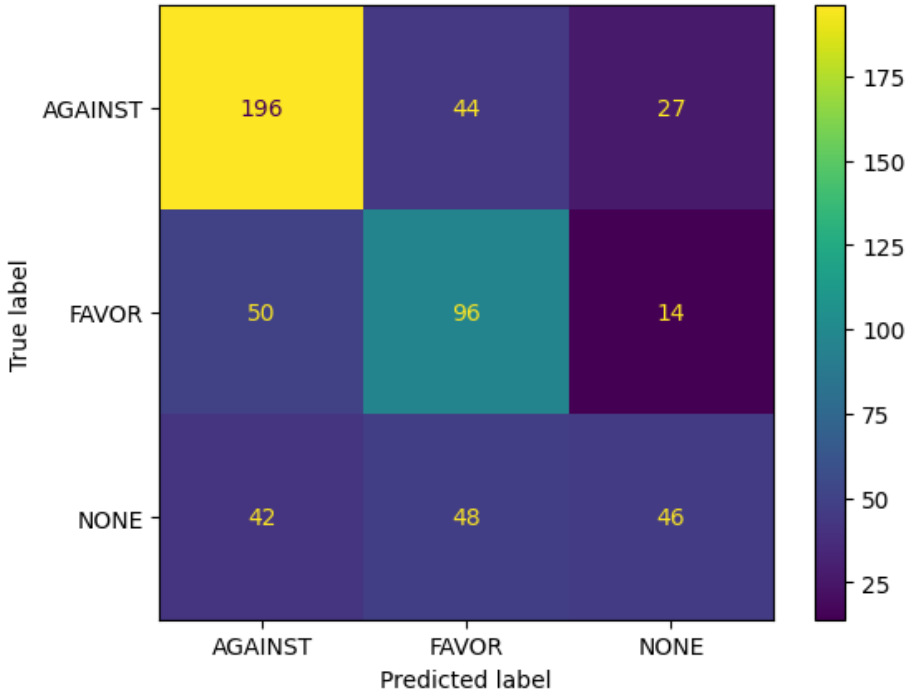
1. **Data Expansion** - Scraping Twitter for more tweets related to the existing topics plus recently trending topics.
2. **Synonym expansion** - Including synonyms to capture diverse opinions understand more nuanced contexts.
3. **Phrase expansion** - Expanding the use of related phrases to improve the comprehensiveness of the analysis.
4. **Query reformulation** - By reformulating the query, the model can become more adaptable to predict the stance regardless of the specific way in which an opinion is expressed.


Evaluation Metrics
0.65%
0.64%
0.63%
0.62%
0.61%
0.60%
0.59%
0.58%
0.57%
0.56%
0.63%
0.58%
0.63%
0.58%
0.63%
0.57%
0.63%
0.58%
Accuracy
Precision
Recall
F1 Score
BERT
Bi-LSTM

Confusion Matrix

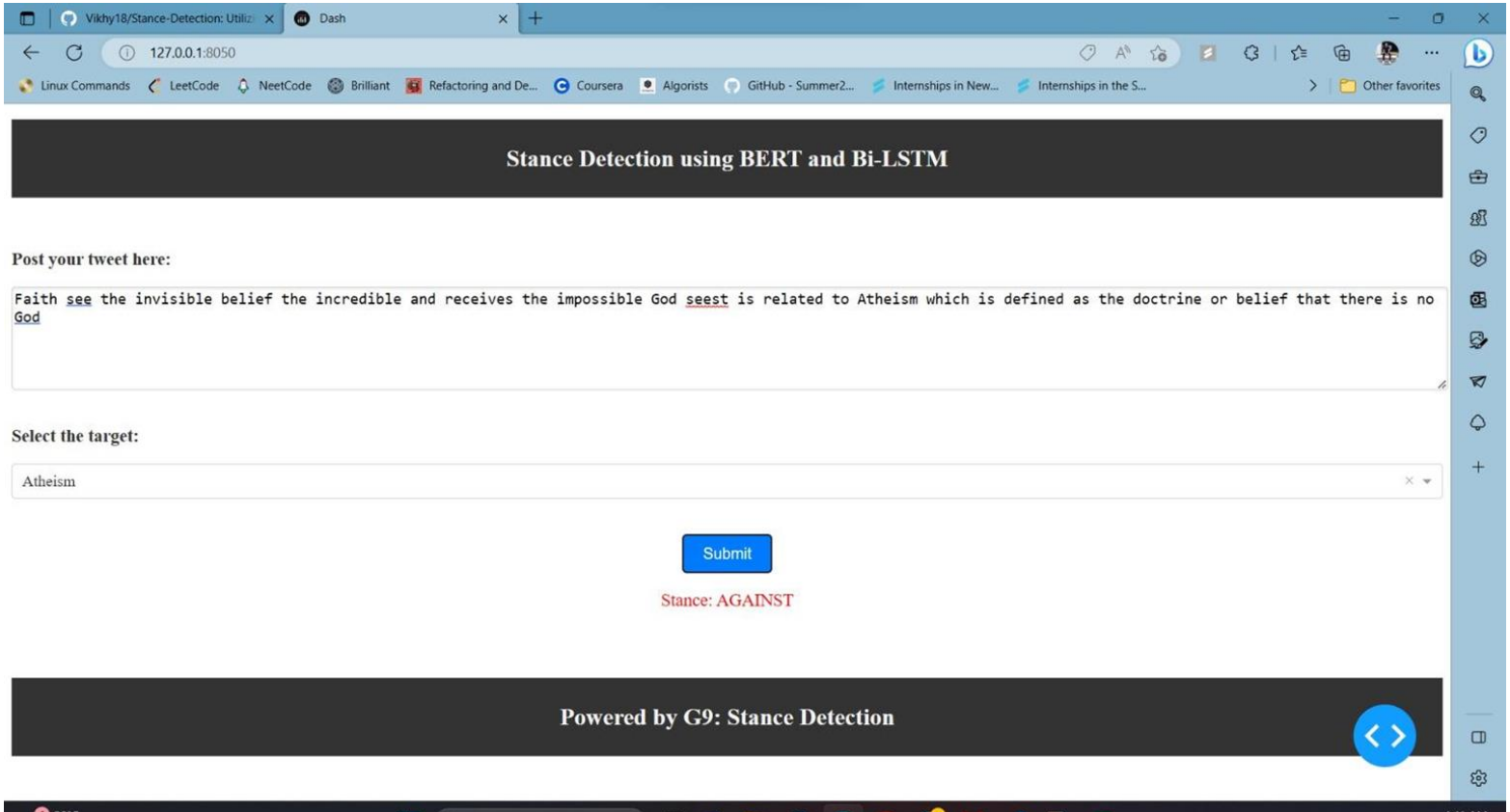


BERT



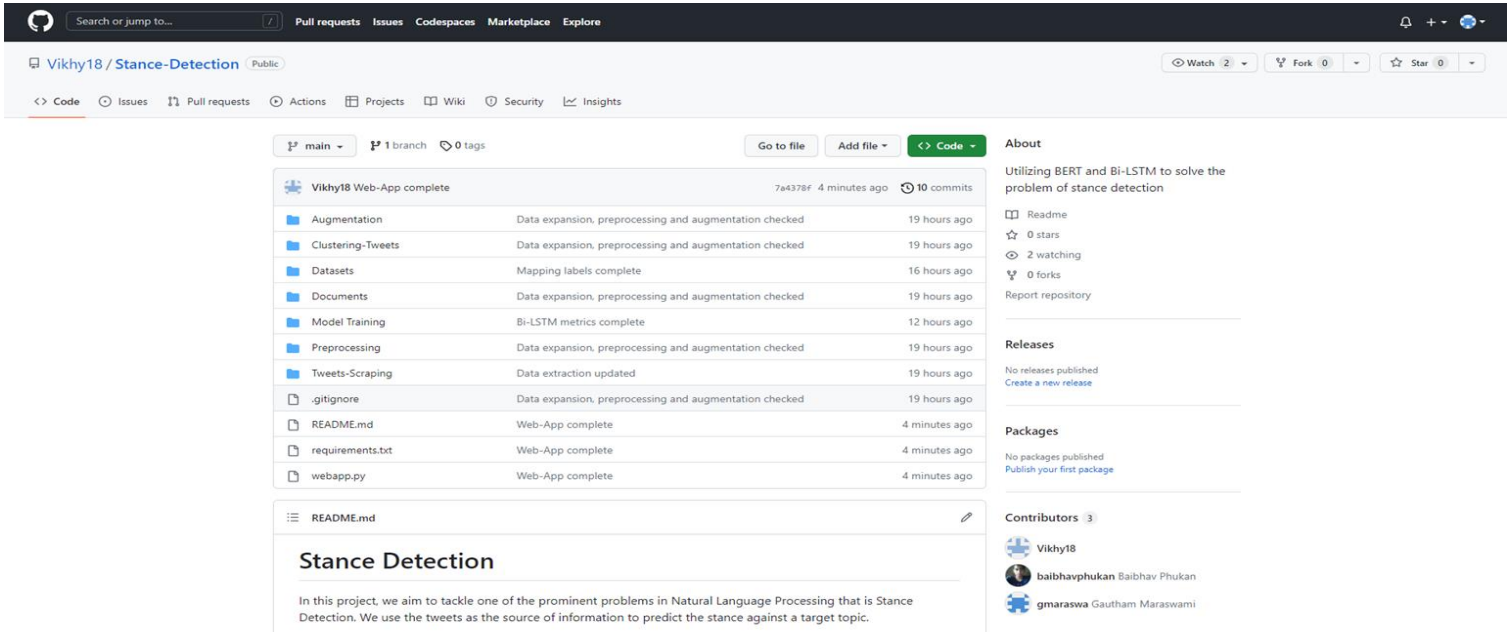
BI- LSTM

Product Screenshots



Github Link

https://github.com/Vikhy18/Stance-Detection



Management Plan

Task	Deadline	Members Responsible	Status
Data cleaning and preprocessing	16 March	Tanuja Renu Sudha	Done
Data Augmentation	23 March	Sai Rathnam Pallayam Ramanarasaiah	Done
Web Scraping	1 April	Baibhav Phukan	Done
Model Training	5 April	Sai Vikhyath Kudhroli, Avish Khosla	Done
Performance Evaluation	8 April	Gautham Maraswami	Done
Final Report	28 April	All	In-Progress

THANK YOU