

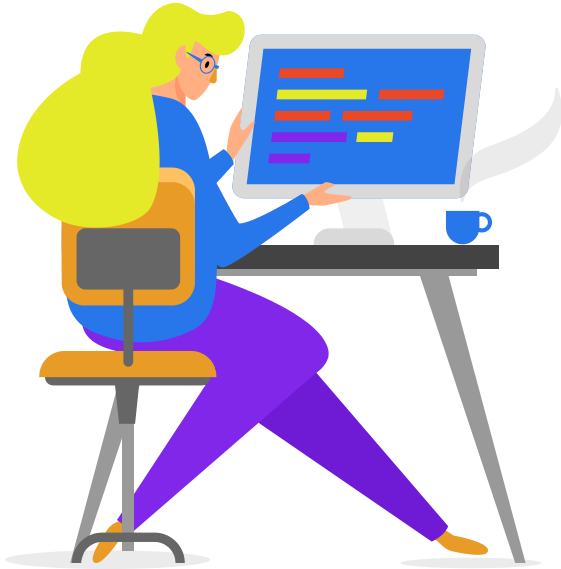
CSE573 Group 10 Project 23:

# Stance Detection



By:  
Aatish Chaudhari  
Arhat Kobawala  
Caleb Panikulam  
Mohak Acharya  
Priyal Desai

# Table of contents



01

**Problem Definition**

02

**System  
Architecture &  
Algorithms**

03

**Datasets**

04

**Evaluations**

05

**UI Interfaces/Data  
Visualization**

06

**Project Plan: Tasks,  
Deadlines, Division of  
Work**

07

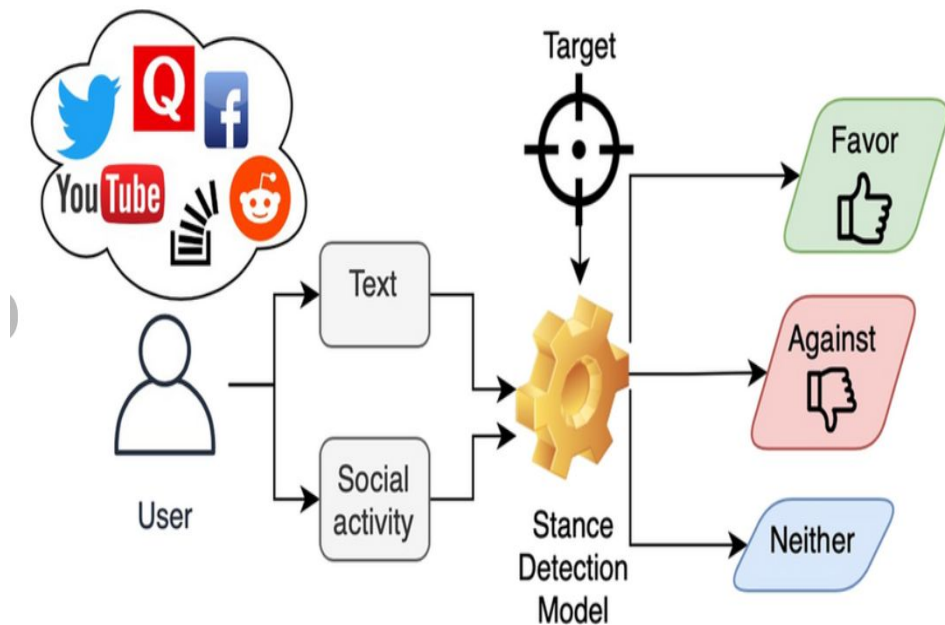
**GitHub Link**



# Problem Definition



- Stance detection is used to determine whether a statement's perspective or position supports, opposes, or is neutral to the topic.
- It enables researchers and businesses in understanding how people feel about a specific issue or topic, as well as in identifying fake news or propaganda.

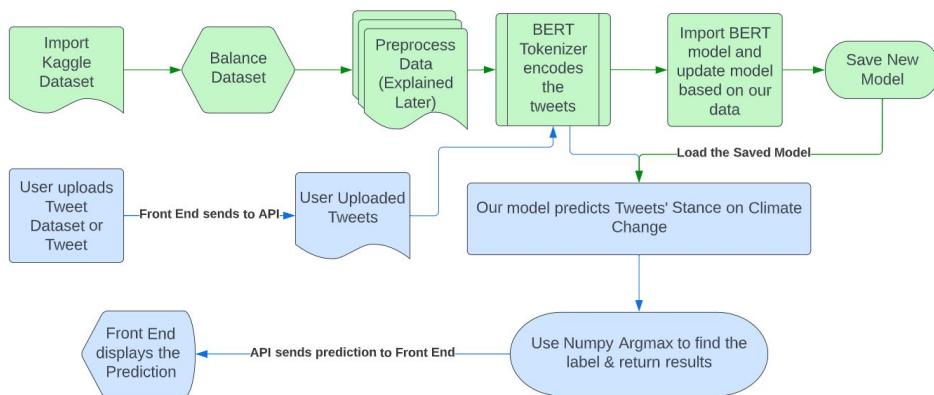




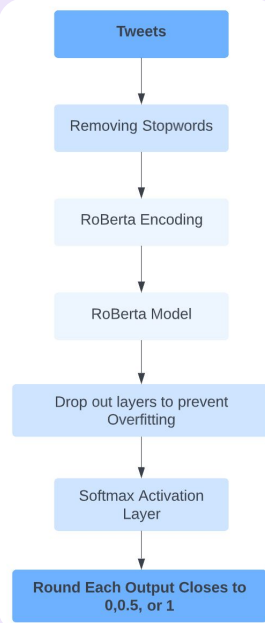
# System Architecture & Algorithms



## System Architecture FlowChart



## BERT Initial Model FlowChart





# System Architecture & Algorithms

## BERT Model FlowChart

input_ids	input:	[(None, 70)]	attention_mask	input:	[(None, 70)]
InputLayer	output:	[(None, 70)]	InputLayer	output:	[(None, 70)]

tf_bert_model	input:	(None, 70)			
TfBertModel	output:	TFBaseModelOutputWithPoolingAndCrossAttentions(last_hidden_state=(None, 70, 768), pooler_output=(None, 768), past_key_values=None, hidden_states=None, attentions=None, cross_attentions=None)			

global_max_pooling1d	input:	(None, 70, 768)
GlobalMaxPooling1D	output:	(None, 768)

dense	input:	(None, 768)
Dense	output:	(None, 128)

dropout	input:	(None, 128)
Dropout	output:	(None, 128)

dense_1	input:	(None, 128)
Dense	output:	(None, 32)

dense_2	input:	(None, 32)
Dense	output:	(None, 3)

# Datasets (Description, Size and Preprocessing Steps)

**Description:** Obtained a twitter stance detection dataset from Kaggle and scraped the tweets using tweet IDs using snsrape.

**Size:**  
52000

## Preprocessing Steps

**01** Clean ASCII

---

**02** Remove Mentions

---

**03** Remove Links

---

**04** Remove Punctuation

---

**05** Remove Stopwords

---

**06** Open Contractions

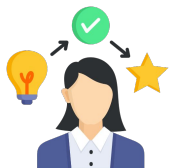
---

**07** Remove Accents

---

**08** Lemmatize Tokens

---



# Evaluations (Metrics, Experiments, Preliminary Findings)

## SVM

Accuracy: 0.71  
Precision: 0.71  
Recall: 0.71  
F1 Score: 0.71

## Logistic Regression

Accuracy: 0.7  
Precision: 0.7  
Recall: 0.7  
F1 Score: 0.7

## Random Forest

Accuracy: 0.66  
Precision: 0.66  
Recall: 0.66  
F1 Score: 0.66

## BERT

Accuracy: 0.91  
Precision: 0.91  
Recall: 0.91  
F1 Score: 0.91

## K-Means

Silhouette Avg: 0.015  
DB index: 4.21

# Evaluations (Metrics, Experiments, Preliminary Findings)

- For improving the model we tried data augmentation and random sampling and verified the model using using Logistic Regression.

## Original Dataset

Accuracy: 0.8  
Precision: 0.76  
Recall: 0.63  
F1 Score: 0.67

## Over Sampled

Accuracy: 0.65  
Precision: 0.72  
Recall: 0.76  
F1 Score: 0.68

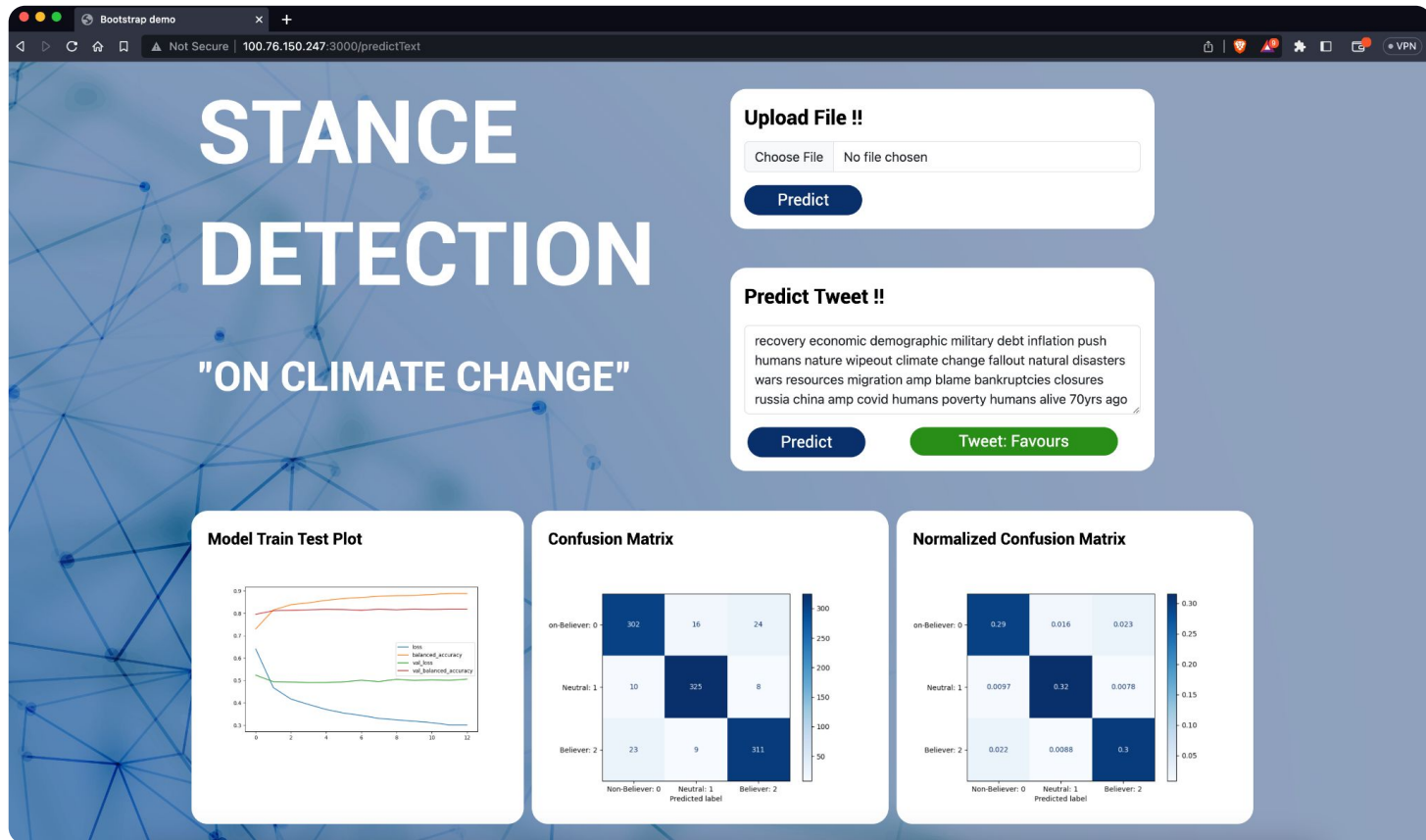
## Under Sampled

Accuracy: 0.61  
Precision: 0.71  
Recall: 0.71  
F1 Score: 0.64



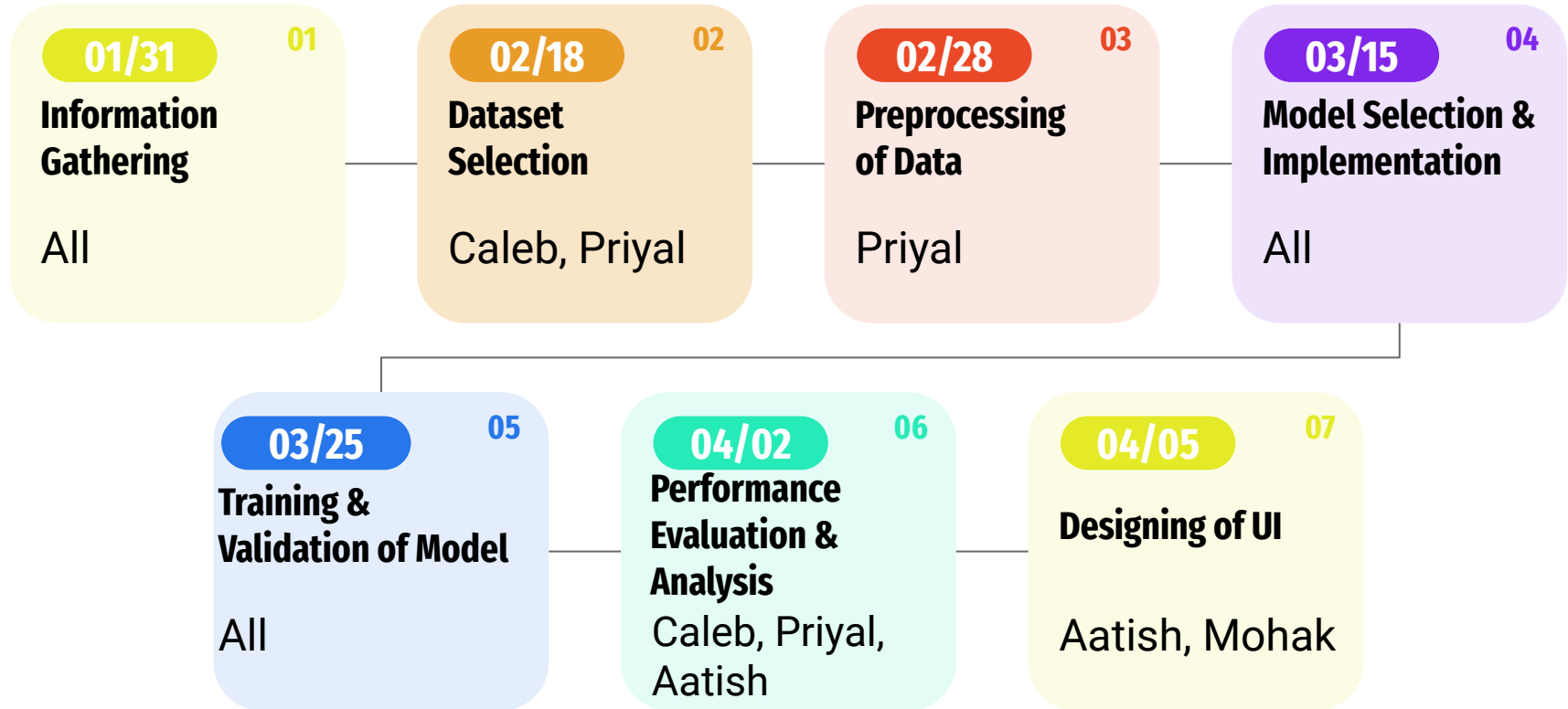


# UI Interfaces or Data Visualization





# Project Plan: Tasks, Deadline, Division of Work



**GitHub:**  
**<https://github.com/Aatish13/StanceDetection>**

**Thank You**