PROJECT

REAL-TIME SENTIMENT ANALYSIS SYSTEM

Team Members

Mina Edwar Dawood

Samaa Abdullah Elsayed

Bavly Wagdy Alfy

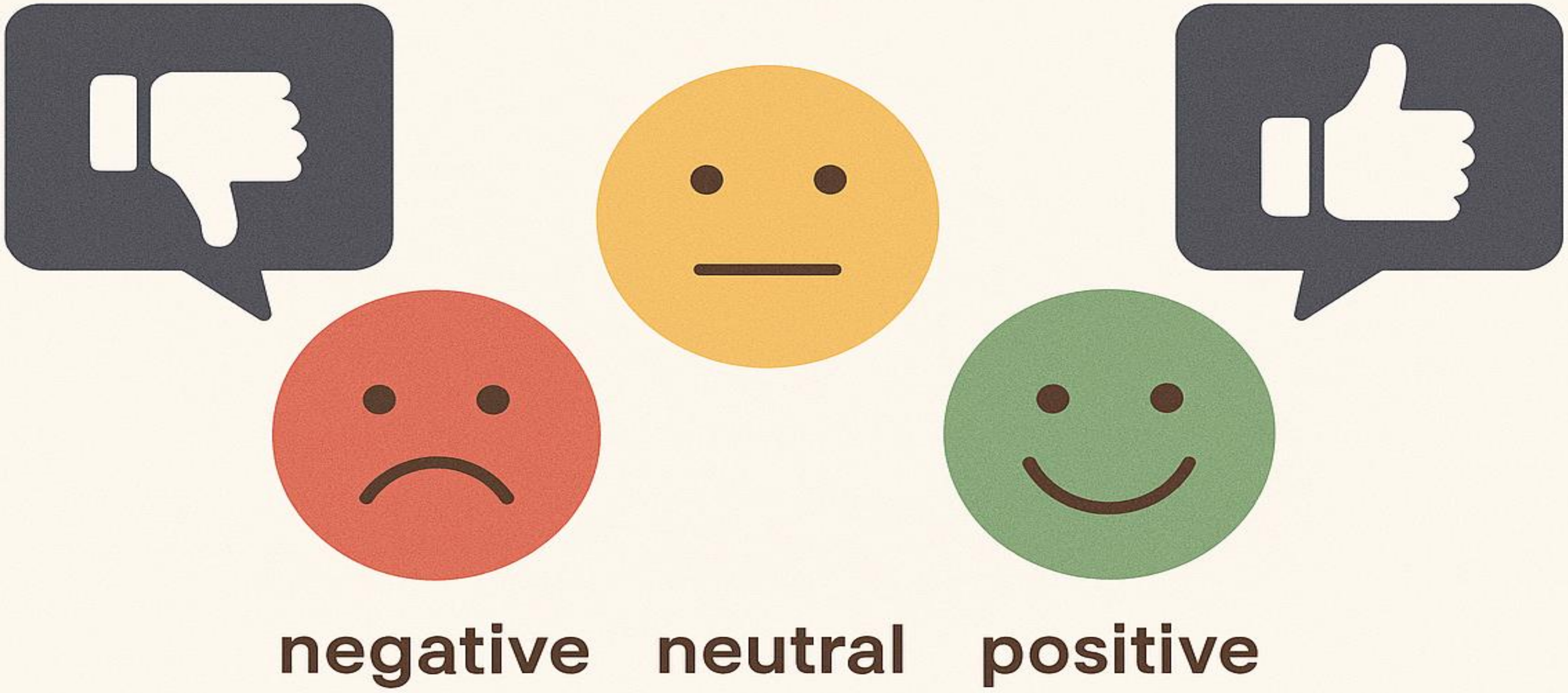
Mohamed Heshmat

Mohaned Saber

Malak Magdy

Table of Contents

Introduction



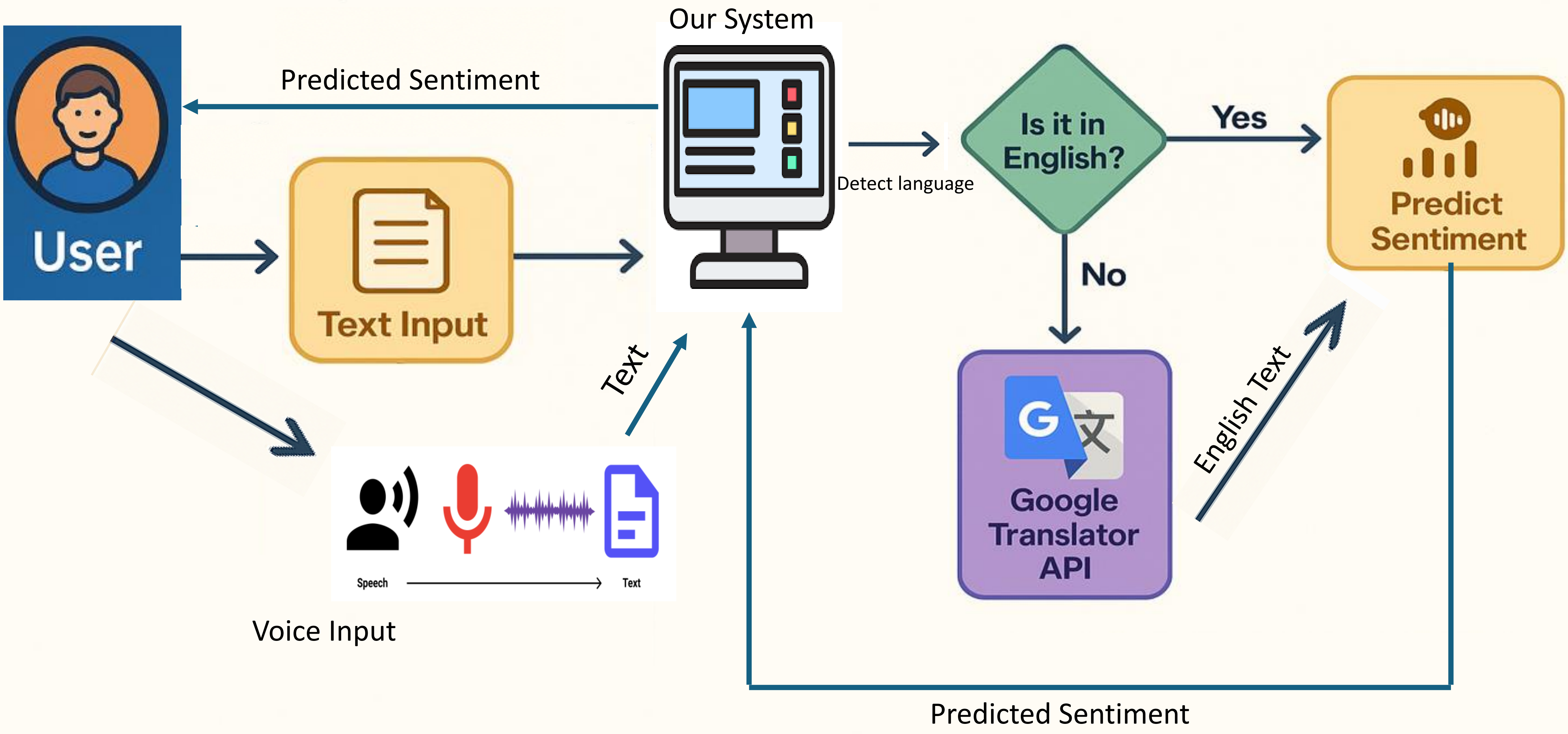
Introduction

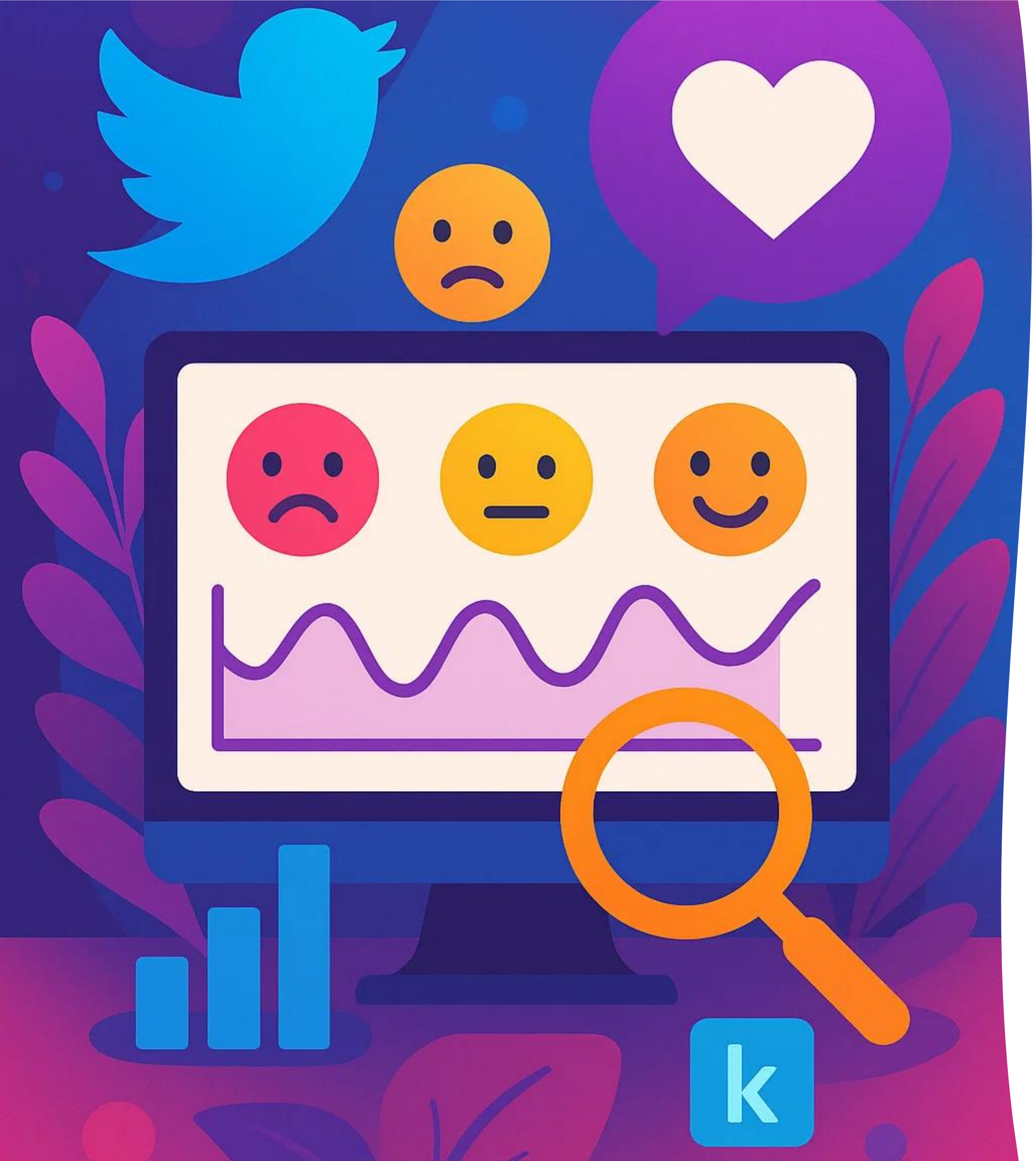
- In today's digital world, understanding public sentiment in real time is crucial for businesses, governments, and researchers.
- Over **500 million tweets are sent per day** – a goldmine of opinions and emotions.
- Existing sentiment tools often ignore **Arabic** or **spoken language inputs**.

Objectives

- Develop a system that predicts sentiment from text or voice inputs in Arabic or English.
- Train the system using Twitter and review datasets for higher accuracy with different models.
- Use translation APIs and speech-to-text tools to process non-English and voice inputs.
- Deliver real-time predictions with an easy-to-use interface.

System Architecture



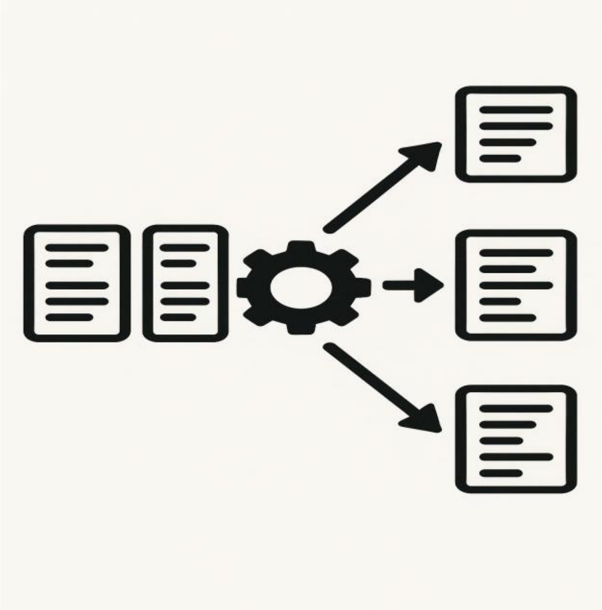


Data Collection

- Data was collected from Kaggle.
- We collected two datasets the first was **Twitter Sentiment Analysis** dataset and the second one was **Social Media-Analysis Sentiment** with overall **65269** input texts and sentiment.

Data Preprocessing

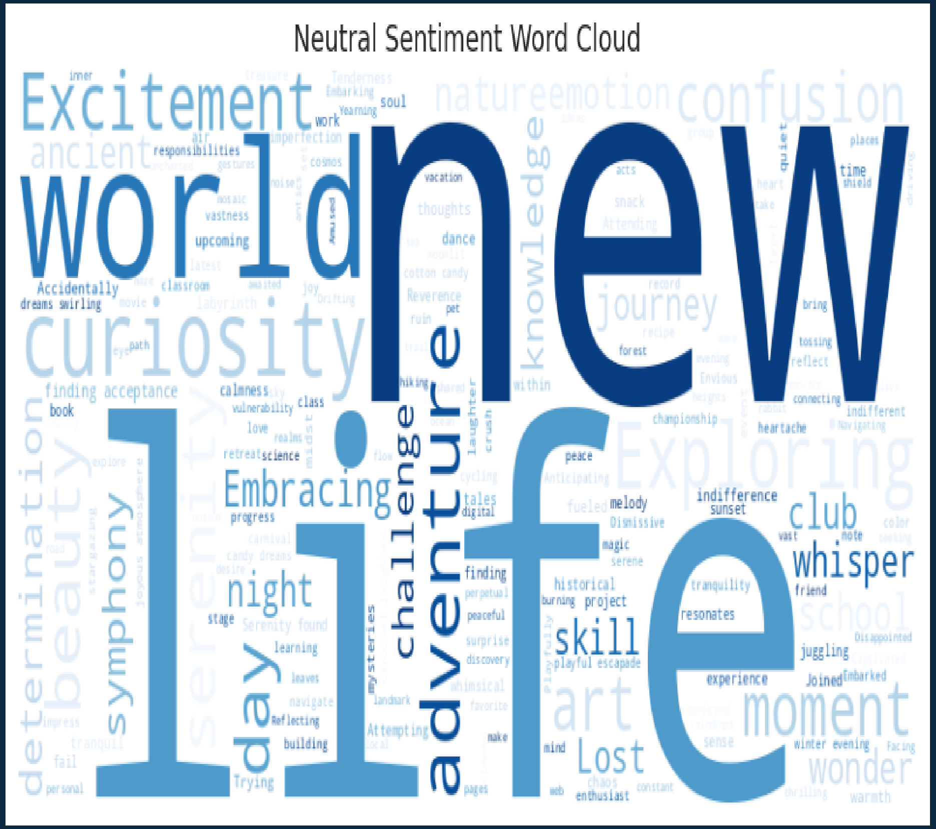
Dataset / Operation	Dataset (1) Twitter Sentiment Analysis	Dataset (2) Social Media-Analysis Sentiment
Removed Nulls & Duplicates	✔	✔
Labels fixed	Mapped some words first to Negative-Neutral-Positive then mapped **Negative: 0 Neutral: 1 Positive: 2**	Mapped **Negative: 0 Neutral: 1 Positive: 2**
Text Processing	Removed: hashtags (#), URLs, emojis, ,@, &, * ,then Lowercase, Tokenize, Stop Words Removal, Lemmatization.	Removed: hashtags (#), URLs, emojis, ,@, &, * ,then Lowercase, Tokenize, Stop Words Removal, Lemmatization.



- After Preprocessing each data alone we combined them into fully dataset with **65269** text inputs.
- We started tokenizing and padding sequences for the combined data with **vocab_size = 30000** and **max_len = 100.**
- We used pre-trained word embeddings **(GLoVe)** to form our embedding layer.
- We finally split our data **87% train,10% valid and 3% test.**

Words Cloud

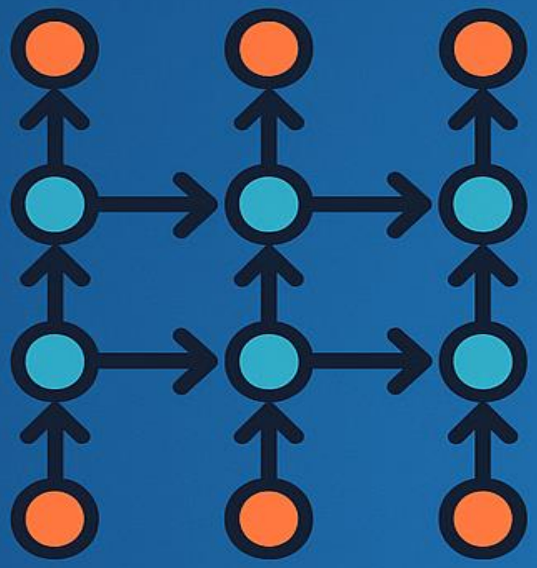

Positive Sentiment Word Cloud
new
joy
friend
moment
nature
gratitude
challenging
life
challenge
creativity
laughter
concert
contentment
community
every
Feeling
art
old
city
beauty
nostalgia
garden
day
time
dance
happiness
memories
kindness
enthusiasm
inspiration
dreams
journey
book
love
Exploring
sky
cup


Neutral Sentiment Word Cloud
new
life
world
Excitement
curiosity
confusion
ancient
adventure
challenge
knowledge
Embracing
day
night
symphony
beauty
determination
journey
Exploring
moment
club
whisper
skill
art
Lost
wonder
school

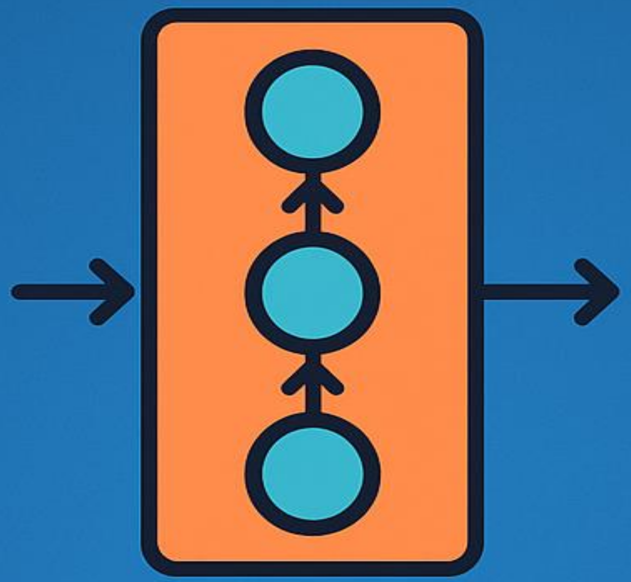

Negative Sentiment Word Cloud
despair
echoes
heart
Feeling
silent
night
emotion
life
Loneliness
grief
solitude
lost
shattered
day
away
thought
betrayal
emotional
Frustration
storm
jealousy
trust
numbness
soul
shadows

Models

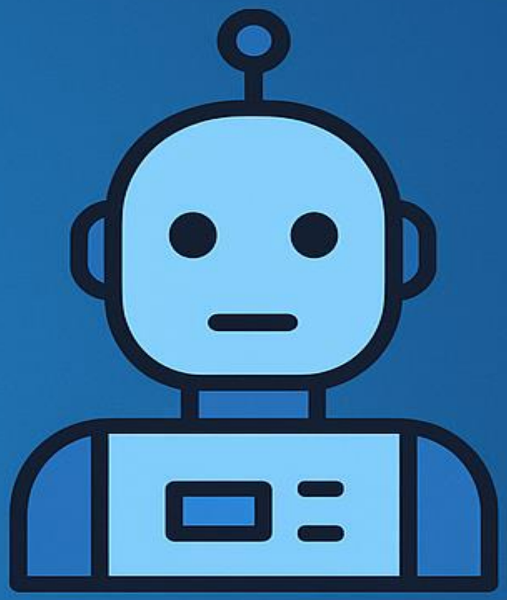
BILSTM



GRU



BERT



Model I: BILSTM Architecture

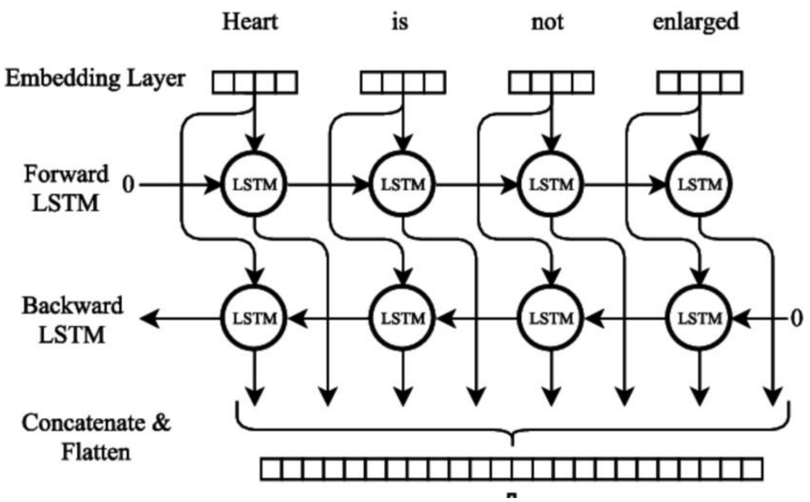
Model: "sequential"

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 100, 300)	9,000,000
bidirectional (Bidirectional)	(None, 100, 512)	1,140,736
global_average_pooling1d (GlobalAveragePooling1D)	(None, 512)	0
dense (Dense)	(None, 128)	65,664
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 64)	8,256
dropout_1 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 3)	195

Total params: 10,214,851 (38.97 MB)

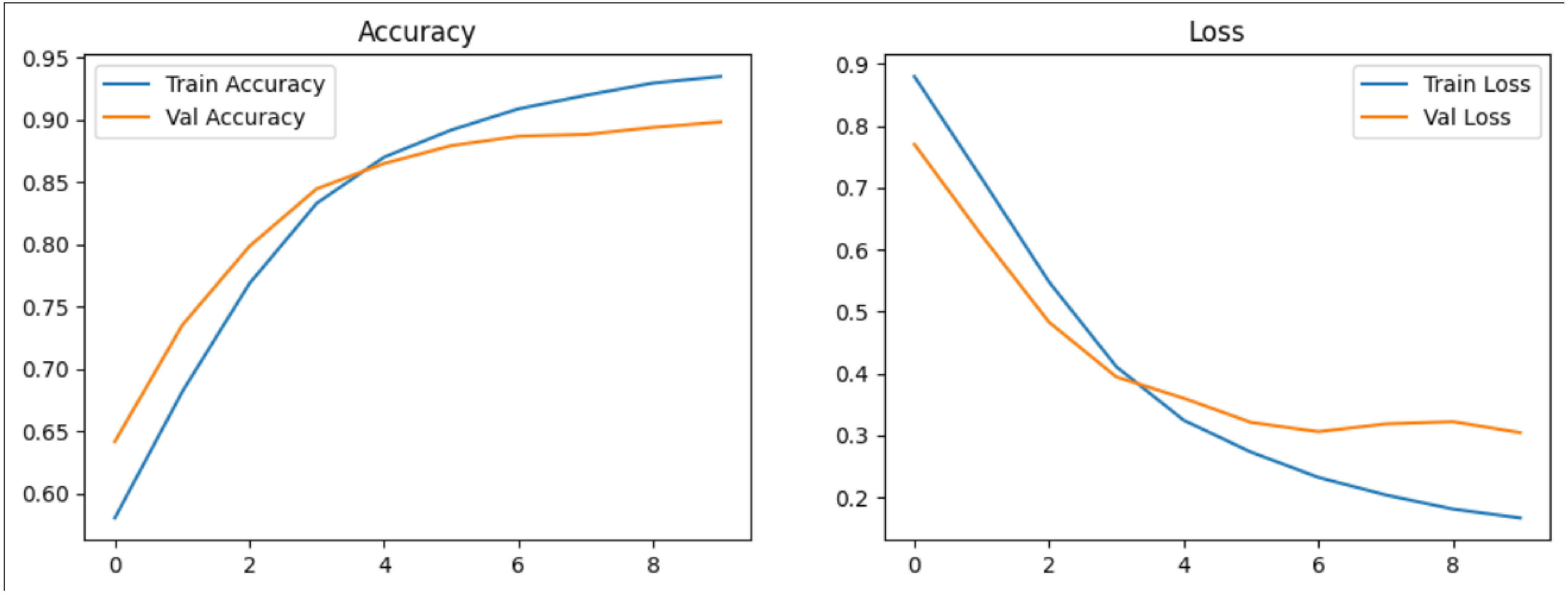
Trainable params: 1,214,851 (4.63 MB)

Non-trainable params: 9,000,000 (34.33 MB)



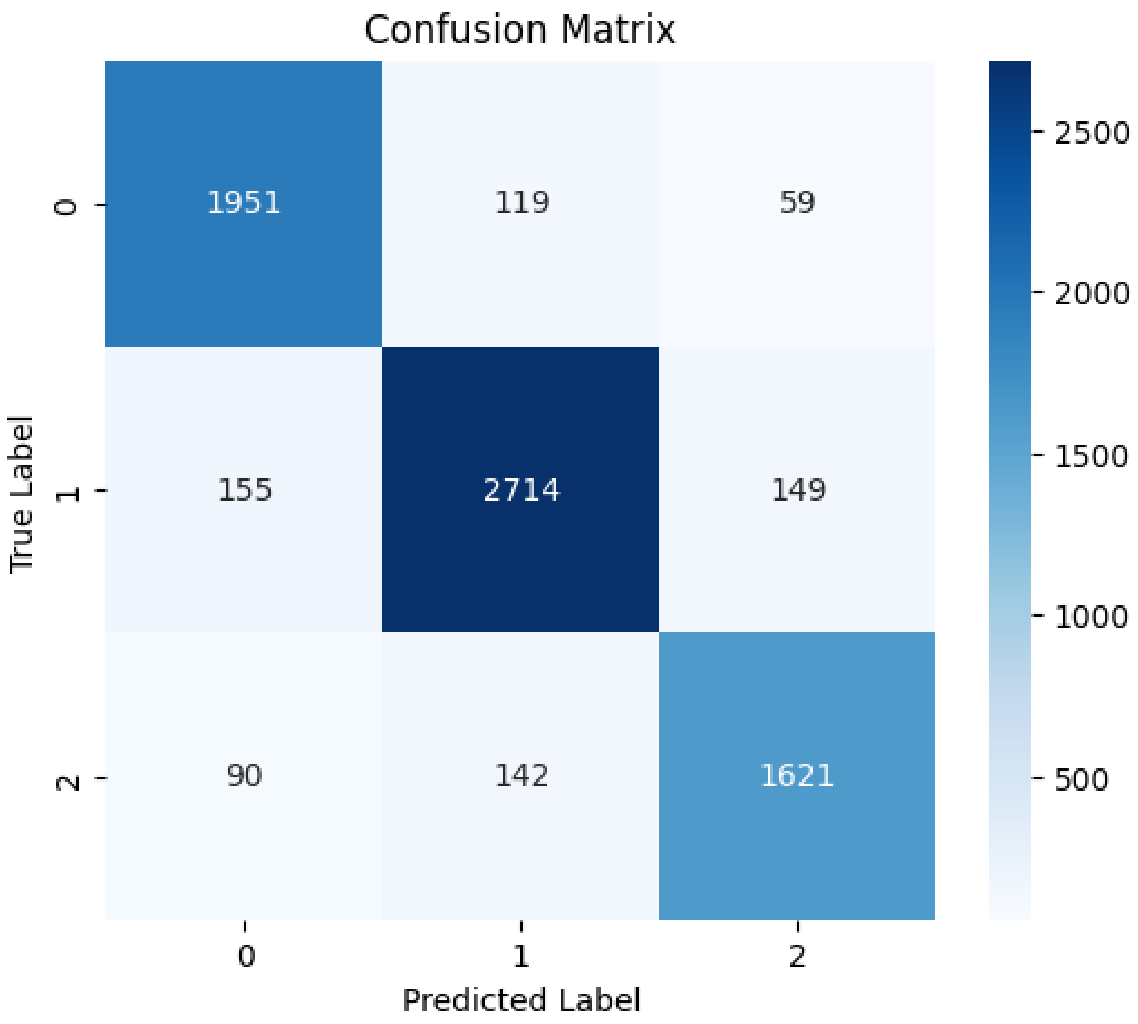
BILSTM: Training

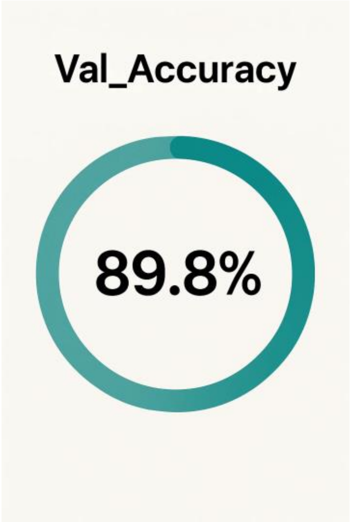
- The model was trained for 10 epochs.




```
Epoch 9/10
1821/1821 ━━━━━━━━━━━━━━━━━━━━ 1030s 565ms/step - accuracy: 0.9310 - loss: 0.1762 - val_accuracy: 0.8937 - val_loss: 0.3221
Epoch 10/10
1821/1821 ━━━━━━━━━━━━━━━━━━━━ 1039s 564ms/step - accuracy: 0.9366 - loss: 0.1611 - val_accuracy: 0.8980 - val_loss: 0.3043
```

BILSTM: Evaluation





Classification Report:

	precision	recall	f1-score	support
0	0.89	0.92	0.90	2129
1	0.91	0.90	0.91	3018
2	0.89	0.87	0.88	1853
accuracy			0.90	7000
macro avg	0.90	0.90	0.90	7000
weighted avg	0.90	0.90	0.90	7000

Model II: GRU Architecture

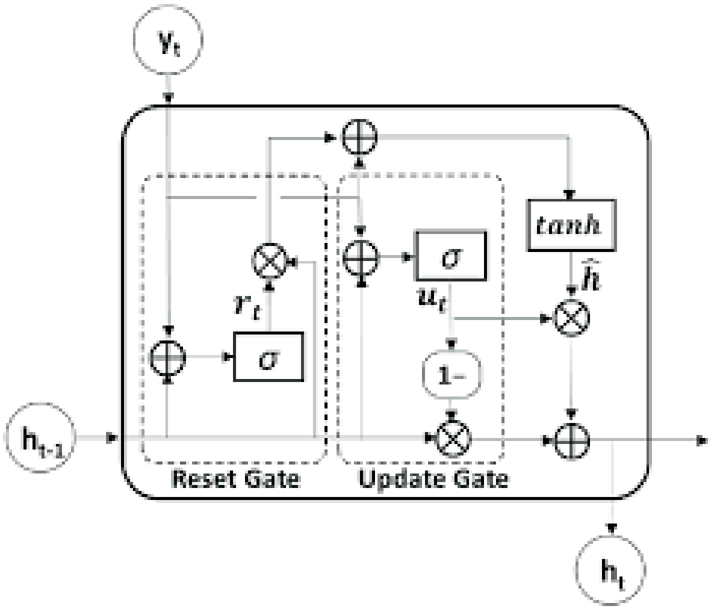
Model: "sequential"

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 100, 300)	9,000,000
gru (GRU)	(None, 100, 256)	428,544
global_average_pooling1d (GlobalAveragePooling1D)	(None, 256)	0
dropout (Dropout)	(None, 256)	0
dense (Dense)	(None, 64)	16,448
dropout_1 (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 3)	195

Total params: 9,445,187 (36.03 MB)

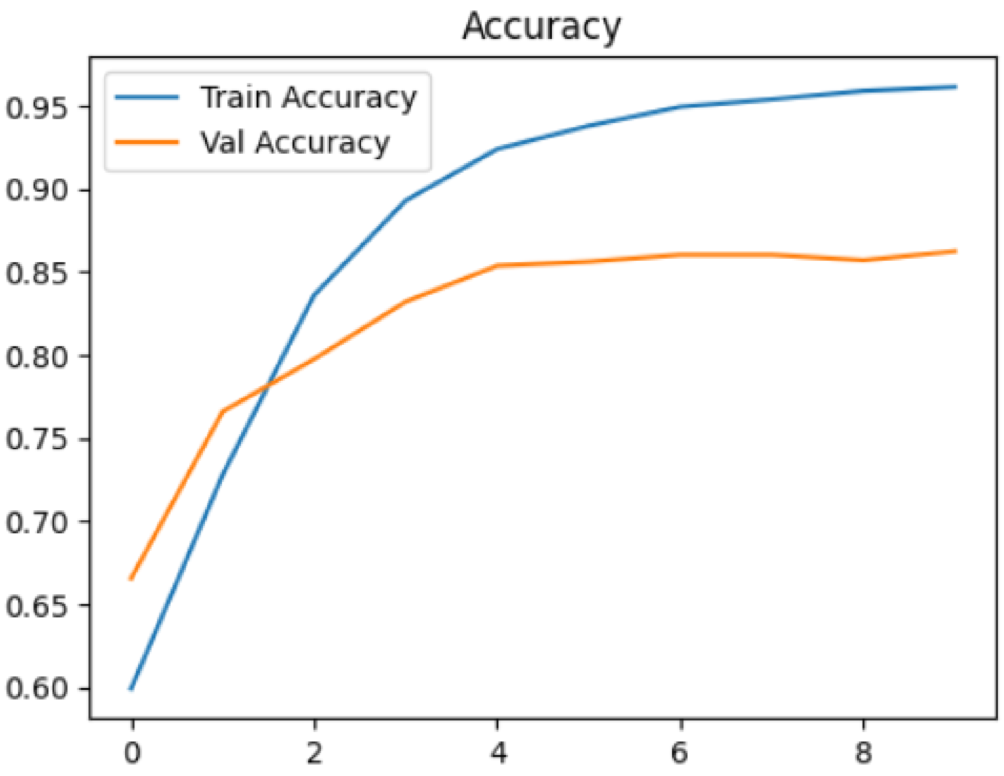
Trainable params: 445,187 (1.70 MB)

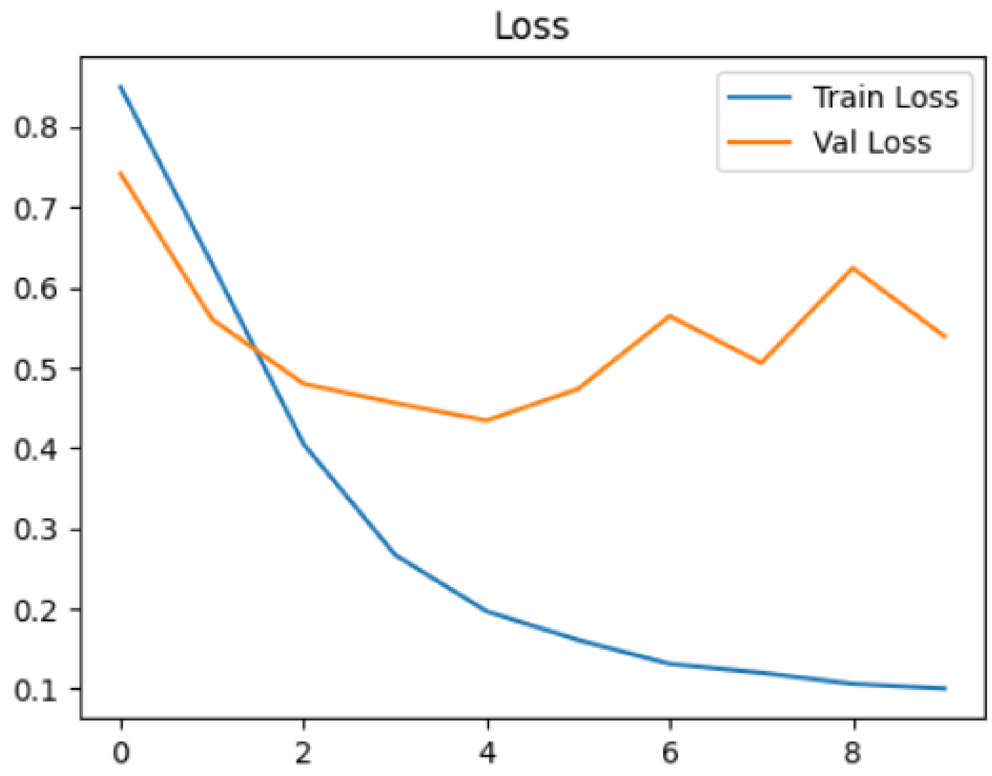
Non-trainable params: 9,000,000 (34.33 MB)



GRU: Training

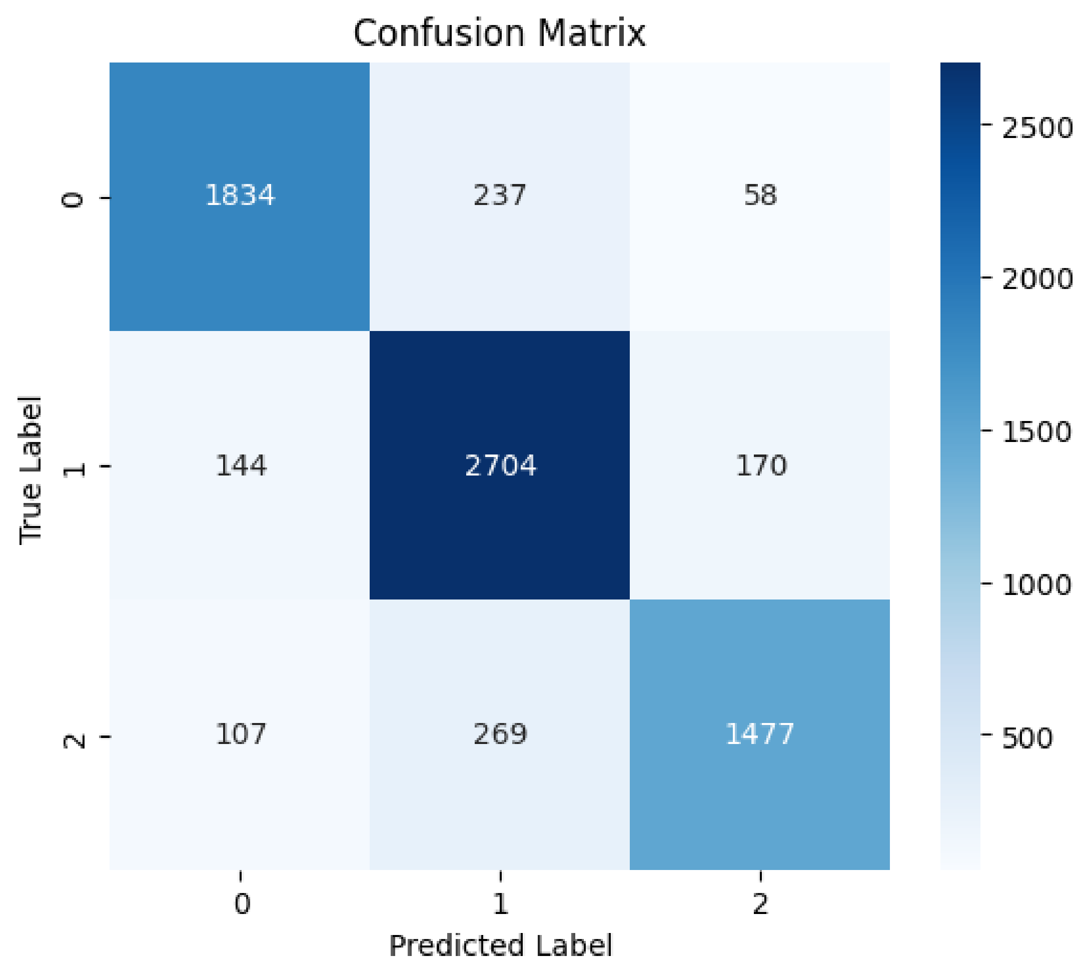
- The model was trained for 10 epochs.

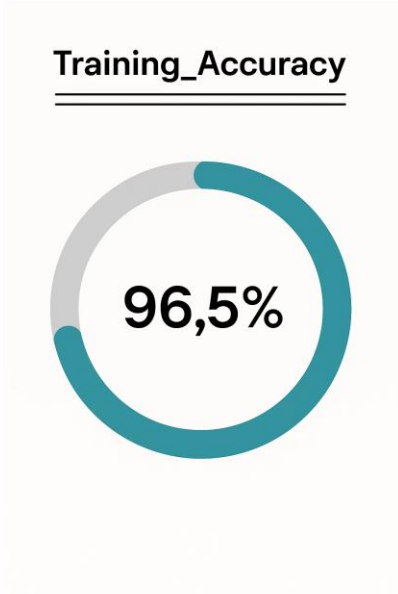


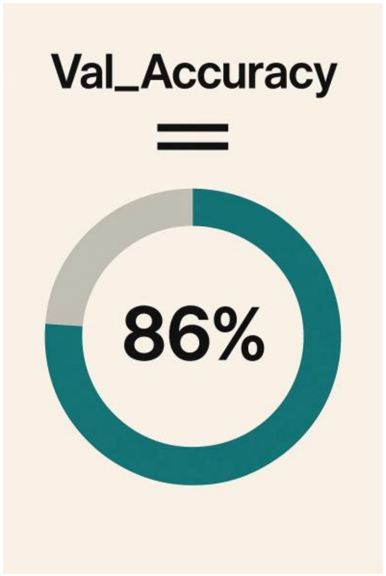



```
Epoch 9/10
1821/1821 ━━━━━━━━━━━━━━━━━━━━ 21s 10ms/step - accuracy: 0.9629 - loss: 0.0978 - val_accuracy: 0.8621 - val_loss: 0.4836
Epoch 10/10
1821/1821 ━━━━━━━━━━━━━━━━━━━━ 21s 10ms/step - accuracy: 0.9650 - loss: 0.0903 - val_accuracy: 0.8593 - val_loss: 0.5185
```

GRU: Evaluation








```
Classification Report:
              precision    recall  f1-score   support

           0       0.88      0.86      0.87      2129
           1       0.84      0.90      0.87      3018
           2       0.87      0.80      0.83      1853

    accuracy                           0.86      7000
   macro avg       0.86      0.85      0.86      7000
weighted avg       0.86      0.86      0.86      7000
```

Model III: BERT

- We fine-tuned bert-base-uncased pretrained model with 3 epochs.
- The following is our configurations with 3 classes as output for Negative, Neutral and Positive.
- We also used the BERT Pre-trained Tokenizer to tokenize

```
# Model Hyperparameters
MAX_LEN = 128
BATCH_SIZE = 32
N_LAYERS = 12
D_MODEL = 768
N_HEADS = 12
D_FF = 3072
VOCAB_SIZE = 30522  # Standard BERT vocab size
NUM_CLASSES = 3  # Sentiment classes
DROPOUT = 0.1
```

BERT Training Process

```
You should probably TRAIN this model on a down-stream task to be a
Epochs:  33%|███       | 1/3 [22:49<45:38, 1369.02s/it]
Epoch 1/3 - Loss: 0.6859 - Val Accuracy: 0.8216
Epochs:  67%|██████    | 2/3 [45:36<22:47, 1367.92s/it]
Epoch 2/3 - Loss: 0.3119 - Val Accuracy: 0.8883
Epochs: 100%|██████████| 3/3 [1:08:23<00:00, 1367.90s/it]
Epoch 3/3 - Loss: 0.1665 - Val Accuracy: 0.9071
```

Bert Evaluation

Arabic: المنتج سيء جدًا وخيبة أمل كبيرة.

Translated: The product is very bad and a great disappointment.

Predicted Label: 0 Sentiment: Negative

Arabic: ما زلت أجربه، لا يمكنني الحكم الآن.

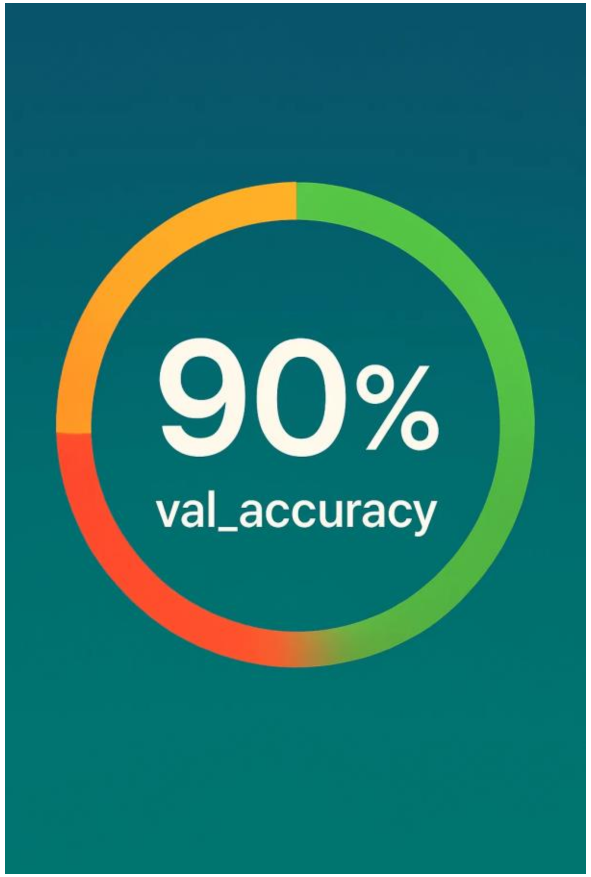
Translated: I still try it, I can't judge now.

Predicted Label: 1 Sentiment: Neutral

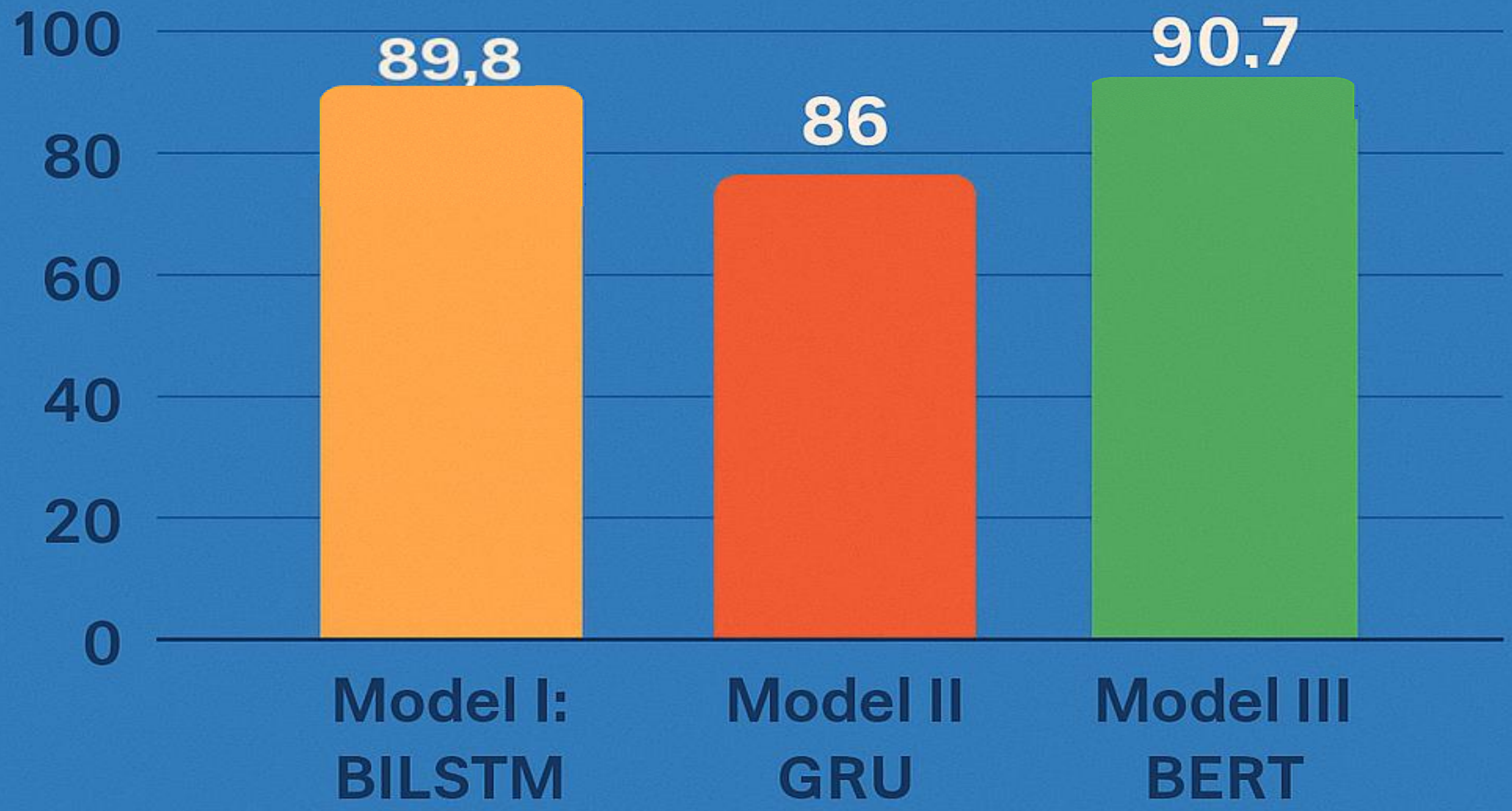
Arabic: المنتج رائع جدًا وتجاوز توقعاتي.

Translated: The product is very great and exceeds my expectations.

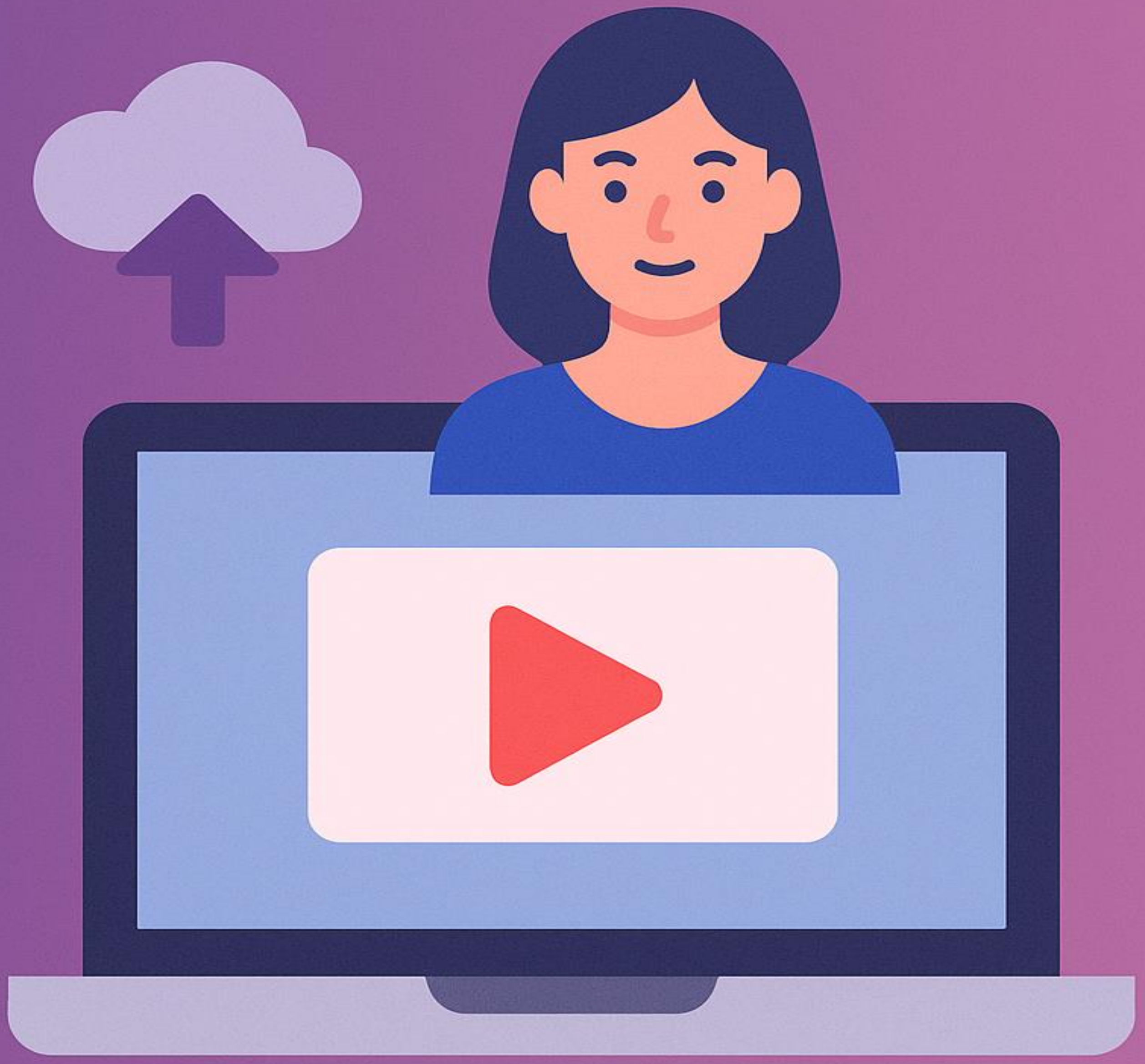
Predicted Label: 2 Sentiment: Positive



Validation Accuracy

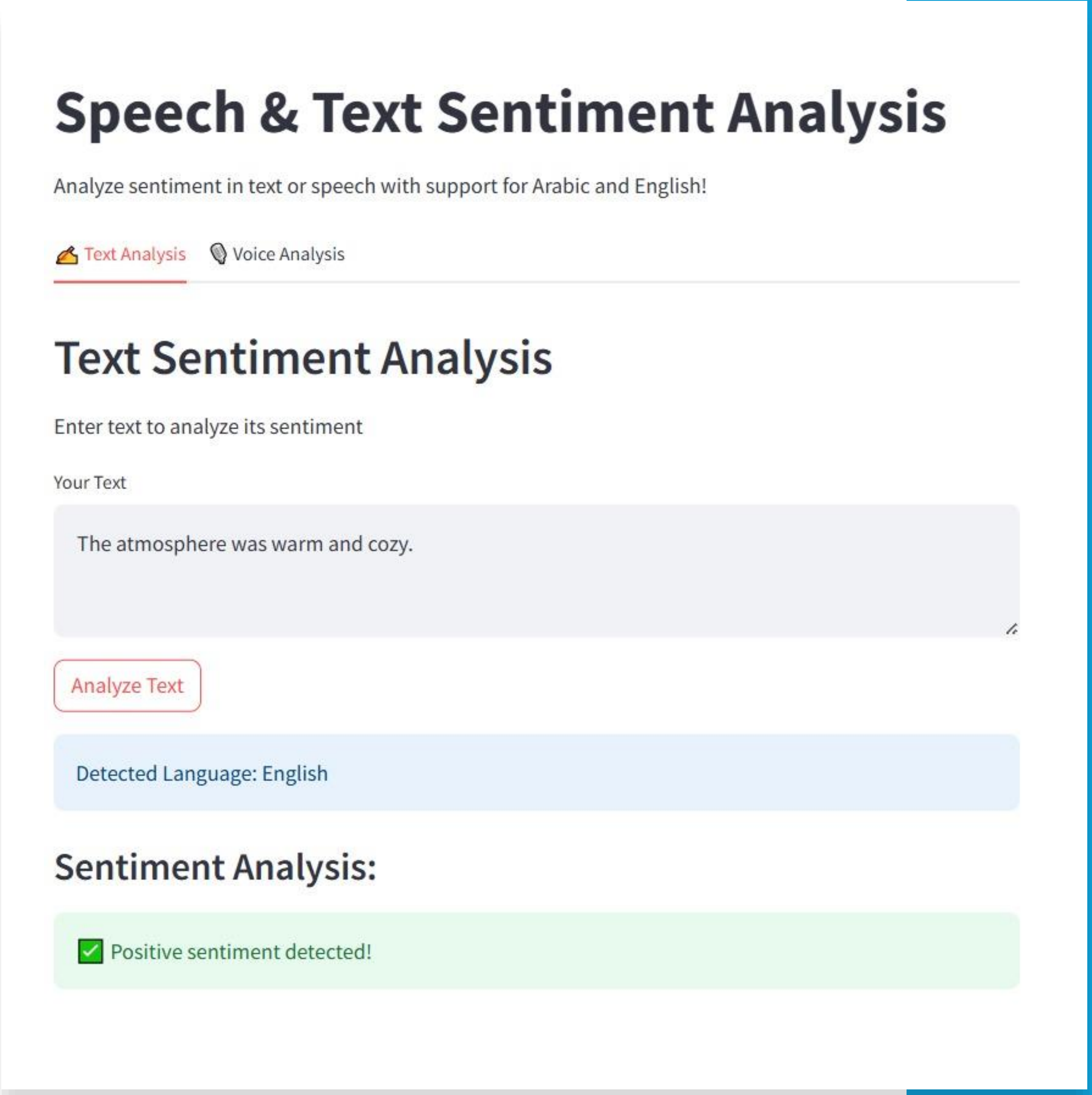


Deployment&
Demo

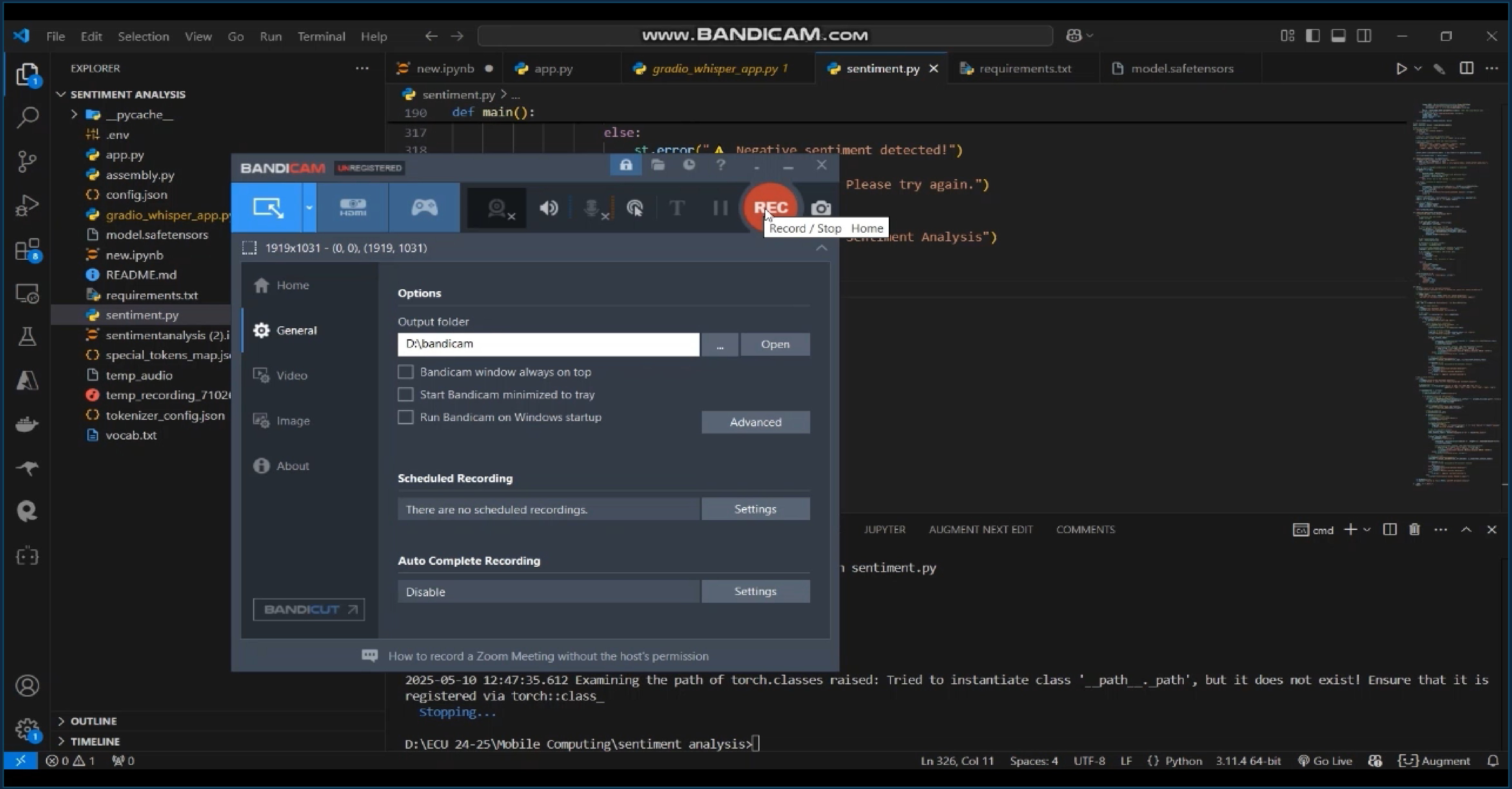


Deployment

- We used streamlit to deploy our model, we built an easy-to use interface.



DEMO



www.BANDICAM.com
EXPLORER
SENTIMENT ANALYSIS
__pycache__
.env
app.py
assembly.py
config.json
gradio_whisper_app.py
model.safetensors
new.ipynb
README.md
requirements.txt
sentiment.py
special_tokens_map.json
temp_audio
tokenizer_config.json
vocab.txt
def main():
else:
st.error("⚠ Negative sentiment detected!")
Please try again.")
BANDICAM UNREGISTERED
REC
Record / Stop Home
1919x1031 - (0, 0), (1919, 1031)
Home
General
Video
Image
About
Options
Output folder
D:\bandicam
Open
Bandicam window always on top
Start Bandicam minimized to tray
Run Bandicam on Windows startup
Advanced
Scheduled Recording
There are no scheduled recordings.
Settings
Auto Complete Recording
Disable
Settings
BANDICUT
How to record a Zoom Meeting without the host's permission
2025-05-10 12:47:35.612 Examining the path of torch.classes raised: Tried to instantiate class '__path__._path', but it does not exist! Ensure that it is registered via torch::class_
Stopping...
D:\ECU 24-25\Mobile Computing\sentiment analysis>
OUTLINE
TIMELINE
Ln 326, Col 11 Spaces: 4 UTF-8 LF Python 3.11.4 64-bit Go Live Augment

NumPy

pandas

matplotlib

Tools & Libraries

PyTorch

Hugging Face

Streamlit

scikit learn

Future Work

We will work on gathering more data of reviews for sentiment analysis

We will work on adding a computer vision feature to allow real-time sentiment analysis using the camera

Thank You!