Project Report

Project Title: Multimodal Sentiment Intelligence Platform using VC-CSA + RAG

Team Member Contributions

Member	Contribution	Percentage
Vinay	VC-CSA implementation, Streamlit integration, fusion logic	29
Shivanjali	SimpleRNN training, Text sentiment + image analyzer module	24
Apurva	Video processing module, Testing, Report writing	24
Divya	YOLOv8 + DeepFace integration, Poster and presentation creation	23

1. Introduction

In today's digital landscape, understanding user sentiment from multimodal sources (text, images, and videos) is critical for various applications including marketing, recommendation systems, and social media analytics. Traditional sentiment analysis systems fall short when it comes to combining multiple modalities to derive context-aware sentiments.

Our project presents a unified **Multimodal Sentiment Intelligence Platform** that integrates advanced AI models such as VC-CSA, GPT-2, DistilBERT, YOLOv8, DeepFace, and SimpleRNN to deliver robust sentiment analysis across various data types.

2. Problem Statement

- Sentiment from text alone cannot capture the full spectrum of user opinions.

- Visual and contextual cues play a key role in understanding true sentiment.

- There is a need to induce sentiment from video context and user comments (Induced Sentiment).

3. Solution Overview

The platform supports analyzing and fusing sentiments from:

- Text → Using SimpleRNN (IMDB trained model)

- Text Explanation → Using GPT-2 based RAG explanation

- Image → Using YOLOv8 (object detection) + DeepFace (emotion detection)

- Video → Using YOLOv8 + DeepFace for frame-by-frame analysis

- Video + Comment (VC-CSA) → Induced sentiment prediction using our PyTorch based VC-CSA model

4. Architecture and Modules

Text Encoder

- DistilBERT model to extract text features

Image Sentiment Analyzer

- YOLOv8 → Object detection
- DeepFace → Facial emotion detection

Video Sentiment Analyzer

- YOLOv8 + DeepFace → Frame level object and facial emotion extraction
- Aggregation → Dominant objects and emotion prediction

Simple RNN Sentiment Analyzer

- Pre-trained SimpleRNN model trained on IMDB dataset to classify text sentiment

VC-CSA Based Induced Sentiment Module

- Video Encoder → CNN-based temporal feature extractor
- Text Encoder → DistilBERT embeddings
- Fusion Classifier → Predict opinion and emotion from combined features

Fusion Strategy

- Rule-based mapping for image/text sentiment fusion
- VC-CSA based deep fusion for video/comment induced sentiment prediction
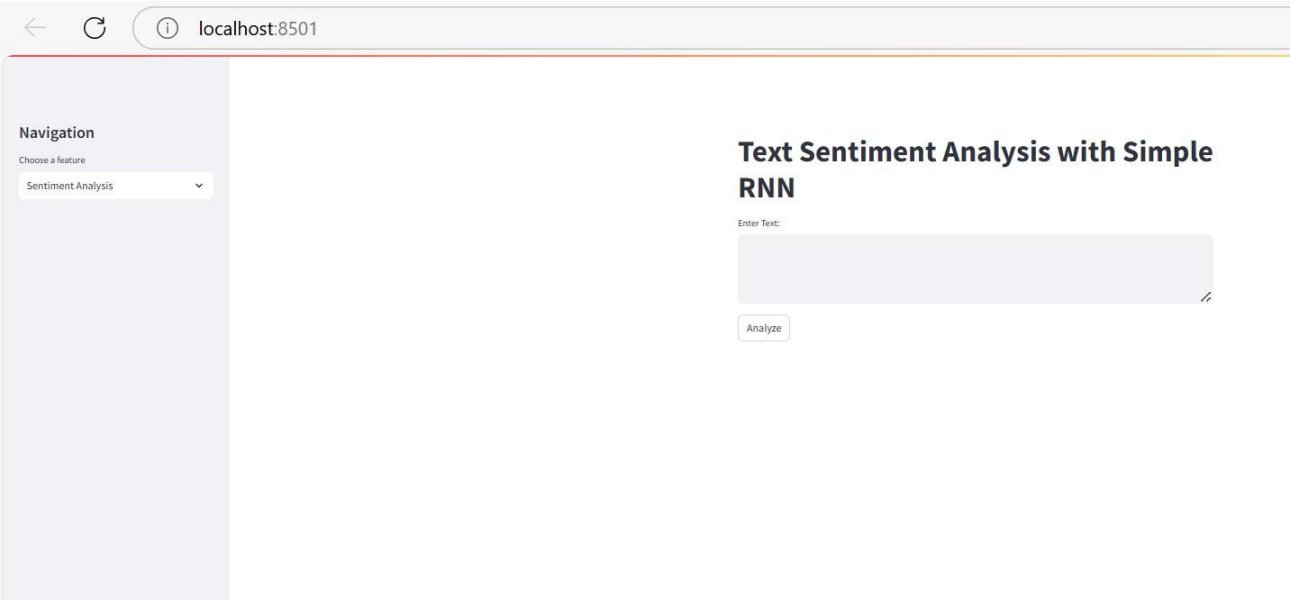
5. Implementation Details

- Developed complete solution using Python and Streamlit
- Backend Models:
    - Tensorflow + Keras (SimpleRNN)
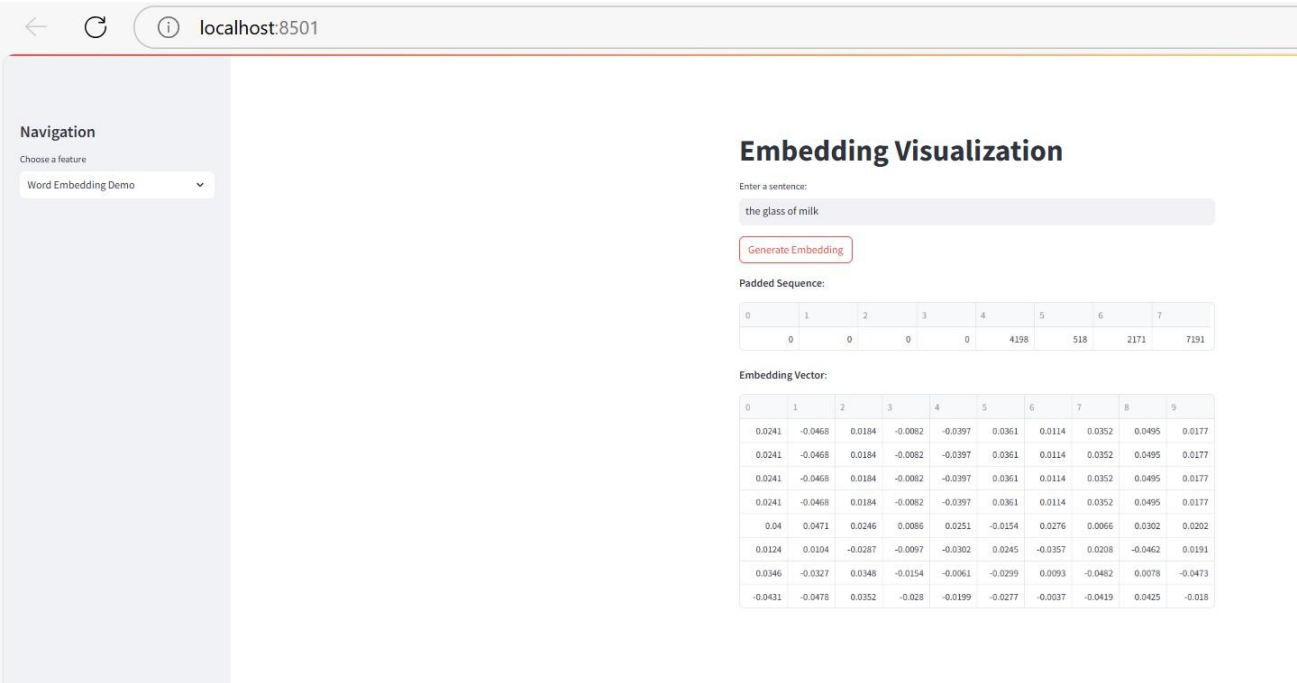   - PyTorch (VC-CSA, DistilBERT)
    - YOLOv8 (Ultralytics)

- o DeepFace (Facial analysis)

- Streamlit app supports file uploads and interactive sentiment analysis

- All models cached using Streamlit caching for performance

Results and Demonstration

Text Sentiment Analysis

Word Embedding Visualization

Sentiment Explanation with GPT-2

Video-Based Sentiment Analysis

VC-CSA Based Induced Sentiment Analysis

Navigation

Choose a feature

VC-CSA Sentiment Analysis

VC-CSA Based Induced Sentiment Analysis

Upload a Video File (.mp4)

⤴ Drag and drop file here
Limit 200MB per file • MP4, MPEG4

Browse files

📄 istockphoto-1424256178-640_adpp_is_h.mp4 454.0KB ✕

Enter Viewer Comment:

Amazing news, couldn't be happier for you

Analyze Sentiment (VC-CSA)

Predicted Opinion: Positive

Predicted Emotion: Anger

7. Challenges Faced

- Managing GPU/CPU memory due to multimodal model usage
- Aligning video frame representations for VC-CSA model
- Integration of Tensorflow + PyTorch + OpenCV + Streamlit
- Streamlit limitations for concurrent heavy model usage

8. Future Work

- Incorporate Audio modality for true multimodal sentiment analysis
- Train VC-CSA with more advanced video transformers
- Move solution to cloud for scalable sentiment analytics service
- Improve fusion strategy with learned fusion techniques (MLP or attention based)

9. Resources

- GitHub Repository: GitHub - VinayAlli/BDA_NeurIps
- Pretrained Models: SimpleRNN, DistilBERT, GPT-2, YOLOv8, DeepFace
- Project Demo Video: https://www.youtube.com/watch?v=HgDqGmlCjos

10. Conclusion

Our project demonstrates a successful implementation of a multimodal sentiment analysis platform that combines text, image, and video to predict context-aware sentiment. By leveraging advanced AI models and multimodal fusion techniques, our system provides valuable insights beyond traditional single-modal sentiment analysis.

This project lays the foundation for scalable multimodal intelligence platforms in real-world applications such as social media monitoring, product review analysis, and customer feedback systems.