Multimodal Sentiment Intelligence Platform for Dynamic Market Insights

Computer Science

Vinay Viswanadh Alli, Apurva Vadlakonda, Shivanjali Dyapa and Divya Mishra

UNIVERSITY OF MISSOURI

Context

The **Multimodal Sentiment Intelligence Platform** integrates text, visual, and contextual data to deliver real-time market insights. It leverages advanced models such as **BERT and GPT-2** for text sentiment analysis, **YOLOv8 and DeepFace** for visual sentiment detection, and advanced **fusion strategies** (feature-level, decision-level, and VC-CSA-based fusion) to improve sentiment prediction accuracy. For robust training and evaluation, the platform utilizes the **CMU-MOSEI dataset**, enabling multimodal sentiment understanding across diverse scenarios. The system is designed for seamless deployment using **Streamlit**, **FastAPI**, or **cloud platforms**, and is targeted at enabling actionable business intelligence in domains such as marketing, finance, customer service, and social media monitoring.

Background

The Multimodal Sentiment Intelligence Platform combines text, visual, and contextual analysis to deliver accurate, real-time market insights. Leveraging BERT, GPT-2, YOLOv8, DeepFace, and RAG, along with advanced fusion strategies and CMU-MOSEI training, it captures human sentiment across modalities. Designed for deployment via Streamlit, FastAPI, or cloud platforms, it empowers industries like marketing, finance, and customer service with actionable sentiment intelligence. This holistic sentiment understanding enables deeper insights into user emotions, opinions, and intentions. The system's flexible architecture allows easy deployment and scaling, making it adaptable for enterprise applications ranging from brand monitoring and customer feedback analysis to financial market sentiment tracking and social media intelligence.

Dataset

We utilize multiple datasets to train and evaluate our multimodal sentiment analysis model:

1. **CMU-MOSEI** (Carnegie Mellon Multimodal Opinion Sentiment and Emotion Intensity)
   Largest dataset for **sentence-level multimodal sentiment analysis.** Contains **over 12 hours of annotated videos** from 1000+ speakers across 250 topics. Provides **text, visual, and audio modalities,** enabling robust sentiment recognition.

2. **Amazon & IMDb Reviews** (For Text Sentiment Analysis)
   Large-scale textual datasets with labeled **positive, neutral, and negative sentiment.** Used to fine-tune **BERT and GPT-2** for sentiment classification.

3. **AffectNet / FER-2013** (For Facial Emotion Recognition)
   Large datasets containing images labeled with **facial expressions** (happy, sad, angry, etc.). Helps train **DeepFace for emotion detection** in video sentiment analysis.

Reference

[1]Das, Ringki, and Thoudam Doren Singh. "Multimodal sentiment analysis: a survey of methods, trends, and challenges." ACM Computing Surveys 55.13s (2023): 1-38. [2] Miah, Md Saef Ullah, et al. "A multimodal approach to cross-lingual sentiment analysis with ensemble of transformer and LLM." Scientific Reports 14.1 (2024): 9603. [3] Liu, Zhicheng, et al. "Ensemble pretrained models for multimodal sentiment analysis using textual and video data fusion." Companion Proceedings of the ACM Web Conference 2024. 2024. [4] Alam, Md Jahangir, et al. "Advancements in Multimodal Social Media Post Summarization: Integrating GPT-4 for Enhanced Understanding." 2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC). IEEE, 2024.

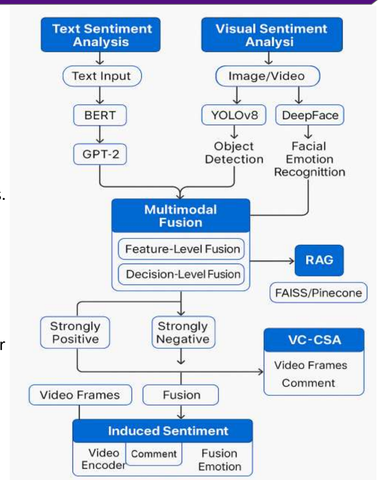
Multi-modal Methodology

Text Sentiment Analysis
- Utilizes **BERT** for sentiment classification and **GPT-2** for explanatory text generation.
- Trained on **IMDb**, **Amazon Reviews**, and **CMU-MOSEI** datasets.
- Predicts sentiment polarity as **Positive** or **Negative** based on textual input.

Visual Sentiment Analysis
- Employs **YOLOv8** for object detection and **DeepFace** for facial emotion recognition.
- Processes images and videos to identify objects and dominant facial emotions.
- Trained on datasets including **CMU-MOSEI**, **AffectNet**, and **FER-2013**.

Multimodal Fusion
- **Feature-Level Fusion**: Integrates embeddings from both text and visual modalities.
- **Decision-Level Fusion**: Combines independent predictions via ensemble techniques.
- Produces sentiment outcomes such as **Strongly Positive**, **Strongly Negative**, or **Mixed Sentiment**.

Retrieval-Augmented Generation (RAG)
- Leverages **FAISS/Pinecone** for retrieving relevant external context.
- Enhances sentiment classification through contextual enrichment.



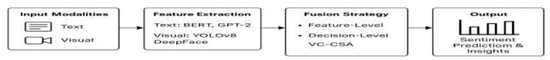
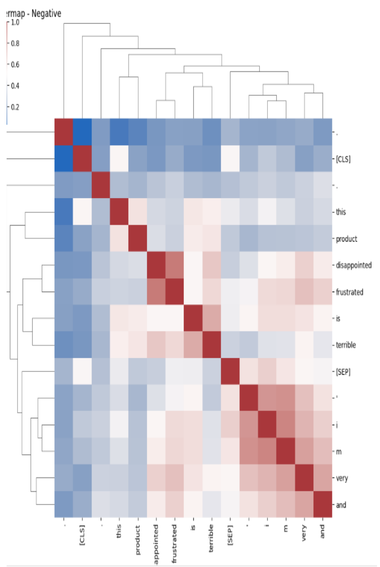
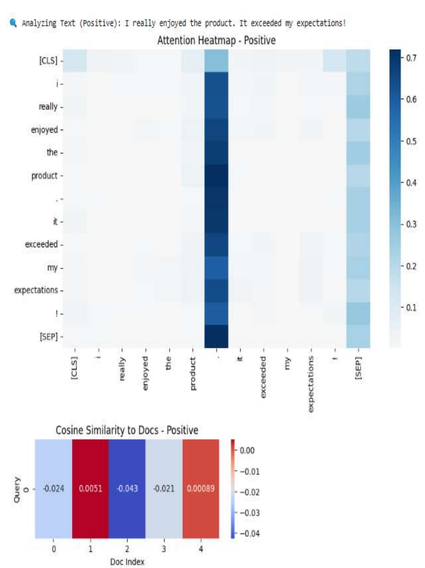
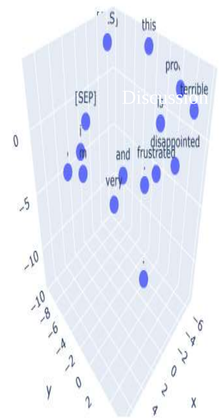


Fig 1 : Multimodal Sentiment Intelligence Workflow: From Input (Text, Image, Video) to Processing (AI Models & Fusion) to Final Sentiment Analysis Display on Web Interface.

Results & Graphs







Visualization

Our platform provides a **real-time sentiment analysis dashboard**, integrating insights from text and visual data. The results are displayed through an interactive web interface, showcasing sentiment trends, classifications, and feature-based insights.

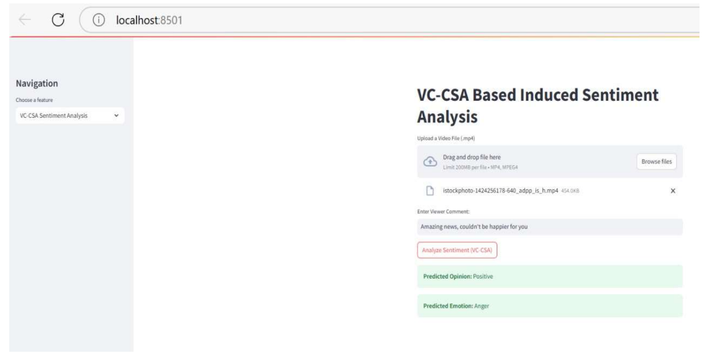


Fig 2: Real-time Sentiment Analysis Dashboard: A web-based interface displaying sentiment predictions from text and visual data.

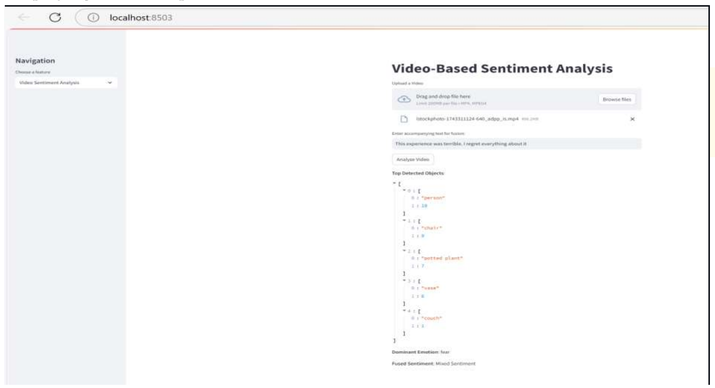


Fig 3: Sentiment Analysis Output: Predicted sentiment classifications with insights derived from multimodal AI processing.

Discussion & Future Work

The **Multimodal Sentiment Intelligence Platform** seamlessly integrates text, visual, and contextual data to significantly enhance sentiment analysis accuracy. By leveraging advanced models like **BERT**, **GPT-2**, **YOLOv8**, and **DeepFace**, the system extracts meaningful insights across modalities, offering a more comprehensive view of sentiment dynamics.

Through **feature-level** and **decision-level fusion**, the platform achieves improved prediction reliability, positioning itself as a valuable tool across industries such as marketing, finance, and customer service. However, challenges remain. Limitations such as sarcasm misinterpretation, low-quality inputs, and inherent data biases impact overall accuracy.

Ultimately, the platform aims to evolve into a scalable solution for **social media monitoring**, **brand sentiment analysis**, and **financial forecasting**, enabling deeper and more actionable sentiment-driven insights.