MULTIMODAL SENTIMENTAL ANALYSIS USING DEEP LEARNING FUSION TECHNOLOGIES AND TRANSFORMERS

Project Reference No.: 48S_BE_0054

College : VSM'S Somashekhar R. Kothiwale Institute Of Technology Nipani
Branch : Department of Computer Science And Engineering
Guide(s) : Prof. Rahul Palakar
Prof. Anup Ganji
Student(s) : Ms. Pranita Kulkarni
Ms. Shreya Patil
Ms. Samruddhi Khot
Ms. Pratiksha Bhatale

Keywords: Data science, Cyber Security, Computing, Pattern Recognition and Image Processing, Signal Processing

Introduction:

The introduction provides a high-level overview of sentiment detection, focusing on the importance of understanding human emotions across various media. Traditionally, sentiment analysis has been applied to text data (e.g., social media, customer reviews), but as multimedia content becomes more prevalent, there is a growing need to interpret sentiment across different formats, such as audio, images, and documents. This project will leverage Django and SQLite to create an integrated system for detecting and analyzing sentiment in multiple media types. The introduction should also set the stage for why this type of analysis is valuable across industries, from customer service to mental health.

Objectives:

1. **Create a Multi-Modal Sentiment Detection System**

   This involves training or integrating specialized models for each modality and ensuring they work in harmony. The results from these models will then be combined (or "fused") for an overall sentiment analysis.

2. **Integrate Support for Different Media Types: Text, Audio, Image, and DOCX**

   This will involve selecting or training separate sentiment detection models for each data type, with preprocessing steps to format each input type for analysis.

3. **Develop a User-Friendly Web Interface**

   Using Django for backend management and front-end frameworks like Bootstrap or Vue.js (if desired) to create an accessible and intuitive design, enabling file uploads, processing, and displaying analysis outputs.

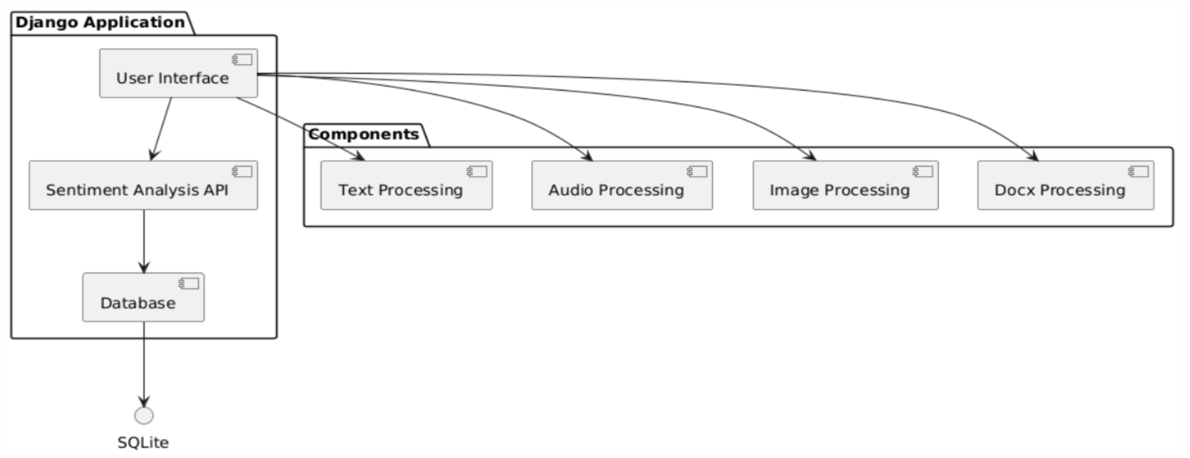
4. **Store and Manage User Data in SQLite**

   In Django, SQLite will act as the backend database to persist user data. Django models will represent and organize data in the database, and views will handle data retrieval and display on the front end.

5. **Display Detailed Results and Insights on Sentiment Analysis**

   Sentiment scores and insights will be displayed in a detailed report format on the interface, using charts, graphs, or text summaries. Libraries like Chart.js or D3.js can be used for dynamic data visualization.

Methodology:



1. **Data Collection**
    - **Goal**: Gather inputs from users in multiple formats: text, audio, image, and DOCX.
    - **Process**: The web interface will provide options for users to upload files in their chosen format, capturing the input for further analysis.
2. **Preprocessing**
    - **Goal**: Prepare each type of data for sentiment analysis by applying specific preprocessing steps.
    - **Process**:
        - **Text**: Tokenize, remove stop words, and standardize the text.
        - **Audio**: Extract audio features like MFCCs, pitch, and tone using Librosa.
        - **Image**: Detect and crop faces from images using OpenCV to focus on facial expressions.
        - **DOCX**: Extract text content and tokenize it for analysis.
3. **Sentiment Detection**
    - **Goal**: Use respective models to detect sentiment based on preprocessed data.
    - **Process**: Each media type is analyzed using its dedicated model. For instance, text is analyzed using NLTK, audio with a TensorFlow model, images with CNNs, and DOCX files with NLP processing.
4. **Data Storage**
    - **Goal**: Save results in SQLite for efficient storage and retrieval.
    - **Process**: Analysis results (including user input and sentiment scores) are saved in an SQLite database, making it easy for users to access past results.
5. **Output**
    - **Goal**: Display sentiment analysis results on a web dashboard.

SOFTWARE USED:

- Django, Python, SQLite
- TensorFlow, OpenCV, NLTK, SpeechRecognition
- HTML, CSS, JavaScript, Bootstrap (for frontend)

HARDWARE USED:

- Processor: Intel i5 or higher
- RAM: 8 GB or more
- Storage: 20 GB or more
- GPU (optional but recommended for model training)

Result and Conclusion:

1. **Accurate Sentiment Detection Across Multiple Media Types**
    - **Goal**: The system is expected to accurately detect sentiment from various input types (text, audio, images, and DOCX), with each media type handled by specialized models.
    - **Outcome**: By combining different algorithms optimized for each media type, the platform will provide accurate sentiment scores or classifications. For example, a user-uploaded text document might yield a "positive" sentiment classification, while an audio recording could indicate "neutral" sentiment.
2. **Real-Time Analysis and Display of Sentiment Insights**
    - **Goal**: The system aims to process inputs quickly and provide sentiment results in real time or near-real time.
    - **Outcome**: Users will experience minimal delay when uploading files and receiving sentiment feedback, making the platform suitable for applications where quick sentiment analysis is essential, such as customer feedback monitoring or social media sentiment analysis.
3. **Integrated Dashboard for Sentiment Visualization**
    - **Goal**: Provide a user-friendly dashboard where sentiment insights for each media type are displayed in an organized, visually intuitive format.
    - **Outcome**: The dashboard will allow users to see sentiment scores, classifications, and potentially visualizations (e.g., pie charts, bar graphs) for each type of media they upload. This visualization enhances user understanding of the analysis results, presenting a comprehensive view of sentiment across all media types.

Conclusion:

This multi-modal sentiment detection project addresses the growing need for comprehensive sentiment analysis across diverse media types. Current sentiment analysis systems are generally limited to text-based analysis, missing valuable insights from audio, images, and DOCX files. This proposed solution bridges this gap by integrating text, audio, image, and document sentiment detection into a single web-based platform.

Project Outcome and Industry Relevance:

Project Outcome:

The project successfully demonstrates how combining multiple data modalities—such as text, audio, video and image using deep learning fusion techniques and transformer models can significantly improve sentiment analysis accuracy. It highlights the power of multimodal data integration, using models like BERT for text and CNN/RNN or transformer variants for other modalities, to create a more holistic and context-aware understanding of human emotions.

Industry Relevance:

1. **Customer Service**: Enhancing chatbots and virtual assistants with emotional intelligence.
2. **Marketing and Brand Monitoring:** Analyzing user feedback from videos, reviews, or social media to understand public sentiment**.**
3. **Healthcare and Mental Health:** Supporting emotion recognition in teletherapy or digital mental health applications.
4. **Entertainment & Media:** Improving user recommendations by understanding emotional reactions to content.
5. **E-learning Platforms:** Adapting content delivery based on student reactions to improve learning outcomes.

Working Models vs. Simulation/Study:

Simulation/Study:

- The project was simulated using real-world multimodal datasets (text, audio, visual) to evaluate the performance of different fusion strategies and transformer models.
- Comparative analysis was conducted between unimodal and multimodal approaches, proving the effectiveness of deep learning fusion in enhancing sentiment detection accuracy.

Project Outcomes and Learnings:

The project successfully developed a multimodal sentiment analysis system that integrates text, audio, and visual data using deep learning fusion techniques and transformer-based models. By leveraging architectures like BERT for text and vision/audio transformers for other modalities, the system achieved higher accuracy in detecting complex and nuanced emotions. The use of early, late, and hybrid fusion strategies enabled a more comprehensive emotional understanding, outperforming traditional unimodal sentiment analysis approaches. This outcome demonstrated the potential of combining multiple modalities for creating emotionally aware AI systems suitable for real-world applications.

Throughout the project, I gained valuable insights into handling multimodal data and the challenges of aligning and synchronizing different data types. I learned how to preprocess textual, audio, and visual features and how to implement effective fusion strategies for model integration. Working with transformer architectures taught me the importance of attention mechanisms and contextual embeddings. Additionally, I developed strong skills in model evaluation, hyperparameter tuning, and interpreting results to improve performance. The experience also enhanced my understanding of the real-world implications of AI systems in fields like customer experience, healthcare, and media analysis.

Future Scope:

1.Multimodal Data: The project combines text, audio, and visual data to improve sentiment Analysis accuracy by capturing diverse emotional cues from different sources.

2.Deep Learning & Transformers: Leverages transformer-based models (e.g., BERT, GPT, ViT) to process and analyze each modality, enabling better contextual understanding.

3.Fusion Techniques: Explores strategies like early, late, and hybrid fusion to combine multimodal data for enhanced sentiment prediction.

4.Real-World Applications: Applicable to areas like customer feedback, social media monitoring, and healthcare for deeper insights into user sentiment.

5.Challenges & Future Trends: Addresses data alignment, imbalances, and computational complexity, with future directions focused on real-time analysis and emotional intelligence beyond basic sentimen