**Multimodal Sentiment Analysis: Project Report**

**Developed by :** Anbhi Thakur

**Date:** 27/03/2025

**1. Executive Summary**

This report outlines the development and implementation of a **Multimodal Sentiment Analysis System**, leveraging advanced Natural Language Processing (NLP) and Computer Vision techniques. The system integrates **BERT-based text sentiment analysis, GPT-2 text generation, YOLO-based object detection, and DeepFace emotion recognition**, enabling a holistic approach to sentiment understanding. A Flask-based API was developed for seamless accessibility and deployment.

**2. Technologies Utilized**

**2.1 Core Development Stack**

* **Programming Language:** Python
* **Frameworks & Libraries:**
  + **NLP & Transformers:** Hugging Face Transformers (BERT, GPT-2)
  + **Computer Vision:** YOLOv8 (Object Detection), DeepFace (Facial Sentiment Analysis)
  + **Data Processing:** Pandas, NumPy, Hugging Face Datasets
  + **Backend & API Development:** Flask
  + **Dataset Management:** Kaggle API, OS operations

**3. Project Workflow & Implementation**

**3.1 Dataset Acquisition & Preprocessing**

* Configured **Kaggle API authentication** to automate dataset retrieval.
* Downloaded and processed **IMDB Sentiment Analysis Dataset**, **Amazon Reviews**, and additional datasets from **Hugging Face Datasets**.
* Utilized **Pandas** for efficient data loading, preprocessing, and exploratory analysis.

**3.2 Text-Based Sentiment Analysis**

* Implemented **BERT-based** sentiment classification using nlptown/bert-base-multilingual-uncased-sentiment.
* Integrated **GPT-2** for **contextual text generation**, enabling AI-driven sentiment-based text predictions.
* Developed a unified function to analyze input text, classify sentiment, and generate AI-driven responses.

**3.3 Visual Sentiment Analysis**

* Integrated **YOLOv8** for object detection to analyze visual context.
* Leveraged **DeepFace** for facial emotion recognition, classifying expressions such as joy, sadness, anger, and surprise.
* Combined both methodologies to derive **visual sentiment insights** from images.

**3.4 Multimodal Sentiment Fusion**

* Designed a **Feature-Level and Decision-Level Fusion Model** to merge textual and visual sentiment data.
* Combined outputs from **BERT, GPT-2, YOLO, and DeepFace** into a structured sentiment representation.
* Created a scoring mechanism for a more **comprehensive sentiment analysis**.

**3.5 Backend Development & API Integration**

* Developed a **Flask-based API** to facilitate real-time sentiment analysis.
* Implemented the /analyze endpoint for multimodal sentiment processing.
* Integrated **Retrieval-Augmented Generation (RAG)** via /rag\_query to enhance query-based sentiment retrieval.

**3.6 Deployment & System Scalability**

* Hosted the Flask server for **real-time sentiment processing**.
* Ensured **scalability and reliability** by optimizing API responses and handling concurrent requests efficiently.
* Designed for potential integration with **business analytics tools, social media monitoring platforms, and chatbot systems**.

**4. Conclusion & Business Relevance**

This **Multimodal Sentiment Analysis System** demonstrates a **cutting-edge AI approach** by combining **textual and visual sentiment analysis** for enhanced consumer and market insights. The project provides a scalable, API-driven framework for applications in **customer feedback analysis, brand sentiment monitoring, AI-driven chatbots, and automated content moderation**.

The integration of **NLP, Computer Vision, and Multimodal AI** ensures a **highly adaptable and robust** solution suitable for enterprise deployment and further expansion into real-world sentiment-driven applications.