Multimodal Sentiment Analysis of Instagram Posts

ETANA DISASA

MS IN DATA SCIENCE

REGIS UNIVERSITY

Introduction

Social media platforms such as Instagram host vast amounts of user-generated content combining images, text captions, and comments.

Understanding public sentiment from these multimodal posts is vital for brands, researchers, and individuals seeking insights into content perception and emotional tone.

Traditional sentiment analysis has focused on text alone, missing the nuanced meaning conveyed through visual content.

Instagram posts are **multimodal**: images + captions + comments.

Traditional sentiment analysis focuses on text alone.

Need: Integrated tools to understand public perception, emotional tone, and unfavorable sentiment.

This practicum addresses the need for integrated multimodal sentiment analysis tools by developing a proof-of-concept interactive application that processes Instagram posts to extract and analyze their sentiment in real time.

Introduction

Social media enables toxic content and negative emotional contagion.

Research Question: *How can a multimodal sentiment analysis system be developed to analyze Instagram post content (images, captions, comments) to provide post sentiment that helps identify harmful content?*

The purpose of this project is to develop an interactive tool to analyze Instagram posts' sentiment (image, video, caption, comments).

Introduction

Approach:

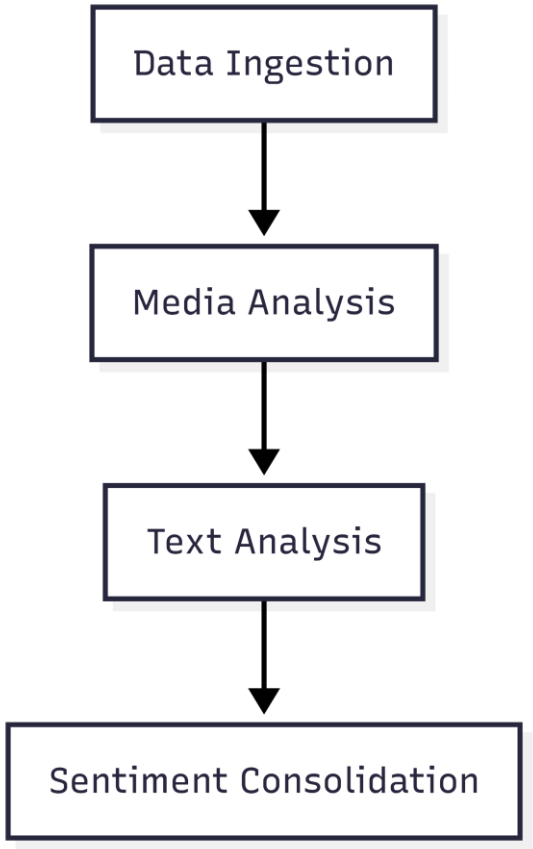
NLP (Natural Language Processing): For analyzing captions, comments, and image descriptions using BERT and emoji handling.

Computer Vision: For analyzing images and video frames, including face emotion detection and NSFW filtering.

Multimodal Fusion / Data Integration: Combining text and visual signals into a single sentiment/emotion output.

Outcome: Real-time insights; supports mental health research, brand monitoring, social media analysis.

Methodology



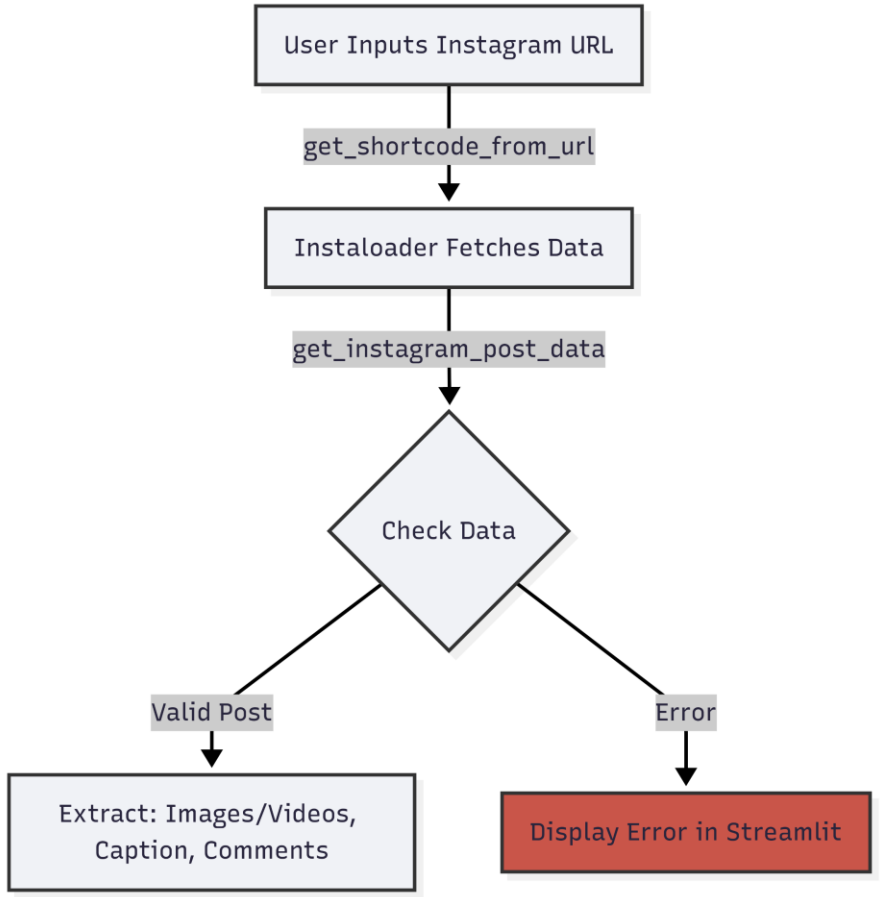
Step 1: Data Ingestion – Collect images, video frames, captions, and comments from Instagram posts.

Step 2: Media Analysis – Filter NSFW content, detect faces, generate image descriptions, and analyze face emotions.

Step 3: Text Analysis – Analyze sentiment of post captions and top 5 comments.

Step 4: Sentiment Consolidation – Combine media, caption, and comment sentiment into a final post-level sentiment (weighted 50% media, 30% caption, 20% comments).

Methodology



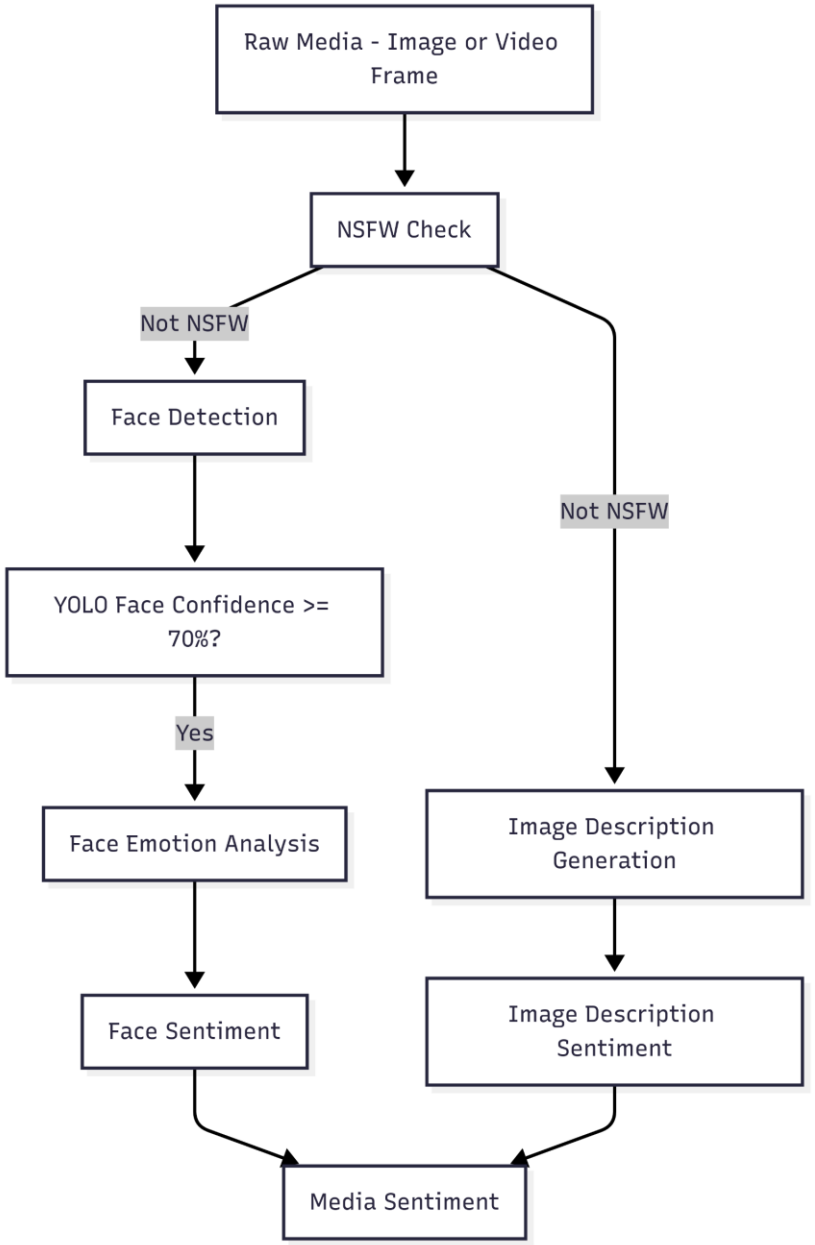
Step 1. Data Ingestion

- User enters a valid Instagram URL to collect public post data.
- Checks Data and validates the link.
- Shows an error if the post is invalid.
- If valid URL is valid, images, videos, captions, and comments (top 5) are extracted for analysis.

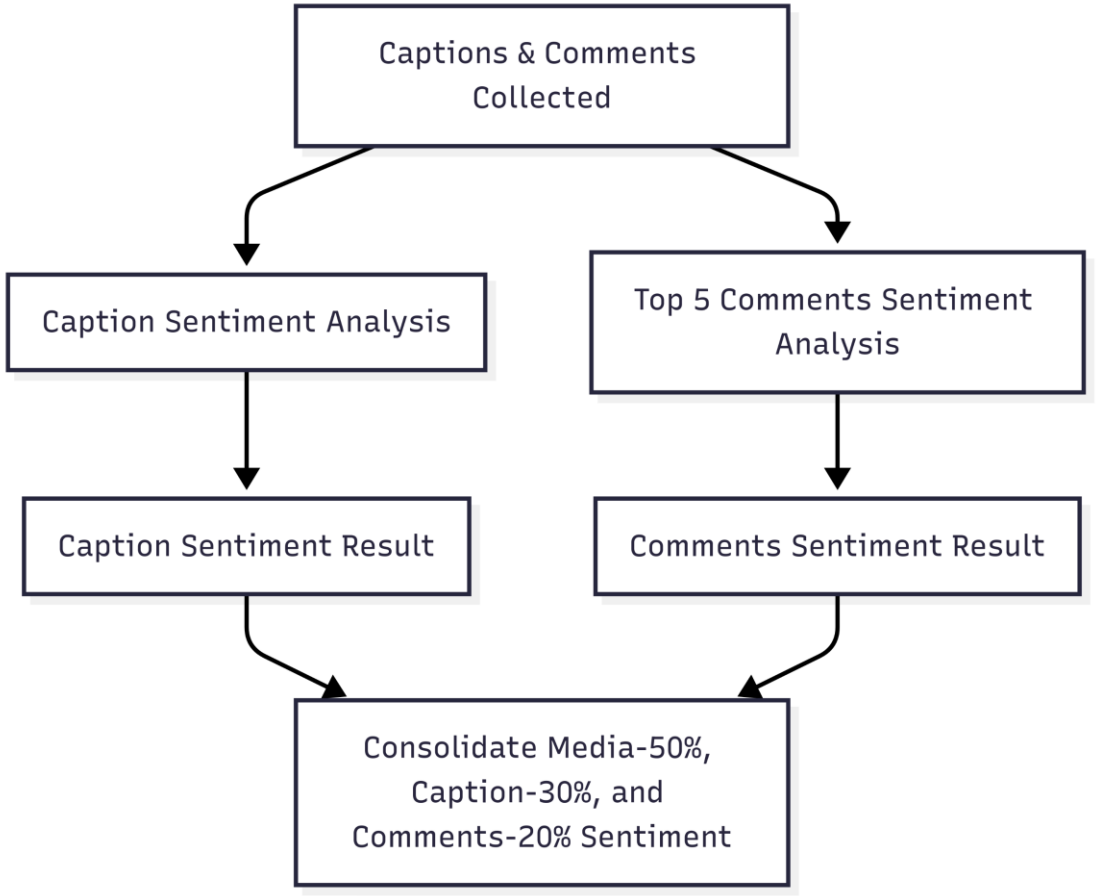
Methodology

Step 2. Media Analysis

- Run NSFW check on all media; skip inappropriate content.
- For non-NSFW media:
- Generate image descriptions.
- Detect faces and run YOLO confidence ≥ 70%.
- If a confident face exists, perform face emotion analysis.
- Consolidate face sentiment and image description sentiment into media sentiment.



Methodology



Step 3. Text Analysis

- Caption Sentiment analyzed using pretrained BERT (bhadresh-savani/bert-base-go-emotion) model.
- Top 5 Comments first demojized (emoji library) to better analyze sentiment from emojis.
- Emotion analysis from Media, Caption and Comments consolidated.

4. Sentiment Consolidation

- Emotion and Sentiment standardized
- Data passed for visualization

Tools Utilized

Data Retrieval & Processing:

Instaloader: Collect Instagram posts, captions, and comments.

OpenCV: Extract video frames as images for analysis.

Computer Vision / Image Analysis:

YOLOv8: Face detection for emotion analysis.

BLIP-2: Generate descriptive captions for images and video frames.

Pretrained Face Emotion Model: Detect emotions from faces.

NSFW Model: Filter inappropriate images.

Natural Language Processing (NLP):

BERT (pretrained models): Sentiment analysis for captions, comments, and image descriptions.

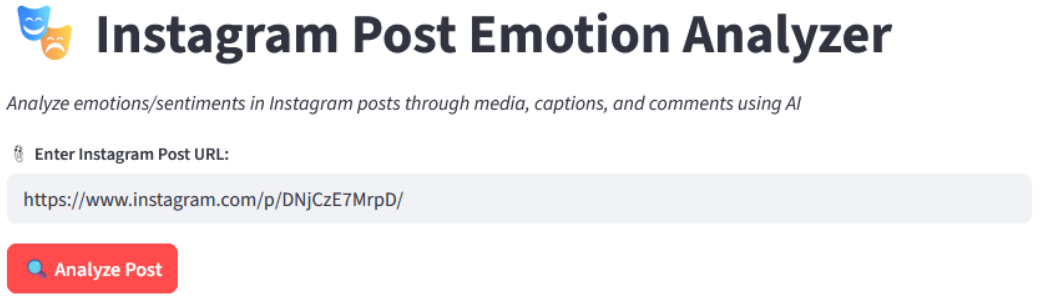
Emoji Library: Convert emojis to text for sentiment interpretation.

Deployment & Utilities:

Streamlit: Build an interactive app for input and results display.

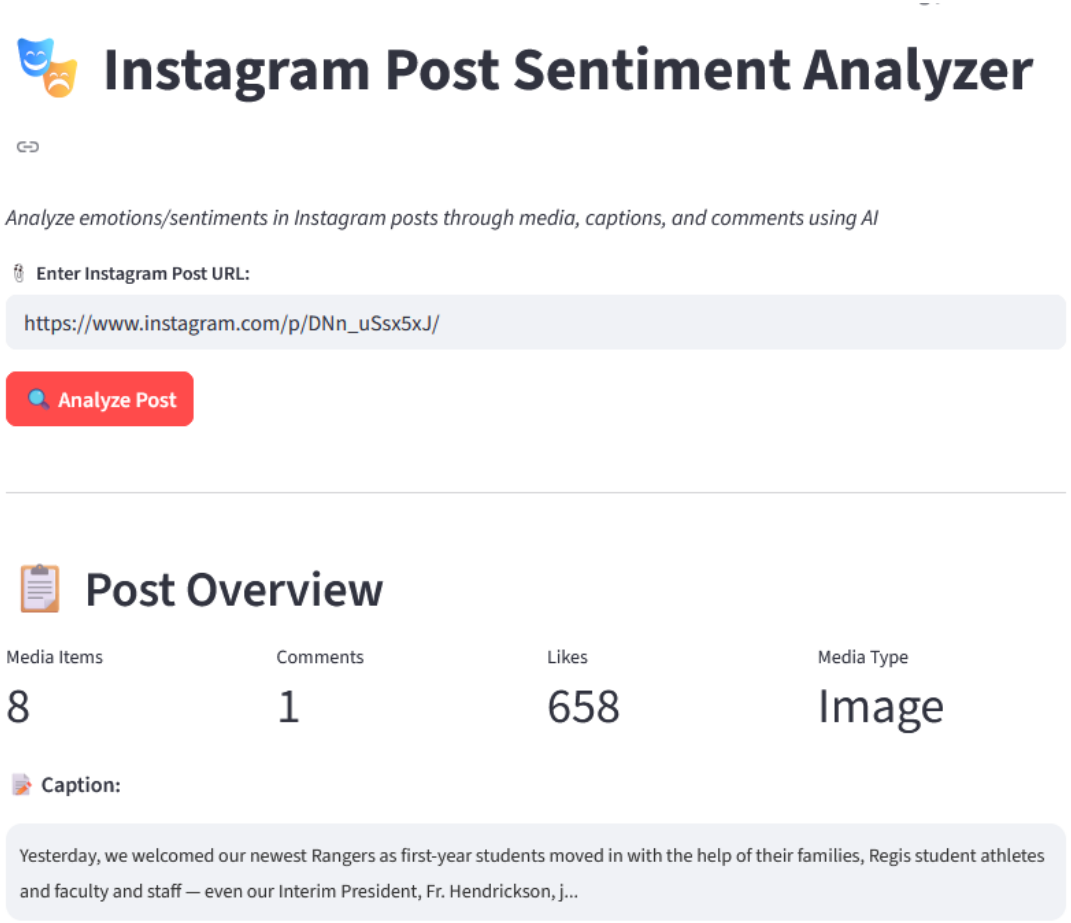
Python Libraries: Pandas, NumPy, Matplotlib / Seaborn for data handling and plotting.

Example Posts



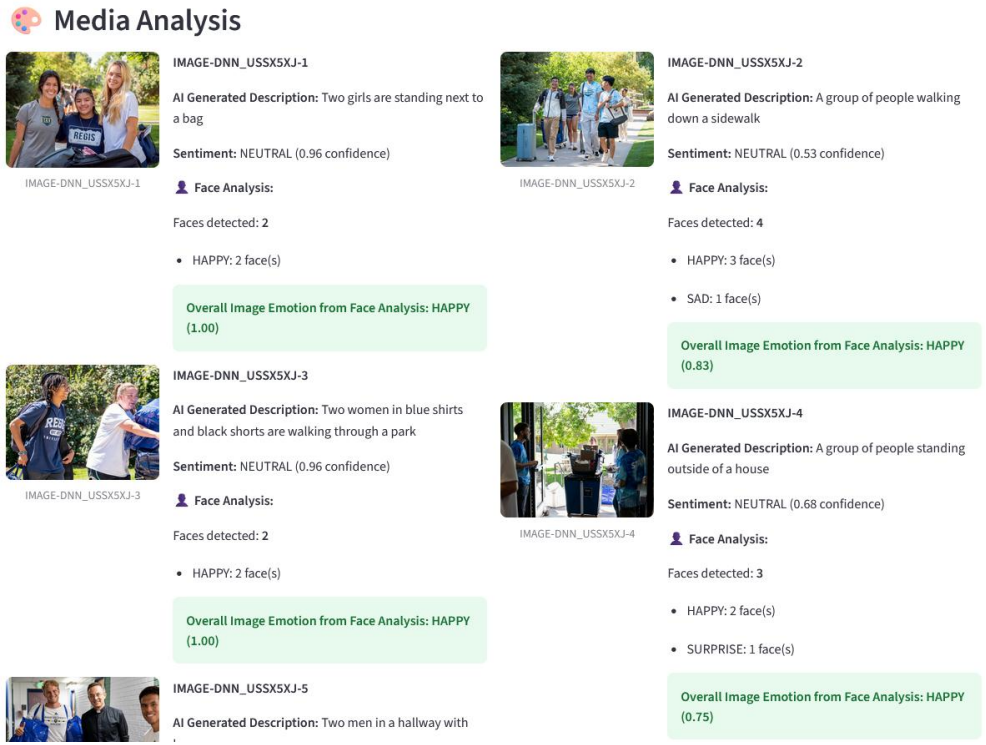
- Regis University's Instagram page posted this post.
- The link is copied from browser as https://www.instagram.com/p/DNn_uSsx5xJ/
- The streamLit UI allows user to enter it.

Example Posts



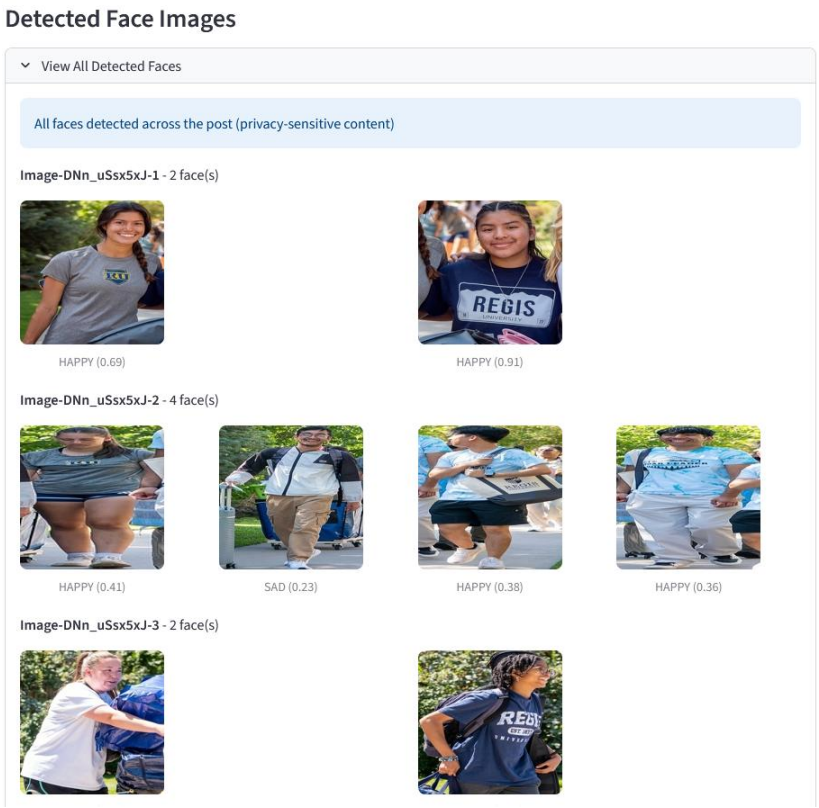
- A brief overview including the type of media posted (Image, Video), number of comments,
- Like and number of media posted is featured.
- Furthermore, an excerpt from the caption of the post is presented.

Example Posts



- If an image/video is appropriate to analyze, a thumbnail of images (pictures or an image from a video) will be populated.
- For each image, an image caption with the unique Instagram ID will be associated.
  - E.g. IMAGE-DNJCE7MRPD-1
- AI Generated Description will be generated.
- Sentiment analysis from the generated description will be displayed.
- Face Analysis shows how many faces detected and emotions captured.
- An overall weighted media sentiment will be displayed for each media.

Example Posts

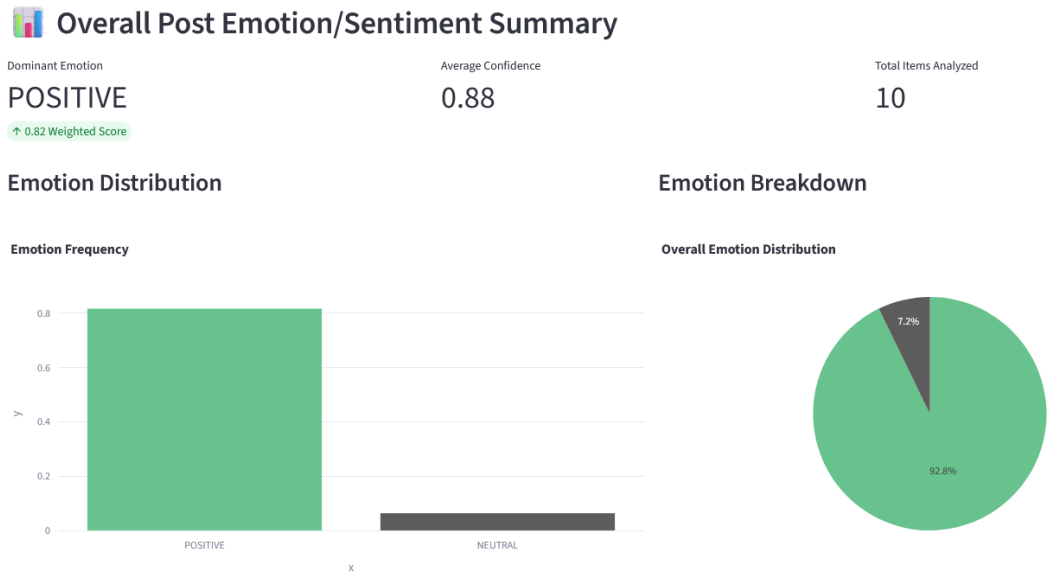


- It will also allow the user to view all detected faces, emotions captured and confidence point.
- This allows the user to pick on details that may not be visible at a first glance and provide a robust face recognition mechanism.
- To avoid mis-capturing faces, the YOLO model was set to pass at least a 70% confidence rate for a face to be included in the analysis.

Example Posts

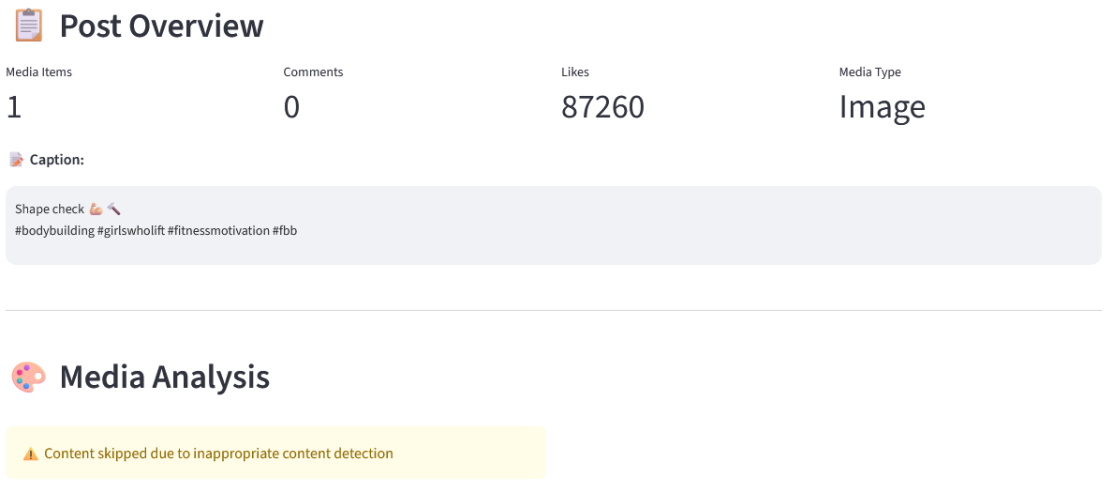
- Captions and comments are analyzed.
- To capture a truer sentiment from comments, emojis are converted to text before being analyzed.
- Caption sentiment and comment sentiment are listed accordingly.

Example Posts



- A final summary is presented in these visuals for an easy understanding.
- Potentially, this app will be used by non-technical users, therefore, a simple presentation and color scheme appropriate to implement

Example Posts



When inappropriate content is identified, media processing is halted.

Project Contributions

- **Multimodal Analysis**: Integrated image, video, captions, and comments into a single sentiment/emotion pipeline.
- **NSFW & Face Filtering**: Automatically filtered inappropriate content and detected face emotions to improve accuracy.
- **Text Sentiment Insights**: Analyzed captions, comments, and image descriptions using pretrained BERT models, including handling emojis.
- **Media Description**: Generated descriptive captions for images and video frames using BLIP-2 for richer text-based analysis.
- **Consolidated Sentiment**: Weighted combination of image, face, media, caption, and top comment sentiments into a final summary.
- **Reusable Workflow**: Demonstrated a reproducible methodology for multimodal social media analysis with Streamlit deployment.

Ethical Considerations

Privacy & Consent:

- Only publicly available Instagram posts were analyzed.
- No private or direct messages were accessed.

Data Sensitivity:

- NSFW and inappropriate content was filtered to prevent misuse.
- Media and text containing sensitive information handled cautiously.

Bias & Fairness:

- Pretrained models may reflect cultural, gender, or racial biases.
- Interpreted sentiment/emotions responsibly.

Responsible AI Use:

- Outputs intended for research, monitoring, or educational purposes.
- Not used for individual profiling, harassment, or discriminatory actions.

Transparency:

- Clear documentation of models, preprocessing, and analysis steps.

Challenges and Opportunities

Challenges:

- Accessing public Instagram post data while maintaining session stability.
- Library compatibility issues (e.g., TensorFlow, OpenCV, Streamlit).
- Loading and integrating multiple pretrained models efficiently.
- Limited documentation or examples for some specialized models (NSFW detection, face emotion).
- Handling diverse media types (images, video frames, captions, comments) in a unified pipeline.

Opportunities:

- Developing a reproducible multimodal sentiment analysis workflow.
- Exploring integration of multiple pretrained models for richer insights.
- Potential applications in social media monitoring, brand research, or digital well-being studies.
- Foundation for future research on ethical AI and content moderation.

Conclusion

Supports Ethical AI Applications

- Demonstrates **responsible and ethical AI deployment** by incorporating transparency, privacy considerations, and harm reduction in automated social media analysis.

Advances Multimodal Research

- Contributes to the growing field of **multimodal sentiment analysis**, which is essential for building AI systems that understand human communication holistically (images + text).

Empowers Brands and Communities

- Enables brands to monitor **public perception and community mood** more accurately, facilitating proactive engagement and ethical marketing strategies.

Raises Awareness of The Social Dilemma

- By highlighting toxic or harmful content, your project aligns with broader societal discussions about **how social media platforms influence mental health, democracy, and well-being**.

Conclusion

Addresses Online Toxicity

- Social media platforms like Instagram can become breeding grounds for **negative sentiments, bullying, and toxic interactions**.
- Your tool provides **real-time detection and insights**, empowering users, researchers, and policymakers to **identify harmful patterns**.

Promotes Digital Well-being

- By revealing the emotional tone of posts and comments, the tool fosters **self-awareness and healthier online engagement**, aligning with mental health promotion efforts.

References

- Hugging Face. BERT Pretrained Models.
- OpenCV Library. https://opencv.org/
- Instaloader. https://instaloader.github.io/
- BLIP-2: Bootstrapped Language-Image Pretraining.
- YOLOv8: Ultralytics Object Detection Models.

References

Baltrusaitis, T., Ahuja, C., & Morency, L. P. (2019). Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(2), 423–443. https://ieeexplore.ieee.org/document/8214216

Chancellor, S., & De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: A critical review. npj Digital Medicine, 3(1), 1–11. https://www.nature.com/articles/s41746-020-0233-7

Cheng, J., Danescu-Niculescu-Mizil, C., & Leskovec, J. (2017). Antisocial behavior in online discussion communities. ICWSM, 11, 61–70.

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Schafer, B. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707.

Kross, E., Verduyn, P., Demiralp, E., Park, J., Lee, D. S., Lin, N., ... & Ybarra, O. (2021). Social media and well-being: Pitfalls, progress, and next steps. Trends in Cognitive Sciences, 25(1), 55–66.

Lup, K., Trub, L., & Rosenthal, L. (2015). Instagram instasad?: Exploring associations among Instagram use, depressive symptoms, negative social comparison, and strangers followed. Cyberpsychology, Behavior, and Social Networking, 18(5), 247–252.

Niu, X., Li, H., Xu, X., & Xu, C. (2016). Visual sentiment analysis on Flickr images. IEEE Transactions on Multimedia, 18(11), 2224–2233.

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. NYU Press.

Orlowski, J. (Director). (2020). The Social Dilemma [Film]. Netflix.

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. Information Fusion, 37, 98–125.

Tsai, Y. H. H., Bai, S., Yamada, M., Morency, L. P., & Salakhutdinov, R. (2019). Multimodal transformer for unaligned multimodal language sequences. Proceedings of the Annual Meeting of the Association for Computational Linguistics, 6558–6569.

Verduyn, P., Ybarra, O., Resibois, M., Jonides, J., & Kross, E. (2017). Do social network sites enhance or undermine subjective well-being? A critical review. Social Issues and Policy Review, 11(1), 274–302

Acknowledgements

I would like to sincerely thank:

My practicum professor Dr. Ghulam Mujtaba for guidance, patience, and insightful feedback throughout this project.

My classmates for their constant support, encouragement, and constructive suggestions.

My wife Ellu for her unwavering support during late nights, troubleshooting sessions, and throughout this entire project.

Your guidance and collaboration made this work possible.