# ANALYTICA POSTER

SUBMITTED BY:
Aditya Srivastava (2201013)
Deepanika Gupta (2201062)
Kartik Saini (2201103)

## ABSTRACT

Analytica is a full-stack tweet analysis platform that performs **real-time sentiment, emotion, and toxicity classification** on tweets fetched using usernames or hashtags. The system supports **multilingual input**, dynamic visualization, a **leaderboard**, and a **Live Wall**. By leveraging **transformer-based models from CardiffNLP** and integrating features like user history and login, Analytica addresses both functional richness and usability for researchers, journalists, and digital marketers.

## ARCHITECTURE OVERVIEW

• **Frontend:** React + Tailwind (modern SPA with components like TweetCard, AnalysisForm).

• **Backend**: Django REST Framework with integrated NLP analysis modules.

• **Models Used**: cardiffnlp/twitter-roberta-base-{task} for sentiment, emotion, and toxicity.

• **Scraping Engine**: Selenium with GeckoDriver to extract tweets dynamically.

• **Visualization**: Badge-based live tweet cards + pie-chart/statistical modules.

• **Database**: SQLite during dev (supports Postgres for scaling).

## PROPOSED TECHNIQUE

Analytica employs a modular pipeline for real-time, multilingual tweet analysis across sentiment, emotion, and toxicity:

• **Tweet Scraping**: Tweets are fetched live using Selenium based on user-input username/hashtag and tweet count.

• **Translation & Preprocessing**: Non-English tweets are auto-translated to English, followed by text cleaning and tokenization for accurate model input.

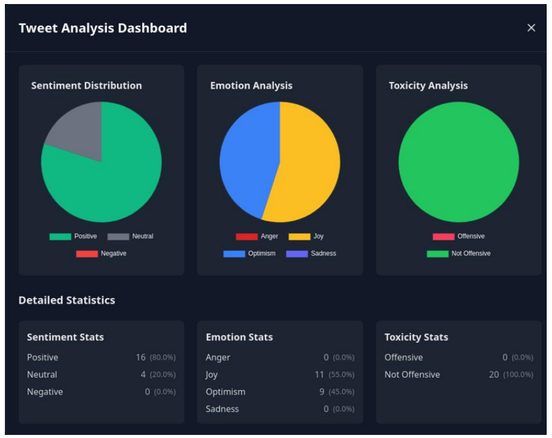
• **Triple NLP Pipeline**: Each tweet is simultaneously analyzed using CardiffNLP's RoBERTa models for:

- (a) Sentiment — Positive / Negative / Neutral
- (b) Emotion — Joy / Sadness / Anger / Optimism
- (c) Toxicity — Offensive / Not Offensive

• **Storage & History**: Each logged-in user's analysis results are stored with timestamps, enabling detailed activity tracking and re-access.

• **Visualization**: The analyzed data is presented through a Live Wall and pie charts using color-coded badges for intuitive understanding.

• **Leaderboard**: Displays top users, queries, and hashtags across various analysis types, promoting engagement and exploration.




## NOVELTY CONTRIBUTIONS

• **Combined Analysis**: Simultaneous sentiment, emotion, and toxicity detection using transformer models.
• **Real-Time & Multilingual**: Live tweet scraping with automatic translation of non-English tweets.
• **Live Wall**: Unique animated badge-based visualization of analyzed tweets.
• **User Personalization**: Login, history tracking, and leaderboard features for interactive user experience.
• **Full-Stack Optimization**: Efficient integration of React, Django, and HuggingFace for responsive, low-latency analysis.

| Feature | VADER-Based System | Analytica |
|---|---|---|
| Emotion Detection | Not supported | **80.2%** accuracy |
| Toxicity Detection | Not supported | **82.2%** accuracy |
| Multilingual Support | English only | **Yes (via translation)** |
| Visualization | Basic static charts | **LiveWall with badges** |
| User History | Not available | **Yes** |
| Leaderboard | Not available | **Yes** |

## KEY RESULTS

• Analytica offers **higher accuracy**, **real-time analysis**, and supports **multilingual tweets**, unlike VADER (English-only) and TweetEval (offline, task-specific).

• Analytica provides **combined sentiment, emotion, and toxicity analysis**, along with **user history**, **leaderboard**, and **Live Wall**—features not available in TweetNLP's limited demo interface.

## FUTURE SCOPE

• **Twitter API Integration**: Replace Selenium with **Tweepy (Twitter API)** for more stable and scalable tweet retrieval. While API access offers higher reliability, it may incur costs under Twitter's latest pricing model.

• **Advanced Language Support**: Fine-tune models on **Indian languages** and regional datasets for improved sentiment and emotion detection in diverse contexts.

• **Cloud Deployment**:
Host Analytica on **AWS / Heroku / GCP** for continuous, global access and data persistence.

## REFERENCES

1. A. R. Pai, M. Prince, and P. C. V. *Real-time Twitter sentiment analytics and visualization using VADER*, **2022 2nd Int. Conf. on Intelligent Technologies (CONIT), IEEE**
2. F. Barbieri, J. Camacho-Collados, et al. *TweetEval: Unified benchmark and comparative evaluation for tweet classification*
3. J. Camacho-Collados, K. Rezaee, et al. *TweetNLP: Cutting-edge NLP for social media*