

Phase 1: Research

Objective:

Investigate tools, technologies, and best approaches for a real-time sentiment analysis pipeline on social media data.

Research Goals:

- Understand sentiment analysis techniques.
- Explore data scraping via snscreap and Twitter API.
- Compare tools for sentiment classification (TextBlob vs VADER).
- Study Streamlit for interactive dashboards.

Key Decisions:

- Use **TextBlob** for sentiment scoring.
- Use **Streamlit** for the front-end dashboard.
- Due to Twitter API/snscreap issues, switch to a **CSV-based data approach** (simulate real-time using pre-collected tweets).

Files Involved:

- sample_tweets.csv (manually scraped Twitter data).

Outcome:

- Technology stack finalized.
- Problem-solution mapping completed.
- Dataset format and structure defined.