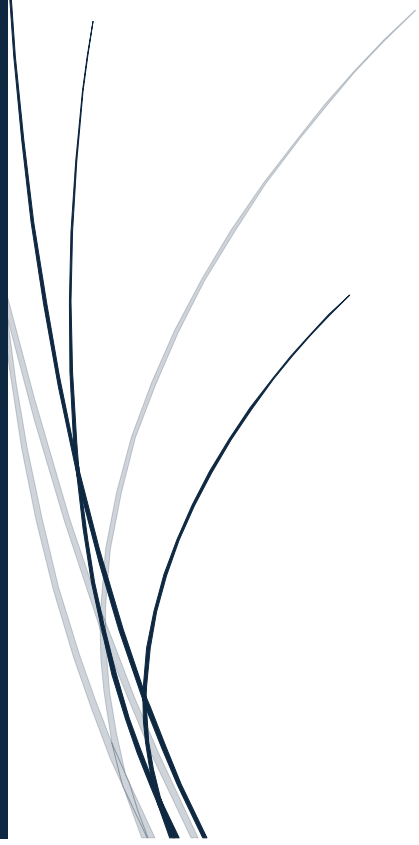
[Date]

PROGRESS REPORT

Big Data Analytics for Social Media Trends



PROGRESS REPORT

Project Title: *Big Data Analytics for Social Media Trends*
Date: April 30, 2025

INTRODUCTION (OVERVIEW, BACKGROUND)

This project investigates how big data analytics can be applied to identify, analyze, and predict social media trends over time. With the rapid growth of platforms like Twitter, Instagram, Facebook, and Reddit, understanding user sentiment and reaction patterns has become essential in fields ranging from marketing to policy-making.

Purpose of the Project

To develop a data-driven framework for analyzing public sentiment and topic trends across multiple social media platforms using big data tools.

Specific Objectives of the Project

- Collect and preprocess social media data from multiple platforms using APIs.
- Build a clean, structured dataset categorized by platform, country, time, and user sentiment.
- Apply machine learning models and natural language processing techniques to detect patterns and trends in social discourse.
- Visualize temporal and geographical insights across key global events.

Scope of the Project

The study will focus on English-language, public posts from Twitter, Instagram, Facebook, and Reddit, spanning from 2008 to 2024. The scope includes sentiment and emotion analysis, topic modeling, trend visualization, and comparative event-based analysis.

Project Timeline

- **Start Date:** April 14, 2025
- **Projected Completion:** May 9, 2025

Project Team

- Zain Baig – 22K4593
- Umer Ahmed – 22K4213
- Arsalan – 22K4614
- Qasim Khan – 22K4380
- Maaz Kashif – 22K4518
- Abdul Rafay Iqbal – 22K4182

Instructor

Mrs. Nazia Imam
[Technical and Business Writing], [FAST - NUCES]

Overview of Contents

This progress report covers all completed and ongoing work, outlines remaining tasks using a Gantt chart, and details challenges encountered. It confirms adherence to the proposed schedule and the plan for successful project completion.

DISCUSSION (FINDINGS, BODY)

(Chronological Organization)

Work Completed

- **Project Planning Finalized**: Timeline, objectives, and tools defined.
- **Data Collection Setup**:
  - APIs configured for Twitter (Tweepy), Reddit (PRAW), and Instagram (Instaloader).
  - Facebook data strategy outlined (limited access via CrowdTangle/manual scraping).
- **Data Collected**:
  - Public posts sampled randomly from 2008–2024, categorized by year and platform.
  - Focused on diverse topics, events, and global user data.
- **Data Preprocessing Completed**:
  - Stopword removal, lemmatization, noise filtering.
  - Language filtering to retain English-only posts.
- **Clean Tables Created**:
  - Structured data tables with fields: `user`, `country`, `date`, `month`, `year`, `platform`, `text`.

Work in Progress

- Setting up **TF-IDF matrices** for all textual data.
- Initial **sentiment polarity scoring** using VADER/TextBlob.
- Preparing data input for **BERTopic** and **LDA** topic modeling.
- Drafting visuals for country-wise and year-wise sentiment trends.

Work Remaining

Task	Status	Deadline
Sentiment + Emotion Analysis	To do	May 3
Topic Modeling (LDA/BERTopic)	To do	May 4
Time-Series & Clustering Analysis	To do	May 5
Visualization (heatmaps, word clouds)	To do	May 6-7
Final Report Writing & Presentation	To do	May 8

Problems Encountered

- **Facebook API Restrictions**: Limited access due to policy changes; workaround involves publicly available data or selected scraping.
- **Data Gaps in Early Years**: Some years (e.g., 2008–2011) lack dense data. Will note these in analysis and adjust visuals accordingly.
- **Platform Differences**: Visual vs. text-heavy platforms require slightly different NLP approaches (e.g., Instagram captions vs. tweets).

CONCLUSIONS / RECOMMENDATIONS

The project is **on schedule**. Planning, data collection, and cleaning stages are complete. We are now entering the model-building and analysis phase. Despite some platform limitations, the core datasets are robust enough for meaningful trend and sentiment analysis.

We are confident that the final report will be completed by **May 09, 2025**, as originally planned. The project is expected to provide valuable insight into historical and geographical patterns of public sentiment and event-based reactions, backed by clean data, solid analytics, and rich visualizations.