Research Proposal: Big Data Analytics for Social Media Trend

Group Members:

- -Umer Ahmed – 22K4213
- -Zain Baig – 22K4593
- -Arsalan – 22K4614
- -Qasim Naveed – 22K4380
- -Maaz Kashif – 22K4518
- -Abdul Rafay – 22K4192

Research Problem

Social media generates massive amounts of data daily, reflecting public opinion, emerging issues, and social shifts. However, accurately predicting trends from this unstructured, fast-paced data remains a challenge due to issues like noise, scalability, and lack of real-time contextual understanding. There is a pressing need for improved methods that can analyze and predict trends across multiple platforms using big data tools.

Purpose

To develop and test a framework using big data analytics techniques (sentiment analysis, engagement metrics, and trend modeling) to identify and predict emerging trends on platforms like Twitter, Instagram, and Reddit.

Relevance and Benefits of the Study

- **Researchers**: Advances methodologies in big data and social media analytics.
- **Industry**: Supports brand management, crisis response, and customer insight generation.
- **Society**: Helps monitor public sentiment, detect misinformation, and inform policy decisions.

Literature Review

The study draws on the following peer-reviewed sources to support its foundation:

1. Chen et al. (2018), *ACM Transactions on the Web* – Sentiment via emoji embedding
2. Bandari et al. (2012), *ICWSM* – Popularity prediction via engagement
3. Liu, B. (2012), *Sentiment Analysis Survey* – Core foundation for sentiment mining
4. Sharma & Ghosh (2017), *Procedia Computer Science* – Real-time analytics with Spark
5. Fan & Gordon (2014), *Decision Support Systems* – Tool comparison for social analytics
6. Morstatter et al. (2013), *ICWSM* – Twitter real-time stream processing
7. Zeng et al. (2010), *Communications of the ACM* – Trends and tools in social analytics

Research Methodology

Design:
Quantitative, big data–driven design incorporating trend modeling, sentiment analysis, and engagement metrics.

Sample:
Publicly available data from Twitter, Instagram, and Reddit collected via platform APIs (Tweepy, PRAW, Instaloader).

Tools:

- **Languages & Libraries**: Python (NLTK, TextBlob, Scikit-learn, Seaborn)
- **Visualization**: Tableau, Matplotlib, WordCloud
- **External Sources**: Google Trends for correlation

Analysis Techniques:

- Sentiment analysis (lexicon + machine learning)
- Regression & time-series models (ARIMA)
- Clustering/topic modeling (LDA, K-Means)
- Visualization for trend interpretation