Team 11: Linear Regression Analysis of Public Sentiment in the 2024 U.S. Presidential Election

1 Introduction

The folder 11 contains the **2024 U.S. Presidential Election Social Media** dataset, which provides real-time insights into public sentiment, political engagement, and policy discussions using data collected from **X (formerly Twitter)**. Covering key candidates from multiple political parties, this dataset allows for an in-depth analysis of how social media interactions influence electoral trends. By leveraging **linear regression**, we can explore the relationship between social media engagement metrics and sentiment scores, uncovering trends that shape voter perceptions and campaign effectiveness.

2 Dataset Description

The dataset contains the following key attributes:

- **Tweet Text** – The textual content of each tweet.
- **Timestamp** – The date and time when the tweet was posted.
- **User Handle** – The username of the account posting the tweet.
- **Sentiment Labels** – Categorical sentiment classification (Positive, Neutral, Negative).
- **Engagement Metrics** – Number of retweets and likes per tweet.
- **Party Affiliation** – The political party associated with the tweet's subject.

The dataset spans from **January to February 2025**, covering key moments in the final phase of the 2024 U.S. Presidential Election.

3 Tasks and Requirements

To analyze and extract meaningful insights using **linear regression**, the following tasks need to be performed:

3.1 Data Exploration and Preprocessing

- Load and inspect the dataset, handling missing or inconsistent values.
- Convert categorical sentiment labels into numerical values for regression analysis.
- Normalize numerical features such as engagement metrics.
- Analyze time-based trends in sentiment fluctuations.

3.2 Regression Analysis

- Model the relationship between sentiment scores and engagement metrics (likes, retweets) using linear regression.
- Investigate how sentiment trends correlate with specific candidates and party affiliations.
- Predict future sentiment trends based on past engagement patterns.
- Evaluate model performance using metrics such as **R-squared** and **Mean Squared Error (MSE)**.

3.3 Visualization and Reporting

- Generate time-series plots to visualize sentiment trends over key election events.

- Create scatter plots to examine relationships between engagement metrics and sentiment scores.

4 Submission Requirements

- A well-structured report detailing the methodology, results, and analysis in a given report format.

- Python code is used for implementation.

- A presentation summarizing key findings and recommendations in a given presentation format.