Sentiment Analysis Report

Title: Analyzing Public Sentiment on X (Twitter)
Author: Samiksha Atul Hujare
Date: 11.03.2025

---

1. Overview

This project applies sentiment analysis—a key technique in natural language processing—to evaluate public opinions expressed on X (formerly known as Twitter). By categorizing tweets into **Positive**, **Neutral**, and **Negative** sentiments, we aim to uncover trends and insights that can guide decision-making.

---

2. Data Summary

The dataset used for this analysis comprises pre-processed tweets collected from X. Each tweet comes with an associated sentiment label, defined as follows:

- **1.0**: Indicates Positive sentiment
- **0.0**: Indicates Neutral sentiment
- **-1.0**: Indicates Negative sentiment

Data Fields:

- **clean_text:** The tweet content after text cleaning (removal of URLs, mentions, hashtags, etc.)
- **category:** The numerical sentiment indicator (-1 for Negative, 0 for Neutral, 1 for Positive)

---

3. Approach

3.1 Data Preparation

Since the dataset is already cleaned, our focus was on mapping the numerical sentiment values to descriptive text labels. This step ensures clarity during analysis and visualization.

3.2 Sentiment Label Conversion

We transformed the numeric labels as follows:

- **1.0 → Positive**
- **0.0 → Neutral**

- **-1.0 → Negative**

This conversion facilitates intuitive understanding and interpretation of the results.

3.3 Visualization Techniques

To present our findings, we employed several visualization methods: - **Sentiment Distribution Bar Chart:** Illustrates the frequency of each sentiment type. - **Word Clouds:** Separate word clouds were generated for positive, negative, and neutral tweets to emphasize frequently occurring keywords.

---

4. Findings

4.1 Sentiment Frequency Analysis

The overall sentiment distribution from the dataset is as follows:

Sentiment	Frequency
Positive	Over 70,000 tweets
Neutral	Over 50,000 tweets
Negative	Over 30,000 tweets

The bar chart (saved as `sentiment_distribution.png`) provides a visual summary of these counts.

4.2 Insights from Word Cloud Visualizations

Word Cloud for Positive Tweets

The word cloud for positive tweets highlights that key terms include: - “india” - “will” - “bjp” - “now” - “people”

These words suggest the dominant themes and public interests within the positive sentiment group.

---

5. Interpretation

- **Neutral Tweets:** The prevalence of neutral tweets suggests that a significant portion of the audience expresses balanced or indifferent opinions.
- **Positive Tweets:** The frequent appearance of words like “india”, “will”, and “people” in positive tweets reflects a trend of affirmative and supportive messaging.
- **Negative Tweets:** Although not detailed in this section, negative tweets tend to include critical terms (e.g., “poor”, “pakistan”, “chowkidar”) and offer insights into areas of public discontent.

These findings can assist stakeholders in understanding audience sentiment and adjusting strategies accordingly.
