**CSCE 5320 - Scientific Data Visualization**

**Project 1 Report: Social Media Sentiment Analysis and Visualization**

**Team Members:**  
Lohitha Konidena  
Sai Sathvika Potluri  
Sai Karthik Nimmagadda

**1. Introduction**

**Project Title:** Social Media Sentiment Dashboard: A Comparative Visualization of Twitter and News Sentiments  
**Domain:** Data Visualization and Social Media Trends

**Objective:**

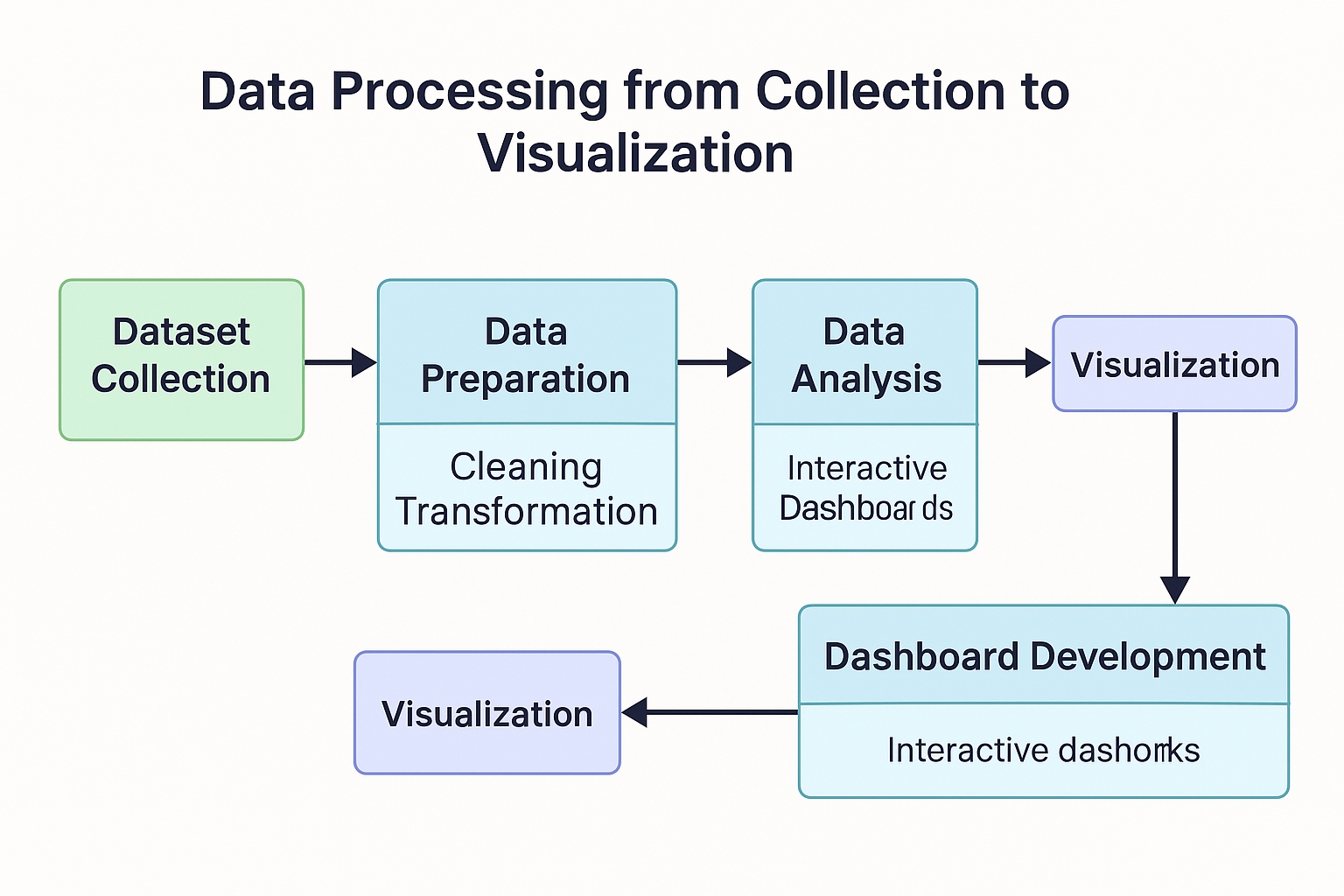
A study analyses variability in public opinions which exists between Twitter's social media platform and standard news headlines. The analysis between sentiment trends was possible through the use of Power BI and Python which created visualizations that revealed information about emotional tone together with platform biases along with sentiment alterations throughout time. This document details the procedure for preparing data along with visual development phases and analysis findings.

**2. Problem Statement:**

The analysis of sentiment transformation between Twitter platforms and traditional news media proves useful for studying public sentiments together with media tonal changes. This research investigates the temporal differences in sentiment distribution between Twitter and News platforms while exploring the detectable data patterns.

**3. Methodology:**

**Workflow Diagram:**



1. Data Collection (Twitter + News)
2. Data Cleaning & Preprocessing
3. Sentiment Mapping (Binary Classification)
4. Merging and Transformation
5. Dashboard Development (Power BI + Python)
6. Visualization and Insight Generation

**Data Cleaning and Integration:**

* Null and duplicate values were removed from both datasets.
* Column headers were renamed for consistency.
* A unified column structure was created to merge both sources.

**Sentiment Mapping:**

* Twitter sentiment labels (0/4) were mapped to “Negative” and “Positive”.
* News headlines were analyzed using rule-based mapping into binary sentiments.

**Exploratory Data Analysis (EDA):**

* Distribution of sentiment across platforms and over time.
* Temporal pattern analysis using line charts and stacked visuals.

**Visualization Strategy:**

* Power BI was used for dashboard and filterable interactive visuals.
* Python (Seaborn/Matplotlib) created more flexible custom plots.
* Text analysis performed using WordCloud to understand context.

**Insight Extraction:**

* Used counts, groupings, and time-based rollups to compare trends.
* Filtered by Source and Sentiment in various visual forms.

**4. Data Abstraction:**

**Dataset Descriptions:**

Sentiment140 Dataset (Twitter Data)

**Source:** Kaggle (<https://www.kaggle.com/datasets/kazanova/sentiment140> )

**Size:** 1.6 million tweets

**Attributes**: target, id, date, query, user, text

**Label Meaning:**

0 = Negative sentiment

4 = Positive sentiment

The model training process begins with this dataset and it also enables the creation of time-based sentiment visualization for Twitter data.

**Custom News Headlines Dataset**

**Source**: Public/open-access dataset containing headline and date fields

**Size**: ~15,000 headlines

**Attributes**: headline, date

**Usage of Datasets:**

* The Sentiment140 (Twitter) dataset served to evaluate public opinions within Twitter posts. The existing sentiment labels in this dataset included negative and positive as well as a numeric rating range from zero to four.
* The News Dataset served the purpose of detecting sentiment direction within news headlines. Person-made instructions classified textual data points into either the Positive or Negative category.
* A combined dataset resulted from cleaning and merging two datasets which received analysis through Power BI and Python visualizations**.**

**Data Transformation:**

* Removed nulls and duplicates
* Mapped sentiment values (0/4 → Negative/Positive)
* Parsed and unified dates
* The datasets were appended together for joint evaluation purposes.

**5. Task Abstraction**

**Target:**

* Analyzing the emotional divergence between information contained in News media and Twitter platforms constitutes a major objective.
* The system should identify emotional shifts that occur over time periods.
* Expose any trends in sentiment bias patterns as well as intensity fluctuations.

**Actions:**

* The analysis uses three categories of grouping and aggregation including source, sentiment, and date.
* The analysis employs line together with area charts to study temporal patterns.
* Keyword extraction using WordCloud.

**6. Implementation Using Tools**

**Tools Used:**

* Power BI Desktop
* Python (Jupyter, Matplotlib, Seaborn, WordCloud)

**Implementation Steps:**

* Loaded datasets in Power BI via Power Query Editor
* Used Python to preprocess data and generate visualizations
* Built bar, donut, line, and area charts in Power BI
* Created heatmaps, grouped bar plots, and pie charts in Python

**How Visualizations Help:**

* Understand sentiment volume per platform
* Track change in sentiment over months
* Compare sentiment intensities visually
* Surface contextual clues through WordCloud

1. **Results and Analysis**

A graph with blue lines

AI-generated content may be incorrect.

**Figure 1: Sentiment Distribution by Date**

The line chart monitors Twitter sentiment changes across time in the dataset. Public response to different types of events is shown through high and low points in the chart.

A screen shot of a graph

AI-generated content may be incorrect.

Figure 2: Sentiment Comparison – Twitter vs News

The comparative area chart establishes divergent sentiment pattern changes between News and Twitter from 2003 through 2020. Twitter exhibits higher volatility since News data maintains a steady negative trend.

A graph of a bar graph

AI-generated content may be incorrect.

Figure 3: Sentiment Count by Source

The two sources show their sentiment frequency results through this grouped bar chart. The news presents primarily negative content while Twitter users demonstrate more positive-bias content distribution.

**A blue and orange squares

AI-generated content may be incorrect.**

Figure 4: Proportional Sentiment per Source

Each source displays its sentiment proportions using 100% stacked column chart visualization. Twitter presents sentiment distribution that extends beyond negative and positive while News primarily consists of negative sentiment only.

A graph with a line and a point

AI-generated content may be incorrect.

Figure 5: Sentiment Trend Over Time

The monthly trend line demonstrates that social media activity keeps growing in positive and negative sentiments due to increasing polarized engagement throughout the month.

A graph showing positive and negative results

AI-generated content may be incorrect.

Figure 6: Publish Date vs Sentiment Line Chart

The representation shows how news headlines display emotional shifts throughout successive years. Throughout most of the timeline negative sentiment consistently stands out over other sentiments.

A graph of a number of words

AI-generated content may be incorrect.

Figure 7: Top 20 Frequent Words in Tweets

This horizontal bar chart identifies the most frequently used words in tweets. Words like “just,” “good,” and “like” indicate common emotional expressions among users.

A screenshot of a computer

AI-generated content may be incorrect.

Figure 8: Combined Dataset View in Power BI

The successful fusion of News and Twitter data appears through visual screenshots from Power BI. The steps for transformation and mapping become visible within the applied steps pane.

**8. Work Management**

**Team Members:**

Lohitha Konidena – Visualizations & Power BI Logic

Sai Sathvika Potluri – Data Cleaning & Integration

Sai Karthik Nimmagadda – Python Visualizations & Report Writing

**Implementation Status:**

* Data cleaning, transformation, and visualization complete
* Report and presentation materials drafted

**9. Next Steps for Project 2**

* Integrate additional platforms (Reddit, YouTube)
* Fully interactive dashboards should be constructed using either D3.js or Plotly.
* The process of sentiment detection needs improvement through machine learning classification techniques.
* HTML/CSS + JS code will be used to design and deploy the dashboard web page.
* The use of animated visuals for tracking time-based trends should be examined.