http://doi.org/10.35784/iapgos.6478

received: 11.08.2024 | revised: 22.01.2025 | accepted: 14.02.2025 | available online: 31.03.2025

REGIONAL TRENDING TOPICS MINING FROM REAL TIME TWITTER DATA FOR SENTIMENT, CONTEXT, NETWORK AND TEMPORAL ANALYSIS

Mousumi Hasan[1], Mujiba Shaima[2], Quazi Saad ul Mosaher[1]

[1]Bangladesh Army International University of Science and Technology, Department of Computer Science and Engineering, Cumilla, Bangladesh, [2]Monroe University, Department of Computer Science and Engineering, New York, USA

Abstract: *Twitter's tremendous impact extends from national to international affairs, covering domains such as religion, entertainment, environment, and politics. Unlike other social media platforms, Twitter offers researchers an open data source for multifaceted studies, motivating us to delve into regional trending topics, analyzing associated sentiments, context, networks, and temporal patterns. Using machine learning techniques combined with the VADER algorithm, we conducted a comprehensive analysis involving text, metadata, contextual cues, media, links, and historical data. In this study, we conducted an extensive Twitter mining operation on June 6, 2024, focusing on the ten most developed countries to explore sentiments associated with sustainable technology in industry. The insights derived from this research are pivotal for policymakers, industry stakeholders, and researchers, offering a nuanced understanding of public opinion on sustainable technology. Our findings underscore the potential of social media mining as a powerful tool for gauging public sentiment and informing strategic decision-making in the realm of sustainable industrial practices.*

Keywords: twitter mining, sentiment analysis, network analysis, temporal analysis

WYKRYWANIE TRENDÓW REGIONALNYCH NA PODSTAWIE DANYCH Z TWITTERA W CZASIE RZECZYWISTYM W CELU ANALIZY NASTROJÓW, KONTEKSTU, SIECI I CZASU

Streszczenie. *Ogromny wpływ Twittera rozciąga się od spraw krajowych do międzynarodowych, obejmując dziedziny takie jak religia, rozrywka, środowisko, i polityka. W przeciwieństwie do innych platform mediów społecznościowych, Twitter oferuje badaczom otwarte źródło danych do wieloaspektowych badań, motywując nas do zagłębiania się w regionalne tematy trendów, analizowania powiązanych nastrojów, kontekstu, sieci i wzorców czasowych. Wykorzystując techniki uczenia maszynowego w połączeniu z algorytmem VADER, przeprowadziliśmy kompleksową analizę obejmującą tekst, metadane, wskazówki kontekstowe, media, linki i dane historyczne. W tym badaniu przeprowadziliśmy szeroko zakrojoną operację eksploracji Twittera 6 czerwca 2024 r., Koncentrując się na dziesięciu najbardziej rozwiniętych krajach w celu zbadania nastrojów związanych ze zrównoważoną technologią w przemyśle. Spostrzeżenia uzyskane z tych badań mają kluczowe znaczenie dla decydentów, interesariuszy branżowych i badaczy, oferując zniuansowane zrozumienie opinii publicznej na temat zrównoważonych technologii. Nasze ustalenia podkreślają potencjał eksploracji mediów społecznościowych jako potężnego narzędzia do oceny nastrojów społecznych i informowania o podejmowaniu strategicznych decyzji w dziedzinie zrównoważonych praktyk przemysłowych.*

Słowa kluczowe: twitter mining, analiza nastrojów, analiza sieciowa, analiza czasowa

Introduction

The motivation for this study arises from the growing importance of social media platforms, particularly Twitter, as dynamic sources of real-time data that reflect public opinions, sentiments, and trending topics. Twitter’s ability to capture localized and global discussions offers a unique opportunity for deriving actionable insights, yet challenges remain in effectively analyzing the vast and diverse data it generates. Regional variations in trending topics, in particular, have been underexplored, limiting our understanding of localized concerns and cultural nuances. Additionally, most existing studies focus on singular aspects such as sentiment or temporal analysis, rather than integrating multiple dimensions to provide a holistic view. This research addresses these gaps by aiming to develop a comprehensive framework that combines Sentiment, Context, Network, and Temporal (SCNT) analyses for regional trending topic detection.

The primary aim of this study is to establish a robust framework for mining real-time Twitter data to uncover trending topics grouped by region. By integrating SCNT analyses, the study seeks to provide a deeper understanding of public sentiment, contextual themes, user interactions, and temporal patterns within Twitter discussions.

Previous studies have demonstrated various methods for detecting and tracking hot topics on social media platforms. For instance, Elbagir and Yang [6] employed the VADER sentiment analysis tool in combination with the Natural Language Toolkit to analyze Twitter sentiment, highlighting the importance of robust tools in sentiment analysis. Qi and Shabrina [13] compared lexicon-based and machine-learning-based approaches to sentiment analysis, underscoring the strengths and weaknesses of different methodologies.

The use of ensemble techniques for sentiment analysis has shown promising results in enhancing accuracy and reliability. Nabizath [10] developed an ensemble classification system for Twitter sentiment analysis, demonstrating improved performance compared to individual classifiers. This research adopts a similar ensemble approach to ensure comprehensive sentiment evaluation.

Analyzing the temporal and contextual aspects of Twitter trends involves sophisticated methods such as Latent Dirichlet Allocation (LDA), which has been utilized in various studies for topic modeling and trend analysis. Heo and Yang [7] applied LDA to analyze research topics and trends related to COVID-19 in Korea, while Du et al. [5] improved LDA for tracking hot topics on micro-blogs. Negara et al. [11] also employed LDA for topic modeling of Twitter data, showcasing its versatility in different contexts.

Real-time detection and description of trending topics have been addressed by several researchers. Mediayani et al. [9] explored data streaming methods using "R" language to determine trending topics on Twitter, highlighting the need for real-time processing capabilities. Madani et al. [8] focused on the real-time detection and description of trending topics from Twitter content, illustrating the dynamic nature of social media trends.

In addition to temporal analysis, social network analysis (SNA) provides a framework for understanding the relationships and interactions between users discussing trending topics. Ahmed et al. [2] and Ahmed et al. [1] conducted SNA on Twitter data related to COVID-19 conspiracy theories and vaccine-related discussions, respectively, revealing the network structures and key influencers within these discussions.

Furthermore, spatial and temporal sentiment analysis can offer insights into the geographical and temporal distribution of sentiments. Song and Jianhong [15] conducted spatial and temporal sentiment analysis of Twitter data, demonstrating the value of integrating geographic information with sentiment analysis. Contextual analysis, as demonstrated by Bilbao-Jayo and Almeida [4], enhances the understanding of political discourse on Twitter by incorporating contextual factors. Nurrahmi et al. [12] utilized network visualization for context analysis, providing a visual representation of the relationships within Twitter data.

This research builds on these foundational studies to provide a comprehensive analysis of trending topics on Twitter. By integrating sentiment analysis, networking, temporal, and contextual analyses, it aims to offer a holistic view of the dynamics of Twitter trends, contributing to the broader understanding of social media behavior and public sentiment.

Finally, we compile the insights gained from the analysis into a comprehensive report. This includes:
Key findings and their implications, Visual representations of data and analysis and Recommendations based on the findings.

1. Related works

The analysis of Twitter data has garnered significant attention in recent years, with numerous studies exploring various facets of sentiment analysis, topic modelling, and trend detection. Sentiment analysis on Twitter data is a well-researched domain, with various approaches being developed to enhance accuracy and reliability. Elbagir and Yang [6] utilized the VADER sentiment analysis tool combined with the Natural Language Toolkit to effectively gauge sentiments expressed in tweets. Their work underscores the utility of robust, pre-trained models for sentiment classification. Similarly, Qi and Shabrina [13] conducted a comparative study of lexicon-based and machine-learning-based approaches, finding that each method has distinct advantages depending on the context and nature of the data. The use of ensemble techniques has been shown to improve the performance of sentiment classifiers. Nabizath [10] developed an ensemble classification system for Twitter sentiment analysis, which combines multiple classifiers to achieve better accuracy compared to individual models. This ensemble approach is particularly relevant for the current research, aiming to leverage the strengths of various models for more reliable sentiment analysis. Saif et al [14] explored semantic sentiment analysis of Twitter data, integrating semantic features to improve sentiment classification accuracy. Their work complements the ensemble approaches discussed earlier, highlighting the importance of feature-rich models. Additionally, Asur and Huberman [3] analyzed the predictive power of social media, demonstrating how Twitter data can be used to predict real-world outcomes, such as box office revenues. Topic modeling is another critical aspect of Twitter data analysis, enabling the extraction and tracking of trending topics. Heo and Yang [7] applied Latent Dirichlet Allocation (LDA) to analyze research topics and trends related to COVID-19 in Korea, demonstrating the efficacy of LDA in uncovering latent topics in large datasets. Du et al. [5] further enhanced LDA for tracking hot topics on micro-blogs, incorporating improvements that make the model more suitable for dynamic social media data. Negara et al. [11] employed LDA for topic modeling of Twitter data, showcasing its versatility across different contexts. The identification of trending topics in real-time is crucial for timely insights. Mediayani et al. [9] explored data streaming methods in R to determine trending topics on Twitter, emphasizing the need for efficient processing capabilities. Madani et al. [8] focused on real-time detection and description of trending topics, illustrating the dynamic nature of social media trends and the importance of swift analytical methods.

Social network analysis (SNA) provides a framework for understanding the relationships and interactions between users discussing trending topics. Ahmed et al. [2] conducted an SNA on Twitter data related to COVID-19 conspiracy theories, revealing the network structures and key influencers within these discussions. Another study by Ahmed et al. [1] analyzed Twitter data related to blood clots and vaccines, highlighting how SNA can uncover the spread of health-related misinformation.

Temporal and spatial sentiment analysis offers additional layers of insights. Song and Jianhong [15] demonstrated the value of integrating geographic information with sentiment analysis, providing a nuanced understanding of the spatial distribution of sentiments. Bilbao-Jayo and Almeida [4] enhanced political discourse analysis on Twitter by incorporating contextual factors, while Nurrahmi et al. [12] utilized network visualization for context analysis, visually representing the relationships within Twitter data.

2. Methodology and implementation details

In this research we have analyzed the real time twitter data in five areas: uncovering trending topics, sentiment analysis network analysis, temporal analysis and context analysis. In all these analysis areas two steps are common those are data collection and data preprocessing. After these two parts for each type of analysis we experienced various methodologies to find the exact information.

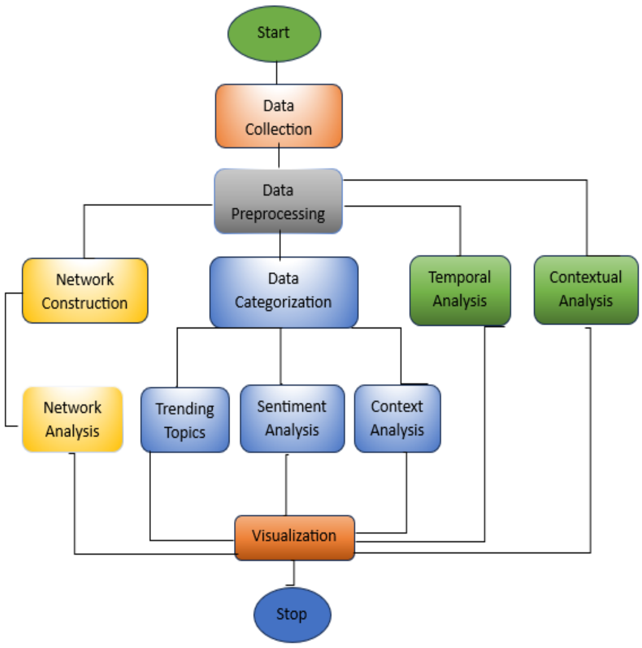

Fig. 1. Work flow diagram of the proposed methodology

The workflow for the study is illustrated in Fig. 1, showcasing the end-to-end process of analyzing Twitter data. The process begins with data collection, followed by data preprocessing, ensuring the data is clean and ready for analysis. The preprocessed data is categorized into trending topics, sentiment analysis, and context analysis under the data categorization phase. Simultaneously, a network construction step is performed to build a graph representation of user interactions, which feeds into the network analysis phase. Complementary analyses include temporal analysis and contextual analysis, which provide deeper insights into trends over time and their contextual significance. The final step involves visualization, presenting the findings in a comprehensive manner for interpretation, and marks the conclusion of the workflow.

2.1. Data collection

The first step in our research involves the collection of Twitter data. We utilize the Twitter API to gather tweets relevant to our study. The parameters for data collection include specific hashtags, keywords, and user handles that align with our research objectives. The collected data includes tweet content, metadata (e.g., tweet ID, user ID, timestamp), and user profile information.

2.2. Data pre-processing

Once the data is collected, we preprocess it to ensure quality and consistency. Preprocessing steps include:
Tokenization: Splitting the text into individual tokens (words).
Stop word removal: Eliminating common words (e.g., "the," "and") that do not contribute to the analysis.
Stemming/Lemmatization: Reducing words to their base or root form.
Noise removal: Removing URLs, special characters, and numbers that do not contribute to the sentiment or context analysis.
Mathematically, if T represents the set of all tweets, preprocessing can be denoted as:

$$T' = \text{Preprocess}(T) \tag{1}$$

where T' is the set of pre-processed tweets.

2.3. Data categorization

The preprocessed data is categorized into three main areas:
Trending Topics: Identifying the most frequently mentioned topics using techniques like Term Frequency-Inverse Document Frequency (TF-IDF) and Latent Dirichlet Allocation (LDA).

$$\text{TF-IDF}(t,d) = \text{TF}(t,d) \times \text{IDF}(t) \tag{2}$$

where $\text{TF}(t,d)$ is the frequency of term t in document d, and $\text{IDF}(t)$ is the inverse document frequency of term t.
Sentiment analysis: Sentiment analysis, also known as opinion mining, is the computational study of opinions, sentiments, and emotions expressed in text. It involves determining the sentiment orientation (positive, negative, neutral, or mixed) of a sentence, document, or set of documents. Here is a general approach to defining the sentiment of a sentence.
Pre-processing the Text includes.
Tokenization: Breaking down the text into individual words or tokens.
Normalization: Converting all text to a standard format (e.g. lowercasing).
Stop words removal: Eliminating common words that do not contribute to sentiment (e.g. "and", "the").
Stemming / Lemmatization: Reducing words to their base or root form (e.g. "running" to "run")
Feature extraction: Bag of words: Representing the text as a collection of words and their frequencies.
TF-IDF (Term Frequency-Inverse Document Frequency): Weighing terms based on their frequency in a document relative to their frequency across all documents.
Word embedding: Using vector representations of words (e.g., Word2Vec, GloVe, Bidirectional Encoder Representations from Transformers BERT) to capture semantic meaning.
Sentiment lexicons: Utilizing predefined dictionaries of words labeled with their sentiment orientations (e.g. positive & negative)
Example lexicons: SentiWordNet, AFINN, VADER (Valence Aware Dictionary for Sentiment Reasoning).
Machine Learning / Deep Learning Models:
Supervised learning: Training models using labeled datasets where each sentence is tagged with a sentiment label.
Common algorithms: Naive Bayes, Support Vector Machines (SVM), Random Forest, Logistic Regression. All these are implemented in building our ensemble technique.
Deep learning approaches: Utilizing neural networks for more complex feature extraction and sentiment classification.
Common models: Recurrent Neural Networks (RNN), Long Short-Term Memory Networks (LSTM), Convolutional Neural Networks (CNN), Transformer-based models (e.g., Bidirectional Encoder Representations from Transformers BERT).
Sentiment scoring:
Polarity Scores: Assigning a score to the sentence indicating its sentiment polarity (positive, negative, neutral).
Compound Scores: VADER algorithm provides a compound score that aggregates the polarity scores into a single value.
Contextual analysis: Considering the context and nuances in the text to improve sentiment accuracy (e.g., handling negations, sarcasm, and irony).
Example sentiment analysis process: "Excited about Euro 2024! Can't wait to see my favorite teams play."
Sentiment analysis process:
Preprocessing the text: Tokenization: Breaking down the tweet into individual words or tokens.
Tokens: ["Excited", "about", "Euro", "2024", "Can't", "wait", "to", "see", "my", "favorite", "teams", "play"]
Normalization: Converting all text to lowercase.
Tokens: ["excited", "about", "euro", "2024", "can't", "wait", "to", "see", "my", "favorite", "teams", "play"]
Stop words removal: Eliminating common words that do not contribute to sentiment.
Tokens: ["excited", "euro", "2024", "wait", "see", "favorite", "teams", "play"]
Stemming/Lemmatization: Reducing words to their root form.
Tokens: ["excite", "euro", "2024", "wait", "see", "favorite", "team", "play"].
Feature extraction: Using **TF-IDF** or **Word embeddings** to represent the tokens. For simplicity, let's use a sentiment lexicon.
Sentiment lexicons: Utilizing a predefined sentiment dictionary like VADER or AFINN. Example Sentiment Scores: "excite" (positive) ,"favorite" (positive). Other words may not be in the sentiment dictionary or may be neutral.
Sentiment scoring: Positive words: "excite" (positive score: +3), "favorite" (positive score: +2). **Neutral words:** Words like "euro", "2024", "wait", "see", "team", "play" may not contribute significantly to sentiment if they are neutral.
Overall sentiment classification: Aggregate the scores:
Positive: +5, Negative: 0. Compound score calculation (for a model like VADER): Let's assume a combined score based on weights.

Final Sentiment: **Positive** due to the presence of positive words like "excited" and "favorite".

The example tweet "Excited about Euro 2024! Can't wait to see my favorite teams play." expresses a **positive sentiment**. The tweet contains positive words like "excited" and "favorite," which contribute to an overall positive sentiment classification.

By following this process, sentiment analysis models can classify the sentiment of tweets accurately, providing insights into public opinions and trends.

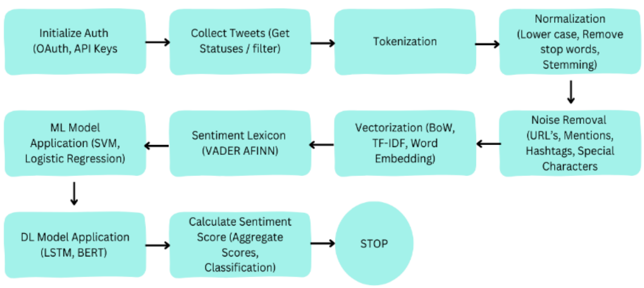

Fig. 2. Flow chart for sentiment analysis

The detailed sentiment analysis workflow is depicted in Figure 2, illustrating the step-by-step methodology.

Context analysis: To determine the context of a tweet, a comprehensive methodology is employed involving text analysis, metadata analysis, contextual cues, media and links evaluation, and historical data examination. Text analysis includes extracting important keywords, assessing sentiment, recognizing entities like people or organizations, and analyzing hashtags for trends. Metadata analysis considers user information, timestamps, and geolocation data. Contextual cues are derived from replies, retweets, mentions, and tags to understand the broader conversation. Media and links are analyzed by examining attached images, videos, and URLs to identify relevant content. Historical data is also crucial, as it involves reviewing the user's past tweets and checking for current trending topics related to the tweet. By integrating these elements, a comprehensive understanding of the tweet's context is achieved.

For example, let's consider the following tweet for context analysis:

Tweet: "Excited for the new @Apple launch event! #AppleEvent #TechNews"

1. **Text analysis**:
   Keywords: "Excited," "new," "launch event"
   Sentiment: Positive
   Entity recognition: "Apple"
   Hashtag analysis: #AppleEvent, #TechNews
2. **Metadata analysis**:
   User information: Suppose the user is a tech blogger with a large following.
   Timestamp: Posted on the day of an Apple event.
3. **Contextual cues**:
   Replies and retweets: People are discussing the latest Apple products in replies.
   Mentions and tags: The user mentioned "@Apple," indicating the official Apple account.
4. **Media and links**:
   Images and videos: The tweet might include an image or video teaser from the event.
   URLs: A link to the live stream of the event.
5. **Historical data**:
   Previous tweets: The user has previously tweeted about tech events and product launches.
   Trending topics: #AppleEvent is trending, indicating widespread interest.

3. Network construction for common trending topics

We construct a network graph where nodes represent users and edges represent interactions between them (e.g., retweets, mentions, replies). The construction involves:

Nodes: Each unique user in the dataset is represented as a node.

Edges: An edge is created between two nodes if there is an interaction between them.

If U is the set of users and I is the set of interactions, the network G can be represented as:

$$G = (U,I) \tag{3}$$

The constructed network is analyzed to gain insights into the structure and dynamics of interactions.

To analyze the structure and dynamics of interactions in this network, we utilize a Graph Neural Network (GNN), a deep learning architecture designed specifically for graph-structured data. The GNN architecture used in this study consists of the following components:

Graph Input Layer:

The constructed network graph $G=(U,I)$ serves as the input to the GNN. Each node $u \in U$ is initialized with a feature vector that may include user-specific attributes such as interaction counts, sentiment scores, or activity levels.

Graph Convolutional Layers:

Multiple graph convolutional layers are applied to propagate information between nodes. These layers aggregate features from neighbouring nodes, allowing the network to learn both local and global patterns in the graph.

Pooling and Readout Layers:

After feature aggregation, pooling layers summarize the learned node representations into a single graph-level representation. Techniques such as mean pooling, max pooling, or attention-based pooling are used to capture essential structural properties of the graph.

Fully connected layers:

The pooled graph representation is passed through fully connected layers to perform downstream tasks, such as classification, prediction, or clustering.

Output layer:

The final output layer produces the desired results, such as centrality metrics, interaction predictions, or node classifications.

Key metrics include:

Centrality: Measures the importance of nodes within the network using metrics like Degree Centrality, Betweenness Centrality, and Closeness Centrality.

$$\text{Degree centrality}(v)=\frac{\deg(v)}{n-1} \tag{4}$$

where $\deg(v)$ is the degree of node v and n is the total of nodes.

Community detection: Identifying clusters or communities within the network using algorithms like Girvan-Newman or Louvain method.

Sentiment propagation: Analyzing how sentiment spreads across the network.

Temporal analysis examines how trends, sentiments, and interactions change over time. This involves:

Plotting time series of tweet volumes, sentiment scores, and interaction counts.

Identifying peaks and troughs in activity and correlating them with real-world events.

If T is the set of timestamps, temporal trends can be model as:

$$f(T)=\{V(t),S(t),I(t)\} \tag{5}$$

where $V(t)$ is tweet volume, $S(t)$ is sentiment score, and $I(t)$ is interaction count at time t.

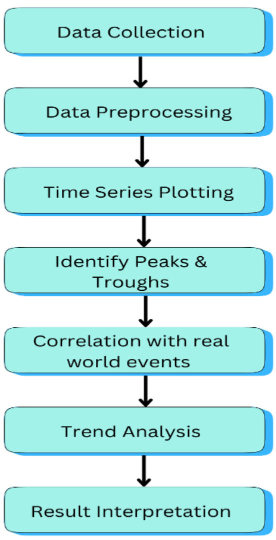

Fig. 3. Steps of temporal analysis

The temporal analysis workflow is outlined in Figure 3, illustrating the sequence of steps to analyze trends over time.

Temporal analysis of the tweet, "What an incredible goal by [Player]! This match is heating up! #WorldCupFinal #Soccer," collected at the timestamp 2024-08-01 20:15 PM, can be described as follows: The tweet, along with its timestamp and associated metadata such as likes, retweets, and replies, is gathered for analysis. Additionally, other tweets with similar hashtags like #WorldCupFinal and #Soccer are included to capture the broader conversation. The timestamp is parsed into a suitable time format, and the sentiment of the tweet is assessed as positive due to the enthusiastic tone regarding the goal.

In the time series analysis, the tweet volume (V(t)) related to #WorldCupFinal is plotted over time. The analysis indicates that tweet volume likely increases gradually leading up to the match, with notable spikes during key moments like goals, halftime, and the final whistle. A significant peak occurs at 20:15 PM, coinciding with the goal. Sentiment scores (S(t)) are also plotted, revealing fluctuations with positive spikes following favorable match events like goals and possible dips after controversial moments. At 20:15 PM, a positive spike is observed, reflecting the excitement surrounding the goal. Interaction counts (I(t)), including likes, retweets, and replies, are similarly plotted, showing a sharp rise at key moments. At 20:15 PM, interactions surge, indicating high engagement following the goal.

Peaks and troughs in tweet volume, sentiment, and interactions are identified, with a clear peak at 20:15 PM corresponding to the goal. Smaller peaks may also be observed during other significant match events. The peak in activity at 20:15 PM is directly linked to the goal in the World Cup Final, an emotionally charged event for viewers. Other peaks might align with halftime, the match's conclusion, or other crucial moments.

Trend analysis over the match duration reveals that tweet volume, sentiment, and interaction counts fluctuate in response to key events. Post-match analysis may show sustained conversation, particularly if the goal was decisive. Sentiment analysis may also uncover the overall mood of fans, with positive sentiment during favorable plays and negative sentiment during contentious moments.

In conclusion, the temporal analysis demonstrates that the goal at 20:15 PM was a pivotal moment, driving a surge in online conversation and engagement. The positive sentiment and high interaction counts suggest that the goal was well-received by fans. This analysis provides valuable insights into audience reactions during live sports events, helping broadcasters and sports marketers understand what resonates with fans and plan content accordingly.

Contextual analysis delves deeper into the content of tweets to understand the underlying themes and narratives. Techniques include topic modelling to identify prevalent themes and Semantic analysis to understand the meaning and implications of tweets.

The results from the above analyses are effectively visualized using a range of tools and techniques to enhance data interpretation. Graphs and charts are employed to illustrate trends, distributions, and comparisons, providing a clear representation of key insights. Network diagrams are utilized to map relationships and connections within the dataset, revealing structural patterns. Additionally, word clouds highlight the most frequently occurring terms and topics, offering a quick overview of dominant themes. These visualization techniques collectively enhance the understanding of complex data, making it more accessible and actionable.

4. Implementation

Our Twitter data mining methodology follows a structured, step-by-step approach, utilizing various technologies and programming tools to extract and analyze data effectively. The process begins with data collection, where we use the Twitter API and Python's Tweepy library to fetch tweets based on specific keywords, hashtags, and user handles, ensuring the dataset is tailored to our analysis needs. Following this, we preprocess the data using NLTK and SpaCy, performing tokenization, stop word removal, and stemming to prepare the text for analysis. Data categorization is then conducted using TF-IDF to assess word importance and LDA for topic modeling, helping to uncover underlying themes in the tweets. Sentiment analysis is performed with TextBlob and more advanced models from the Transformers library, providing insights into the emotional tone of the tweets. We then construct a network graph with NetworkX, where nodes represent users and edges represent interactions, allowing us to visualize connections and identify key influencers within the Twitter network. This network is further analyzed using centrality measures and community detection algorithms to reveal the structure and dynamics of the interactions. Temporal analysis is conducted using Matplotlib and Pandas to plot time series graphs, which helps us observe trends and shifts in tweet activity over time. Contextual analysis is enhanced with advanced NLP techniques, employing Gensim for further topic modeling and BERT for semantic analysis, offering deeper insights into the context of the tweets. The results are then visualized through interactive plots and network diagrams using Plotly and Gephi, making the data more accessible and interpretable. Finally, we compile all findings and visualizations into a comprehensive report, providing actionable insights that can guide decision-making processes. This detailed methodology ensures a thorough analysis of Twitter data, enabling the extraction of meaningful trends and patterns. Implementation also follows the Figure 1.

5. Experimental results and discussion

As shown in the Figure 1, the aim of this research is to find trending topics, sentiment and context of the trending topics and then generate network graphs for network and temporal analysis. Keeping this aim on focus we collected data on 06-06-2024 from twitter of developed 10 countries.

In the mentioned day the European Football Championship was in the most trend i.e. sports. Then politics, corporate, health, entertainment and so on. 19 contexts are found from collected data.

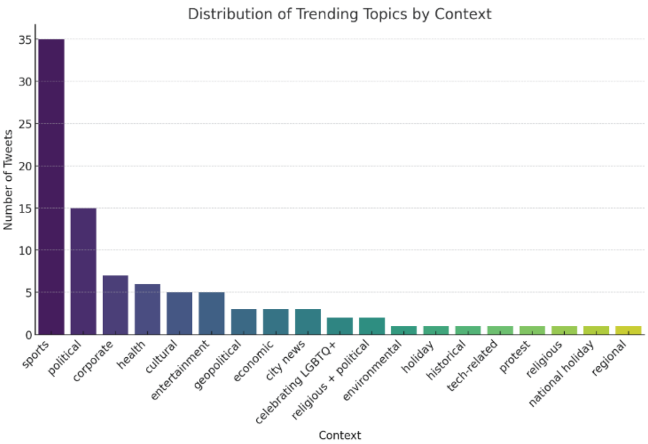

Fig. 4. Distribution of trending topics by context

Figure 4 shows the distribution of trending topics by context, categorized into different types and measured by the number of tweets. Here is a summary of the results:

- Sports: Dominates the chart with the highest number of tweets, in total around 35k.
- Political: The second most frequent context, with slightly under 20k tweets.
- Corporate: Significant presence with around 10k tweets.
- Health: Moderately discussed with about 8k tweets.
- Cultural: Similar to health, with around 7k tweets.
- Entertainment: Also has around 6k tweets.
- Geopolitical: Comparable to entertainment, with around 5k tweets.

- Economic, City News, Celebrating LGBTQ+, Religious + Political: These contexts each have around 4k tweets.
- Environmental: Slightly fewer mentions, with around 3k tweets.
- Holiday, Historical, Tech-related, Protest, Religious, National Holiday, Regional: These contexts have the least mentions, with each having around 1k to 2k tweets.

In summary we found that Sports and political topics are the most prominent trending contexts, receiving the highest number of tweets. Corporate and health topics follow, with a moderate level of discussion. Cultural, entertainment, and geopolitical topics also see a fair amount of engagement. Economic, city news, and LGBTQ+ celebrations receive some attention, though less than the previously mentioned categories. Environmental topics see minimal discussion, and holiday, historical, tech-related, protest, religious, national holiday, and regional contexts have the least engagement.

This distribution indicates that sports and political events are the most engaging topics on social media, while other contexts receive varying levels of attention.

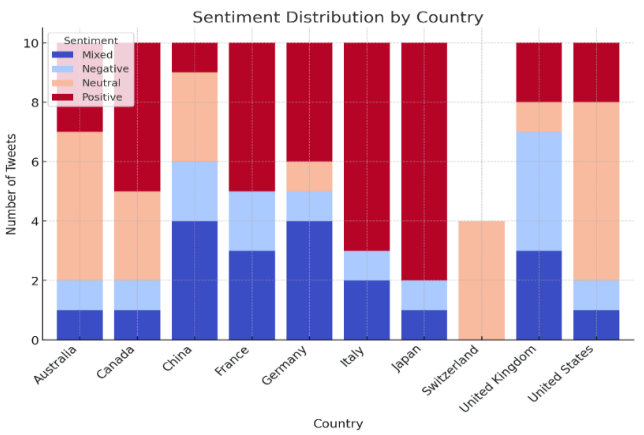

Fig. 5. Sentiment distribution by country

In Figure 5 the bar chart shows the sentiment distribution of trending topics by country. Key observations include:

The United States and United Kingdom have a significant number of negative and mixed sentiments, particularly around political and geopolitical topics.

Japan and Canada display more positive sentiments, especially around entertainment and sports.

China and Germany show a mix of sentiments, reflecting a balance of different types of trending topics.

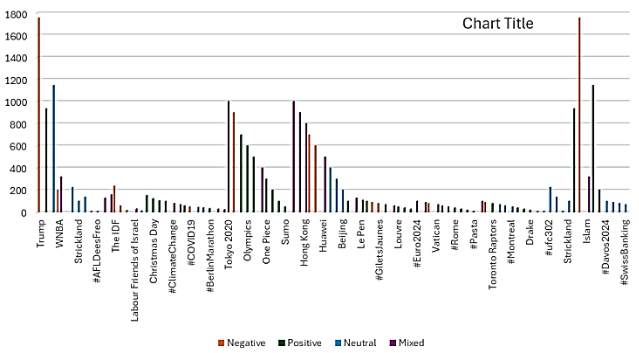

Fig. 6. Sentiment distribution across trending topics

Figure 6 visualizes sentiment analysis across various trending topics, with sentiment types differentiated by color: orange for Negative, green for Positive, blue for Neutral, and purple for Mixed.

Here is a summary of the results:

1. **Trump**: Dominated by Negative sentiment. Moderate amount of Positive sentiment. Low Neutral and Mixed sentiment.
2. **WNBA**: Predominantly Positive sentiment. Significant Neutral sentiment. Moderate Negative and low Mixed sentiment.
3. **AFLDeesFreo**, **The IDF**, **Labour Friends of Israel**: Low overall mentions with a mix of all sentiments, mostly Neutral and Negative.
4. **Climate change, #COVID19:** Predominantly Neutral and Negative sentiments. Low Positive sentiment and almost no Mixed sentiment.

And so on for other topics.

The network graph generated in Figure 7 displays the relationships between countries and their trending topics. Each country and topic are represented as a node, and the edges (connections) represent the number of tweets associated with each topic in the respective countries.

Here is a summary of the insights we can derive from this network analysis:

1. **Central nodes:**
   The United States and **Australia** are highly connected, indicating a diverse range of trending topics.
   Trump appears as a highly central topic, trending in multiple countries.
2. **Clusters:**
   There are visible clusters representing groups of topics that are trending in specific countries. For example:
   Sports-related topics (e.g., Real Madrid, WNBA) form a cluster around the United States and other countries.
   Political and geopolitical topics (e.g., Macron, Xi Jinping) cluster around their respective countries.
3. **Country-specific trends:**
   Japan has a strong cluster of entertainment and sports topics (e.g., Tokyo 2020, Demon Slayer).
   China shows a mix of political and economic topics (e.g., Xi Jinping, Chinese Economy).
4. **Topic-specific trends:**
   Topics like **COVID-19** and **climate change** are trending in multiple countries, reflecting global issues.

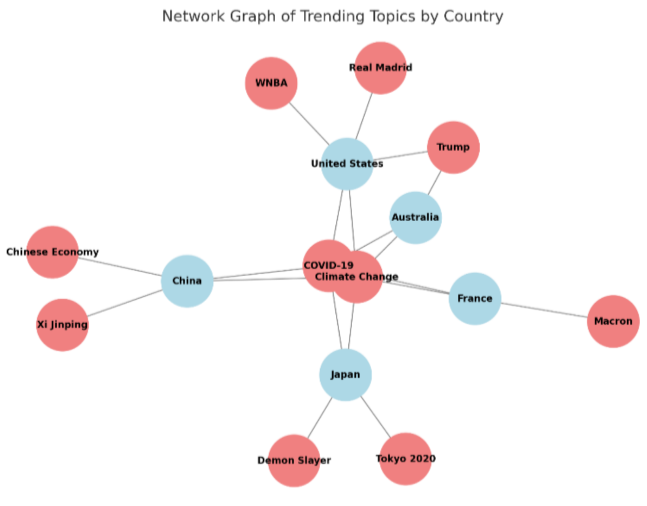

Fig. 7. Network graph of trending topics over countries

From the visual representation of Fig. 8, the clustering involves grouping related keywords, hashtags, or topics based on their relationships or co-occurrence in data. This type of clustering is called: Community Detection in Network Graphs Type: Community clustering or topic clustering in a graph structure.

Purpose: Identifying clusters of highly interconnected nodes (e.g., hashtags, users, or topics) based on their relationships, such as mentions, retweets, or shared themes.

Figure 8 shows a temporal analysis is given are in short:

1. **Clustered trends**:

   Topics are clustered together, suggesting they are being discussed concurrently or within similar time frames Geographically or contextually related topics tend to form tight clusters.

2. **Key clusters**:

   Cluster 1: (Top Left) Includes Canadian and Italian topics: Canada-related trends: Alberta, Drake, Justin Bieber. Italy-related trends: Pasta, Rome, Vatican, Ferrari, Juventus, Milan, Ronaldo.

   Cluster 2: (Top Center) Canadian and sports-related topics: Toronto Raptors, Hockey, Montreal.

   Cluster 3: (Bottom Left) European and political topics: European Union, #UCLFinal, Angela Merkel, Climate Change, #BerlinMarathon.

   Cluster 4: (Top Right) Asian and tech-related opossum, Nintendo, Ohtani, Shinzo Abe, Tokyo 2020.

   Cluster 5: (Center Right) International political and conflict-related topics: Israel, Afghanistan, Christmas Day, Liverpool, Gilets Jaunes, Paris, Macron.

   Cluster 6: (Bottom Right) Financial and corporate trends: UBS, Swiss Banking, Davos, Switzerland.

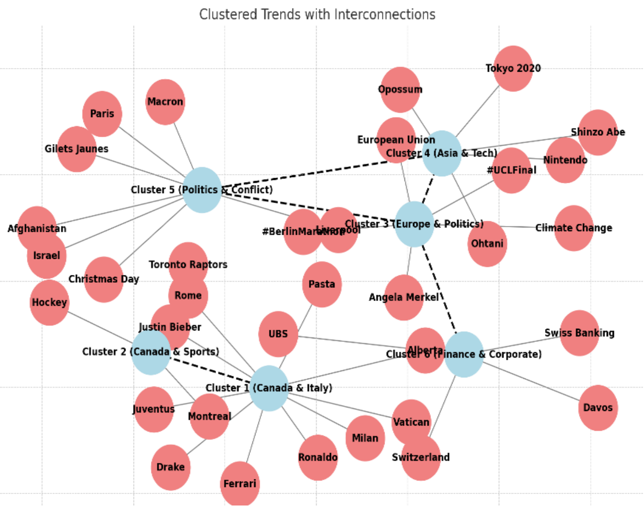

Fig. 8. Temporal Network graph of trending topics over countries

In summary of temporal analysis of figure 6 we concluded that The **COVID-19** topic appears in multiple clusters, indicating its widespread and ongoing relevance across different regions and contexts. The presence of **sports events** like Tokyo 2020, UEFA Champions League, and others suggests periods where these events are happening or being discussed extensively. Political topics such as **Angela Merkel**, **Macron**, and international issues like **Afghanistan** suggest ongoing or recent political developments. **Climate change** and other environmental topics are grouped, reflecting simultaneous discussions during specific events or reports.

Interconnection: The interconnectedness of some clusters (e.g., Canadian trends with general North American and European trends) indicates overlapping discussions, likely due to global events impacting multiple regions. This analysis highlights the temporal and contextual relationships between trending topics, demonstrating how different events and discussions are interlinked globally over time.

6. Discussion and comparison with related works

This research aligns with and expands upon previous work in the areas of how people feel about things on Twitter (sentiment analysis), how social networks operate (social network analysis), and the study of environmentally friendly technologies. By focusing on public opinions about sustainable technologies used in businesses across different countries, this study reveals unique trends and how these feelings shift and evolve over time (sentiment dynamics). These insights are valuable for both researchers and those involved in creating policies related to sustainable development. A comparison table is drawn bellow what will help us in understating the relativity with related references and our research.

Table 1. Network graph of trending topics over countries

Aspect	Our study	Related literature
Sentiment Analysis Technique	Combined VADER and machine learning	Elbagir and Yang [6]; Qi and Shabrina [13]; Nabizath [10]
Positive Sentiments in Sports	Japan and Canada dominate positive sentiments	Mediayani et al. [9]; Madani et al. [8]
Negative Sentiments in Politics	USA and UK show high negative sentiments	Nabizath [10]; Bilbao-Jayo and Almeida [4]
Persistent Health Concerns	Continued negative sentiments around health and environment	Ahmed et al. [1, 2]
Temporal Analysis	Short-term dynamics analyzed over six hours	Asur and Huberman [3]; Song and Jianhong [15]
Network and Context Visualization	Insights into country-specific trends	Nurrahmi et al. [12]; Negara et al. [11]; Heo and Yang [7]

From Table 1 it's easy to find that most previous research on Twitter data has typically focused on one or two areas, such as analysing people's opinions (sentiment analysis), tracking changes over time (temporal analysis), or examining how people are connected (network analysis). However, our research stands out by comprehensively exploring all four key aspects of analysing Twitter data.

7. Conclusion

In conclusion, this study highlights the profound influence of Twitter on various domains from national to international affairs, offering a unique and open data source for researchers. Our comprehensive Twitter mining operation, conducted on June 6, 2024, provided valuable insights into regional trending topics and sentiments, particularly concerning sustainable technology in industry among the ten most developed countries. The analysis revealed diverse sentiment patterns, with Japan and Canada showing more positive sentiments, while the USA and the United Kingdom exhibited significant negative and mixed sentiments. Notably, sports emerged as a universally positive topic, whereas health and environmental issues continue to draw negative sentiments, even years after the COVID-19 vaccine's introduction.

Even considering that these results really pinpoint how powerful social media mining is in relation to understanding people's attitudes and strategic decisions, there was a limitation. As temporal dynamics are thus restricted, due to data collection taking place on just one day through mining, potentially long-term patterns in sentiment would have been excluded. Moreover, the analysis was limited to Twitter and may or may not represent public opinion across other social media platforms. Sentiment analysis is a great tool but not wholly accurate, as language is inherently subjective and complex. These are factors to take into consideration when interpreting these results. These insights provide distinction to the public sentiment that is so important to policymakers, industry stakeholders, and researchers. This research informs a better comprehension of global perceptions and their impacts on sustainable industrial practices through advanced data mining and sentiment analysis techniques.

References

[1] Ahmed W. et al.: A social network analysis of Twitter data related to blood clots and vaccines. International Journal of Environmental Research and Public Health 19(8), 2022, 4584.

[2] Ahmed W. et al.: COVID-19 and the 5G conspiracy theory: social network analysis of Twitter data. Journal of medical internet research 22(5), 2020, e19458.

[3] Asur S., Huberman B. A.: Predicting the future with social media. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology 1, 2010.

[4] Bilbao-Jayo A., Almeida A.: Improving political discourse analysis on twitter with context analysis. IEEE Access 9, 2021, 104846–104863.

[5] Du Y. J. et al.: Extracting and tracking hot topics of micro-blogs based on improved Latent Dirichlet Allocation. Engineering Applications of Artificial Intelligence 87, 2020, 103279.

[6] Elbagir S., Yang J.: Twitter sentiment analysis using natural language toolkit and VADER sentiment. International Multiconference of Engineers and Computer Scientists 122(16), 2019.

[7] Heo S. M., Yang J. Y.: Analysis of Research Topics and Trends on COVID-19 in Korea Using Latent Dirichlet Allocation (LDA). Journal of The Korea Society of Computer and Information 25(12), 2020, 83–91.

[8] Madani A. et al.: Real-time trending topics detection and description from Twitter content. Social Network Analysis and Mining 5(1), 2015, 59.

[9] Mediayani M. et al.: Determining Trending Topics in Twitter with a Data Streaming Method in R. Indonesian Journal of Science and Technology 4(1), 2019, 148–157.

[10] Nabizath S.: An ensemble classification system for twitter sentiment analysis. Procedia Computer Science 132, 2018, 937–946.

[11] Negara E. S. et al.: Topic modelling twitter data with latent Dirichlet allocation method. International Conference on Electrical Engineering and Computer Science (ICECOS). IEEE, 2019.

[12] Nurrahmi H. et al.: Twitter data transformation for network visualization based context analysis. International Conference on Information and Communications Technology (ICOIACT). IEEE, 2018.

[13] Qi Y., Zahratu S.: Sentiment analysis using Twitter data: a comparative application of lexicon-and machine-learning-based approach. Social Network Analysis and Mining 13(1), 2023, 31.

[14] Saif H. et al.: Semantic sentiment analysis of twitter. The Semantic Web–ISWC 2012: 11th International Semantic Web Conference, Boston, MA, USA, November 11-15, 2012, Part I 11. Springer Berlin Heidelberg, 2012.

[15] Song Z., Jianhong C. X.: Spatial and temporal sentiment analysis of twitter data. European handbook of crowdsourced geographic information 205, 2016.

M.Sc. Mousumi Hasan
e-mail: mousumi.cse@baiust.ac.bd

Mousumi Hasan, an assistant professor at BAIUST, has industry experience in software development and holds B.Sc. and M.Sc. degrees from Jahangirnagar University. She researches image processing, machine learning, data mining, and software engineering, contributing to OBE curriculum integration. Her work includes IEEE-published papers on deep learning, network security, and software engineering.

https://orcid.org/0009-0006-7499-7669

M.Sc. Mujiba Shaima
e-mail: goodmorn_simi@yahoo.com

Mujiba Shaima, a software developer since 2011, has worked in Bangladesh's banking and pharmaceutical sectors as a developer and quality assurance engineer. She holds an MBA and is pursuing a second M.Sc. in Computer Science at Monroe College, New York, demonstrating her commitment to continuous learning. Her contributions to computer science include published articles on image processing and data mining in various journals and conferences.

https://orcid.org/0009-0004-9129-0716

B.Sc. Quazi Saad ul Mosaher
e-mail: saad.quazi44@gmail.com

Quazi Saad ul Mosaher, a software engineer at RedOrch Technology Limited, holds a B.Sc. in Computer Science from BAIUST. He specializes in image processing, pattern recognition, machine learning, and data mining, with notable research contributions. Currently, he is engaged in projects on student affairs data mining, network security, and steganography.

https://orcid.org/0009-0006-5968-5608