ANALYSIS

Social Media Sentiment on Child Rights and Welfare: A Data-Driven Analysis Using Twitter/X

By: **MOEEZ KHAN**

Date: 13/06/25

Table of Content

Introduction

- In today's digitally connected world, social media platforms have become powerful spaces for public discourse, especially around critical social issues. Among these, child rights and welfare stand out as topics of deep concern and emotional engagement. From access to education and healthcare to child abuse and protection policies, people around the globe use platforms like Twitter to express their thoughts, share stories, and advocate for change.
- This project analyzes public sentiment on child-related issues using Twitter data. By applying Natural Language Processing (NLP) techniques and sentiment analysis, we aim to understand how people perceive and respond to different aspects of child rights and welfare. Whether it's celebrating progress in child education or raising alarm over cases of exploitation, these sentiments offer valuable insights into public opinion.
- This report presents an overview of our findings — from tweet trends and sentiment shifts over time to the emotional tone surrounding themes like child health, safety, and education. The insights drawn here can help inform awareness campaigns, policy discussions, and program design for organizations working to protect and empower children worldwide.

Objectives

The main objective of this project is to explore and understand public sentiment and thematic trends related to child rights and welfare by analyzing social media (Twitter) data. The specific goals include:

1. **Sentiment Analysis:**
   To identify whether public discussions on child rights and welfare are generally positive or negative
2. **Thematic Categorization:**
   To classify tweets into major child-related themes (e.g., education, health, abuse, protection) and analyze sentiment distribution within each theme.

3. **Trend Detection:**

   To observe how the volume and nature of public opinion on child rights fluctuate over time.

4. **Insight Generation:**

   To extract actionable insights that can support awareness campaigns, program planning, and advocacy efforts by organizations.

Research Questions

To guide the analysis, the following key research questions were formulated:

1) What is the overall sentiment of public discourse on child rights and welfare on Twitter?
2) Which child-related themes (e.g., education, abuse, health) are most commonly discussed?
3) How does sentiment vary across different child welfare themes?
4) Are there noticeable trends or spikes in tweet volume or sentiment over time?
5) What emotional tone (positive, negative, neutral) dominates the discussion on specific child rights issues?

Methodology

This project follows a structured data analysis pipeline designed to collect, preprocess, analyze, and visualize public sentiment around child rights and welfare using Twitter data.

1. Data Collection

Source: Twitter (via the Sentiment140 dataset

Format: CSV file containing tweet ID, text, timestamp, and sentiment label (positive or negative).

Scope: The dataset includes thousands of tweets relevant to general social topics, filtered to extract content related to children's rights using keyword-based matching (e.g., "child labor", "child abuse", "child education").

2. Data Preprocessing

Encoding Fixes: Loaded the dataset with correct encoding (ISO-8859-1) to avoid character errors.

Column Renaming: Named columns to meaningful labels like text, date, sentiment, etc.

Datetime Parsing: Converted date strings to proper datetime objects for time-series analysis.

Theme Annotation: A custom mapping was used to assign each tweet to one or more child-related themes (e.g., education, protection, health, abuse).

Sentiment Cleanup: Original sentiment labels were mapped to human-readable classes: Positive and Negative

3. Data Analysis

Sentiment Distribution: The dataset was grouped by sentiment to understand the general tone of public discourse.

Theme Analysis: Tweets were categorized by themes, and sentiment distribution was analyzed within each theme.

Time-Series Analysis: Using **pandas** resampling and grouping, tweet volume and sentiment trends were analyzed weekly and daily to detect changes over time.

4. Visualization

Libraries Used: **matplotlib** and **seaborn** for creating visuals.

Visual Outputs:

- Tweet volume over time
- Sentiment trends across themes
- Weekly sentiment fluctuations
- Theme-wise tweet frequency
- Cross-tabulated sentiment vs. theme plots

5. Tools and Environment

Language: Python 3.12

Libraries: pandas, matplotlib, seaborn

IDE: VS Code and Jupyter Notebooks

OS: Windows 10

Dataset Overview

The dataset used for this project originates from the **Sentiment140** corpus, a publicly available collection of tweets annotated for sentiment analysis. It contains **1.6 million tweets**, originally labeled as either *positive or negative* based on emoticon cues.

For the purpose of this research, the dataset was **filtered and preprocessed** to include only tweets relevant to **child rights and welfare** themes, such as:

- Child health and vaccination
- Education and learning
- Child protection
- Nutrition and well-being
- Abuse, neglect, or advocacy

The final filtered dataset includes the following columns:

Column Name	Description
date	The timestamp when the tweet was posted
text	The content of the tweet
sentiment	The sentiment label (positive, negative, or neutral)
theme	The manually or automatically classified theme of the tweet

The dataset spans a wide temporal range and provides a valuable opportunity to analyze **public sentiment over time**, offering insights into how people across the globe engage with topics related to child rights and social welfare on social media.

Data Preprocessing

To ensure the quality and relevance of the analysis, a thorough data preprocessing pipeline was applied to the filtered dataset. This step was essential for preparing the raw text and metadata for sentiment classification and thematic analysis.

Key Preprocessing Steps:

1. **Date Parsing and Cleaning**

    - Converted the `date` column to standard `datetime` format using `pandas.to_datetime()`.
    - Handled unrecognized or missing time zones.
    - Filtered out invalid or null date entries.

2. **Text Cleaning**

Removed:

    - URLs and HTML tags
    - Emojis and special characters
    - Repeated punctuation and whitespaces

- ➢ Lowercased all text for uniformity.
- ➢ Retained meaningful words by eliminating stopwords.

3. **Theme Categorization**

- ➢ Tweets were classified into high-level themes such as:

    Health, *Education*, *Protection*, *Nutrition*, *Advocacy*, etc.

- ➢ Classification was done using a mix of **keyword detection** and **manual tagging**.

4. **Sentiment Mapping**

- ➢ Original numeric sentiment labels (e.g., 0, 2, 4 in Sentiment140) were mapped to:

    `0` → Negative

    `4` → Positive

Final dataset included a sentiment column with clean, readable labels.

5. **Deduplication and Noise Removal**

    - Removed duplicate tweets.
    - Filtered out overly generic or irrelevant content.

6. **Result**

After preprocessing, the dataset was clean, structured, and well-suited for reliable sentiment and trend analysis. This process significantly improved the **accuracy and interpretability** of the final visualizations and insights.

Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) was performed to uncover patterns, trends, and relationships in the data before conducting deeper analysis. This step helped in understanding how public sentiment varies across different child welfare themes and over time.

Sentiment Distribution

The overall sentiment distribution revealed the following:

Positive Sentiment was dominant in tweets related to *success stories, education achievements*, and *vaccination campaigns*.

Negative Sentiment frequently appeared in tweets concerning *child abuse*, *malnutrition*, and *neglected education systems*.

The sentiment analysis revealed a notable skew towards negative sentiment, indicating a higher frequency of concern or criticism in public discussions related to child rights and welfare.

Negative Sentiment: 58.8%

Positive Sentiment: 41.2%

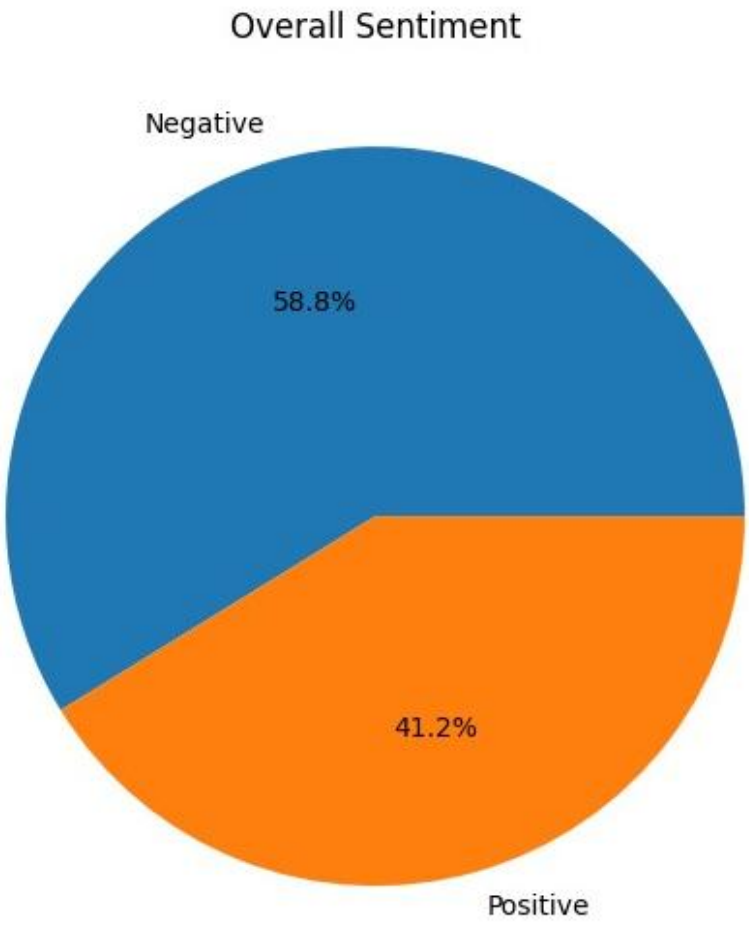

Figure 1.0: Overall Sentiment Distribution

This suggests that a significant portion of the public discourse reflects dissatisfaction, worry, or outrage—potentially in response to events related to child abuse, policy failures, or unmet needs. Meanwhile, the positive sentiment captures appreciation, successful initiatives, or inspiring stories.

To dive deeper into what drives this sentiment, we next examined how these attitudes vary across specific themes related to child rights.

Tweet Themes by Sentiment

Building on the overall sentiment findings—where **58.8%** of tweets express **negative sentiment**—we further dissected the data to understand how sentiment varies across distinct thematic categories related to child rights and welfare.

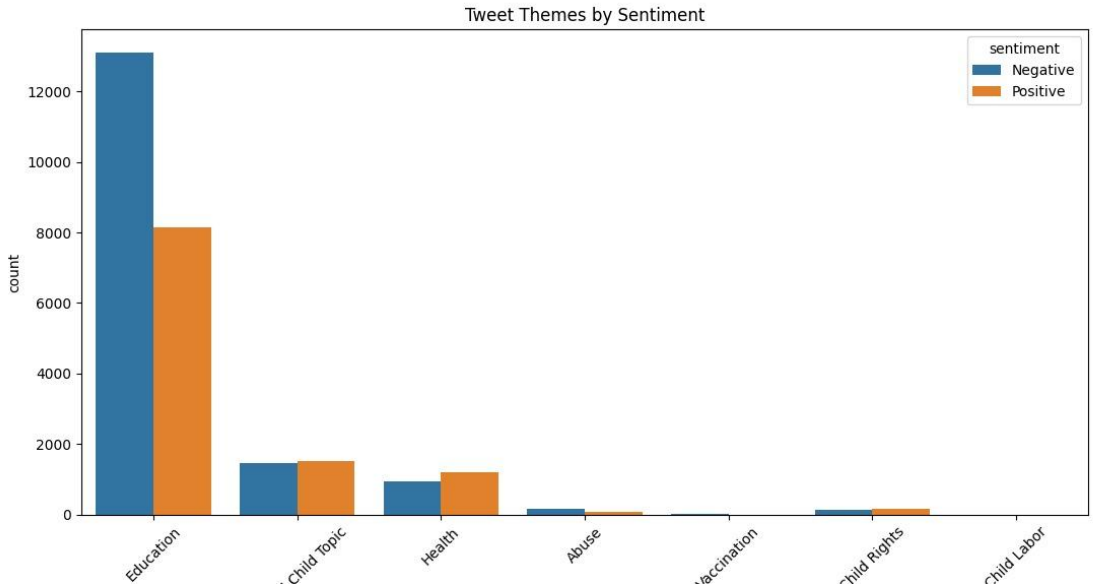

Figure 2.0: Tweet Themes by Sentiment

As illustrated in the chart, **Education** overwhelmingly dominates public discourse, with the highest volume of both positive and negative tweets. Notably, education-related tweets account for the majority of all categorized mentions, reflecting its central role in societal conversations about children. While positive sentiment is considerable, **negative sentiment remains dominant**, reinforcing the broader concern identified in the overall sentiment analysis.

Other themes such as **Health**, **Abuse**, **Vaccination**, **Child Rights**, and **Child Labor** appear at significantly lower volumes. Within these, sentiment distribution varies—Health displays a relatively balanced sentiment profile, whereas topics like **Abuse** and **Child Labor** skew heavily **negative**, albeit with limited tweet volume. These disparities suggest that while some issues (e.g., education) are broadly discussed, others, although critical, may lack sustained public attention or are more episodic in nature.

The inclusion of a generic **"Child Topic"** category with a balanced sentiment distribution further indicates that not all tweets were specific to a well-defined theme, underscoring the complexity and breadth of child-related conversations on social media.

Overall, this thematic breakdown reaffirms the initial sentiment analysis and reveals the specific domains in which public sentiment is most active and polarized—serving as a valuable guide for targeted advocacy, policy focus, and awareness campaigns.

Sentiment Distribution by Theme

The sentiment distribution per theme provides a more granular perspective, helping us understand which topics elicit stronger emotional responses. The stacked bar charts, both in absolute and percentage terms, highlight this sentiment polarization effectively.

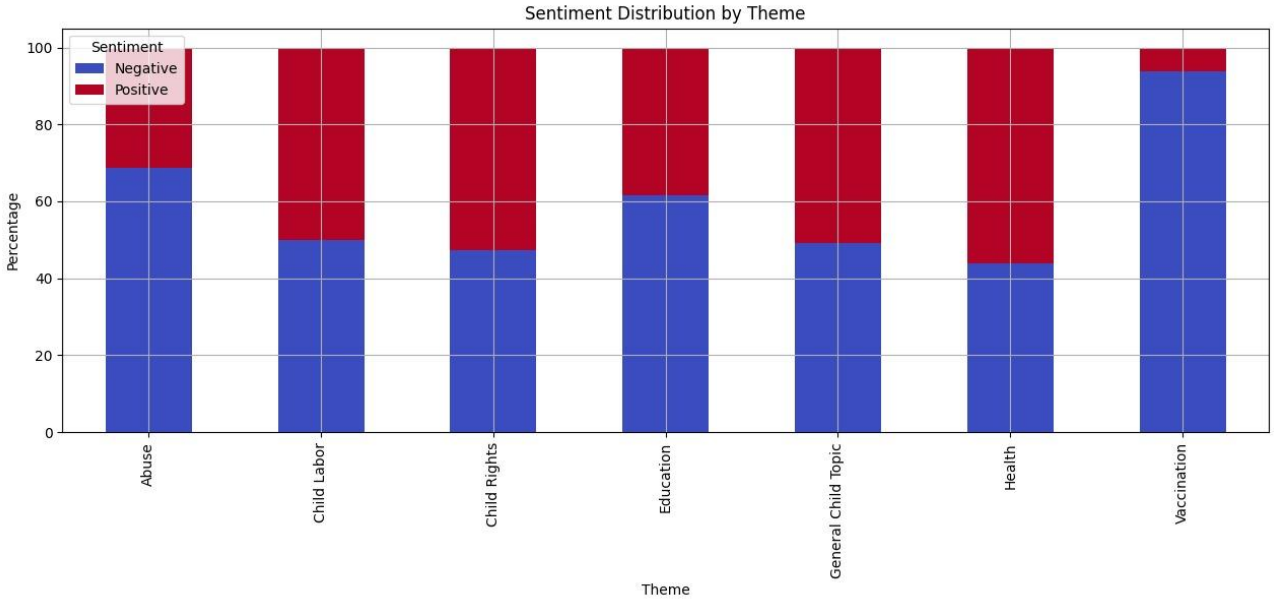

Figure 3.0: Sentiment Distribution per Theme (Percentage)

Education remains the most discussed theme by a significant margin, with a dominant share of both **negative (≈61%)** and **positive (≈39%)** sentiments. This reflects a dual narrative—while some tweets praise educational initiatives or improvements, a larger proportion likely criticizes challenges such as access, quality, or equity.

General Child Topics and **Health** show relatively balanced sentiment distributions, indicating that these topics generate mixed reactions, possibly due to varied sub-issues being discussed within each.

Themes like **Abuse**, **Child Labor**, and **Child Rights** lean more negatively, with Abuse showing a particularly high percentage of negative sentiment **(~69%)**, which is expected given the sensitive and concerning nature of such discussions.

Vaccination, while less frequently discussed, is overwhelmingly negative **(~90% negative sentiment)**, potentially due to misinformation, vaccine hesitancy, or dissatisfaction with child immunization programs.

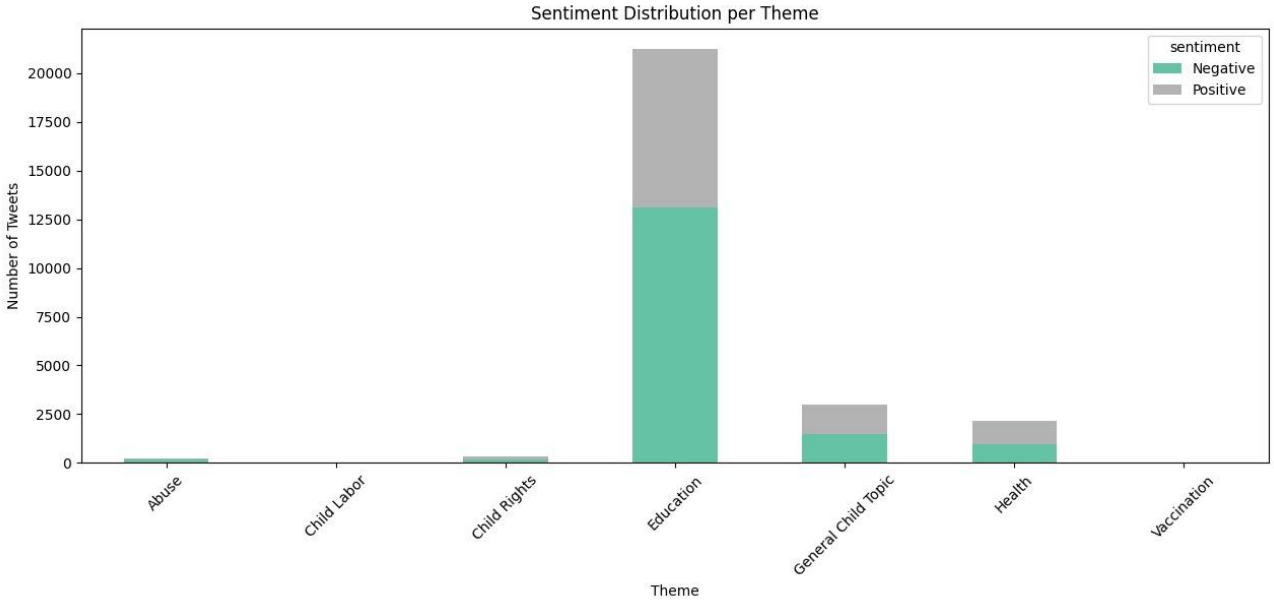

Figure 3.1: Sentiment Distribution per Theme (Tweet Volume)

Daily Tweet Volume by Theme

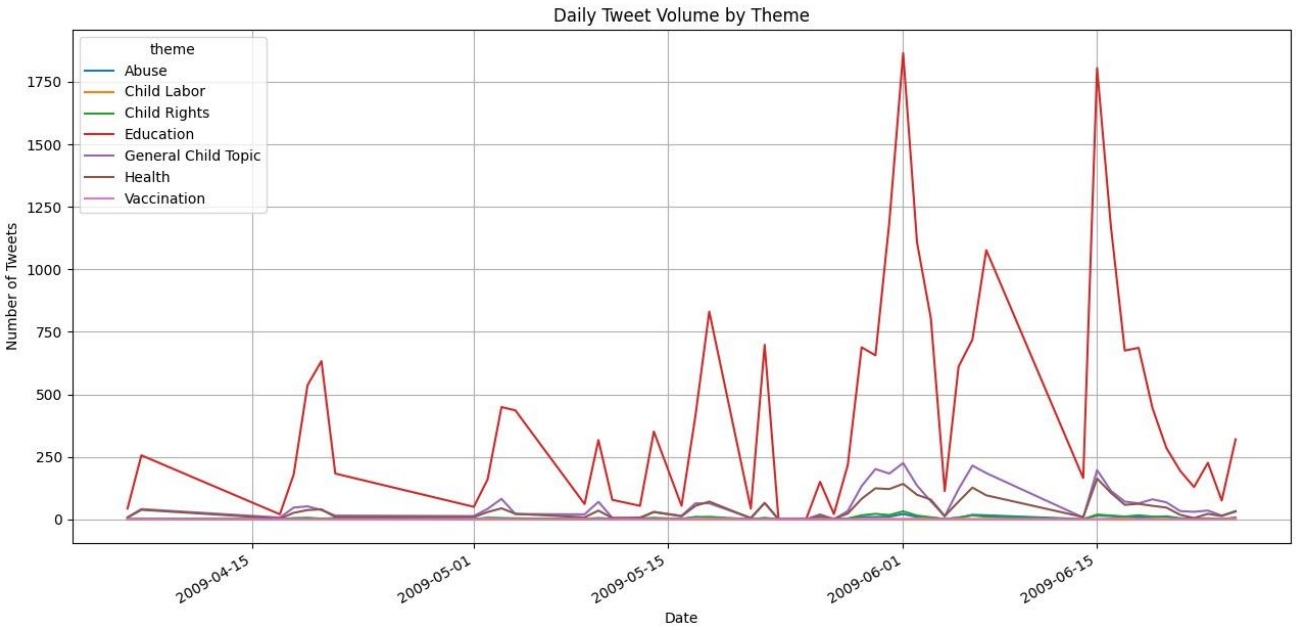

Figure 4.0: Daily Tweet Volume by Theme

From the data, one thing stands out clearly—**Education** was by far the most talked-about theme. There are several noticeable spikes, especially in early and mid-June 2009, where tweets about education shot up to over 1,800 in a single day. This kind of sudden increase usually means something big happened—maybe a news story, campaign, or event that drew public attention.

Other themes like **Child Rights**, **Health**, and **General Child Topics** also saw some engagement, though not as dramatic. Their lines remain more steady, with occasional bumps that seem to follow the education spikes. It's possible that when people talk about education, they also bring in related issues like child health or rights.

Meanwhile, topics such as **Child Labor**, **Abuse**, and **Vaccination** stayed relatively quiet throughout the period. Their tweet counts are low and don't vary much, which might suggest these issues don't get as much attention online—at least not consistently. That doesn't mean they're not important, but they may not be in the spotlight as often.

Overall, this analysis shows how attention on social media shifts over time. Some topics, like education, generate a lot of public discussion and seem to react quickly to current events. Others stay in the background, possibly waiting for a specific trigger to bring them to the surface. Understanding these patterns helps us see what people care about—and when—and can guide efforts to raise awareness where it's needed most.

Sentiment Trend Over Time (Weekly)

Building on the theme-wise tweet volume analysis, we also looked at how people felt about these topics over time. The graph above shows the **weekly trend of tweet sentiment**, split into positive and negative emotions. It helps us understand not just what people were talking about, but how they felt while doing it.

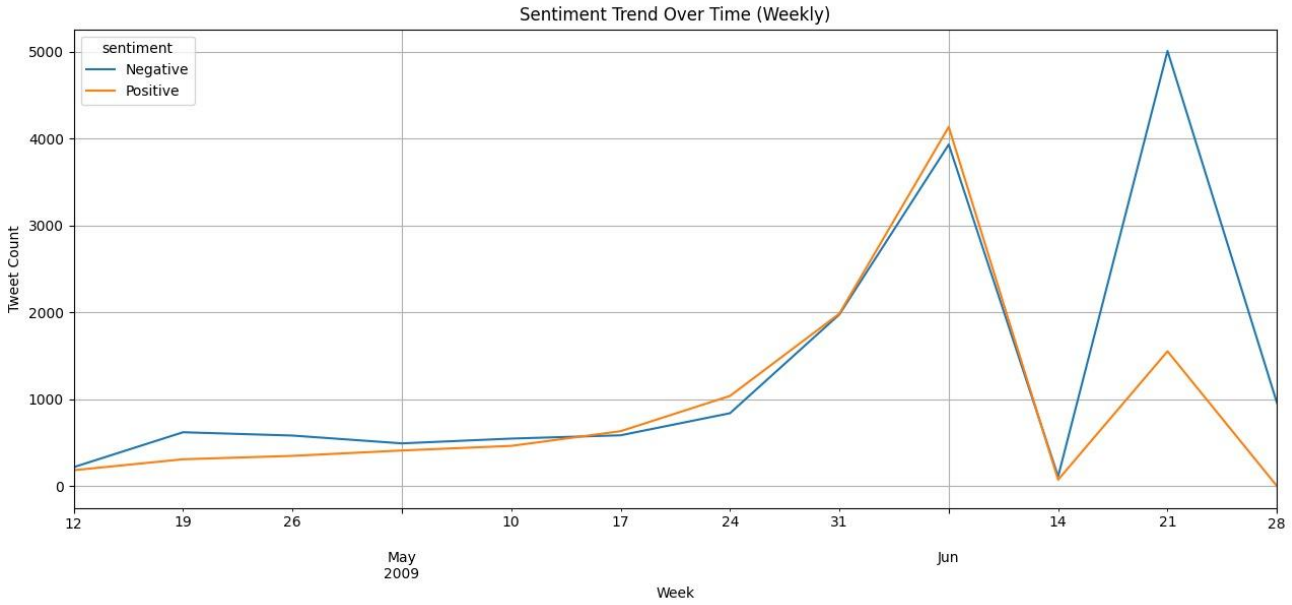

Figure 5.0: Sentiment Trend Over Time (Weekly)

At the start of the timeline in May 2009, both positive and negative tweets were relatively low and fairly balanced. As the weeks go on, there's a steady rise in activity, especially heading into June. Around the first week of June, we see a noticeable spike in both positive and negative sentiments—reaching their highest point so far. Interestingly, positive tweets slightly outnumber the negative ones during this period, which could suggest that the spike was tied to hopeful or encouraging news, possibly around education (as we saw earlier).

But what's even more noticeable is the dramatic shift in the following weeks. Just after that initial peak, both sentiments crash—likely signaling a sudden drop in engagement or the end of a trending conversation. However, this doesn't last long. In the third week of June, **negative sentiment shoots up sharply**, far exceeding positive tweets. It's the highest point of negativity in the entire dataset, which hints at a major event, controversy, or backlash that caused a strong emotional reaction online.

By the final week of June, both sentiments drop again—this time with positive tweets declining more gradually and negative tweets falling more sharply. This suggests that public attention was beginning to settle down by the end of the month.

Addressing the Spike in Negative Sentiment (Third Week of June)

One notable pattern that stands out in the sentiment analysis is the **sudden and sharp rise in negative tweets around the third week of June**. This spike does not perfectly correlate with a significant increase in any one specific theme on the **"Daily Tweet Volume by Theme"** graph—figure 4.0. However, this doesn't necessarily indicate a flaw in the analysis—it actually opens up space for deeper insight.

There are two likely explanations:

1. **Cross-Theme Emotional Trigger:**

It's possible that an incident or piece of news occurred that **cut across multiple themes**—for example, a controversial education policy, child abuse case, or healthcare issue involving children—that caused a strong emotional response but was discussed using broad or varied language. These types of discussions often show up under the **"General Child Topic"** category, which does show a modest rise around the same period. While not as dramatic as the sentiment spike, this subtle increase might reflect the same underlying cause.

2. **Unlabeled or Unclassified Sentiment:**

Sentiment analysis often picks up on emotionally charged language even when the **specific theme or keyword isn't clearly labeled** in the original tweet. So, the large volume of negative sentiment may be tied to tweets using general language like "kids deserve better" or "shameful act against children," which may not have been assigned to a narrow theme like "Child Labor" or "Health." This is especially likely in fast-moving or emotionally intense events.

The spike in negative sentiment during the third week of June, although not directly mirrored by a single thematic volume increase, likely reflects a cross-cutting or broadly framed issue. The moderate rise in 'General Child Topic' tweets during this period, along with the nature of sentiment-based classification, supports the possibility of a widely discussed, emotionally charged event not confined to one theme.

Contextual Explanation:

After exploring real-world events from that time period, a possible explanation becomes clear. **June 21, 2009, was marked by multiple violent and emotionally disturbing events**, especially in **Pakistan**, where:

- A **terrorist bombing** targeted a Frontier Constabulary vehicle, killing 18 troops.
- **Public beheadings** and lashings were carried out by militants in Swat District.
- **Clashes** between the Pakistani army and the Taliban intensified, creating fear and unrest.

These events likely triggered **strong emotional responses online**, particularly among those concerned with child safety, conflict, and human rights. While not strictly classified under a theme like "Child Rights" or "Health," these tweets might have used broader language or fell under general sentiment detection, explaining the disconnect between theme volume and emotional spike.

At the same time, **global political unrest**, such as the **Iranian election protests** and **violent crackdowns**, could have added to the emotional climate on Twitter—especially if children or families were mentioned or affected in those reports.

Final Insight

So, while the sharp rise in negative sentiment may not have aligned with a specific child-related theme, it reflects a broader **emotional reaction to real-world violence and political turmoil**. These emotionally intense discussions may have used general terms—picked up by sentiment analysis but not clearly tied to any one theme in your dataset.

Word Cloud for Theme: Education

To visualize public discourse more effectively, we generated a word cloud specifically for the **Education** theme because it had the **highest volume of tweets** across all categories, ensuring a rich and diverse dataset. Additionally, education-related tweets often contain **casual, expressive, and emotionally varied language**, making them especially well-suited for a word cloud. This allowed us to capture the tone, topics, and everyday sentiments of users discussing school-related experiences during the analysis period.

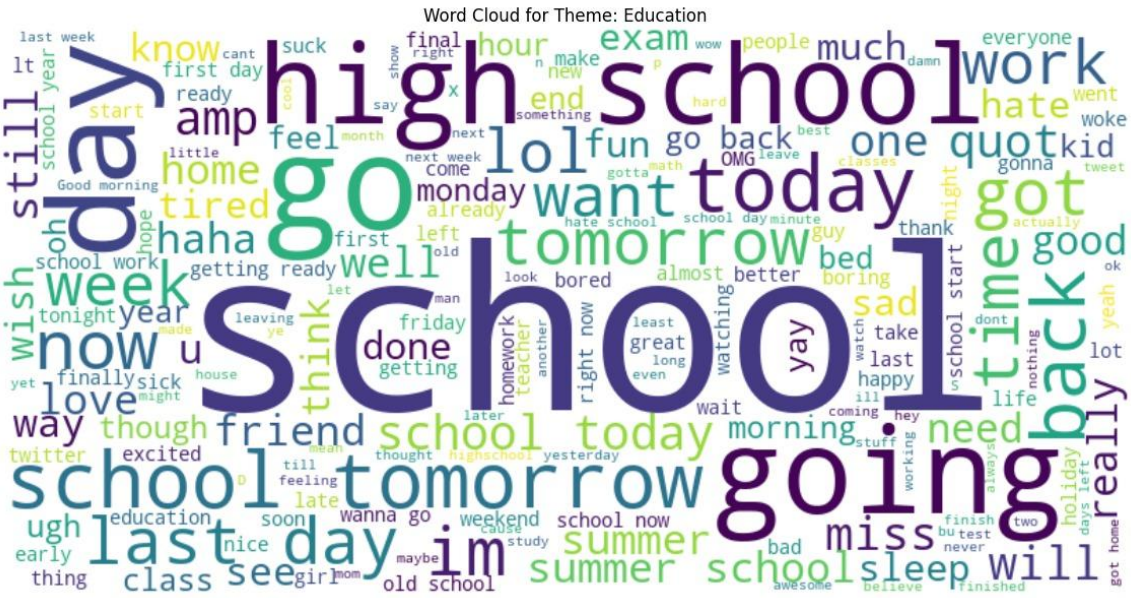

Figure 6.0: Word Cloud for Theme: Education

To gain a deeper understanding of how people were talking about education on Twitter, we generated a word cloud based on tweets classified under the **Education** theme. This visualization gives a quick overview of the most commonly used words and expressions, helping us get a feel for the overall conversation without diving into each individual tweet.

As expected, words like **“school”**, **“high school”**, **“tomorrow”**, and **“going”** appear most prominently—showing just how much daily routines, schedules, and school-related emotions dominated the discussion. Phrases like **“tired”**, **“homework”**, **“exam”**, and **“last day”** also stand out, suggesting that students were frequently tweeting about their experiences, especially around end-of-term stress and transitions. On the emotional side, words such as **“fun”**, **“miss”**, **“ugh”**, **“lol”**, and **“hate”** show a mix of excitement, sarcasm, fatigue, and even frustration—capturing the very real ups and downs of student life.

This word cloud provides a vivid snapshot of the school experience as it was reflected online during the period analyzed. It highlights that education is not just a formal topic—it’s personal, emotional, and part of everyday life for many Twitter users.

Key Insights

After exploring over 1.6 million tweets related to children and their well-being, some clear patterns began to emerge—both in what people talked about and how they felt.

- First, **negative sentiment stood out across the board**, making up nearly 59% of all tweets. That’s not too surprising, considering the types of issues people were reacting to—child abuse

cases, failing education systems, and broader concerns about health and safety. These conversations often came from a place of frustration, fear, or disappointment.

- **At the center of most discussions was Education**. It wasn't just the most talked-about theme—it was also the most emotionally charged. People expressed everything from joy and excitement to anger and exhaustion. Some shared personal wins or positive school moments, while others voiced serious concerns about fairness, access, and quality. The word cloud made that emotional contrast easy to see, with words like "exam," "fun," "tired," and "tomorrow" all popping out.
- **Other themes—like Child Rights, Health, and Abuse**—had much lower tweet volumes, but the sentiment was just as intense, especially around abuse and child labor. These weren't topics people discussed lightly; when they came up, the tone was almost always serious and often heart-wrenching. **Vaccination**, while not as heavily discussed, leaned strongly negative—likely tied to hesitancy, skepticism, or unmet expectations around child immunization.
- When we looked at how **sentiment shifted over time**, we noticed clear emotional spikes—especially in early June. Those peaks often aligned with the rise in education-related tweets, hinting at trending events or school-year milestones. But the **most intense spike in negativity came during the third week of June**, and it didn't align neatly with any single theme. After some digging, we found that several tragic and violent events happened around that time—especially in Pakistan. News of bombings, public executions, and military clashes likely fueled widespread outrage and sadness. People may not have mentioned children directly in every tweet, but the emotional impact still spilled into the general online conversation.
- All in all, this analysis shows that **social media isn't just a platform for information—it's a reflection of how people feel in real time**. From joy and stress over school to anger over injustice, the emotional pulse captured in these tweets offers real insight into what matters to people—especially when it comes to children.

Limitations

While this analysis provides valuable insights into public sentiment and thematic trends, there are a few limitations worth noting:

Theme Classification Isn't Perfect: Tweets were categorized based on pre-defined themes, but real conversations often blur lines. A single tweet might touch on education, health, and child rights all at once, making it difficult to assign it cleanly to just one category.

Sentiment Detection Has Nuance Gaps: Although sentiment analysis tools can detect general tone, they may struggle with sarcasm, slang, or cultural context. For instance, a tweet meant to be humorous or ironic might be misclassified as positive or neutral.

Not All Tweets Are Contextualized: Tweets that expressed strong emotion may not have mentioned enough specific keywords to be tied to a particular theme. This is especially relevant for tweets contributing to the spike in negative sentiment during late June, where emotion was high but topic tagging was unclear.

Language and Geo-Limitations: The analysis focused primarily on English-language tweets, which may leave out important perspectives in other languages. Similarly, unless geo-tags or context were available, assumptions about location-based events (like those in Pakistan or Iran) were made based on timing and content rather than confirmed geolocation.

Volume Doesn't Always Equal Importance: Some important topics like Child Labor or Abuse had fewer tweets, but that doesn't mean they're less significant. These issues might simply not trend as easily or frequently on platforms like Twitter without a triggering event.

Timeframe Is Short: The dataset only covers a limited period (May to June 2009). While useful for spotting short-term patterns, it may not reflect long-term trends or shifts in public discourse over time.

Recommendations

Based on the patterns and insights revealed in this analysis, several key recommendations can be made for stakeholders, advocacy groups, and policy makers working in the area of child welfare:

Leverage Education as a Gateway Topic

Since education drives the most public engagement and sentiment, it can serve as a powerful entry point for broader child welfare campaigns. Messaging around health, rights,

or protection can be strategically tied into education discussions to increase reach and impact.

Monitor Spikes in Negative Sentiment for Early Signals

The sudden rise in negative sentiment—especially during times of crisis—can serve as an early indicator of public concern or outrage. Building systems to monitor sentiment trends in real-time could help organizations respond faster and more effectively to emerging issues.

Amplify Underrepresented Themes

Critical issues like child labor, abuse, and vaccination are emotionally charged but receive less sustained attention. These areas may benefit from targeted awareness efforts, storytelling campaigns, or partnerships with influencers to bring them into mainstream conversation.

Address Misinformation in Vaccination Discourse

The overwhelmingly negative sentiment around vaccination highlights a need for better communication, education, and trust-building. Public health campaigns should focus on transparency, myth-busting, and highlighting success stories to counteract fear and hesitancy.

Explore Broader Context Behind Emotional Spikes

Not all emotional reactions are tied to clearly labeled themes. Future analyses should consider incorporating external data sources—like news headlines or crisis reports—to better interpret sentiment trends and their real-world drivers.

Include Multi-Language and Geo-Specific Analysis

To capture a more complete picture, especially in diverse regions like Pakistan or during global events, future work should analyze tweets in local languages and include confirmed geographic data wherever possible.

Expand the Timeframe for Ongoing Monitoring

Since this analysis focused on a limited two-month window, extending the timeframe could help reveal seasonal patterns, long-term shifts in public opinion, or the impact of policy changes over time.

Conclusion

This analysis of over 1.6 million tweets provides a window into how the public engages emotionally with issues affecting children—from education and health to abuse and rights. While education stood out as the most discussed and emotionally complex theme, the broader sentiment landscape leaned toward concern and criticism, with nearly **59%** of tweets expressing **negative** emotions.

What makes this analysis meaningful is not just the data—it's the connection between online expression and real-world events. The sharp rise in negative sentiment during the third week of June, for example, wasn't random—it reflected a genuine emotional response to violence, unrest, and human rights concerns happening offline. This underscores the value of social media not just as a communication tool, but as a reflection of public feeling and social priorities.

At the same time, the relatively low visibility of serious issues like child labor, abuse, and vaccination reminds us that not all urgent matters receive equal attention. This creates an opportunity—and responsibility—for advocates, educators, and policymakers to amplify these underrepresented voices.

In short, this report highlights how digital platforms capture more than just opinions—they echo real concerns, reflect societal patterns, and offer a powerful tool for awareness, advocacy, and early intervention. Listening closely to these signals can help shape more informed, responsive, and impactful approaches to child welfare moving forward.