Paper Number:

(For office use only)

**SENTIMENT ANALYSIS ON USER PERSPECTIVES OF THE ANDROID EARTHQUAKE EARLY WARNING ALERT SYSTEM THROUGH TWITTER DATA**

**ABSTRACT**

*Google's Android Earthquake Alerts System, launched in August 2020, uses Android smartphones as compact seismometers to detect seismic activity and transmit data to Google's server. The system notifies users and provides seismological data for research. However, challenges include user knowledge, availability, technical issues, trust, and attitudes towards technology. To improve, it needs user-centric design, educational initiatives, system enhancements, and effective communication strategies. This study uses Twitter data to analyze user opinions on the Android Earthquake Early Warning (EEW) Alert System. It aims to assess user responses and factors influencing their feedback. The findings will guide research methodology, provide insights into public perception, and propose strategies to enhance user responses and optimize the system's use. Twitter provides real-time insights into user experiences and reactions to alerts. Data collection involved using the Python SNScrape library to extract relevant tweets and filtering to include countries with Android EEW systems. To ensure valuable insights and a meticulously curated dataset, exploratory data analysis and sentiment analysis were performed following data cleaning and preprocessing. The study found disparities in tweets generated by countries regarding the Android EEW system, with the US showing the highest awareness due to familiarity with the system. New Zealand showed a slower rate of awareness, while Greece showed slower progress. Other countries showed similar patterns. The study identified areas for improvement, including technical issue resolution, user education, and cultural context considerations to foster acceptance and trust.*

**Keywords**: Awareness, Android EEW, Public perception, Sentiment analysis.

1. **Introduction**

In August 2020, Google introduced the Android Earthquake Alerts System, a pioneering initiative designed to provide rapid earthquake detection and alerts through Android smartphones. Leveraging the built-in accelerometers of Android devices, this system transforms them into compact seismometers capable of detecting seismic activities. The captured data is then relayed to Google's earthquake detection server, enabling timely alerts to users in earthquake-prone areas. Moreover, the system contributes valuable earthquake data to seismological organizations, fostering research efforts in the field. Although the initial response to this system has been largely positive, there remain certain concerns and uncertainties that warrant further investigation and resolution. Notably, a study conducted in New Zealand identified the need for improved communication and user comprehension of the Android Earthquake Alerts System. This system's efficacy is contingent on factors such as user awareness, accessibility, technical challenges, trust-building, and prevailing attitudes towards technology. To overcome these challenges, it is imperative to adopt a user-centric approach to design, implement comprehensive educational campaigns, refine system functionalities, and establish effective communication strategies. These measures collectively aim to bolster the system's acceptance and utilization.

The primary objective of this study is to perform sentiment analysis on Twitter data to gain insights into user perspectives regarding the Android Earthquake Early Warning (EEW) Alert System. By analyzing public sentiment on social media, this research seeks to achieve several key objectives. Firstly, it aims to gauge individuals' reactions to the EEW alert system and identify the fundamental factors influencing their evaluations and feedback. Additionally, sentiment analysis of Twitter data offers a unique window into the public's perception of the EEW system, enabling a nuanced understanding of the sentiment landscape. Based on the findings, the study intends to propose strategies and concepts that can optimize user responses and maximize the utility of the EEW system. The establishment of these objectives guides research methodology, facilitates a comprehensive grasp of user viewpoints, and ultimately contributes to the enhancement of the Android EEW Alert System.

To conduct this investigation, the study employs Twitter as a platform for accessing real-time insights into personal experiences and reactions to earthquake alerts. Leveraging the Python SNScrape1 library, the research collects tweets containing relevant keywords, focusing on countries where the Android EEW system has been implemented. The ensuing analytical process encompasses several crucial steps. Initial sentiment analysis utilizes the VADER2 sentiment lexicon, followed by exploratory data analysis to yield statistical summaries. Moreover, data preprocessing techniques are employed to optimize textual analysis, and data cleaning procedures ensure a refined dataset. These methodologies collectively ensure the acquisition of valuable insights and a meticulously curated dataset for in-depth analysis.

The outcomes of this study unveil intriguing disparities in tweet volume across different countries, indicative of varying levels of awareness and engagement with the Android EEW system. Temporal analysis further uncovers an incremental increase in awareness, particularly evident between 2022 and 2023. Disparate levels of awareness are observed among countries, with the United States showcasing notably higher awareness owing to prior exposure to similar EEW systems. New Zealand's trajectory transitions from initial surprise to gradual acceptance, while other countries like Greece demonstrate a slower pace of awareness development. The sentiment analysis underscores an overall positive sentiment towards the Android EEW system.

1 <https://pypi.org/project/snscrape/>

2 <https://pypi.org/project/vaderSentiment/>

1. **Literature Review**

The convergence of technology and disaster management has resulted in the development of creative solutions, such as Android-based earthquake early warning systems, in the present day where smartphones have become an essential part of daily life. By providing prompt alerts to people and communities, these technologies have tremendous potential to lessen the catastrophic effects of earthquakes. A lot of research is done to show how individuals react to the warning notice and the reasons they leave such reviews.

Cremen et al. (2020) have explored the real influence of these systems on human responses and behaviors is vastly different from what is currently understood about their potential. This obvious discrepancy highlights the critical need for additional empirical research and thorough analysis to close the gap between theoretical predictions and actual, real-world facts. The prospect of producing a more complete understanding of these systems' capabilities and constraints lies in further in-depth research initiatives in this field. This, in turn, will play a crucial role in facilitating informed decision-making processes and, more importantly, improving their overall effectiveness in influencing human reactions and behaviors in a variety of different circumstances. To fully utilize the capabilities of these systems for many practical applications, such study is essential.

Becker et al. (2020, a) have examined the success of public education initiatives in the context of disaster preparedness. Despite massive initiatives to educate the public about risks and warnings, people's actual conduct during catastrophes can frequently deviate from what they have been taught. This discovery emphasizes the necessity of a thorough analysis of the body of literature to understand the causes of this divergence. Improving disaster response tactics and guaranteeing community safety depends on an understanding of the variables behind this gap. To close the theory-practice gap in disaster preparedness education, this review will clarify the existing state of knowledge and point out areas where additional study is necessary.

Becker et al. (2020, b) have highlighted a puzzling yet crucial aspect of how people behave in catastrophe situations: despite extensive public education campaigns, people's behavior can still lead to behaviors that are improper in terms of disaster preparedness. This key discovery captures the complicated interplay between mental operations, emotional states, environmental stressors, and the subtle complexity of decision-making under pressure. This study critically investigated the broad implications, resulting in a significant contribution to a comprehensive knowledge of the multifaceted interactions between public education campaigns, human behavior, and the confluence of elements that influence individuals' actions.

Since social networks have become more and more popular over the past 10 years, Twitter has expanded dramatically on a global scale. It is not surprising that many academics began exploring methods to use a platform akin to microblogging to discover fresh applications across a variety of fields. Online sentiment analysis has gained more psychological and social acceptance, but there are a lot of uses for it. Pang et al. (2002) were intrigued by the idea of studying people's behavior.

Formen (2003) has employed many machine-learning techniques that can be used to categorize text documents. The support vector machine and the Naïve Bayes classification are the most well-known. These techniques enable classification by instructing a computer on annotated datasets so that texts can be automatically tagged, which is different from unsupervised learning clustering.

Text mining relies on written content, and for this purpose, tweets are a preferred choice among academics. Tweets are brief and provide accurate updates on current events, making them a prime focus. They are more accessible to gather compared to content from other social networks. An open Application Programming Interface (API) is now available for use by any developer who wishes to use data from Twitter.

Additionally, hashtags and other symbols that facilitate the search for relevant content are widely used in tweets. Writers employ a variety of machine learning techniques after gathering a large enough sample, which they can validate using certain techniques like cross-validation and then assess using accurate metrics like the F1-measure (Giachanou et al., 2016).

**3. Methodology**

This study is primarily concerned with understanding the perspective of users on the Android EEW system. This is achieved through the analysis of datasets to gain significant insights. The primary intention of the study is to analyze feedback collected from the social media platform, Twitter. Twitter users are recognized for their tendency to predominantly center their attention on personal experiences, encompassing their emotions and responses to notifications. This characteristic renders the social media platform an optimal tool for gathering contemporaneous observations regarding the EEW system. The analysis of a substantial amount of self-reported data obtained from tweets is an opportunity to derive insights about the user experience that may not be replicated by traditional survey methods.

* 1. **Data Source**

To procure data from the Twitter social media platform, we utilized the Python SNScrape library. During the data collection phase, we encountered a series of notable challenges. Typically, Python Tweepy library3 and the Twitter API serve4 are reliable tools for extracting data from Twitter's platform. However, our data collection efforts were significantly impacted by recent changes in Twitter's ownership and associated developments. These alterations introduced a plethora of modifications and required adjustments in the syntax and functionality of the Twitter API.

3 <https://pypi.org/project/tweepy/>

4 <https://developer.twitter.com/en/docs/twitter-api>

Consequently, we found ourselves unable to employ the conventional Twitter API for our data retrieval purposes, prompting us to turn to the SNScrape library as an alternative solution. This decision was necessitated by the evolving landscape of Twitter's interface,

and it allowed us to effectively gather the data required for our research endeavors.

The data collection methodology prioritized the extraction of tweets that incorporated pertinent keywords, including 'earthquake', 'alert', 'android', 'warning', and 'notification'. Following the successful extraction of data from Twitter, a significant dataset was acquired, consisting of various columns including 'Date', 'User', 'Tweet', and 'Location'. To maintain data consistency, specific filters were implemented to refine the dataset. In particular, the 'Location' column was subjected to a filter to extract tweets that originated from countries where the Android EEW system had been introduced.

In this study, the data collection procedure entailed the retrieval of tweets during the launch of the initial Android EEW system in April 2021. The period of data collection extended from the inception of the system in April 2021 until March 2023. Table 1 displays the number of months for each year the data was collected, and Table 2 and Figure 1 show the breakdown of the total number of Tweets extracted from each country and the proportion of total Tweets from all contributed countries, respectively.

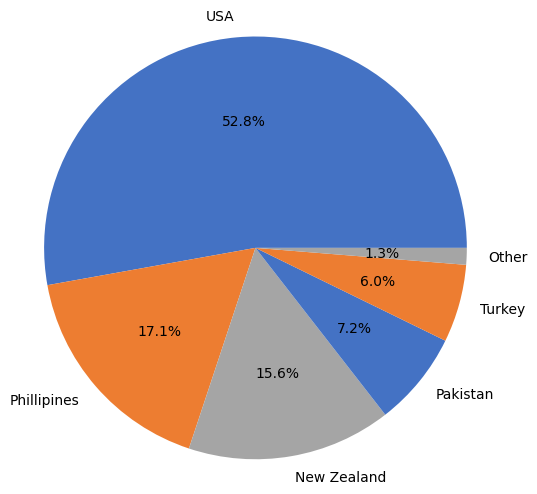
**Table 1. The number of months for each year the data was collected.**

|  |  |  |
| --- | --- | --- |
| Year | Number of months | Period |
| 2021 | 09 | April - December |
| 2022 | 12 | January - December |
| 2023 | 03 | January - March |

**Table 2. The breakdown of the total number of Tweets extracted from each country where**

**Android EEW was launched.**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Country | 2021 | | 2022 | | 2023 | | Total | |
| USA | | 510 | | 528 | | 159 | | 1197 | |
| Philippines | | 263 | | 94 | | 30 | | 387 | |
| New Zealand | | 194 | | 129 | | 31 | | 354 | |
| Pakistan | | 31 | | 44 | | 89 | | 164 | |
| Turkey | | 15 | | 20 | | 100 | | 135 | |
| Greece | | 5 | | 9 | | 3 | | 17 | |
| Bangladesh | | 2 | | 3 | | 1 | | 6 | |
| Tajikistan | | 2 | | 0 | | 2 | | 4 | |
| Uzbekistan | | 2 | | 0 | | 0 | | 2 | |
| Kazakhstan | | 0 | | 0 | | 0 | | 0 | |
| Turkmenistan | | 0 | | 0 | | 0 | | 0 | |
| Kyrgyz Republic | | 0 | | 0 | | 0 | | 0 | |
| Total | **784** | | **1060** | | **416** | | **2266** | |



**Figure 1. The proportion of total Tweets from all contributed countries.**

* 1. **Data Cleaning and Preprocessing**

The Twitter data collected often includes superfluous components, such as tweets that lack relevance or are redundant, retweets, or unsolicited messages. The data cleaning process involves several steps, including the identification and removal of duplicate entries, the filtering of retweets, the exclusion of irrelevant or promotional content, and the resolution of any missing or erroneous data.

Given the concise and contemporary character of Twitter data, it is critical to use preprocessing techniques to optimize textual data analysis. To guarantee that the tweet data was suitable for study, it was subjected to several preprocessing methods. Irrelevant punctuation, special characters, and excessive spaces were initially removed from the tweet to maintain the pertinent phrases and assure consistency.

The normalization of the text was facilitated by the reduction of repeated characters. The elimination of URLs and HTML tags was implemented to focus solely on the textual content. Due to their lack of relevance in the context of sentiment analysis, numerical digits were removed from the dataset. The contractions were enlarged to their full forms, ensuring that terms were represented consistently. By changing all terms to lowercase, the content was normalized.

Tokenization is the technique of separating a given text into discrete components or tokens that may be evaluated individually. Certain terms, such as commonly used articles and conjunctions, were removed to reduce interference. Lemmatization is a linguistic procedure that attempts to standardize words by reducing them to their simplest form, assuring consistency. Finally, any duplicated tokens were removed, ensuring that each phrase contributes only one contribution to the analysis. The preprocessing processes used converted the tweet data into a more refined and homogeneous structure, making it acceptable for later analysis and sentiment classification.

* 1. **Data Analysis**

The data analysis process for sentiment analysis on Twitter data involves a series of sequential stages meticulously designed to extract meaningful insights and gain a comprehensive understanding of the inherent characteristics of the collected data. These stages constitute a systematic and iterative approach, allowing for a thorough exploration of the dataset's nuances and sentiment-related attributes.

**3.3.1. Exploratory Data Analysis**

The exploratory study utilized a rigorous statistical methodology, employing descriptive statistics to thoroughly examine the fundamental aspects of our dataset. The statistical analysis involved a thorough investigation of counts, distributions, and frequencies related to important variables of interest. Through this analytical procedure, we have successfully revealed significant findings, which encompass the measurement of the number of tweets and their temporal distribution, the occurrence and popularity of hashtags, the frequency of mentions, and noticeable patterns in user involvement. The systematic investigation facilitated the acquisition of a more profound comprehension of the intrinsic dynamics of the data, establishing a robust basis for further sentiment analysis.

**3.3.2. Sentiment Analysis**

The primary focus of our research pertains to the analysis of sentiment. The purpose of this analysis is to ascertain the sentiment conveyed within the tweets. Analysis of sentiment distribution and trends can offer valuable insights into the prevailing public opinion or sentiment regarding topics or entities. The sentiment analysis of the study was conducted using the VADER sentiment lexicon from the Python NLTK5 (Natural Language Toolkit) and TextBlob6 libraries. The utilization of pre-constructed sentiment dictionaries by the libraries facilitated the assignment of sentiment scores and labels to every tweet that underwent processing. Through the utilization of these robust instruments, we acquired sentiment scores that accurately depict the polarity and magnitude of emotions conveyed within the tweets. The sentiment labels furnished a categorical depiction of the general sentiment expressed by every tweet, classifying them into positive, negative, or neutral categories. The employed methodology facilitated the quantification and analysis of the sentiment patterns present in the dataset, thereby yielding significant insights into the overall sentiment of the subject matter under consideration.

**4. Results/Analysis and Discussion**

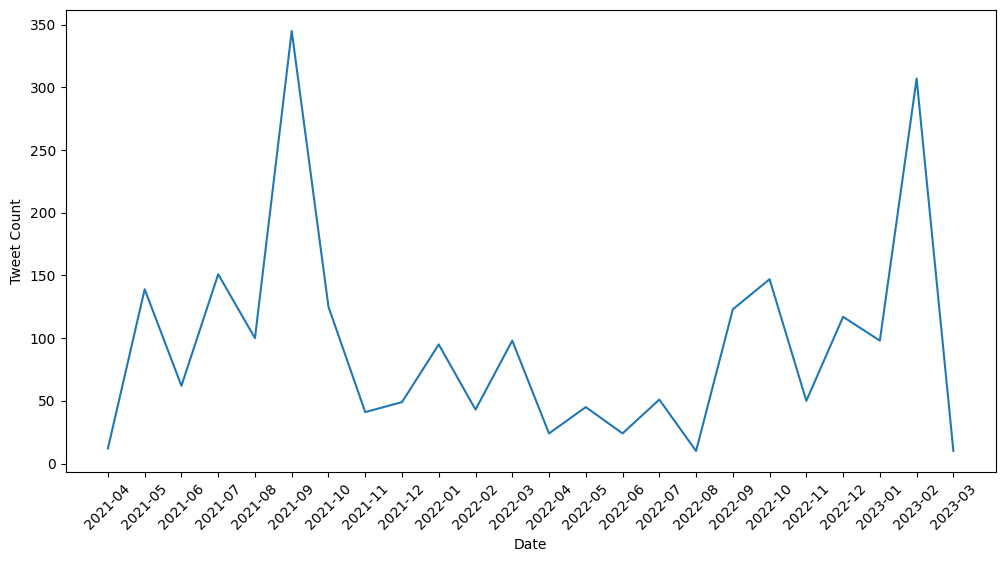
* 1. **Exploratory Data Analysis Result**

Earthquake Early Warning systems can provide valuable time for people to take

5 <https://www.nltk.org/>

6 <https://pypi.org/project/textblob/0.9.0/>

protective action before the arrival of strong shaking. However, the effectiveness of EEW systems depends on the public's awareness of these systems and their willingness to participate in them. Figure 2 displays notable disparities in the frequency of tweets generated over time, offering light on the diverse degrees of awareness and engagement in the deployment of the Android EEW system.

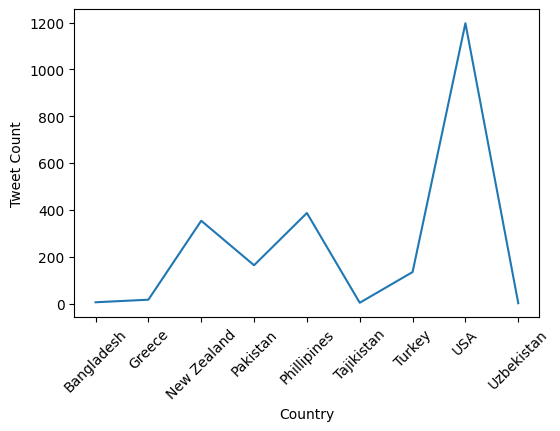
****

**Figure 2. The fluctuation of the frequency of tweets over time.**

A variety of factors can influence the level of awareness and involvement in EEW systems. The occurrence of real seismic occurrences is one of the primary drivers of public awareness of EEW systems. These technologies rely on the detection and monitoring of naturally occurring earthquakes. As a result, the efficiency of EEW systems is conditional on such events occurring. When seismic activity activates the warning system, it acts as a powerful stimulus for informing the public. In the lack of recent earthquakes, people may perceive a reduced necessity for EEW notifications, which might lead to a reduction in public awareness.

Twitter data analysis also indicated a significant decrease in debates and mentions pertaining to EEW in 2022. The decrease in attention towards seismic preparedness can be ascribed to the predominant global emphasis on the COVID-19 pandemic, which has overshadowed deliberations on other urgent global concerns. The outbreak of the pandemic garnered significant public attention and allocated resources, resulting in a temporary decline in the prominence of EEW posts on social media platforms.

Figure 3 shows significant differences in the number of tweets created by individual countries. Despite the simultaneous deployment of the Android EEW system in a few countries, it became clear that the level of awareness and involvement varied across different countries.



**Figure 3. The fluctuation of the frequency of tweets over countries.**

The observed variation in tweet volumes highlights the need to take cultural and environmental aspects into account when evaluating the acceptance and deployment of the Android EEW mechanism. Understanding how people think can help with the distribution and operation of earthquake early warning systems in several locales.

Indeed, awareness levels vary among countries due to a variety of circumstances. Seismic activity levels, historical experiences, cultural views, technological access, and public education are examples of these. These characteristics have a substantial impact on earthquake early warning awareness and readiness. Figure 4 illustrates the Tweet count distribution by Location and Year.

Despite the simultaneous launch of the EEW system in the USA, New Zealand and Greece on April 28, 2021, there are significant variations in awareness levels among these countries. The United States stands out with a high awareness rate, attributed to the previous implementation and familiarity with EEW systems like Shake Alert. In New Zealand, the initial shock was observed among the population, but over time, they appear to be adapting to the concept of EEW. However, Greece is lagging in terms of EEW system development and awareness.

In contrast, Pakistan, the Philippines, and Turkey have experienced notable increases in awareness. The nations stated above have exhibited a growing understanding and recognition of the EEW mechanism. It is evident that while their degree of awareness is significantly lower than that of the United States, there is still a recognizable level of awareness when compared to the prevailing circumstances in Tajikistan and Uzbekistan. The levels of awareness in these countries are similar to that of Greece, despite the differences in the timing of their respective initiations.

* 1. **Sentiment Analysis Result**

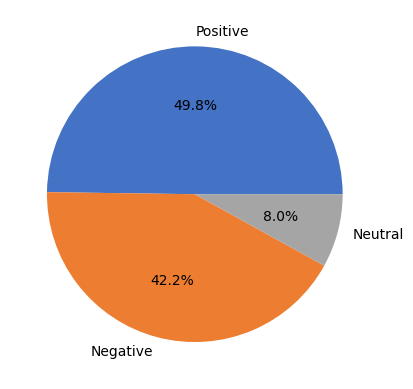
The use of pre-existing sentiment dictionaries provided by the Python NLTK library aided in the process of assigning sentiment scores and labels to individual processed tweets. The sentiment scores accurately captured the polarity and intensity of the tweets' emotions. Based on the sentiment labels, the tweets were classified as positive, negative, or neutral, providing a categorical representation of the overall sentiment represented in each tweet.

A graph of different countries/regions

Description automatically generated

**Figure 4. Tweet count distribution by Location and Year.**

Using this methodology, we were able to successfully quantify and investigate the prevalent patterns of sentiment within the given dataset. The current study yielded notable data about the general sentiment toward the topic under consideration. Figure 5 displays the summary of overall sentiments based on Twitter data.

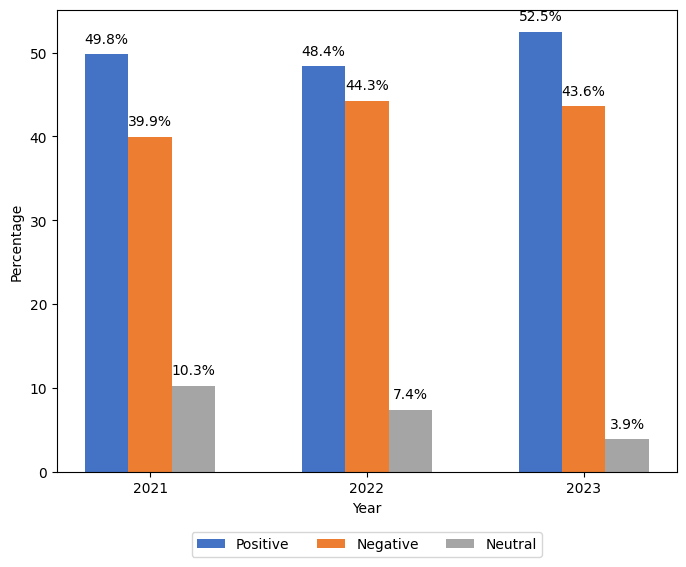
****

**Figure 5. The summary of overall sentiments.**

It can be seen that the general sentiment regarding the Android EEW system is positive. Furthermore, the study found only a little variation between positive and negative feelings toward the Android EEW system. This indicates that the public's response to the system is largely balanced, with both positive and negative opinions expressed.

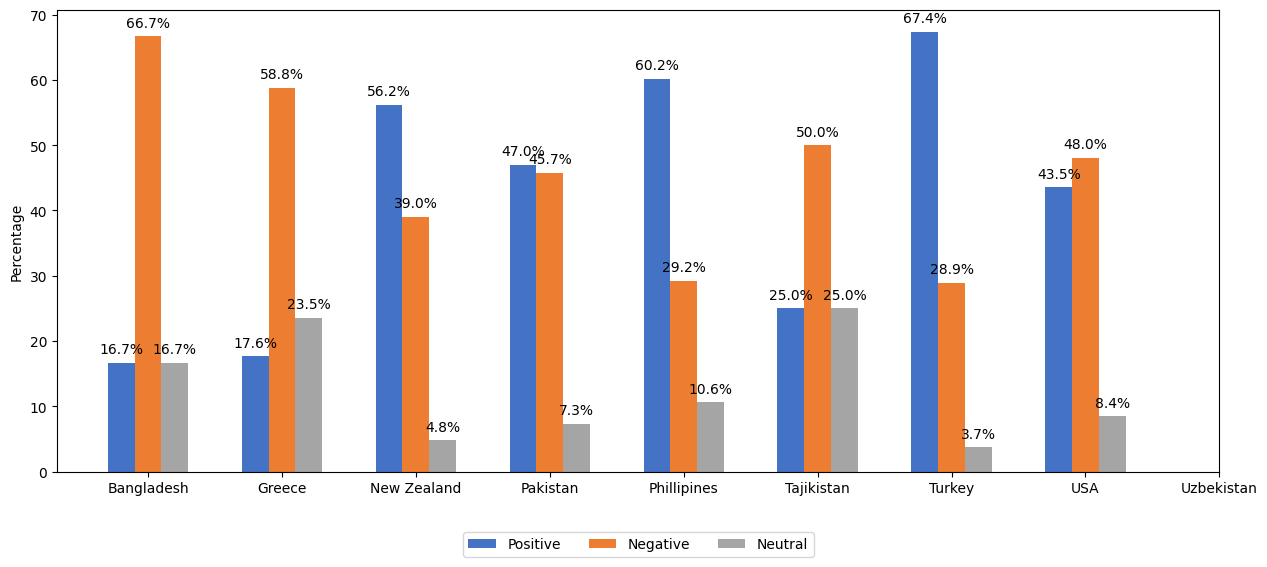
The alert system's accuracy and timeliness were regarded as major variables influencing people's attitudes toward the Android EEW system. Users' trust in the system emerged as an important factor as well. These findings emphasize the importance of these variables in affecting public opinion and satisfaction with the Android EEW system.

Upon classifying the sentiment distribution by year in Figure 6, it was observed that the overall sentiment towards the Android EEW system has shown a positive trend over time. However, it is important to note that the negative sentiment has also been increasing. In particular, the percentage of negative emotions in 2023 is higher compared to the negative emotions expressed in 2021. Interestingly, while the overall sentiment has become more positive, the proportion of neutral sentiment has been decreasing over time. This suggests that people are becoming more confident in expressing their opinions, whether they are positive or negative, about the Android EEW system. These findings indicate that there is a growing level of engagement and feedback from users regarding the Android EEW system. The increasing negative sentiment highlights areas where improvements may be needed, while the increasing positive sentiment suggests that the system is also meeting the expectations and needs of users.



**Figure 6. Sentiment distribution over time.**

When the sentiment distribution was classified by year, it was discovered that certain countries, such as the United States, Bangladesh, Greece, and Tajikistan, exhibited a higher rate of negative emotion. This implies that people in these nations had more negative feelings or thoughts about the Android EEW system. These findings are illustrated in Figure 7.

****

**Figure 7. Sentiment distribution by country.**

Countries with a greater rate of favorable emotion included the Philippines, New Zealand, Pakistan, and Turkey. This suggests that people in these nations had a more favourable attitude regarding the Android EEW system or had pleasant sentiments about it. Surprisingly, the neutral sentiment rate was consistently low in all countries. This shows that people in these countries were certain of their beliefs and had a strong preference for either a positive or negative state of mind.

These results provide useful insights into sentiment distribution across nations and demonstrate people's differing perspectives and opinions about the Android EEW system. It shows that the system's reception and public attitude can vary greatly between countries. Understanding these country-specific emotional trends can aid in adapting strategies and addressing regional problems.

**5. Conclusion and Implications**

The Android Earthquake Alerts System stands as an innovative solution in the realm of earthquake detection and early warning, harnessing the capabilities of Android smartphones to provide prompt alerts and contribute to seismic research. While the initial response has been promising, this study has shed light on crucial insights that can further refine and optimize the system's effectiveness.

The analysis of Twitter data has revealed intriguing patterns in user sentiments and perceptions. The varying tweet volumes across different countries suggest disparities in awareness and engagement levels, prompting the need for tailored outreach strategies. Temporal analysis underscores the system's growing awareness over time, with significant shifts in consciousness occurring in recent years. Notably, the United States emerged as a high-awareness country due to prior exposure to similar systems.

Crucially, the sentiment analysis paints a positive overall picture, demonstrating user appreciation for the Android EEW system. However, areas for improvement have also been identified. Addressing technical issues, enhancing user education, and considering cultural contexts are all pivotal in fostering trust and acceptance. By focusing on these aspects, the Android EEW system can not only continue to deliver effective early warnings but can also better align with user expectations and needs.

This study underscores the significance of analyzing user sentiments in shaping and refining technological solutions. By gaining a comprehensive understanding of user perspectives, we can tailor interventions that address challenges, bridge gaps, and enhance the usability of such systems. Moving forward, the insights provided by this study can serve as a foundation for refining the Android Earthquake Alerts System, ultimately contributing to the advancement of both technology and public safety.

**Acknowledgement**

The authors would like to acknowledge Massey University of New Zealand Research Fund for supporting this research.

**References**

Becker, J. S., Potter, S. H., Prasanna, R., Tan, M. L., Payne, B. A., Holden, C., Horspool, N., Smith, R., & Johnston, D. M. (2020, a). Scoping the potential for earthquake early warning in Aotearoa New Zealand: A sectoral analysis of perceived benefits and challenges. *International Journal of Disaster Risk Reduction*, *51*, 101765.

Becker, J. S., Potter, S. H., Vinnell, L. J., Nakayachi, K., McBride, S. K., & Johnston, D. M. (2020, b). Earthquake early warning in Aotearoa New Zealand: A survey of public perspectives to guide warning system development. *Humanities and Social Sciences Communications*, *7*(1), 1-12.

Cremen, G., & Galasso, C. (2020). Earthquake early warning: Recent advances and perspectives. *Earth-Science Reviews*, *205*, 103184.

Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. *J. Mach. Learn. Res.*, *3*(Mar), 1289-1305.

Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of twitter sentiment analysis methods. *ACM Computing Surveys (CSUR)*, *49*(2), 1-41.

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine learning techniques. *arXiv preprint cs/0205070*.