RESEARCH ARTICLE

ANALYZING SENTIMENTS TOWARDS E-LEVY POLICY IMPLEMENTATION IN GHANA USING TWITTER DATA

Peter Appiahene*[1] | Stephen Afrifa[2] | Emmanuel Akwah Kyei[3] | Isaac Kofi Nti[4] | Kwabena Adu[5] | Patrick Kwabena Mensah[6]

[1]Department of Computer Science and Informatics, University of Energy and Natural Resources, Sunyani, Ghana

[2]Department of Electrical and Information Engineering, Tianjin University, Tianjin, China, and Department of Computer Science and Informatics, University of Energy and Natural Resources, Sunyani, Ghana

3Department of Computer Science and Informatics, University of Energy and Natural Resources, Sunyani, Ghana

[4]Department of Computer Science and Informatics, University of Energy and Natural Resources, Sunyani, Ghana

[5]Department of Computer Science and Informatics, University of Energy and Natural Resources, Sunyani, Ghana

[6]Department of Computer Science and Informatics, University of Energy and Natural Resources, Sunyani, Ghana

Correspondence
*Peter Appiahene,
P. O. Box 214, Sunyani, Ghana.
Email:
peter.appiahene@uenr.edu.gh

Funding Information
This is a self-funded project

Abstract

Government policies face challenges whenever it is implemented or proposed. The ordinary Ghanaian always feels the down side of government policies. This paper ponders on the government of Ghana proposed electronic levy on mobile money transactions which was announced in the 2022 budget on November 17, 2021. Using the concept of sentiment analysis and Twitter data, we have carefully studied this government policy from the perspective of the ordinary citizen. Aside from conducting a non-bias examination, a full data analysis has also been performed on the data to further expound the reasons associated with the discontentment among people in the country. The results showed that data on sentiments are enormous on social media (twitter) and it serves as means for people to share their views. Our analysis depicted that phase five recorded the highest number of data, thus, 18,423 data with 24.43%, 59.29%, and 16.28% been positive, neutral and negative respectively. In our results, phase one recorded the least amount of data with 8.93%, 89.29%, 1.78% been positive, neutral and negative respectively in a total of 1,400. In the full data analysis on 38,771 tweets, 25.50%, 59.02%, and 15.48% were positive, neutral, and negative respectively.

KEYWORDS

E-Levy, policies, sentiment analysis, taxation

1 | INTRODUCTION

“Taxation is the price we pay for living in a civilized society” – Justice Oliver Wendel Holmes, U. S. Supreme Court Justice. The history of taxation in Ghana dates back to the colonial era. The British colonial authority initiated taxation in Ghana, called the Gold Coast, in September 1943, during World War II [1]. Several attempts had been made previously, for example, in April 1852, the poll tax ordinance was passed under the then-Governor Major Hill to earn money to support the growing expense of British administration (Abdallah,2008). The Income Tax Ordinance (No.27) of 1943 was thus the first Income Tax Law. Taxation has taken a different swift with the advent of technology and charged electronically which is termed as E-Levy. Technology has altered our world and daily lives over the years. Multi-functional devices such as the QR codes, E-Zwich, and smartphone have been made possible by modern technology to make transactions. The Electronic Transaction Levy (also known as the E-Levy) is a tax levied on transactions conducted over electronic or digital platforms. The government of Ghana has proposed an electronic levy on mobile transactions. The E-levy is a new tax on fundamental transactions related to digital payments and electronic platform transactions that was proposed by the government in the 2022 Budget. A charge of 1.75% will apply to electronic transactions that are more than GH¢100 on a daily basis.

There have been numerous reactions from high profile and ordinary people in the country at large on the issues of E-Levy. With the advent of social media, people resort to share their opinions which leads to data on sentiments. Sentiments have been shared on various tabloids, the mainstream and social media. It is obvious that people would have shared their opinions on the mammoth anti-government protest in 1995 dubbed “Kume Preko”, literally translated “kill me now” demo, which was held to express displeasure at the Jerry John Rawlings’ government introduced Value Added Tax (VAT) idea as well as the untold economic hardships experienced by Ghanaians at the time [2], all these were sentiments. Sentiment Analysis is a term that refers to the application of Natural Language Processing (NLP), text analysis, and computational linguistics to determine a speaker's or writer's opinion toward a certain issue (Singh et al., 2018) [3]. It basically aids in determining whether a text expresses good, negative, or neutral thoughts. Sentiment analysis is one of the most popular research topics in the field of Natural Language Processing presently. Opinion mining, recommender systems, and event detection are just a few of the scientific and commercial applications of sentiment analysis (Manguri et al., 2020) [4]. Ghana had approximately eight million social media users as of January 2021 according to Statista.com. Twitter and other social media platforms play a crucial part in expressing our thoughts (sentiments) about an event and the Ghanaian people have expressed their feelings on the proposed E-Levy. Anguish, as well as pleasure, can be used to indicate approval or rejection of specific regulations. Therefore, the need to conduct a sentiment analysis on the proposed E-Levy introduced by the government of Ghana. The primary goal of this study is to use sentiment analysis to form a fair judgment regarding the E-Levy government policy. This will help to analyze government policies based on sentiments and also assess the impact of sentiments on government policies. This study proposes a sentiment analysis concept on the proposed E-Levy policy by the government. This study is conducted on tweets on E-Levy in Ghana using the notion of sentiment analysis on social media data as a case study.

The remainder of the paper is structured as follows: Section 2 examines the Literature of related works. The methodology and dataset are explained in Section 3. Section 4 analyze the results of the methodology and limitations followed by the conclusion in Section 5.

2 | LITERATURE REVIEW

To buttress our research with solid backgrounds, we conducted a number of literature review which are related to study. In a study by Singh et al. (2018) [3] they considered one such government policy is the Indian government's demonetization of high-denomination money, which went into effect at midnight on November 8, 2016. They used the concept of sentiment analysis and Twitter to examine this government policy from the standpoint of the average citizen. In another study by (Manguri et al., 2020) [4] they used Tweepy library was used to extract data from Twitter social media, and then the sentiment analysis operation was completed using the TextBlob library in Python. Chowdary et al. (2020) [5] used POS-N-gram tokenization technique to extract word tokens for identifying sentiments or opinions in key utterances (tweets) about GST. The goal of their suggested research was to increase the sentiment classification accuracy of the review data by using the least number of phrases possible. Kaurav et al., (2020) [6] underlined the concerns and focus of NEP 2020 in another study. To understand crucial areas of concentration in policy texts, the authors used qualitative data analysis methodologies and computer-assisted qualitative data analysis software. Balakrishnan et al. (2012) [7] discussed a strategy in which a stream of tweets from the Twitter microblogging site were pre-processed and classified as positive, negative, or irrelevant based on their emotional content; it examined the performance of several classifying algorithms in such cases based on their precision and recall. Abeywardena (2014) [8] aimed to learn more about public attitudes of OER, MOOCs, and their complimentary functions. Within a 12-month period, raw Twitter data in the domains of OER and MOOC was analyzed using a text mining approach.

Pagolu et al. (2017) [9] in a study investigated the link between stock market movements of a firm and sentiments in tweets using sentiment analysis and supervised machine learning concepts on tweets retrieved from twitter. Pokharel (2020) [10] also took into account tweets from twelve countries in a study. The tweets were collected between March 11th and March 31st, 2020, and were somehow related to COVID19. The researcher looked at how people in other countries are dealing with the problem. In another related work by Zhang et al. (2021) [11] the researchers used a sampling of Twitter data to uncover the fundamental factors of carbon tax discussions to cluster tweets based on phrases that indicate people's attitudes, the bisecting k-means algorithm and correspondence analysis were utilized. Sharma et al. (2016) [12] during the campaigning season for general state elections in 2016, in a study performed data (text) mining on 42,235 tweets collected over a month that referenced five national political parties in India. They analyzed how Twitter users felt about each of the Indian political parties under consideration.

3 | METHODOLOGY

Our approach is based on the premise that people share their feelings, thoughts, and attitudes (sentiments) on social media platforms. These feelings and thoughts are expressed in short phrases, containing terms and words, which symbolize their secret beliefs and attitudes to policies, concepts, and other things. In our case, the terms and words used to symbolize the hidden driving factors of people’s attitudes towards the electronic levy is considered. In our methodology as represented in figure 1, we first identify specific hashtag in the collected tweets sample, then we preprocessed the data. We use text extraction techniques to extract the needed features. The polarity on each categorized tweet is determined to measure the sentiment in each sample. Finally, we compare the sentiments in different phases as well as the full data collected for the study.

3.1 | COLLECTION OF TWEET DATA

We gathered data using Twitter. In the study, snscrape python library was utilized to scrape tweets without requiring personal API (Application programming interface). Moreover, in snscrape thousands of tweets are returned in seconds, and extensive search features allow for highly customizable searches. In this paper we used the hashtag #elevy in all the tweets posted between November 17, 2021 to January 31, 2022 using Ghana as a targeted country, because the government has proposed an electronic levy tax on mobile transactions and sentiments have been shared from the general public. We have made several attempts to gain access to as many posts as feasible where we divided them in phases. During the first phase, which lasted from November 17 to November 30, 2021, a total of 1,400 tweets were collected. In addition, the second phase, 4,554 tweets were collected, December 01 to 14, 2022. The third phase of the data was collected from December 15 to 30 with a total of 7,679 tweets. Furthermore, the fourth and fifth phases were from December 31, 2021 to 15 January 2022, and January 16 to 31, 2022 with a total tweet of 1,701 and 18,423 respectively. In this study a total of 33,757 tweets were used for our analysis. The 2-week interval tweet collection is shown in Table 1 below;

Table 1 2-Week Interval Tweet Collection

Phase	***Date Interval***	***Total Tweets***
Phase 1	17/11/2021 – 30/11/2021	1,400
Phase 2	01/12/2021 – 14/12/2021	4,554
Phase 3	15/12/2021 – 30/12/2021	7,679
Phase 4	31/12/2021 – 15/01/2022	1,701
Phase 5	16/01/2022 – 31/01/2022	18,423
Total	November 17, 2021 – January 31, 2022	33,757

3.2 | PREPROCESSING OF THE TWEETS

Pre-processing the data is the process of cleaning and preparing the text for classification. Online texts contain usually lots of noise and uninformative parts such as HTML tags, scripts and advertisements (Haddi et al., 2013). In this study, we reduced the noise in the text in order to help improve the performance of the classifier and speed up the classification process, thus aiding in real time sentiment analysis. The terms or phrases (features) that strongly express the opinion as positive or negative are extracted. The procedure for the preprocessing stage is: the transformations, which are, online text cleaning, white space removal, expanding abbreviation, stemming, stop words removal, negation handling and finally, the filtering stage which is known as feature selection. The preprocessed stage is described in the equation below;

$$TF - IDF = \mathrm{FF} * \log(N|DF) \quad (1)$$

where N indicates the number of documents, and DF is the number of documents that contains the feature. FP takes the value 0 or 1 based on the feature absent/presence in the document.

3.3 | PREPROCESSING OF THE TWEETS

Sentiment Analysis (SA) is a specified issue in the Natural Language Processing (NLP) discipline that has been a source of concern. There are a variety of approaches for determining the state of sentiment (positive or negative feeling) in a text (Manguri et al., 2020). The procedure involved in Sentiment Analysis has a way it follows. The information (tweets) was gathered for this investigation from twitter. Data pre-processing was performed and the feature extraction is applied to extract the required data. The polarity was calculated to check the neutral, positive and negative counts in the data.

$$f_m = \begin{cases} \mathrm{f(posScore)}, & if\ f(\mathrm{posScore}) \leq\ f(\mathrm{negScore}) \\ -\mathrm{f(negScore)}, & otherwise \end{cases} \quad (2)$$

where f_m computes the absolute maximum of the two scores. It is worth noting that f(negScore) is always positive by construction. To obtain a final prior polarity that ranges from -1 to 1, the negative sign is imposed. Lastly, in the phase of sentiment polarity and subjectivity, decisions are made/conducted where the analysis is plotted. The figure 1 below represents the procedure used in this study.

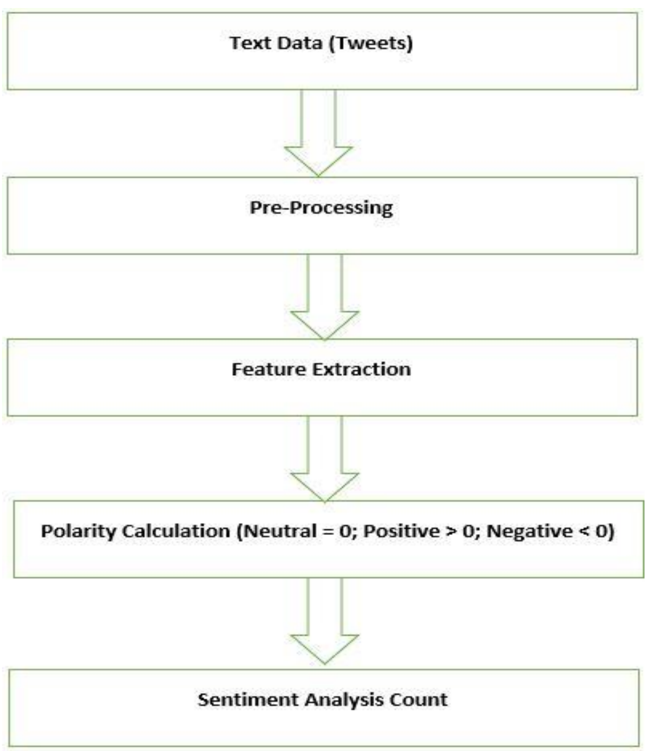

Figure 1 The procedure used in the study

The table 2 below shows sample tweets used in this study.

Table 2 Sample tweets used in the study

No	***Tweets***
1	E-Levy: Where we've gotten to, borrowing is not an option; taxing is the way. But the question is, can we trust government to make good use of our taxes? - Hassan Ayariga \n\n#OneOnOne
2	Showboy @NAkufoAddo, please increase the E-Levy and add internet tax too. \n\nThey've enough money to moan on Spaces
3	The e-levy is the most long-term approach to improve public budgets. â€“ John\xa0Boadu https://t.co/HGX5iOOgXz'
4	'Please find another way to tax and forget about this e-levy. 3nfa https://t.co/oUoEVJlwES'
5	'NPP clutching at straws now. the â€œyÉ›te sika so no...â€• ein this? ein be the E-levy no? https://t.co/OoY7jI8APN'

4 | RESULTS AND FINDINGS

4.1 | GHANAIAN INTERNET USERS

According to Statista.com (2022), Ghana had around 16 million internet users as of January 2021, up from 14.76 million the previous year. The figure 2 shows the internet trend in Ghana.

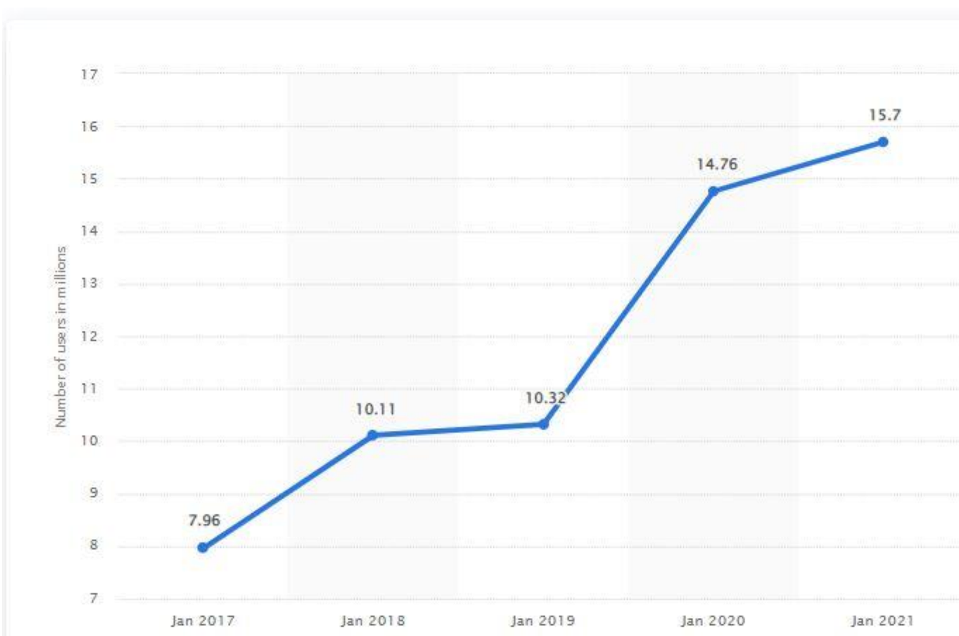

Figure 2 Internet usage trends in Ghana (Source: Statista.com, 2022)

4.2 | SOCIAL MEDIA USAGE TREND IN GHANA

Ghana had approximately eight million social media users as of January 2021. According to data from Statista.com (2022). The figure 3 below is a trend of social media message in Ghana.

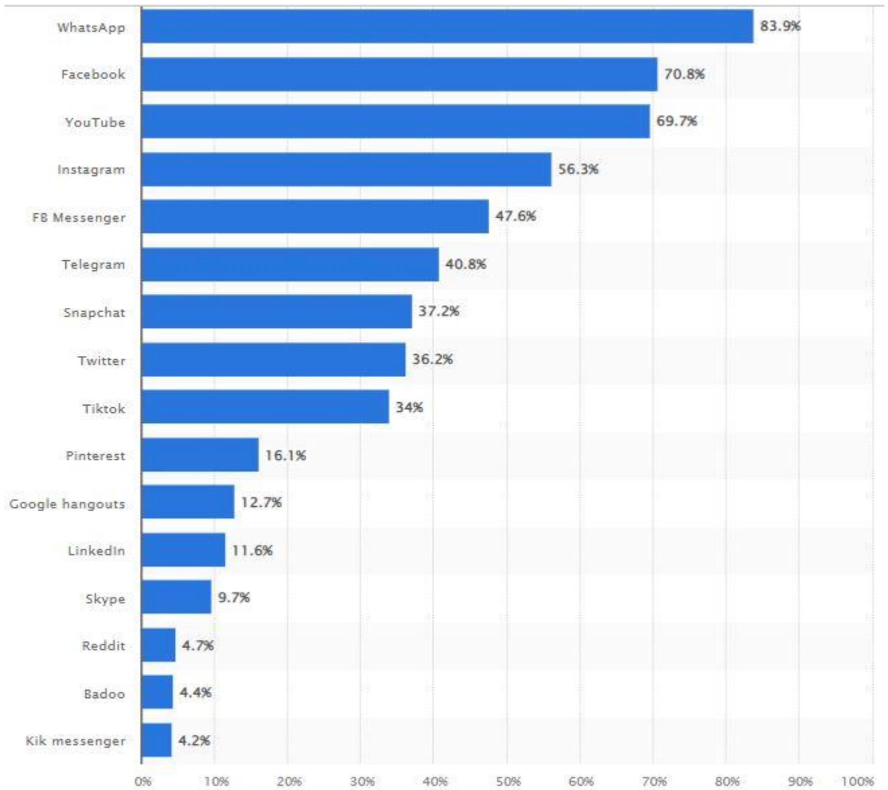

Figure 3 Social media trends in Ghana (Source: Statista.com, 2022)

In this study we considered data from twitter which represents 36.2% among social media users in Ghana.

4.3 | SENTIMENTAL ANALYSIS

Sentiment analysis was applied on the data that were collected within the 2-week interval period. Sentiment analysis is the method of identifying sentiments from a given text. This enables us to comprehend the emotions of someone who has written a text about the E-Levy policy in Ghana. The results on the sentiment analysis are as follows: P (positive), N (negativity), and Neutral. The outcomes of the study with their associated word cloud are classified in five phases, thus, the first, second, third, fourth, and fifth phases.

4.3.1 | THE FIRST PHASE

The first data was collected from November 17 to 30, 2021 when the E-Levy was first announced in parliament. A total of 1,400 tweets were collected. The figure 4 and figure 5 represents the sentiment analysis and word count respectively.

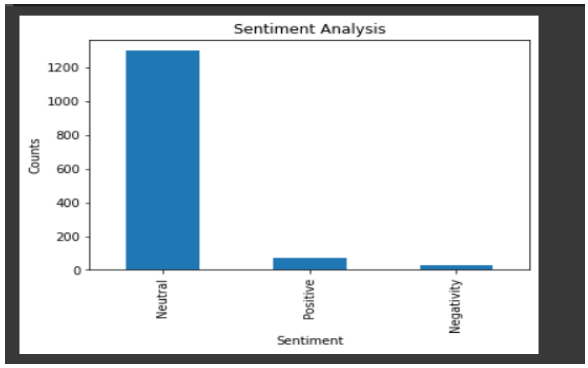

Figure 4 Sentiment analysis for phase 1

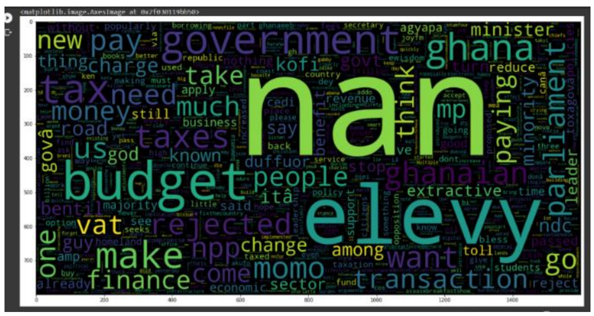

Figure 5 Word count for phase 1

The Phase 1 recorded a total of 1,400 tweets from when it was announced on November 17, 2021 to 30 November, 2021. Out of the total, 1,250 (89.29%) tweets were neutral tweets, while 125 (8.93%) and 25 (1.78%) tweets were positive and negative tweets. People did not have a clear understanding of the policy.

4.3.2 | THE SECOND PHASE

The second phase was when the general public have become aware/heard much of the E-Levy policy. From December 1 to 14, 2021 a total of 4,554 tweets were collected. Figure 6 and figure 7 are the sentiment analysis and word count of the second phase.

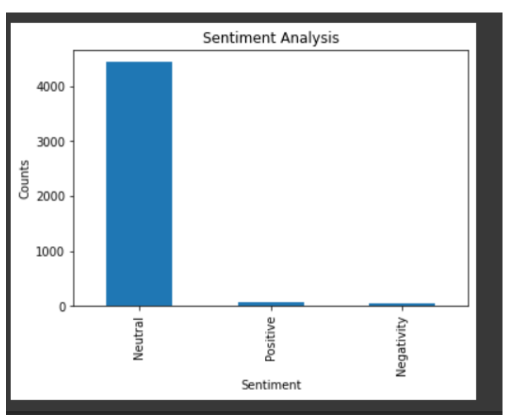

Figure 6 Sentiment analysis for phase 2

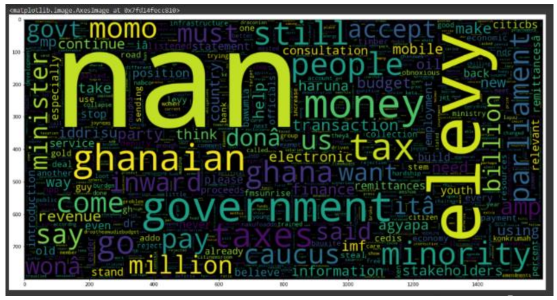

Figure 7 Word count for phase 2

It can be observed that, the data increased during the second phase. Out of the total tweets, 4500 (98.81%) were neural, 38 (0.83%) were positive, and 16 (0.36%) were negative.

4.3.3 | THE THIRD PHASE

A total of 7,679 tweets were gathered during the third phase from December 15 to 30, 2021. During this time period, the bill was called before parliament to be considered. The sentiment analysis and word count is represented in figure 8 and figure 9.

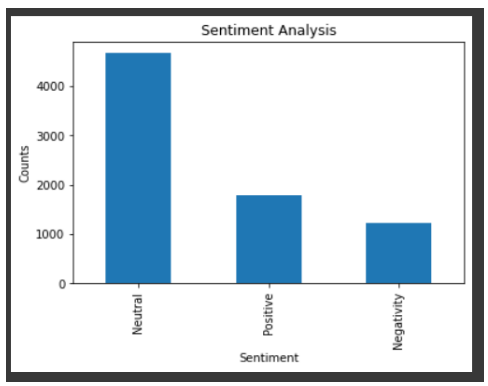

Figure 8 Sentiment analysis for phase 3

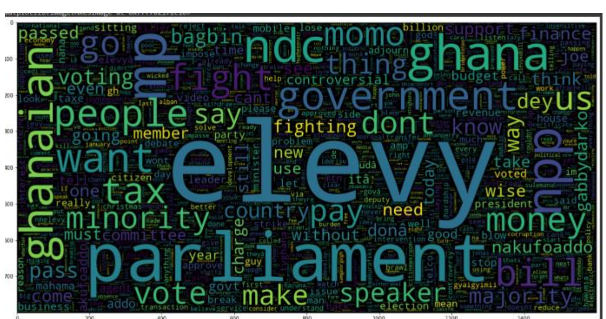

Figure 9 Word count for phase 3

From the results in Phase 3, 4647 (60.52%) tweets were neutral. 1980 (25.79%) tweets and 1052 (13.69%) tweets were positive and negative respectively. There was an increase in the positivity counts which can be attributed to the controversies the policy garnered in the mainstream media after the brouhaha in parliament.

4.3.4 | THE FOURTH PHASE

Between December 31, 2021 to January 15, 2022, 1,701 total tweets were collected in phase 4 with figure 10 and figure 11 below as the sentiment analysis and word count.

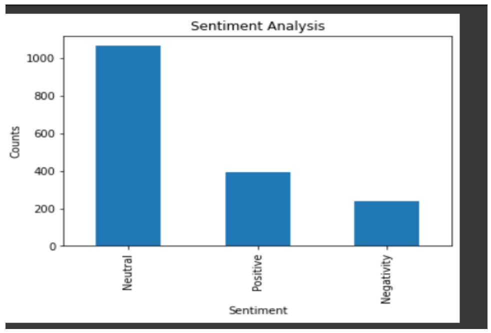

Figure 10 Sentiment analysis for phase 4

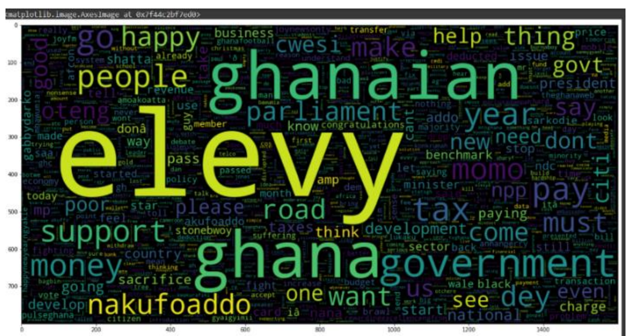

Figure 11 Word count for phase 4

During this period parliament was on recess and the issue of E-Levy was not popular in the mainstream media, mostly due to the Christmas and New Year festivities. Positive tweets were 400 (23.52%), negative tweets were 200 (11.5%), and 1101 (64.73%) tweets were neutral. There was a slight drop in the positivity count as compared to the third phase.

4.3.5 | THE FIFTH PHASE

The fifth phase recorded a total of 18,423 tweets. This took place between January 16 to 31, 2022, and the figure 12 and figure 13 represents the sentiment analysis and word count below;

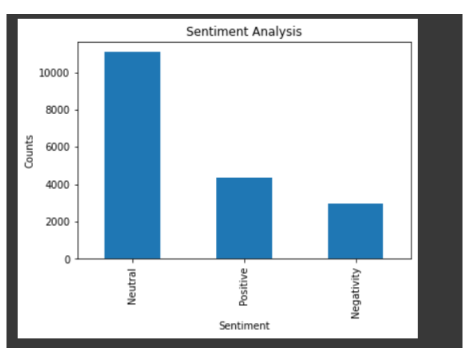

Figure 12 Sentiment analysis of phase 5

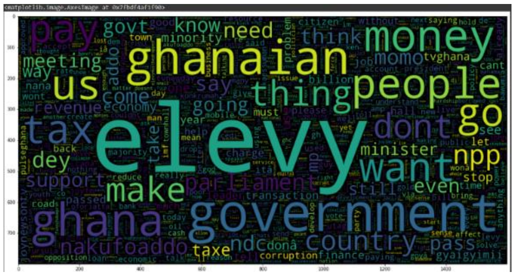

Figure 13 Word count of phase 5

From the analysis 10,923 (59.29%) tweets were neural, 4,500 (24.43%) tweets were positive, and 3,000 (16.28%) tweets were negative. It can be seen from the analysis that; the biggest data was collected during the fifth phase with the highest positivity. During this period, parliament recalled the policy to be discussed, and town hall meetings were also held by government to explain the policy to the general public.

4.3.6 | EXPERIMENTAL ANALYSIS OF THE PHASES

Table 3 shows the results of the analysis.

Table 3 Analysis of the phases

Phase	***Date***	***Pos***	***%***	***Neu***	***%***	***Neg***	***%***	***Total***
Phase 1	17/11/2021 – 30/11/2021	125	8.93	1,250	89.29	25	1.78	1,400
Phase 2	01/12/2021 – 14/12/2021	38	0.83	4,500	98.81	15	0.36	4,554
Phase 3	15/12/2021 – 30/12/2021	1,980	25.79	4,647	60.52	1,052	13.69	7,679
Phase 4	31/12/2021 – 15/01/2022	400	23.52	1,101	64.73	200	11.75	1,701
Phase 5	16/01/2022 – 31/01/2022	4,500	24.43	10,923	59.29	3,000	16.28	18,423
Total	November 17, 2021 – January 31, 2022	7,043		22,421		4,293		33,757

The figure 14 below represents the trend in the number of tweets recorded in each phase graphically.

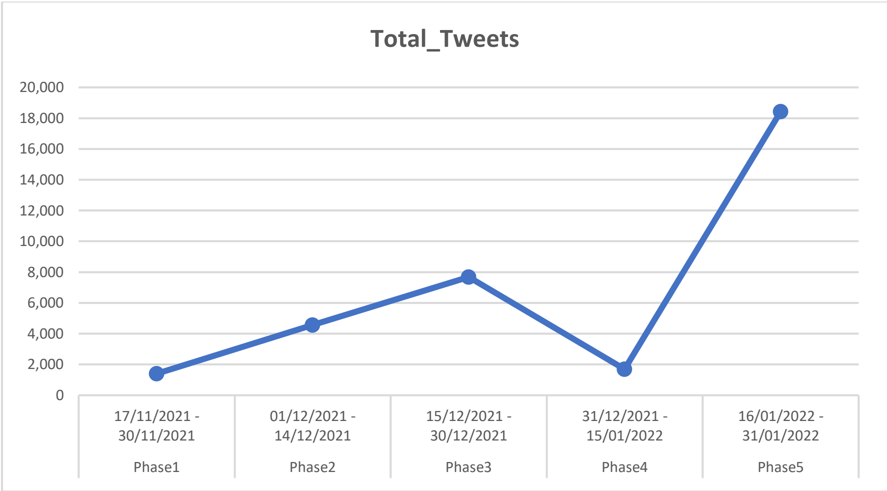

Figure 14 graphical representation of the tweets trend

4.4 | SENTIMENT ANALYSIS OF TOTAL DATASET

In order to determine the efficacy of our model and check for possible biasness, we performed sentiment analysis in figure 15 and word count in figure 16 on the full dataset (38,771 tweets) collected from November 17, 2021 to January 31, 2022.

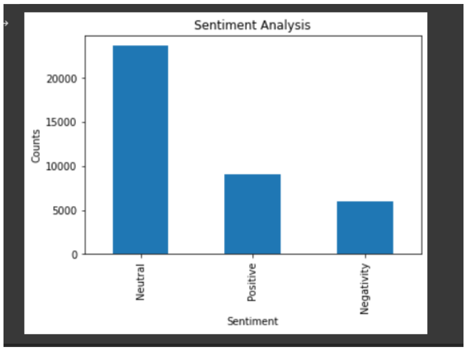

Figure 15 Sentiment analysis for full dataset

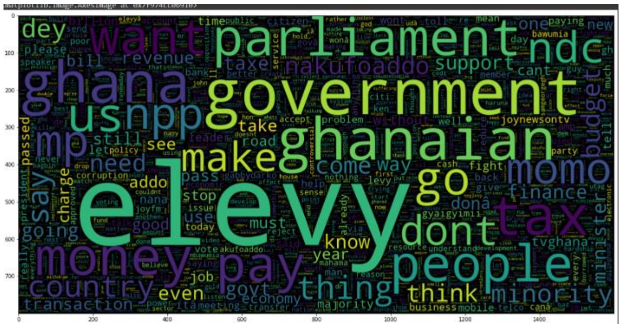

Figure 16 Word count for full dataset

In the full analysis, 9,887 (25.50%), 6,000 (15.48%), and 22,884 (59.02%) were positive, negative and neutral. The bigger the dataset, the positive sentiment increases.

4.5 | DISCUSSION

Based on the results of our analysis, the driving factors of the Ghanaian people's attitudes toward the implementation of the e-levy can be categorized into three: positive, neutral, and negative. Positive refers to the section of the populace who are in support of the proposed e-levy. Neutral means the people do not understand the proposed e-levy (neither in support nor against), while negative are people strongly against the proposed e-levy. Individuals around the globe support policies that minimize their expenses and maximize their benefits. For this study, we categorized our dataset into five different phases to ascertain the real sentiments of the people using the twitter data. Moreover, a full dataset analysis was also conducted to prevent any possible biasness in the data collection. From the analysis, it can be observed that the number of positives increase steadily whenever the total amount of data increases in each phase. From this study, it can be observed that large portion of the populace have not clearly understood (neutral) the e-levy policy. In order for the government to gain the support of the citizens, we recommend that the townhall meetings be held nationwide and translated into the local dialects. We strongly believe that, when the proposed e-levy policy is understood, the more positive the sentiments to support the policy.

4.6 | LIMITATIONS

In this study we collected data from the microblogging site twitter, however, our research shows that larger portion of social media users in Ghana are on Facebook. It was challenging to access data from Facebook which can be used for this study.

5 | CONCLUSION

With the announcement and implementation of new government policies, there is aways some positive and negative repercussions. A typical example is the proposed E-Levy policy to charge 1.75% on mobile transactions. The goal of this study was to use sentiment analysis to assess the impact of the proposed e-levy policy. Our research found that a considerable percentage of Ghanaians were neither delighted nor unhappy (neutral) with the policy, because the average person had little or no awareness of the policy. Eventually, as the policy gained popularity on mainstream and social media, the positive sentiments of the people increased steadily.

The full data analysis led us to conclude that, in broader terms, the large number of Ghana people are neutral on the proposed policy with little difference in positive and negative. Based on the aforementioned, we suggest that the government take a critical look at the policy before its implementation. This is to curb the effects of possible drop in mobile transactions to sabotage our vision as a cashless country. The government should increase its town hall meetings in the local dialects for easy understanding. Moreover, social media users in Ghana use Facebook a lot and access to the data can be used.

ACKNOWLEDGMENTS

We are grateful to the microblogging site twitter for making the data available for this study.

DATA AVAILABILITY

The data that support the findings of this study are openly available in figshare at http://doi.org/10.6084/m9.figshare.19209747

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest in this paper.

REFERENCES

[1] GlobalSecurity.org, “British Gold Coast - Economic Development,” 2019. [Online]. Available: https://www.globalsecurity.org/military/world/africa/gh-history-3-4.htm. [Accessed: 14-Feb-2022].

[2] GhanaWeb.com, “Today in History: The 1995 ‘Kume Preko’ protest that rocked Ghana,” 2020. [Online]. Available: https://www.ghanaweb.com/GhanaHomePage/NewsArchive/Today-in-History-The-1995-Kume-Preko-protest-that-rocked-Ghana-948484. [Accessed: 14-Feb-2022].

[3] P. Singh, R. S. Sawhney, and K. S. Kahlon, “Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by Indian government,” *ICT Express*, vol. 4, no. 3, pp. 124–129, 2018.

[4] K. H. Manguri, R. N. Ramadhan, and P. R. Mohammed Amin, “Twitter Sentiment Analysis on Worldwide COVID-19 Outbreaks,” *Kurdistan J. Appl. Res.*, pp. 54–65, 2020.

[5] E. Deepak Chowdary, S. Venkatramaphanikumar, and K. Venkata Krishna Kishore, “Aspect-level sentiment analysis on goods and services tax tweets with dropout DNN,” *Int. J. Bus. Inf. Syst.*, vol. 35, no. 2, pp. 239–264, 2020.

[6] R. P. S. Kaurav, K. G. Suresh, S. Narula, and R. Baber, “Qualitative (Contents) Analysis and Twitter Mining (Sentiment Analysis),” *J. Content, Community Commun.*, vol. 12, no. November 1956, pp. 4–13, 2020.

[7] As. P. Balakrishnan Gokulakrishnan, Pavalanathan Priyanthan, Thiruchittampalam Ragavan, Nadarajah Prasath, “Opinion Mining and Sentiment Analysis on a Twitter Data Stream,” p. 229, 2012.

[8] I. S. Abeywardena, “Public opinion on OER and MOOC: A sentiment analysis of Twitter data,” *Emerg. modes approaches open Flex. Educ.*, no. October, pp. 296–307, 2014.

[9] V. S. Pagolu, K. N. Reddy, G. Panda, and B. Majhi, “Sentiment analysis of Twitter data for predicting stock market movements,” *Int. Conf. Signal Process. Commun. Power Embed. Syst. SCOPES 2016 - Proc.*, pp. 1345–1350, 2017.

[10] B. P. Pokharel, “Twitter Sentiment Analysis During Covid-19 Outbreak in Nepal,” *SSRN Electron. J.*, no. March, pp. 1–9, 2020.

[11] W. I. Yaming Zhanga, Majed Abbas, “Analyzing sentiments and attitudes toward carbon taxation in Europe, USA, South Africa, Canada and Australia,” *Sustain. Prod. Consum.*, vol. 28, pp. 241–253, 2021.

[12] P. Sharma, “Prediction of Indian Election Using Sentiment Analysis on Hindi Twitter,” pp. 1966–1971, 2016.

[13] E. Haddi, X. Liu, and Y. Shi, “The role of text pre-processing in sentiment analysis,” *Procedia Comput. Sci.*, vol. 17, no. December, pp. 26–32, 2013.

[14] Statista.com, “Ghana: active social media users 2017-2021 | Statista,” 2022. [Online]. Available: https://www.statista.com/statistics/1171445/number-of-social-media-users-ghana/. [Accessed: 14-Feb-2022].