Social Network Analysis and Mining (2024) 14:91
https://doi.org/10.1007/s13278-024-01248-3

ORIGINAL ARTICLE

Sentiment analysis of tweets on social security and medicare

Unmesh Kumar Chakravarty[1] · Shaikh Arifuzzaman[2]

Received: 27 June 2023 / Revised: 14 March 2024 / Accepted: 19 March 2024

Abstract

The field of politics can be greatly aided through the use of natural language processing techniques on places like social media. Twitter hosts extensive discussions on political topics like Social Security and Medicare. We gathered almost 90,000 tweets on Social Security and Medicare for sentiment analysis. Positive sentiment was higher than negative sentiment, and public discussion was polarized with relatively little neutral sentiment. We conducted named entity recognition and found that entities of people and organizations were most prevalent in discussion regardless of sentiment. Topic modeling was also performed, and we determined the dominant topics in tweets. We compared word counts of keywords to the probabilities that each belongs to their respective topic in order to gauge the impact of keywords in their topics. The discussion of Social Security and Medicare on Twitter is objectified through named entity recognition and topic modeling. The experimentation conducted can be broadly applied to politics to better understand objects and themes of key interest in various complex issues that are debated and discussed on Twitter. This study provides a comprehensive structure to the public discussion of Social Security and Medicare on Twitter and assists politicians and lawmakers in making significant, relevant decisions and policies.

Keywords Twitter · Sentiment analysis · Named entity recognition · Topic modeling · Latent dirichlet allocation

1 Introduction

Language has been a fundamental construct for much of human history. Although its origins are up for debate (Hauser et al. 2014; Washburn and Lancaster 2017; Dunbar 1998; Deacon 1998; Power 1998; Burling, et al. 1993; Laland 2017), language has evolved into the rich medium of expression it is today. Machine learning and natural language processing techniques can yield deeper insights amid everyday conversation and remarks. Social media boasts an extensive inventory of daily language filled with the thoughts and feelings of users across the world, effectively becoming a platform hosting information, opinions, and arguments of all kinds.

Social media platforms act as archives recording public responses to events and ideas. Twitter provides a database made easily accessible through its API. Politicians have increasingly adopted Twitter as a means of broadcasting their ideals as well as assessing the voices of their constituents. Welfare programs are a product of politics and aim to provide aid financially or by other means to those who cannot support themselves. Social Security and Medicare offer benefits to seniors and retirees, and talks of cuts, expansion, and privatization of these programs have been around for over a decade. The discussion of the matter on Twitter has persisted to a new degree recently with inflation reaching a record high, projected trust fund depletions approaching, and Congress circulating plans to sunset Social Security and Medicare.

This paper intends to gauge and analyze public sentiment on Social Security and Medicare. Tweets mentioning Social Security or Medicare are retrieved on the basis of assessing public response to the matter. This provides lawmakers and politicians the opportunity to understand and cater to the thoughts of their constituents through meaningful policy and decision-making.

Sentiment analysis is the interpretation of sentiment in textual data. Utilizing natural language processing and machine learning techniques, sentiment analysis aims to discern the underlying sentiment or emotional tone conveyed

✉ Unmesh Kumar Chakravarty
unmeshchak70@gmail.com

Shaikh Arifuzzaman
shaikh.arifuzzaman@unlv.edu

1 Haynes Academy for Advanced Studies, Metairie, LA, USA

2 University of Nevada, Las Vegas (UNLV), Las Vegas, NV, USA

within a given text, typically categorized as positive, negative, or neutral. Named entity recognition is a natural language processing task that involves the identification and classification of named entities within textual data. These entities encompass specific instances of predefined categories, such as persons, organizations, locations, dates, and other relevant terms. Topic modeling is a computational technique within natural language processing that aims to identify latent thematic structures inherent in a corpus of textual data. Through probabilistic modeling and unsupervised learning methods, topic modeling endeavors to discover and characterize distinct topics or themes present in a collection of documents. These identified topics are represented as sets of co-occurring words, allowing for the extraction of meaningful semantic structures within the text. Widely employed algorithms for topic modeling, such as Latent Dirichlet Allocation (LDA) (Blei et al. 2003), contribute to the enhanced comprehension of the underlying content and facilitate the organization and categorization of large textual datasets.

Sentiment analysis is typically approached by two primary methodologies: machine learning and lexicon-based approaches. Machine learning models leverage algorithms and annotated datasets to autonomously learn patterns and associations between textual features and sentiment labels. These models, including support vector machines, neural networks, and ensemble methods, excel in capturing nuanced and context-dependent sentiments. On the other hand, lexicon-based methods rely on predefined sentiment lexicons or dictionaries containing words annotated with sentiment scores. These approaches assign sentiment based on the presence of sentiment-bearing terms in the text, enabling a rule-based analysis. Due to the lack of pre-labeled training data, we opted to use a lexicon-based approach. Additionally, while machine learning approaches showcase adaptability to diverse contexts and evolving language nuances, lexicon-based methods offer transparency and interpretability which is beneficial for the purposes of providing quantitative structure to the public discussion on Twitter regarding Social Security and Medicare.

After tweets are scraped and have undergone preprocessing procedures, sentiment analysis is performed to return the polarity of a tweet and classify it into one of three classes: positive, negative, or neutral. To understand the public's objects of interest, named entity recognition is performed, recognizing entities of particular types such as people, organizations, documents, products, and events in each of the sentiment classes. LDA is also performed to provide topic models across all sentiment classes. The topic models considered each tweet as its own separate document and probabilistically assigned words to each topic. These techniques provide structure to the public discussion regarding Social Security and Medicare.

Leveraging sentiment analysis, named entity recognition, and topic modeling can provide a comprehensive understanding of public sentiment, key entities, and relevant topics related to Social Security and Medicare. By analyzing tweets, politicians and lawmakers can understand the public's emotions and opinions regarding these policies: positive sentiments may indicate public support, while negative sentiments may highlight areas of concern or dissatisfaction. Named entity recognition can identify and extract key entities surrounding Social Security and Medicare and provides a quantitative measure of how frequently these entities are being discussed in the public domain. Lawmakers can use this information to prioritize policy discussions and allocate resources accordingly. Topic modeling can help identify main themes and topics within discussions related to Social Security and Medicare on Twitter which enables lawmakers to gain insights into specific issues, concerns, and aspects of these programs that are being highlighted by the public. This information can guide policymakers in addressing key concerns and shaping policies that align with the needs and preferences of the public. Ultimately, the methodologies profiled in this study can be invaluable for policymakers in making informed decisions, addressing public concerns, and shaping policies that are responsive to the dynamic nature and significance of a broad spectrum of political ideas and issues.

2 Related work

The Twitter dataset has been used to analyze the sentiment of different topics. Mittal and Goel (2012) used Twitter data and past Dow Jones Industrial Average values to determine a connection between public mood and stock market movements. Mood was divided into four classes: Calm, Happy, Alert, and Kind, and sentiment was scored using an expanded word list from the Profile of Mood States questionnaire from which words were scored on a given day. Out of the four learning algorithms Linear Regression, Logistic Regression, Support Vector Machines, and Self Organizing Fuzzy Neural Networks (SOFNN), SOFNN performed the best with 75.56% accuracy. Rasool et al. (2019) used Naïve Bayes and a lexicon dictionary to analyze positive, negative, or neutral sentiment in tweets to compare the apparel brands Nike and Adidas. They found that Adidas had more positive reviews than Nike and little difference in negative reviews. They also found that other brands were referenced as a means of comparison by consumers. Zhou et al. (2020) used Twitter data to analyze sentiment dynamics during the COVID-19 pandemic in people living in the state of New South Wales (NSW) in Australia to focus on local government areas (LGAs) instead of a country-level basis. The results show that the people had an overall positive sentiment which decreased through the pandemic and found profound changes in sentiment in some LGAs. Analyzing hot topics on Twitter at the time showed that policies and events did impact sentiment. Sattar and Arifuzzaman (2021) collected Twitter data on COVID-19 vaccines and lifestyle after vaccination and used lexicon-based tools TextBlob and VADER for sentiment analysis. COVID-19 vaccination had received more positive sentiment with 20–25% of tweets being positive with only 10% being negative and the majority being neutral. They found that for them, VADER had a tendency to classify neutral tweets as positive or negative. They provided a timeline of sentiment of vaccines and were able to visually distinguish the side effects of some vaccines. Positive sentiment was always ahead of negative sentiment when it came to lifestyle choices after vaccination and a sentiment timeline was provided as well. They also used different machine learning models (Support Vector Machine, k-Nearest Neighbor, Linear Regression, Random Forest, M5 Model Tree, Gaussian process for regression, and Multilayer Perceptron) to forecast the population of those in the USA who are vaccinated. They decided that the Multilayer Perceptron model performed best for a forecast time of 7 days and the Support Vector Machine model (which predicted that around 48% of the population would be fully vaccinated by 25 July 2021) worked best for a forecast time of 2 months.

Several works have used the Twitter dataset for political settings. Wang et al. (2012) provide a system for the sentiment analysis of presidential candidates in the 2012 US election with real-time Twitter data. They used a crowdsourcing approach using Amazon Mechanical Turk for sentiment annotation on data to be used for model training and testing and a Naïve Bayes model for sentiment classification which performed at 59% accuracy which exceeded a baseline of 56% for classifying everything as negative, the most prevalent sentiment class. Pla and Hurtado (2014) provide an approach to determining the political tendencies of Twitter users and used a general corpus developed at TASS2013 workshop for Spanish. Tweets are first preprocessed with tokenization, lemmatization, named entity recognition, and part-of-speech tagging. Tweets are then heuristically segmented by political entities to determine sentiment polarity for each entity using a Support Vector Machine approach. They determined political tendency by indicating that a positive opinion on a party of a certain political orientation signals that the user is of similar political orientation while a negative opinion signals the opposite. They reported the best result so far for their corpus. Bermingham and Smeaton (2011) investigated whether political sentiment on Twitter was a good indicator of election results and used an Irish General Election as a case study. They used classifiers trained on data specific to this election and had nine annotators for annotating sentiment. They used an Adaboost Multinomial Naïve Bayes that had 65.09% accuracy. They concluded that Twitter did appear to have predictive value with volume possibly being a stronger indicator than sentiment.

The Twitter dataset has been used for analyzing public discourse on a breadth of topics such as healthcare, energy policy, and higher education. Ainley et al. (2021) used Twitter content to understand health and care delivery in the United Kingdom because of COVID-19, particularly focusing on remote care delivery. 637 tweets posted in the United Kingdom between January 2018 and October 2020 were assigned sentiment and distributed into categories using NVivo software after extraction. The results showed that the volume of tweets on remote care delivery increased markedly following the COVID-19 outbreak. Five main themes were identified: access to remote care, quality of remote care, anticipation of remote care, online booking and asynchronous communication, and publicizing changes to service or care delivery. They found mixed public attitudes and experiences to the changes in service delivery. The proportion of positive tweets regarding the access to and quality of remote care was higher in the immediate period following the COVID-19 outbreak than the time before the COVID-19 onset and when restrictions from the first lockdown eased. They determined that due to this decline in positive sentiment, there is a need for continued examination of people's preferences. Kim et al. (2021) examined public sentiment toward solar energy in the United States using Twitter. Sentiment classification was performed using Robustly optimized Bidirectional Encoder Representations from Transformers. Their model achieved 80.2% accuracy for the classification of tweets as positive, neutral, or negative. They analyzed 266,686 tweets during the period of January to December 2020 and found public sentiment to vary widely across states with a percent contribution of 164.66%. The results show that the Northeast U.S. region had more positive sentiment than the South U.S. region. States with a larger share of Democratic voters in the 2020 presidential election were more positive in public opinion on solar energy. Public sentiment toward solar energy is more positive in states with consumer-friendly net metering policies and a more mature solar market. Lasri et al. (2023) proposed a deep learning approach to analyze people's sentiment in tweets on distance learning in higher education. They extracted 24,642 tweets from July 20, 2022 to November 6, 2022. For sentiment classification (positive, negative, or neutral), they used a self-attention-based Bi-LSTM model with GloVe word embedding which was compared to LSTM (Long Short Term Memory), Bi-LSTM (Bidirectional-LSTM), and CNN-Bi-LSTM (Convolutional Neural Network-Bi-LSTM). The model obtained the best test accuracy of 95% on a stratified 90:10 split ratio. Their results showed that there was overall neutral sentiment in regard to distance learning for higher education, followed by positive sentiment, particularly in psychology and computer science, and negative sentiment in biology and chemistry. Their approach outperformed state-of-the-art methods.

To our knowledge, sentiment analysis of Social Security and Medicare with Twitter data has not been done. Political elections, given their dynamic nature and high stakes, make an interesting and popular focus of study. However, a study involving a specific policy or program can also be interesting and insightful. Meaningful policies and programs often expand and outlive their makers while impacting the lives of millions. Times that allude to the possible end of these significant pieces of history in the future can spark controversy and debate within the public. These discussions can be analyzed to understand the extent of influence these policies and programs have had and help direct the course of action to be taken. Social Security and Medicare are prime examples of such. Social Security began in 1935 during the troubled times of the Great Depression in the US as a limited form of insurance to the elderly and families with dependent children. Since then, it has expanded to provide hundreds of billions of US dollars to tens of millions of Americans. Its origins stem from widespread poverty among the elderly which has since been effectively reversed. Medicare started in 1965 under the Social Security Administration as a form of government national health insurance and, like Social Security, has provided hundreds of billions of US dollars to tens of millions of Americans. With prospects of the end to these wide-reaching and long-lasting programs, Social Security and Medicare provide a unique choice of study through their respective Twitter data.

3 Methods

In this section, we describe the extraction and preprocessing procedures for our Twitter dataset and mention the Python libraries used.

3.1 Dataset extraction

Twitter API (Accessed 2022) gives users the opportunity to request tweets from their archive that fit queries specified by the user. These queries can contain keywords to be found in resulting tweets as well as other classifications like geolocation and time. We used the Tweepy library (Roesslein 2020) to gather 89,998 unique tweets. These tweets contained the keywords "social security", "medicare", and "medicaid". We opted for the use of "medicaid" as a keyword due to Medicaid's general association with Medicare. We also filtered out any retweets as part of the query to Twitter API.

3.2 Preprocessing

After gathering the tweets, the following preprocessing procedures were done in preparation for sentiment analysis: (1) We removed any punctuation and digit characters, (2) Any links were removed, (3) The result was then tokenized, (4) We removed any stopwords provided by the NLTK library (Bird et al. 2009), and (5) All words were lemmatized using the NLTK WordNetLemmatizer. No word grouping, n-grams, multiwords, sentences, or other forms of textual structure were done, and sentiment would be scored on a word-by-word basis. Figure 1 provides a visual description of the dataset preparation workflow. Table 1 shows an example of preprocessing one of the tweets from the dataset.

3.3 Sentiment analysis

Sentiment analysis is performed with the intent of classifying text based on its sentiment, namely the feeling, emotion, or opinion attached to the text. For our purposes, sentiment is classified into one of three classes: positive, negative, and neutral. Sentiment analysis can be done through two main approaches: machine learning or deep learning using techniques such as Support Vector Machines and Artificial Neural Networks, or lexicon-based approaches like using a sentiment dictionary. Due to not having a prelabeled dataset, we decided to use a lexicon-based approach. We used VADER (Hutto and Gilbert 2014) as it was designed for sentiments found specifically in social media. VADER returns a polarity score for each tweet and each sentiment class is determined as follows: a tweet is positive if the polarity score is greater than or equal to 0.05; if the polarity score is exclusively between −0.05 and 0.05, the tweet is neutral; otherwise, when the polarity score is less than or equal to −0.05, the tweet is negative. Table 2 shows some examples from the dataset with their respective polarity scores.

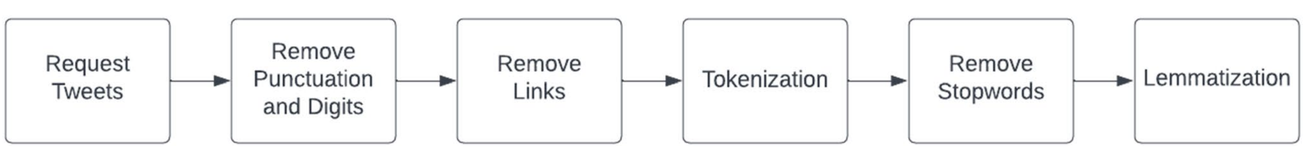


Fig. 1 Dataset preparation visualized: the workflow behind the preprocessing procedures for the dataset

3.4 Named entity recognition

Named entity recognition is the process of identifying named entities and classifying them into predefined categories. We used the spaCy library (Honnibal and Montani 2017) for named entity recognition. spaCy has trained pipelines beyond that of named entity recognition for many languages.

3.5 Topic modeling

Topic modeling is an unsupervised machine learning problem that is interested in determining the abstract topics in a set of documents and has several existing approaches. We decided to use a LDA model, with each tweet being its own separate document. To do so, we used the gensim library (Rehurek and Sojka 2011). First, we create a dictionary through gensim that maps all the words to a unique id. We filter out any words that appear in less than 100 documents and any words that appear in more than 75% of the documents. We used this dictionary to create a bag-of-words corpus which we use for the gensim LDA model. This was done for each sentiment class, with each model having 5 topics. The models returned the keywords of each topic along with the probability of each keyword belonging to that topic. Figure 2 visually summarizes the steps taken for topic modeling.

3.6 Data visualization

Plots were generated using matplotlib (Hunter 2007). Word clouds were generated using the Wordcloud for Python library (Oesper, et al. 2011). Word clouds show a cluster of words from input text in such a way that more frequent words appear larger than less frequent words. We generated word clouds based on the log-likelihood ratio,

Table 1 Preprocessing workflow example: the result of each preprocessing step is shown for an example tweet from the dataset

Preprocessing steps	Tweet
Original tweet (Before preprocessing)	They want to privatize social security. They oppose diversity, equity and inclusion. Recognize these are culture war issues. https://t.co/gPfGjTIgA7
Clean tweet (Remove punctuation and digits, remove links)	they want to privatize social security they oppose diversity equity and inclusion recognize these are culture war issues
Tokenized tweet (Tokenization)	['they', 'want', 'to', 'privatize', 'social', 'security', 'they', 'oppose', 'diversity', 'equity', 'and', 'inclusion', 'recognize', 'these', 'are', 'culture', 'war', 'issues', '']
Final tweet (Remove stopwords, lemmatization)	['want', 'privatize', 'social', 'security', 'oppose', 'diversity', 'equity', 'inclusion', 'recognize', 'culture', 'war', 'issue', '']

Table 2 Example tweets with polarity scores: polarity scores outputted from VADER are shown for example tweets from each sentiment class

Original tweet	Polarity score	Sentiment classification
"It's time to stop treating universal healthcare or single-payer… as some sort of fringe, left-wing issue. No, it's a public health issue!" Me, on @MSNBC, on a new study showing 338,000 Covid deaths could have been avoided if we had Medicare-for-All: https://t.co/SVpFs8mGon	−0.7003	Negative
@mcuban @JoeBiden @LeaderMcConnell @SpeakerPelosi @GOPLeader @SenSanders @RonDeSantisFL @AOC Can you fix medicare next ?	0	Neutral
Don't pay attention to any media, that says the Dems are going to lose the House and Senate. We have great candidates, better policies, and the truth on our side. Republicans would INCREASE TAXES on working families AND seniors; gut Social Security, and END Medicare #FreshWords	0.8909	Positive

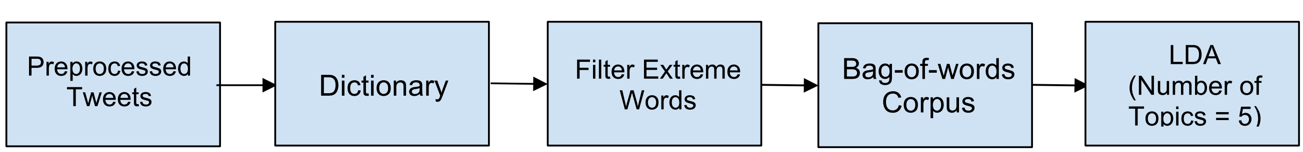


Fig. 2 Topic modeling summarized: the workflow behind the topic modeling procedures facilitated with gensim

$$\log \frac{P(w_i|x)}{P(w_i|\overline{x})},$$

unless specified otherwise. Here, w_i denotes the word, x denotes the sentiment class, and $\overline{x}$ denotes all other sentiment classes.

4 Results and discussion

In this section, we provide and summarize the results of our experimentation. The main experimental scheme was first classifying each tweet into one of three sentiment classes: positive, negative, or neutral. All other experimentation was done to compare results for each sentiment class.

4.1 Sentiment analysis

Figure 3 plots the distribution of polarity scores returned by VADER for all tweets.

Figure 4 plots polarity over time. To do this, we sorted tweets by their tweet id, effectively sorting them in chronological order. Plotting this as a scatterplot would not generate useful results due to the volume of data. Instead, we used a hexagonal bin plot that plots clusters of nearby data points as a single hexagon. Figure 4 displays hexagons that contain at least 15 data points.

An interesting finding found in Figs. 3 and 4 is the volume of tweets scored at a polarity of around 0.35. Figure 3 shows a large peak at that value while Fig. 4 shows that there is a large and constant persistence of tweets of that polarity. Generally, with the distribution shown in Fig. 3 being left-skewed and the majority of hexagons in Fig. 4 having a positive polarity, sentiment has been largely positive when it comes to Social Security and Medicare. Figure 5 clearly shows this.

The low number of neutral tweets compared to positive and negative tweets reveals how disputed the discussion of Social Security and Medicare is on Twitter.

Figure 6 shows the word clouds of the tweets in each sentiment class which were generated to visually represent the frequencies of words in each sentiment class. Words in the positive word cloud such as "benefit," "care," and "support" not only reflect a sense of assurance and satisfaction but also underscore the positive impact of these programs on the lives of millions. Conversely, the negative sentiment word cloud highlights terms like "cut," "pay," and "fraud," shedding light on areas of dissatisfaction and potential challenges perceived by the public. It is evident that certain concerns, perhaps related to government policy and management, have resonated strongly within this sentiment class. The neutral sentiment word cloud, meanwhile, presents a mix of terms without strong emotional connotations, including "health," "need," and "program," and suggests a measured discourse around Social Security and Medicare through an objective lens.

Overall, policymakers can leverage the insights found in the collective polarities of tweets and their respective word clouds to address specific concerns within negative tweets, enhance aspects mentioned in positive tweets, and navigate the otherwise complex landscape surrounding Social Security and Medicare. However, it is crucial to acknowledge the limitations inherent in Twitter sentiment analysis,

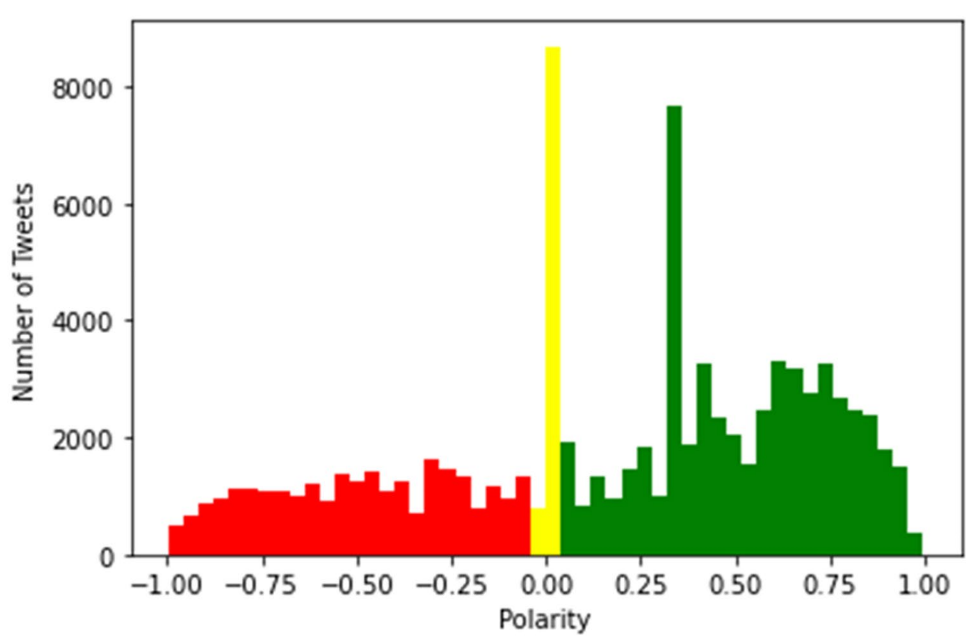


Fig. 3 Number of tweets per polarity: the figure indicates peaks at certain polarities and a left-skewedness towards positive polarity

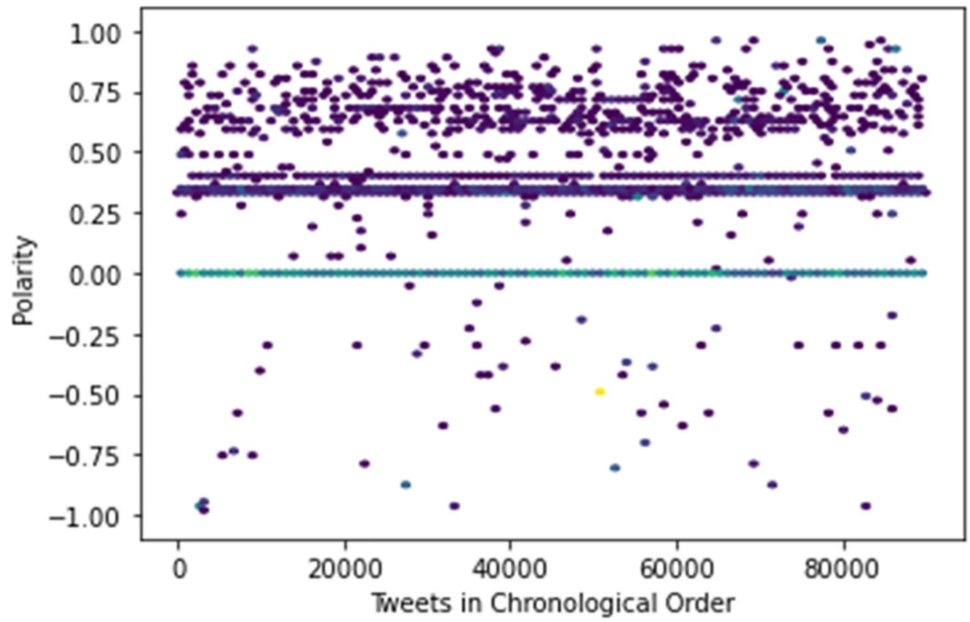


Fig. 4 Polarity of tweets plotted in chronological order: particular polarities have persisted chronologically among the tweets

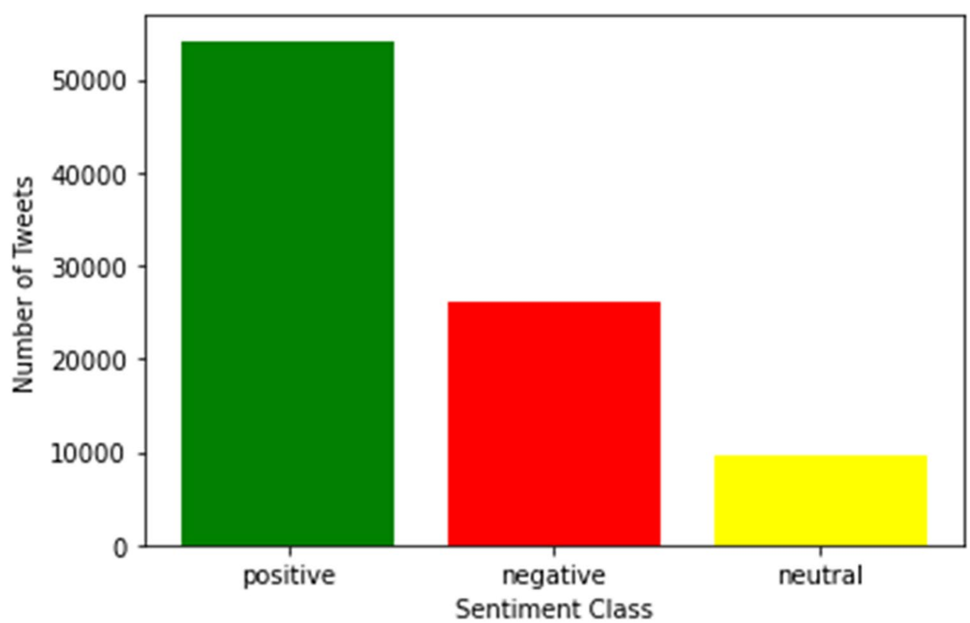


Fig. 5 Number of tweets per sentiment: sentiment is generally positive; relatively little neutral sentiment indicates polarization

(a) Positive (b) Negative (c) Neutral

Fig. 6 Sentiment class word clouds: word clouds highlight points of satisfaction (positive) and concern (negative) in regards to social security and medicare recognizing that the platform may not capture the full spectrum of public sentiment.

4.2 Named entity recognition

Figure 7 compares the prevalence of each entity type between the sentiment classes. The number of occurrences for each entity type was counted for each sentiment class and was the metric for prevalence in Fig. 7. The results were normalized by dividing them by the total number of recognized entities in the respective sentiment class.

Figure 7 clearly illustrates that some entity types had little to no prevalence in each of the sentiment classes. Other entities (such as DATE, ORDINAL, and CARDINAL) could have an inflated prevalence due to their greater tendency to appear in everyday language. The comparative analysis presented in Fig. 7 serves as a valuable tool for understanding the linguistic landscape surrounding Social Security and Medicare discussions in different sentiment contexts. Policymakers can use this information to discern which entities are consistently discussed across sentiment classes and tailor their strategies accordingly. Furthermore, recognizing the potential inflation of certain entities can inform a more nuanced interpretation of sentiment, helping policymakers avoid drawing misleading conclusions based on the overrepresentation of specific types.

Figure 8 summarizes these results by displaying the word clouds for selected entity types in each sentiment class. By selecting specific entity types for visualization, Fig. 8 facilitates a targeted examination of the entities that carry significant weight in shaping sentiment within each class. Policymakers can leverage the insights gained from Fig. 8 to tailor their strategies based on the entities consistently discussed across sentiment classes. This targeted approach enables a more nuanced understanding of the specific topics that resonate with the public, guiding policymakers in addressing concerns or reinforcing positive aspects associated with Social Security and Medicare. Additionally, the visual representation in Fig. 8 serves as a complementary tool to the comparative analysis of entity prevalence in Fig. 7, enhancing accessibility and aiding policymakers in making informed decisions grounded in the linguistic nuances of public sentiment.

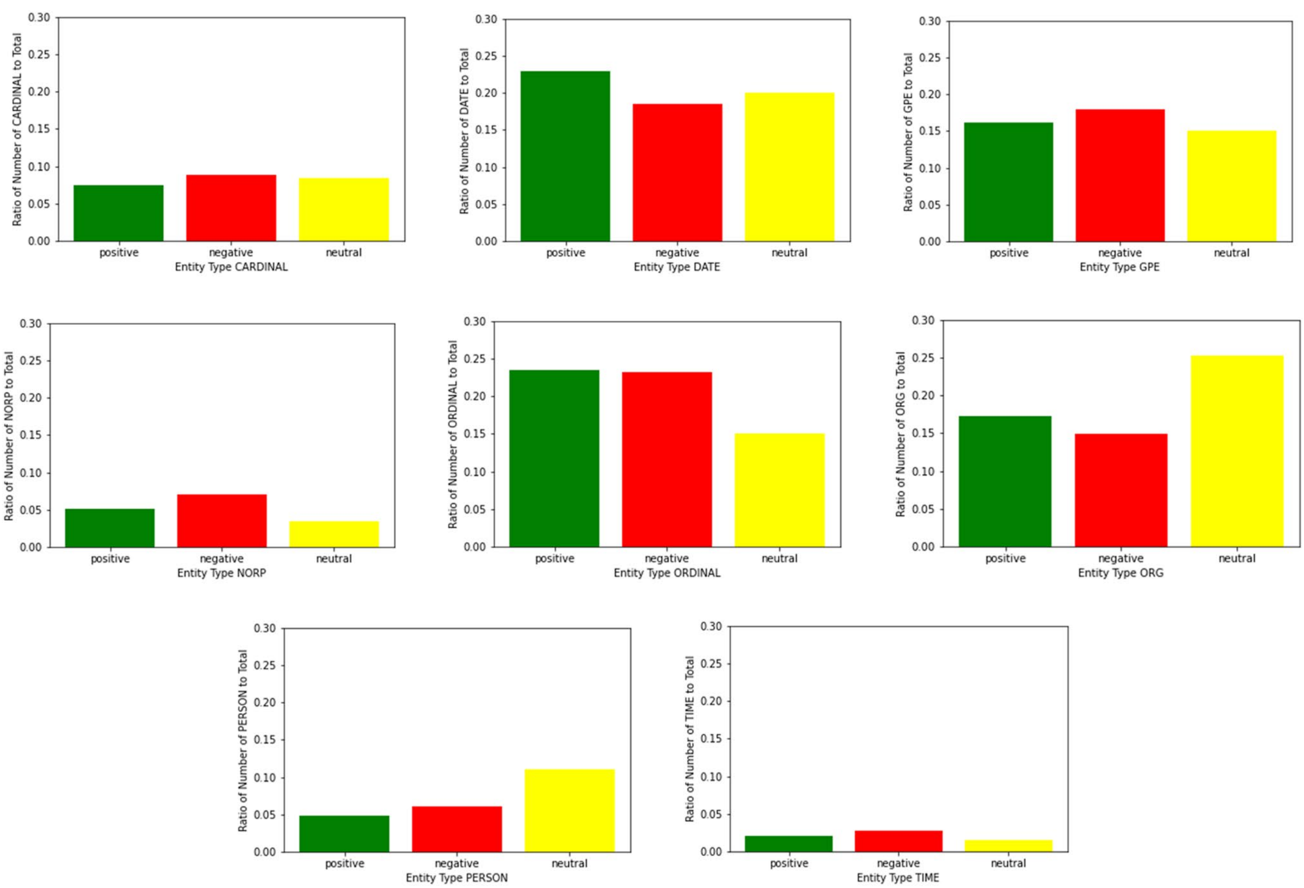
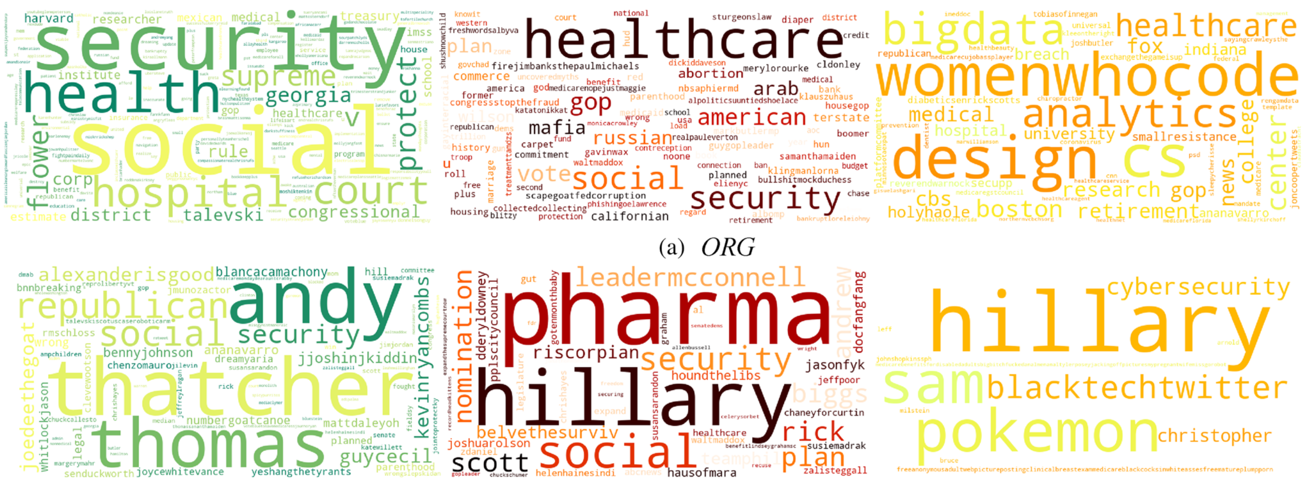


Fig. 7 Prevalence of entity types in each sentiment class: the figure displays varying degrees of prevalence across sentiment class for different entity types



(a) *ORG*

(b) *PERSON*

Fig. 8 Word clouds of named entities for selected entity types in each sentiment class: the figure highlights the entities of the densest word clouds

4.3 Topic modeling

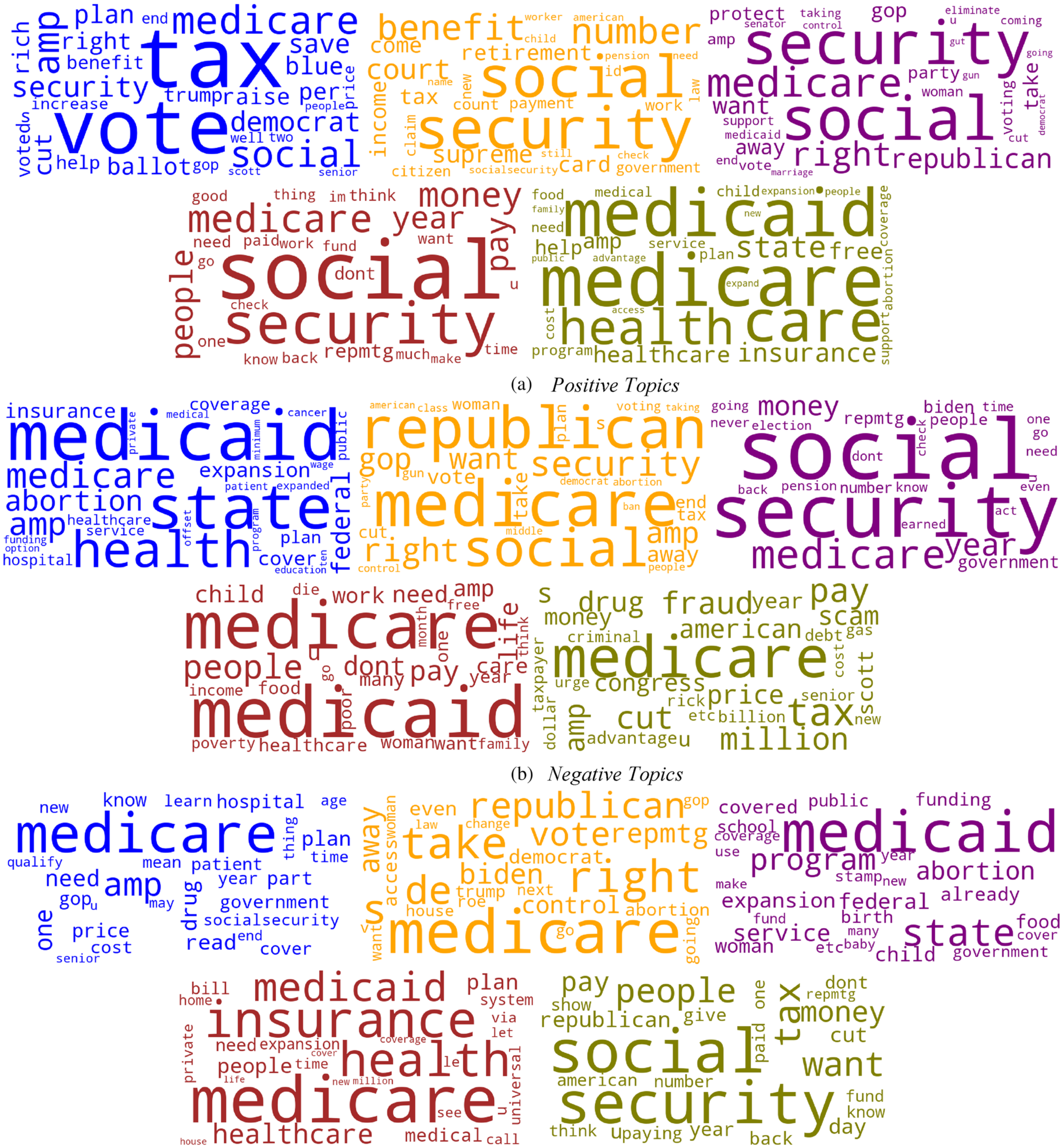
Figure 9 displays the word clouds containing the top 30 keywords of each topic in each sentiment class. These word clouds were generated without the log-likelihood ratio. Instead, a greater probability associated with the keyword in the topic corresponds to a larger size of the keyword in the word cloud.

Figure 9 notably shows across sentiment classes the separation and overlap of Social Security and Medicare as central topics. For positive topics, Topic 2 seems to distinctly focus on Social Security whereas Topic 5 seems to distinctly focus on Medicare. For negative topics, Topics 1, 4, and 5 seem to distinctly focus on Medicare. For neutral topics, Topics 2 and 4 seem to distinctly focus on Medicare, and Topic 3 seems to distinctly focus on Medicaid. Neutral Topic 1 seems to largely focus on Medicare more so than Social Security. For the topics that haven't been mentioned for each sentiment class thus far, Social Security and Medicare jointly share contributions.

Figure 9 would enable lawmakers to discern and decipher abstract topics that surface within the public discussion of Social Security and Medicare and can gain key insights into various focal points of discussion. This visual representation enables a more accessible interpretation of the interplay between sentiment classes and associated topics. Policymakers can identify key thematic areas that resonate positively or negatively with the public, informing strategies for addressing concerns or reinforcing positive aspects. Additionally, the absence of log-likelihood ratio ensures a straightforward visual hierarchy, making it easier for policymakers to prioritize and focus on keywords that hold greater importance within each topic. While there may be potential for oversimplification, presenting topic keywords as word clouds like in Fig. 9 gives policymakers the opportunity to understand and take action through the more-focused insights of topic modeling as opposed to named entity recognition and sentiment analysis.

We determined the dominant topic for each tweet and the percent contribution of the dominant topic in the tweet. The dominant topic of a tweet is the topic that has the greatest percent contribution to the tweet among the other topics. Figure 10 shows the number of tweets in each topic of each sentiment class. The topics were colored the same as they were colored in the word clouds of Fig. 9. Considering the distribution of tweets across their dominant topics, Fig. 10 reveals a large peak in concentration of positive tweets whose dominant topic is positive Topic 2. Negative Topics 3, 4, and 5 were more dominant among negative tweets than negative Topics 1 and 2. Neutral Topic 2 is noticeably less dominant among neutral tweets than the other topics.

Figure 10 serves as a comprehensive visualization of the distribution of dominant topics within each sentiment class. The number of tweets associated with each dominant topic is depicted, offering a quantitative perspective on the prevalence of specific themes across sentiments. Importantly, each tweet is counted only for its dominant topic, ensuring clarity in the representation of dominant themes within the dataset. The insights derived from Fig. 10 can inform policymakers about the dominant themes that resonate most strongly within each sentiment class, providing a foundation for targeted interventions and policy adjustments. This approach not only aids in identifying overarching topics but also in discerning specific nuances associated with varying sentiments.



(a) *Positive Topics*

(b) *Negative Topics*

(c) *Neutral Topics*

Fig. 9 Word clouds of topics in each sentiment class: keywords for each topic are sized proportionally to their contribution to the topic as returned from the LDA model

Figure 11 displays the distribution of the percent contribution of each tweet's dominant topic. Given the large volume of data, we used a hexagonal bin plot. Hexagons in Fig. 11 contain at least one data point. Tweets were plotted in chronological order according to their tweet id.

The distributions for positive and negative tweets are similar. However, an interesting result is in the distribution of neutral tweets. The volume of neutral tweets at percent contributions of around 0.2, 0.6, and 0.73 are peculiar

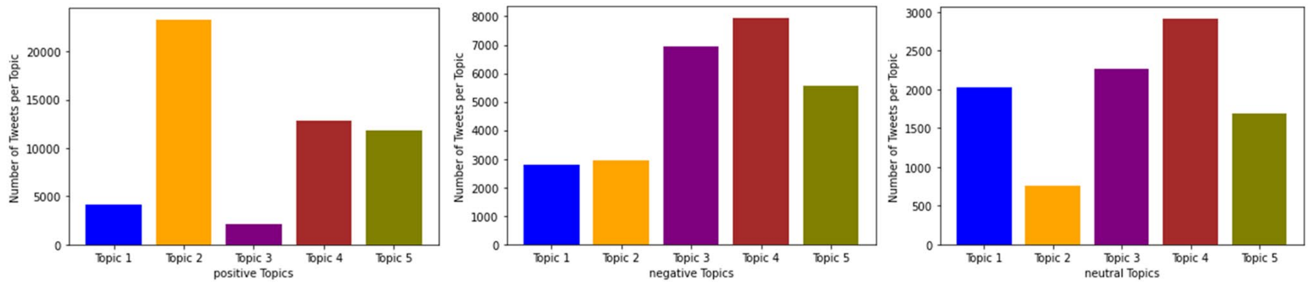


Fig. 10 Number of tweets per topic in sentiment class: different distributions of tweets across topics for each sentiment class indicates particular concentrations on certain topics or the lack thereof

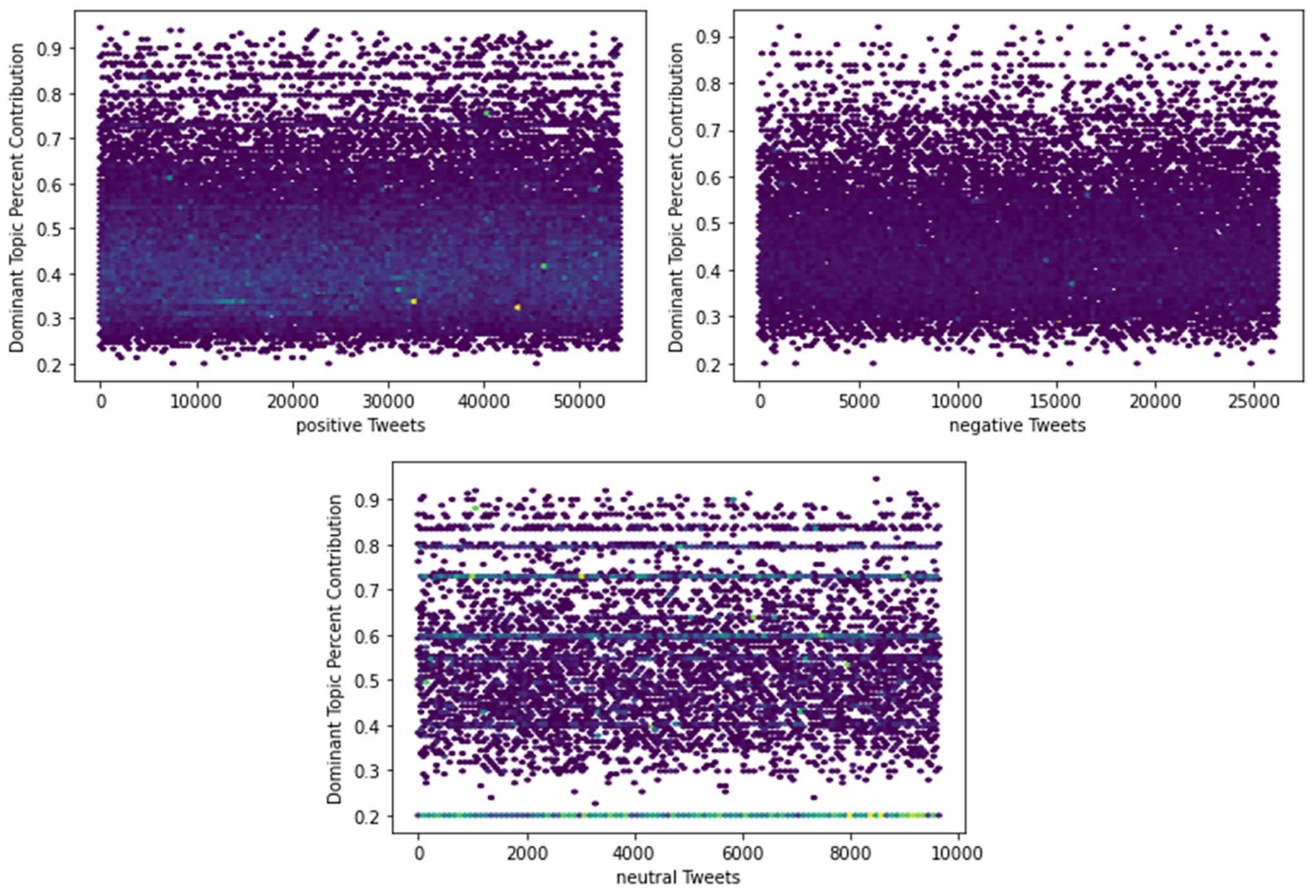


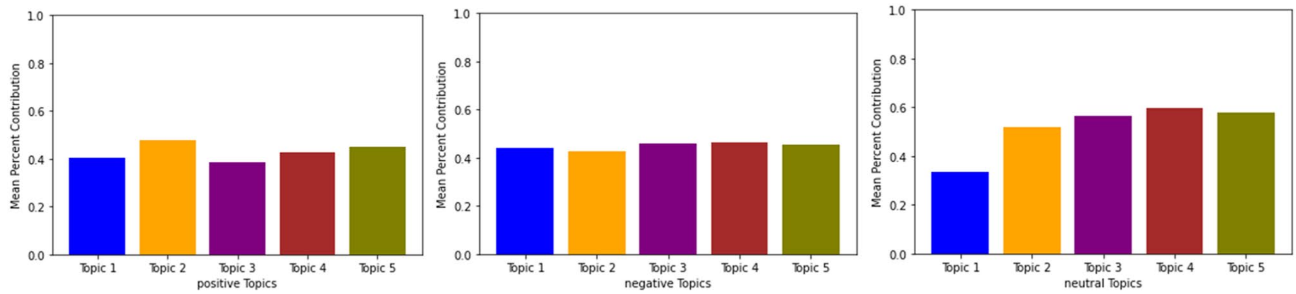
Fig. 11 Dominant topic percent contribution of each tweet: anomalies are revealed in the chronological plotting of dominant topic contribution in neutral tweets which may indicate a poor choice in the number of neutral topics for topic modeling

outliers to an otherwise similar distribution to those of positive and negative tweets.

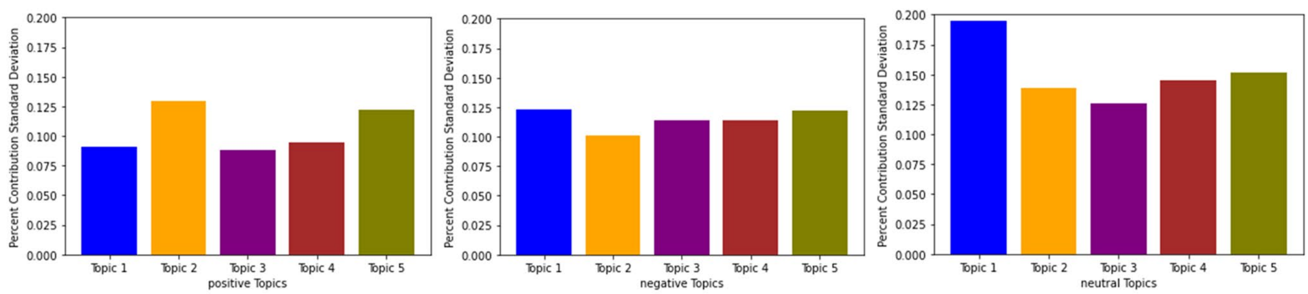
Figure 12 shows the mean and standard deviation of the percent contribution of each topic. This was done by only considering the percent contributions of dominant topics.

The mean for neutral Topic 1 is noticeably lower than the means of other neutral topics whereas the standard deviation of neutral Topic 1 is noticeably higher than the standard deviations of the other neutral topics. Coupled with the abnormalities in Fig. 11, this could allude to a poor choice in the number of neutral topics for the gensim LDA model.

Figure 13 provides the word counts of the top 10 keywords for each topic along with the associated probability of the keyword belonging to that topic returned by the gensim LDA model. The word counts shown account for all instances of the keyword in tweets of the respective sentiment class and not necessarily in tweets whose dominant topic contains that keyword. This was done with the



(a) *Mean*



(b) *Standard Deviation*

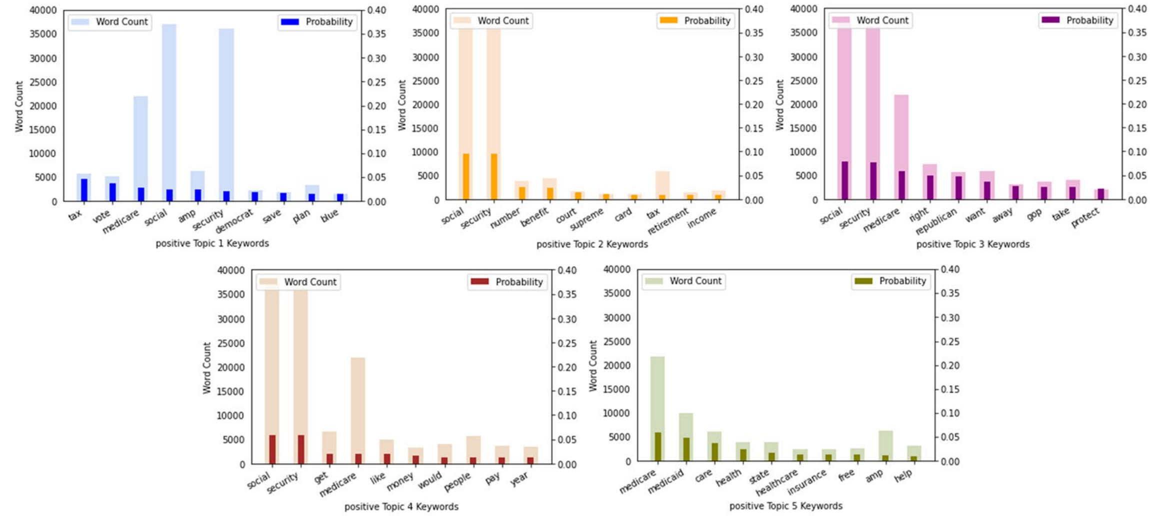
Fig. 12 Mean and standard deviation of percent contribution: differences in the distributions of mean and standard deviation in neutral tweets as opposed to positive and negative tweets further contribute to poor choice in neutral topics

purpose of not comparing the topics as a whole but rather comparing keywords within each topic separately. A key insight gained from Fig. 13 is that keywords that appear with more frequency within their topic's sentiment class tend to have a greater contribution to its topic.

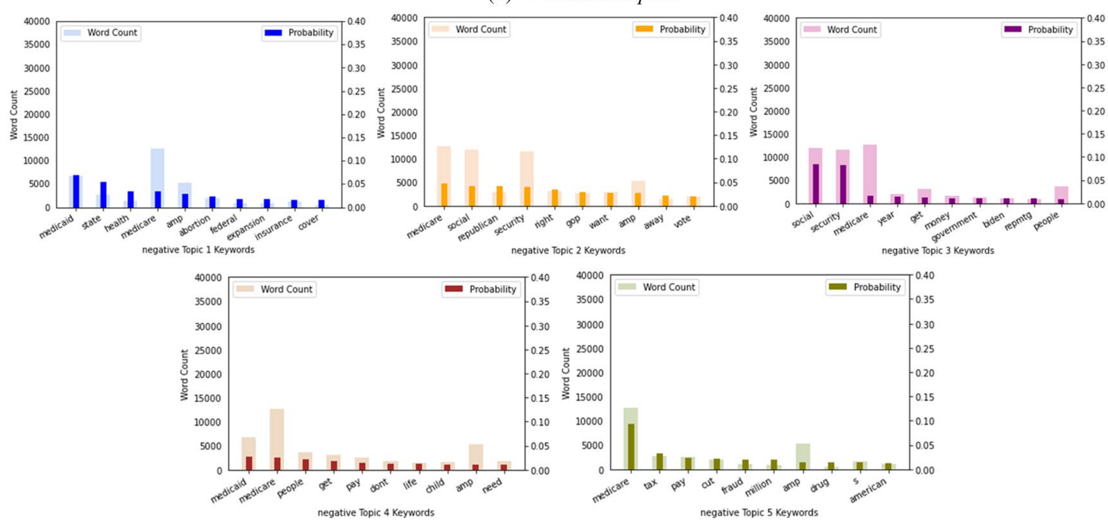
By considering all instances of a keyword across tweets within the respective sentiment class—rather than restricting the count to tweets where the dominant topic contains that keyword—Fig. 13 provides a more comprehensive perspective on the frequency and relevance of specific terms. The inclusion of topic association probabilities adds an additional layer of insight, offering a measure of confidence in the likelihood of a keyword being closely tied to a particular topic. This nuanced approach not only refines our understanding of the prominence of keywords within a sentiment class but also enables a more informed interpretation of the topics identified by the LDA model. Policymakers and researchers can leverage the information presented in Fig. 13 to discern the key terms that carry substantial weight within each topic, understanding both their frequency and the degree of association with the overarching themes. This level of detail allows for a targeted examination of language use and topic association probabilities, facilitating a more comprehensive understanding of sentiment, topics, and individual keywords in the public discussion surrounding Social Security and Medicare.

5 Conclusion

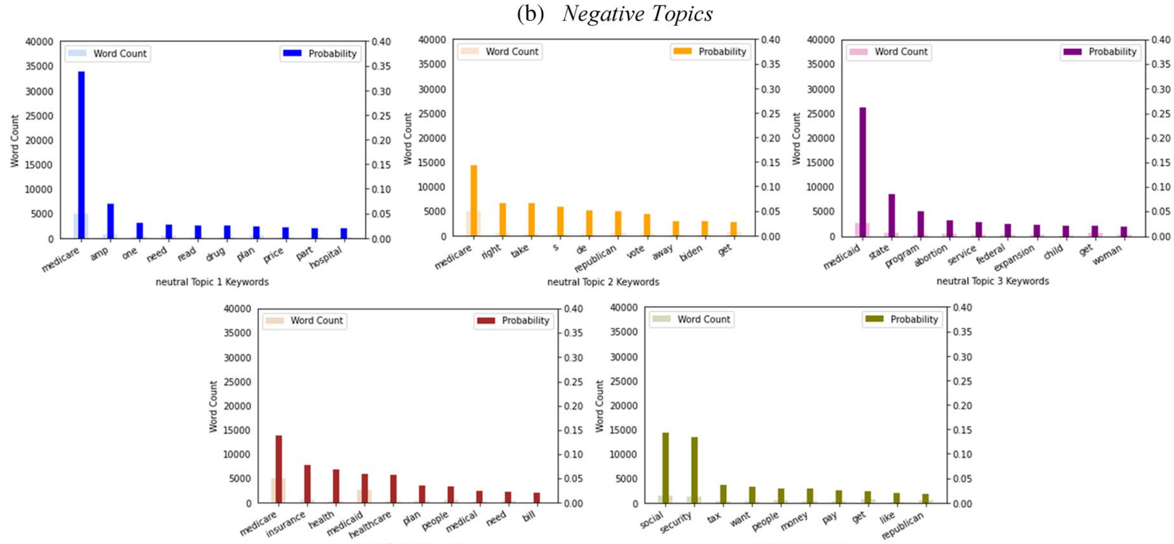
This paper analyzes public sentiment on Social Security and Medicare as a means to provide structure to the discussion of the matter on Twitter. Politically involved individuals are able to quantitatively identify objects of discussion as well as understand general public trends. Given the small number of neutral tweets (11%), the topic of Social Security and Medicare is rather controversial. Regardless, sentiment has been largely positive, with the percentage of tweets being positive (60%) being more than double that of negative (29%). People and organizations were mentioned in variety, having the most dense wordclouds out of the named entity types. Geographical entities and nationality/religious/political groups were also prevalent albeit less diverse than people and organizations. It is interesting that, despite their association with politics, legal entities and law were not discussed as much as those mentioned above. Five topics were extracted unsupervised for each sentiment class to distinguish general trends of discussion. Experimentation compared the prevalence of each topic for each sentiment class and assessed the impact of keywords. The most prevalent positive topic focuses on Social Security and includes keywords such as "benefit," "supreme court," and "retirement," suggesting that such subjects are a positive, recurring theme regarding Social Security, whereas



(a) *Positive Topics*



(b) *Negative Topics*



(c) *Neutral Topics*

Fig. 13 Topic keyword counts: compares the keyword contribution to the topic with the overall frequency of the keywords across all tweets in the sentiment class which brings to focus the individual keywords as opposed to topics/sentiment as a whole the most prevalent negative topic highlights keywords like "pay," "life," and "child" as a negative, recurring theme for Medicaid and Medicare. Meanwhile, keywords such as "insurance," "healthcare," and "plan" are a recurring theme in the most prevalent neutral topic. Experimentation also revealed a poor choice in the number of neutral topics to be determined and the experimentation could be improved through a more careful choice on the number of topics. This inadequacy points to the need for careful consideration and calibration of parameters to achieve optimal results. The inadequate choice in the number of neutral topics likely resulted in a suboptimal representation of the nuanced discourse present in the dataset which could manifest in several ways: either by categorizing disparate themes under a single neutral topic and masking important distinctions or by failing to adequately capture certain neutral topics altogether. Consequently, the insights may be incomplete or skewed and not accurately portray the full spectrum of topics across all sentiments. The experimentation could benefit from more thorough exploratory analyses to gauge the complexity and diversity of the dataset as well as considering specific nuances within the topics themselves. The classification of emotions like anger and sadness can also provide another avenue of experimentation worth exploring. The amalgamated results yield a comprehensive analysis of public sentiment and discussion on Social Security and Medicare. The experimentation can be broadly applied in politics. Politicians and lawmakers can observe the public response to actions and policies in real time. Election candidates can campaign considering named entities mentioned on social media and understanding insightful trends and inherent issues in public discussions.

Author contributions Chakravarty chose what to study, gathered relevant data, performed analysis, compiled and visualized results, and wrote the manuscript. Arifuzzaman inspired the idea behind the work and revised the manuscript.

Funding No funding was received.

Data availability The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Ainley E et al (2021) Using twitter comments to understand people's experiences of UK health care during the COVID-19 pandemic: thematic and sentiment analysis. J Med Internet Res 23(10):e31101

Bermingham A, Smeaton A (2011) On using Twitter to monitor political sentiment and predict election results. In: Proceedings of the workshop on sentiment analysis where AI meets psychology (SAAIP 2011)

Bird S et al (2009) Natural language processing with python. O'reilly, Beijing

Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. J Mach Learn Res 3:993–1022

Burling R et al (1993) Primate calls, human language, and nonverbal communication [and comments and reply]. Curr Anthropol 34(1):25–53

Deacon TW (1998) The symbolic species: the co-evolution of language and the brain. No. 202. WW Norton & Company, New York

Dunbar R. (1998) Theory of mind and the evolution of language. Approach Evol Lang 92–110

Hauser MD, Yang C, Berwick RC, Tattersal I, Ryan MJ, Watumull J, Lewontin RC (2014) The mystery of language evolution. Front Psychol 5:401

Honnibal M, Ines M (2017) spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To Appear 7(1):411–420

Hunter JD (2007) Matplotlib: a 2D graphics environment. Comput Sci Eng 9(03):90–95

Hutto C, and Eric G. (2014) Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the international AAAI conference on web and social media. Vol. 8. No. 1

Kim SY et al (2021) Public sentiment toward solar energy—opinion mining of twitter using a transformer-based language model. Sustainability 13(5):2673

Laland KN (2017) The origins of language in teaching. Psychon Bull Rev 24:225–231

Lasri I, Anouar R, Mourad E (2023) Self-attention-based bi-LSTM model for sentiment analysis on tweets about distance learning in higher education. Int J Emerg Technol Learn 18(12):119–141

Mittal A, Arpit G (2012) Stock prediction using twitter sentiment analysis. Stanford University, CS229 (2011 https://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf) 15:2352

Oesper L et al (2011) WordCloud: a Cytoscape plugin to create a visual semantic summary of networks. Source Code Biol Med 6(1):1–4

Pla F, Hurtado LF (2014) Political tendency identification in twitter using sentiment analysis techniques. In: Proceedings of COLING 2014, the 25th international conference on computational linguistics: Technical Papers

Power C (1998) 7 old wives' tales: the gossip hypothesis. In: Approaches to the Evolution of Language: Social and Cognitive Bases, Cambridge University Press, p. 111

Rasool A et al (2019) Twitter sentiment analysis: a case study for apparel brands. In: Journal of physics: conference series, vol 1176. No. 2. IOP Publishing

Rehurek R, Petr S (2011) Gensim–python framework for vector space modelling. NLP Centre, Fac Inform, Masaryk Univ, Brno, Czech Repub 3(2):2

Roesslein J (2020) Tweepy: twitter for python! URL: https://github.com/tweepy/tweepy

Sattar NS, Shaikh A (2021) COVID-19 vaccination awareness and aftermath: public sentiment analysis on Twitter data and vaccinated population prediction in the USA. Appl Sci 11(13):6128

Wang H, et al. (2012) A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In: Proceedings of the ACL 2012 system demonstrations.

Washburn SL, Lancaster GS (2017) The evolution of hunting. In: Man the hunter, Routledge, pp 293–303, England

Zhou J et al (2020) Examination of community sentiment dynamics due to covid-19 pandemic: a case study from Australia

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.