Analyzing Sentiments and Topics on Twitter Towards Rising Cost of Living

Yanyi Li, Nian Ran, Yifu Chen, Renhua Zhou and Riza Batista-Navarro

July 21, 2023

Analyzing Sentiments and Topics on Twitter towards Rising Cost of Living

Yanyi Li, Nian Ran, Yifu Chen, Renhua Zhou, Riza Batista Navarro

University of Manchester

Abstract. In September 2022, the United Kingdom experienced an unprecedented 40-year high in its inflation rate, resulting in a cost of living crisis that has significantly impacted British citizens. To assess public opinion on this issue, we developed a social media analytics pipeline to collect and analyze microblogs posted on Twitter. Our primary objective was to conduct sentiment analysis on the collected tweets to determine the dominant sentiment towards the topic of the cost of living. Additionally, we performed named entity recognition to identify the entities most frequently mentioned and used topic modeling to uncover the most discussed topics. Our approach employed a hybrid sentiment analysis method that utilized three lexicons for preliminary tweet labeling and fine-tuned a RoBERTa model. Our results demonstrate the superior effectiveness of our methods, which provided an in-depth analysis of the cost-of-living situation in the UK.

Keywords: Text Mining · Sentiment Analysis · Topic Modelling

1 Introduction

Social media analytics is a prevalent method for understanding and analyzing public opinions on products or services, as well as perceptions of social events and discussions about political or economic news[18,19,25]. The primary objective of this project is to apply a social media analytics pipeline to assess British people's perceptions of the cost of living in the United Kingdom using tweets retrieved from Twitter[26,27]. The Consumer Prices Index including owner occupiers' housing costs (CPIH) is a measure of inflation. According to the Office for National Statistics, the CPIH index rose from 0.9% in January 2021 to 8.8% in March 2023 [3], indicating that the cost of living crisis is a serious problem. To gain insight into British people's perspectives on this issue, we conducted sentiment analysis, topic modeling and named entity recognition tasks on tweets collected from Twitter and analyzed the results obtained.

2 Related Work

2.1 Sentiment Analysis

Twitter sentiment analysis represents a novel and challenging domain within the field of sentiment analysis. The length limitations and informal nature of tweets make them particularly difficult to analyze [4,7,9,24,26,28]. There are four predominant methods for conducting sentiment analysis on Twitter: lexicon-based, graph-based, machine learning-based, and hybrid methods that combine lexicon-based and machine learning-based approaches [4]. Deep learning techniques have been applied to Twitter sentiment analysis tasks with great success, demonstrating high levels of precision and effectiveness[10]. Hybrid methods that utilize learning-based approaches to help the model learn new rules and lexicons have proven to be a practical approach for labeling and analyzing sentiment[13]. In our study, we combine lexicon-based approaches and machine learning and achieve better accuracy and more meaningful and authentic results.

2.2 Topic Modeling

Topic modeling is a fundamental technique to uncover latent topics from a given document, it can also be used for fast recommendation by hashtags on Twitter[29]. The two primary types of approaches used for topic modeling are statistical-based and machine-learning-based methods. Among them, Latent Dirichlet Allocation (LDA) [5] has emerged as the most successful statistical one, assuming that a document can be represented as a distribution of topics, while a topic can be represented as a distribution of words. To better capture the complexities of modern social media such as Twitter, Author Topic Model (ATM) [6] and conversation-based LDA [8] were proposed to aggregate documents by authors and conversations, respectively. Recent advancements in pretraining word vectors have led to more granular topic modeling approaches such as Top2Vec [11] and BERTopic [12] using word vector pretraining and deep learning techniques. A comparison between these methods and LDA has been conducted in [14]. In this study, we employ LDA as our topic model due to its prominent performance and applications on Twitter data [15,16], as well as social prevalent topics such as energy companies [18] and COVID-19 [19].

2.3 Aspect-based Sentiment Analysis

Sentiment analysis can be categorized into three levels: document-level, sentence-level and aspect-level. Aspect-based sentiment analysis (ABSA) associates sentiments with specific aspects or entities within a sentence[20]. Named entity recognition can be used to extract entities from sentences and then combined with a sentiment analyzer model to conduct sentiment analysis at the aspect level[21].

3 Data collection and preprocessing

3.1 Data Collection

Twitter API for academic research was used to search for tweets from 2021 to 2023 that contains keywords or hashtags such as cost of living, food prices, energy bills, etc. We retrieved over 30,000 tweets posted in the United Kingdom by specifying geological coordinates. The raw data contained a wealth of information; replies and retweets were also collected as they contain important expressions of sentiment and perceptions. Additionally, we took into account the like count of each tweet when conducting sentiment analysis.

3.2 Pre-processing

All tasks require certain common pre-processing procedures, such as the removal of usernames, URLs and punctuation. However, different tasks have specific requirements for input data. For instance, while emojis and emoticons can contribute to polarity scores in sentiment analysis, they are irrelevant to topic modeling and named entity recognition (NER) tasks. Additionally, topic modeling is case-sensitive whereas sentiment analysis is not. As a result, different pre-processing steps were conducted for each task. The *clean-text* package [1] was used to filter out emojis, usernames, hashtags and web links from tweets for fundamental data cleaning. These elements have certain patterns; hence we used regular expressions to match the corresponding formats and remove them.

4 Methodology

4.1 Sentiment Analysis

Due to the fact that tweets collected from Twitter are naturally unlabeled and the volume of data is vast, manually labeling all tweets is impractical. As such, supervised learning is unlikely to be feasible for this task. However, a hybrid method represents a suitable approach. In this project, we initially used *Valence Aware* Dictionary and Sentiment Reasoner (*VADER*), a lexicon and rule-based sentiment analysis tool that is particularly accurate and effective for analyzing sentiment expressed in social media. However, *VADER* has certain limitations; it calculates valence scores for each word in a sentence and then adjusts the score according to predefined rules. This means that it cannot fully comprehend the context and semantics of sentences [17]. As a result, sarcasm and irony may be misinterpreted and corresponding tweets may be labeled with the opposite sentiment.

Traditional hybrid methods as a solution to above issue use only one sentiment lexicon to annotate data before employing machine learning techniques to increase precision. In contrast, we employed an ensemble of three different sentiment lexicons - *VADER* [2], *TextBlob* [3] and *SentiWordNet*[4] - to improve annotation accuracy and provide a solid training dataset for our machine learning model.

[1] https://github.com/prasanthg3/cleantext
[2] https://github.com/cjhutto/vaderSentiment
[3] https://github.com/sloria/TextBlob
[4] https://github.com/aesuli/SentiWordNet

Tweets that received consensus from a minimum of two lexicons were assigned a corresponding sentiment label and utilized as training data. On the other hand, tweets that failed to achieve consensus across the lexicons were employed as test data without any sentiment annotations. This approach effectively capitalizes on the individual strengths of each lexicon to establish a robust consensus regarding the sentiment expressed within each tweet.

However, setting appropriate thresholds for the sentiment score of each lexicon is a crucial step in improving accuracy. The commonly used threshold for VADER in classifying a sentence is >0.05 as 'positive' and <-0.05 as 'negatives'[1]. However, this threshold was not accurate or suitable for our data and resulted in numerous misclassification and counterintuitive sentiments. Similarly, the commonly used threshold for TextBlob and SentiWordNet to separate 'positive' and 'negative' sentiments were not accepted by our dataset due to low F1-score.

To address this particular concern, we randomly selected a subset of 500 data samples from the entire dataset. Each individual data sample was then annotated by two members of our team. In order to enhance the sentiment interpretation of each data sample, we introduced an additional "Neutral" tag, acknowledging that certain tweets may convey factual information or serve as mere reports. Consequently, for each lexicon, there are two thresholds that need to be determined and set. To ensure the reliability of our annotated data, we utilized the Cohen's kappa coefficient [30], treating the annotations provided by the first annotator as the "Ground truth" and those provided by the second annotator as the "Reference". This allowed us to calculate metrics such as "True positives", "True negatives", "False negatives", and "False positives", ultimately yielding the Cohen's kappa coefficient. Our objective was to maximize the Cohen's kappa coefficient, aiming for a value close to 1 since a value of 1 indicates perfect agreement between the annotators. Additionally, we implemented a validation process within our team, involving an independent auditor who scrutinized our labeled tags whenever the Cohen's kappa coefficient fell below a predetermined threshold.

After obtaining the 500 data samples as our gold standard, we proceeded to adjust the threshold values for each lexicon until the predicted labels achieved a relatively high F1-score when compared to the gold standard. To determine these optimal thresholds, we conducted a random search[31] within a range of -1 to 1. Specifically, we initialized 5000 samples within this range for each lexicon and iteratively evaluated our objective function, the F1-score, to identify the highest achievable value. Subsequent to the random search process, we obtained the best thresholds and their corresponding F1-scores for each lexicon, which are presented in table 1.

Subsequently, we proceeded to apply the thresholds associated with the highest F1-scores for each lexicon to the complete dataset, thereby generating preliminary predictions using the three lexicons. Data points that achieved consensus among two or more lexicons were selected to form the training set. As a result, our training set exhibited a sentiment distribution comprising 14,833 instances

Lexicons	Threshold1	Threshold2	F1-score
VADER	0.809	-0.102	0.588
TextBlob	0.262	-0.126	0.408
SentiWordNet	0.122	-0.121	0.338

Table 1. Thresholds to reach highest F1-scores and highest F1-scores

analyzed as neutral, 9,622 instances analyzed as negative, and 2,129 instances analyzed as positive.

Finally, we used **RoBERTa** [5] [2], a more robust version of **BERT** that is pre-trained on a larger corpus of English data, as our machine learning model. We fine-tuned the pre-trained **RoBERTa** model on our training dataset to learn latent patterns in the tweets and make predictions about the sentiment of tweets in our test dataset.

The methodology employed for conducting sentiment analysis is delineated in Figure 1.

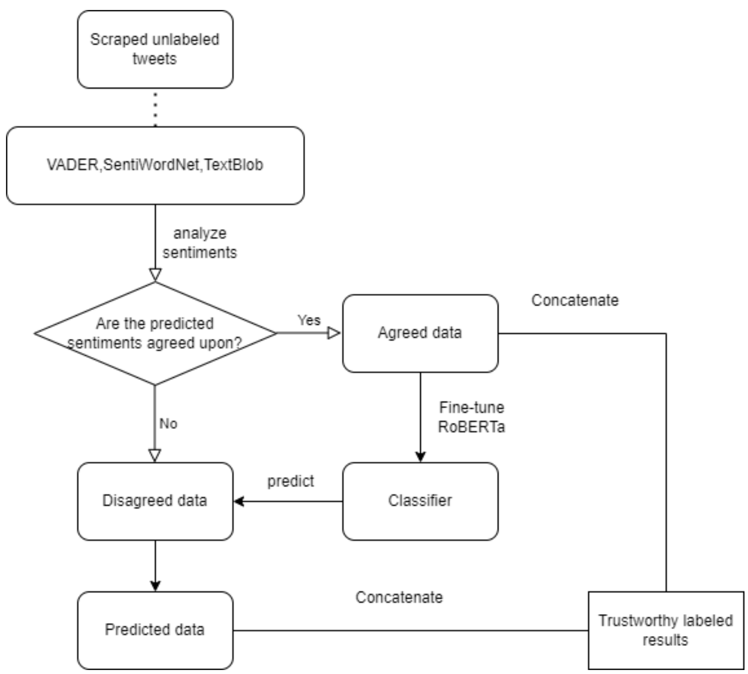

Fig. 1. Sentiment Analysis Workflow

[5] https://huggingface.co/RoBERTa-base

4.2 Topic Modeling

Effective data preprocessing is critical for accurate topic modeling. In addition to the preprocessing steps mentioned in 3.2 Pre-processing, we follow the suggested order in [22] and make some empirical modifications. After data cleaning, we first apply lemmatization, followed by the removal of stopwords. This approach is necessary because some words can become stopwords after lemmatization. For example, 'where's' becomes 'where' and 's.' After preprocessing, bi-grams are constructed to provide more context for topic analysis.

To obtain the best performance from the LDA model, a Bayesian Optimization technique is employed to search for optimal hyperparameters, including the number of topics, training passes, and the parameters of two distributions in LDA. The search is guided by the topic model coherence score, which uses both normalized pointwise mutual information (NPMI)[23] and cosine similarity.

4.3 Aspect-based Sentiment Analysis

This study employs the Stanford NER Tagger, a 7-class model, to extract named entities including location, person, organization, money, percent, date, and time. Tagged data is saved for future use due to the time-consuming tagging process. We retain uppercase tags and word features because our tagger is case-sensitive. To remove irrelevant words, those with the 'O' label are removed. The top 20 most frequent tokens are then selected for further named entity sentiment analysis.

5 Evaluation and results

5.1 Sentiment Analysis

For our evaluation, we employed the hold-out method. Specifically, we partitioned a portion of the consensus data into a 70% training set and a 30% evaluation set. After fine-tuning critical hyperparameters[6] such as learning rate, batch size, and epochs, our model yielded compelling results with 0.75 macro F1 score confronting an unbalanced dataset. We use "WeightedRandomSampler" [7] from PyTorch to address unbalanced dataset problem, it assigns weights to each sample in the dataset based on the class unbalance, allowing for more frequent sampling of minority class examples and less frequent sampling of majority class examples.

Sentiment details are elaborated upon Figure 6 and Table 4 in the appendix. We found that after conduting our mothod, there's a promising improvment compared with initial single lexicon shown on table 2.

[6] Learning rate:5e-5, batch size:64, epochs:10
[7] https://pytorch.org/docs/stable/data.html

Lexicons	Threshold1	Threshold2	F1-score	Improved_F1	Improve rate (%)
VADER	0.809	-0.102	0.588	0.650	10.544
TextBlob	0.262	-0.126	0.408	0.650	54.762
SentiWordNet	-0.122	-0.121	0.338	0.650	95.783

Table 2. Improvement rate for each lexicon

We later employed the fine-tuned model to predict the sentiment of the discordant data. Finally, we concatenated the predicted data with the original consensus data to generate our conclusive results. To discern the dominant sentiments, we took into consideration the count of likes, as they convey the support and endorsement of the original tweets and consequently express the same sentiment. Therefore, the frequency of a sentiment is the sum of all likes associated with that sentiment. (We omitted retweets and replies as they may express disparate sentiments from those of the original tweets.) In Figure 2 and Figure 3, we present the sentiment frequency and monthly sentiment frequency, respectively, as well as statistics without counting likes in Figures 7 and 8 in the appendix. We also provide WordClouds for both positive and negative sentiments in Figure 9, 10 in the appendix for a more clear view of these sentiments

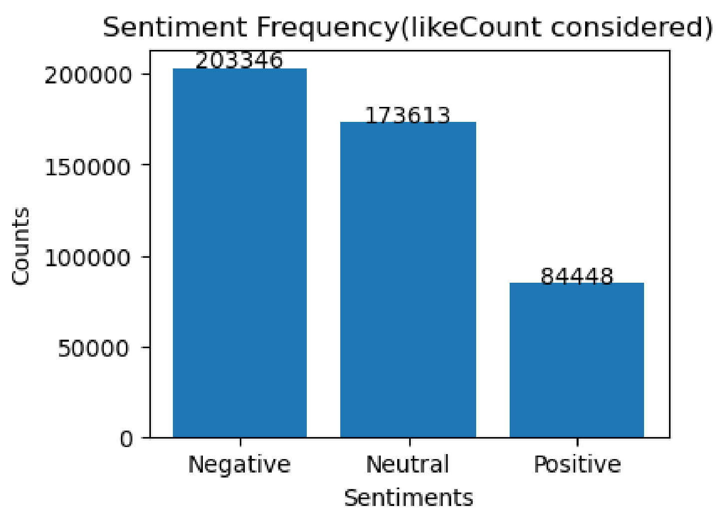

Fig. 2. Sentiment Frequency

The figures presented in the analysis reveal a clear predominance of negative sentiment towards the rising cost of living, exceeding positive sentiment by over twofold, combined with the number of likes. Notably, this negative sentiment exhibits a consistent upward trend from January 2021 to August 2022,

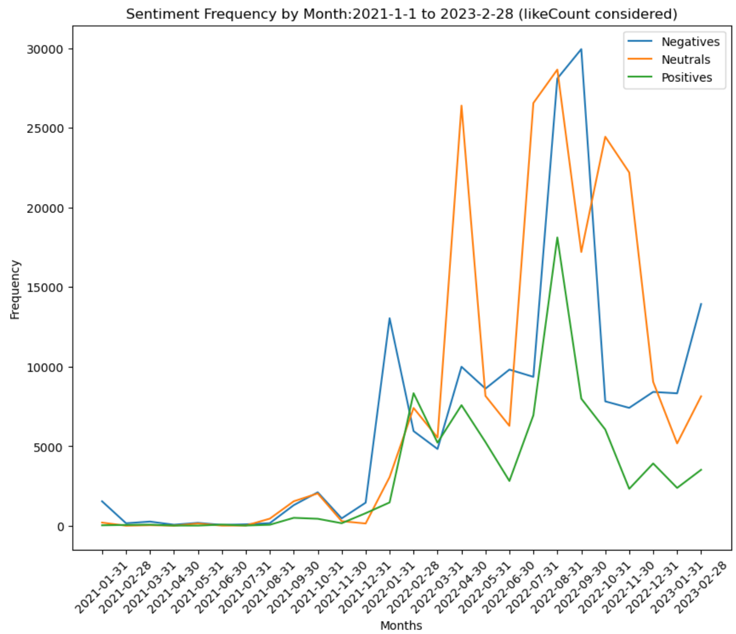

Fig. 3. Sentiment Frequency monthly

with a surge observed between June and August 2022, coinciding with the surge in inflation rates. However, a decline in negative sentiment was observed until November 2022, which may be attributed to the government's *Energy Bills Support Scheme(EBSS)*[8] that was introduced from October 2022 to March 2023. Nevertheless, the negative sentiment still remained dominant over positive sentiment, and there is an observed recent trend of a resurgence due to the yet unresolved issue of high living costs.

Additionally, during the timeframe spanning from February 2022 to December 2022, an intriguing phenomenon emerged, characterized by a significant fluctuation in the number of neutral tweets, ranging from approximately 6,000 to 27,500. This period witnessed an increased inclination among individuals to share factual information, occasionally accompanied by a heightened expression of sentiments related to the cost of living. Additionally, as the timeline progressed, a substantial number of tweets centered around reporting on cost-of-living policies, thereby contributing to a higher prevalence of neutral sentiments within the dataset.

5.2 Topic Modeling

We extracted the top 7 topics with the highest occurrence in our dataset, as shown in Figure 4. The remaining topics have infrequent occurrences, so it's

[8] https://www.gov.uk/get-help-energy-bills

trivial to list them. Each topic is represented by a set of the most significant keywords. Topics 6, 23, 7, 20, and 3 indicate public concern about the rising cost of living, including the prices of energy, heat, and food. People are struggling to pay their bills, and they expect relevant policies from the government and energy companies to address this energy crisis. To provide a more granular visualization, we created WordClouds for each topic, some of which are displayed in Figure 5, while the full image is available in Figure 12 in the Appendix. The WordClouds further reveal that people are seeking support from the government and energy companies to alleviate their problems. The WordCloud in Figure 12 also indicates that people are concerned about the cold, inflation, food demand, and resources for kids.

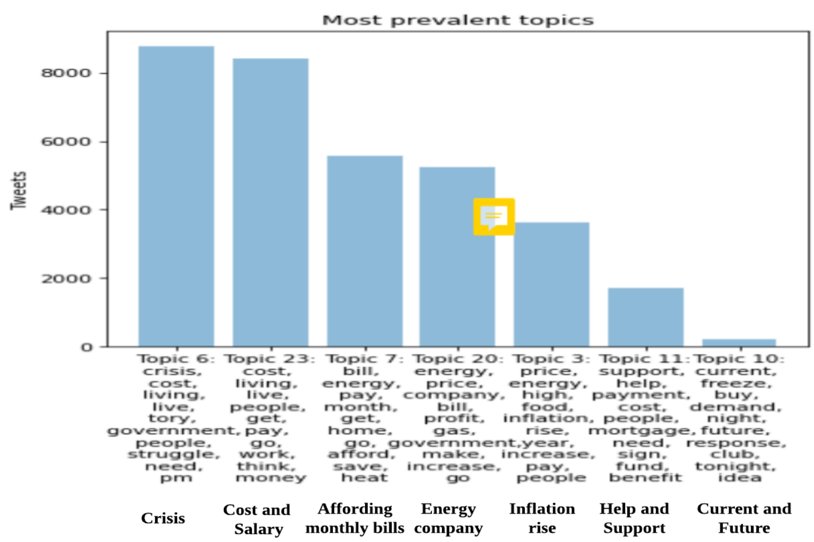

Fig. 4. Most Prevalent topics from LDA, the top 7 topics are extracted, each of the them is represented by a word that best summarizes the topic, which is in black bold.

5.3 Aspect-based Sentiment Analysis

To conduct a thorough sentiment analysis of the most-discussed entities, we extracted the 20 entities with the highest occurrence. The results show that organizations and locations such as UK, NHS, Government, Scotland, London, Johnson, Britain, Labour, and Rishi are the most mentioned. This suggests a correspondence between our sentiment analysis and topic modeling with regard to government policies and related organizations. On the other hand, Ukraine,

Fig. 5. WordCloud of the topics most talked about(partial)

EU, Europe, Energy, winter, and Russia are all related to international politics, which is a significant factor in the current energy crisis. October and April are mentioned because of the Energy Bills Support Scheme (EBSS) and the rise of bills starting from October 2022, as reported by the London government[9], and the energy price cap which will be carried out in April 2023[10]. The appearance of BBC as a frequently mentioned entity indicates that it is the main source of news for our topics. Moreover, we provide the dominant sentiment towards these entities in Table 3 according to the most frequent entities in Table 5. It presents an interesting discovery: despite the proliferation of negative sentiments prompted by the escalating cost of living, there remains a noteworthy degree of positive sentiment directed towards governmental institutions.

Entity	Sentiment
NHS	Positive
Ukraine	Positive
UK	Positive
EU	Positive
April	Positive
Government	Positive
Russia	Positive

Table 3. The dominant sentiment towards mostly mentioned entities

6 Conclusion

In this paper, we propose a novel approach for sentiment classification on Tweets. Due to the inherent complexity of the language used in tweets, we employ three lexicon-based approaches for preliminary sentiment labeling, which are

[9] https://www.london.gov.uk/city-hall-blog/rising-energy-prices-latest-advice

[10] https://www.ofgem.gov.uk/news-and-views/blog/what-april-2023-price-cap-means-consumers

then refined using the state-of-the-art RoBERTa language model. Our model achieves a macro 0.94 F1 score, indicating high accuracy in sentiment classification. Through our analysis, we have uncovered significant insights into the sentiments expressed on Twitter and its trend and causes, as well as a range of topics, including energy bills, and government policies and support. Specifically, we found that users frequently express negative sentiments towards rising living costs, citing underlying causes such as Russia and Ukraine. We also observed that people seek support from the government, energy companies, and related organizations, and that government policies aimed at alleviating energy bills have had a positive impact. In addition, our topic modeling and named entity recognition have enabled us to draw further connections between different aspects of the data.

References

1. A. A. Collomb and Lionel Brunie and Crina Costea.: A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation (2013)
2. Liu, Yinhan and Ott, Myle and Goyal, Naman and Du, Jingfei and Joshi, Mandar and Chen, Danqi and Levy, Omer and Lewis, Mike and Zettlemoyer, Luke and Stoyanov, Veselin.: Roberta: A robustly optimized bert pretraining approach. In: arXiv preprint arXiv:1907.11692 (2019)
3. Office for National Statics.: CPIH ANNUAL RATE. https://www.ons.gov.uk/economy/inflationandpriceindices/timeseries/l55o/mm23. Last accessed 2 Mar 2023
4. Giachanou, Anastasia and Crestani, Fabio.: Like It or Not: A Survey of Twitter Sentiment Analysis Methods. In: Association for Computing Machinery. New York, USA. vol. 49, num. 2. https://doi.org/10.1145/2938640. (2017)
5. Blei, David M. and Ng, Andrew Y. and Jordan, Michael I.: Latent Dirichlet Allocation. In: JMLR.org. vol. 3. (2003)
6. Steyvers, Mark and Smyth, Padhraic and Rosen-Zvi, Michal and Griffiths, Thomas.: Probabilistic Author-Topic Models for Information Discovery. In: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, USA. pp. 306-315. https://doi.org/10.1145/1014052.1014087. (2004)
7. Zimbra, David and Abbasi, Ahmed and Zeng, Daniel and Chen, Hsinchun.: The State-of-the-Art in Twitter Sentiment Analysis: A Review and Benchmark Evaluation. In: ACM Trans. Manage. Inf. Syst. vol. 9. num. 2. https://doi.org/10.1145/3185045. (2018)
8. Alvarez-Melis, David and Saveski, Martin.: Topic Modeling in Twitter: Aggregating Tweets by Conversations. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 10. num. 1. https://ojs.aaai.org/index.php/ICWSM/article/view/14817. (2021)
9. Kouloumpis, Efthymios and Wilson, Theresa and Moore, Johanna.: Twitter Sentiment Analysis: The Good the Bad and the OMG! In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 5. https://ojs.aaai.org/index.php/ICWSM/article/view/14185, DOI=10.1609/icwsm.v5i1.14185. (2021)
10. Severyn, Aliaksei and Moschitti, Alessandro.: Twitter Sentiment Analysis with Deep Convolutional Neural Networks. In: Proceedings of the 38th International

ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 959-962. https://doi.org/10.1145/2766462.2767830. (2015)

11. Angelov, Dimo.: Top2Vec: Distributed Representations of Topics. https://arxiv.org/abs/2008.09470. (2020)
12. Grootendorst, Maarten.: BERTopic: Neural topic modeling with a class-based TF-IDF procedure. https://arxiv.org/abs/2203.05794. (2022)
13. Zhang, Lei and Ghosh, Riddhiman and Dekhil, Mohamed and Hsu, Meichun and Liu, Bing.: Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis. (2011)
14. Egger, Roman and Yu, Joanne.: A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. In: Frontiers in Sociology. vol. 7. https://www.frontiersin.org/articles/10.3389/fsoc.2022.886498. (2022)
15. Zhao, Wayne Xin and Jiang, Jing and Weng, Jianshu and He, Jing and Lim, Ee-Peng and Yan, Hongfei and Li, Xiaoming", editor="Clough, Paul and Foley, Colum and Gurrin, Cathal and Jones, Gareth J. F. and Kraaij, Wessel and Lee, Hyowon and Mudoch, Vanessa.: Comparing Twitter and Traditional Media Using Topic Models. Springer Berlin Heidelberg. pp. 388-349. (2011)
16. Hayder M Alash and Ghaidaa A Al-Sultany.: Improve topic modeling algorithms based on Twitter hashtags. In: Journal of Physics: Conference Series. vol.1660, num. 1, pp. 012100. https://dx.doi.org/10.1088/1742-6596/1660/1/012100. (2020)
17. Hutto, C. and Gilbert, Eric.: 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, num. 1, pp. 216-225. https://ojs.aaai.org/index.php/ICWSM/article/view/14550. (2014)
18. Ikoro, Victoria and Sharmina, Maria and Malik, Khaleel and Batista-Navarro, Riza.: Analyzing Sentiments Expressed on Twitter by UK Energy Company Consumers. In: 2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS). pp. 95-98. https://doi.org/10.1109/SNAMS.2018.8554619. (2018)
19. Murimo Bethel Mutanga and Abdultaofeek Abayomi.: Tweeting on COVID-19 pandemic in South Africa: LDA-based topic modelling approach. In: African Journal of Science, Technology, Innovation and Development. Routledge. vol. 14, pp. 163-172. https://doi.org/10.1080/20421338.2020.1817262. (2022)
20. Nazir, Ambreen and Rao, Yuan and Wu, Lianwei and Sun, Ling.: Issues and Challenges of Aspect-based Sentiment Analysis: A Comprehensive Survey. In: IEEE Transactions on Affective Computing. vol. 13, pp. 845-863. https://doi.org/10.1109/TAFFC.2020.2970399. (2022)
21. A, Vasanthi and B, Harish Kumar and R, Karanraj.: AN RL APPROACH FOR ABSA USING TRANSFORMERS. In: 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI). pp. 354-361. https://doi.org/10.1109/ICOEI53556.2022.9776915. (2022)
22. Daniel Maier and A. Waldherr and P. Miltner and G. Wiedemann and A. Niekler and A. Keinert and B. Pfetsch and G. Heyer and U. Reber and T. Häussler and H. Schmid-Petri and S. Adam.: Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. In: Communication Methods and Measures. vol. 12, pp. 93-118. https://doi.org/10.1080/19312458.2018.1430754. (2018)
23. Bouma, Gerlof.: Normalized (Pointwise) Mutual Information in Collocation Extraction. In: Proceedings of the Biennial GSCL Conference 2009. (2009)

24. Qi, Yuxing and Shabrina, Zahratu.: Sentiment analysis using Twitter data: a comparative application of lexicon- and machine-learning-based approach. In: Social Network Analysis and Mining. vol.13, pp. 31. https://doi.org/10.1007/s13278-023-01030-x. (2023)
25. Kaur, Pardeep and Edalati, Maryam.: Sentiment analysis on electricity twitter posts. https://arxiv.org/abs/2206.05042. (2022)
26. Wang, Yili and Guo, Jiaxuan and Yuan, Chengsheng and Li, Baozhu.: Sentiment Analysis of Twitter Data. In: Applied Sciences. vol. 12. https://www.mdpi.com/2076-3417/12/22/11775. (2022)
27. Rodrigues, Anisha and Fernandes, Roshan and A, Aakash and B, Abhishek and Shetty, Adarsh and K, Atul and Lakshman, Kuruva and Rajasaheb, Mahammad Shafi.: Real-Time Twitter Spam Detection and Sentiment Analysis using Machine Learning and Deep Learning Techniques. In: Computational Intelligence and Neuroscience. vol. 2022, pp. 1-14. https://doi.org/10.1155/2022/5211949. (2022)
28. Hassan Saif and Yulan He and Miriam Fernandez and Harith Alani.: Contextual semantics for sentiment analysis of Twitter. In: Information Processing & Management. vol. 52, pp. 5-19. https://www.sciencedirect.com/science/article/pii/S0306457315000242. (2016)
29. Godin, Fréderic and Slavkovikj, Viktor and De Neve, Wesley and Schrauwen, Benjamin and Van de Walle, Rik.: Using Topic Models for Twitter Hashtag Recommendation. In: Proceedings of the 22nd International Conference on World Wide Web. pp. 593-596. https://doi.org/10.1145/2487788.2488002. (2013)
30. Kraemer, Helena C.:Kappa coefficient. In: Wiley StatsRef: statistics reference online. pp.1-4. https://doi.org/10.1002/9781118445112.stat00365.pub2.(2014)
31. Bergstra, James and Bengio, Yoshua.:Random search for hyper-parameter optimization. In: Journal of machine learning research. vol.13, pp. 2.
meme

7 Appendix

	precision	recall	F1-score	support
0	0.74	0.86	0.79	2348
1	0.85	0.74	0.79	4457
2	0.67	0.78	0.72	1370
accuracy			0.78	8175
macro avg	0.75	0.79	0.77	8175
weighted avg	0.79	0.78	0.78	8175

Table 4. Classification Report for Fine-tuning RoBERTa the Consensus data (0-negative,1-neutral,2-positive)

	token	label	count
0	NHS	ORGANIZATION	760
1	Ukraine	LOCATION	607
2	UK	ORGANIZATION	424
3	EU	ORGANIZATION	409
4	Government	ORGANIZATION	402
5	Europe	LOCATION	374
6	April	DATE	349
7	Energy	ORGANIZATION	346
8	BBC	ORGANIZATION	346
9	winter	DATE	341
10	Scotland	LOCATION	321
11	October	DATE	321
12	of	ORGANIZATION	316
13	London	LOCATION	292
14	Johnson	PERSON	288
15	Britain	LOCATION	247
16	Labour	ORGANIZATION	221
17	Russia	LOCATION	220
18	Rishi	PERSON	210
19	today	DATE	209

Table 5. Top 20 most frequent words in cost of living topic

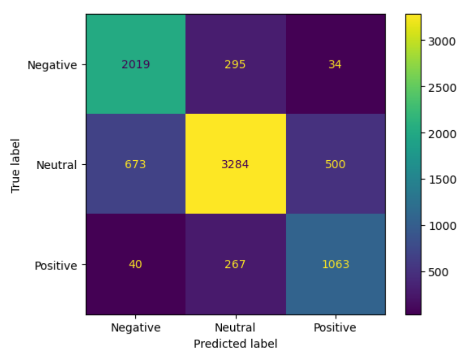

Fig. 6. Confusion Matrix after Fine-tuning RoBERTa on Consensus data

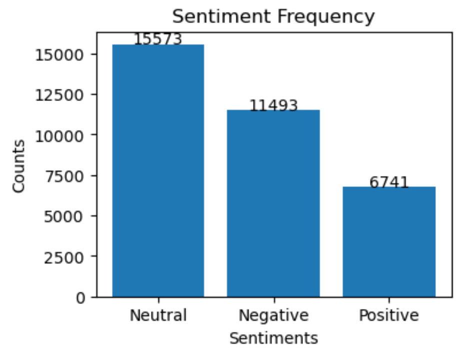

Fig. 7. Sentiment Frequency without Counting Likes

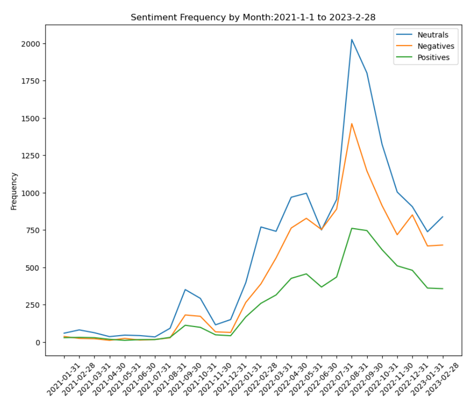

Fig. 8. Sentiment Frequency Monthly without Counting Likes

Fig. 9. WordCloud for positive sentiment

Fig. 10. WordCloud for negative sentiment

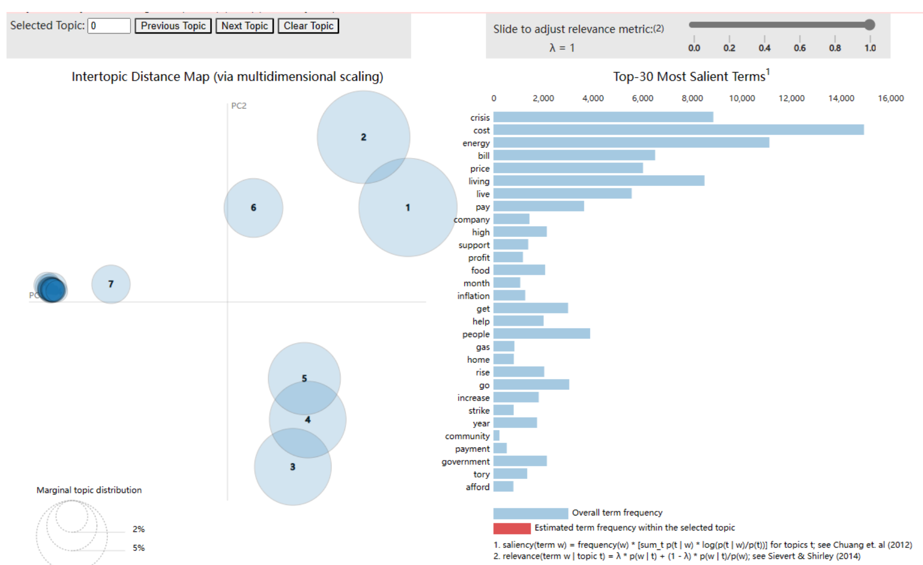

Fig. 11. visualization of the topics from LDA with rankings of the words most talked about

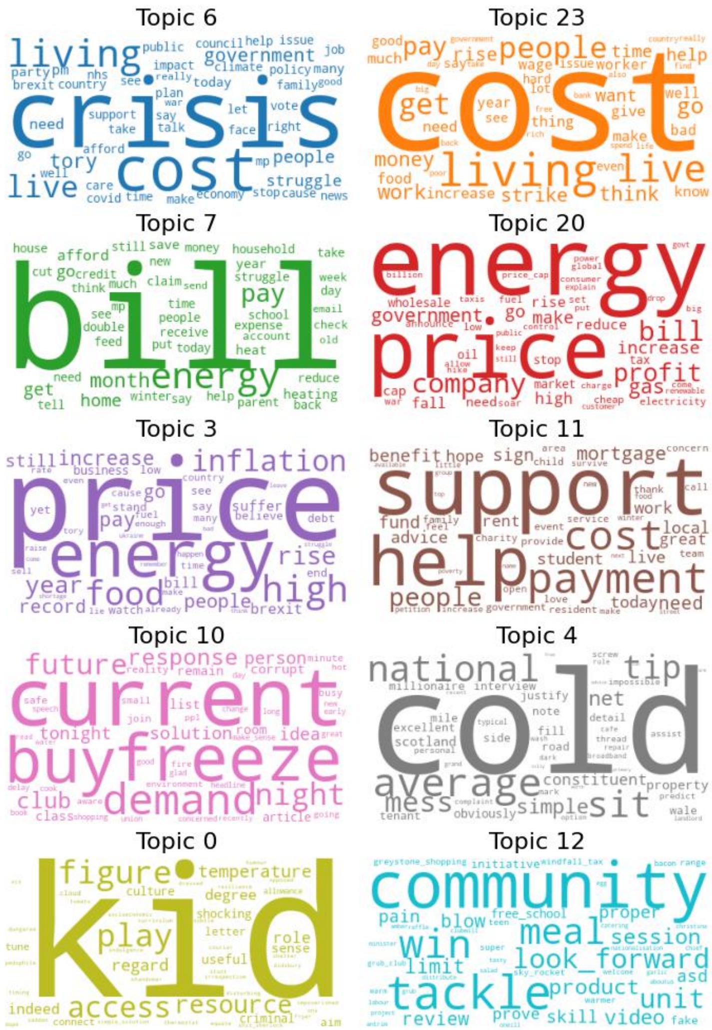

Fig. 12. WordCloud of the topics most talked about