ROMANIAN JOURNAL OF INFORMATION SCIENCE AND TECHNOLOGY
Volume **26**, Number 3–4, 2023, 312–322

Twitter's Mirroring of the 2022 Energy Crisis: What It Teaches Decision-Makers – A Preliminary Study

Horia-Nicolai L. TEODORESCU[1, *] and Mironela PIRNAU[2]

[1]Institute of Computer Science, Romanian Academy
[2]Titu Maiorescu University, Bucharest, Romania
E-mail: hteodor@etti.tuiasi.ro*
* Corresponding author

Abstract. The paper pertains to the field of opinion mining on social networks in view of decision-making. It was designed to determine the effect of the energy crises on the socio-ethical and ethical worries and concerns related to the role of *Information and Communication Technology* (ICT) in the crisis, as reflected on Twitter. Knowing public opinions during crises is a key component of decision-making. Other purposes include the determination of the perception of the correlations between these categories of concerns and the determination of specificities of distributions of such perceptions, as represented by words. A mixture of tools from lexical, topical, and semantical analysis were applied for generating the bag of words and for analyzing a set of tweets collected during the initial stage of an energy crisis, with special emphasis to coherence indices. The results confirm that energy crises may be associated with such concerns, as expected, but the association probability is low for the 2022 energy crisis, indicating at least a partial success of the mitigation decisions of the respective authorities. Some surprising relationships also occurred from the study. The paper is dedicated to Acad. Florin Gheorghe Filip, at the 75th anniversary.

Key-words: Correlations of concerns; decision-making; energy crises; ethics; natural language processing (NLP); opinion extraction; social networks; Twitter; worries.

1. Introduction

Concerns are a process of human cognition [1]. According to [1], worries expressed in words help focus on the most important fears, "greater attentional bias to threat". Therefore, expressing worries on social networks specifically related to a major cause of anxiety could reveal related concerns not expressed under normal conditions. It has been observed in the literature that social issues elicit high attention and emotions over the Internet and that the relationship between attention and emotions is intricate and change over time [2]. From this point of view, the public attention ("what issues the general public thinks about, *i.e.*, public attention" [2]) was provoked by the issue of energetic crisis, while the public emotion relates to eliciting messages involving the concerns derived from that and related with ethical and socio-ethical issues, or concerns related to the role of technology (ICT), as these reflect how people feel about the crisis (public emotion). The study of the tweets related with ethical and socio-ethical issues, or with the role of technology in crisis reveals manifestations of the interdependency of emotion and cognition [2]. The energy crises may prompt, through association, concerns on social polarization and equity, that is, rich vs. poor division and social or economic discrimination, concerns on ethics, such as bias, honesty, and integrity, and may elicit concerns about the causes of the crisis, such as increased energy consumption by computers, Internet and the recent AI development, briefly by ICT. We analyze these effects as they occur on Twitter and we study their degree of association. The results of this study confirm that energy crises may be associated with the mentioned concerns, as expected. However, the results are surprising in several ways; first, in that the rank distribution of words related to socio-ethical and ethical worries and to ICT concerns differ from the typical Zipf's law; moreover, the correlations between these concerns are very weak. The different exponents in the power laws of selected vocabularies related to concerns inform how the worries are aggregated and how much the probability distributions of the preoccupations expressed by the words in the lists are different from the uniform distribution. A high topical coherence would mean that the distribution is almost uniform, while a typical Zipf rank distribution would mean very low topical coherence. The values of the exponents also inform on the degree of specificity and distributivity of attention to concerns under the stress provoked by the energy crisis; topical specificity is indicated by lower coefficients in Zipf's law for one topical list of words and larger coefficients for other topical lists. Distributivity of attention means that more than one topical list of words has low exponents, meaning a higher topical coherence with more than one topic. The distributions are very different from usual distributions of words and somewhat different between them; that shows a specific degree of focusing due to (and during) worries. The research is significant because of the importance of the energy crises for the well-being of people and countries and because of the impact of these crises on the economic and social life, as proved during the several previous energy crises. Energy poverty is hurting countries and people [3]. Lack of modern sources of energy or their lack of dependability is an indicator of poverty, according to [3]: "It concerns people that have low income, low energy consumption and no access, or limited access, to modern energy fuel". One has even pleaded for "a place for energy poverty in the agenda in energy economics" Birol (2007) [4]. Previous literature provided us with only a limited picture of the reflections of energy crises on individuals. An understanding of these reflections would show ways to better educate the public for facing such crises. After describing the research context in the next section, we present the method in Section 3, show what worries can be derived from linguistics features of the messages, and discuss the results and conclude in the last section.

2. Research Context

Opinion mining on social networks is already an established field aged more than one decade [5]. It has been developed particularly in relation with financial transaction [6], marketing strategies [7], and security [8]. Opinions may be expressed as and detected by sentiments [9, 10], emotions [11], collocations, associations, and other relationships between linguistic items and diffuse sentiments that can be inferred from texts ("sentiment orientation in a structured form from a set of unstructured data" [10]). Opinions may be derived using supervised learning, with neural networks, or similar means [12–14]. Despite the progress, there are still difficulties in reliably determining opinions [15], especially in brief messages as on Twitter [9]; new tools for opinion extraction continue to appear.

From an early stage of this domain development, determination of the public opinions on social networks in view of improving decision-making, including at the governmental level, has been identified as a line of research of great interest [16, 17]. Our research is in line with this direction, also covering the context of dangerous situations. Specifically, we deal with opinion mining in tweets that are related to the energy crisis at the beginning of 2022. The primary tools we use are based on lexical co-occurrences in the same tweet, where the occurrences are from specified sets (bags) of words. The sets of words have been selected to relate each to a specific type of concern the public may have in the given circumstances. Not all possible concerns are dealt with; of interest have been economic worries, ethical concerns, and concerns related to the use of information and communication technology (ICT) use in those circumstances, specifically the power used by ICT infrastructures.

3. Method

3.1. Data collection procedure

The tweets were collected using two methods, generating two sets of data. The first method of data collection used Twitter API V1, with the application in PHP, using the sets of keywords (word strings) $sir[0]='energy crisis'; sir[1]='energy crisis machine learning'; $sir[2]='energy crisis discrimination'; $sir[3]='energy crisis ethics'. The second data acquisition method used Twitter API V2, in Node.js, applying a combination of keywords and tags: {value: 'energy crisis', tag: 'rule_1'}, {value: 'energy crisis machine learning', tag: 'rule_2'}, {value: 'energy crisis discrimination', tag: 'rule_3'}, {value: 'energy crisis ethics', tag: 'rule_4'}.

A total number of a little more than 0.3 million tweets (total number of tweets is 308926 in the analyzed list) have been collected in late February 2022, just after the energy crisis started in late 2021, but before subsequent crises emerged (the crisis due to the conflict in Ukraine). The use of a limited period of time is justified by the consistency of the data, because the public opinion and the issue (the level of crisis) evolve largely during time; see for example [18] (Javaheri, 2020).

The data include many *repeated tweets* (RT); we have not removed them because some of them are most probably answers to other tweets, as proved at a manual checking; otherwise, their re-tweeting was a conscious act with a meaning or intent of the re-sender, and thus the repetitions have to be taken into account. The descriptive statistics of the total, original and re-tweets, is given in Table 1. Only tweets in English have been analyzed.

The correlation between the number of original tweets/day and the number of re-tweets per day is high, 0.823, meaning that the number of retweets per day is well predictable.

Table 1. Descriptive statistics of the number of tweets per day in the collected database

	Number tweets / day	Number RE-tweets (RT @)	Original tweets/day (without RT @)	Original tweets/day vs. total number of tweets	Original tweets/ day vs. RE-tweets
average	8581.3	6215.7	2365.6	0.29	0.41
STDEV	4535.5	3586.0	1101.3	0.06	0.12

3.2. Data collection procedure

The collected data was analyzed for determining the occurrences of the words from three main lists and several secondary lists derived from the first three ones. The words from the first lists are associated with general ethical concerns, such as bias, honesty, and discrimination. The second list includes words related to socio-economical concerns, such as poverty, wealth, and ethics of rich people. The third list includes words related with the ICT and was designed to determine how people connect ICT activities with the energy crises. Shorter lists with the main words in each category were used for checking what the distributions of the "core" notions are.

In some respects, the methods used in this study parallel those used by Bittermann, (2021) [19]. We also follow the desideratum / necessity of putting the analysis in the right sociological and socio-economic context for gaining a meaningful perspective, see Murthy (2012) [20], and we also look to managerial implications. While we agree with Araque *et al.* (2020) [21] that a good starting point for selecting a list of keywords for the analysis is to use a list derived from WordNet synsets or from WordNet-based lists such as Moral Foundations Dictionary and lexicons such as MoralStrength [21], we preferred to build our own lists because we found that some synonyms and words have, for the analyzed database, null or very low probabilities. We believe that these specificities are related with the "detailed subjectivity relations that exist between the actors" exchanging messages on Twitter, as discussed by Maks & Vossen (2012) [22], and that these specificities may be further explained by the relatively low number of actors and tweets in the collected data, which regards a very specific topic. Because of numerous adjustments of the lists of combinations of words and lemmas used, we could say that the lists were manually built, with guidance from the other lists and the current literature. The lists were built in a first version based on extant word lists, grounded in synonym lists and using our own judgment and experience. After checking the occurrence frequencies of words in the original list, some words were dropped because of no appearance and other words were replaced by their lemmas (word roots) for generality. Also, after manually checking some of the tweets where the selected words appear and finding them related to words of no interest (for example, rich and Richard), instead of the initial words "W" we used in the search the version "_W" or "W_", that is, the words with a space before or after.

The probability distributions and the corresponding Zipf's distributions of the words in the three main lists were computed. Then, the probabilities of the joint occurrence in the tweets of words from different lists have been determined and their equality with the values predicted by the hypothesis of independent apparition was checked. In addition, we computed the probabilities of joint occurrence of two or three words from the same list in the messages and verified the hypothesis of their independent apparition. Some of the coherence indices discussed by Röder, *et al.*, (2015) [23], particularly the ones due to Mimno *et al.*, (2011) [24], are computed in view of determining the interplay of the worries in response to the energy crisis. Our purpose is not to select a specific coherence index or to compare coherence indices, but to characterize through NLP tools the relations between the issue (energy crisis) and the concerns of the public, as reflected on Twitter. The topic coherence proposed in [24] is, in their notations,

$$C\big(t, V^{(t)}\big) = \sum_{m=2}^{M} \sum_{l=1}^{m-1} log \frac{D\big(v_m^{(t)}, v_l^{(t)}\big) + 1}{D\big(v_l^{(t)}\big)}, \tag{1}$$

where t is a topic, v_m, v_l are word types, $D\big(v_l^{(t)}\big)$ is the absolute frequency of documents containing at least a word of type $v_l^{(t)}, D\big(v_m^{(t)}, v_l^{(t)}\big) = D(v_m, v_l)$ is the absolute frequency of documents containing at least a word of type $v_l^{(t)}$ and at least one word of type $v_m^{(t)}$ and $(v_l^{(t)}, \ldots, v_m^{(t)}) = V^{(t)}$ is the list of the M most probable words in the topic. We reinterpret the definition and adjust it to the concept of coherence of two topics (here, two concerns); as the coherence concept is better expressed as adherence of topic t_2 to topic t_1, for example adherence of ethical concerns to the socio-economic ones, we will use this term. We define the adherence of two topics in a set of documents (tweets) as

$$A(t_1, V(t_1); t_2, U(t_2)) = \sum_{m=2}^{M} \sum_{l=1}^{m-1} log \frac{D(v_m, u_l) + 1}{D(u_l)}, \tag{2}$$

where $V(t_1)$ and $U(t_2)$ are the lists of the most frequent words in topic t_1 and t_2, $v_m \in V$, $u_l \in U$. The value of M can be chosen, according to [24], in the range 5 to 20. The measure A shows how frequent words from the list U characterizing topic t_2 associate with words in list V, with associations of the most frequent words in the lists given a higher weight. Further, extending the concept of UCI coherence (which is based on pointwise mutual information (PMI), see Newman *et al.* (2010) and Rosner *et al.* (2013) [25, 26], to an adherence notion, the (UCI-type) adhesion of topics defined by [27] as

$$A_{UCI} = \frac{2}{M(M-1)} \cdot \sum_{m=1}^{M-1} \sum_{l=m+1}^{M} log \frac{D(v_m, u_l) + \varepsilon}{D(u_m) D(u_l)}, \tag{3}$$

where $log \frac{D(v_m, u_l) + \varepsilon}{D(u_m) D(u_l)}$ is the pointwise mutual information and ε is recommended to be $1/\Delta$, where Δ is the number of documents [26], in our case, the number of tweets. This version of definition of adherence has the benefit of being a symmetrical relation, with the adherence of t_1 to t_2 equal to that of t_2 to t_1.

The pointwise mutual information between two lists was computed as

$$PMI(V_1, V_2) = log \frac{D(v_m, u_l) + \frac{1}{\Delta}}{D(u_m) D(u_l)}. \tag{4}$$

Both these indices were computed for the lists used and the results are shown and discussed in the next section.

We have not attempted to build a graph of partial correlations because of the very low probabilities of joint apparition of words from different lists, meaning that the graphs essentially reduce to the nodes represented the three lists connected only with main node, energy crisis. In other words, the three lists are greatly incoherent at the topic level, in the sense of topical coherence of sets of words of [23, 24]. We posit that the notion of adherence of one topic to another is a more suitable concept than that of partial correlations for the subject we deal with.

A further step in the analysis was informed by semantic analysis, with adaptation to this research. Because the presence of antonyms may enforce a message, we searched for specific antonyms in the topics of ethics (honest – dishonest) and socio-economic ethics (poor / deprive – luxury; poor / deprive – rich; poor / deprive – wealth) and in contrasting pairs even if not antonyms, such as poor / deprive – dishonest. This approach is believed to have not been applied previously. In addition, judging that some meronyms (has-part relation) may help create antithetical relations, for example "yacht" is "part of wealth" and thus suggests the semantic antithesis with "poor / deprived", we searched for such co-occurrences. In this respect, we depart from the literature that favors the strict use of the WordNet thesaurus for categorization schemes.

4. Inferring Worries from Linguistic Features

We assume that preoccupations and worries are indicated by the linguistic-level content, moreover that public worries that specific causation relations appear may be derived from connections (co-occurrences, correlations) between words from specific sets of words. Specifically, we assume that associations of words pertaining to ethics and ethical relations and words related to specific economic worries in tweets indicate a connection made by the tweeters between ethics and the specific economic situation. The association needs not to be a collocation or a syntactic association as used in word associations.

4.1. Power laws for the main lists

The rank distributions of the words in the two main lists are well approximated by a Zipf type distribution. However, the corresponding power laws are far from the law for the English language, which has an exponent close to 1. The absolute frequencies are rich 11270; pover 4195; poor 3874; wealth 1132; destitut 420; justice 416; bias 334; honest 318; ethic 111; dishonest 89; luxury 46; injustice 37; integrity 31; discriminat 21; unjust 20; unbias 3; opulence 0; opulent 0. The words with null frequencies do not appear in the graph. The power law has an exponent of –0.21, which indicates a strong topical coherence of the set of tweets from the point of view of socio-economic topic. Almost similar results are obtained for the general list related to ICT; the conclusion on coherence is similar, because the exponents are –0.2 and –0.22. (The natural logarithm was used in the linearization of the power law; all words with no apparition were removed.)

4.2. Probabilities of co-occurrences of words from different lists

The absolute and relative frequencies of the words in the main list are given in Tables 2–4. The results are surprising in that negligible correlation was found between these concerns; the frequency of finding terms from ICT and general ethics categories in the same tweet is less than 5.7 E-6, lower than 10-4 as expected from the probability product, showing a relative dichotomy (separation) of these concerns. It is unclear if that is due to the specific circumstances of energy crises. The manual analysis of the tweets and the statistics show that knowledge is limited to a few topics, such as the relation between the power consumption and cryptocurrencies and that of (less power of) new computers, but also shows that there is a total lack of understanding of power consumption of communication networks, Internet, Big Data, data mining, and neural networks. There is proof of serious lack of understanding of several topics by the public using Twitter, such as the bias of algorithms [28]. For example, one tweet #79159 in our collection states that "... because it's code, and therefore unbiased." Yet, the general concept of bias is frequently mentioned, for 103 times. The level of awareness about the role of ICT in the general consumption is average; for example, no one names communications as a major power use; the Internet is named unrelated to power consumption (except data centers). Surprisingly, we found no occurrence of several words often associated with lack of ethics of wealthy people; the words (roots) opulence, opulent, _lush, plush, and lavishness do not appear in the collected tweets, although 'luxury' appears 46 times. (We used "_lush" instead of "lush" because there are other words including "lush", *e.g.*, "Halushka"). The word "deprive" occurs 71 times; however, at manual checking we found that all apparitions are misleading, as they are not related with socio-economic polarization of people (some tweets refer to depriving states, but none to people). Therefore, the true number of apparitions as a word related to people's ethics or socio-economic status is 0; we corrected accordingly the results of the automatic analysis. The words and roots from our initial lists, such as "destitut" (420 occurrence) and "pover" (4195 occurrences) are significant for the purpose of the analysis.

Table 2. Absolute and relative frequencies of words in the various bags of words used, for all the tweets (including RT) and for original tweets only (in brackets)

	In Ethics 3	In CS	In CS core	In Ethics 0	In Ethics, more than on match	In Ethics selected 1	In Ethics selected 2	In Ethics 3B
Absolute frequency	6219	23424	2777	16723	1045	10539	1109	10792
Probabilities (relative frequency)	0.020	0.076 (0.0719)	0.009 (0.0119)	0.054 (0.0575)	0.003 (0.0032)	0.034 (0.0417)	0.004 (0.005)	0.035

Table 3. Absolute and relative frequencies of words simultaneously in various couples of bags of words used, for all the tweets (including RT)

	in CS core and in ethics 0	in CS and in ethics 0	in CS and in Ethics selected 2	in CS core and in Ethics selected 1	in CS core and in Ethics selected 2	in Ethics 3 and ethics 2
Absolute frequency	175	1503	39	165	8	**21**
Probabilities (relative frequency)	0.001	0.005	0.0001	0.001	0.0000	**0.000041**

Table 4. Absolute and relative frequencies of words simultaneously in various couples of bags of words related to ethics

Simultaneously Ethics 3 and in Ethics 2	Ethics 3B in A, B, C or D	Q*CC = CS and Ethics 3B	if in Antonyms semantic I	if in Antonyms semantic II	ce*cf = antonyms 1 AND Antonyms 2
10792	10792	23	4440	7226	685
0.035	0.035	0.0001	0.014	0.023	0.002

The probability of co-occurrence of words from the CS core list and of words from Ethics lists (for all the tweets, including RT) are negligible (73 E-6 overall, 37/(2882+14196)=0.00216653 $\approx$ 0.2%). Even lower are the probabilities co-occurrence of words in CS core and from the lists Ethics selected 1 or Ethics selected 2. Therefore, combined worries in ethical and CS categories have a negligible probability, which is surprising, being given that ICT has fast-growing energy consumption. This may be explained by the public unawareness of the mentioned technical aspect. Also, the numbers of tweets showing combined worries in different categories (ethical 2, ethical 3) are very low. The number of tweets with semantically strengthening words in the socio-economic category is high, 1.4% of the tweets expressing this type of worry, see Tables 5-6. This is supported by the value of the PMI index between the two lists, PMI (Antonym semantic I; Antonym semantic II), is 1.89.

Table 5. Tweets with words from the antonym lists related with socio-economic concerns, for all the tweets (including RT)

	A	B	C	D
	Tweets with words from the Antonym semantic I list	Tweets with words from the Antonyms semantic II list	Tweets with words from both Antonym lists	C/ A ratio
Number	4440	7226	685	0.154
Relative frequency	0.014	0.023	0.002	

The analysis for all tweets and for the not repeated (original) tweets is given in Tables 5 and 6. The PMI for the lists related to computers and ICT (ICT-general and ICT-core lists) and for the ethics-related lists are given in Table 7.

There is virtually no mutual information in the pair of lists with the PMI value less than 1; the large value for the couples (ICT-general; ICT-core) is expected, as one is a subset of the other.

The large PMI value for the couple (ICT general; Ethics selected 1) is explained by the numerous uses of the common word "just" with meanings unrelated to ethics ("has just been hired", "I just think" etc.). The situation is quite different for the couple (ICT-core; Ethics selected 1), showing that the use of adverbial "just" (with the meaning "in recent past" or "exactly") is strongly connected with general terms in ICT-general, but less connected with more specific terms in ICT-core, while the terms "rich, poor, wealth" in topic Ethics 0 are strongly connected with terms in ICT-core but less connected with terms in ICT-general that do not appear in ICT-core. Such specificities might be of interest for the linguists and the sociologists.

The methodology exposed, especially when applied to ethical and social concerns under crises, may be used in multi-participant decision support systems [29] to inform authorities about the public unease and the results of the authorities' policies.

Table 6. Results for original tweets only

in CS AND original tweet	in CS core AND original tweet	in ethics 0 AND original tweet	in ethics more than on match	in CS AND in ethics 0 AND original tweet	in Ethics selected 1 AND original tweet	in Ethics selected 2 AND original tweet	in CS AND in Ethics selected 2 AND original tweet	in CS core AND in Ethics selected 1 AND original tweet	in CS core AND in Ethics selected 2 AND original tweet
6112	1013	4881	269	256	3542	421	26	29	8
0.0719	0.0119	0.0575	0.0032	0.003 (0.005)	0.0417	0.0050	0.0003 (0.0001)	0.0003 (0.0001)	0.0001

Table 7. PMI values for selected lists, for all the tweets (including RT)

	ICT core	Ethics 0	Ethics selected 1	Ethics selected 2	Ethics 3	Ethics 3B
ICT (general)	2.580	0.171	2.709	-0.743	-1.737	-3.529
ICT core		2.567	0.158	-0.102	-1.626	-0.102

5. Discussion and Conclusions

The general ethical concerns were present in 10% of the tweets; socio-economic concerns alone were expressed in 1% of the messages; concerns related with core ICT concepts in the context of energy crises had a probability of 0.6%. The results are interesting in that, despite the hardship during the first months of the energetic crisis, the percentage of the twitting population that expressed concerns regarding ethical issues is low, while the percentage expressing socio-economic worries was very low. This means that, despite the hardship, the population tweeting (in English) seems to have understood and regarded the governments' policies as fair (for the analyzed period). This is interesting, as distributive fairness is a sensitive aspect of fairness that often creates discontent [30]. Also, the results are surprising in that a negligible correlation was found between these concerns, lower than that expected from the probability product, showing a relative separation of these concerns under energy crisis circumstances. At the linguistic level, the results obtained are surprising in that the rank distribution of words related to socio-ethical and ethical worries and to ICT concerns differ from the typical Zipf's law; moreover, the correlations between these concerns are very weak, contradicting one of our initial hypotheses.

In future studies, the topic coherence and topic adherence of tweets will be methodically investigated.

Acknowledgements. We thank Prof. Mike Teodorescu for the idea and main lines of the study, for numerous hints and the proposal of the UCI-type adhesion of topics, and for checking a preliminary version of the paper. We also thank the reviewers for their thorough analysis.

Conflict of interest: Authors are not aware of any reason (economic, funding, or affiliation interests) that might be perceived as affecting the objectivity of this study.

Authors' contribution: HNT has planned, designed with help of Prof. M. Teodorescu, performed data analysis, and written the paper with inputs from MP. The analyzed data have been extracted by MP. Both authors contributed to the optimization of the computer applications for data collection and have not used any automated means to write the paper or to write the computer applications by means of which the tweets were collected. The entire contribution to the present study is human.

References

[1] M. O. WILLIAMS, A. MATHEWS and C. R. HIRSCH, *Verbal worry facilitates attention to threat in high-worriers*, Journal of Behavior Therapy and Experimental Psychiatry **45**(1), 2014, pp. 8–14.

[2] T. Q. PENG, G. SUN and Y. WU, *Interplay between public attention and public emotion toward multiple social issues on Twitter*, PLOS ONE **12**(1), 2017, paper e0167896.

[3] J.-M. CHEVALIER and N. S. OUÉDRAOGO, *Energy poverty and economic development*, in J.-M. Chevalier and P. Geoffron (Eds.), The New Energy Crisis, Palgrave Macmillan, London, pp. 102–132, 2013.

[4] F. BIROL, *Energy economics: a place for energy poverty in the agenda?*, The Energy Journal **28**(3), 2007, pp. 1–6.

[5] K. CORTIS and B. DAVIS, *Over a decade of social opinion mining: a systematic review*, Artificial Intelligence Review **54**(7), 2021, pp. 4873–965.

[6] B. YOON, Y. JEONG and S. KIM, *Detecting a risk signal in stock investment through opinion mining and graph-based semi-supervised learning*, IEEE Access **8**, 2020, pp. 161943–161957.

[7] J. DEIGHTON, A. M. CIECHANOVER and M. H. TEODORESCU, *Shopkick: The power of shopper data*, Harvard Business School, Product #: 517069-PDF-ENG, 2017, pp. 1–17.

[8] N. A. M. RAZALI, N. A. MALIZAN, N. A. HASBULLAH, M. WOOK, N. M. ZAINUDDIN, K. K. ISHAK, S. RAMLI and S. SUKARDI, *Political security threat prediction framework using hybrid lexicon-based approach and machine learning technique*, IEEE Access **11**, 2023, pp. 17151–17164.

[9] S. TAN, Y. LI, H. SUN, Z. GUAN, X. YAN, J. BU, C. CHEN and X. HE, *Interpreting the public sentiment variations on Twitter*, IEEE Transactions on Knowledge and Data Engineering **26**(5), 2014, pp. 1158–1170.

[10] Y. W. LO and V. POTDAR, *A review of opinion mining and sentiment classification framework in social networks*, Proceedings of 3rd IEEE International Conference on Digital Ecosystems, Istanbul, Turkey, 2009, pp. 396–401.

[11] A. M. ABD AL-AZIZ, M. GHEITH and A. S. ELDIN, *Lexicon based and multi-criteria decision making (MCDM) approach for detecting emotions from arabic microblog text*, Proceedings of First International Conference on Arabic Computational Linguistics, Cairo, Egypt, 2015, pp. 100–105.

[12] F. H. KHAN, S. BASHIR and U. QAMAR, *Twitter opinion mining framework using hybrid classification scheme*, Decision Support Systems **57**, 2014, pp. 245–257.

[13] M. LI, X. ZHANG, M. LI, Y. CHEN, Y. BAI, B. ZHANG and R. YANG, *Modeling group opinion evolution on online social networks: a gravitational field perspective*, IEEE Transactions on Computational Social Systems, 2023, DOI: 10.1109/TCSS.2023.3256892.

[14] K. B. VAMSHI, A. K. PANDEY and K. A. P. SIVA, *Topic model based opinion mining and sentiment analysis*, Proceedings of 2018 International Conference on Computer Communication and Informatics, Coimbatore, India, 2018, pp. 1–4.

[15] A. V. MOHAN KUMAR and A. N. NANDKUMAR, *A survey on challenges and research opportunities in opinion mining*, SN Computer Science **1**(3), 2020, paper 171.

[16] G. STYLIOS, D. CHRISTODOULAKIS, J. BESHARAT, M. A. VONITSANOU, I. KOTROTSOS, A. KOUMPOURI and S. STAMOU, *Public opinion mining for governmental decisions*, Electronic Journal of e-Government **8**(2), 2010, pp. 202–213.

[17] K. SCHNATTINGER and H. WALTERSCHEID, *Opinion mining meets decision making: towards opinion engineering*, Proceedings of 9th International Conference on Knowledge Discovery and Information Retrieval, Funchal, Madeira, Portugal, 2017, pp. 334–341.

[18] A. JAVAHERI, N. MOGHADAMNEJAD, H. KESHAVARZ, E. JAVAHERI, C. DOBBINS, E. MOMENI-ORTNER and R. RAWASSIZADEH, *Public vs. media opinion on robots and their evolution over recent years*, CCF Transactions on Pervasive Computing and Interaction **2**(3), 2020, pp. 189–205.

[19] A. BITTERMANN, V. BATZDORFER, S. M. MÜLLER and H. STEINMETZ, *Mining Twitter to detect hotspots in psychology*, Zeitschrift für Psychologie **229**(1), 2021, pp. 3–14.

[20] D. MURTHY, *Towards a sociological understanding of social media: theorizing Twitter*, Sociology **46**(6), 2012, pp. 1059–1073.

[21] O. ARAQUE, L. GATTI and K. KALIMERI,*MoralStrength: Exploiting a moral lexicon and embedding similarity for moral foundations prediction*, Knowledge-Based Systems **191**, 2020, paper 105184.

[22] I. MAKS and P. VOSSEN, *A lexicon model for deep sentiment analysis and opinion mining applications*, Decision Support Systems **53**(4), 2012, pp. 680–688.

[23] M. RÖDER, A. BOTH and A. HINNEBURG, *Exploring the space of topic coherence measures*, Proceedings of 8th ACM International Conference on Web Search and Data Mining, Shanghai, China, 2015, pp. 399–408.

[24] D. MIMNO, H. M. WALLACH, E. TALLEY, M. LEENDERS and A. McCALLUM, *Optimizing semantic coherence in topic models*, Proceedings of 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, Scotland, UK, 2011, pp 262–272.

[25] D. NEWMAN, Y. NOH, E. TALLEY, S. KARIMI and T. BALDWIN, *Evaluating topic models for digital libraries*, Proceedings of 2010 Joint International Conference on Digital Libraries, Gold Coast, Queensland, Australia, 2010, pp. 215–224.

[26] F. ROSNER, A. HINNEBURG, M. RÖDER, M. NETTLING and A. BOTH, *Evaluating topic coherence measures*, Proceedings of 27th Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 2013, pp. 1–4.

[27] M. TEODORESCU, *UCI-type adhesion of topics*, personal communication.

[28] M. TARAFDAR, M. TEODORESCU, H. TANRIVERDI, L. P. ROBERT and L. MORSE, *Seeking ethical use of AI algorithms: challenges and mitigations*, Proceedings of 41st International Conference on Information Systems, Hyderabad, India, 2020, pp. 1–7.

[29] F. G. FILIP, *Collaborative decision-making: concepts and supporting information and communication technology tools and systems*, International Journal of Computers Communications & Control **17**(2), 2022, paper 4732.

[30] L. MORSE, M. H. M. TEODORESCU, Y. AWWAD and G. C. KANE, *Do the ends justify the means? Variation in the distributive and procedural fairness of machine learning algorithms*, Journal of Business Ethics **181**, 2022, pp. 1083–1095, https://doi.org/10.1007/s10551-021-04939-5.