Urban response towards tropical cyclone using twitter in Indonesia

U Widyanarko¹, D R Hizbaron¹*

¹Faculty of Geography, Universitas Gadjah Mada, Sekip Utara, Bulaksumur, Yogyakarta 55281, Indonesia

*Corresponding author: dyah.hizbaron@ugm.ac.id

Abstract. Tropical Cyclone Cempaka struck the southern coast of Java Island in November in 2017, causing floods and landslides in Yogyakarta, Wonogiri, Pacitan, Ponorogo, and the surrounding areas. These disasters devastated urban and rural areas and disrupted community livelihoods as they disconnected inter-city road access, flooded agricultural land, and damaged public facilities. During these hazardous events, many people actively left a “trace” on social media, especially Twitter. This research was designed to describe urban respond using Twitter during Tropical Cyclone Cempaka (TCC). The research data sources were tweets filtered with five keywords with geotagging. After the preprocessing stage, a total of 457 tweets were categorized as “urban” and “rural”. Then, the content was analyzed qualitatively by identifying the most frequently appearing words, the number of response (like, reply, and retweet), and the accounts mentioned on the tweets. The results showed that urban response to these catastrophic events was positive by posting words of prayer and sympathy, disseminating information for raising awareness of disasters, and organizing fundraising for the victims. Turned out that those able to respond was not mainly urban dwellers, there were many responds which based at rural areas. It was marked by similar words that mostly appeared in two comparative areas.

1. Introduction

Tropical Cyclone Cempaka (TCC) is one of many meteorological phenomena occurring at the end of 2017. It was first identified from the emergence of a cyclonic disturbance on November 26, 2017, on the southwest of Cilacap [1]. This low-pressure system grew and moved closer to Java Island for three days (November 26-29) before it dissipated. It induced heavy rainfall, strong winds, thunderstorms, and flooding and landslides that devastated thousands of houses, thousands of hectares of agricultural land, and public facilities. The TCC track shown in Figure 1. The most destructive effects were found in areas closest to its passage, namely Pacitan, Wonogiri, Yogyakarta, and Ponorogo [2]. Although the impacted area hit West Java area (Jakarta, Bandung), Central Java (Yogyakarta, Semarang, Solo and many other urbanized area), and East Java as the prone hazard area, Kangean Island, Bali, Nusa Tenggara, and several other southern part of Java Sea. The occurrence were predicted to last for two or three days, and the alert went to 2.5-6 meter height surge.
Figure 1. Regencies that are directly affected by Tropical Cyclone Cempaka. Source: Bureau of Meteorology Australia, 2017 [3] and Debora, 2017 [2]. (Edited).

Tropical Cyclone Cempaka also disrupted the livelihoods of the community. With the extensive damages to property and infrastructure, access to resources was obstructed [4]. Examples included farmers losing their job because the agricultural fields and plantation areas were either decimated or submerged. Landslides buried road facilities, impeding the distribution of goods. Also, the cyclone-induced disasters damaged school buildings and disturbed teaching and learning activities.

In the era of Industrial Revolution 4.0, the internet of things has recently doing so many things including disaster monitoring system. During emergency situation, there were many activities on social media. Social media has been recognized and used widely as a means of communication and exchange of information [5]. Besides, it is highly effective in collecting information for creating situation awareness, early warning, risk assessment, damage evaluation, and command and control [6]. An example of social media is Twitter, which was launched in 2006 to facilitate users to upload short messages, or called tweets, of no more than 140 characters [7]. As one of the popular social media in Indonesia, it has 50 million local users, with approximately 4.1 billion tweets per year [8, 9].

This research is aim to identify especially urban response over Tropical Cyclones Cempaka (TCC) occurred at Pacitan, East Java Province through social media Twitter. There are numbers of population of social media Twitter with no Geo-Tagging information. This refers to social media without location information. The research focuses to conduct content analysis for social media information relevant to TCC with Geo-Tagging or location information. The arguments of the research are as follow:

1. Those area with geo tagging relevant to indicate distance from hazard prone area, therefore, the research argues, the more the distance of the geo-tagging twitter, the less the intensity and frequency to response to the TCC.
2. Those people who happen to shout out using Twitter, generally urban people, which by all means, has more access to be measured its capacity to deal with disaster using social media.
3. The content analysis indicates that respond was not originated solely from urban areas, it also involves rural areas. Additionally, the content analysis also indicates that there were significant livelihood sectors each in urban and or rural that has been hugely impacted by TCC.

The social media for disaster risk reduction has wide range of interest, although it originated from limited source of literature. Social media is a place where people can freely share their thoughts, feelings, and opinions [10]. Its users can establish contacts and acquire valuable content, information, and even entertainment. When disasters hit, it transforms into a means for people to gain emotional support, create collective awareness, and share information; all of which are crucial in disaster recovery. In doing so, people leave digital footprints on social media, and these records can significantly assist in describing the latest condition of a community. The advantage of social media for disaster risk reduction are crucial especially to monitor up to date situation, which sometime failed to assure its accuracy however it can be easily confirmed once huge amount of responds said so [11].

To add, social media also function as a crowd-sourcing. Crowdsourcing for disaster management has many advantages, for instance, it collects not only near real-time data but also information directly sourced from disaster victims and provides an overview of collective awareness, and most importantly it is cost-effective [12]. However, some drawbacks must be addressed prior to its application in disaster-related analyses, namely data quality, validity, and accuracy as noted earlier [13]. Flaws in crowdsourced data include fake reports, deceiving news and issues (i.e., hoaxes), duplicated information that complicates its processing, and inaccurately geotagged data [12].

The social media in disaster generally captures local to global perception towards the disaster occurrence, specified impacts of disaster towards segmented groups (local people, indigenous community, inclusive groups, etc.), collected humanitarian actions for the disaster, possible access for emergency situation, rehabilitation and reconstruction, existing adaptation strategy and many more [14]. By content, social media covers four types of responses, such as emotional response, instrumental resources support, informational assistance and appraisal mechanism [15]. The research also explore many possibility of using several social media platform. Evidently from the previous research that has been documented, Twitter is the most mentioned social media platform compared to other such as Facebook, Instagram, Youtube, Wikis, Blogs and many other message board [16]. The aforementioned facts described in the following Figure 2.

As noted above, in the intermediate phase of disaster, there are many ways of people express responses. The remaining question is who are those people? Again, derived from the previous studies, most of users of social media platform especially Twitter are citizens (in general), public or government institution, business organization, students, NGOs, media/journalist, public relations or communication staff, and university/college [16]. Partaking to the idea, urban society who happen to have more access to technology and information, generally accustom to social media, such as Twitter. Meanwhile, those rural society who have rare access to internet of things generally get less access to social media. This research raised the argument, who is the social media person during disaster occurrence in Indonesia, especially that relate to TCC. The geo tagging become prior to this statement. Geographical distribution of social media response by all means important to indicate the raised argument [11,16].

The concept of sustainable livelihoods states that the process of livelihood includes the aspects of society, assets, abilities, and ways of life. Tangible assets include natural resources, while intangible assets are claims and access. People use their access to assets as part of strategies to achieve sustainable livelihoods [17]. For each community, there are potentially different asset conditions and, consequently, varied livelihood strategies. Moreover, assets in normal condition are likely to change in the face of disasters. If properly selected, asset utilization strategies can help affected livelihoods to bounce back to and even become stronger than their normal condition [4].
2. Method

2.1. Data collection
Tweets related to Tropical Cyclone Cempaka in 2017 were crawled with several Bahasa Indonesia keywords, which translate to "Cempaka", "tropical cyclone", "flood", "landslide", and "disaster". Bahasa Indonesia was selected to represent the most used language by Twitter users in this multicultural country. The Meteorological, Climatological, and Geophysical Agency first issued an early warning of Tropical Cyclone Cempaka on November 26, 2017. Therefore, the crawling started on the date until the following two weeks, on December 11, 2017. This period was divided into three phases, namely the standby or evaluation phase (November 26-27), emergency phase (November 28-29), and rehabilitation-reconstruction phase (November 30-December 11). Only geotagged posts were analyzed in this research to understand the spatial distribution of disaster-related tweets within the crawling time.

The tweets were crawled using the Twint module (https://github.com/twintproject/twint) with the Python 3.7 programming language. Jupyter Notebook and PyCharm were used to run the Python module and script. After the script for tweet crawling was executed, the tweets were stored in .csv format for further analysis. The tweets provide information like the date of the post, username, used hashtag, and tweet location (or Twitter Place).

2.2. Data processing
The successfully filtered tweets were random and raw (not ready to process). Aside from irrelevant posts, they still contained tweets with arbitrary upper/lowercase combination and unnecessary punctuation. Therefore, further analyses were preceded with a preprocessing stage to eliminate unrelated tweets and correct the letter case and punctuation.

In this study, tweet locations helped to distinguish posts from urban and rural areas. Urban areas are defined with urban features, namely high population density, extensive built-up area, narrow agricultural land, and a large number of economic centers. In Indonesia, urban areas usually occupy the capitals of provinces and several other cities. Located outside the urban periphery, rural areas have
no urban characteristics. Instead, they are typified by extensive agricultural areas, low population density, and dominant agricultural practices.

Furthermore, based on the date of the post, the tweets were grouped into standby, emergency, or reconstruction-rehabilitation phase. The standby phase started right after the Meteorological, Climatological, and Geophysical Agency (BMKG) issued an early warning. The emergency phase marked the incidence of floods and landslides in many places due to the effect of Tropical Cyclone Cempaka and the resulting damages. Disaster management in this phase focuses on searching for casualties and evacuating residents to safer places. The next phase was reconstruction-rehabilitation, in which improvements to affected livelihoods commenced. Rehabilitation concerned on providing psychological and social interventions for distressed communities, while reconstruction dealt with the restoration of damaged physical facilities.

The most frequently appearing words in each group of tweets (urban and rural) were identified and compared to pinpoint any differences between the conditions of urban and rural livelihoods during disasters. The livelihood dynamics were defined by analyzing tweets in three phases of disasters. Aside from the most frequently appearing words, this research also identified the number of retweets, likes, and replies to find out which types of tweets received responses from the public. Also, the mentioned Twitter accounts were observed to understand the type of account that was central to the dissemination of information in the online community.

2.3. Expected results

The data processing generated information about the most frequently mentioned words in urban and rural areas. These words reflect resonance or similarity of thoughts among the society, and therefore, they are assumed to closely represent the actual condition of the affected community [6]. With these words, the research qualitatively described and compared the livelihood conditions in urban and rural areas.

As stated in Blank’s research [18], “no social media platform is representative of the general population”. It means that every analysis that include social media data will not be perfectly described the general population. Not every single person have access to social media. The social media users tend to have higher education, higher income rate, and have more access to the information and technology.

The consistency of social media data could be defined by comparing them to other media sources. The report about damages, victims, and activities occur during the disaster could be found in news media or government institute report. Social media data consist a valid content result if social media data consistently shows a similar information with other official sources [6].

3. Results and discussion

The Twint module successfully crawled a total of 627 tweets. For each crawling process, a keyword was used. Therefore, to filter the tweets with five Indonesian keywords, which translate to "tropical cyclone", "cempaka", "flood", "landslide", and "disaster", the crawling process was executed five times. Furthermore, in the tweet identification process, any duplicated tweets were deleted. Then, irrelevant tweets were erased manually, reducing the number of tweets to 468.

Based on their locations, these tweets were then divided into two, i.e., posted from either the urban or rural area. A total of 136 tweets came from rural areas, and the other 321 tweets were from urban areas. However, the remaining 11 tweets were indefinite because they did not specifically mention from which regencies they were posted (only "Indonesia", "East Java", or "Central Java"). Most tweets were from Java Island, with a high concentration of tweets coming from big cities, such as Jakarta and its surroundings (Jakarta-Bogor-Depok-Tangerang-Bekasi), Surabaya, and Yogyakarta. The first argument which stated that the more it distance from its hazard area, the least of the frequency of tweets, were not proven in the research. Guan and Chen in their research stated that there were a close relationship between disaster damage with Twitter activities [7]. The difference between Guan and Chen’s and this research influenced by information and technology condition. Eastern coast of
America, where Guan and Chen’s research occur, has good information technology condition and evenly distributed availability. Meanwhile, information technology condition in Java where this research occur has worse condition and unevenly distributed availability.

The urban responds were not accumulated in the area where disaster occurs, it originated from numerous scattered urban areas in Java Island. For example, as indicated by a large number of tweets, the high tweeting activity in Yogyakarta was attributable to the direct impact of Tropical Cyclone Cempaka on this city. The same level of tweeting activities was identified in Jabodetabek and Surabaya. Although Cempaka did not directly devastate these regions, there was a high concentration of tweets coming from them. Besides the above urban areas, many tweets were sent from one rural area, namely Pacitan Regency. Cempaka directly affected Pacitan, although the concentration of tweets originating from this regency was not as high as Yogyakarta.

The temporal variation of the number of tweets is charted in Figure 3. It fluctuated from time to time and peaked on December 2. During the standby phase (November 26-27), there was not a single tweet matching the five keywords. In other words, there was no discussion among the community regarding the early warning of an emerging tropical cyclone issued by the BMKG. In the emergency phase (November 28-29), the number of tweets slightly increased. Then, it rose significantly in the next three days and peaked on December 2 before it fluctuated and eventually decreased at the end of the rehabilitation-reconstruction phase (November 30-December 11).

![Figure 3. Graph of Number of Tweets during the Three Phases of Disasters.](image)

After searching for the most frequently referred words through each group of tweets, this research obtained an overview of the condition of the community. Raw tweets data were first processed to remove special characters like "/", ",", ",@", and conjunctions. Afterward, every word that appeared in each group of tweets was calculated for its frequency of occurrence. The most frequently occurring words in urban and rural areas are listed in Figure 4, while rural areas in Figure 5. The bigger the word size in word cloud, the more frequent that word occur in Twitter. Because there was little difference between the words frequently posted from both areas, this study analyzed the top twenty and ten words for urban and rural areas, respectively. This procedure particularly complicated the process of distinguishing typical responses to disasters in these regions. The words "disaster", "cempaka", "cyclone", "tropical", "flood", and "landslide" were included in the top 20 list. Meanwhile, "instagram" and "twitter" were the other two words in this list because users often included links from other posts on different social media as additional information.
There are only a few activities on Twitter in rural areas, as marked by the low number of responses in each tweet (likes, replies, and retweets). Tweets with the most response are those related to disasters, their impact, rehabilitation-reconstruction activities, and donations for disaster relief. Rural communities can share the conditions of their surroundings and disseminate information about the impact of disasters from other sources. Thus, social media can be used to track process of disaster effectively [19]. It illustrates the growing public awareness of information. Community leaders and government institutions are not infrequently mentioned on tweets to let other parties realize the conditions of the affected society and create a concerted effort to provide immediate assistance. Tweets on reconstruction-rehabilitation and donation for relief mostly came from outside the affected areas of floods and landslides (Yogyakarta, Pacitan, Wonogiri, and Ponorogo) and received a tremendous number of responses from the Internet user society. It demonstrates that the unaffected population has excess livelihood resource that allows them to raise or mobilize assistance to disaster victims.

Tropical Cyclone Cempaka did not induce floods or landslides that devastated the majority of big cities observed in this study, except for Yogyakarta. The words “flood” and “landslide” appeared less
frequently in urban society than in rural communities. The urban society mostly responded to tweets on disaster information, donations for disaster relief, reconstruction-rehabilitation activities, and politics. It shows a similar result with Deng et al research. In their research, Guangxi citizens concern about recovery after disaster and their daily activities [6]. Citizens of Guangxi have low damage from typhoon Haiyan, so they talk a little about the damage itself.

Urban society seems unable to escape from the political theme, as evidenced by many tweets coming from the community and political figures that linked disaster events with political topics. An example includes political figures tweeting about raising aid for affected areas or posting political senses of humor that compare the recent election of governor Anies Baswedan and deputy governor Sandiaga Uno of Jakarta to a disaster. Similar tweets often gained responses from the public, raising discussion about disaster awareness. When a community talks non-disaster topic more than disaster related, such as urban community in this research, it indicates that the community is on a stable condition [7].

The behavior of urban and rural communities on Twitter shows differences. Many rural people express the effects of floods and landslides, while their urban counterparts talk about increasing awareness of extreme weather conditions due to Tropical Cyclone Cempaka. An interesting keyword that appeared in urban communities was "dahlia", which is the name of a tropical cyclone that the BMKG expected to develop after Cempaka. It signifies how urban society has been able to anticipate disasters, supporting the assumption that it has a high capacity. Another difference is that rural people often mention community leaders and institutions on their tweets—to ask for assistance, while urban people tend to mention the accounts of news provider and community groups to disseminate information and build awareness of disasters.

4. Conclusion
The conditions of people in urban and rural areas, as seen from their activities on Twitter, have similarities and differences. The similarity lies in how both communities have shared much information about disasters, reconstruction-rehabilitation activities, and fundraising. Meanwhile, they show different interests in disasters: the rural people post about the direct disaster impact that they experience, while the urban people discuss how to increase awareness and preparedness of future disasters.

This research has several limitations to be considered in preparing the next related research. The data used in this research is depends on social media users. Every single tweet produced by users is what to be processed and analyzed in the research. The researcher could not ask for any specific data or information to the users. Thus, the community description from this research really depends on what social media users post in their Twitter.

Twitter data used in this research is only geotagged tweet data. Non-geotagged tweet are excluded from this research because it cannot be located in the map, so the spatial analysis is not possible using those data. Depends on Twitter [20], only 1-2% tweets around the world has geotag information included. It means that this research took only small part of all community’s tweet into analysis process. Community’s tweet produces low representativeness of general population, but it shows consistency with other media sources.

Acknowledgment
This research was one of the beneficiaries of the grant provided by Universitas Gadjah Mada under the 2019 Final Assignment Recognition (RTA) scheme. Authors would like to thank Universitas Gadjah Mada for supporting this research. The topics discussed in this study are part of a broader topic discussed in the final assignment of the first author.

Reference
[1] Magee, Verdon-Kidd D C, Kiem A S and Royle S A 2016 Tropical cyclone perceptions,
impacts and adaptation in the Southwest Pacific: an urban perspective from Fiji, Vanuatu and Tonga *Nat. Hazards Earth Syst. Sci.* **16** 5 pp 1091–1105

[2] MacEachren *et al* 2011 SensePlace2: GeoTwitter analytics support for situational awareness in *2011 IEEE Conference on Visual Analytics Science and Technology (VAST)* 2011 pp 181–190

[3] Bureau of Meteorology Australian Government Southern Hemisphere Tropical Cyclone Data Portal 2017 [Online] Available: http://www.bom.gov.au/cyclone/history/tracks/index.shtml. [Accessed: 31-Jul-2019]

[4] Alexander 2014 Social Media in Disaster Risk Reduction and Crisis Management *Sci. Eng. Ethics* **20** 3 pp 717–733

[5] Blank and Lutz C 2017 Representativeness of Social Media in Great Britain: Investigating Facebook, LinkedIn, Twitter, Pinterest, Google+, and Instagram *Am. Behav. Sci.* **61** 7 pp 741–756

[6] Herman 2017 Indonesia Masuk Lima Besar Pengguna Twitter Beritasatu 2017 [Online] Available: https://www.beritasatu.com/iptek/428591/indonesia-masuk-lima-besar-pengguna-twitter [Accessed: 10-Jul-2019]

[7] Scoones 1998 Sustainable Rural Livelihoods. A Framework For Analysis. IDS Working Paper 72 *Ids* **72**

[8] Fang, Hu J, Shi X and Zhao L 2016 Assessing disaster impacts and response using social media data in China: A case study of 2016 Wuhan rainstorm *Int. J. Disaster Risk Reduct* **2** pp 1241–1256

[9] Li J, Stephens K K, Zhu Y and Murthy D Using social media to call for help in Hurricane Harvey: Bonding emotion, culture, and community relationships *Int. J. Disaster Risk Reduct* **38** p 101212

[10] Rasmussen and Ihlen Ø Risk, Crisis, and Social Media *Nord. Rev* **38** 2 pp 1–17

[11] Goodchild 2007 Citizens as sensors: the world of volunteered geography *GeoJournal* **69** 4 pp 211–221

[12] Jamali, Nejat A, Ghosh S, Jin F and Cao G 2019 Social media data and post-disaster recovery *Int. J. Inf. Manage* **44** pp 25–37

[13] Deng, Liu Y, Zhang H, Deng X and Ma Y 2016 A new crowdsourcing model to assess disaster using microblog data in typhoon Haiyan *Nat. Hazards* **84** 2 pp 1241–1256

[14] Chambers and Conway G 1991 Sustainable rural livelihoods: practical concepts for the 21st century *IDS Discuss*

[15] Ogie, Clarke R J, Forehead H and Perez P 2019 Crowdsourced social media data for disaster management: Lessons from the PetaJakarta.org project *Comput. Environ. Urban Syst.* **73** pp 108–117

[16] Prihadi 2019 Berapa Jumlah Pengguna Facebook dan Twitter di Indonesia? CNN Indonesia 2015 [Online]. Available: https://www.cnnindonesia.com/teknologi/20150327061134-185-42245/berapa-jumlah-pengguna-facebook-dan-twitter-di-indonesia? [Accessed: 10-Jul-2019]

[17] Kurniawan 2017 Siklon Tropis ‘CEMPAKA’ Lahir, Siaga Cuaca Ekstrem 3 Hari Ke Depan BMKG 2017. [Online]. Available: http://www.bmkg.go.id/press-release/?p=siklon-tropis-cempaka-waspada-hujan-lebat-disertai-angin-kencang-dan-gelombang-tinggi-di-wilayah-selatan-indonesia&ctag=press-release&lang=ID [Accessed: 26-Jan-2019]

[18] Twitter 2019 Tweet geospatial metadata [Online]. Available: https://developer.twitter.com/en/docs/tutorials/tweet-geo-metadata. [Accessed: 21-Sep-2019]

[19] Guan and Chen C 2014 Using social media data to understand and assess disasters *Nat. Hazards* **74** 2 pp 837–850

[20] Debora 2017 Dampak Siklon Cempaka Yogyakarta dan Asal-Usul Nama Badai Tirto.id 2017. [Online]. Available: https://tirto.id/dampak-siklon-cempaka-yogyakarta-dan-asal-usul-nama-badai-cAVz [Accessed: 10-Jul-2019]