Social media insights for sustainable development and humanitarian action in Indonesia

Imaduddin Amin, Zakiya Pramestri, George Hodge and Jong Gun Lee
Pulse Lab Jakarta - United Nations Global Pulse
E-mail: {imaduddin.amin, zakiya.pramestri, george.hodge, jonggun.lee}@un.or.id

Abstract.
Tracking human development and humanitarian action has been enhanced by the growth of social media. Twitter is a data source with potential, when used alongside data from surveys, especially the national census, to understand the situation on the ground and track changes. In Indonesia, a country with one of the highest Twitter penetration rates, we seize this opportunity by using Twitter data to produce more timely insights and to enhance evidence-based decision-making. Despite social media’s limitations, namely representativeness and validity, we are able to show its potential by looking at case studies on five different topics; (a) food and agriculture, (b) public health (c) economic well-being (d) urban resilience and (e) humanitarian action. We observe that the insights gained by using Twitter data were derived not only from the content of posts such as understanding public opinion or sentiment, but also from activities related to it, for instance the location and time-stamp of the post, which furthers our real-time understanding of the situation and user behavior changes. In this paper, we also briefly explain “social listener”, a social media monitoring tool that used by Government of Indonesia to understand citizen opinions in social media related to government priorities.

1. Introduction
Traditional statistics, surveys and census data are used to understand public policy issues and to track development. These traditional data, however, are less effective at capturing a real-time snapshot of the situation on the ground in order to inform timely action. New data sources hold the opportunity to complement and supplement traditional data to inform and enhance decision-making in the public sector.

As a core component of the data revolution, social media generate useful data on a variety of issues. Social media platform enables public discourse on a variety of topics, including development and public sector issues. With over 20 million of users, Indonesia ranks third in the world in the number of active Twitter users. This wealth of data has huge potential to provide real-time insights on public interests and issues, including for the Government to understand citizens’ priorities, perceptions and behavior in aggregate.

On Twitter, a user can post content such as text, images, and short videos on any topic. Alongside the content of the post, data is captured on the time, time zone, location, device, and other variables. Others also include gender information, users language and users demographics such as age and gender to extend that basic information. The user can choose whether to attach certain data to the post, such as the location where the tweet was posted.
We group the attributes available from Twitter into two classes, namely “what people say”, and “what people do”. “What people say” contains text-based information shared by a user, such as a tweet text or profile narrative, while “What people do” concerns the information associated with the post or Twitter profile that is related to user activities such as location information, time information, or device used.

In this paper, we would like to address the usage of Twitter data, specifically in the development and humanitarian sectors. We will explore five case studies including (a) food and agriculture, (b) public health (c) economic well-being (d) urban resilience and (e) humanitarian action. In addition, we briefly explain “social listener”, a tool developed by Pulse Lab Jakarta and used by Government of Indonesia (GoI) to monitor public discourse on Twitter.

This paper is organized as follows: Section 2 describes related work on the application of Twitter data to issues of international development and public policy in general. In Section 3, we explain how Pulse Lab Jakarta and GoI use Twitter for sustainable development and humanitarian action, followed by a description of social listener in section 4. In Section 5 we conclude this work.

2. Related Work
The online activities in social media have been subject to a number of studies. Various methodologies have been examined including text mining, sentiment analysis, spatiotemporal analysis, network analysis, among others, to be applied in different fields such as health, humanitarian actions, economy and urban dynamics.

A set of works has demonstrated the potential of using Twitter for investigating public health. A topic model was used to track illness symptoms and awareness, with the quantitative correlations with public health data [1, 2]. Real-time content analysis were explored to provide more real-time insights on public concerns [3, 4]. Forecasting and ”nowcasting” model were also developed using surveillance signals from Twitter to predict and track level of diseases, such as influenza [5, 6, 7, 8, 9, 10, 11], cholera outbreak [12], and mental illness [13]. Another work by Sadilek et al. explored interaction of social activity, human mobility inferred from Twitter, and the spread of infectious disease in a large real-world population [14]. Besides, Twitter streams also used to investigate public concerns related to efforts in improving disease surveillance, for instance, on vaccination [15, 16, 17, 4, 18]. Public sentiments on vaccination were assessed, to investigate the implication on likelihood of disease outbreak [16]. The Twitter pattern behavior were also found to be highly correlated with vaccination behavior based on official government data [17], and the Twitter narrative related to particular vaccination were analyzed, to explore the structure of discussions, connections, and its correlation to the geographical spaces [18].

Prior works also investigated the potential of social media to provide insights during critical situation, such as in natural and man-made disasters. A machine learning methods was proposed to extract relevant information for disaster response [19, 20, 21, 22, 23]. A study proposed framework for using Twitter on event detection over crisis situation [24, 25, 26, 27]. Aside from the data mining methods, other study suggested visual analytic systems based on spatiotemporal analysis to support disaster management [28]. Geographical approached also proposed to incorporate social media analysis for gaining better insight including the spatial patterns of the disaster, particularly in flood [29], and wildfire [30]. A set of works on data-mining and machine-learning techniques that has been applied for disaster management were also summarized in this study [31, 32].

A number of research works attempted to utilize social media data for understanding urban dynamics and social media has been studied for capturing the social behavior in real world events. Wu et al. analyzed social media data to cluster urban hot spots, and found the relationship between the hot spots and housing prices in Shenzhen, China [33]. Another work attempted to analyze geo-tagged tweets for understanding the social-inequality in the city [34]. A
probabilistic topic model to investigate a set of topics related to citizens activities [35]. Gallegos et al. measured public sentiment in tweets to gain insights on land use and services in specific areas [36]. Related to environment, an interesting work by Jiang et al. suggested geo-targeted social media analytic method to monitor air quality dynamics [37]. Other studies analyzed the spatio-temporal dynamics for inferring human mobility and traffic insights [38, 39, 40, 41, 42, 43]. Origin-destination of people’ movement was inferred from the social media mobility network in global-level [44], and local-level [45, 46, 47]. Furthermore, Chan et al. proposed a unified framework combining linguistic model with Hinge-loss Markov Random Fields to identify traffic congestion locations [48].

In Indonesian context, a set of works utilizing social media data also has been conducted. A study extracted information from authority’ Twitter account on reporting traffic condition, to be visualized in a map [49]. The sentiment of tweets using Bahasa Indonesia was classified by applying pre-processing noisy text and machine learning for classification [50]. Further, a sentiment analysis method has been implemented to detect sarcasm messages [51]. Text mining also used to investigate user personality [52] and cyber-bullying pattern in Twitter [53].

3. Twitter Data for Development in Indonesia
In this section, we explain how Pulse Lab Jakarta and GoI use Twitter data for development with five case studies in five sectors, (a) food and agriculture, (b) public health (c) economic well-being (d) urban resilience and (e) humanitarian action.

3.1. Humanitarian actions
Forest and peatland fires continue to affect not only Indonesia but also in regional of Southeast Asia. This situation leads to extensive environmental destruction, economic loss, health problem and others. At present, Indonesia disaster management authorities manage the haze situation by utilizing hot spot data from satellite, sensors deployed in the affected area and official statistics such as population density and distribution. All of those information will inform about the peat fire condition but do not capture a ground truth situation from citizen perspectives such as citizen needs, opinion or citizen behaviour. In order to complement the information gap as well as having more frequent data update, Pulse Lab Jakarta and GoI test the potential of using social media, specifically twitter data.

With the help of domain expert, we defined a set of initial keyword or taxonomy that related to haze situation. For instance ‘kabut asap’ for a general topic on haze, ‘masker’, ‘sesak napas’ for a health-related issue. These keywords are then used to capture relevant message from twitter space to generate insights. From our experiment, we show that these messages are strongly correlated with hotspot dynamics [54].

To understand user mobility, we use location information that available in a geo-stamp tweet. Each user will have set of places that he/she already visit and mobility characteristics inferred from these data. We then compare the user typical behaviour during the normal (non-haze) situation and haze situation. Our result shows that residents make more short-distance movements on normal situation compare to the weeks with haze. Furthermore, we also able to show that some people increase significantly their mobility by hundreds of kilometers during the evacuation week on haze disaster [55].

In addition, we then combine this insights with other dataset used by authority such as hotspot information, place of interest, demographic information and others into one platform1. This platform is now used by authority and opened for public access.

1 http://hazegazer.org
3.2. Urban Dynamics

In 2014, the Indonesian Central Bureau Statistics (BPS) conducted their first commuting survey for Greater Jakarta area that covers 13,120 households. The aims of the survey are to understand the commuting behaviour of citizens in 10 cities and 3 regencies in Greater Jakarta area. This survey result filled the initial data gap needed by the government but the process takes one year from design to delivery.

In order to have similar statistics that relatively easy to frequently reproduce, we then test the possibility of using social media data. To this end, we use more than 38 millions of tweets from 1.4 million of users. Per-user, we inferred two most important locations: origin and destination in sub-district level. Origin location was inferred as the most tweeted sub-district location between 9 pm and 7 am destination location was determined as the most tweeted sub-district location during weekdays, excluding the origin location. Using this approach, we found 305,761 users who meet the criteria. This number of users represents 2.8 percent of the whole population or 14 percent of the total commuting population in Greater Jakarta.

This simple approach produces good results without significant differences. A cross-correlation score between official statistics result and our approach shows 0.97 similarity. We also show that only 3 out of 25 origin-destination pair that has the rank difference between official statistics and our approach [56]. Even though three origin cities have different rank pairs, it is worth to mention that the score of those cities is close each other from official statistics.

3.3. Food and Agriculture

Understanding price dynamics in near real-time is an important aspect concerning the government as a basis to have good policy. Currently, GoI through Ministry of Trade rely on people that deployed in a number of local markets to daily report about price dynamics, including food price as basic commodities. This initiative has good implication to fill the data gaps and reduce the time delay, but has a scalability issue. Driven by this need, we then explore the usability of Twitter data to nowcast food price in national level.

For this purpose, we collect two different types of dataset. First, twitter data as a primary source for nowcasting and official statistics data as a secondary or ground-truth dataset. Twitter data was collected using a set of keyword that related to food commodities such as price and commodity name. For this research, four commodities (beef, chicken, onion and chili) were selected. These commodities were chosen considering data availability and the country-level priorities for food security monitoring in consultation with the Ministry of National Development Planning.

From all relevant tweet messages, we extract two main information, a price mentioned in tweet messages and the volume of relevant messages over time. The numerical filter was applied to extract the first information while the latter was calculated based on tweets captured by taxonomy. We then develop a model to nowcast each commodity based on price information extracted in tweet messages, total tweets today and total tweets yesterday. Once we have the price for a commodity we then compare the result with official data.

The result shows that a price built by the model has a high correlation with the official price such as 0.76 for chili and more than 0.8 for other commodities\(^2\) [57, 58]. This method is later improved by incorporating Google trends data, it and shows higher correlation compared to the previous model.

3.4. Public Health

There are numerous challenges in implementing immunization program including misperceptions about vaccines and dissemination of incorrect information. These factors lead to

\(^2\) [http://nowcasting.unglobalpulse.org/](http://nowcasting.unglobalpulse.org/)
low vaccine coverage and high dropout rates, which means that children do not receive full immunization. Hence, understanding the situation in real-time is important to identify the misinformation disseminated on the ground and taking rapid response accordingly.

Pulse Lab Jakarta together with UNICEF and Government of Indonesia, conduct a project to understand public perception on immunization. We capture all conversation related to immunization and identify the influencer of a topic discussed by the citizen.

To refine the focus of the project, four main sub-topics were chosen based on relevance to immunization program; religious concerns on immunization, conversations around disease outbreaks, symptoms or health conditions discussed as vaccine side effects, and the launch of the new vaccine product. All sub-topics has related keyword or taxonomy that defined by our domain expert. By using the keywords, 88,368 relevant public tweets were collected from January 2012 to December 2013 for Bahasa Indonesia language. From these relevant messages, some analysis was conducted such as understanding the conversation dynamics, word clustering to identify the top keyword for each sub-topic and which content is shared the most by the citizen. We also identify the network of influencers, which we could leverage for rapid response to public concerns and misunderstood information about vaccines and immunization [59].

3.5. Economic Well-being
Public sentiment towards policy changes is one of the topic that can be identified through social media, particularly to the policies which is directly has impact on citizen’ economic well-being. For example, a hashtag #OccupyNigeria was triggered, when government cut fuel subsidy and cause young people mobilize themselves to protest via social media3. Another case was in Yemen, where a number of citizens turned to social media to express their anger on fuel hikes4. Fuel pricing has been one of sensitive issues for citizens, as it may shake political stability as well as hurt the economy of a country.

In order to understand the public sentiments on fuel subsidy in Indonesia, we analyze tweets in 10 - 30 November 2014 using social media monitoring tools namely Crimson Hexagon5. The result shows that the relevant tweets’ volume has significant spike in 17 November 2014, the day when the GoI started to cut the fuel subsidy, with more than 600,000 tweets per day. It is also interesting to see that public sentiment was changed over time toward this issue. The negative sentiments was raised significantly from one week before announcement, then decreased after two weeks as shown in Figure 1

4. Social Listener
Pulse Lab Jakarta developed “Social listener”, a platform that collects, analyze and present information from text data such as social media data, citizen feedback and others. It allows a user to launch a dashboard that monitor conversation from selected data set. To launch the dashboard, user need to provide a list of topic and sub-topics of interest. Each sub-topics is represented by keywords or taxonomy in Boolean search format. This keywords later used as a filter to capture relevant messages from data firehouse, which the insights will be presented in an easy way.

GoI use this platform to monitor citizen opinion related to the government national priorities. For this purpose, two datasets are selected; a citizen complain mechanism, Lapor6 and Twitter. Lapor! is considered as active participation from public as they need to submit to specific platform, while Twitter is considered as passive participation from public. The dashboard

3 http://www.bbc.com/news/world-africa-16591389
4 http://gulfnews.com/news/gulf/yemen/yemenis-protest-against-fuel-price-hike-1.1367338
5 https://www.crimsonhexagon.com/
6 https://lapor.go.id/
launched by Government of Indonesia is an extended version of citizen dashboard that we built for mining citizen feedback in city level [60]. Currently, the platform is able to provide messages distribution and its dynamics, topic cluster and trending topics.

Figure 1. Sentiment classification on fuel subsidy cut related tweets

Figure 2 shows the snapshot of Social Listener for National Citizen Feedback Dashboard from Twitter data for a year. From the image, we are able to see that the monitor has 10 sub-topics of interest. The map shows geographic distribution of relevant messages. A user then able to go deeper in city level to filter only relevant messages from selected city. The darker color means the area has more relevant messages. In this case, Java island has most relevant messages compare to other main islands.

In social listener, a topic of interest can be group into two levels such as topic and sub-topics. In Figure 3 it is shown the proportion of relevant messages and the keyword associate with that. Figure 2 shows that poverty category dominate the conversation and keyword such as “raskin” and “subsidy” are the most common word mentioned in messages. It is also possible that a message will have two or more categories. The relation between a category and other is then presented as shown in Figure 4.
5. Conclusions
Social media have great potential as sources of aggregated citizen perspectives in real-time. However, there are some limitations and challenges that we need to consider before using social media as primary sources, such as representativeness and validity [61]. Despite its limitations, the case studies described in this paper show that Twitter can be used to complement existing traditional data such as surveys and the national census. In addition, it is shown that the potential of Twitter data is not limited to only textual information or “what people say”, but also “what people do” such as location, time and other variables.

References
[1] Paul M J and Dredze M 2011 You are what you tweet: Analyzing twitter for public health. ICWSM
[2] Lamb A, Paul M J and Dredze M 2013 Separating fact from fear: Tracking flu infections on twitter In NAACL
[3] Chew C and Eysenbach G 2010 Pandemics in the age of twitter: Content analysis of tweets during the 2009 h1n1 outbreak vol 5 (Public Library of Science) pp 1–13
[4] Mollema L, Harmsen I A, Broekhuizen E, Clijnk R, Melker H D, Paulussen T, Kok G, Ruiter R and Das E 2015 Disease detection or public opinion reflection? content analysis of tweets, other social media, and online newspapers during the measles outbreak in the netherlands in 2013 vol 17 p e128
[5] Lampos V and Cristianini N 2012 Nowcasting events from the social web with statistical learning vol 3 pp 1–22
[6] Paul M J, Dredze M and Broniatowski D 2014 Twitter improves influenza forecasting (Public Library of Science (PLoS))
[7] Gesualdo F, Stilo G, Agricola E, Gonfiantini M V, Pandolfi E, Velardi P and Tozzi A E 2013 Influenza-like illness surveillance on twitter through automated learning of naive language vol 8
[8] Broniatowski D A, Paul M J and Dredze M 2013 National and local influenza surveillance through twitter: An analysis of the 2012-2013 influenza epidemic vol 8
[9] Signorini A, Segre A M and Polgreen P M 2011 The use of twitter to track levels of disease activity and public concern in the u.s. during the influenza a h1n1 pandemic vol 6 pp 1–10
[10] Nagar R, Yuan Q, Freifeil C C, Santillana M, Nojima A, Chunara R and Brownstein J S 2014 A case study of the temporal and spatiotemporal perspectives vol 16 p e236
[11] Lee K, Agrawal A and Choudhary A 2017 Forecasting influenza levels using real-time social media streams
[12] Chunara R, Andrews J R and Brownstein J S 2012 Social and news media enable estimation of epidemiological patterns early in the 2010 haitian cholera outbreak vol 86 pp 39–45
[13] Choudhury M D, Gannon M, Counts S and Horvitz E 2013 Predicting depression via social media ICWSM ed Kiciman E, Ellison N B, Hogan B, Resnick P and Soboroff I
[14] Sadilek A, Kautz H and Silenzio V 2012 Modeling spread of disease from social interactions In Sixth AAAI International Conference on Weblogs and Social Media (ICWSM)
[15] Dunn A G, Leask J, Zhou X, Mandl K D and Coiera E 2015 Associations between exposure to and expression of negative opinions about human papillomavirus vaccines on social media: An observational study vol 17 p e144
[16] Salath M and Kandelwal S 2011 Assessing vaccination sentiments with online social media: Implications for infectious disease dynamics and control vol 7 pp 1–7
[17] Huang X, Smith M C, Paul M, Ryzhkov D, Quinn S, Broniatowski D and Dredze M 2017 Examining patterns of influenza vaccination in social media AAAI Joint Workshop on Health Intelligence (W3PHIAI)
[18] Radzikowski J, Stefanidis A, Jacobsen K H, Croitoru A, Crooks A and Delamater P L 2016 The measles vaccination narrative in twitter: A quantitative analysis vol 2 p e1
[19] Mendoza M, Poblete B and Castillo C 2010 Twitter under crisis: Can we trust what we rt? Proceedings of the First Workshop on Social Media Analytics SOMA ’10 (New York, NY, USA) pp 71–79
[20] Qu Y, Huang C, Zhang P and Zhang J 2011 Microblogging after a major disaster in china Proceedings of the ACM conference on Computer supported cooperative work - CSCW ’11 (ACM Press)
[21] Imran M, Elbassuoni S, Castillo C, Diaz F and Meier P 2013 Extracting information nuggets from disaster-related messages in social media ISCRAM (ISCRAM Association)
[22] Imran M, Elbassuoni S, Castillo C, Diaz F and Meier P 2013 Practical extraction of disaster-relevant information from social media Proceedings of the 22nd International Conference on World Wide Web - WWW’13 Companion (ACM Press)
[23] Ashktorab Z, Brown C, Nandi M and Culotta A 2014 Tweedr: Mining twitter to inform disaster response 11th Proceedings of the International Conference on Information Systems for Crisis Response and Management ed Hiltz S R, Plotnick L, Pfaf M and Shih P C
[24] Terpstra T, de Vries A, Stronkman R and Paradies G 2012 Towards a realtime twitter analysis during crises for operational crisis management Proceedings of the 9th International ISCRAM Conference Proceedings of the 9th International ISCRAM Conference
[25] Zhou X and Chen L 2013 Event detection over twitter social media streams vol 23 pp 381–400
[26] Salsak T, Okazaki M and Matsu Y 2010 Earthquake shakes twitter users: real-time event detection by social sensors Proceedings of the 19th international conference on World wide web (ACM) pp 851–860
[27] Gupta A, Kumaraguru P, Castillo C and Meier P 2014 Tweetcred: A real-time web-based system for assessing credibility of content on twitter vol abs/1405.5490
[28] Chae J, Thom D, Jang Y, Kim S, Ertl T and Ebert D S 2014 Public behavior response analysis in disaster events utilizing visual analytics of microblog data vol 38 pp 51–60
[29] de Albuquerque J P, Herfort B, Brenning A and Zipf A 2015 A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management vol 29 pp 667–689
[30] Wang Z, Ye X and Tsou M H 2016 Spatial, temporal, and content analysis of twitter for wildfire hazards vol 83 pp 523–540
[31] Houston J B, Hawthorne J, Perreault M F, Park E H, Hode M G, Halliwell M R, McGowen S E T, Davis R, Vaid S, McElriddy J A and Griffith S A 2014 Social media and disasters: a functional framework for social media use in disaster planning, response, and research vol 39 pp 1–22
[32] Li T, Huang Y, Chen S C, Navlakha J, Iyengar S S, Xie N, Zeng C, Zhou W, Zheng L, Jiang Y, Yang Y, Ha H Y and Xue W 2017 Data-driven techniques in disaster information management vol 50 pp 1–45
[33] Wu C, Ye X, Ren F, Wan Y, Ning P and Du Q 2016 Spatial and social media data analytics of housing
prices in shenzhen, china vol 11 ed Choo K K R

[34] Shelton T, Poorthuis A and Zook M 2015 Social media and the city: Rethinking urban socio-spatial inequality using user-generated geographic information

[35] Kling F and Pozdnoukhov A 2012 When a city tells a story: Urban topic analysis Proceedings of the 20th International Conference on Advances in Geographic Information Systems SIGSPATIAL ’12 pp 482–485

[36] Gallegos L, Lerman K, Huang A and Garcia D 2015 Geography of emotion: Where in a city are people happier? vol abs/1507.07632

[37] Jiang W, Wang Y, Tsou M H and Fu X 2015 Using social media to detect outdoor air pollution and monitor air quality index (aqi): A geo-targeted spatiotemporal analysis framework with sina weibo (chinese twitter) vol 10 pp 1–18

[38] Jurdak R, Zhao K, Liu J, AbouJaoude M, Cameron M and Newth D 2015 Understanding human mobility from twitter vol 10 ed Wu Y

[39] Pan B, Zheng Y, Wilkie D and Shahabi C 2013 Crowd sensing of traffic anomalies based on human mobility and social media Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - SIGSPATIAL ’13

[40] Tejaswin P, Kumar R and Gupta S 2015 Tweeting traffic Proceedings of the 2nd IKDD Conference on Data Sciences - CODS-IKDD ’15

[41] Wu L, Zhi Y, Sui Z and Liu Y 2014 Intra-urban human mobility and activity transition: Evidence from social media check-in data vol 9 pp 1–13

[42] Liu Y, Sui Z, Kang C and Gao Y 2014 Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data vol 9 pp 1–11

[43] Chua A, Marcheggiani E, Servillo L and Moere A V 2015 FlowSampler: Visual analysis of urban flows in geolocated social media data Lecture Notes in Computer Science pp 5–17

[44] Hawelka B, Sitko I, Beinat E, Sobolevsky S, Kazakopoulos P and Ratti C 2014 Geo-located twitter as proxy for global mobility patterns vol 41 pp 260–271

[45] Gabrielli L, Rinzivillo S, Ronzano F and Villatoro D 2013 From tweets to semantic trajectories: Mining anomalous urban mobility patterns Lecture Notes in Computer Science pp 26–35

[46] Yang F, Jin P J, Cheng Y, Zhang J and Ran B 2014 Origin-destination estimation for non-commuting trips using location-based social networking data vol 9 pp 551–564

[47] Gao S, Hu Y and Janowicz K 2014 Detecting origin-destination mobility flows from geotagged tweets in greater los angeles area

[48] Chen P T, Chen F and Qian Z 2014 Road traffic congestion monitoring in social media with hinge-loss markov random fields 2014 IEEE International Conference on Data Mining

[49] Endarnoto S K, Pradipta S, Nugroho A S and Purnama J 2011 Traffic condition information extraction visualization from social media twitter for android mobile application Proceedings of the 2011 International Conference on Electrical Engineering and Informatics

[50] Naradhipa A R and Purwarianti A 2011 Sentiment classification for indonesian message in social media Proceedings of the 2011 International Conference on Electrical Engineering and Informatics

[51] Lunando E and Purwarianti A 2015 Indonesian social media sentiment analysis with sarcasm detection vol abs/1505.03085

[52] Pratama B Y and Sarno R 2015 Personality classification based on twitter text using naive bayes, KNN and SVM 2015 International Conference on Data and Software Engineering

[53] Margono H, Yi X and Raikundalia G K 2014 Mining indonesian cyber bullying patterns in social networks Proceedings of the Thirty-Seventh Australasian Computer Science Conference - Volume 147 pp 115–124

[54] Kibanov M, Amin I and Lee J G 2015 Supporting peat fire management using social media

[55] Kibanov M, Stumme G, Amin I and Lee J G 2017 Mining social media to inform peatland fire and haze disaster management

[56] Amin I, PP N L P S, Riyadi Y, Lee J G, Pramana S and Kurniawan R 2017 Inferring commuting statistics in greater jakarta from social media locational information from mobile devices Netmob Book of Abstracts pp 24–26

[57] 2014 Nowcasting food prices in indonesia using social media signals Global Pulse Project Series no. 1 (UN Global Pulse)

[58] Kim J, Cha M and Lee J G 2017 Nowcasting commodity prices using social media vol 3 (PeerJ) p e126

[59] 2015 Understanding immunisation awareness and sentiment through social and mainstream media Global Pulse Project Series no. 19 (UN Global Pulse)

[60] 2017 Fostering Innovation in the Public Sector (OECD Publishing)

[61] Tufekci Z 2014 Big questions for social media big data: Representativeness, validity and other methodological pitfalls