

Journalist System using Tweeter News Media Traffic



Abba Suganda Girsang, Sani Muhamad Isa, Ikrar Harvy

Abstract: *Journalists have some responsibility as their duty. Besides delivering the right news to public, they also have to deliver fast. Therefore getting the potential news fast is one important key for journalist. This study tries to use the information public from social media tweeter. Social media is considered as public information which every people can be involved. This system, starts collecting tweets which mention formal tweeter media. The application will group each tweets into various themes based on the cosine similarity. If cosine similarity is big, it can be indicated as one group news. The tweets which are more than the certain threshold, it can be indicated as the candidate news. This candidate news is then classified into one of ten categories using support vector machine (SVM). This classification has 85% accuracy. The category classified is used for assigning the specific journalist to cover this candidate news. This system has two important modules. One in center server which is responsible for collecting and processing data, the other one application mobile which is installed in journalist is called Journes. The results show the Journes give notification to journalist for getting some candidate news.*

Keywords : *candidate news, journalist, classification, tweeter.*

I. INTRODUCTION

Social media currently takes attention very much for many researchers. One of reasons is the social media offers many data from users. This data is very interesting because this data is generated people who feel the same positions. Therefore data media social is considered as a fair data form people without driven any person or purpose. The use of this social media has influenced the growth of the development communication technology in some aspects such as social-politic, business and government [1].

By writing 140 characters in Twitters, each people can share, judge, deliver the data, ideas, opinion. Therefore, social media is able to bridge between users to share and feedback openly in a short time

In traditional, people share the news by share the link from media web in their user social media. It means many news are found by journalist by collecting or investigate directly in location. Contrarily, journalist currently needs the comments traffic to get the potential news. It is because the news can happen in every where in many times which cannot be covered by the journalist in one time [2][3].

By social media, user can deliver fast and analyze the event which occurs in their around life. The research about the professional journalist which uses media social to get news are hard to find. However, many contemporary journalist might use traffic tweeter for their professional duty [4] [5].

Some researches shows social media has become very famous for its ability to get various domain data such as politics, economics, celebrity and sports [6][7][8][9]. Journalist uses Twitter function to keep contact with their friends to get the professional network, sharing news, and some issues [10].

The contribution of this paper is provide the application for journalist based on traffic twitters media officially. This system starts by collecting data tweet which calls/mentions the media tweeter, then grouping them into potential news or no. At last, the system classifies the potential news into some categories based on journalist' task. This system is implemented to help the journalist getting the potential news fast.

II. RELATED WORK

Social Media is a technology or means that allows us to communicate with others more effectively [11]. Its function is actually the same as other communication media such as telephone, television, correspondence, billboard, radio, but this internet technology-based social media has far better effectiveness. Media is one form of online media, which has special characteristics, others [12]: social media participation allows each user to contribute, comment and provide feedback on a theme posted, openness Most social media services have the feature of voting in the form of opinions, suggestions, as well as sharing information and knowledge, conversation In conventional media, news or material is distributed to people in one direction. But with social media communication can be done both ways (bidirectional), community social media allows people to communicate quickly and effectively. Communities share the same interests, such as love of photography, political issues or a favorite television program. Connectedness Generally, social media develops because there are links to people, sites, another resource.

Revised Manuscript Received on November 30, 2019.

* Correspondence Author

Abba Suganda Girsang*, Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480, Email : agirsang@binus.edu

Sani Muhamad Isa, Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480, Email : sani.m.isa@binus.ac.id

Ikrar Harvy, Computer Science Department, BINUS Graduate Program – Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480, Email : ikrar.harvy@binus.ac.id

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

From the objectives and facilities in it, social media can be classified as follows [13]. Blog is a variant of the content management system (CMS) that makes it easy for everyone to publish short articles commonly referred to as posts. Blogs can be integrated with almost every platform and other tools. For example, blog.detik.com, WordPress, microblogging is the development of blogging where the characters for each post are limited, for example, twitter, where each post will be limited to one hundred forty characters. This limitation has given birth to a truly unique set of features, protocols, and behaviors, social networks are sites where users can connect with other people. In its development, many features exist in social networks such as application plug-ins, groups, and fan pages. For example, Facebook, LinkedIn, media sharing allows users to share multimedia content. Examples of sharing media sites include youtube, Instagram, vine, flicker, SlideShare. social bookmarking websites allow users to filter interesting content from various other websites. Examples: Digg, Reddit, delicious, niche sites, This type of site provides advice, reviews, comments on a topic or product. From this site, users can get feedback from other users' experiences on products or services. For example, TripAdvisor, Citysearch, Yelp, Forums are community development bulletin boards. Forum users can post posts on the forum and later others will respond to the post. Usually, forums are united by a common interest or similarity in relationships. For example automotive forums, school alumni forums, hiking, robotic, planting, and others. For example blackberry groups, forum.detik.com.

Many researches about social media references in newspaper Facebook, Twitter and YouTube are used as sources in newspaper journalism. In 2014, Stave Paulus and Raymond conducted research on social media references in the news papers Facebook, Twitter and YouTube as sources in newspaper journalism, which is a Flemish-quality newspaper that only publishes five articles per day that refer to Facebook, Twitter or YouTube, both as (apart from) the subject or as a source, or both. Of the seven out of ten articles, social media sites serve as sources of information. Facebook and Twitter are more popular as journalistic sources in newspapers than YouTube. Content analysis shows that social media seems to change the pyramid in reverse: although newspaper reporters use social media to extract information from official politicians, institutions, and experts, they mainly refer to social media related to ordinary citizens, celebrities and sports people.

Reconfiguring journalism research about Twitter, one tweet at a time, [5] in 2013 conducted research that social media in general, and Twitter in particular, has increased dramatically in recent years because individuals and institutions have developed; especially by journalists and media organizations. Most journalistic research on Twitter has focused on the dynamics of professional news practices on social media platforms.

Social media as an information source for rapid flood inundation mapping, [12]. In 2015 conducted a research on Methodologies and tools to automatically filter and efficiently support manual extraction of information from the post (content) social media for fast puddle mapping that was presented. In the first step, the processing chain allows a number of potentially interesting social media to be filtered in seconds. In the case of the Dresden application, eighty-four potentially interesting posts were chosen from almost sixteen million posts. In the second step, PostDistiller supports this

process by manually assessing and filtering entries that are derived automatically according to the relevance and plausibility of their content. Finally, information about the inundation depth is extracted. Estimates of the inundation depth can be reduced within three to four hours in the Dresden example. Compared to traditional data sources such as satellite data, social media can provide data faster. The results of previous studies are as expected. The strength of the proposed procedure is that information to estimate fast inundation depth is available, especially in urban areas that have high population densities, and are very valuable because alternative information sources such as remote sensing data analysis do not work well. The photos provided represent an overview of the current situation and thus also help increase awareness and assessment of the situation

III. PROPOSED METHOD

There are some steps in this research as shown in Figure 1. Some important steps are collecting tweet data, preprocessing, determining the top news, assigning journalist. Each steps can be described as sub section 3 of Proposed Method.

A. Collecting tweet data

The collecting data is conducted by gathering the tweets which mention the ten tweet media officially. They are Kompascom, Detikcom, BeritaSatu, MediaIndonesia, Tempo.co, BreakingNews, MediaIndonesia, MncNewsChannel, TribunIndonesia and CnnIndonesia. This media are chosen because it can be considered as big media in Indonesia. This action is done because people usually mention / call the tweet media if they find the event occurs around them. This system uses the API tweet service which is provided by tweeter. This system will be available in all the time. However, the data is evaluated each three hours. This collecting data is built by phyton language and all data is kept in mysql database.

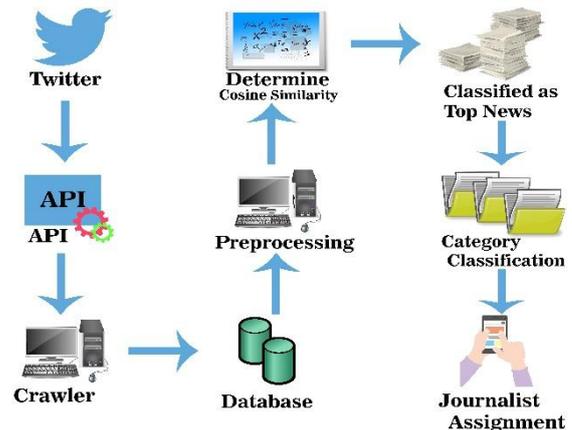


Fig. 1. Process of report news journal application

B. Preprocessing

Data from tweeter mostly cannot be used directly. It can be said the data is unclean to process. Therefore, the data should be preprocessed. The preprocessing consists some processes. They are tokenization, case folding, stemming, and removing stop word. Stopword which is considered with no influence in the classification process, such as the, and, or, to, from, and others will be removed

from data. There are 754 words will be removed from the data.

C. Getting potential / Top News

As aforementioned, the tweet data will be processed each three hours.

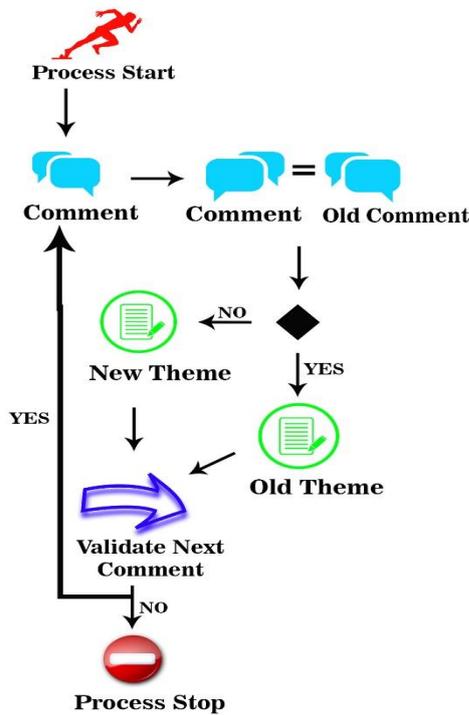


Fig.2 Process of category classification

It means the system will generate eight batches. One batch which consists many kinds comments from various tweet. It could be some tweets are similar meaning or content, even they have a different sentences. Therefore, the cosine similarity is used to evaluate two tweets are similar or no. In this study, the threshold of cosine similarity is 0.7. If two tweets have cosine similarity at least 0.7, it will be indicated as one meaning or content. The system will group all tweets based on this process. This process can be shown in Figure 2. If number twitters which is one group topic more than 20, it can be admitted as a potential or top news.

D. Assigning Journalist

After getting potential news, the system should be able to classify the news. Resulting of this classification will assign the categories journalists. In this study there are ten news categories. To simplify, the media company also have ten journalist categories. The categories are travel, technology, lifestyle, automotive, economy, politic, entertainment, food, health and football. After top news classified into a category, the journalist will be notified based on its category. For example, if the top news is classified as politic category, the application for the politic journalist will be notified. The application for journalist, Journes is built in mobile application because of its flexibility. It is evaluated in each batch (or each three hours) by removing the previous batch data as shown Figure 3

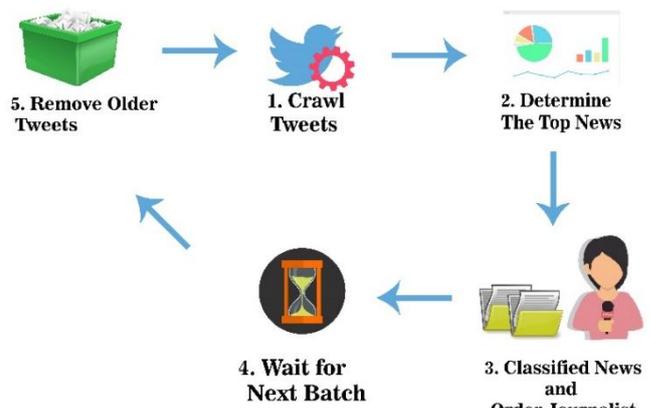


Fig.3. Cycle of the process finding trending news

IV. ANALYSIS RESULTS

A. Setting Variables

There are some variables used in this study as shown Table 1.

Table I Setting Parameter

Variable	Value	Description
Number account twitter	10	Kompascom, Detikcom, BeritaSatu, MediaIndonesia, Tempo.co, BreakingNews, MediaIndonesia, MncNewsChannel, TribunIndonesia and CnnIndonesia
Duration	3 days	
Batch	8/days	Each batches is three hours
Threshold Cosine	0.7	For measuring similarity of two comments
Threshold number comments	10	If number comments more than 10, it can be considered as top news

B. Implementation application

The Journes consists some parts programs. Before using the application, the user is required to fill in the login information consisting of a username and password that has been registered on the application. This user name is for journalists / journalists according to their categories (travel, technology, lifestyle, automotive, economy, politic, entertainment, food, health and football.) . Figures 4 is a sample of filling in username and password information where the username is filled in according to the required news category.

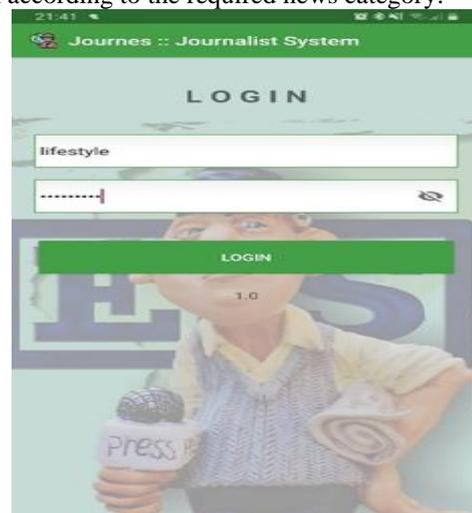


Figure 4 Login Page



Figure 5 Notification for Journalist

If the cellphone sounds and displays as shown in Figure 5, the journalist can see the latest news that is displayed. On this page, users can press the show button to see the latest news. If the user presses on one of the news information in the application, the journalist can see some of the comments reported by Twitter users on the official news media as in Figure 6(a). Some reports to the official news media that have been successfully filtered out are ten among them are @kompascom, @Detikcom, @BeritaSatu, @MediaIndonesia, @Tempo.co, @BreakingNews, @TMCPoldaMetro, @MncNewsChannel, @TribunIndonesia, and @CnnIndonesia. The sample filter of @Kompascom is shown in Figure 6 (b). The news has been grouped based on the level of similarity that has been measured with cosine similarity and categorized using the SVM (Supply Vector Machine) method.

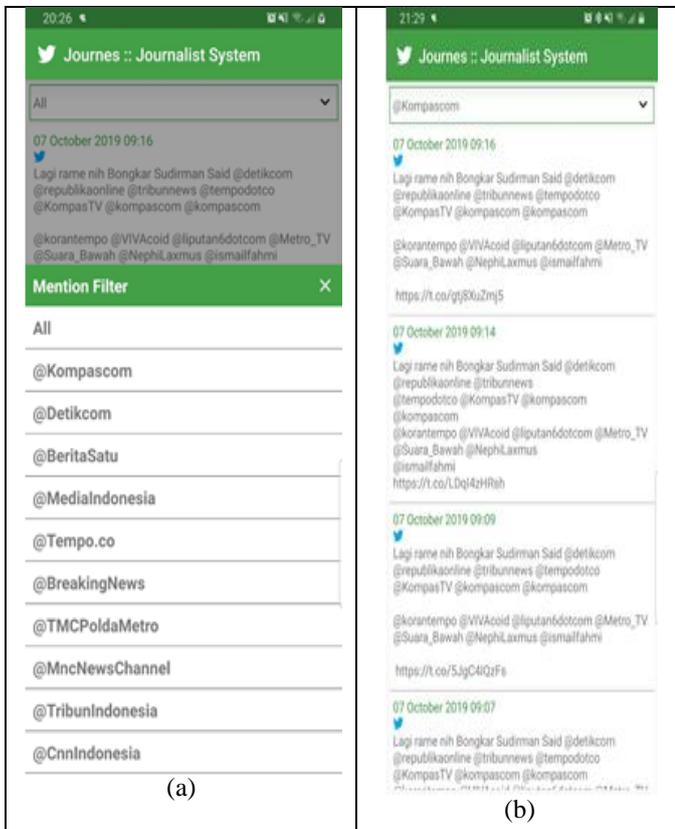


Figure 6 (a) Comments Filter, (b) Comment filter by @Kompascom

C. Performance Analysis

Each three hours, called a batch, data is evaluated. The batches are at 0-3, 3-6, 6-9, 9-12, 12-15, 15-18 and 21-24 which totally is eight batch each day. In this study data is taken during on three days. We got 364-1967 comments each batch. The average is about one thousand comments as shown Table II.

The tweets data are grouped by considering the score of cosine similarity. If the cosine similarity is more than 0.7, the tweets data can be grouped into one group. Each group whose 10 comments are considered as the hot news. Table III shows the detailed of the first day which consists 8 batch. Each batch has some top news. This top news classified into ten categories. We have data training with about 24.584 records with 10 category as shown Table 3. Support vector machine is used for classification the category. By using SVM, the accuracy of classification is 80 %. This machine is then used to classify the comments tweeter as shown in Figure 7.

The detailed of the result can be shown in Table IV. For example, batch 3 has three top news. They are entertainment, entertainment, and politics category which have number 22, 19 and 19 comments respectively. By using this result it can be used for order a journalist to cover the candidate news

Table II Detail Data Tweet in Three Days

Batch	Created		Tweet count	Group count
	Day	Time		
1	1	00.01-03.00	711	28
2	1	03.01-06.00	1704	72
3	1	06.01-09.00	1633	58
4	1	09.01-12.00	1511	34
5	1	12.01-15.00	1732	77
6	1	15.01-18.00	1222	52
7	1	18.01-21.00	509	24
8	1	21.01-24.00	694	21
9	2	00.01-03.00	1920	97
10	2	03.01-06.00	1967	92
11	2	06.01-09.00	1359	2
12	2	09.01-12.00	1515	44
13	2	12.01-15.00	1414	41
14	2	15.01-18.00	1159	49
15	2	18.01-21.00	364	15
16	2	21.01-24.00	473	21
17	3	00.01-03.00	997	49
18	3	03.01-06.00	1482	48
19	3	06.01-09.00	1524	68
20	3	09.01-12.00	1283	51
21	3	12.01-15.00	1441	71
22	3	15.01-18.00	1100	47
23	3	18.01-21.00	289	10
24	3	21.01-24.00	479	23

Table III Distribution data training

Category	Number news
Economy	4721
Health	88
Entertainment	5690
Culinary	181
Lifestyle	2634
Otomotive	369
Politics	4857
Football	3578
technology	2149
travel	317
TOTAL	24.584

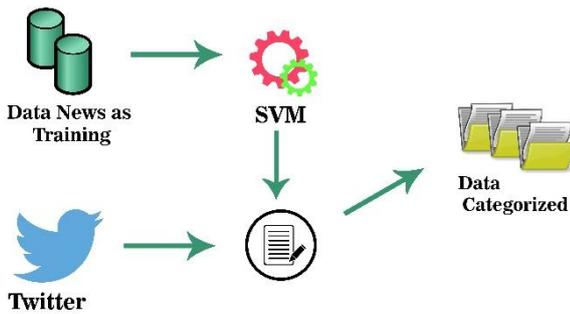


Fig. 7 Classified tweets into a category

Table IV Distribution Top News in First Day

Day	Batch (Time)	Number top news
Day 1	Batch 1	2 (Entertainment (18), politics (14)) =
	Batch 2	4 (Entertainment(29), politics(22), politics (21), politics (14))
	Batch 3	3 (Entertainment (22), entertainment (19), politics (15))
	Batch 4	3 (Entertainment (1414), politics (18), politics (13))
	Batch 5	3 (Entertainment (17), entertainment (15), politics (10))
	Batch 6	3 (Entertainment (1074), politics (15), politics (10))
	Batch 7	1 (Entertainment (16))
	Batch 8	-

V. CONCLUSIONS

Tweeter data can be used to get the top news to ease the journalist gets the potential news. To get the focus news in an appropriate topic, the cosine between comments can be used to measure the similarity comments. Many approaching can be done for improve this system, such as filtering the similarity posting, adjusting number cosine and number threshold of number comments for one top news.

ACKNOWLEDGMENT

This work is supported by the Directorate General of Strengthening for Research and Development, Ministry of Research, Technology, and Higher Education, Republic of Indonesia, as a part of Penelitian Terapan Unggulan Perguruan Tinggi Research Grant to Binus University titled “Sistem Penugasan Jurnalis Berdasarkan Trafik Media Sosial” or “Journalist Assignment System Based on Social Media Traffic” with contract number: 23/AKM/MONOPNT/2019 and contract date: 27 March 2019.

REFERENCES

- Budiyono, M. (2017). Media Sosial Dan Komunikasi Politik: Media Sosial Sebagai Komunikasi Politik Menjelang Pilkada Dki Jakarta 2017. *Jurnal Komunikasi*, 11(1), 47–62. <https://doi.org/10.20885/komunikasi.vol11.iss1.art4>
- Nugraha, R. P. (2013). Pengaruh New Media Terhadap Perkembangan Cyber Society. *Prosiding Seminar Besar Nasional dan Kongres ISKI VI Padang*, 681-686
- Maryulis.(2014). Pengaruh Aktivitas di Media Sosial Terhadap Rutinitas Blogger Sumatera Barat The Influence of Activity in Social Media Toward Writing Habit to Sumatera Barat Bloggers. *Padang. Jurnal Pekommas*, Vol. 17 No. 2
- Oriella PR Network. (2013). “The New Normal for News. Have Global Media Changed Forever?” The 6th Annual Oriella Digital Journalism Survey. <http://www.oriellaprnetwork.com/research>.
- Hermida, Alfred. (2013). “#Journalism. Reconfiguring Journalism Research about Twitter, One Tweet at a Time.” *Digital Journalism*. Online first version. doi:10.1080/21670811.2013.808456.
- Zamith, R., & Lewis, S. C. (2014). From public spaces to public sphere: Rethinking systems for reader comments on online news sites. *Digital Journalism*, 2, 558-574. <https://doi.org/10.1080/21670811.2014.882066>
- Axel Bruns & Jean Burgess (2012) *Researching News Discussion on Tweeter, Journalism* Studies, 13:5-6, 801-814, DOI: 10.1080/1461670X.2012.664428
- Marcel Broersma & Todd Graham (2012) *Social Media as Beat Journalism Practice*, 6:3, 403-419, DOI: 10.1080/17512786.2012.663626
- Hermida, Alfred (2010) *From TV to Twitter: How Ambient News Became Ambient Journalism*, *Media-Culture Journal*, 13
- Mackintosh, E. (2012). *Twitter’s Role as the New Newswire; Why journalists should break news online.*
- Safko, L., Brake, D., David, K. (2009) *The Social Media Bible: Tactics, Tools, and Strategies for Business Success*, NY: Wiley
- Antony, Mayfield 2008. *What is Social Media?*. London: iCrossing.
- Zarella, D. (n.d.) (2010). *Beijing · Cambridge · Farnham · Köln · Sebastopol · Taipei · Tokyo the social media marketing book.*
- Ahmad, Ali N. (2010). “Is Twitter a Useful Tool for Journalists?” *Journal of Media Practice* 11 (2): 145–155. doi:10.1386/jmpr.11.2.145_1.
- Cozma, Raluca, and Kuan-Ju Chen. (2013). “What’s in a Tweet? Foreign Correspondents’ Use of Social Media.” *Journalism Practice* 7 (1): 33–46. doi:10.1080/17512786.2012.683340.
- Schultz, B., & Sheffer, M. Lou. (2016). *An Exploratory Study of How Twitter Is Affecting Sports Journalism*. *International Journal of Sport Communication*, 3(2), 226–239. <https://doi.org/10.1123/ijsc.3.2.226>
- Marwick, A., & Boyd, D. (2011). *To see and be seen: Celebrity practice on twitter*. *Convergence*, 17(2), 139–158. <https://doi.org/10.1177/1354856510394539>
- Allan, Stuart. (2012). “Online News Reporting of Crisis Events: Investigating the Role of Citizen Witnessing.” In *The Handbook of Global Online Journalism*, edited by Eugenia Siapera and Andreas Veglis, 331–352. Boston: Wiley-Blackwell.
- Becker, Howard S. (1967). “Whose Side Are We On?” *Social Problems* 14 (3): 239–247. <http://www.jstor.org/stable/799147>.
- Fohringer, Dransch, Kreibich, Schröter. (2015). *Social media as an information source for rapid flood inundation mapping*. *Nat. Hazards Earth Syst. Sci.*, 15, 2725–2738, 2015.
- Paulussen.s dan Harder,R.(2014). *Social Media References in Newspapers*. London.

AUTHORS PROFILE



Abba Suganda Girsang is currently a lecturer at Master in Computer Science, Bina Nusantara University. He got Ph.D. in the Institute of Computer and Communication Engineering, Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan, He graduated bachelor from the Department of Electrical Engineering, Gadjah Mada University (UGM), Yogyakarta, Indonesia, in 2000. He then continued his masters degree in the Department of Computer Science in the same university in 2006–2008.



Journalist System using Tweeter News Media Traffic

He was a staff consultant programmer in Bethesda Hospital, Yogyakarta, in 2001 and also worked as a web developer in 2002–2003. He then joined the faculty of Department of Informatics Engineering in Janabadra University as a lecturer in 2003-2015. His research interests include swarm, intelligence combinatorial optimization, and decision support system.



Sani Muhamad Isa is currently a lecturer at Master in Computer Science, Bina Nusantara University. He got Dr. in Department of Computer Science, Indonesia University. His research interests big data, datamining, and machine learning.



Ikrar Harvy is currently a student at Master in Computer Science, Bina Nusantara University. His research interests internet of thing, data mining and advanced database