Journal of Information Systems and Informatics

Vol. 1, No. 2, September 2019 e-ISSN: **2656-4882** p-ISSN: **2656-5935**

http://journal-isi.org/index.php/isi Published By DRPM-UBD

Social Media Analytics: Data Utilization of Social Media for Research

Ria Andryani[1], Edi Surya Negara[2*], Dendi Triadi[3]

[1]Information System Departement, Data Science Interdiciplinary Research Center, Bina Darma University, Palembang, Idnonesia
[2,3] Informatic Departement, Data Science Interdiciplinary Research Center, Bina Darma University, Palembang, Idnonesia

Email: [1]ria.andryani@binadarma.ac.id, [2]e.s.negara@binadarma.ac.id, [3]dendi.triadi@binadarma.ac.id

Abstract

The amount of production data generated by social media opportunities that can be exploited by various parties, both government and private sectors to produce the information. Social media data can be used to know the behavior and public perception of the phenomenon or a particular event. To obtain and analyze social media data needed depth knowledge of Internet technology, social media, databases, data structures, information theory, data mining, machine learning, until the data and information visualization techniques. In this research, social media analysis on a particular topic and the development of prototype devices software used as a tool of social media data retrieval or retrieval of data applications. Social Media Analytics (SMA) aims to make the process of analysis and synthesis of social media data to produce information can be used by those in need. SMA process is done in three stages, namely: Capture, Understand and Present. This research is exploratorily focused on understanding the technology that became the basis of social media using various techniques exist and is already used in the study of social media analytic previously..

Keywords: Social Media, Data Analytics, Data Mining, Research, Twitter.

1. INTRODUCTION

Development of technology and information which very fast are allowed society to have communication directly, face to face and it has also come into online communication, such as the using of online media. By using the online, society is able to communicate even they have been separated by a mile or more than it. Actually, these phenomena tend to be called as the using of social media in which it is known as online media that connect the users entire the world with the development of population too.

The users of social media increase every year. This can be evidence-based on the data which promote by communication and information department at

kominfo.go.id and it showed that the use of internet in Indonesia attained 63 million with 95% of them takes the internet to access social media[12]. Based on the survey of Global Web Index in January 2014, the user of the internet in Indonesia filled 72.700.000 person from 251.160.124 person in Indonesia. The survey also showed the active users of social media attained 79.7% from the user of the internet in Indonesia[21].

Looking at the condition above, it can be said that the user of the internet tended to use social media. The uses of social media data are to produce information which can be reviewed or it tends to call Social Media Analytics (SMA) and it has been done in current days. Actually, there are many fields which function SMA, such as economy, business[5], healthy and epidemiology[9][6][2][13], development of social[21] until fields of dynamics city[10][24], academic social network[16], Geospatial data analysis[25][26] etc. Social media is often used to express phenomena which happen in a certain location. Therefore, the data which have been produced from social media is known widely and bigger information. If the data tend to wider form, it will have an opportunity to produce accurate information.

The process of SMA has been done by retrieving, storing, processing, and visualizing data[19] by utilizing various clustering and classification algorithms such as the community detection algorithm that is used to detect communities on social media[15]. Actually, the process is taken as a method to make social media to be one of the research media which is possible to analyze data and it is used as a source of the data in conducting research.

The challenge which is taken in using data of social media is: each site of social media uses different platform volume, complexity from information and the data unstructured[20]. SMA gets the challenge by providing tools and framework to collect, evaluates, analyze, conclude, and visualize data of social media[7].

To use data of social media in supporting research activities which relate to perception society and social behavior are dealing with tools and framework that developed in specific ways to collect, evaluate, analyze, conclude, and visualize the data. The problem which needed to response through this research is the tools and framework, such as: what are things which needed to support research activities on social phenomena which have data on social media. Considering the complexity and kinds of the platform which is used by each media social, thus this research will discuss SMA at micro-blogger site. Twitter or it is often called with Twitter Data Analytics.

This research takes a phenomenon, namely crash of Air Asia with flying code QZ8501 at 28th January 2014 in Strait of Karimata and it looks like an object to

look at perception and social behavior toward the phenomenon. Then it will be processed by using SMA, especially for the data on Twitter which discusses the problem.

2. LITERATURE REVIEW

2.1. Social Media

Internet and Web 2.0 provides a platform which is used to improve services with functionally to creating and sharing ideas and story (Blogger and Twitter); sharing information and links (Delicious, Digg, and Twine); sharing multimedia (YouTube and Flickr), making and sharing knowledge (Wikipedia, Yahoo! Answer and SlideShare) and sharing partnership (Facebook, MySpace and LinkedIn) by big groups. These services are known as social media[18].

Social media is a platform which gives service through two ways, namely making and sharing are taken as tools of new communication in digital era that can be used to reform networking into community and it gives possibility to have communication through online communication in which it tends to make, manage, edit, comment, tagging, discussion, grouping, connecting, and sharing any information which are included on it. Currently, Twitter is one of the familiar social media. Actually, Twitter is a micro-blogging which can be used to send a message to 140 characters by fasting through platforms. In this case, there are 90% interactions of the Twitter but it comes from another website, namely mobile message, fast message or desktop application[4].

Current days, there are many kinds of social media, such as Social networks, Blogs, Wikis, Podcast, Forums, Content communities, Micro-blogging, etc that can be used to get certain aims[4][11]. By implementing related theories on social presence, media richness, and social process Kaplan and Haenlein has claimed that social media can be divided into six types, namely: 1) Collaborative projects, 2) Blogs and micro-blogs, 3) Content communities, 4) Social networking sites, 5) Virtual game worlds, and 6) Virtual communities (Kaplan and Haenlein, 2010). But, if it looked from the category, social media has been divided into four categories, namely: 1) Social Networking, 2) Social Collaboration, 3) Social Publishing, and 4) Social Feedbacks[1].

2.2. Social Media Analytics

Social Media Analytics (SMA) tends to act in developing and evaluating tools of information and framework to collect, evaluate, analyze, conclude, and visualize data of social media[22][7]. Gartner Research defines SMA is a process to look at, analyze, measure, and predict digital interaction, relationships, topic, ideas or contents in social media (Gartner Research, Social analytics). SMA aims to have an analysis process and synthesize data of social media until it gives information which functionally for the stakeholders. The process of SMA takes three steps,

namely Capture, Understand and Present[7]. The steps of SMA can be seen in Figure 1.

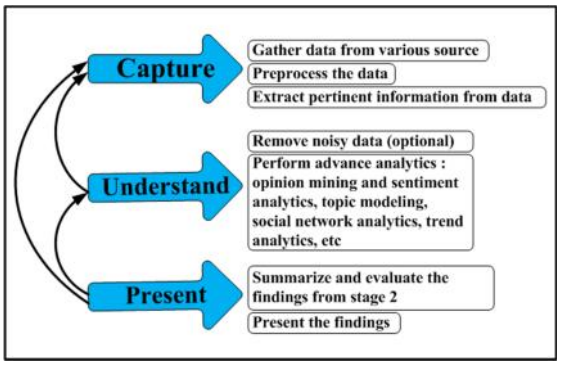

Figure 1. Social Media Analytics Process[22]

The capture on the process of SMA tends to process in collecting data of social media which relevant toward needs by using crawler tools that have been connected to the Application Programming Interface (API) of social media, such as: Facebook, Twitter, LinkedIn, YouTube, Pinterest, Google+, Tumblr, Foursquare, Internet forums, blogs and micro-blogs, Wikis, news sites, picture sharing sites, podcasts, and social bookmarking sites, etc. The data which has been produced by the Capture process will be saved into the database and it is provided to further process, namely Understand the process. In this step, the data is processed to give its information which relevant to needs and it includes creating a model for the data from[22].

After finishing the Capture process, the next is Understand the process. Actually, the Understand process at SMA is a process to choose the data which relevant to apply data modeling, break of noise which includes at the data, selecting the data with high quality and making a process to analyze the data in which it takes to have the best information[22]. In this step, a process of analyzing data gets statistic method, text mining, data mining, natural language processing (NLP), machine translation, machine learning, and network analysis[22]. Many techniques analysis data of social media which are possible to use in producing information: Opinion mining (or sentiment analysis), Topic modeling, Social network analysis, Trend analysis, and Visual Analytics[23].

The last step in the process of SMA is Present. Initially, Present is a process to show or visually information which gets from the step of Understand [22]. In

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

this case, there are many kinds of visualization that can be used to show the information from the analysis process.

2.3. Micro-Blogger Twitter Terminology

Twitter is one of the media social which popular and it has 8th rank at Alexa rank [3]. Initially, Twitter comes from the idea of Jack Dorsey in 2006 and he looks habitual of society which wants to share their activities when they have quality time with other people[14]. In the development of Twitter, Jack Dorsey combined pattern of communication from one to be more and it took as basic of the pattern for communication on Twitter. It gives the possibility to the users of Twitter to share information with other people.

Nepplenbroek, et.al describe the architecture of Twitter development take model "4+1" in which it had been developed by Kruchtens[14]. This model is used to describe the architecture of software by focusing on logical, process, physical, and development and scenario views. By using a model of Kruchtens, Nepplenbroek, et.al. describe the architecture of Twitter with a Logical view, Process view, Physical view, Development view and Scenario view[14].

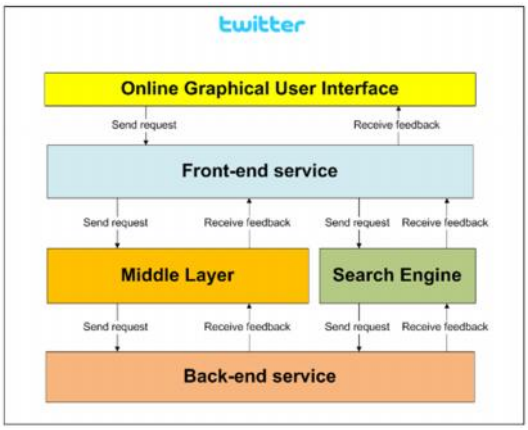

Figure 2. The architecture of Twitter [22]

The architecture of Twitter development can be seen in Figure 2[14]. Appendix of Back-end Service from Twitter saved all of the Tweets which has been posted by members with MSQL as a database of saving data. In the line of Search Engine, Twitter used Apaches Lucene. Search Engine at the Twitter used invert indexing method in which it is used to separate Tweets to be words of a sentence. Actually, a line of Middle Layer at the architecture of Twitter is taken as a requesting system until it cannot burdening Back-end Service. In the first,

Line of Middle Layer is implemented by Starling by using program language Ruby on Rails[14].

3. RESULTS AND DISCUSSION

3.1. Capture Data

Based on the research which had been conducted, it can be seen there was a prototype application of retrieval data and framework for social media analytics. In this research, the writer used Twitter as an object of the research. To analyze the data on Twitter, it needed certain techniques in order for the data (text) on Twitter to be information which took as objects of the research.

The first step that should be done in social media analytics is retrieving data from Twitter. To have retrieved, it should register the application which developed toward Twitter and it needs to be done in order there is permission to have user credential in which it will be used into the process of retrieving data of Twitter. This process aims to get authentication from Twitter toward data access on Twitter. The process of authentication can be shown in Figure 3. After the process of retrieving the Twitter data successes, the next step is saving the data to a database of MongoDB, like Figure 3.

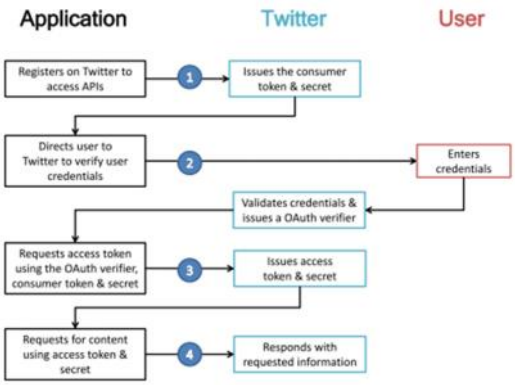

Figure 3. Oauth Proses [11]

In the process of retrieving, the data which got are the username, re-tweet count, tweet followers count, source, and tweet mentioned count, tweet ID, and tweet text. Username is a name of the user or Twitter, re-tweet count describes how many times the status re-tweeted by the other user. Tweet followers count belongs to total of the follower from the user of its account, the source is media that has been used to update the tweets, tweets mentioned count shows how many the tweet which followed, tweet ID is ID of the Twitter user and the tweet text is content of the tweet.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

In the process of retrieving data, there are many factors which affect it, namely the connection of the internet, time to collect data, and up-data new news that will be done. The connection of the internet is very important to process retrieving. The connection of the internet which is stable will support the process of retrieving data Contrary, if the connection unstable, it will be error or process of retrieving is to be slow and break-down.

The second factor is the longest of the process for retrieving data. If the retrieving data of Twitter is done in the longest time, it will have more data from the process.

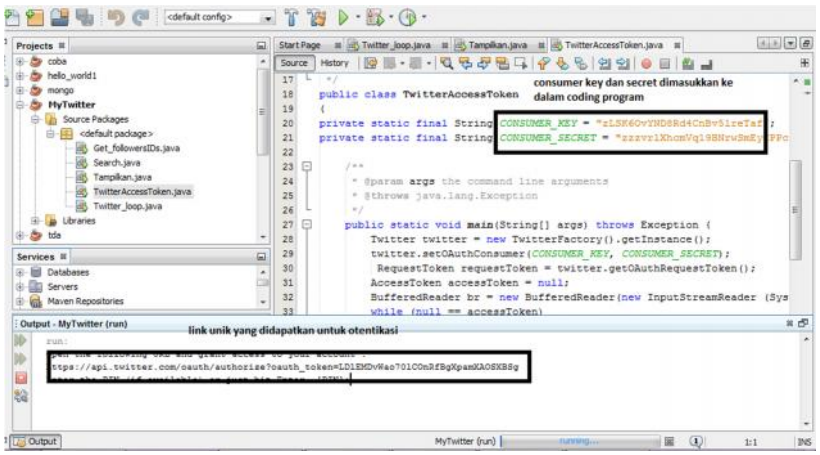

Figure 4. Code PIN Authentication

Then, the third factor is up-data new news. This has happened because retrieving data which has been done by Twitter relate to real-time. Therefore, in this process streaming API is needed. In this case, the data which is taken is real-time data. Retrieving data can be done at the new news which posted 10 days before. Furthermore, the method which is used for retrieving data is the REST API. Thus, the new news which is happening or has been happened will be easy to know the process of development.

Representational State Transfer API (REST-API) is one of architecture model of software to distribute hypermedia system like WWW. The term is promoted firstly by Roy Fielding in his doctoral dissertation in 2000 and he is a major writer of HTTP specification[17].

Specifically, REST tends to a collection of principles networking architecture which stress the role of definition and keep of sources. This term is often used with widely understanding to describe all of simple face-to-face which convey

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

the data on HTTP specific domain without an additional line like SOAP or tracking session which used HTTP cookies.

REST-API on the Twitter can be used to access status or timelines of the Twitter user. REST-API can take 3.200 new tweets from the user include re-tweet[19].

- Main parameter: every page, we can take 200 Tweet from the user.
- Rate limit: an application is allowed to have a request into 300 requests.

3.2. Understanding Data

In this step, the data which took from Twitter entered into the database. The database used in this research was MongoDB.

Furthermore, the data which success saved would be analyzed to get clear data without noises. Clear data can be used as data for the research. To get the comprehension of the data which showed as information, it needs to be visualized into bubble graph form or the other graphs based on the needs.

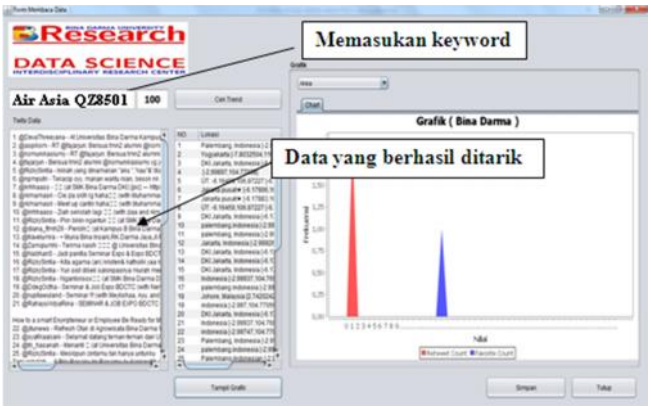

Figure 5. Application of Data Retrieval

The process saving of the data should be done directly or it tends to call as direct storing. It needs to be done because the data which got from retrieving data are real-time data on Twitter. Thus, it needs to use a database which possible to save the data directly.

The result of retrieving data which got from the process of retrieving data tended to the data in text form or document. More than 100 data texts which took through retrieving data. It will be difficult to create primary data manually because the process of retrieving data indicate that the data which got should be managed into the database directly.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

In this research, MongoDB or Mongo database is taken because it tends to document or text database. The ease of using the database is the primary key should not use to insert the data. Because of MongoDB will make ID object or primary key automatically on the process of submitting the data into a database.

In this step, it needs to do field analysis based on Tweet Text and User Name. The process of analysis toward Tweet Text will show the possibility of the countries from the language of the user. The countries have been symbolized with a code of two-letter country based on ISO639-1. The result of the analysis of language identification showed three columns, namely text, language, and rank. Field text showed Tweet or text which up-data by the user of Twitter, field language showed the possibility of the country, and field rank showed the rank of the possibility of the country.

Furthermore, the field which will be analyzed is the User Name. The process of analysis on the username will show the sex of the user, characteristics of the user, and age of the user of Twitter. In this case, the result of the analysis can be called as user demographics.

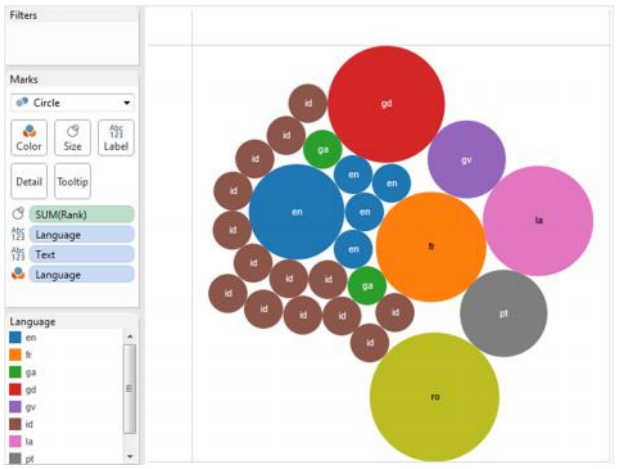

Figure 6. The user of Twitter based on the Countries, Source: Twitter (Airasia QZ 88501)

3.3. Present Data

Visualization is a certain way to convert data into visual format or table until characteristics of the data and relationships between items of the data or attributes can be analyzed or reported. Visualization of the data is one of technique that appropriate and interest to use in exploring the data. Moreover, the visualization can be used to describe the general pattern that happens, a trend which comes as hot issues, or the other phenomena.

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

After the process of analysis finish, the next step is a visualization of the data. It is used to get an interesting form and easy to understand the data ascertain information.

To look at the data which produces information on the easy way, it should be visualized into bubble graph form and the others based on needs. Like previous aims of the research, the result of this research/finding is to make crawling data on Twitter by using Application Programming Interface (API) and it has been provided by Twitter. Based on data of Twitter, it will be processed to be information in which it can be used as the object of the research.

The information is also used to reflect how the behavior of society toward a phenomenon on social media which happen in human life. It shows what the phenomena have influence toward a global society. The influence of the phenomena can be seen from the tweets which up-date by the user of Twitter. In this case, the language uses will be used to look at the possibility where the users come from (country). Based on the analysis on data of Twitter in which it discusses phenomena crash of Air Asia QZ8501, it can be seen that most of the users on Twitter who give attention come from Indonesia.

Out of the language uses, the behavior of society can look from the age of the Twitter user. Based on the one phenomenon, it can be seen that frame of the user's age in which the frame shows the active users on the Twitter. Moreover, sex, personality, and organization which includes on the user can be known through the result of this research.

Figure 7. Sex of the User on Twitter, Sources Twitter (Airasia QZ 88501)

p-ISSN: **2656-5935** http://journal-isi.org/index.php/isi e-ISSN: **2656-4882**

In this research, the writer conducted a research about the crash of Air Asia QZ8501 which happened on 28th January 2014 in the Strait of Karimata-Indonesia. Moreover, the writer will show were come from (country) the users that look at the phenomenon based on language use on their tweets.

It is not just a country, it is also used to show the frame of age which active to observe the phenomenon from the users. Moreover, sex and characteristic of the users whether they come from an organization or person are shown in the result of this research.

Out of the identification toward language uses, this research is also used to show characteristics, the frame of ages, sex of the users. The results/findings of this research can be shown in Figure 6. and Figure 7. The frame of age which discuss the crash of Air Asia QZ8501 are 25 34 years old and the sex is dominated by male.

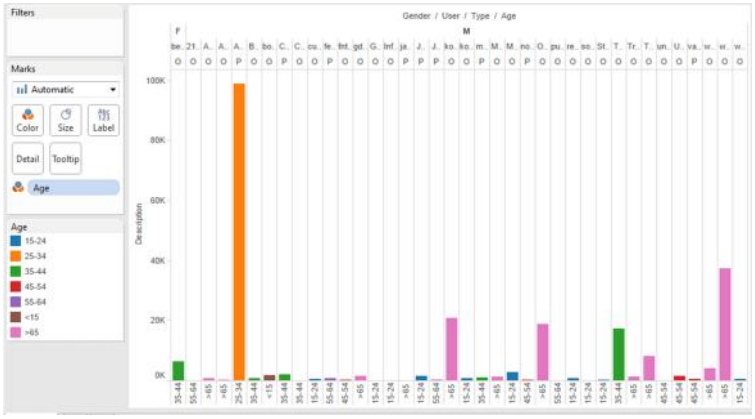

Figure 8. Age of Twitter User, Sources Twitter (Air asia QZ 88501)

4. CONCLUSION

Based on the research on social media analytics, it is Twitter, it can be concluded:

1. Crawling process toward data of Twitter by using Application Programming Interface has been done successfully and it showed that informative data processed through capture, understand, and present.
2. By having the data, it can be identified sex, age, characteristics of the active user and the country which has the use of Twitter, it is known from the language uses.
3. The society gets response toward phenomena. The respond can be measured to know how the extent of the phenomena effects for them.

4. This research can be used to look at how the extent of someone influence or certain issues and the current phenomena.

This research has the limitations in analyzing real-time social media data, therefore it needs a better algorithm to be able to produce more accurate information in analyzing social media data. In addition, this research can be continued by adding other analysis variables such as adding geospatial data and analysis sentiments.

REFERENCES

[1] A. J. Bradley. (2010) Becoming a social organization: Taking a strategic approach to social media. Gartner Inc.
[2] A. Culotta. (2010) Towards detecting inuenza epidemics by analyzing twitter messages, in Proceedings of the First Workshop on Social Media Analytics (SOMA10), Association for Computing Machinery. Association for Computing Machinery, 2010, pp. 115122.
[3] Alexa.com. (2015) Alexa internet. twitter.com. [Online]. Available: http://www.alexa.com/siteinfo/ twitter.com
[4] Antony. (2008) What is social media? Icrossing.
[5] C. Holsapple, S. Hsiao, and R. Pakath. (2014) Business social media analytics: Denition, benefits, and challenges, in Proceedings of the 20th Americas Conference on Information Systems (AMCIS2014), Association for Information Systems. Association for Information Systems.
[6] C. Corley, D. Cook, A. Mikler, and K. Singh. (2010) Using web and social media for inuenza surveillance, in Advances in Computational Biology, ser. Advances in Experimental Medicine and Biology, H. Arabnia, Ed. New York: Springer, vol. 680, pp. 559564.
[7] D. Zeng, H. Chen, R. Lusch, and S.-H. Li. (2010) Social media analytics and intelligence, Intelligent Systems, IEEE, vol. 25, no. 6, pp. 1316.
Global Web Index. (2014) Survei data global web index. [Online]. Available: https://www. globalwebindex.net/
[8] Gartner Reasearch, Social analytics. [Online]. Available: http://www.gartner.com/ it-glossary/social- analytics
[9] H. Carneiro and E. Mylonakis. (2009) Google trends: a web-based tool for real-time surveillance of disease outbreaks, vol. 49, no. 10, pp. 15571564.
[10] J. Cranshaw, R. Schwartz, J. Hong, and N. Sadeh. (2012) The livehoods project: Utilizing social media to understand the dynamics of a city, in Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media (ICWSM) 2012, Association for the Advancement of Articial Intelligence. Association for the Advancement of Articial Intelligence, June 2012.

[11] J. Sterne and D. M. Scott. (2010) Social Media Metrics: How to Measure and Optimize Your Marketing Investment. John Wiley, March.
[12] Kominfo.go.id. retrived September, 2017 from: http://kominfo.go.id
[13] M. Paul and M. Dredze. (2011) You are what you tweet: Analyzing twitter for public health, in Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media (ICWSM) 2011, Association for the Advancement of Articial Intelligence. Association for the Advancement of Articial Intelligence, 2011, pp. 265272.
[14] M. Neppelenbroek, M. Lossek, R. Janssen, and T. de Boer. (2011) Twitter an architectural review.
[15] Negara, E.S. and Andryani, R., 2018. A Review on Overlapping and Non-Overlapping Community Detection Algorithms for Social Network Analytics. Far East Journal of Electronics and Communications, 18 (1), 1-27
[16] Negara, E.S., Kerami, D., Wiryana, M., Kusuma, TM. 2017. Researchgate data analysis to measure the strength of Indonesian research. Far East Journal of Electronics and Communications, 17 (5), 1177-1183
[17] R. T. Fielding and R. N. Taylor. (2002) Principled design of the modern web architecture, ACM Transactions on Internet Technology (TOIT), vol. 2, no. 2, pp. 115150.
[18] R. Brussee and E. t. Hekman. (2015) Social media are highly accessible media.
[19] S. Kumar, F. Morstatter, and H. Liu. (2013) Twitter Data Analytics. [Online]. Available: www. tweettracker.fulton.asu.edu
[20] S. Stieglitz and D. Linh. (2013) Social media analytics and political communication; a social media analytics framework, Social Network Analysis and Mining, vol. 3, no. 4, pp. 12771291.
[21] UN Global Pulse. (2014) Mining indonesian tweets to understand food price crises, UN Global Pulse, Methods Paper.
[22] W. Fan and M. D. Gordon. (2014) The power of social media analytics, Communications of the ACM, vol. 57, no. 6, pp. 7481.
[23] W. Fan, L. Wallace, S. Rich, and Z. Zhang. (2006) Tapping the power of text mining, Communications of the ACM, vol. 49, no. 9, pp. 7682.
[24] X. Long, L. Jin, and J. Joshi. (2012) Exploring trajectorydriven local geographic topics in foursquare, in Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp12), Association for Computing Machinery. Pittsburgh, United States: Association for Computing Machinery, 2012, pp. 927934.
[25] Negara, E. S., Andryani, R., & Saksono, P. H. (2016). Analisis Data Twitter: Ekstraksi dan Analisis Data G eospasial. INKOM Journal, 10(1), 27-36.
[26] Antoni, D., Negara, E. S., & Suweno, S. (2015). Ekstraksi Data Geo-Spatial Twitter (Studi Kasus: Badan Penyelenggara Jaminan Sosial Kesehatan).