

Analyzing Civic Satisfaction on Public Service Using Twitter Data

Ade Hendraputra¹
Ministry of National Development Planning/BAPPENAS

Abstract

In recent years, the digital technology has transformed rapidly, allowing social media to become an important place for people to express their thoughts and opinion on any topic that they want. In this paper, we present an analysis of twitter data regarding civic satisfaction and opinion of public service in Indonesia. We try to capture the tweets which discussed several topics related to traffic condition, public service and driving license service. Then we present an analysis process using a framework and present the result which may help government have a better picture of certain problems.

¹ Ade Hendraputra is a Planner Staff at Directorate of Public Private Partnership and Financial Engineering, Ministry of National Development Planning/BAPPENAS. Email address: adehendraputra@yahoo.com

Analyzing Civic Satisfaction on Public Service Using Twitter Data

Ade Hendraputra, BAPPENAS

I. Introduction

Today internet network is crowded by mobile device which boost internet user up to half of the human population (3.649 billion internet users) and 75 percent of internet users is using their mobile device like a cell phone and tablets to access information on the internet. The simplicity of the internet makes people really close to their gadgets. This fact leads people to enter a new era of communication. Social media is a tool that people use to communicate and discuss the wide spread of topics.

Furthermore, the Government which has a mandate from their citizen to develop their country should recognize this new communication technology as a complimentary way to assess their performance on serving the citizen and measure their citizen satisfaction level.

This research provides a literature review on how social media can be used and leveraged as a tool to analyze civic opinion and satisfaction on several interested topics on government program and simulation as well as statistical analysis on twitter data which crawled in a certain time period.

As a initial hypothesis, we also want to prove that social media analysis is easy and instant for government officers to implement and can be a good tool during public policy planning and evaluation process.

II. Related Work

2.1 Public Sector and Big Data

Governments used to receive a huge demand and pressure to increase their productivity. Especially after we meet global recession in the last 2008, many governments has a difficulty to provide an excellent public service which at a same time, they also tied into the budgetary constraints that force them to reduce the large number of deficits and high level of national debt. Moreover, public money also needed to stimulate economic growth which can help their economy run well. Furthermore, this economic problem has also caused by high demand for medical and social services to serve their citizens.

As recent MGI research has shown, the way that government need to overcome with such constraints is by walking through a change in their productivity which sometimes relatively has a low level. While the productivity level in the public sector is hard to measure, MGI research revealed an evidence that public sector productivity growth has fallen behind that of the private sector in many (or most) economies. Furthermore, UK office of National Statistics experiment reported declining productivity in the public sector from 1995 to 2005, while MGI research reported increasing growth of productivity in the market sector. In the big picture, the declining productivity performance of the UK public sector was caused by the government's increasing employment.

Is it possible big data and social media can help the public sector to win its battle of productivity? To attempt to answer this question, we studied the civic, social media activities of the public sector performance, with a focus on Indonesia country. McKinsey research found that big data levers, for example, increasing transparency

and applying advanced analytics, offer the public sector a powerful tool of strategies and techniques to increase productivity and achieving higher levels of efficiency and effectiveness. In McKinsey research case study shows Europe’s public sector could potentially reduce the costs of administrative activities by 15 to 20 percent, which make the equivalent of €150 billion to €300 billion (\$223 billion to \$446 billion)—or even higher—in newest research data. This estimate includes both efficiency gains and a reduction in the distance between the actual and potential collection of tax revenue. Furthermore, these effort could accelerate annual productivity accretion by up to 0.5 percentage points over the next ten years.

We believe that big data can play a similar role in other countries and regions, too. Other governments around the world face many of the same social and economic challenges, and the big data opportunities we discuss above will apply elsewhere, particularly in other developed countries Indonesia. Private sectors are familiar with this technology for several purposes such as their market strategy and decision support system. So we believe that in the simplest way, this big data can help the government to get an instant input from their citizen in social media.

2.2 Indonesia Technology Demographic

Indonesia is undoubtedly one of the most attractive markets in Southeast Asia. With a population of over 250 million – the largest in the region and the fourth largest in the world, after China, India, and the US. According to We Are Social’s compendium of world digital stats, Indonesia now has 88.1 million active internet users, up 15 percent over the past 12 months.

Its mobile market has exploded over the past couple of years. SIM subscriptions in Indonesia stand at 326.3 million, way more than its population. This means each mobile phone user owns an average of two SIM cards. 85 percent of the population own mobile phones, while 43 percent carries smartphones.

With nearly 79 million social media users or 30% of the total population, we believe that social media should be a valuable asset for finding out people statements regarding some issue that matter for the government.

Twitter comes in third as most popular social media platform in Indonesia, with close to 17 million users. In recognition of the importance of this region, Twitter opened an office in Jakarta in March 2015.

Figure 1. Twitter Users

Twitter Users and Penetration in Select Countries in Southeast Asia, 2014-2019
millions and % of internet users

	2014	2015	2016	2017	2018	2019
Twitter users (millions)						
Indonesia	12.0	14.3	16.8	18.9	20.9	22.8
Philippines	5.5	6.6	7.5	8.5	9.5	10.4
Vietnam	4.0	4.7	5.4	6.1	6.7	7.3
Thailand	2.7	3.1	3.4	3.7	3.9	4.1
Malaysia	1.5	1.7	1.9	2.1	2.2	2.4
Singapore	0.4	0.4	0.5	0.5	0.5	0.6
Twitter user penetration (% of internet users)						
Indonesia	14.3%	15.4%	16.3%	16.8%	17.0%	17.1%
Philippines	11.4%	12.2%	12.8%	13.2%	13.7%	14.1%
Malaysia	11.1%	11.9%	12.4%	12.7%	12.8%	12.9%
Thailand	11.0%	11.8%	12.3%	12.6%	12.8%	12.9%
Vietnam	9.8%	10.5%	11.1%	11.6%	12.1%	12.2%
Singapore	9.2%	10.3%	10.8%	11.3%	11.7%	12.1%

Note: internet users who access their Twitter account via any device at least once per month
Source: eMarketer, Nov 2015

200380 www.eMarketer.com

The social media platform has also been credited for its big role in the Indonesian presidential elections in 2014. Today Indonesian president Mr. Joko Widodo's which sponsored tweets by one of his election consultants had 2014's second most retweeted tweet. So we believe that if the government can still use this platform to deliver government campaign, they may have a better citizen involvement and support of their people.

2.3 Twitter

Twitter is a microblogging service and belongs to the category of online social networks. Users of Twitter have the possibility to send short messages with a maximum length of 140 characters to communicate with each other or just to state their opinions. In Twitter's wording such a message is called a status, whereby it is also common to call it a tweet. To subscribe to the tweets of another user, users have to set up a follow-relationship, in example, user A will receive all tweets of user B and therefore A is called a follower of B. As a consequence, user B is called a friend of user A. The difference in the naming of the direction of the relationship between two users can be explained by the unidirectional manner of the relationship in Twitter's social network.

Twitter has been growing in popularity since its opening in 2006. According to twitter (2016), on 31 Dec 2013 there were already more than 313 million active users per month which cover nearly every country in the world. Approximately 1 billion tweets are created every day (Twitter, 2016). Like other online social networks, Twitter is also extensively used to gain interesting information about trends and opinions as well as marketing tools. Moreover, one can list, in example, companies utilizing the dialogic capacity of social media for communication with stakeholders (Rybalko and Seltzer, 2010), (Lovejoy et al., 2012), emergency services, obtaining up-to-date information for disaster relief (Chatfield, et al., 2014), and health organizations detecting (Aramki, et.al., 2011) and predicting (Achrekar, et al., 2011) outbreaks of diseases, to mention just a few instances of Twitter's widespread utilization. The number of areas of application is increasing steadily and tools that rely on Twitter become more and more sophisticated (MacEachren, et al., 2011). For example, in the case of an earthquake, it takes only seconds until first related tweets appear, which in turn can be utilized to visualize earthquake activities on a map (Sakaki, et al., 2013). Such kind of news or, to be more general, information, is propagated much faster via Twitter than by common news agencies.

For researchers and trend analysts Twitter offers two technically different possibilities to access this huge amount of data. There is the option to collect tweets happening in real-time by using a stream access to the public tweets. The other option is a using Twitter API, which provides the possibility to request free selectable information, e.g., details about a special user or a user's tweet timeline. The Twitter API suffers from limitations in respect of the frequency of allowed data requests, while the tweet amount of the free-of-charge stream access is limited by a public streaming cap. That's the way after serial experiment using open source application and found that limitation hinder us to get a complete dataset, finally we found Twitter Achiever offers as an add-on in the Google Spreadsheet which offer simple and cloud based computing service.

III. Methods

3.1. Dataset

One of the main issues of the analysis is the availability of adequate datasets. A method that called "Firehose" access, offers the complete data, but is very expensive to get. The other variant which we called "Spritzer" is using a sample stream which, according to Twitter and other researcher, offers only one percent of all tweets, even

though it is free of charge. Alternatively, one can access the stream by applying a filter like keywords or geo-coordinate bounding boxes. The output of this filtered stream is still limited to not more than 1% of the global count of tweets. In the past, there was another sample stream available, called "Gardenhose", which offered 10% of all tweets. Today, signing up for a free Gardenhose access is no longer possible, nevertheless the existing accounts are still valid. Hence, there are still users of the Gardenhose stream, today. In this paper, we examine the Twitter's streams and direct our attention to Twitter's statement of these streams being lively collected by selecting keywords. To obtain the dataset for our studies, we accessed Twitter's streaming Twitter API "Twitter Archiver" using the Google sheets API.

Twitter Archiver lets us easily save tweets for any search keyword or hashtag in a Google Spreadsheet. Enter a search query, or a hashtag, and all matching tweets are automatically saved in the Google Sheet. You can write simple queries, use boolean search or include advanced Twitter search operators to create more complex queries. Twitter Archiver can be used for saving tweets around trending hashtags, conference tweets, your brand mentions, Geo-tagged tweets, and more. It polls, twitter every hour and pulls all the matching tweets into the Google Spreadsheet.

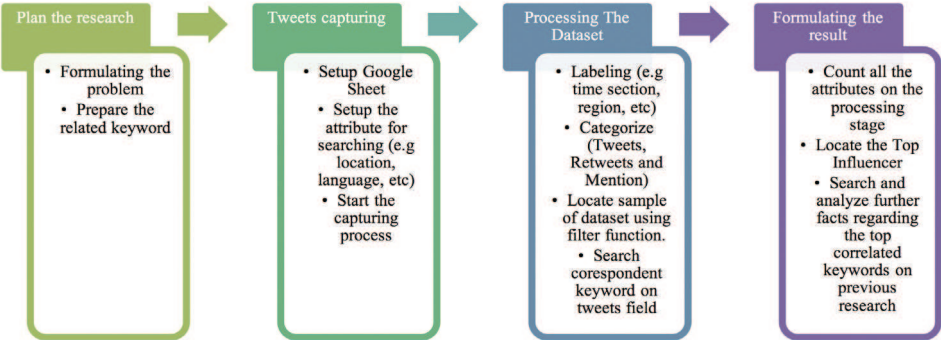
The output file of our dataset is based on spreadsheet format. Twitter Archiver generates a new worksheet with the following fields which appear in the following order.

- Date
- Screen Name
- Full Name
- Tweet Text
- Tweet ID
- App
- Followers
- Follows Retweets
- Favorites
- Verified User Since
- Location
- Bio
- Profile Image
- Google Maps

3.2 Our Approach And Findings

Our experiment starts with creating a research procedure to examine the dataset for social media analysis. Below is our concept diagram based on our iteration in this research.

Figure 2. Research Process Diagram

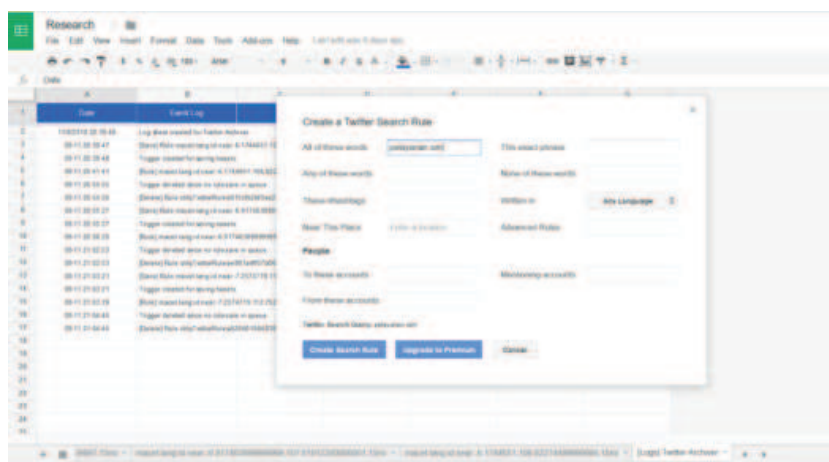


Based on PDCA framework, which is commonly used by most of government project management, we suggest that this technique can be used on both Plan and Check stage. In our research we try to simulate using three problems, understand the problem based on public opinions and facts on the social media and show the formulating suggestion to overcome the problem.

Furthermore, we test this framework using 3 problems (two specific problems and one general problem). The first two problems are traffic jam and driving license service represent a specific problem, then public service as a general problem. So basically, we want to find the actual public opinion on those three issues.

Then we continue capturing tweets process. In this process, thank to Labnol which provide a robust Google Sheet API to search and easy to set up (Agarwal, 2015). Moreover, during the capturing process, we can maintain our anonymity aspect. This aspect is important for public policy planner in the very beginning process because we could test the policy without concern of effecting public behavior before the policy tested or implemented.

Figure 3. Twitter Archiver Search Interface

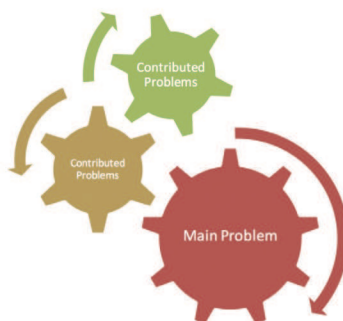


The next process is processing the dataset. At this stage, we also tried to feed the dataset into several Social Media Analytics tools, but most of them has their own structure so we had a difficulty to use it and takes time to reformat it especially in a big number of data. For people who has not or low level of computer skills, it would be a nightmare for them and hinder people to buy-in to this issue. So learn from predictive modeling concept, we offer this simple and straightforward concept to analyze the data. Basically, first people need to be labelled data based their output. For example, if you need a further research on time frame and geo-tag, you label the dataset by adding new columns and fill it. Then, people examine few correlated keywords on sample data, means that read a few tweets on the dataset and write on a piece of paper which later it would be used to analyze the whole data set. Literally, we believe that user just needs an spreadsheet skills to examine and analyze the data.

Finally, in the last stage, which is formulating the result, we execute and count the correlated keywords on the whole dataset so we can see how much that suspected keywords affect the main problem. Then use the previous research and facts regarding that correlated keywords to explain the contributed problems of the main problems.

Moreover, by having these contributed problems we may have a better understanding of the main problem so policy maker can create a policy to overcome the contributed problems and gain the success rate of solving the main problem.

Figure 4. Solving Problem Concept



Top Influencer also has a significant role in this problem solving process. By knowing who are the top influencers, policy maker could use it to advertise and gain civic engagement to be together overcome the problems. Civic participation has a huge impact on the success of government policy implementation. Next, we are going to describe and share our experience of these three problems.

IV. Analysis

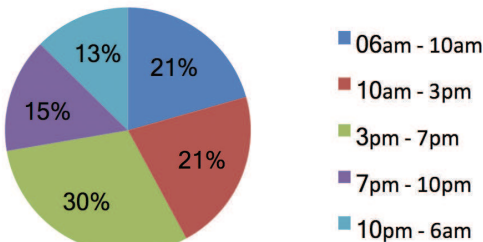
4.1 Traffic Jam

Firstly, in traffic jam case problem, we use keyword “macet” to represent it in Indonesian language and start the capturing process on Google Spreadsheet. Then for the final result, we achieved 11,248 tweets which posted from November 6, 2016 at 12pm to November 10, 2016 at 10 pm (West Indonesian Region Time).

Figure 5. Output Dataset

Date	Screen Name	Full Name	Avatar	Text	Date	Time	Age	Gender	Location	Category	Verified	Retweet	Reply	Like	Profile Image	Image
11/06/2016 11:04:28	@yulianita	Yulianita		macet di cikampeun?? dikira macetnya aje??	11/06/2016	11:04	25	F	Bandung	Travel	0	0	0	0	0	0
11/06/2016 11:04:49	@DianDian12	Dian Dian		macet di cikampeun?? dikira macetnya aje??	11/06/2016	11:04	25	F	Bandung	Travel	0	0	0	0	0	0
11/06/2016 11:05:21	@mamiandika	Mami Andika		macet di cikampeun?? dikira macetnya aje??	11/06/2016	11:05	25	F	Bandung	Travel	0	0	0	0	0	0

Figure 6. Time Section



The next process is processing the dataset. By this stage, we added time section and location label to the dataset. We divide the time into 5 sections which are morning commute time (6pm - 10am), working time (10am - 3pm), return time (3pm - 7pm), after work time (7pm - 10pm) and night time (10pm - 6am). From figure 4, we find out that after working hours become the most active tweet time by 30% of tweets compare to night time that only 13% of tweets. The interesting fact is, during the morning commute time and working time, twitter users still experience the traffic jams on their trip.

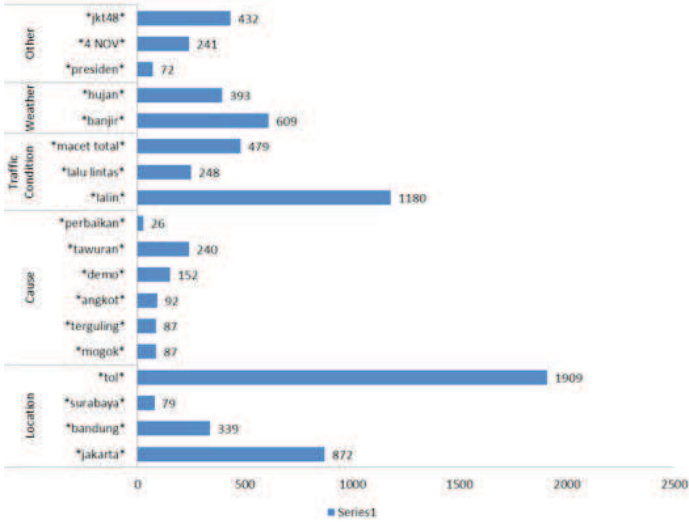
Figure 7. Geo Information



From figure 5, we found the facts that most of tweets come from 3 major cities in Indonesia which are Jakarta (Capital city of Indonesia), Bandung (Capital city of West Java Province), and Surabaya (Capital city of East Java Province). Moreover, location field in the Dataset also cannot be whole used to filter the data based on the location because the twitter user sometimes does not state the real position of them. Then most of the google maps field, also return blank because most of the users turn the GPS location off by default. So if you want to search down the topic based on a certain location, you need to do in the very beginning process which is set up the Twitter Achiever parameter.

Next, we enter the result formulating result process. At this point, we select sampling data and read several tweets as well as gather some correlated keywords. Then, we are grouping of those correlated keywords and count the appearance of those keywords on the whole dataset.

Figure 8. Correlated Keywords



We found 5 groups of topics based on that correlated keywords, which are Location, Cause, Traffic Condition, Weather and Other. Next we analyze and test this correlated keywords using previous research and facts.

From location group, reveal facts that it is consistent with Ministry of Transportation research at 2014. This research stated that the highest traffic cities are Bogor, Jakarta, Bandung and Surabaya. Bogor area is a part of the greater Jakarta region so the number of tweets consistent with this research. Still in this group, the "tol" word which means highway/toll road fulfills the tweet conversation by nearly 20% of tweets. So from those facts, we can suggest that government need take an action by having main road and toll road improvement and advancement especially in this three location.

Furthermore, in the cause group, we found "perbaikan" (means repair process), "tawuran" (means peoples riot), "demo" (means protest/demonstration), "angkot" (means small bus transportation, "terguling" (means overthrowing vehicle/traffic accident) and "mogok" (means break down vehicle). Even though the total number of those keywords mentioned is low, we can understand the cause of a traffic jam so we can prepare the right action to eliminate the effect. From that fact, we can suggest that toll road operator and transportation department should be ready and prepare a procedure for "perbaikan", "mogok" and "terguling" situation because it is a part of road service.

Moreover, in terms of small bus transportation, Jakarta and other cities in Indonesia still using a small bus transportation, which have a low capacity so it is not compatible and efficient for travel within the Greater Jakarta Region so people change their movement style using private car and motorcycle. Again still from previous research, shows that road capacity as has reached a peak of their maximum capacity. So from that circumstance, we suggest to the government to change these "angkot" and develop a higher capacity transportation as well as perform a reliable transit transportation.

The last correlated keywords in cause group are "tawuran". This is a kind of social problem happened in the big city in the developing country, when a group of people acting like criminals and attack another group of people in the middle of the road. Especially in Jakarta when all Indonesian ethnic migrate and try to find a better life, they can easily trigger a conflict just because of small things that happen in the road. We suspect that this problem is coming from social problem in the big city so from this research, we are not only found the correlated keywords, but also another main problem in the big city so solving this social problem would effect to traffic problem too. We suggest that the government could have a better solution for this social problem, like creates a new job opportunity, intensify police patrol and approach the religious leader in the areas.

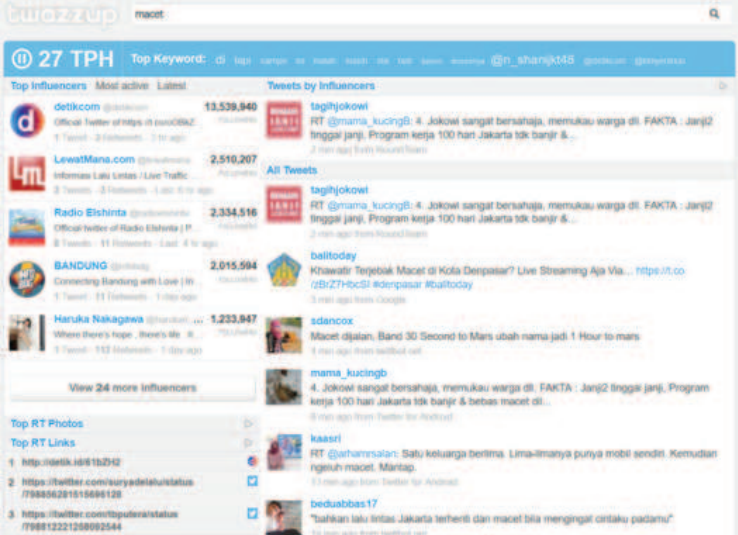
The traffic condition group is a group sentiment which indicates people report traffic condition. "lalu lintas" and "lalin" words has a same meaning which is traffic. Then "macet total" words mean indicating a traffic is jammed or stuck with no or less movement. In this issue, right now Indonesian government agency and other traffic information service already use this Twitter platform to report the traffic condition combined with placing CCTV along the main route.

The next group is weather. Located on the equator, Indonesia has 2 main season, which are dry season and wet season. Especially now, when wet season, "banjir" (means water flood) and "hujan" (means rain) words become a popular keywords in the tweets. This condition happens because a poor condition of drainage which cannot handle a high level of rain water. We suggest in the dry season, government can do normalization and improvement in the drainage especially along the main road. Then distribute the weather warning on social media and other ways so people will be ready and well prepared.

The last group is other groups. In our analysis process we find out these three words which are “presiden” (mean president), “4 NOV” and “JKT48”. Regarding president, we find out from the tweets, that president convoy may affect the traffic, but we think that is okay because he is a president and need to be respected on the road. For “4 NOV” word, this is an abbreviation for an event of recent protest regarding religious issues. Lastly, “JKT48” is a group of girl band idol who are famous in the youth segment. Surprisingly, the youth really admires with them so while they tweet regarding traffic, they used to mention their idol too.

We also mentioned top influencers in the last stage. Using Twazzup app, it would help us locate the top influencers. As we described before that this account can help government to share, address and advertise their approach in the right channel so it would have a greater impact to their follower.

Figure 9. Top Influencers



From figure 9, we find out that the top influencers in the traffic jam problem are detikcom, lewatmana.com, radio elshinta, infobandung and Haruka Nakagawa. The first four of the top influencers are the news media, traffic info provider, news radio and city information provider. However the last top influencer is still coming from a member of JKT48 group. So beside the media and information provider, government also needs to consider a popular artist as their ambassador to solve their problem.

4.2 Public Service & Driving License

Figure 10. Tweets Category for Public Service

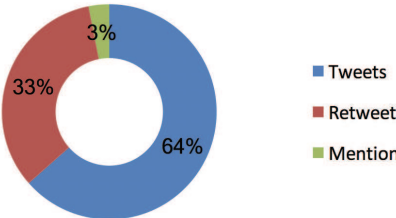
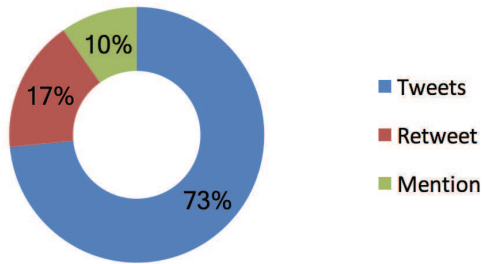


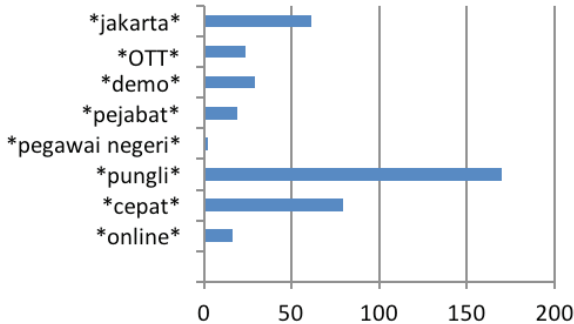
Figure 11. Tweets Category for Driving License



From both above figures, even though, the tweets are dominating the conversation, it does not mean that most of them are commenting and representing the actual condition of both keywords. When we did further investigation, most of tweets just share the other source of news.

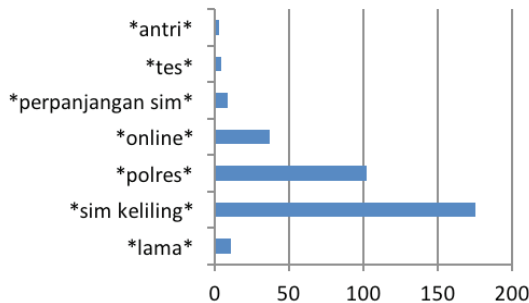
Furthermore, for labeling process also both keywords cannot represent the actual time and situation of the problems. However for the analyzing the dataset using correspondent keywords, we still able to figure out the most significant keywords that related to the main problems.

Figure 12. Correspondent Keywords for Public Service



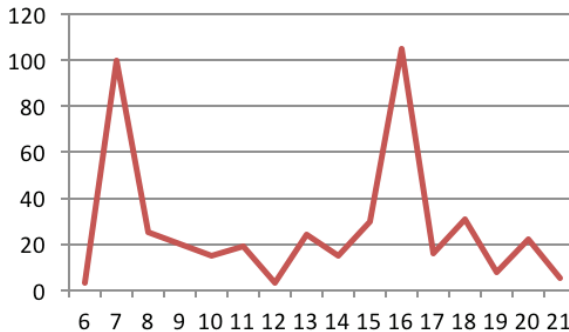
The top three of the corresponding keywords is “pungli” (means gratification), “cepat” (means fast response) and “Jakarta” (is a capital city of Indonesia). For public service problem, we believe that citizen demands the service without any corruption activities on it, fast response and Jakarta as a capital city can be a role model for the other local government.

Figure 13. Correspondent Keywords for Driver Licensing



The top three of the corresponding keywords is “sim keliling” (means mobile service), “polres” (means city/municipality level police and “online” (means online service). For public service problem, we believe that citizen demands more online info on mobile service. Especially in Indonesia, when driver licensing service is prepared by local police, people the excellent service from this local institution. Other anomalies from this data represents below.

Figure 14. Number of Tweets per day



The number of the tweets on date 6 and 16 was really high compared to other days. In the beginning, we thought that this is because of the days itself, but when we dug further, this anomaly triggered by tweets from local police regarding the information of mobile service on that day. So we believe that this data support our previous thoughts that mobile service is the most demanded info for the citizen.

Finally, from both problems, our frameworks still work for finding out the corresponding keywords. However, for unpopular topics like this, we cannot have a real comment from the twitter users compare to the first problem. So based on the number of outputs from searching process, still anybody can use this framework to have a further information regarding a main problem.

V. Conclusion And Future Research

Having this research really open our mind regarding how deep big data from social media can be used for government as an input for the public policy plan and evaluation. In this paper we have shown the simple and instant, but effective process of archiving tweets regarding an issue that want to be solved. We also create a framework to analyze it, find the correlated problem and using previous knowledge and research, we can overcome the problem in holistic ways. Also, we can easily locate the top influencers who can be a leverage to gain the civic engagement and participation in solving problem together.

Furthermore, again, this method cannot necessarily judge whether government program is failing or success, but this social media can be used as a tool to gain the success level. Learning from private entities which already used this big data analytics on social media, government also has an opportunity to exploit this approach for a better service and greater success in the future.

Hence, with this work, future publications and research may escalate this approach and analysis using proper sentiment analysis as an application of predictive analysis for social media and creates a social media dashboard which focus on government necessity. Especially for Indonesian language sentiment, there is still rare to find it.

References

- _____. (2016) About twitter, inc. | about. Accessed November 16, 2016. [Online]. Available: <https://about.twitter.com/company>
- Achrekar, Harshvardhan, A. Gandhe, R. Lazarus, S.-H. Yu, and B. Liu, "Predicting flu trends using twitter data," in Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on. IEEE, 2011, pp. 702-707.
- Agarwal, Amit. (2015) *How to Save Tweets for any Twitter Hashtag in a Google Sheet*. Accessed November 16, 2016. [Online]. Available: <http://www.labnol.org/internet/save-twitter-hashtag-tweets/6505/>
- Aramaki, Eiji, S. Maskawa, and M. Morita, "Twitter catches the flu: detecting influenza epidemics using twitter," in Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2011, pp. 1568-1576.
- Chatfield, Akemi, H. Scholl, and U. Brajawidagda, "#sandy tweets: Citizens' co production of time-critical information during an unfolding catastrophe," in System Sciences (HICSS), 2014 47th Hawaii International Conference on, Jan 2014, pp. 1947-1957.
- Lovejoy, Kristen, R. D. Waters, and G. D. Saxton, "Engaging stakeholders through twitter: How nonprofit organizations are getting more out of 140 characters or less," *Public Relations Review*, vol. 38, no. 2, pp. 313-318, 2012.
- MacEachren, Alan M., A. Jaiswal, A. C. Robinson, S. Pezanowski, A. Savelyev, P. Mitra, X. Zhang, and J. Blanford, "Senseplace2: Geotwitter analytics support for situational awareness," in Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on. IEEE, 2011, pp. 181-190.
- Rybalko, Svetlana and T. Seltzer, "Dialogic communication in 140 characters or less: How fortune 500 companies engage stakeholders using twitter," *Public Relations Review*, vol. 36, no. 4, pp. 336 - 341, 2010. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0363811110000792>
- Sakaki, Takeshi, M. Okazaki, and Y. Matsuo, "Tweet analysis for realtime event detection and earthquake reporting system development," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 25, no. 4, pp. 919-931, April 2013.