Big Data Analysis on Social Media Messages as Public Discourse

Sanhari Prawiradiredja[1] Budi Santoso[2]
[1]Dr. Soetomo University, Surabaya, Indonesia
[2]Dr. Soetomo University, Surabaya, Indonesia
Email: [1]sanhari.prawiradiredja@unitomo.ac.id [2]santosobudi.id@gmail.com

Abstract

We know that expression or opinion through the media is a public discourse. This public discourse is massive along with the development of new media through the internet. With the convenience associated with technological developments, this public opinion has become widespread and has a digital footprint. This phenomenon is now called big data, People comment anything via Twitter, YouTube, Facebook, Instagram, Telegram, WhatsApp.People argue about the ongoing political process or issue. As consumers, they comment on the products they use. Of course, the interests related to social and cultural affiliation cannot be separated from the public's attention.

With the ability of artificial intelligence to recognize words, data mining can be done in this case for research purposes in terms of data analysis. Likewise with natural language process, the corpus as a source of data analysis can be defined and described, the taxonomy of words can then be determined. While the 'N-gram model' is used to predict the use of words in a sentence. This paper will discuss the possibility of data analysis to be used in reading the phenomenon of massive messages. That one can relate to public opinion on policy or other political issues. For marketing communications related to opinions about products, production processes or institutional image. Of course, many cultural issues are commented on in public discourse. And this is very interesting because researchers can better understand how the issue is described in public discourse. Because public discourse is basically the use of language to express anything in the public sphere. We try to discuss the possibilities in analyzing the phenomena of this massive message.

Keywords *big data analysis; public discourse; social media; artificial intelligence*

A. Introduction

The social behavior of the Indonesian people has changed, especially in terms of choosing the media to voice opinions and address a problem. Currently, some people have shifted towards online social media on the internet compared to other conventional forms of media which tend to contain one-way information. Millions of audiences, whether personal or institutional, have used social media as a form of existence in cyberspace. They use social media to establish social relationships, share information and events (Sivarajah, 2019).

This is supported by massive developments in terms of the availability of internet network infrastructure, mobile technology, and adequate public purchasing power. Whereas previously the channels for voicing opinions were dominated by limited print and electronic media, nowadays there are various forms of online social media available and free to use. While online media has the advantage of being fast and complete, online media has a weakness in terms of news accuracy. (WAN-IFRA, 2009).

This includes every policy of the Government of the Republic of Indonesia, will always be escorted by control through public voice responses, one of which is through this social media. For example, the problem that is currently happening is the outbreak of the Covid-19 virus pandemic in almost all of the world and also Indonesia. The COVID-19 pandemic or the corona virus that spread began at the end of 2019 in Wuhan, China. In Indonesia, the first case of the Covid-19 virus was officially announced on March 1, 2020 (Covid-19 Task Force, 2020).

Although previously social media users, especially Twitter, had been busy discussing the possibility that this virus had entered Indonesia, this was denied by statements by several Indonesian Government officials. Currently (as of September 11, 2020) the number of positive cases of Covid-19 in Indonesia has reached 207,203 cases, with 8,456 cases of deaths, and 147,510 people declared cured (DetikHealth, 2020).

What's interesting is that not all social media users simply believe the official statements of government officials in this case of the Covid-19 pandemic. Currently, social media platforms are outside the control of the government, so that people do not encounter difficulties in voicing their aspirations. Meanwhile, the government must be careful and careful in conveying official information, because the data and facts are easily accessible to the public. So carelessness in delivering official information will bring criticism, even public blasphemy through online social media.

Of course, this is just an example, public discourse is not only

about disease and government policies. All things can be responded to by the public so that analysis on discourse will show tendency for patterns of opinion variations or conflicts that exist in it. In general, anything can be discussed by the public in discourse. Whether it's about a particular object, event or certain person or individual. This form of public discourse will have certain implications for institutions operating in the public sphere. Because the discourse conveyed by the public can be a basis for action for a person or institution if it is related to its public image. In this case, public discourse functions as public opinion that is described in an open public sphere. Public opinion about something is of course useful for institutions operating in the public sphere to understand it because it is very meaningful for policy planning that must be carried out next.

Public opinion is an input for institutions to respond. The input can be in the form of support, positive views, approval of institutional actions. But it could also be the opposite way, the public can also

reject, harass, give a negative view of what the institution is doing. This positive and negative images in public relations activities show the success of an institution's public communication performance. If there is a lot of positive discourse, it means that public communication institutions have succeeded in creating good cognition and affection among people. In this article, the author will not look from this side, how the influence of public communication on the image of the institution is reflected in public discourse or sentiment. But what the author sees is the opposite, namely how to analyze public discourse to design the best and relevant response to form good public relations actovities for the next step.

B. Result and Discussion

Discourse in the context of communication is actually the way people convey something. The message conveyed by someone always emphasizes something in a certain way. Therefore, discourse is not only a matter of language. Language is used by humans in daily

social interactions in various contexts of situations. In a critical approach, the use of language as a discourse is always considered as part of social action.

Two functions of language in social actifities are to scaffold the performance of social activities (whether play or work or both) and to scaffold human affiliation within cultures and social groups and institutions (Gee, 1999). A person performs social actions every day and that is reflected in the language he uses. He tries to fulfill his needs and he cannot do it alone. He must interact to be accepted as affiliation in his social group. He acts in a cultural context in social association. In this case, he must understand and develop an understanding of the symbols of his identity to be able to play a social role optimally.

In this perspective, the use of language is seen as a social act or in other words in the view of social semiotics. In Halliday's tradition of systematic - functional linguistics the structure of speech using language is seen in the context of the situation in the categories of (1) discourse field,

(2) meaning and (3) form or mode. In this case, language has a two-way domain. Language determines the context, but on the other hand language is also determined by the context (Ida, 2014).

In living their social life, individuals use language to participate in social interactions which are related to the distribution of social goods. Social goods are meant here is a social valuation related to the position and role of individuals. Social valuation in the communication process lies in social interaction discourses. In discourse analysis, researchers examine the context of the communication, such as who is communicating with who and why; in what types of audiences and situations; through what medium; how the different types and developments of communication; and relationship typology for each party. We understand how important the role of context in determining the meaning of an utterance in a discourse. When the context changes, the meaning also changes.

All discourses that appear in the form of texts, conversations, or

anything is seen as a form of power struggle. Concept of power is one of the key relationships between discourse and society. Relationship between power and discourse can be seen from what is called control. Control in a discourse can be in the form of control over context, and control over discourse structure. Control over context, for example, can be seen from who can or have to speak while the other position as a listener or who agree. While the control over the structure of discourse can be seen from someone who has more power can decide which part which should be displayed and which parts should not be and how it should be displayed.

Power is always present at every level of social relations. Discourse produced by social actors plays a role in interactions at the level of social relations. Power does not only unite social togetherness or uniformity. Discourse as a variety of expression of power also emphasizes the power for subordination in actions, processes and social relations. Discourse studies thus provide the possibility to see how individuals or

components in society voice or articulate messages as part of a social entity.

In Foucoult's model there are five stages in analyzing the realm of discursive events: (1) understanding statements according to truly unique events, (2) determining the conditions of their existence, (3) determining their boundaries, (4) correlating them with other possible statements. related to the statement (links to political, cultural, social, economic and other discourses) (5) shows another form of the statement conveyed (Lubis, 2014). In the public sphere, we now know that the existence of new media especially in social media supported by technological developments related to the internet and gadgets – has become a massive channel for citizens to express their opinions. Any aspect can be discussed and commented on by residents who are now known as netizens.

This becomes interesting for us to discuss, because it is also related to how the social process takes place in society. For the government, industry, social groups, for example, the public

voice is very important to assess their existence and institutional function. Institutions or organizations can assess whether what they have done is accepted by the community or vice versa, they get rejected. The social functions they perform then also show something significant or not. The comments and opinions of netizens are reflected in the discourses that appear on their accounts. Of course, the available data relating to the opinions of these netizens is large and very massive. Therefore, when this large public discourse will be analyzed, the analysis carried out is called big data analysis. The results of the analysis will be very useful for assessing the social function of an institution, both regarding products, processes or individuals who carry out functions within the institution. Of course, this generally shows the existence of social institutions.

This new information and communication technology did initially raise fears. This concern is associated with the nature of its virtuality which can reduce face-to-face encounters. as well as the anonymity inherent in this new

technology, so that individuals may have little or no responsibility for the discourse that they convey. But of course, life can't go backwards, technology is a tool created to make it easier for its users. The existence of new media is also the same, for users of course need to be equipped with an understanding of media literacy. For institutions, the existence of this new media provides new opportunities to more quickly understand their social role in society. Of course, by doing big data analysis to describe the discourses about their products, services, work and performance. The discourses were exposed through the communication of netizens via Twitter, Facebook, Instagram, Youtube, Telegram or WhatsApp.

C. BIG DATA ANALYSIS

Social media has become one of the dominant social interaction media today to get opinions and reactions from the audience through comments, following, likes/dislikes, ratings, and choosing favorites. One of the interesting things is the positive and negative comments section given by users as a reflection of the public's

mindset towards the information presented. This information is very useful in studying the treatment of users and the public on content material on social media.

Big data is a collection of processes consisting of large volumes of structured and unstructured data and is used to assist various activities. The definition of big data can also be divided into 3 parts, which are commonly referred to as 3V:

(1) Volume, the size of data owned by big data has a large capacity. (2) Velocity, data transfer speed is also very influential in the process of sending data effectively and stably. (3) Variety, the types of data variations owned by big data are more than using conventional database systems.

Big data analysis is divided into four stages, namely data collection, pre-processing, feature extraction, and classification. If needed, as an interaction with the reader added visualization of the results in graphic form.

a. Data Collection Stage.

In its implementation, the data collection proces applies several concepts about data mining. Text mining is also known as data mining because the same algorithm can be used for both mining concepts. The significant difference between the two is that data mining involves more structured data, while text mining deals with certain features and is relatively unstructured and requires preprocessing. According to (Yogapreethi, 2016), the main process of Text mining consists of:

• Information Extraction: a preliminary analysis of unstructured text by extracting information, including searching for relationships between phrases.

• Clustering: the process of separating text into one group and by giving weight to each word.

• Classification: the process of finding the main topic of a document by specifying metadata in the form of class labels.

• Information visualization: the process of displaying the results of the analysis of information patterns, used for user interaction with documents.

At the data collection stage, the process of collecting text data from streams of tweets sent by the public as a reaction to certain issues related to the public interest is carried out. In order to obtain a collection of text data that is in accordance with what is needed in this study, several keywords that represent the process of filtering text data streams are set. These keywords will be combined as phrases and used to limit the research topics taken in this research. The data collection process is carried out through a tool in the form of a computer program that is connected to the internet and uses the Python programming language.

b. The Pre-processing Stage

The pre-processing stage or initial handling is needed to ensure that the text data generated from the previous process has met the requirements for further processing. Generally, the data sent by the account owner is in the form of abbreviations, or other characters that complicate the analysis process. Therefore, the word abbreviations and characters are omitted without affecting the weighting of the response to the sentence as a whole.

c. The feature extraction stage

The feature extraction stage is the process of searching for a collection of words from tweet data that has passed pre-processing. In other words, the text data entering this stage is clean of unnecessary abbreviations or characters so that word-weighting analysis can be carried out.

d. Classification Stage

This stage is the determination stage where the weighting results will be calculated and grouped into positive, neutral, or negative categories.

THE BENEFITS OF BIG DATA ANALYSIS

Big data has several important functions in the process of developing and perfecting an application. The following are some of the functions related to big data:

1. Can determine the cause of a problem, failure in real time.

2. Making a smart and right decision.

3.Reduce costs, time, and improve activity performance.

From these functions, it appears that big data analysis is especially useful for seeing the current reality of an institution in the eyes of the public. The public is a group of people who have an interest. And this public interest is related to the existence of the institution in carrying out its social functions.

We will look at some of the uses or functions of big data analysis in some institutional contexts. The public's reaction to institutional elements can be seen from the response of individual account owners. This collection of opinions or responses through big data research can be abstracted into various typologies that can be interpreted for institutional purposes. This institutional interest can be related to several aspects. In general, there are three institutional elements that can be judged from the public's response. The three are products or services from institutions, individuals who play a role in institutional activities and the third is operational processes that exist within the institution itself.

For products or services, we can see the government as an example. The product of government is policy. From political issues related to government policies, we can understand the emerging public sentiment. In the current era of the Covid-19 pandemic, there are many issues related to the handling of the government to overcome its impact. We know that some public responses are trending topics. Let's say, for example, related to vaccination, PCR and economic aspects related to the handling of covid. Based on tests and analyzes conducted on public sentiment data through Twitter social media on several topics related to the Covid-19 pandemic, it can be concluded that: the public has a greater tendency to convey neutral and positive sentiments in responding to the Covid-19 pandemic event and its handling (Prawiradiredja and Santoso, 2021).

Although the scope of the research is only among Twitter users. This research is quite interesting because in addition to providing an overview of the public's response. This study also found that a number

of people through their accounts expressed negative sentiments through harsher criticism than positive sentiments. This is indicated by the higher average negative sentiment score, although the number is less than the positive sentiment.

This is just an example of big data analysis research on an issue raised from a particular social media, namely Twitter. We know that the variety of products and services varies and the number is very large depending on the type of institution, be it government, economic or social institutions that are cultural in general. As an illustrative example, the recent series of accidents on toll roads has brought a lot of public responses related to toll road management agencies, police agencies, as well as comments related to public figures. The quality of toll roads, driving safety controls, aspects of vehicle security equipment are the issues raised from these recent accidents. Relevant institutions - based on this public response - can immediately anticipate it by making relevant decisions or policies. Again, this is just an example because in

reality the variety of institutional products or services that are responded to by the public is very diverse from commercial products for daily needs, lifestyle, industrial goods, services from economic institutions such as banking, religious communal activities, special interest groups and many others.

The public's response to the use of social media is also related to the personal aspects involved in institutions. For example, in handling Covid, names that stand out and are widely commented on include Minister Luhut Pangaribuan, Erick Tohir and the Chair of the Covid Handling Task Force, of course there are other names that are commented on by the public. Moreover, if the regional level from the national level is lowered to the regionality below it, the names mentioned will increase. And we can easily predict that if the issue involves other institutional variations outside of governance, more and more names will appear. These personal aspects are important not only because of their unique individuality but also because they are representatives of the institution.

Therefore, once again, institutions need to respond to this to demonstrate their capability in assessing and responding to situations that are relevant to the role and development of their institutions.

The third aspect of public response that deserves to be analyzed by big data is the aspect of the institutional process. This time we take the example of a commercial product. An example is the traditional herbal medicine product which shows that the herbal product process is carried out using advanced technology products and with high hygienic standards. This activity is certainly posted to get a positive response from the public through the use of social media accounts.

What the herbal medicine industry will do next, of course, is to observe whether the creative content they create is seen and understood by the public? Second, how is the public's response to the content whether positive, neutral or even negative. The third, of course, is to provide relevant and comprehensive institutional communicative action for the public response. Once again,

this is just an example, if the institutional process is changed, variations in terms of commercial, political, cultural or communal institutions are seen, the cases of institutional processes or trending topics will increase more and more.

What needs to be noted is that big data analysis needs to be developed and complemented by more qualitative research to show the depth level of the message so that institutional responses will be more effective and efficient. This is increasingly important because digital discourse is related to social change and multiculturalism in addition to more 'younger' styles of behavior and social appearance because social media users are mostly young people (Thurlow and Mroczek, 2019). This critical study of the media is also important because it shows an emancipatory effort in the praxis of communication in the sense of taking into account the power aspect in the balance of power between communication participants at the personal and institutional levels (Sugiharto, 1996) (Hardt, 1992). Institutions that are sensitive to

changes in opinion mediation through social media will use big data analysis to make themselves functional and relevant to their existence in public view. Because the implementation of social functions in accordance with deontological or teleological ethics is the underlying reason for the existential institutions.

References

DetikHealth, F. A. (2020). "*Update Corona Indonesia 10 September: Rekor 3.861 Kasus Baru, Total 207.203*". Retrieved from DetikHealth: https://health.detik.com/berita-detikhealth/d-5167585/update-corona-indonesia-10-september-rekor-3861-kasus-baru-total-207203

Gee, James Paul, (1999), *Introduction to Discourse Analysis Theory and Method*, London: Routledge

Ida, Rachmah, (2018), *Metode Penelitian Studi Media dan Kajian Budaya*, Jakarta: Prenada Media Grup

Hardt, Hanno, (1992), *Critical Communication Studies, Sebuah Pengantar Komprehensif Sejarah Perjumpaan Tradisi Kritis Eropa dan Pragmatis Amerika*, Yogyakarta: Jalasutra

Lubis, Akhyar Yusuf, (2014), *Postmodernisme Teori dan Metode*, Jakarta: PT. RajaGrafindo Persada

Prawiradiredja, Sanhari and Budi Santoso, (2021) "*Analisis Sentimen Publik Kebijakan Pemerintah dalam Menangani Pandemi Covid-19 di Indonesia*"**,** paper pada Konferensi Nasional Ilmu Komunikasi, Fikom Universitas Ciputra

Sivarajah, K., U. I. (2019). "*Role of big data and social media analytics for business to business sustainability: A participatory web context*". *Ind. Mark. Manag.*

Sugiharto, Bambang I., (1996), *Postmodernisme Tantangan Bagi Filsafat*, Yogyakarta: Penerbit Kanisius

Thurlow, Crispin and Kristine Mroczek, (2019), *Wacana Digital Digital Discourse*, Jakarta: Prenada Media Grup

Yogapreethi, N. (2016). "A Review on Text Mining in Data Mining". *Int. Journal Soft Computing*