Personality of Politicians as the Object of Public Assessment

Sigit Apriyanto, Adelina Anum
{sigitteduh89@gmail.com[1], Adelina Anum[2]}

Universiti Tun Hussein Onn Malaysia, Johor Baharu, Malaysia[1], University of Sang Bumi Ruwa Jurai, Bandar Lampung[2]

Abstract. The information contained in the social media can be processed and used to predict the user's personality. Personality information can be used as a basis for public assessment of the leaders in a government. In obtaining it is required the existence of a system of personality prediction to bridge between social media with a person's personality. The feature is obtained from the extraction method Linguistic Inquiry and Word Count (LIWC) is a text analysis program that calculates a word in psychological category into percentage value. The purpose of this study is to make a personality prediction system based on information from social media Twitter. Twitter's social media was chosen because of its text-shaped data, so it fits in with this paper using the word analysis program. The outcomes of this study show that 47% of 13 politician's parties tend to have depressed meaning, lack of confidence and disappointed that influence to their emotion. It was also found that party's communication system generally allowed their politicians to give statements freely to the social media, which is 62% of politicians stick out negative emotion on public. These factors were reducing the public trust as citizenry's representatives.

Keywords: Personality, Social language, Positive and Negative Emotion, Public Trust, LIWC

1 Introduction

Personality is one of the known methods in the world of psychology to interpret the personality of a person, especially to find the relationship between personalities and job's performance. Personality can be seen from various aspects they are openness, extraversion and others personality. Personality measurements can be performed by using shared methods. One of the methods is a questionnaire question that contains the characteristics of the scale-from the most agree to the most disagree. But this method has a drawback because the prediction result of the questionnaire is less valid [1].

This case will be made by a model of personality classification based on social media status that is Twitter with Linguistic Inquiry Word Count (LIWC). By looking at 3 parameters to select the personality those are positive emotion, negative emotion and social word. The Linguistic Inquiry and Word Count (LIWC) method is a transparent text analysis program that computes

words in a meaningful psychological category or word analysis method that converts the status into values. The LIWC results show its ability to detect meanings in a variety of experimental settings, including to show the focus of attention, emotionality, social relationships, thinking styles, and individual differences [2].

Social networking on the website has grown rapidly in the last decade. In January 2005, a survey of social networking websites has reached 115 million members and continues to grow to date [3]. When creating social media profiles, users tell themselves about self-description, status updates, photos and interests, so that personality can be observed through social media. Twitter is one of the social media that was used in many countries to express feelings, activities and written in one or two sentences. As multi-leveled equations, graphics, and tables are not prescribed, although the various table text styles are provided. The formatter will need to create these components, incorporating the applicable criteria that follow.
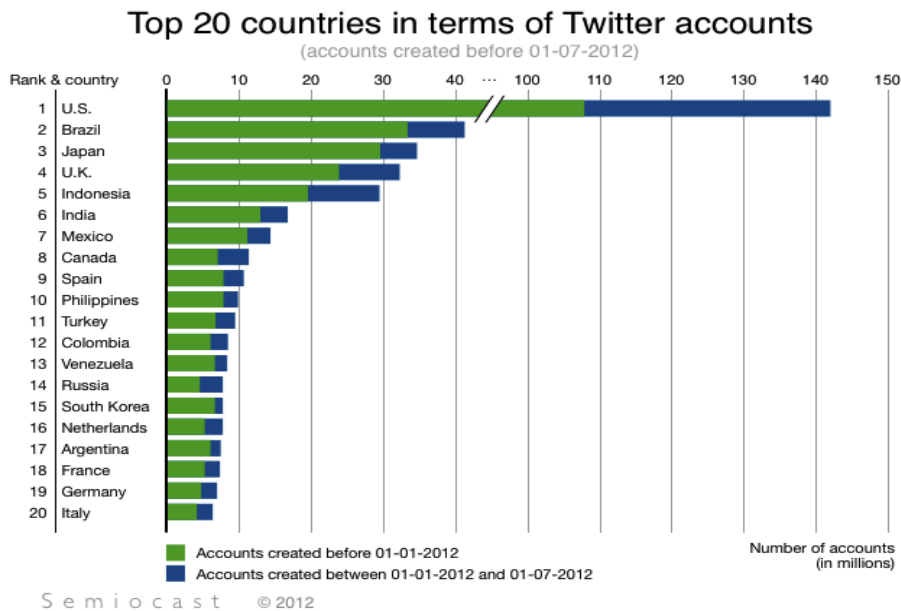

Fig. 1. Top 20 countries in terms of Twitter accounts

Based on fig. 1 above, it can be seen that the Twitter users in US, 2012 has reached 140 million users and continues to grow to date. Seeing the number of Twitter users and its habits, it can be used as a reference in predicting the personality. There is a relationship between human nature and their habit towards psychological disorders, performance in work, even in relation to the partner [4].

Therefore, making the system to predict the personality of a person through social media is considered necessary as a bridge between social media and personality paper. If a person's personality can be accurately predicted through this system, then the system will help many things, such as determining the level of success and compatibility of a relationship with the partner [5], determining the type of music heard by personality [6], and individual personality types can be used to predict whether they prefer McCain or Obama as a US presidential candidate in 2008[7].

Based on the previous description, it can be seen that the personality is very influential to some aspects of life, so the system of personality prediction is considered necessary. Where by

using this prediction system, the personality of a person can be known by the users quickly and accurately. In addition, this prediction system can minimize the error of personality prediction due to the subjective analysis of the testers.

Software-based personality prediction systems have been developed by several researchers until now and have been applied to some aspects of life and social media (such as Facebook, Twitter, and Instagram). The purpose of this paper is to know the accuracy of personality classification based on twitter tweets by using word analysis from LIWC which calculation accuracy level expressed into percentage.

2 Literature Review

2.1 Linguistic Inquiry and Word Count (LIWC)

To find a more efficient evaluation method due to make an accurate evidence from the suspects or witnesses, it turned to the promise of computerized text analysis programs to assess the resources text. LIWC (Linguistic Inquiry and Word Count, often pronounced 'Luke') is a lexical resource developed by social psychologist James Pennebaker and his team at the University of Texas [8].

The LIWC program has two central features; the processing component and the dictionaries. The processing feature is the program itself, which opens a series of text files which can be essays, poems, blogs, novels, and so on and then goes through each file word by word. Each word in a given text file is compared with the dictionary file [9]. Shortly, LIWC is a transparent text analysis program that counts words in psychologically meaningful categories.

Linguistic Inquiry and Word Count (LIWC) is a word counting software widely used for quantitative text analysis in social, linguistics, and medical sciences. Although LIWC is able to measure features in text that allow text classification and prediction for behavioral outcomes, it is mainly used to identify informative word features about the psychological state of the author or the speaker or the underlying group. LIWC was originally developed to address content analytics issues in experimental psychology [10].

2.2 How Linguistic Inquiry and Word Count (LIWC) works

The use of Linguistic Inquiry and Word Count (LIWC) is quite simple. Basically, it reads certain texts and calculates the words into percentage that reflect emotions, thinking styles, social problems, and the different parts of the conversation. Because LIWC is developed by researchers who have an interest in social, medical, health, and cognitive psychology, the language category was created to capture the social and psychological state of society [9].

2.3 Twitter

As many people know that Twitter is one of the most popular social media in Indonesia, not only in Indonesia but also in the world. Twitter popularity is already worldwide. Twitter usefulness is needed for the users. The people use Twitter to keep in touch with friends, get the latest news, motivation, as business media, and find out more about anything. Twitter has a number of features that can interact with the users.

Personalized predictions on Twitter, generally using LIWC features. The dataset used in the research is data obtained from the Twitter public timeline[11], as many as 13 public figures in the spotlight of trending topics. This dataset is then called "personal-tweets" and contains about 130 posts. The extraction of the features are used to process the dataset, list of feature as follows:

1. Username/Account
2. Post
3. Date
4. Platform from which the user posted
5. Personality model
6. Accuracy of the personality model
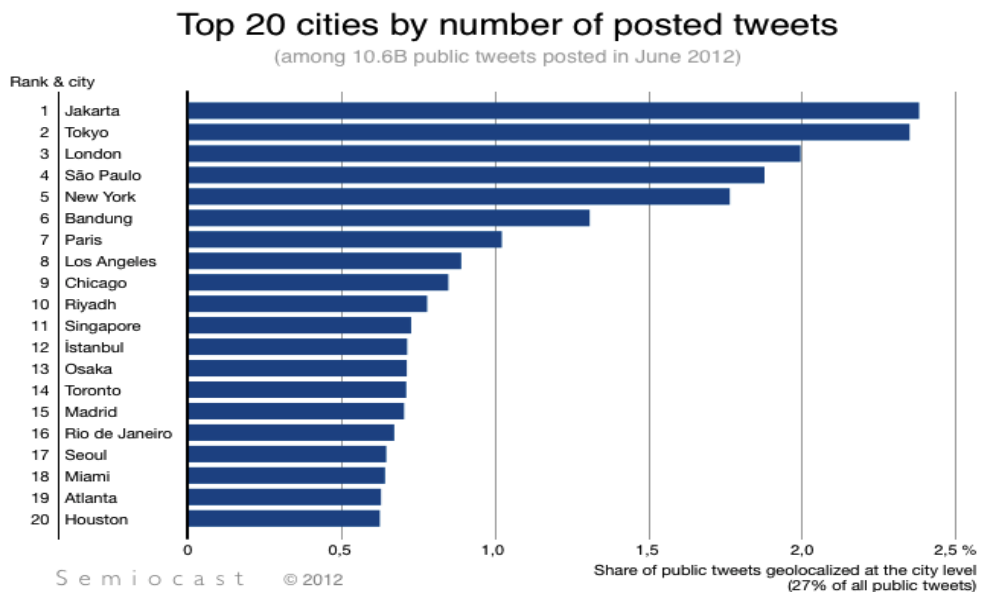7. Validity of personality model
8. Language

Top 20 cities by number of posted tweets
(among 10.6B public tweets posted in June 2012)

Rank & city	
1	Jakarta
2	Tokyo
3	London
4	São Paulo
5	New York
6	Bandung
7	Paris
8	Los Angeles
9	Chicago
10	Riyadh
11	Singapore
12	İstanbul
13	Osaka
14	Toronto
15	Madrid
16	Rio de Janeiro
17	Seoul
18	Miami
19	Atlanta
20	Houston

Semiocast © 2012

Share of public tweets geolocalized at the city level
(27% of all public tweets)

Fig. 2. Top 20 cities by number posted tweets

Based on fig. 2 above, it shows that the users from Jakarta were the most active within all cities ranked by the number of posted tweets. Using Semiocast's technology, 27% of all public tweets were geo-localized at the city level. More than 2% of all these tweets were posted from the capital of Indonesia, witnessing the activity of Twitter in the country. Bandung, its second largest metropolitan area, ranks 6th.

3 Methodology

Here is a flowchart of the system design in classifying personality based on the information on social media accounts such as Twitter. The purpose of this classification is to make it easier to identify, compare and study. Comparing means looking for similarities and differences of the things characteristics.
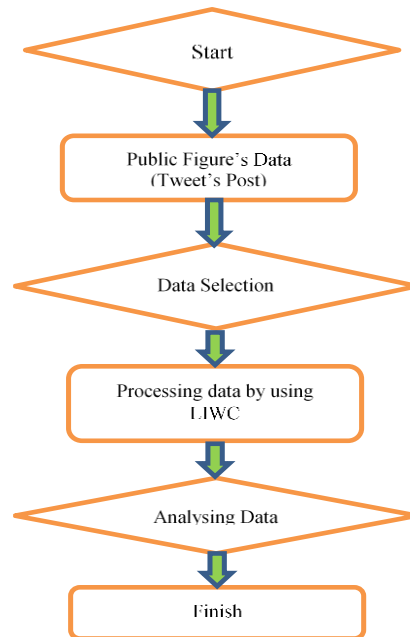

Fig. 3. Flowchart Flow Design System

Based on fig. 3, it shows the flowchart of the system created. Which has several processes. Where the first process is to prepare the Tweet post data of the three accounts. Then the data must be analyzed through the selection process. Where the data should be selected to be translated into English and eliminate the abbreviation of Tweet, and the content of the status itself will be used in testing. After completing the data selection, then the data will be input into the system; LIWC to be counted. The output data (in percentage form) will be processed and analyzed by using Pennebaker theories[12 13 14 15]. In the final stages of the system will be obtained in the form of a written text that shows the meaning of each feature (social, positive, and negative emotion). The results will be used as a basis for assessing the public confidence towards politicians.

The data collected are based on the Tweets from 13 politicians Twitter accounts and from 13 political parties in Indonesia. In this paper only use 1 political figures who are being trending topic and represent his/him political party. The number of Tweets Post consist of maximum 10 Tweets from each account. So, there are 130 data Tweets from the 13 political figures.
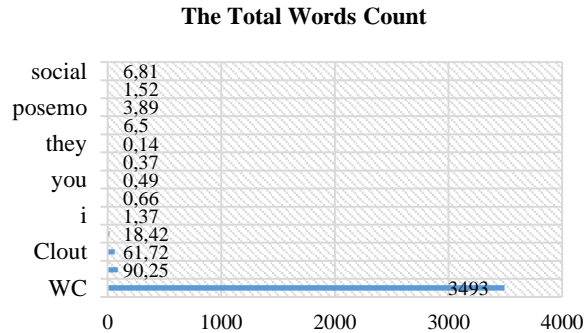
The Total Words Count

Fig. 4. Words Count Data

Fig. 4, presents words count that most represented from the total accounts (13 accounts). A basic data (130 Tweets from 13 accounts) has been counted and was pre-processing, in which the extracted text is cleaned by removing punctuations, numbers, common words, etc. and focused on three features keys; self-references, emotion, and social words in order to increase the accuracy and efficiency of the results. Even though any appear words from others features generally reflect what might be shown from the three features.

4 Result and Analyse

Based on the results of the tests that have been done, it was obtained a percentage of each feature and Tweets from each account. The percentage value is derived from the calculation of LIWC2015 words count program.

Table 1. Summary of Self-References - 13 Politicians Accounts in Indonesia

Pronouns	Mr. A	Ms. Y	Mr. B	Mr. C	Mr. D	Mr. E	Mr. F	Mr. G	Mr. H	Mr. I	Mr. J	Mr. K	Mr. L
I	0.58	1.02	0.72	1.16	0	2.15	0	3.73	0	0.41	0	2.49	4.4
You	0.29	0.77	1.43	0.58	0	1.08	0	0.47	0	0	0.85	0	0.55
They	0.29	0	0.36	0	0	0	0	0.47	0	0.41	0	0	0
We	0.29	0.26	0.36	0.58	0.37	0.54	4.95	0.7	0	0.81	0.85	0.83	1.1
She/He	0	0.51	0.36	0	0	2.15	0	1.17	0	0	0	0	0
Article	7.54	7.91	5.02	5.23	6.72	4.84	4.95	8.16	2.17	9.35	5.08	5.81	4.95

Based on table 1 shows that the use of personal pronouns in making sentences has different meanings and tendencies. [8 14 16] There were 6 politicians (47%) who use "I, me, my" repeatedly and tend to have depressed meanings, lack of confidence, disappointed, and feelings of approbation. This is so contrary to the attitude they show in public. On the other hand, there were only 2 politicians (15%) who use "they, she, he" repeatedly and tend to have "good" meaning in managerial. Shortly, it has correlation with their leadership in their institution (DPR). In contrast of using "we", when a person is more likely to use the word "we" and tend to avoid in using first person then that person is indicated "lie". In this case there were 3 politicians (23%) indicated "lie". Lying here is related to the content of the statements they post in public. However, it would be different when someone not only uses "we" but also uses "you

and me" then that person is indicated "powerful". In this case there were 2 politicians (≥1%) or 15% of total politicians who have positive power.
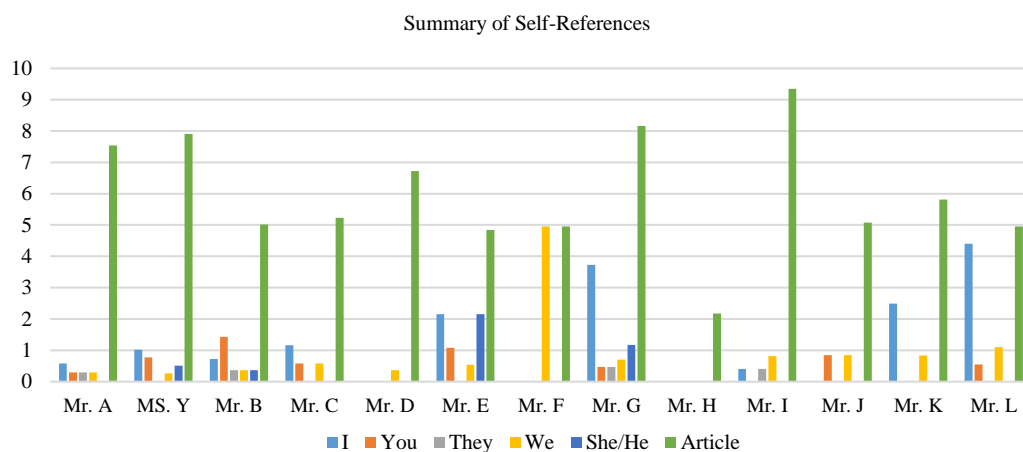
Summary of Self-References

Fig. 5. Summary of Self-References

Fig. 5, shows that each politician of each party has a tendency to use any articles. It means that the meaning of the statements are very clear and shown to someone who is clear as well.

Table 2. Summary of Social-Emotion - 13 Politicians Accounts in Indonesia

No	1	2	3	4	5	6	7	8	9	10	11	12	13
Emotions	Mr. A	Ms. Y	Mr. B	Mr. C	Mr. D	Mr. E	Mr. F	Mr. G	Mr. H	Mr. I	Mr. J	Mr. K	Mr. L
Positive	3.19	4.34	1.08	6.4	1.49	3.23	0.99	2.1	4.35	2.03	4.24	8.71	7.14
Negative	0.87	2.55	2.15	1.74	1.49	1.08	0.99	1.86	0	0	3.81	1.24	0.55
Social Words	5.51	7.14	7.14	3.49	5.97	10.22	5.94	10.72	3.62	4.88	6.36	3.32	7.69

Table 2, shows the results of Tweets' calculations and analysis of 13 politicians (13 political parties) who passing the 2019 election verification. The results show that 8 politicians or 62% of the thirteen parties (above 1%) write statements and have negative meanings on each their posts. Moreover, 38% of the remaining parties tend to be normal (looking for safe spots). The negative content of the material is caused by several factors, including refuting others' judgment of a certain politician, self-defenses from the negative issues, and opposition party resistance. It can be concluded but it cannot be generalized that politicians in each party have bad personalities which are correlated to the quality of the party's human resources. It would be a consideration to the public appraisal from the citizenry.

On the other hand, there were 7 politicians who tend to write and make statements on positive meaning (more than 2% average of positive and negative value). Positive personality value was created to the goal of positive correlations on social image. The point is the politicians tend to make positive statements to build positive social relationship as well. Although between positive and negative emotions of all the party politicians have a common goal tendency in creating good relationship with the social society.
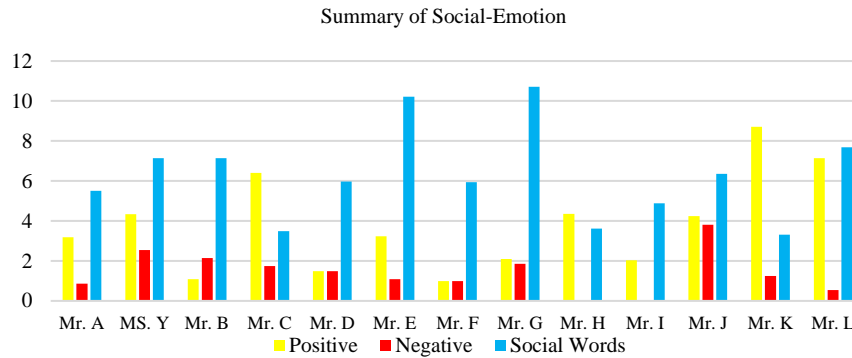
Summary of Social-Emotion

Fig. 6. Summary of Social-Emotion

Fig. 6, shows that each politicians of each party has emotional variations that ultimately focus on creating a positive relationship with the social communities.

5   Conclusion

The outcomes of this study show that 47% of 13 politician's parties tend to have depressed meaning, lack of confidence and disappointed that influence to their emotion.

It was also found that party's communication system generally allowed their politicians to give statements freely to the social media, which is 62% of politicians stick out negative emotion on public. These factors were reducing the public trust as citizenry's representatives. Based on the analysis and testing that has been done, the conclusions that can be drawn is overall for the dataset of this research experiment, the politician's personality can be predicted well by using linguistics features of psychology from LIWC program. In addition, it can be concluded also that linguistics and social features can recognize a person's personality well. Keep in mind that predicted results is depend on experiments performed so that other similar studies can produce different result according to the parameters used.

References

[1] Damanik AT, Kunci K. Prediksi Kepribadian Big 5 Pengguna Twitter dengan Support Vector Regression. 1999;(1981):14-22.

[2] Pennebaker JW, King LA. Language Use as an Individual Difference. J Pers Soc Psychol. 1999;77(6):1296-1312. doi:10.1037/0022-3514.77.6.1296

[3] Golbeck J a. Computing and applying trust in web-based social networks. Ann Phys (N Y). 2005;54(1):199. doi:10.1016/j.jpsychires.2008.02.003

[4] Saulsman LM, Page AC. The five-factor model and personality disorder empirical literature: A meta-analytic review. Clin Psychol Rev. 2004;23(8):1055-1085. doi:10.1016/j.cpr.2002.09.001

[5] Chamorro-Premuzic T, Furnham A. Personality, intelligence and approaches to learning as predictors of academic performance. Pers Individ Dif. 2008;44(7):1596-1603. doi:10.1016/j.paid.2008.01.003

[6] Rentfrow PJ, Gosling SD. The Do Re Mi's of Everyday Life: The Structure and Personality Correlates of Music Preferences. J Pers Soc Psychol. 2003;84(6):1236-1256. doi:10.1037/0022-3514.84.6.1236

[7] Jost JT, West T V., Gosling SD. Personality and Ideology As Determinants of Candidate Preferences and "Obama Conversion" in the 2008 U.S. Presidential Election. Du Bois Rev Soc Sci Res Race. 2009;6(1):103. doi:10.1017/S1742058X09090109

[8] Chung CK, Pennebaker JW. Linguistic Inquiry and Word Count (LIWC). Appl Nat Lang Process. 2015;(2015):206-229. doi:10.4018/978-1-60960-741-8.ch012

[9] Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol. 2010;29(1):24-54. doi:10.1177/0261927X09351676

[10] Faliagka E, Ramantas K. Application of Machine Learning Algorithms to an online Recruitment System. ICIW 2012 Seventh Int Conf Internet Web Appl Serv. 2012;(c):215-220.

[11] Ahmad N, Siddique J. Personality Assessment using Twitter Tweets. Procedia Comput Sci. 2017;112:1964-1973. doi:10.1016/j.procs.2017.08.067

[12] Pennebaker JW. The secret life of pronouns. New Sci. 2011;211(2828):42-45. doi:10.1016/S0262-4079(11)62167-2

[13] Groom CJ, Pennebaker JW. Words. J Res Pers. 2002;36(6):615-621. doi:10.1016/S0092-6566(02)00512-3

[14] Pennebaker JW, Boyd RL, Jordan K, Blackburn K. LIWC Manual. Environ Plan D Soc Sp. 1983;1(2):163-180. doi:10.1068/d010163

[15] Abe JAA. Positive emotions, emotional intelligence, and successful experiential learning. Pers Individ Dif. 2011;51(7):817-822. doi:10.1016/j.paid.2011.07.004

[16] Wynne HSZ, Wheeldon L, Lahiri A. Compounds , phrases and clitics in connected speech. J Mem Lang. 2018;98:45-58. doi:10.1016/j.jml.2017.08.001