

IDENTIFYING THE MOST INFLUENCING CHARACTERISTICS OF FAKE NEWS

FERGYANTO EFENDY GUNAWAN AND VINCENT SUWANDI

Industrial Engineering Department, BINUS Graduate Program – Master of Industrial Engineering
Bina Nusantara University
Jl. K. H. Syahdan No. 9, Kemanggisian, Palmerah Jakarta 11480, Indonesia
fgunawan@binus.edu

Received July 2019; accepted October 2019

ABSTRACT. *Nowadays, social media have facilitated fake news or hoax to reach wider audience at a high speed. Many studies have been conducted to establish the characteristics of such news. This work intends to identify the characteristics that make the news convincing and influential. To achieve the objective, we relate the fake news characteristics, identified from previous publications, to Klout score, a well-established indicator describing the impacts of news to their audience. A piece of news having a high Klout score is regarded to be influential. For the current study, the data are collected from Twitter messages containing issues related to politics, business, and entertainment during the period of January-June 2017. A sample of 165 tweets is collected, and then, is identified for the fake news characteristics and Klout score. Finally, a logistic regression function is established relating the characteristics to the Klout score. The results suggest that the fake news characteristics of the use of fake photos/images and the absence of the publishing time are significant factors for the news to reach wider audience.*

Keywords: Fake news, Hoax, Klout score, Social media, Twitter

1. Introduction. From the perspectives of communication and journalism study, the debate has always been about how media play their role as a news carrier to people. Meanwhile, based on the business perspective, all media's efforts in publishing news must be focused on attracting the attention of the readers or audience. This situation becomes increasingly challenging at the present time where online media such as social media are the main sources of news for the world community [1]. The glory of social media makes stories and events widely spread to millions of users instantly, yet at the same time, the number of articles describing similar events multiplies [2].

Hoax, a form of misinformation, is certainly a serious problem. Fake news or hoax utilizes existing information technology infrastructure to spread quickly and to generate big impacts. The results of the National Hoax Outbreak Survey by Indonesian Telematics Society (Mastel) show that 92.4% respondents received fake news from social media, including Twitter [3]. From early 2016 until February 13, 2017, a review team, established by the Ministry of Communication and Informatics of Republic of Indonesia, received 3252 reports about negative contents in Twitter. During the same period, the team acquired 1572 negative-content reports in Facebook and Instagram.

Regarding fake news, attention of many scientists has been directed on developing methodologies for detection. For the purpose, [4] suggested the use of linguistic and network-based approaches. [5] identified fake news possessing certain linguistic patterns such as the use of suspenseful language, unresolved pronouns, reversed narrative style, forward referencing, the use of images, and other cues of clickbait. Moreover, [6] advised the

use of credibility checking by a hierarchical credibility propagation analysis. [7] suggested the use of a news certification system. Finally, [8] evoked the use of Twitter Trails.

Determining the relevant characteristics of fake news is important in detecting hoaxes. Understanding those characteristics can facilitate the development of a detection system and may prevent the spread of fake news later. Detecting and preventing the spread of hoax are indeed very challenging [9]. Therefore, tools or methods are needed to effectively implement them. Klout score can be the solution for the matter.

Klout score is a metric for measuring social influence in online platforms such as social media [10]. Relevant data are retrieved from online source and then processed with a specific algorithm to produce a value that represents influence. People who share fake news in social media have certain social influences [9]. Therefore, Klout score is used in this study to measure the relevance of factors characterizing false news, in contrast to previous studies where linguistic methods were used to identify fake news.

This work intends to identify characteristics of fake news that make them influential. The results are a ranking of fake news characteristics from most to less prominent. These results are of importance for academia and law enforcers in particular and the current society in general.

We structure this paper as the following. Section 2, Theoretical Background, provides definitions and explanation regarding fake news, social media, and Klout score. They are essential for understanding the whole paper. Section 3, Research Method, presents the data collection and analysis method. Section 4, Results, shows the collected data in the form of descriptive statistics, the established logistic regression model, and discussion. Finally, Section 5, Conclusion, briefly restates the research objectives and the most important findings, and proposes a topic potential for future research.

2. Theoretical Background.

2.1. News. News is a report on the latest ideas or facts that are true, interesting, and important for most audiences through regular media such as newspapers, radio, television, or Internet [11]. News does not refer only to the press in a narrow sense but also on radio, television, or the Internet. News has become a basic need of modern society around the world. News is also defined as the latest report of facts or opinions that are important or interesting for the audience and published through the mass media [12].

2.2. Fake news. Fake news is a deliberate and proven false news article, and can mislead readers [13]. Fake news is also defined as all the news about a situation or event that at the time of broadcast has proven its contents contradict the truth [14]. Fake news mostly highlights inconsistencies in political rhetoric and insinuates the norms that govern general news media through an ironic inversion of recent news [15]. Fake news is usually made by individuals or groups who are motivated by profit and/or fame and have a low level of ethics and responsibility towards society [16].

Fake news stories are aimed at influencing or manipulating readers' opinion on a topic to achieve a particular goal [17]. In this age of information, hoaxes are likened to counterfeit money circulating in cyberspace. Some people come to use it by spreading the fake news again. Of course, the fake news maker benefits the most from that regardless of the motive behind the making [18].

Fake news has shown its influence in various fields over the last few years. Some of them are fake news in social media influential to the victory achieved by Donald Trump in the presidential election of United States in 2016 [13]. Star News made fake news about a conversation recorded between actor Salman Khan and actress Aishwarya Rai which proved to be successful in gaining public attention for itself [19]. There was also an old article about the bankruptcy of United Airlines parent company that reappeared on the Internet without publishing time and was mistakenly believed to be the new one about

bankruptcy filing by the company, resulting in United Airlines’ stock price rapidly fell by 76 percent before Nasdaq stopped the trading [20].

Social media users can identify fake news with the characteristics which include hyperbolic and clickbait headlines; suspicious website domains that spoof legitimate news media; misspellings in content and awkwardly laid out website; doctored photos and images; absence of publishing time; lack of author, sources, and data [17].

2.3. Klout score. Klout score is a metric for measuring the influence of users in online platforms, such as social media and community forums [10]. In other words, Klout score is a potential scoring system that can be used as a measure of credibility. Klout score is calculated based on the user’s ability to push actions in social networks. Users’ Klout scores are measured based on three components: True Reach (how many people are affected by the user), Amplification (how much the user affects them), and Network Impact (the influence of the user’s network). Users whose networks consist of other influential users who respond or share the posts will have a higher Klout score. The information is collected daily, and every morning, the Klout score of the user is updated. Klout scores range from 1 to 100, with a higher score indicating a higher level of influence as well. Klout average score is 40. Figure 1 illustrates the Klout scoring process.

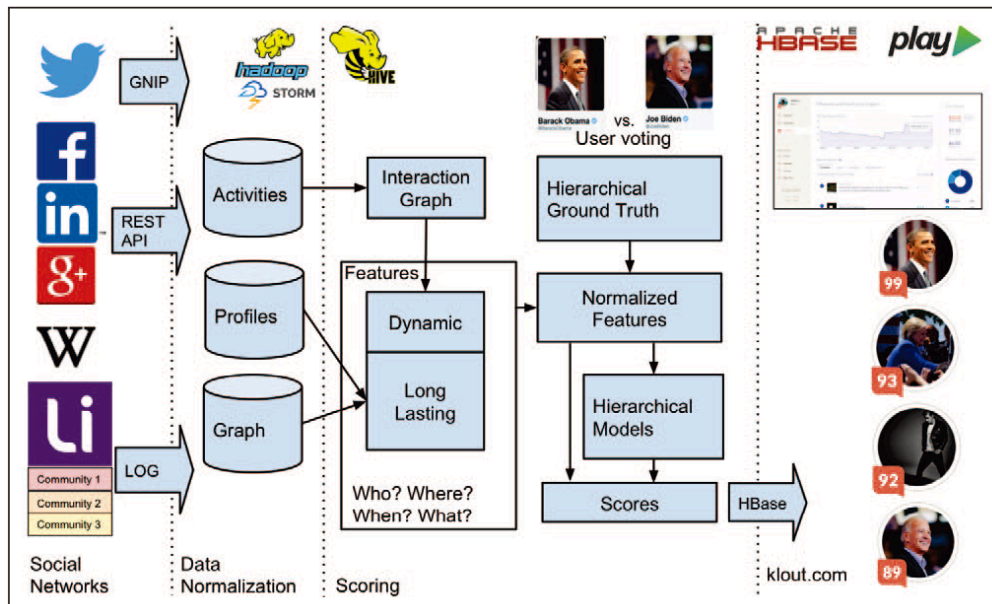


FIGURE 1. The Klout scoring process [10]

To produce a Klout score, a feature generation framework gathers signals across multiple dimensions for each user, creating a large feature set that contains more than two thousand features. In addition to combining signals from social networking interactions, they also include factors that provide proxies for real-world influences. Weights obtained from the supervised model are applied to these features to generate network or community values. This score is then combined hierarchically to obtain a final Klout score.

2.4. Social media. According to Kaplan and Haenlein [21], social media are defined as a group of Internet-based applications with ideological foundations and Web 2.0 technologies, which enables the creation and exchange of user-generated contents. Social media allow democratization of information and a shift in the role that people play in the process of reading and disseminating information. Social media empower people to become information spreaders. Therefore, it can be said that social media signify the shift of information dissemination from broadcast mechanism (one-to-many) to many-to-many mechanism [22].

Twitter is a social medium that allows users to send and read text-based messages up to 140 characters known as tweets. Tweets are not displayed on the user profile only, but can also be brought to followers via instant messages, Short Message Service (SMS), Really Simple Syndication (RSS), email, or to other social media. In addition, Twitter also has other features like Home or Feed, Notifications, Direct Message, Followers, Following, Likes, Retweets, Pinned Tweets, and Trending Topic. As of 2016, Twitter already has over 319 million monthly active users.

3. Research Method.

3.1. Research design. The method used for this research is quantitative. This research is started with formulation of the problem and a review of the relevant theoretical background, which has been presented in Section 2. In the following, we discuss the data sampling and collection method, fake news' characteristics extraction, and the statistical analysis method. From the six characteristics of fake news on social media, see Section 2.2, the most relevant ones to the Klout score are firstly identified. The presence or absence of each characteristic in each hoax is determined by a dummy variable. For example, the characteristics of the existence of the publication, we set its value to 0 if it exists and to 1 otherwise. In addition, we also record the Klout score of the individual who distributes the news.

3.2. Sampling. The sample is selected by purposive sampling [23]. Only those in accordance with the research objective are selected. Purposive sampling is classified as a type of non-probability sampling. This study takes samples of fake news that have the following criteria: fake news tweeted on Twitter, fake news released during the period of January 2017 through June 2017, fake news in politics, business, and entertainment. In [13], a sample size of 156 was considered sufficient for this study case. We adopt the same approach.

3.3. Data collection. This research requires two types of data. The first data are tweets containing fake news in the fields of politics, business, and entertainment during the period of January 2017 through June 2017. The second data are the Klout scores of Twitter users who post fake news with the criteria mentioned before.

Twitter is chosen in this study because according to the Mastel's survey [3], 92.4% of respondents received fake news in the Twitter platform. The review team of the Ministry of Information Republic of Indonesia reported that the majority of the complained were related to tweets. This social medium platform is vulnerable to fake news because the information may flow non-stop from one person to another [24]. Another reason is that Twitter's Gnip APIs are open-source, so it is possible to see Klout score from other Twitter users who do not have Klout accounts by using the Klout add-on in the web browser. It cannot be done in other social media that have closed API like Facebook.

Fake news on Twitter is searched by using the search feature. The search is assisted by the B.S. Detector extension as well as specific searches on the Internet. The B.S. Detectors extension that can be installed in Google Chrome marks fake news with a red warning sign. Specific searches can be done in Google and fact-checking websites like Snopes.com, TurnBackHoax.id, and Gossip Cop. As for Klout score data, the Klout extension in Chrome needs to be enabled first, so Klout score from user can directly appear next to its user-name.

3.4. Data analysis. The collected data are processed and arranged in a table format to be reprocessed for a statistical analysis. This is done to facilitate the data processing using Statistical Package for the Social Sciences (SPSS) software.

The data are analyzed by using descriptive statistics and logistic regression analysis. Descriptive statistics analysis is an analysis that gives general description of various variables in the study consisting of mean, median, mode, standard deviation, maximum value, minimum value, sum, range, kurtosis, and skewness. This analysis describes the data that has been collected as is without the aim of making conclusions for generalization. Therefore, descriptive statistics analysis facilitates understanding of the variables used in the study. Logistic regression analysis is an analysis used to see the relationship between categorical response variables with categorical and continuous predictor variables. Logistic regression analysis calculates conditional probabilities for objects in the information table [25]. This analysis is done by multinomial logistic regression in SPSS 23. The relevant statistics such as the coefficient of determination, correctness of classification model, and individual parameter significance test are also computed.

4. **Results.** To understand which fake news characteristics that strongly contribute to the spread of the news, a logistic regression model is established. The model is written

$$Y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_6 X_6)}}, \tag{1}$$

where Y denotes the Klout score, X_1 is the fake news characteristic of hyperbolic and clickbait headlines, X_2 is the fake news characteristic of suspicious website domains that spoof legitimate news media, X_3 is the characteristic of misspellings in content and awkwardly laid out website, X_4 is the characteristic of doctored photos and images, X_5 is the characteristics of absence of publishing time, and X_6 is the characteristic of lack of author, sources, and data.

The total number of tweets collected during the period of January-June 2017 is 165. Those data are obtained by purposive sampling method within three fields: politics, business, and entertainment. The statistical descriptions of the responses for Klout score variable are presented in Figure 2. It shows that the majority of the collected tweets are from users who have high impact. The number of tweets is distributed uniformly across the three fields: politics, business, and entertainment.

The descriptive statistics of the number of the fake news characteristics are presented in Table 1 for high impact tweets and in Table 2 for low impact tweets. For the former case,

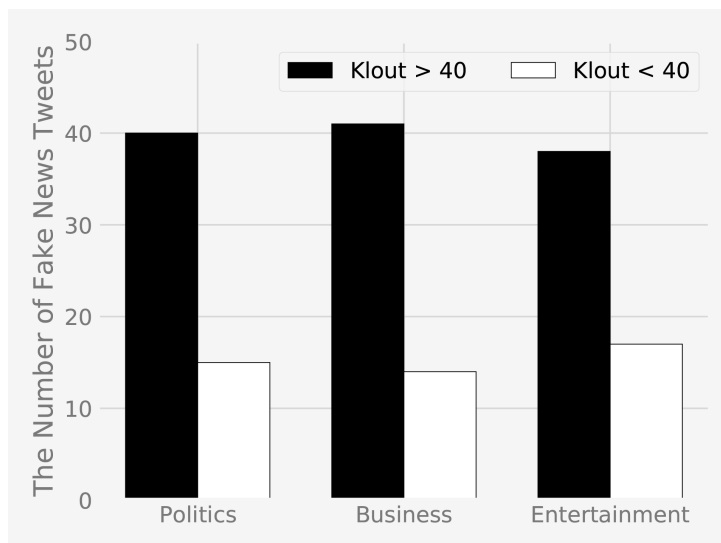


FIGURE 2. The number of tweets of fake news having high impacts (Klout score > 40) and low impacts (Klout score < 40) in politics, business, and entertainment

TABLE 1. The descriptive statistics of the responses associated with the high impact tweets for the six fake news characteristics

Characteristics	N	Mean	Standard deviation
X_1	119	0.849	0.360
X_2	119	0.067	0.252
X_3	119	0.824	0.383
X_4	119	0.261	0.441
X_5	119	0.244	0.431
X_6	119	0.849	0.360

TABLE 2. The descriptive statistics of the responses associated with the low impact tweets for the six fake news characteristics. See Table 1 for the descriptions of those characteristics.

Characteristics	N	Mean	Standard deviation
X_1	46	0.696	0.465
X_2	46	0.087	0.285
X_3	46	0.891	0.315
X_4	46	0.500	0.506
X_5	46	0.630	0.488
X_6	46	0.957	0.206

TABLE 3. The results of the logistic regression analysis

	B	SE	Wald	Df	Sig.	Exp(B)	95% CI for Exp(B)	
							Lower bound	Upper bound
Intercept	-0.556	0.774	0.517	1	0.472			
X_1	-0.325	0.469	0.480	1	0.489	0.723	0.288	1.812
X_2	0.059	0.723	0.007	1	0.935	1.061	0.257	4.373
X_3	0.666	0.590	1.274	1	0.259	1.947	0.612	6.189
X_4	0.882	0.392	5.063	1	0.024	2.416	1.120	5.210
X_5	1.397	0.407	11.771	1	0.001	4.044	1.820	8.984
X_6	0.899	0.799	1.264	1	0.261	2.456	0.513	11.763

the dominant characteristics are the first and sixth characteristics. For the latter case, the recorded number of characteristics are rather uniform, except the second characteristics.

The results of the logistic regression analysis is depicted in Table 3. It shows that the fourth and fifth characteristics are having statistical significant relations to the Klout score. The fourth characteristic, namely, the factor of doctored photos and images, has the coefficient of $\text{Exp}(B_4) = 2.416$ with a p -value = 0.024. The fifth characteristic, namely, the factor of absence of publishing time, has the coefficient of $\text{Exp}(B_5) = 4.044$ with a p -value = 0.001. Clearly, the fifth characteristic is the strongest predictor of the Klout score. These findings suggest that fake news with doctored photos or images or without the data of publication date and publisher have greater impact.

In addition, the analysis also shows that the remaining characteristics seem to possess low influence of the fake news impacts. The accuracy of this model is shown in Table 4. The indicator measures the model capability to accurately predict the fake news Klout score on the basis of the fake news characteristics. The accuracy is about 80%.

In addition to the model prediction accuracy, the model fitness is also measured by using R^2 statistics. The results are 0.165 for the Cox and Snell pseudo R^2 and 0.237 for the Nagelkerke pseudo R^2 .

TABLE 4. The accuracy of the regression model (1) in predicting Klout score

Observed	Predicted		
	Klout Score > 40	Klout Score < 40	Percent Correct
Klout Score > 40	114.0	5.0	95.8%
Klout Score < 40	28.0	18.0	39.1%
Overall Percentage	86.1	13.9	80.0%

The results of this research show that the fifth fake news characteristic which is the absence of publishing time is one of the most relevant characteristics to Klout score based on significance level of 0.001 and the Wald value of 11.771. In addition, its Odds Ratio is 4.044, so it is a characteristic of fake news that most likely exists in hoaxes by users with Klout scores above 40. These results make sense even though the descriptive statistics analysis does not show this characteristic was used a lot in the fake news circulated during the period from January 2017 to June 2017, because the effect caused by the absence of publishing time is very harmful as in the case of United Airlines. In the case, an old article about the bankruptcy of United Airlines parent company that reappeared on the Internet without publishing time and was mistakenly believed to be the new one about bankruptcy filing by the company, resulting in United Airlines' stock price rapidly fell by 76 percent before Nasdaq stopped the trading [20].

The fourth characteristic (doctored photos and images) is the other most relevant characteristic to Klout score. Although it did not appear a lot on fake news during the period from January 2017 to June 2017 according to descriptive statistics analysis, the significance level of this characteristic is 0.024 and the Wald value is 5.063. In addition, the fourth characteristic's Odds Ratio is 2.416, so it also most likely exists in the hoaxes by users with Klout scores above 40. These results can be understood, because fake photos and images can have a major impact on the right time as in the case of the photo of a shark swimming in the streets of Houston, USA that was flooded due to Hurricane Harvey. The photo proved to be fake, but have succeeded in triggering furor and fear from people.

This research also shows the first (hyperbolic and clickbait headlines), second (suspicious website domains that spoof legitimate news media), third (misspellings in content and awkwardly laid out website), sixth characteristic (lack of author, sources, and data) are not relevant to Klout score. That can be said, because Wald values and significance levels of these 4 characteristics are below standard. Nevertheless, the overall descriptive statistics analysis shows the 1st, 3rd, and 6th characteristics existed a lot in fake news circulated during the period of January 2017 to June 2017. The conclusion that can be drawn is that fake news which contains spoofings is not taken seriously by netizens. Another conclusion is the hyperbolic and clickbait headlines; misspellings in content and awkwardly laid out website; and lack of author, sources, and lack of author, sources, and data are seen by netizens as mainstream elements that are common in fake news, so they are less likely to be trapped by hoaxes with those characteristics.

Klout score helped this research to conclude that a hoax that has the fourth fake news characteristic (doctored photos and images) and/or the fifth characteristic (the absence of publishing time) is a convincing hoax. Thus, Klout score can be recommended for entry into a large group of fake news identification supporting tools and methods which have previously encompassed a variety of solutions from earlier studies such as linguistic and network-based approaches [4], three types of fakes approach [26], linguistic pattern consideration approach [5], hierarchical propagation model [6], real-time news certification system [7], as well as Twitter Trails [8].

5. **Conclusion.** This work brings our focus to fake news characteristics that may help the news to spread widely. To identify them, firstly, we study the fake news characteristics described by [17]: hyperbolic and clickbait headlines, suspicious website domains that spoofing legitimate news media, misspellings in content and awkwardly laid out website, doctored photos and images, absence of publishing time, and lack of author, sources, and data. Then, we link those characteristics to Klout score [10], a metric measuring the influence of news in online platform such as social media and community forum, by a logistic regression. Higher Klout score means higher impact. From the analysis, we find that fake news having the characteristics of containing doctored photos or images and without the publication time tends to spread to wider audience.

The propagation of fake news through social networks can also be studied by means of graph theory. In fact, the approach, for example, by using agent-based simulation, potentially provides a more detail description of the issue. We offer this as future research where one may establish the propagation networks of fake news with and without doctored photos and then, analyze the network sizes and the propagation velocities from node to node.

REFERENCES

- [1] Reuters, *Digital News Report 2016*, Tech. Rep., Reuters Institute, 2016.
- [2] J. H. Kim, A. Mantrach, A. Jaimes and A. Oh, How to compete online for news audience: Modeling words that attract clicks, *Proc. of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.1645-1654, 2016.
- [3] Mastel, *Survey Results of National Hoax Outbreak*, Tech. Rep., Indonesia Telematics Society, 2017 (in Indonesian).
- [4] N. J. Conroy, V. L. Rubin and Y. Chen, Automatic deception detection: Methods for finding fake news, *Proc. of the Association for Information Science and Technology*, vol.52, no.1, pp.1-4, 2015.
- [5] Y. Chen, N. J. Conroy and V. L. Rubin, Misleading online content: Recognizing clickbait as false news, *Proc. of the 2015 ACM on Workshop on Multimodal Deception Detection*, pp.15-19, 2015.
- [6] Z. Jin, J. Cao, Y.-G. Jiang and Y. Zhang, News credibility evaluation on microblog with a hierarchical propagation model, *IEEE International Conference on Data Mining (ICDM)*, pp.230-239, 2014.
- [7] X. Zhou, J. Cao, Z. Jin, F. Xie, Y. Su, D. Chu, X. Cao and J. Zhang, Real-time news certification system on sina weibo, *Proc. of the 24th International Conference on World Wide Web*, pp.983-988, 2015.
- [8] P. T. Metaxas, S. Finn and E. Mustafaraj, Using twittertrails.com to investigate rumor propagation, *Proc. of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*, pp.69-72, 2015.
- [9] D. Lazer, M. Baum, N. Grinberg, L. Friedland, K. Joseph, W. Hobbs and C. Mattsson, Combating fake news: An agenda for research and action, *Harvard Kennedy School, Shorenstein Center on Media, Politics and Public Policy*, vol.2, 2017.
- [10] A. Rao, N. Spasojevic, Z. Li and T. DSouza, Klout score: Measuring influence across multiple social networks, *IEEE International Conference on Big Data*, pp.2282-2289, 2015.
- [11] A. H. Sumadiria, *Indonesia Journalism: Writing News and Features: Practical Guide for Professional Journalists*, Simbiosis Rekatama Media, Bandung, 2008 (in Indonesian).
- [12] J. Yosef, *To Be Journalist: Becoming Professional Journalist for TV, Radio, and Newspapers*, Graha Ilmu, 2009 (in Indonesian).
- [13] H. Allcott and M. Gentzkow, Social media and fake news in the 2016 election, *Journal of Economic Perspectives*, vol.31, no.2, pp.211-236, 2017.
- [14] A. Ferdian, Harmonization of special criminal arrangements for the object of fake writing and news, *Journal of Law Panoramic*, vol.1, no.1, pp.27-42, 2016 (in Indonesian).
- [15] D. G. Young, *Two Presidential Candidates Walk into a Bar. Late-Night Political Humor: Cognitive Processes, Political Consequences and Normative Implications*, Ph.D. Thesis, University of Pennsylvania, 2007.
- [16] G. Nomad, *The Great Radio Hoax: With Audio-Link to the Infamous Broadcast*, Dreaming Billabong Books, 2016.
- [17] L. Gu, V. Kropotov and F. Yarochkin, *The Fake News Machine*, Trend Micro: A TrendLabs Research Paper, Tech. Rep., <https://www.a51.nl/sites/default/files/pdf/wp-fake-news-machine-how-propagandists-abuse-the-internet.pdf>, 2017.

- [18] S. E. Hafiz, Internet: Building awareness of new public spaces, *Proc. of National Conference of Young Indonesian Psychology Researcher*, vol.2, no.1, pp.5-7, 2017.
- [19] D. K. Thussu, The 'murdochization' of news? The case of star TV in India, *Media, Culture & Society*, vol.29, pp.593-611, 2007.
- [20] C. Carvalho, N. Klagge and E. Moench, The persistent effects of a false news shock, *Journal of Empirical Finance*, vol.18, no.4, pp.597-615, 2011.
- [21] A. M. Kaplan and M. Haenlein, Users of the world, unite! The challenges and opportunities of social media, *Business Horizons*, vol.53, pp.59-68, 2010.
- [22] B. Solis and D. K. Breakenridge, *Putting the Public Back in Public Relations: How Social Media is Reinventing the Aging Business of PR*, FT Press, 2009.
- [23] T. Palys, Purposive sampling, in *The Sage Encyclopedia of Qualitative Research Methods*, L. M. Given (ed.), Los Angeles, Sage, pp.697-698, 2008.
- [24] H. Situngkir, *Spread of Hoax in Social Media*, Tech. Rep., Fe Institute, Bandung, BFI Working Paper No. WP-4-2011, 2011.
- [25] D. Liu, T. Li and D. Liang, Incorporating logistic regression to decision-theoretic rough sets for classifications, *International Journal of Approximate Reasoning*, vol.55, no.1, pp.197-210, 2014.
- [26] V. L. Rubin, Y. Chen and N. J. Conroy, Deception detection for news: Three types of fakes, *Proc. of the Association for Information Science and Technology*, vol.52, no.1, pp.1-4, 2015.