Text Mining of UU-ITE Implementation in Indonesia

1Lukmanul Hakim, Tien F. Kusumasari2 and Muharman Lubis2*
1International Islamic University Malaysia, Jalan Gombak, Kuala Lumpur, 50728
2Telkom University, Jalan Telekomunikasi No. 1, Bandung, 40257

E-mail: *muharmanlubis@telkomuniversity.ac.id

Abstract. At present, social media and networks act as one of the main platforms for sharing information, idea, thought and opinions. Many people share their knowledge and express their views on the specific topics or current hot issues that interest them. The social media texts have rich information about the complaints, comments, recommendation and suggestion as the automatic reaction or respond to government initiative or policy in order to overcome certain issues. This study examines the sentiment from netizens as part of citizen who has vocal sound about the implementation of UU-ITE as the first cyberlaw in Indonesia as a means to identify the current tendency of citizen perception. To perform text mining techniques, this study used Twitter Rest API while R programming was utilized for the purpose of classification analysis based on hierarchical cluster.

1. Introduction
Recently, the rapid growth of social media has led many users to generate web-based contents for various subjects such as sharing information, online chatting, and expressing their opinions, thereby causing the amount of text information become abundance. Due to this phenomenon, the need to analyse unstructured text in social media cannot be ignored. Indeed, analysing social media has become a compulsion for academia and industry [1, 2] in order to discover untapped knowledge. It also supported with a diverse techniques, tools, and platforms that available publicly. The reasons why it is become a major research and business activity today are because of commercial pressures and the potential of social media for analysis. Furthermore, to ensure legal certainty in cyber transaction, encouraging economic growth and as a means to prevent cybercrime for protecting service to the citizen, government regulate UU-ITE in 2008 and has been revised in 2016 involves explanation of insulting and contaminating verse interpretations, decreasing the threat of punishment, setting up intercepts, synchronization with KUHAP, strengthen the role of civil servant investigators, the provision of right to be forgotten or the obligation to remove irrelevant content for e-system providers and the role of government. Meanwhile, there are many controversies regarding the implementation of this Act such as case on defamation involves Prita, Arsyad, Ervani, Florence, Fadhli and other hundred victims, threats to BBM from minister of informatics and SARA (ethnic, religion, race, group) website blocking to prevent radicalism.

In general, the UU-ITE (Act No.14/2008) has main objective to manage information and transaction electronically as well to stipulate the prohibited acts without specifying the principle of privacy. The role of this regulation is to tackle the growing of cybercrime, so the evidence, statement and proof in the digital form or from device became legitimate and be allowed in the court. Despite the lack of agreement between legal experts in interpreting this regulation [5], this act has the vital role as
the basis for the direction of personal data protection and cybercrime prevention in Indonesia. The most important thing in the e-commerce transaction is the protection of personal data in every activity involves accessing, browsing, comparing or storing, which can be transformed into secured and trusted e-voting. However, many people considered the revision of the Act did not alter the substance of the potential problem. The provision of defamation or defamation in article 27or so-called "rubber articles" becomes the core of the objection. Some weaknesses regarding UU-ITE has been reported such as digital evidence, jurisdiction procedure, locus delicti and tempus, issues on context, content and concept [6]. This study want to identify the netizens perception on the implementation of UU-ITE based on text mining methods to see how does certain ideas or concepts has been transferred through social media and become a means to encourage government to revise and improve the regulation.

2. Literature Review
Indonesia is one of populous country with highest internet penetration has been struggling to cover all activities in cyberspace with a comprehensive cyber law, recognized as the Electronic Information and Transaction Bill (UU ITE). It is intentionally to give legal certainty in cyberspace which is the principle that should include in national and international law. However, the implementation of ITE getting various responses from Indonesian citizens, particularly from internet users since it was signed at 2008 and after the recent amendment in November, 2016. The presence of social media like Twitter easier for them to involve in a discussion, to share political contents base on their interest, or to create an online forum to criticize government policies with many participants involves. Moreover, the large number of Twitter users also impacted socio-political transition such as Arab Spring, and the Occupy Wall Street movement [16]. With simple features, short, and interactive comments, Twitter attracts many users to join in their network. In 2010, Indonesia categorized as the highest Twitter penetration, and ranked as the most tweeting country in 2014 [13]. This is due to the fact that, there are many cases convicted unprecedented numbers of citizens in Indonesia who interacting in the cyberspace. Such cases unfolded the intention of internet users, particularly in the Twitter that day by day became a trend. In majority, the social media users are the stakeholder who are the most susceptible to be criminalized with the articles of UU ITE rather than the journalism. According to Southeast Asia Freedom of Expression Network (SAFEnet) noted that there are more than 200 reports about the libels, desecration of religions and threats based on the UU ITE [8].

Text mining could be used to discover useful information, patterns, and knowledge in social media [4] although it still needs human intervention to interpret the output of text mining. It gives new possibility to researchers from wide range of disciplines to use text mining as vehicle of analysis [7], and also constitutes as an effective way to meet user’s diverse information needs [10]. Some scholars from multiple studies have exploited the text analytics in social media for their research development, such as political communication [7], security and defence[9], policy implementation [11, 12]. In another hand, some study tried to compare between the official poll or surveys and text mining analysis, they found that the social media data (e.g. twitter) could be used for social study because of strong correlation between them [7]. However, some criticism derived from other studies justifying that social media like Twitter is not a representative sample of any population [12]. Furthermore, the richness of text mining features could also utilize by government agencies to understand public responses, feedbacks, and reactions upon a policy implementation or initiative [15]. In order to achieve the transparency and engagement in government agencies, it is compulsory for them to get closer to the citizens by initiating to glean ‘citizen sourcing’. Most of netizens express their opinions in social media because of efficient channel for creation and exchange of political content [11].

Actually, many ways to represent the data in text mining; one of them is regarding of bag-of-words [10, 14]. This text mining approach has been used mostly in text analysis, Natural Language Processing (NLP), and Information Retrieval (IR) [1] rather than an algorithmic point of view regardless of its drawbacks [3]. The features of bag-of-words (BOW) also used by researchers in their study for some purposes, for example, the feature of BOW designed to detect location expression of Twitter users [9]. Therefore, this study engaged a bag-of-words model to retrieve an interesting topics,
patterns, and terms from social media Twitter in relation with ITE implementation in Indonesia. The intention of having security policy was not to persuade users but to convince them, by letting the users reflect, on their own terms, on why information security is important and on how to react in certain circumstances [18].

3. Research Methodology
The text messages (tweets) collected from the availability of web-based Application Programming Interface (API) provided by Twitter. To retrieve relevant text data, some parameters used to filter the need information such as keywords search, hashtags, phrases, regions, and usernames or user ID. There are 1843 tweets present in the corpus.

![Figure 1. Terms frequency of ITE tweets.](image)

This study applied text mining techniques to discover the perceived issues surrounding ITE implementation in social media (Twitter). The first stages are to scrape tweets in Twitter by using Twitter Rest API, in which R-Programming as the tool that was used in the entire processes of text analysis. The next stage, when the data that was extracted from Twitter API completed, the next stage called pre-processing techniques such as removing stop-word, stemming and TF/IDF algorithm. The motive that stop-words should be removed from a text is that they make the text look heavier and less important for analysts. Removing stop words reduces the dimensionality of term space. The most common words in text documents are articles, prepositions, and pro-nouns that does not give the meaning of the documents. Thus, the identification of root word should be done to remove various suffixes, to reduce the number of words, to have accurately matching stems, to save time and memory space. The third step is to build the term-document matrix where the features of terms were selected in a collection of the documents. In the next section the terms are plotted with word association, which available in R. Finally, the data then stratified with clustering techniques such as hierarchical cluster for further analysis. By building cluster tree or dendrogram to represent data, each node links to two or more successor node, which are nested and organized as a tree, ideally ends up as a meaningful classification scheme. Each node in the cluster tree contains a group of similar data; nodes group on the graph next to other, similar nodes. Clusters at one level join with clusters in the higher level by using a degree of similarity. The process carries on until all nodes are in the tree, which gives a visual snapshot of the data contained in the whole set. The total number of clusters is not predetermined before starting the tree creation.
Figure 2. Hierarchical cluster of ITE tweets.

The figure 1 presented obviously the frequency of each term which simply to show what the popular terms that discussed in the Twitter are. The word “ite” is the most used term in the tweets followed by “pasal/article”, “karet/elastic”, “nama/name”, “pencemaran/defamation”, and “korban/victim”, while the clusters of each term could be seen in the figure 2 which correlated to another term. These words grouped according to the distance of each term based on the method of hierarchical clustering. The distance between terms are calculated with dist() after scaling, then clustered with hclust() and dendrogram is cut into 5 clusters. The agglomeration method is set to ward to indicate the increase in variance when two clusters are merged. Other than that there are many other options can be use such as single linkage, complete linkage, average linkage, median linkage, and centroid. Some of sparse terms are removed to prevent the plot with a crowd words. These two graphics are used to facilitate the researchers easy to interpret the issues surrounding ITE implementation in Indonesia from Twitter users that vocally expressed their opinion, especially the case related to freedom of expression. Based on the figure 2, from the first cluster that the case is occurred involving a person named “Acho”, which perceived against consumer right, so that they blamed some article s in the UU ITE has potential to be interpret improperly. The second cluster denoted about the libel cases that still prevailed in Indonesia, particularly in social media that attracted the “netizens” attention. The remains clusters discussed about the issues of infringement related UU ITE that became popular among the netizens in the twitter.

4. Discussion

4.1 Legal Consequence
The Twitter users denoted that they confronting such a fear issue in the internet, whether to communicate, to interact, or to express their opinions by considering the legal effect upon certain conducts. This is due to the fact that particular cases happened in Indonesia that prosecuted some perpetrators, received many responses of the netizens to the articles which charged the suspects with improper articles. Moreover, the defamation of others reputation is still prevailing among the internet users in Indonesia. It is recorded in the cluster dendrogram that the words “pencemaran” and “nama” closed together. The action of libel often involves parties and individuals or vice versa that potential to be misuse by certain parties to disadvantages others. It is because the vague definition of defamation leading to multi-interpretation among citizens, or certain parties. Furthermore, the perception of doubt upon certain articles such as the article 27 section (3) has disclosed public trustworthy while article 28 also became the consideration due to possibility to be misinterpreted. Even though the first amendment has been made, the anxiety is still leaving a big trauma of the citizens. There as suggestion
to remove this article and use article 310-311 of the Criminal Code (KUHP). Therefore, it can be threaten freedom of expression in the cyberspace. There are also high number of hateful messages in the social media such as Luna Maya comment’s in her twitter account says that the degrees of infotainment is more despicable than prostitutes and murderers. Indeed, the attempts to meddle in private matters should be discourage through regulation to preserve the individual privacy. In short, the issues of UU-ITE based on text mining relate to the importance of defamation article, the uniformity of insulting case, vagueness of “good name” and strange proof attempt in the court. Some problematic verses has been rejected by constitutional court after the Act initiated, so the other judicial review cannot be conducted. Thus, government should take initiative by revise plan or develop new regulation to give clear definition of certain vague term and to remove distrust among citizen on the proper intention of this Act.

4.2 Precedence
There are many cases convicted unprecedented number of netizens who interacting in the cyberspace. Based on the recent cases discussed in the Twitter, some of these issues attracted the internet users, particularly in the Twitter that day by day became a trend. The presumption of innocent went viral among them. It is begin with tweets of someone or news portal in the Twitter somehow shared quickly to other users. The sympathy channelled to the particular person who presumes commits libel based on themselves perceptions. In short, the assumption of arbitrary power abuse by some parties encouraged them to express their opinion in the Twitter. It is aligned with other research [17] found that only a minority understood the UU ITE really well and applied it in their daily operations either in the individual task or at company operation consistently and strictly while the majority of Indonesians knew nothing about it at all. There were so many companies in Indonesia, which had just started operations and were still ignorant of the rules and regulations governing this sector and at the same time did not feel obliged to consult legal experts. The worst part was that there was no proper control or enforcement in cases of fraud. Occasionally, most Indonesians had made bad experiences in using electronic devices, often in the form of spam delivered to mobile phones or personal computers, either to exploit the emotion in the form of cheating or allegation of defamation [6], which might result in the reluctant support of the government policy and regulation implementation to overcome cyber issues. Moreover, since fraud occurred freely without proper response from the government, Indonesians rightfully and understandably doubted in honest, fair and proper handling of cybercrime based on UU-ITE ever to take place. Indeed, government should not be worried to get criticism from citizen while at the same time provide the way to remind every components include social and mass media to consider ethics and norm in sharing and delivering their mature opinion and comprehensive thought. Certain restriction can be implemented in other type of regulation to prevent other citizen’s right be violated by the others such as cyberbullying or cyber stalking. The plan to draft criminal code in information technology should be supported instead of commenting on the negative effect on certain regulation implementation. Government should take a note that strict restriction should be imposed on specific sensitive issues only such as SARA, houx, hateful message, spam and other negative activity, not to whole content of communication from citizen about policy or their experience.

5. Conclusion
To maintain the enforcement of personal data protection, government is encouraged to create the periodical auditory, proper compliance and privacy impact assessment towards the implementation of the policy. Therefore, it is really resourceful to understand the factor that shape the negative sentiment from netizens. At certain extent, educational level, environment location, economic condition and social status play essential role to determine on how people to perceive government attempt in solving national problems. By understanding these issues through indicators, government can anticipate several possibilities to avoid worst case scenarios, in which certain party might exploit for their own interest. Thus, the policy should measure the attribute of the process, data subject behavior and characteristics. It is recommended that the Indonesian government created a special unit with the task
of monitoring the transaction in the digital and handle the cybercrime investigation. Nevertheless, the trust is essential to create good relationship between netizens and government, which focus to the ground issue not the politics as the sign of good intention and concern [19]. For the sake of realizing the ideal and admirable level of society, the presence of the UU-ITE Law and its revisions need to be appreciated. It forms the responsibility of the state to its people to protect its people from the negative things due to freedom in the virtual world, the application of the UU-ITE and its revisions could be used as a tool for the government to silence all forms of criticism.

References
[1] Batrinca, B & Treleaven, P. C. (2015). Social media analytics: A survey of techniques, tools and platforms. AI & Soc., 89-116.
[2] Palipopoulou, V., & Maragoudakis, M. (2013). On Mining Opinion on Social Media. EANN (pp. 474-484). Halkidiki, Greece: Springer-Verlag Berlin Heidelberg.
[3] Aggarwal, C. C., & Zhai, C. X. (2012). An Introduction to Text Mining. In C. C. Aggarwal, & C. X. Zhai, Mining Text Data (pp. 1-10). Springer US.
[4] Guidi, M., Ruiz-Agundez, I., & Canga-Sanchez, I. (2012 ). Knowledge Mining from the Twitter Social Network: The case of Barack Obama. In A. Abraham, Computational Social Networks: Mining and Visualization (pp. 211-229). London: Springer-Verlag.
[5] E, Kusumasari TF, Lubis M and Lubis AR. Study to the Current Protection of Personal Data in the Educational Sector in Indonesia. ICCAI 2017. IOP Conference Series.
[6] LubisM and Maulana F A. 2010. Information and Electronic Transaction Law Effectiveness (UU-ITE) in Indonesia. IEEE ICT4M, pp. C-13-19.
[7] Ampofo, L., Collister, S., O'Loughlin, B., & Chadwick, A. (2015). Text mining and social media: When quantitative meets qualitative, and software meets humans. In P. Halfpenny, & R. (. Procter, Innovations in Digital Research Methods (pp. 161-192). London: Sage.
[8] Southeast Asia Freedom of Expression Network (SAFEnet) (2017). Daftar Kasus Netizen yang Terjerat UU ITE – SAFEnet Voice. Retrieved from http://id.safenetvoice.org/daftarkasus/.
[9] Inkpen, D. (2016). Text mining in social media for security threats . In R. Abielmona, R. Falcon, N. Zincir-Heywood, & H. A. Abbass, Recent advances in computational intelligence in defense and security (pp. 491-517). Springer, Cham.
[10] Hu, X., & Liu, H. (2012). Text analytics in social media. In C. C. Aggarwal, & C. X. Zhai, Mining Text Data (pp. 385-414). LLC: Springer Science+Business Media.
[11] Charalabidis, Y., Maragoudakis, M., & Loukis, E. (2015). Opinion mining and sentiment analysis in policy formulation initiative: The EU-community approach. International Conference on Electronic Participation (pp. 147-160). Thessaloniki: Springer, Cham.
[12] Arunachalam, R., & Sarkar, S. (2013). The new eye of government: Citizen sentiment analysis in social media . Workshop on Natural Language Processing for Social media (SocialNLP) (pp. 23-28). Nagoya : IJCNLP.
[13] Carley, K. M., Malik, M., & Kowalchuk, M. (2015). Twitter usage in Indonesia. Pittsburgh: Institute for Software Research, School of Computer Science Carnegie Mellon University.
[14] Witten, I. H. (2004). Text Mining. In M. P. Singh, The practical handbook of internet computing (pp. 1-22). Washington: DC: Chapman & Hall/CRC.
[15] Finlay, S. (2014). Predictive analytics, data mining and big data. Steven Finlay.
[16] Kumar, S., Morstatter, F., Liu, H. (2014). Twitter Data Analytics. New York: Springer.
[17] Lubis M, Kartiwi M and Zulhuda S. 2017.Privacy and Personal Data Protection in Electronic Voting: Factors and Measures. Telkomnika, March vol. 15(1) pp. 512-521.
[18] Ahlan A, LubisM and Lubis A R. Information Security Awareness at the Knowledge-Based Institution: Its Antecedents and Measures. Procedia Computer Science, 2015. 72, 361–373.
[19] Lubis M and Kartiwi M. Privacy and Trust in the Islamic Perspective: Implication of the Digital Age. International Conference on Information and Communication Technology for the Muslim World, 2013.
Acknowledgement
This work would not have been possible without the academic support from my supervisor, Associate Professor Mira Kartiwi and the financial support from Telkom University.