

# Analyzing Main Topics Regarding the Electronic Information and Transaction Act in Instagram Using Latent Dirichlet Allocation

*Hans Kresnawan<sup>1</sup>, Sola Graciana Felle<sup>2</sup>, Hanna Gloria Mokay<sup>3</sup>, and Nur Aini Rakhmawati<sup>4</sup>*

*<sup>1,2,3,4</sup>Faculty of Intelligent Electrical and Informatics Technology, Department of Information Systems, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*

**Abstract.** Indonesia is currently experiencing its fourth industrial revolution in the 21st century. With the introduction of the internet, Indonesia is expected to gain more than a hundred billion US Dollars and twenty-six million job openings by 2030. The rising usage of information technology prompts regulators to develop The Electronic Information Transaction Act to protect the populace from cybercrime. However, the law attracts numerous criticism due to its vague interpretation. This resulted in numerous arrests of innocents throughout Indonesia. Thus, the public is trying to voice their opinions on social media for the sake of preventing any more cases in the future. The usage of Latent Dirichlet Allocation could provide numerous benefits for this research. The separation between latent topics among random mixtures helps to identify the common ground and the correlation between each post. These latent topics will be elaborated with a sample post to provide insights and expectations of the public towards the law.

**Keyword:** Instagram, Latent Dirichlet Allocation, Text Mining, Topic Analysis, UU ITE

Received 11 April 2021 | Revised 28 July 2021 | Accepted 28 July 2021

## 1 Introduction

The 4.0 Industrial revolution has sparked Indonesia into the digital age. A survey conducted by The Indonesia Internet Service Provider association supported the notion, in which 73,7% of the population used the internet [1]. The Ministry of Information, Communication, and Technology of Indonesia also estimates a revenue increase of around one hundred and forty billion US Dollars and twenty six million job openings in 2030 as the impact of the fourth industrial revolution [2]. In essence, the availability of the internet provides life-changing benefits to Indonesian society. The rising usage of the internet prompts regulators to figure out how to prevent the usage of the

---

\*Corresponding author at: Faculty of Intelligent Electrical and Informatics Technology, Department of Information Systems, Institut Teknologi Sepuluh Nopember, Jalan Raya ITS, Surabaya, Indonesia

E-mail address: [hansk.public@gmail.com](mailto:hansk.public@gmail.com)

technology for malicious intent. Thus, The Electronic Information and Transaction Act (UU-ITE) serves as a solution for the problem.

However, The Electronic Information and Transaction Act is one of the most controversial laws in Indonesia. The notion that the law can be interpreted from multiple angles occurred when numerous victims have been imprisoned due to the vague interpretation of the law. A director named Dandhy Dwi Laksono was arrested in 2019 for spreading hate speech according to the law [3]. Another case involves Baiq Nuril Maknun, a Teacher from Mataram, who was arrested for spreading explicit content when discussing her sexual harassment experience to her peers according to the law [4]. The ever-growing number of victims of the law provokes numerous people to voice their opinion on social media for the sake of preventing similar cases in the future. To understand the common consensus of the public regarding the law, Latent Dirichlet Allocation (LDA) can be used to extract latent topic from random mixtures between each post.

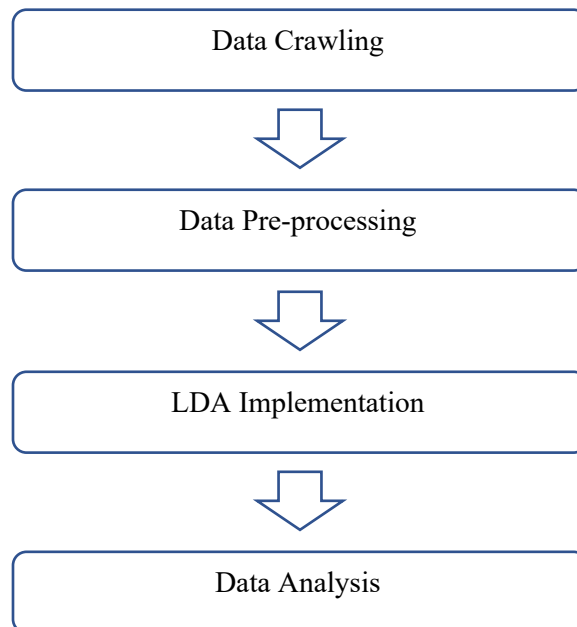
LDA is a probabilistic model that is generated from a text document [5]. According to the theory, there are two components that comprised the document [5]. Those components are latent topics and random mixtures [5]. The goal of LDA is to remove the latent topics from the document to obtain underlying information from the document [5].

Numerous studies have shown the usage of LDA as a way to understand trending topics on social media. A study about K-Pop topic modelling provides useful insights for content creators to align their content with popular sentiments [6]. Another study also suggests that LDA reveals primary concerns and fears regarding the COVID-19 pandemic [7]. The paper provides insight regarding topics revolving around COVID-19 in social media to explore the psychological reactions of the users during the early stages of the COVID-19 pandemic [7]. LDA also helped researchers to identify research topics related to their works from library electronic references [8]. Additionally, the model also helps in market research by providing topics that could increase performance of consumer choices of their products [9]. On another research, the model identifies tourist satisfaction that could be beneficial for the tourism industry [10]. Besides that, LDA also proven beneficial for modelling customer complaints [11], online review analysis [12], identifying ontology of enterprise technological innovation [13], image segmentation [14], research trends [15], and more.

The rising opposition against the law pressures the government to find a solution which could help satisfy numerous parties involved. Thus, this paper is aiming to simplify the problems faced by the people into main topics. This enables the government to analyze inputs easier from numerous sides and factions, including legal professional, university students, public figures, and more.

## 2 Research Method

This section summarizes the research method of this paper. There are six steps that needs to be taken to execute this experiment. Those six steps include literature studies, data crawling, data pre-processing, LDA implementation, data analysis, and summary. The diagram can be seen from Figure 1.



**Figure 1** Research Methodology

The research method of this manuscript revolves around the following points:

### 2.1 Data Crawling

The process of data crawling is executed with the help of the selenium library. To start with, the algorithm provides authentication in Instagram using a throwaway account. Furthermore, the algorithm automatically types the target hashtag that is desired to the Instagram search box. The algorithm later sends it to the website and opens every available post. By the time it runs out of posts to open, the algorithm scrolls down to load more posts to continue the process of data crawling. This process is repeated numerous times until the desired amount of loop is achieved. The algorithm has obtained data from 1137 posts on Instagram.

**Table 1** A Sample of the Collected Data

No	username	text	date
4	suara.aruna	<i>Jadi ini what the maksud guys?#uuite</i>	2021-02-24T01:34:33.000Z

## 2.2 Data Pre-processing

Before the data can be used for analysis, it needs to be checked regarding the amount of missing and duplicate values within the dataset. The process utilizes `pandas.DataFrame.Duplicated().sum()` and `pandas.DataFrame.isnull().sum()` from the python library. The texts for each post are also slightly modified by removing the newlines using. This process is necessary to make data analysis easier. After ensuring that the data is safe to use for analysis, the data will be exported to CSV to ensure reusability. Furthermore, slang words were also part of the problems that were faced during the research. Since slang words were not approved academically in Indonesia, this resulted in a destitute amount of both legitimate and documented data surrounding those words. In essence, the distribution of these slang words primarily relies on hearsay. Thus, the data was normalized using the Indonesian slang word dataset from [16]. Due to the massive amount of slang words that was documented, only the slang that occurs at each post will be mapped for conversion. Additionally, each text in the list is converted to lowercase using `lower()` function from python library. The text then tokenized using `RegexpTokenizer` from the Natural Language Toolkit library. The tokenized documents are then converted into a dictionary.

## 2.3 LDA Implementation

LDA is used to extract the main topics from the dataset. LDA works by extracting the latent topics from the document which is filled by random mixtures to acquire important information of the text [5]. The process of acquiring these topics is by treating each document as a set of topic probabilities. Furthermore, each topic is treated as a mixed set of words. The output of LDA provides the entire topics which comprised of numerous words alongside its probabilities. There are  $D$  number of documents, alongside  $K$  number of topics. Each document has correlation probabilities with each topic, visualized in  $\theta$ . Additionally, each topic has certain distribution of words that correlates with the topic, visualized in  $\beta$ . Each word is represented in  $w$ , and  $Z$  represents topics in each document. Thus, the probability of a document correlated with each topic can be summarized in equation(1).

$$p(D|\alpha, \beta) = \prod_{d=1}^M \int p(\theta_d|\alpha) \left( \prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn}|\theta_d) p(w_{dn}|z_{dn}, \beta) \right) d\theta_d \quad (1)$$

The process of this phase uses the dictionary that was created during pre-processing phase. The model is generated using the *gensim* library to compute the coherence score of the documents. The model will obtain the main topics for each text and presented them using plots.

## 2.4 Data Analysis

Data analysis revolves around identifying main topics regarding the #uuite hashtag. The topics will be elaborated on using keywords according to the similarity of each post. Each topic will be visualized with a bar plot to see the frequency distribution of the topics. The results will be constantly monitored to avoid any previously unidentified stopwords becoming the main topic of the posts.

## 3 Results and Discussion

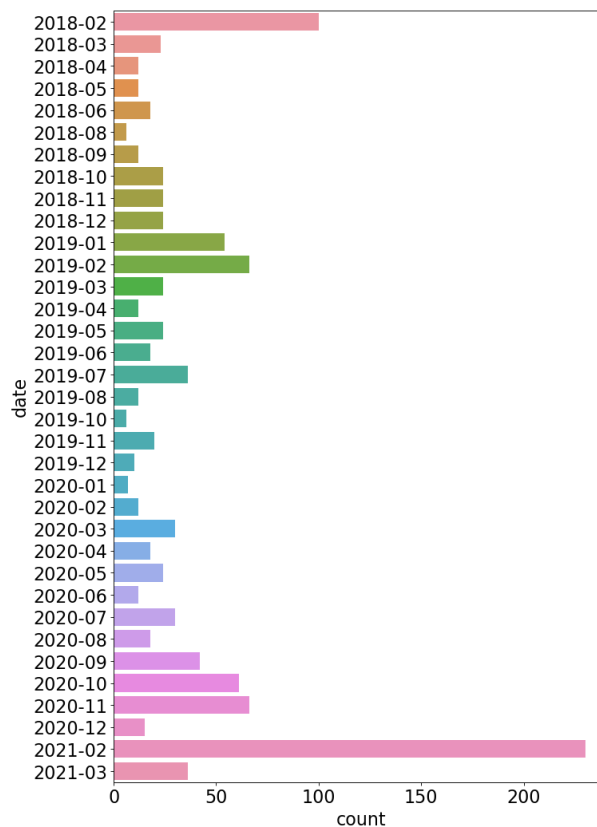
### 3.1 Data Crawling Results

The results of the crawling provide 1137 posts from hundreds of accounts in Instagram. Based on the data of the top 20 usernames that are actively post about #uuite, it can be concluded that the #uuite topic is not artificially driven from a certain account. Instead, numerous civilian activist accounts like gue.netizen and cerita\_negeri187 comprised of 5% of the total posts that has been collected. This means that the #uuite topic is driven by the common consensus of the populace.

**Table 2** Top 20 Accounts Using the #uuite Hashtag on Their Posts

No	Username	Frequency
1	gue.netizen	44
2	cerita_negeri187	22
3	ditjenpktn	22
4	udinsapar.171	14
5	berita168	9
6	disperindag_provinsi_jambi	8
7	telaahhukum.id	7
8	intanpaulina_lol	7
9	inilah_com	7
10	redaksirealita	7
11	andddiez	7
12	safawiahmad123	6
13	safenetvoice	6
14	ardian.arif.96	6
15	humas.polsek_tegalrejo_yka	6
16	dionsrn151	6
17	jibrilsoeharjo	6
18	semuatentangbatam	6
19	herman_vista	5

Furthermore, the distribution plot shows an increasing trend recently regarding the law. The data shows that more than 200 posts using the hashtag #uuite in February 2021. This pattern can be considered as anomaly, in contrast to the data of previous months where the number of posts could reach less than 100.



**Figure 2** The Time Frequency of the Posts Using the #uuite Hashtags

### 3.2 Data Pre-processing Results

The results of the data crawling process were converted into a CSV file for reusability purposes. After that, the data was preprocessed to ensure compatibility with the algorithms. One of the obstacles during this process occurs when the texts contain excessive amounts of newlines. This problem would cause serious consequences for the next process if not handled properly. Thus, the removal of the newlines is necessary to ensure the readability of the data for the next process.

Furthermore, the python library has a built-in function to detect any missing and duplicated data throughout the data collection. Those function includes `pandas.DataFrame.Duplicated().sum()` and `pandas.DataFrame.isnull().sum()`. However, both functions return zero value when it was used to the collected data. This indicates that each data is unique and none of the data is missing. Loanwords also pose a challenge during the pre-processing phase of the dataset. This can be easily

mitigated by comparing each text with the `words()` function from the NLTK library. However, some technical terms around the usage of the internet, such as “upload”, “download”, and others, will not be removed. This decision is supported with the widespread usage of those terms, even in the networking industry [17].

The text normalization process was done using the dataset from [16]. Furthermore, the frequency of the text was calculated by loops(`for`) and comparisons (`if-else`). While there are 1147 slangs that was collected from the website, the results surprisingly show miniscule number of slangs used on each post. These slang words are then replaced with the formal form using the `pandas.DataFrame.at` variable. However, slangs tend to have intertwined meaning with a formal word. For example, *cabut* often refers as a slang for “going away”, while also serves as “remove” in formal connotation. This prompts manual inspection to ensure that such words are collected correctly as a slang. Words that are not accurately identify as slang is deleted. The appropriate shown at Table 3.

**Table 3** Slang Words Collected from the Dataset

<b>index</b>	<b>slang</b>	<b>formal</b>	<b>frequency</b>
<b>995</b>	sob	sahabat	5
<b>334</b>	galau	kacau	1
<b>375</b>	goblok	bodoh	1

Additionally, numerous posts have been misusing the `#uuite` hashtag to promote their products and services. Thus, these data need to be processed. The process of removing irrelevant posts involves identifying the index of the DataFrame which contains price tags. Specifically, any posts that contain “Rp.” were removed by identifying the index of the text. This process involves the usage of `iloc()` function from pandas library which returns the index of the data. Moreover, the conditionals were put inside the list to automatically obtain the specific row of the data.

The ‘date’ column of the dataset is also replaced with the appropriate format according to python notebooks. This includes the selection of characters within the ‘date’ column and transforming the dataset into the “datetime64” format. During the analysis, the data will provide insight regarding the frequency of the posts that use `#uuite` hashtag throughout the time.

**Table 4** Samples of Converted Data into datetime64

<b>No</b>	<b>date</b>
0	17/02/2021
1	13/10/2020
2	16/02/2021
3	13/03/2021

---

4	24/02/2021
5	12/03/2021
6	17/02/2021
7	26/03/2021
8	12/03/2021
9	29/03/2021
10	29/03/2021

---

Additionally, this process filters any unnecessary words that were present in the dataset. The removed words comprised of stopwords such as prepositions and pronouns. This process also removed any words related to date and time, including day, month, year, Sunday, January, etc. Any words related to the law (such as “uu”, “ite”, “pasal”, and “ayat”) are also removed from the dataset. While it seems counter-intuitive to remove these, the purpose of this paper is to summarize the primary concern of the populace regarding the law. Otherwise, the results will only center around the law itself instead of the problems regarding the law.

Due to the sensitivity to casing, the text in the dataset was converted into lowercase. This process is executed to ensure that the model would be able to analyze the topic without considering the casing of the characters. Finally, the dataset is ready to be analyzed using LDA Model.

### 3.3 LDA Implementation results

The pre-processed data were processed further using LDA. The data was first converted into a term dictionary using the Corpora function, which allocates each word using their respective id. Later, the algorithm creates the document-term matrix via topic modeling from the *gensim* library. This process also involves the creation of the LDA model using the *LdaModel* library. Additionally, the coherence of the main topic was calculated using the *CoherenceModel* library. It also provides the perplexity score to decide the optimal number of topics that can be shown. A lower score of perplexity indicates a good LDA model.

### 3.4 Data Analysis Results

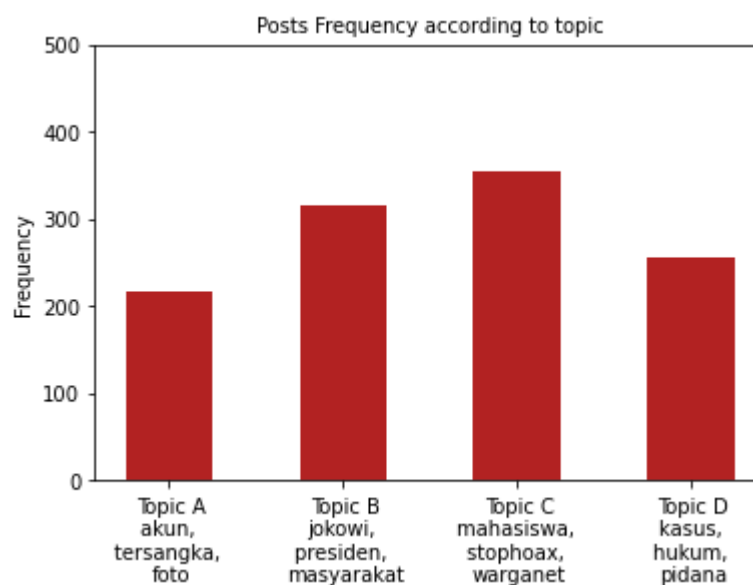
Data analysis process reveals the perplexity of the LDA model and the primary topics within the dataset. It includes the comparison between perplexity and coherence to decide how many topics inferred from the dataset. The ideal amount of topic is decided with the lowest perplexity score and the highest coherence score. The comparison between perplexity and coherence to the number of topics can be seen in Table 5.



**Table 5** Comparison of Perplexity and Coherence to the Amount of Topics

Topic amount	Perplexity	Coherence
2	-8.99186790420366	0.2689242314262426
3	-8.94251491693526	0.3488919964565997
4	-8.89033682048207	0.512773356902045
5	-8.82283471586389	0.46410981948450214

Table 5 shows decent differences of perplexity and coherence between each amount of topic. Two topic amount provides the highest perplexity score against every contender. However, the four topics amount has the highest coherence score than any other contenders in the table. Due to the lack of ideal candidates that could fill both criteria, it is decided that four topics are sufficient to present the dominant topics of the dataset. The decision is based on the gaps in the perplexity and the coherence of the topic amount. The perplexity of each topic amount does not have a noticeable gap amongst each other. In contrast, the coherence of each topic has an enormous gap between each topic amount. Thus, it is sensible that four topic number is sufficient to present the dominant topics in the dataset. The coherent topics are shown in Figure 3.

**Figure 3** Posts Frequency According to Topic Diagram

The main topic of the 1137 Instagram posts have been successfully collected. The detailed distribution of each topic can be seen at Table 6, which described each keyword in the topic and the frequency of the words.

**Table 6** Detailed Statistics of Each Dominant Topics

Dominant Topic	Keywords	Frequency
A	<i>akun, tersangka, foto</i>	216
B	<i>jokowi, presiden, masyarakat</i>	314
C	<i>mahasiswa, stophoax, warganet</i>	353
D	<i>kasus, hukum pidana</i>	255

To make a better connection between these keywords, a comparison is made between the keywords and the documents. The sample documents for each topic can be seen in Table 7.

**Table 7** Comparison between Topic Keywords and a Sample Post for Each Category

Dominant Topics	Keywords	Text
A	<i>akun, tersangka, foto, badak, video, hukum, jenar, pelaku, sragen, masih</i>	<i>['pesan', 'khusus', 'nih', 'bang', 'coki', 'pardede', 'gaes', 'bijak', 'aja', 'bersosial', 'media', 'posting', 'media', 'sosial', 'pasti', 'rekam', 'jejaknya', 'bersosial', 'media', 'seperlunya', 'saja', 'baik', 'gaes', 'ingat', 'media', 'sosial', 'bukanlah', 'buku', 'diary', 'darikuuntukyou', 'bersamakuyou', 'cokipardede', 'cokipardedereza', 'podcastdeddycorbuzier', 'komedian', 'standupcomedy', 'infoupdate', 'infoseleb', 'infounik', 'viral', 'viralpost', 'trendingtopic']</i>
B	<i>jokowi, presiden, masyarakat, kritik, revisi, berita, kata, pemerintah, polri, harus</i>	<i>['revisi', 'tak', 'dilakukan', 'tak', 'masuk', 'program', 'legislasi', 'nasional', 'prolegnas', 'prioritas', 'wacana', 'revisi', 'sebelumnya', 'mengemuka', 'setelah', 'presiden', 'joko', 'widodo', 'mengeluarkan', 'pernyataan', 'soal', 'rencana', 'revisi', 'eits', 'tapi', 'jangan', 'salah', 'presiden', 'memakai', 'kata', 'kalau', 'melontarkan', 'wacana', 'revisi', 'iniditeruskan', 'narasinewsroom', 'narasinewsroom', 'news', 'beritaterkini', 'undangundang', 'revisi', 'jokowidodo', 'jokowi']</i>
C	<i>mahasiswa, stophoax, warganet, netizen, lucu, viralkan, stopisusara, rizieqpenipu, stopadudombai slam, mncbabel</i>	<i>['akui', 'isu', 'nasional', 'vol', 'part', 'polisi', 'virtual', 'bebas', 'bersuara', 'upaya', 'preventif', 'polisi', 'virtual', 'terhadap', 'kasus', 'pidana', 'digital', 'ternyata', 'membawa', 'dampak', 'cukup', 'signifikan', 'bagi', 'mahasiswa', 'kebebasan', 'berekspresi', 'serta', 'hak', 'hak', 'privasi', 'mahasiswa', 'telah', 'direnggut', 'demi', 'tercapainya', 'dunia', 'digital', 'sehat', 'produktifsebagian', 'mahasiswa', 'unpar', 'pun', 'merasa', 'resah', 'lantaran', 'takut', 'bentuk', 'ketidakadilan', 'jelas', 'pelaksanaannya', 'kecemasan', 'kian', 'membesar', 'sebagian', 'mahasiswa', 'unpar', 'masih', 'mempertanyakan', 'karet', 'lantas', 'diutarakan', 'mahasiswa', 'unpar', 'menanggapi', 'hal', 'selengkapnya',</i>

'disimak', 'melalui', 'penjelasan', 'atas', 'kebermanfaatan',  
'unparkebermanfaatan', 'virtualpolice', 'polisivirtual',  
'prokontra', 'medsos', 'cybercrime', 'beritanasional',  
'unpar', 'mahasiswa']

D kasus, hukum, [ 'repost', 'rockygerungofficialterkait', 'pengacara',  
pidana, 'hotman', 'paris', 'hutapea', 'menyurati', 'presiden', 'joko',  
penjara, artis, 'widodo', 'jokowi', 'menko', 'polhukam', 'mahfud', 'serta',  
dhani, 'komisi', 'iii', 'dpr', 'isi', 'suratnya', 'hotman', 'meminta',  
tersangka, baik, 'pemerintah', 'dpr', 'menghapus', 'itesurat', 'diunggah',  
polda, kepada 'hotman', 'paris', 'akun', 'instagram', 'pribadinya', 'sabtu',  
'hotman', 'meminta', 'agar', 'kasus', 'pencemaran', 'nama',  
'baik', 'semula', 'masuk', 'unsur', 'pidana', 'dijadikan',  
'unsur', 'perdata', 'usulan', 'kopi', 'joni', 'pemerintah', 'agar',  
'dihapus', 'agar', 'pencemaran', 'nama', 'baik', 'dijadikan',  
'murni', 'perdata', 'tulisi', 'hotman', 'caption', 'akun',  
'instagramnya', 'setelah', 'berkirim', 'surat', 'diketahui',  
'hotman', 'hari', 'bertemu', 'mahfud', 'berdiskusi',  
'kesempatan', 'hotman', 'mempertemuan', 'mahfud', 'salah',  
'satu', 'warga', 'korban', 'kasus', 'bernama', 'vivi', 'nathalia',  
'vivi', 'adalah', 'salah', 'seorang', 'mantan', 'terpidana',  
'kasus', 'gal', 'mahfudmd', 'hotmanparis']

Based on Table 7, the first topic of the hashtag focused on the usage of social media. This is supported with certain keywords from the dominant topic, such as accounts, photos, and videos. Moreover, the sample post of topic contains words such as “bersosial”, “media” and “posting”. This topic is related to 216 posts.

The second topic focused on the president of Indonesia's decision to amend the law. The sample of the topic talks about how the president decides not to consider the amendment as part of his national programs. From this topic, it can be concluded that the public is consistently waiting for the government to take any action. The sample post is correlated with some keywords in the topics, such as “Jokowi”, “Presiden”, and “revisi”. The sample post has the topic correlates with 315 other posts from the dataset.

The third topic focused on college students' actions to pressure the government to amend the law. The sample is filled with the students' resentment of the law. They feel that the law could potentially limit the freedom of speech and expression due to the vague interpretation of the law. Furthermore, they also argue that the law exposes the injustice within the system for its failure to protect innocent people while the criminals can walk free from repercussion. The sample post has linguistic correlation with certain keywords of the topic, such as “mahasiswa”, “warganet”, and “netizen”, which coincides with “mahasiswa”, “digital”, and “virtual” from the sample post. This topic has the highest frequency of all topics, reaching 353 posts related to the topic.

Finally, the last topic focused on an Indonesian lawyer named Hotman Paris. As one of the most influential lawyers in Indonesia, he decided to send a message to the president of Indonesia. The

message urges the president to amend the law due to the rising number of victims. His action garners positive sentiment amongst the public, with 255 posts from the dataset are related to this topic. The sample post also has close correlation with the keywords of the topic, such as “kasus”, “hukum”, and “pidana”, which is within the scope of his career.

## **4 Conclusions and Recommendations**

### **4.1 Conclusions**

Based on the research that was conducted, four primary topics mold the perception of the populace regarding The Electronic Information and Transaction Act. To start with, the first topic primarily revolves around the usage of social media. Based on the sample post of the topic, it can be summarized that the law raises concern amongst the public on how they could use their social media. Furthermore, the law sparks discontent among the students, which becomes the driving force of the #uite hashtag. Based on the sample post that was related to the topic, it can be summarized that the law could undermine the way how they can express their opinion on the digital landscape. Additionally, the third topic focuses on the public complaints regarding the undermining attitude from the government to amend the law. Finally, the law attracts the attention of the legal industry. Where an influential Indonesian lawyer named Hotman Paris becomes the center of attention. He supports the notion to revise the law, proven by his willingness to pressure the president to change the law. These four topics allow future scholars and the government to identify the problems within the law more effectively. Furthermore, future scholars can suggest better recommendations to the government regarding the law, which could provide meaningful feedback to the government. Hopefully, these topics could help the government on enacting better policies in the future.

### **4.2 Recommendations**

There are numerous rooms for improvement that can be beneficial for future research. The details are described as follows:

1. It is recommended for future researchers to find a more effective data crawling method other than selenium. The usage of selenium prompts the user to constantly load the entire page for data crawling, which consumes a lot of bandwidth. Thus, an alternative method needs to be found to collect more data efficiently.
2. Data crawling activities require a consistent success rate of packet delivery. Thus, an adequate internet facility is required. Any failure to fulfill these criteria could cause connection timeouts. Consequently, the data crawling process needs to be repeated.
3. Instagram often conducts experimental research to collect user feedback regarding their design. This resulted in an inconsistent design template for each Instagram user. Any future

research that uses selenium should consistently modify the methods from any previous research to accommodate the data crawling process in the future.

## REFERENCES

---

- [1] Asosiasi Penyelenggara Jasa Internet Indonesia, "LAPORAN SURVEI INTERNET APJII 2019 – 2020 (Q2)," 2020.
- [2] Kementerian Komunikasi dan Informatika Republik Indonesia, "Rancangan Rencana Strategis Kemenkominfo 2020-2024."
- [3] B. Galih, "Dandhy Dwi Laksono Ditangkap Polisi atas Tuduhan Menebarkan Kebencian Halaman all - Kompas.com," 2019.  
<https://nasional.kompas.com/read/2019/09/27/00462591/dandhy-dwi-laksono-ditangkap-polisi-atas-tuduhan-menebarkan-kebencian?page=all> (accessed Apr. 09, 2021).
- [4] CNN indonesia, "Kronologi Kasus Baiq Nuril, Bermula dari Percakapan Telepon," 2018. <https://www.cnnindonesia.com/nasional/20181114133306-12-346485/kronologi-kasus-baiq-nuril-bermula-dari-percakapan-telepon> (accessed Apr. 09, 2021).
- [5] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, no. 4–5, pp. 993–1022, 2003, doi: 10.1016/b978-0-12-411519-4.00006-9.
- [6] A. Rahmawati, N. L. Nikmah, R. D. A. Perwira, and N. A. Rakhmawati, "Analisis topik konten channel YouTube K-pop Indonesia menggunakan Latent Dirichlet Allocation," *J. Ilm. Sist. Inf.*, vol. 11, no. 1, pp. 16–25, Jan. 2021, doi: 10.26594/teknologi.v11i1.2155.
- [7] J. Xue, J. Chen, C. Chen, C. Zheng, S. Li, and T. Zhu, "Public discourse and sentiment during the COVID 19 pandemic: Using latent dirichlet allocation for topic modeling on twitter," *PLoS One*, vol. 15, no. 9 September, Sep. 2020, doi: 10.1371/journal.pone.0239441.
- [8] D. Fang, H. Yang, B. Gao, and X. Li, "Discovering research topics from library electronic references using latent Dirichlet allocation," *Libr. Hi Tech*, vol. 36, no. 3, pp. 400–410, 2018, doi: 10.1108/LHT-06-2017-0132.
- [9] O. Toubia, G. Iyengar, R. Bunnell, and A. Lemaire, "Extracting Features of Entertainment Products: A Guided Latent Dirichlet Allocation Approach Informed by the Psychology of Media Consumption," *J. Mark. Res.*, vol. 56, no. 1, pp. 18–36, 2019, doi: 10.1177/0022243718820559.
- [10] Y. Guo, S. J. Barnes, and Q. Jia, "Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation," *Tour. Manag.*, vol. 59, pp. 467–483, 2017, doi: 10.1016/j.tourman.2016.09.009.
- [11] K. Bastani, H. Namavari, and J. Shaffer, "Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints," *Expert Syst. Appl.*, vol. 127, pp. 256–271, 2019, doi: 10.1016/j.eswa.2019.03.001.
- [12] S. (Sixue) Jia, "Toward a better fitness club: Evidence from exerciser online rating and review using latent Dirichlet allocation and support vector machine," *Int. J. Mark. Res.*, vol. 61, no. 1, pp. 64–76, 2019, doi: 10.1177/1470785318770571.
- [13] Q. Zhang, S. Liu, D. Gong, and Q. Tu, "A latent-dirichlet-allocation based extension for domain ontology of enterprise's technological innovation," *Int. J. Comput. Commun. Control*, vol. 14, no. 1, pp. 107–123, 2019, doi: 10.15837/ijccc.2019.1.3366.
- [14] Y. Wang and L. Xu, "Unsupervised segmentation of greenhouse plant images based on modified Latent Dirichlet Allocation," *PeerJ*, vol. 2018, no. 6, 2018, doi:

- 10.7717/peerj.5036.
- [15] P. Shah, D. Sharma, and R. Sekhar, "Analysis of Research Trends in Fractional Controller Using Latent Dirichlet Allocation," *Eng. Lett.*, vol. 29, no. 1, 2021.
  - [16] "Lampiran: Bahasa gaul dalam bahasa Indonesia - Wiktionary bahasa Indonesia." [https://id.wiktionary.org/wiki/Lampiran: Bahasa\\_gaul\\_dalam\\_bahasa\\_Indonesia](https://id.wiktionary.org/wiki/Lampiran: Bahasa_gaul_dalam_bahasa_Indonesia) (accessed Jul. 27, 2021).
  - [17] "Telkom." [https://www.telkom.co.id/sites/home-service/id\\_ID/page/paket-3p-\(internet+tv+phone\)-701](https://www.telkom.co.id/sites/home-service/id_ID/page/paket-3p-(internet+tv+phone)-701) (accessed Jul. 28, 2021).