Sentiment analysis of hoax news toward the election 2019 based on student perspective

IOP Conf. Series: Journal of Physics: Conf. Series **1217** (2019) 012125 doi:10.1088/1742-6596/1217/1/012125

Sentiment analysis of hoax news toward the election 2019 based on student perspective

D G F A Sumardi[1], Y Nurmalasari[2], B D Kurnianto[3], A Kesumawati[4]
[1,2,3,4] Department of Statistics, Faculty of Math and Science, Universitas Islam Indonesia, Yogyakarta, Indonesia.
Email: 15611101@students.uii.ac.id

Abstract. In the era of science and technology development and the high number of social media usage in Indonesia instead of becoming a media provocation and hoax dissemination media that is troubling and judged can be a severe threat to national stability. Students are referred to as learned society in the hope of being a golden generation who is always capable of being a change agent in the structure of society. Therefore, in this research, the source of data is UII student's opinion that was collected, using sentiment analysis, the researcher wants to know how the sentiment based on their response to the hoax news ahead of the 2019 election. Which hopefully can give solution and anticipate the heat of the political situation in Indonesia ahead of the next election 2019 and can educate students to care about social and political problems in Indonesia for the common good. Based on 395 respondents, it was concluded that UII student sentiment toward hoax news spread before the 2019 election was negative is 72,1%. The conclusion is the sentiment of UII students tends to be negative towards the spread of hoax news ahead of the 2019 election.

1. Introduction

Today, the development of information technology in Indonesia is proliferating, according to online news tribunnews [1] President, Jokowi said that internet users in Indonesia today are 132 million, or 52% of the total population of Indonesia. Based on this number 129.2 million have active social media accounts. Science, technology, and various means used to educate the nation, freedom of unity and unity, and as "free learning," but in the development, development of science and technology and the proliferation of social media in Indonesia become media and media of deception. Lately, hoaxes have become a serious conversation in Indonesia along with the rampant false news in the social media world, its existence is increasingly troubling and can be the basis for federal supervision. Based on empirical experience, in Indonesia hoax news circulations appear at strategic times, in this case, ahead of the ELECTION is the main target for discussing hoax news whose contents can mutually load groups or individual candidates who will nominate themselves. Based on the problems that exist, there needs to be assisted by students as social managers. Darmayadi [2] Students' social participation becomes more because of the assumption of having a better understanding concept as part and level of learning at the college level. In this research entitled "*Sentiment Analysis of Hoax News Toward the Election 2019 Based on Student Perspective*" hope can be to know the sentiment of student responses to the spread of hoax news ahead of the 2019 ELECTION and anticipate the heat of the political situation in Indonesia ahead of the 2019 ELECTION by educating students to care about social and political issues in Indonesia for the sake of mutual benefit.

2. Literature Review

Based on previous research, by Nurhuda F [3] entitled Analysis of Community Sentiment towards 2014 Indonesian Presidential candidates based on Opinion from Twitter Using the Naive Bayes Classifier Method. In their research, sentiment analysis of 2014 Indonesian candidates and vice presidents was carried out through Twitter social networking. Where the stages used, including data collection, data pre-processing, POS tagging, and opinion extraction using the naive Bayes classifier method. While in this study the method used is manual using an algorithm generated after the input sentiment dictionary. Furthermore, based on previous research by Rozi I F [4] entitled the Implementation of Opinion Mining for Extraction of Public Opinion Data at Higher Education. In their research, an analysis of public opinion was carried out in higher education which has stages of research including POS tagging using the Hidden Markov Model; then opinions are classified into negative and positive opinions using the Naive Bayes Classifier to get the value of public sentiment in college.

3. Methodology

3.1. Research Sources

The data source is in the form of primary data directly obtained from respondents (UII students) who have filled out the questionnaire. As for the sampling time that carried out for two weeks at the beginning of July, taken at the Indonesian Islamic University campus in Yogyakarta with the respondents is the student of UII in D3 and S1 program for all majors.

3.2. Data Collection

In this study used the data collection is done by collecting answers from online questionnaires that have been distributed to respondents, the researcher asked respondents to fill out questionnaire forms in the form of respondents' opinions on the spread of hoax news ahead of the 2019 election.

3.3. Research Method

The research method that I do is an experimental research method. The research stages presented in the form of a flow chart as follows:

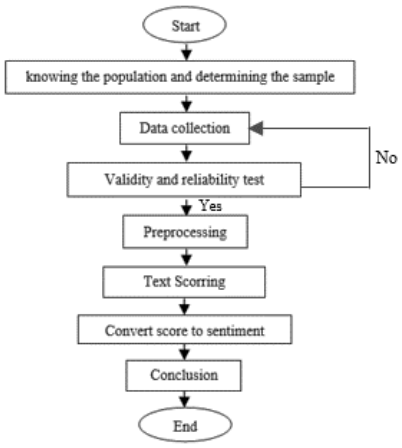

Figure 1. Research flow chart

3.4. Population and sample

Sampling in this study was carried out using the Slovin formula, according to Narimawati [5] the formula is as follows:

$$n = \frac{N}{1+Ne^2} \quad (1)$$

Where,
n = number of samples;
N = number of population; and

e = tolerated error (5%)

3.5. Validity test

Validity according to Sugiyono [6] is "a measure that shows the degree of determination between data that occurs in objects with data that can be collected by researchers." The test of statistics as follows, Validity test decision above is obtained from the comparison of Pearson correlation value (R count) with (R table) or r product moment table, which is at a significance level of 5% and the number of samples is 392 or rounded 400.

3.6. Reliability test

Reliability according to Sugiyono [7] is a series of measurements or a series of measuring instruments that have consistency when the measurements made with the measuring instrument carried out repeatedly. According to Santoso in Niswah [8] the value of reliability is done by comparing the reliability coefficient (alpha) with (r table), the test of statistics is alpha > r table so, the question is considered reliable.

3.7. Preprocessing

This stage is the initial stage of text mining, which includes all routines, and processes for preparing data that will be used in knowledge discovery operations of text mining systems. As for the action taken at this stage is to erase the character and delete the word additions or words that are not meaningful.

3.8. Sentiment Analysis

According to Haddi in Muthia [9] sentiment analysis is the task of finding the opinion of the author about a particular entity. Sentiment analysis in the review is the process of investigating product reviews on the internet to determine opinions or feelings towards a product as a whole. According to Haddi in Muthia, sentiment analysis is treated as a classification task, which classifies the orientation of text into positive or negative.

3.9. Text Scoring

This section is the essential words are match with the dictionary that has been inputted, using 1473 positive words and 2960 negative words, where from the results of the next scoring will be converted into positive or negative sentiments.

3.10. Convert Score to Sentiment

The classification was carried out from the scores obtained, following the results of the classification of text into sentiments, where if the sentence based on text scoring is negative, it will enter the classification of negative sentiments, and if it is positive, then the sentence is in a positive classification.

4. Result and Discussion

4.1. Population and sample

Based on data obtained from the academic division, UII rectorate, it has known that the active student population of UII 2017/2018 school year is as many as 20,669 students. Sampling in this study was carried out using the Slovin formula, according to Narimawati [5] with n = number of samples; N = number of population; e = tolerated error (5%), using formula (1) we get the number of samples needed is equal to 392.4 or rounded to 392 respondents.

4.2. Descriptive Data

In this section, the characteristics of respondents who have filled out online questionnaires are distributed, based on the results of the survey, of 395 students, it is known that the number of respondents is dominated by women with a ratio of 34.7% of men, while 65.3% of female respondents.

Figure 2. Respondents' results based on gender

Then for the origin of the faculty, the respondent obtained the results are as follows,

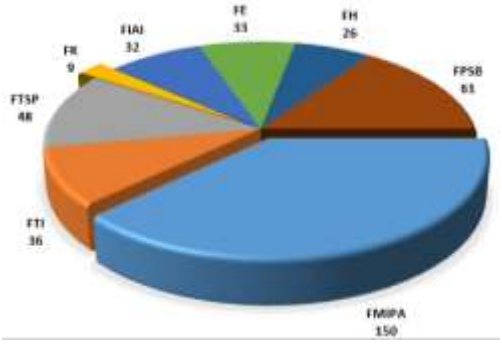

Figure 3. Respondents' results based on faculty

Based on the diagram above, of the 395 respondents, dominant came from FMIPA as many as 150 and the least was from FK as much as nine students. Then, because of respondents' opinions on the spread of hoax news ahead of the 2019 election, of the 395 respondents obtained the following responses,

1	Berikan tanggapan anda tentang penyebaran berita hoax menjelang pemilu 2019 (min 1 kalimat)
2	apa apaan ya gitu tuh, kurang terpuji, dan tidak sehat, meskipun wajar, tetap saja hal seperti itu akan merugikan dirinya sendiri dan juga orang lain. sebaik
3	Siapapun yang menyebarkan berita hoax tersebut hendaknya memperhatikan situasi dan kondisi. Untuk para pembaca atau pendengar jangan asal percay
4	semua berita yang didapatkan harus ditelusuri terlebih dahulu sebelum menyebarkan berita tersebut. jadilah pembaca yang bijak agar tidak menimbulka
5	Berita Hoax menjelang pemilu semakin banyak dengan sasaran orang awam mengenai pengetahuan politik ataupun generasi milenial yang terkadang per
6	Penyebaran berita hoax sangat merugikan kita dapat menimbulkan fitnah dan buruk sangka
7	Dengan terselenggaranya pemilu serentak 2018 di setiap daerah tentu saja banyak calon wakil rakyat (katanya) memiliki strategi dengan cara mengeksisto
8	Tidak terlalu meperhatikan politik
9	saya tidak pernah menemukan berita hoax
10	Saya belum ada baca berita tentang pemilu maaf ya mbakk
11	Banyak berita hoax yang tersebar di kalangan masyarakat mengenai calon pemimpin yang akan mengganggu jalannya pemilihan yang objektif, terlebih jik
12	Karena saya tidak terlalu menyukai politik jd saya tdk perduli dngan isu apapun ttg pemilu 2019. Tanpa isu pun sudah dpt diketahui seperti apa calon presi
13	Sebenernya aku gak suka tentang politik gtu, tapi menurut ku kalau ada berita hoax tentang pemilu kemungkinan juga ada orang terselubung yang ingin m
14	Menurut pendapat saya untuk mengurangi berita - berita hoax menjelang pemilu ini, sebaiknya sering dilakukan sosialisasi kepada masyarakat tentang pe
15	Tidak terlalu memperhatikan dan mengerti politik

Figure 4. Sample results of the respondent's opinions

Then the data obtained will be analyzed with text mining to processing the text data.

4.3. Validity test

To test the validity, the Pearson correlation technique was used, with the help of Excel, the Pearson correlation results obtained as follows,

Table 1. Validity Test

Questions	Correlation *Pearson score*	Test of Statistic *(α=5%,df=393)*	Conclusion
Item 1	0.417	$0.417 > 0.098_{(Rtab)}$	Valid
Item 2	0.496	$0.496 > 0.098_{(Rtab)}$	Valid
Item 3	0.593	$0.593 > 0.098_{(Rtab)}$	Valid
Item 4	0.574	$0.574 > 0.098_{(Rtab)}$	Valid
Item 5	0.354	$0.354 > 0.098_{(Rtab)}$	Valid

Table 1. R table used is equal to 0.098. The conclusion of this test is valid because R counts or *Pearson correlation* in item1 to item5 is higher than the R table or 0.098.

4.4. Reliability test

In this study, the reliability test was used using alpha, using the IBM SPSS program assistance, obtained reliability test results as follows,

Table 2. Reliability Test

Cronbach's alpha value	Test of statistics *(α=5%,df=393)*	Conclusion
0.174	$0.174 > 0.098_{(Rtab)}$	Reliable

Table 2. We know that data was reliable, because the number of Cronbach's alpha greater than r table, so the question questionnaire is considered reliable.

4.5. Preprocessing

Next stage is preprocessing, in this stage only using data from the student's opinion. In this stage, data will be processed to clean the word and eliminate the emoticons or punctuation in the sentence of the opinion; the following are examples of preprocessing results that done,

Table 3. Preprocessing Result

	Before	*After*
1	"apa apaan ya gitu tuh, kurang terpuji, dan tidak sehat, meskipun wajar, tetap saja hal seperti itu akan merugikan dirinya sendiri dan juga orang lain. sebaiknya emang harus ada upaya untuk memberantasnya, tapi gimana ya"	"kurang terpuji, tidak sehat, meskipun wajar, tetap saja akan merugikan orang lain. sebaiknya ada upaya untuk memberantasnya"
2	"b aja :)"	-
3	"Tidak terlalu memperhatikan politik"	"tidak memperhatikan politik"
4	"#GantiPresiden2019"	"gantipresiden2019"
…	…	…
395	"Kalau dari saya tidak pernah menemukan berita hoax"	"saya tidak pernah menemukan berita hoax"

4.6. Text Scoring

In this section, the net data generated from the pre-processing process then classified into score positive and negative. In this section text scoring is done on each sentence based on each word element that is owned, which is done by matching the base words with the input dictionary, using 1473 positive words and 2960 negative words from the system, from which the next scoring results will be converted into sentiments positive or negative.

Table 4. *Text Scoring Result*

	score	***Text***
1	-9	"kurang terpuji, tidak sehat, meskipun wajar, tetap saja akan merugikan orang lain. sebaiknya ada upaya untuk memberantasnya,"
2	0	"Tidak meperhatikan politik"
…	…	…
395	-2	"saya tidak pernah menemukan berita hoax"

4.7. Convert score to sentiment

This classification then carried out from the scores obtained, from the results of the text to negative or positive sentiments,

Table 5. Sentence Classification Result

	Classification	*score*	*Text*
1	Negative	-9	"kurang terpuji, tidak sehat, meskipun wajar, tetap saja

			akan merugikan orang lain. sebaiknya ada upaya untuk memberantasnya,"
2	Positive	0	"Tidak meperhatikan politik"
…	…	…	…
395	Negative	-2	"saya tidak pernah menemukan berita hoax"

Table 5. Based on the results obtained, it is known that the negative sentiment is the result of the conversion of negative values or less than zero, while the positive sentiment is the result of the conversion of positive values or greater than zero. The results of this sentence classification then used to conclude the sentiments.

4.8. Sentiment Result based on responses to hoax news spread before the 2019 election

By the purpose of sentiment analysis, out of 395 respondents, it was concluded that the tendency of student responses to the spread of hoax news ahead of the 2019 election was negative, while the comparison of sentiments obtained was presented in the graph in the following,

Figure 5. Comparison of sentiment results based on student responses

Then the separation of each word from the existing responses made by a word cloud. According to Qeis [8], word cloud is a system that can visualize words based on the frequency of existing words. So that emotional students can quickly know when responding to the spread of hoax news ahead of the 2019 ELECTION, the following word cloud results have obtained.

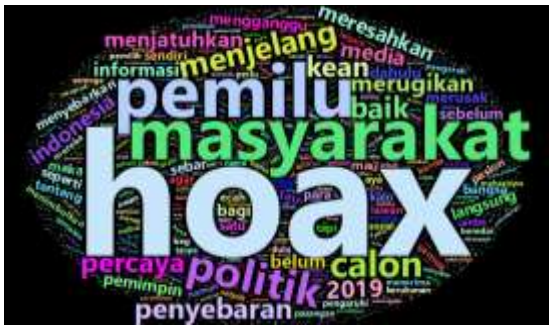

Figure 6. Word cloud result

Based on word cloud results, we have known that student responses have a critical and negative emotional tendency. Words that describe it such as "dropping", "damaging", "breaking", "influencing", "disturbing", "harming", etc.

5. Conclusion

Based on 395 student respondents, most of the students who gave their opinions were from FMIPA is 150 students and at least 9 were from FK just nine students, and female respondents were dominated by 65.3% while men were only 34.7%. So, by using the algorithm of sentiment analysis, it concluded that the tendency of student responses to the spread of hoax news ahead of the 2019 election was negative at 72.2%. Moreover, then, there is recommendation which hope can be improved by the next researcher; the existing suggestions are separating student groups based on their academic type,

increase the number of samples based on faculty, and make more specific questions related to the aspects to be studied.

Acknowledgments
Thank you to the ristekdikti and Universitas Islam Indonesia who have helped fund this research, and to Ms.Ayundyah Kesumawati, S.Si., M.Si as the supervisor who has provided direction in this research process, as well as friends who have helped

References

[1] Tribunnews 2016 *Jokowi Minta Tegakkan Hukum Terhadap Media Online Yang Produksi Berita Bohong dan Fitnah.* [*Online*] Available at: http://www.tribunnews.com/nasional/2016/12/29/jokowi-minta-tegakkan-hukum-terhadap-media-online-yang-produksi-berita-bohong-dan-fitnah.
[2] Darmayadi, A 2011 *Jurnal Unikom* **9**
[3] Nurhuda F, Sihwi S W, and Doewes A 2013 *Jurnal ITSMART* **2** (2)
[4] Rozi I F, Pramono S H, and Dahlan E A 2012 *Jurnal EECCIS* **6** (1)
[5] Narimawati U 2010 *Dasar Penyusun Penelitian Ekonomi* (Jakarta: Genesis) p 38
[6] Sugiyono 2015 *Metode Penelitian Kuantitatif Kualitatif R&B* (Bandung: AlFABET)
[7] Sugiyono 2005 *Memahami Penelitian Kualitatif* (Bandung: ALFABET)
[8] Niswah, U 2016 *Hasil dan Pembahasan Validitas dan Reliabilitas* [Online] Available at: http://repository.umy.ac.id/bitstream/handle/123456789/2245/8.%20BAB%20IV.pdf?sequence=8&isAllowed=y.
[9] Muthia D A 2016 *Jurnal Teknik Komputer* **2** (1)