Online News Analysis of Majene Public Figure Electability with NLP (Natural Language Processing)

Sundari1*, S Cokrowibowo1, and N Zulkarnaim1

1Departement of Informatics, Faculty of Engineering, Universitas Sulawesi Barat, Majene, Indonesia

*Email: sundarisyukur1998@gmail.com

Abstract. Electability is one of the concerns as public figures who will take part in the elections, various things are done to improve their electability. Now, there are many methods to measure electability, one of which is the Natural Language Processing method. In this study, an electability analysis of public figures was conducted on online news to determine the electability level of a public figure. Where information obtained by the public is through online news sites. According to the Indonesian Internet Service Providers Data (APJII) in 2014, internet users in Indonesia will reach 107 million. This number has become one of the potential Public Figures, both institutionally and individually in capturing their masses through new media. 100 million of the masses will be contested by candidates who will compete in the political stage. Therefore, candidates who will advance will carry out digital political imaging. The public figures used in this study are 12 figures who are rumored to be going forward in the 2020 elections based on the online news site MAJENE 2020 ELECTION.

1. Introduction

Advances in science and technology, especially in the field of computer science make information exchange can be done anywhere and anytime. This causes an abundance of information circulating on the internet [1]. The internet is the second medium that is often used by people after television. This can be seen from internet users in Indonesia, which continue to increase from year to year. One of the information obtained by the public is through an online news site. Hall predicts that within five years there will be more people in developing countries accessing news from the internet than newspapers.

Electability is a political compilation of unpopular elections and can be a benchmark for measuring the level of one's electability. Electability is a concern as a public figure who will participate in the elections, various things are done to increase their electability. According to the Indonesian Internet Service Providers Data (APJII) in 2014, internet users in Indonesia will reach 107 million. This number is one of the potentials for election participants both institutionally and individually in capturing their masses through new media. 100 million of the masses will be contested by candidates who will compete on the political stage. Therefore, candidates who will advance will carry out digital political imaging. Imaging is done through various tactics; fanpage, twitter, blog, mainstream online media publications, and other social media [2].

The media has become one of the important tools used to increase electability, one of which is the online news media. News can be used as an assessment of political figures in the form of electability analysis. However, it is not easy to analyze the electability of various online news, because news
contains unstructured texts, especially in Indonesian texts. Text pre-processing in text mining is an important part of getting the basic information contained in the news.

The method for processing text on online news to obtain information on the electability of a public figure can be done with the Natural Language Processing (NLP) method. Natural Language Processing in its application relates to how computers can be used to understand and manipulate natural language texts (Natural Language) to obtain certain information. With this natural language intermediary, humans can interact with computers [3].

The public figures used in this study are 10 figures who are rumored to be going forward in the 2020 elections based on the online news site Majene Pilk 2020. Of the 10 figures will be analyzed their electability based on online news ahead of the upcoming Majene elections in November 2020.

2. Literature Review

2.1. Electability
Electability is someone's interest in choosing. In terms of electability electability is the result of the absorption of English, namely "electability" which means electability, electability can be applied to goods, services, or people, agencies, or parties [2].

2.2. Online News
Online News is online news or news presented on the internet media, including news sites. Online news is a new type of news after the news is presented in print media (newspapers, magazines) in the form of text and images and in broadcast media (radio, television) in the form of audio and video [4].

2.3. Natural Language Processing (NLP)
Natural Language Processing commonly called Natural Language Processing is the Field of Artificial Intelligence that deals with understanding human (natural) language (English, German, Indonesian and others) how computers can be used to understand and manipulate natural language texts (natural language) for get certain information [3].

3. Research Methodology

3.1. Research Stages
The stages of the research to be carried out are:
- Identification of problems
  At this stage, the problem will be identified and how to implement the Natural Language Processing (NLP) method in conducting electability analysis of majene public figures based on online news.
- Data collection
  The data that will be used in this research is online news contained on the website about majene public figures who were rumored to advance in the regent election in 2020 in the Majene district online media. Because the amount of news obtained is very large and complex, to facilitate electability online analysis, the Natural Language Processing (NLP) method is used.

4. System Design
This research uses testing data of 100 web pages, it is assumed that the data used is a collection of news about majene public figures who are rumored to be going forward in the regent election in 2020 in the Majene district online media. Because the amount of news obtained is very large and complex, to facilitate electability online analysis, the Natural Language Processing (NLP) method is used.
Based on Figure 1, it can be seen that the processing of online news data uses the Natural Language Processing (NLP) method, which will be described in the explanation below.

4.1. Data Collection
Data collected is online news contained on online media sites. All data related to the keyword "MAJENE ELECTION 2020" the news data will be extracted, then through processing such as case folding, tokenizing, filtering, electability level calculation which will produce electability percentage of public figure majene.

4.2. Case Folding
All online news data that has been extracted will produce text, where the text is processed in the case folding stage which is replacing all capital letters in the text into lowercase letters [5].

4.3. Tokenizing
At this stage, processing will divide text documents in the form of sentences or paragraphs into tokens [5].

4.4. Filtering
Filtering processing will take words that are considered important and get rid of all the words that are not needed, this stage is known as Stopwords [5].

4.5. Calculating Electability
After processing is done using the Natural Language Processing (NLP) method, the final result will be determined in the form of the calculation of the level of electability of the public figure itself.
5. Results and Discussion

This research began by extracting online news data using the keyword "MAJENE ELECTION 2020".

Figure 2. Online news extraction results

Can be seen in Figure 2 is the result of online news data extraction using the keyword "MAJENE ELECTION 2020" where all online news containing these keywords will be processed in the form of online news extraction. After the extraction is complete, then specify keywords the names of majene public figures. This keyword is proposed to be able to do Processing with the Natural Language Processing (NLP) Method.

Table 1. Keywords public figure names

| Public Figure         | Keywords          |
|-----------------------|-------------------|
| Fahmi Massiara        | Fahmi, Massiara   |
| Lukman                | Lukman            |
| A Achmad Syukri Tammalele | Tammalele     |
| Itol Syaifil Tonra    | Itol, Tonra       |
| Harun                 | Harun             |
| Asnuddin Sokong       | Sokong            |
| Saggaf Katta          | Katta             |
| Muhammad Irfan Syarif | Irfan, Syarif     |
| Amril S               | Amril             |
| Idris                 | Idris             |
Based on Table 1, there are several keywords for each name of a Majene community figure that will make it easy to determine the level of electability. This keyword is obtained from the names of community leaders and greeting names that are also conveyed to the leaders themselves. This keyword will be used to carry out the Case Folding, tokenizing and filtering process. After processing is done using the Natural Language Processing (NLP) method, the final result will be determined in the form of calculating the level of public figure election using the diagram below.

The diagram in Figure 3, shows the percentage of electability of majene public figures, in which Fahmi Massiara obtained (33%), Lukman (18%), Muhammad Irfan Syarif (11%), Itol Syaiful Tonra (10%), Idris (7%), Harun (6%), Asnuddin Sokong (5%), A Achmad Syukri Tammalele (4%) are the same as Amril S (4%) and Saggaf Katta (2%).

6. Conclusions
In the description above, it can be concluded that the electability analysis of Majene public figures who were rumored to advance in the 2020 elections on online news can be known through training data in the form of news from online media and processed using natural language processing methods.
References

[1] F. D. D. Hakimi, “Public Sentiment Analysis System on the Election of East Java Regional Head 2018 on Twitter Documents Using Naive Bayes Classifier”, 2018. [In Bahasa]

[2] N. Indrawati, “Natural Language Processing (NLP) Indonesian as a Preprocessing in Text Mining.”. [In Bahasa]

[3] D. Rustandi, “Indonesian Natural Language Processing (Nlp) as Preprocessing in Text Mining Online Political Imaging: Strategies to Win Digital Mass Towards 2014 Election”, 2013. [In Bahasa]

[4] S. Kurniawan, W. Gata, D. A. Puspitawati, Nurmalasari, M. Tabrani, K. Novel, “Comparison of Classification Methods of Analysis of Political Figure Sentiments in Online News Media Comments”, Jurnal Sistem dan Teknologi Informasi, vol. 3 no. 2, 2019, p. 176-183. [In Bahasa]

[5] DosenPendidikan.Com, “Elektabilitas Adalah” [Online]. https://www.dosenpendidikan.co.id/elektabilitas-adalah/, [Accessed 4 December 2019].