

Analysis of Sentiments on Indonesian YouTube Video Comments: Case Study of The Indonesian Government's Plan to Move the Capital City

1st Risky Ade Maisal
Fakultas Ilmu Komputer
Universitas Indonesia
Jakarta, Indonesia
risky.ade@ui.ac.id

2nd Achmad Nizar Hidayanto
Fakultas Ilmu Komputer
Universitas Indonesia
Jakarta, Indonesia
nizar@cs.ui.ac.id

3rd Nur Fitriah Ayuning Budi
Fakultas Ilmu Komputer
Universitas Indonesia
Jakarta, Indonesia
nurfitriah@cs.ui.ac.id

4th Zaenal Abidin
Fakultas Ilmu Komputer
Universitas Negeri Semarang
Semarang, Indonesia
z.abidin@mail.unnes.ac.id

5th Ayi Purbasari
Fakultas Ilmu Komputer
Universitas Pasundan
Bandung, Indonesia
pbasari@unpas.ac.id

Abstract—The Indonesian government's plan to move the capital city has received pros and cons in the community. Information related to it has become a trending topic in cyberspace, the internet, including YouTube. Many national news channels discuss this through videos on YouTube in the form of news, talk shows and interactive interviews with certain experts. Sentiment analysis that focuses on social issues is interesting to do on this topic. In contrast to the sentiment analysis carried out on product and service topics which usually uses a lot of adjective as the type of words analyzed to attract opinion sentiments, this study focuses on the verb as a focus to get the results of opinion sentiments related to social issues. With the Naive Bayes and Decision Tree to classifying sentiment, show that 91% and 93% is supporting the planned transfer of the capital city.

Keywords—Sentiment Analysis, Naive Bayes Classifier, Decision Tree Classifier, Social Issues, Indonesian Government

I. INTRODUCTION

Globally, quoted from [1] that YouTube is accessed by more than 1.9 billion people every month, more than 1 billion people watch every day and more than 1 billion times the video aired every day. YouTube is also available with local versions of more than 91 countries with 80 different languages, YouTube claims this covers 95% of internet users.

In addition to information and knowledge shared between internet users through videos on YouTube, they can also provide feedback and feelings about the videos watched through the comments page. [2] estimate that from the video on YouTube there are 60-80% of comments in the form of opinions. And according to [3], relevant comments containing opinions about videos can be sentiment analysis.

Sentiment analysis can be done on opinions on products, services, and social problems. Research [4] and [5] focus on product analysis and service sentiments. Technically there are two approaches to conducting sentiment analysis on products and services, namely Bag Of Words (BOW) and Feature-Based Sentiment (FBS). With BOW, the sentence is separated into per word then only each word is judged by its sentiment. This makes the semantic information between the relationship of one word to another missing [6]. Not suitable for analyzing opinion sentiments about a product

that has features. Whereas with FBS can analyze the opinion sentiment of a product and its features. BOW and FBS have similarities in the use of adjectives, adverbs, and nouns as a feature to classify the polarity of sentiments.

In [7] sentiment analysis focused on social issues because according to him, public opinion related to social issues was important for the government to make decisions. Unlike products and services, more social problems use verbs to express opinions. The research has proven that the use of verbs is more effective than adjective in expressing opinions on social problems.

Policies made by the government but reaping the pros and cons are a social problem. This will be widely commented on by the public regarding the policy. Recently [8] the Indonesian government planned to move the country's capital. This plan raises the pros and cons of the community. Through this research, sentiment analysis of YouTube viewers' comments on videos related to the government's plan to move the capital by classifying verbs, using the Naive Bayes classifier and Decision Tree classifier, to know public opinion that supports or rejects government policies.

II. LITERATURE STUDY

A. Sentiment Analysis of Social Problems

According to [9], facts and opinions are the two main types of textual information. Facts express the objectivity of entities, events, and properties. While opinions, express the subjectivity of sentiment, judgment or feeling towards an entity, events, and property. The results of classifying sentiments in various formats according to the domain taken [7]: positive/negative, good/bad, like/dislike, or support/against. In this study classification of sentiment became support/against (supporting/resisting) against the Indonesian government's plan to move the capital.

Social problems are problems that arise from two consequences, first, due to social change and second, the results of social development [10]. The public can provide opinions about social problems that can influence decision-makers related to social issues, such as the government. Research on the analysis of sentiments with the social domain has been carried out by [11], he built a system that

is trained (supervised) from sentiment and opinion to prepare a sentence used in the debate.

Traditional machine learning uses adjectives, adverbs, and nouns as a feature to classify sentiment. However [7] has validated the hypothesis that the use of verbs is more than adjectives, adverb, and nouns in sentiment analysis with the domain of social problems.

B. Naive Bayes Classifier

Naive Bayes classifier is a classifier with the Naive Bayes algorithm. This classifiers are parameterized by two probability distributions [12]: $P(\text{label})$ gives the probability that an input will receive each label, given no information about the input's features; $P(\text{fname} = \text{fval} | \text{label})$ gives the probability that a given feature (fname) will receive a given value (fval), given that the label (label).

[13] uses the Naive Bayes classifier to classify author of tweets which aims to validate the identity of the author and avoid fake news generated from tweets.

C. Decision Tree Classifier

A demonstrate classifier that chooses which name to allot to a token on the premise of a tree structure, where branches compare to conditions on highlight values, and leaves compare to name assignments [14]. In other hand, [15] uses the Decision Tree classifier to predict clients from companies such as credit cards, insurance, banks, retail industries that require direct marketing whether to subscribe to a term deposit, defined that Decision Tree is a choice tree calculation makes a tree demonstrate by utilizing values of as it were one quality at a time. At to begin with, the calculation sorts the dataset on the attribute's esteem.

D. Plan to Move of Capital City

Reported by [16] that the transfer of the capital is planned for the government period 2019-2024 by carrying out the concept of smart green or an environmentally friendly city. This plan is not necessarily supported by all levels of society with various groups. Although under the pretext of the current capital city, Jakarta, severe traffic congestion, flooding and increasing urbanization, urban planning experts from Trisakti University, Nirwono Yoga, through [17] said that Jakarta was still worthy of being a capital city because of the massive investment that had been made, even for the next few years.

III. RESEARCH METHODOLOGY

This research was completed using a Python application. The process of data collection is done by accessing the YouTube API provided by Google. The data taken is comments from a video about the transfer of the capital city. Then label the comments containing verbs into several classes manually to prepare training data according to the research theory [7]. The class in question, like the example in Table 2, defines the label for comments as an expression of sentiment towards the government's plan to move the capital city.

A. Data

The random comment data is from 40 YouTube videos published in the range March 2019 - May 2019 in the News and Politics category. The total comments are 18418.

B. Preprocessing

The preprocessing stage is carried out in several stages including the removal of emojis, deletion of punctuation, deletion of numbers, correcting non-standard words, and POS-Tagging. The results of this stage are as shown in Table 1.

Removal of emojis, punctuation, and numbers is done to get a good POS-Tagging result. POS-Tagging is the stage of labeling types to each word. This study used IPOSTagger_v1.1 [18] for labeling word types. Correcting non-standard words is also intended so that when doing POS-Tagging you get a valid or standard type of word.

TABLE I. Preprocessing Results

Preprocessing	
1.	Removal Emoji
2.	Deletion of punctuation
3.	Deletion of number
4.	Correcting non-standard words
5.	POS-Tagging
↓	
Classification	
Original text	
Tolong yaa pak untuk d seriuskan... supaya tdk mnjadi wacana lg 🙏	
Preprocessing Results & POS-Tagging	
tolong ya pak untuk diseriuskan supaya tidak menjadi wacana lagi	
[[('tolong', 'VB'), ('ya', 'NN'), ('pak', 'NN'), ('untuk', 'SC'), ('diseriuskan', 'VB'), ('supaya', 'SC'), ('tidak', 'NEG'), ('menjadi', 'VB'), ('wacana', 'NN'), ('lagi', 'RB')]]	

C. Classification

The classification used are Naive Bayes Classifier and Decision Tree Classifier, they are Python library. After performing the preprocessing steps like Figure 1, the next step is classification.

Fig. 1. Preprocessing and Classification

This classification uses training data as many as 2000 lines which are grouped into two labels namely Support (support) and Against (reject/resist) as exemplified in Table 2 Training Data. And then about 16418 comments will be a testing data. The final result of this classification is to label all comments that have verbs and adjectives from the class that has been determined. Then calculate the frequency of each label produced.

TABLE II. Training Data

Text	Label
<i>pindah hore hore setuju banget</i> move hooray hooray really agree	Support
<i>jangan pindah ibukota nya pak karena jakarta saksi bisu sejarah indonesia menyatakan kemerdekaan sayang sekali pak</i> don't move the capital, Sir, because Jakarta is a silent witness to Indonesia's history, declaring independence	Against

IV. RESULT

A. Data Crawling

TABLE III. Data Crawling

Tag	Number of Comments
Verb	2945
Adj	517
Without Verb and Adj	13019

Table 3 showed the crawling data process resulted 16,418 total of comments, consisting of 2945 comments with verb, 517 comments with adjective and 13,019 comments without verb and adjective. In percentage terms, showed as Figure 2, the most comments are comments without verb and adjective, about 79%. These comments are generated from comments that use slang, which cannot be translated during the tagging process. The second and third most are comments that only contain verbs, about 18% and adjective only about 3%.

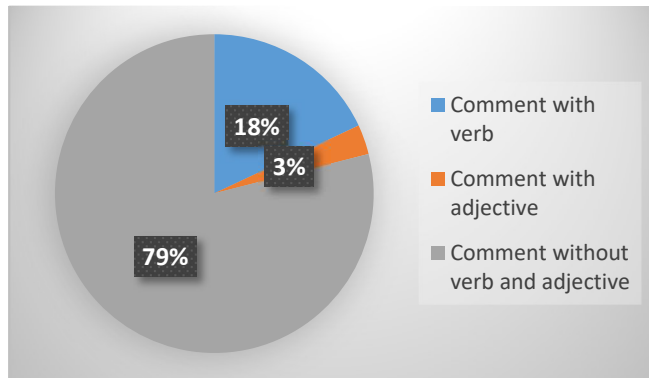


Fig. 2. Data Crawling Graph

B. Sentiment Analysis

TABLE IV. Sentiment Analysis Table

Classifier	Comments with	Support	Against	Sentiment
Naïve Bayes	Verb	2685	260	2425
	Adjective	477	40	437
Decision Tree	Verb	2741	204	2537
	Adjective	460	57	403

Table 4 shows, based on 2945 comment crawling data containing verbs that is supporting the government's plan to move the capital, with Naïve Bayes classifier, as many as

2685 support sentiment or around 91% and with Decision Tree classifier 2741 or around 93%. While the remaining 260 comments with Naïve Bayes or about 9% against the plan and 204 against sentiment with Decision Tree or about 7%. Meanwhile, based on 517 comment crawling data containing adjectives that is supporting the government's plan to move the capital, with Naïve Bayes classifier, as many as 477 support sentiment or around 92% and with Decision Tree classifier 460 comments or around 89%. Then only 40 comments or around 8% were against for Naïve Bayes classifier and 57 comments or around 11%.

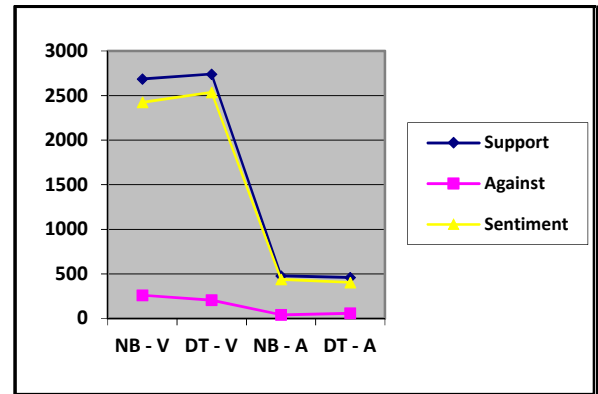


Fig. 3. Sentiment Analysis Graph

Using both Naïve Bayes and the Decision Tree, support sentiment for the government's plan to move the capital is greater than against sentiment. In the use of verb or adjective, the percentage of support sentiment is greater than against sentiment. Figure 3 shows, Naïve Bayes – Verb (NB-V), Naïve Bayes – Adjective (NB-A), Decision Tree – Verb (DT-V), and Decision Tree – Adjective (DT-A), that YouTube viewer sentiment on this topic tends to support.

Indeed, every policy will get pros and cons from the community. In depth, this research found public reasons why it supports and opposes government policies that will move the capital.

Supporting communities consider economic issues and strategic locations. As the following comments: *sangat mendukung sekali agar pemerataan semakin terwujud* | very supportive so that even distribution can be realized; *semoga pemikiran pindah ke papua biar tingkat pendidikan pembangunan perekonomian merata* | hopefully the thought of moving to Papua so that the level of education in economic development is evenly distributed; *sangat setuju ibu kota indonesia dipindah walaupun perekonomian merata terpusat pulau jawa bangunlah lapangan pekerjaan setiap daerah* | strongly agree that the capital city of Indonesia is moved even if the economy is evenly centered on the island of Java, build jobs in each region; *saya tinggal di jakarta sangat setuju pindah ibu kota, jakarta padat perlu pemerataan pembangunan, pak jokowi sangat cerdas sekali setuju pindah ibu kota* | I live in Jakarta, I strongly agree to move the capital city, Jakarta needs to be balanced with development, Mr. Jokowi is very smart to agree to move the capital; *ide bagus Kalimantan pak presiden soalnya pulauanya bagus posisinya strategis* | Kalimantan is a good idea, Mr. President, because the island has a good strategic position.

Meanwhile, those who oppose the policy plan are concentrated on political issues and environmental damage. Like the following comments: *pengalihan isu banyak*

karena kasus kecurangan pemilu | the transfer of many issues due to cases of election fraud; padahal utang negara belum terbayar kenapa malah langsung pindah ibukota | even though the state debt has not been paid, why instead moved the capital city directly; dari pada pindah ibu kota lebih baik bayar utang negara | instead of moving the capital city, it is better to pay the country's debt; Kalimantan jadi ibu kota, maka gak ada lagi istilah indonesia paru paru dunia | Kalimantan is the capital, so there is no longer the Indonesian term for the world's lungs; jangan kalimantan entar rusak habitat dan penyebab banjir | do not Kalimantan because it will damage the habitat and cause of flooding; goodbye hutan Kalimantan | goodbye Kalimantan forest; jangan pindah ke kalimantan karena kan kalimantan paru paru dunia | do not move to Kalimantan because Kalimantan is the world's lungs.

V. CONCLUSIONS AND RECOMMENDATIONS

Based on process of analysis of sentiment carried out in this study, it can be gathered that public opinion delivered through comments on YouTube largely supported the government's plan to move the capital city. Even so, opinions that do not agree with the plan still have a small portion.

The recommendation for the government is to continue the plan to move the capital city. In general, the community supports its policies. To be balanced, also reassure the public that this is not just a matter of politics. This policy is in the interest of the nation. Regarding environmental issues, convey steps to prevent environmental problems from occurring. These steps will become the community's guideline that relocating the capital will be safe from environmental problems.

Technically, most YouTube viewer comments are those that do not contain verbs and adjectives. While comments that contain more verbs than comments that contain adjectives. This is in line with research [7] which validates that the use of verbs is more than adjective, adverb and noun. The performance of the Naive Bayes classifier and the Decision Tree classifier also produced the same sentiment, dominated by those supporting government plans.

The weakness of this research is that it does not capture the sentiments of the YouTube video content commented on, we only chose random videos based on titles that fit the topic discussed, namely the capital transfer plan. Comments from YouTube viewers may also be influenced by the content watched.

The researchers hope that there will be further research related to sentiment analysis that focuses on social problems on YouTube by involving types of words other than verb and adjective. And technically, making YouTube video content is a domain which can also influence opinion.

ACKNOWLEDGEMENT

This study was supported by PIT9 Research Grant No NKB-0005/UN2.R3.1/HKP.05.00/2019 Universitas Indonesia.

REFERENCES

- [1] "YouTube for Press," YouTube, [Online]. Available: <https://www.youtube.com/yt/about/press/>. [Accessed 20 05 2019].

- [2] A. Severyn, A. Moschitti, O. Uryupina, B. Plank and K. Filippova, "Multi-lingual opinion mining on YouTube," *Information Processing and Management*, vol. 52, no. 1, pp. 46-60, 2016.
- [3] A. U. R. Khan, M. Khan and M. B. Khan, "Naïve Multi-label classification of YouTube comments using," in *Symposium on Data Mining Applications, SDMA2016*, Riyadh, Saudi Arabia, 2016.
- [4] M. Gamon, A. Aue, S. Corston-Oliver and E. Ringger, "Pulse: Mining customer opinions from free text," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Redmond, WA 98052, United States, 2005.
- [5] S. Shandilya and D. Jain, "Automatic opinion extraction from web documents," in *International Conference on Computer and Automation Engineering, ICCAE 2009*, Bangkok; Thailand, 2009.
- [6] B. Kasthuriarachchy, K. De Zoysa and H. Premaratne, "Enhanced bag-of-words model for phrase-level sentiment analysis," in *14th International Conference on Advances in ICT for Emerging Regions, ICTer 2014*, Colombo 07, Sri Lanka, 2014.
- [7] M. Karamibekr and A. Ghorbani, "Sentiment analysis of social issues," in *2012 International Conference on Social Informatics*, Lausanne, Switzerland, 2013.
- [8] R. Bonasir, "BBC," BBC News Indonesia, 29 04 2019. [Online]. Available: <https://www.bbc.com/indonesia/indonesia-48093451>. [Accessed 01 06 2019].
- [9] B. Liu, "Sentiment analysis and subjectivity," in *Handbook of Natural Language Processing, Second Edition*, Chicago, IL, United States, CRC Press, 2010, pp. 627-666.
- [10] "Wikipedia bahasa Indonesia," Wikipedia, Ensiklopedia bebas, 11 09 2018. [Online]. Available: https://id.wikipedia.org/wiki/Masalah_sosial. [Accessed 01 06 2019].
- [11] S. Somasundaran and J. Wiebe, "Recognizing stances in ideological online debates," *Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pp. 116-124, 2010.
- [12] E. Loper, "NLTK 3.4.4 documentation," nltk.org, 2001. [Online]. Available: https://www.nltk.org/_modules/nltk/classify/naivebayes.html#NaiveBayesClassifier. [Accessed 13 05 2019].
- [13] O. Aborisade and M. Anwar, "Classification for authorship of tweets by comparing logistic regression and naive bayes classifiers," in *19th IEEE International Conference on Information Reuse and Integration for Data Science, IRI 2018*, Salt Lake, 2018.
- [14] E. Loper, "NLTK 3.4.4 documentation," nltk.org, 2001. [Online]. Available: https://www.nltk.org/_modules/nltk/classify/decisiontree.html#DecisionTreeClassifier. [Accessed 13 05 2019].
- [15] R. M. R. Masud Karim, "Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing," *Journal of Software Engineering and Applications*, vol. 6, pp. 196-206, 2013.
- [16] R. Widiastuti and E. Adyatama, "Kontan," KONTAN, 13 05 2019. [Online]. Available: <https://nasional.kontan.co.id/news/jokowi-ada-tiga-kandidat-lokasi-pemindahan-ibukota>. [Accessed 01 06 2019].
- [17] G. Intan, "VOA Indonesia," VOA Indonesia, 05 05 2019. [Online]. Available: <https://www.voaindonesia.com/a/pemindahan-ibu-kota-perlu-kah-4904322.html>. [Accessed 01 06 2019].
- [18] A. Wicaksono and A. Purwarianti, "HMM Based Part-of-Speech Tagger for Bahasa Indonesia," in *Proceedings of 4th International MALINDO (Malay and Indonesian Language)*, 2010.