Visualizing Tweets Text Analysis of Indonesian Muslim Religious Figures on Moderate and Accusations of Radicalism using Sentiment Analysis and Word Clouds

Tresna Maulana Fahrudin, Amri Muhaimin, Prismahardi Aji Riyantoko, Kartika Maulida Hindrayani
Department of Data Science, Faculty of Computer Science
Universitas Pembangunan Nasional "Veteran" Jawa Timur
Surabaya, Indonesia
tresna.maulana.ds@upnjatim.ac.id, amri.muhaimin.stat@upnjatim.ac.id,
prismahardi.aji.ds@upnjatim.ac.id, kartika.maulida.ds@upnjatim.ac.id

Abstract

Indonesia is a country that has a large Muslim population. The country has many religions, such as Islam, Christianity, Buddhism, Confucianism, and Hindi, but Muslims are the majority. Many Indonesia people who embraced Islam, they need a role model such as a Muslim religious figure who can teach religious studies, be moderate, tolerance, respect differences of opinion and belief, be sociable, be fair, and not be extreme. The problem is that currently the notions of extremism, radicalism, and terrorism are spread by religious figures through social media where the target is mostly young people to adults. Therefore, this study proposes to explore tweets uploaded by Indonesian Muslim religious figures or certain communities on Twitter that lead to moderate or radicalism. Text visualization analysis and word cloud were used to reveal all the things hidden behind the tweet. The data collection is obtained through Twitter's API (Application Programming Interface) authentication using the Consumer Key, Consumer Secret, access token, and access token secret. The polarity scores of each tweet are applied using VADER sentiment analyzer to get the polarity, subjectivity, and sentiment labels. The words are conveyed by the two Muslim religious figures with the highest frequency of words were quite positive. The word frequency of Muslim religious figure A and B tweets which conveys the word such as bless (370), Allah (250), good (237), and upon (204), God (102), Islam (48), Muslim (41), and love (36). However, we found some use of vocabulary related to radicalism as found in Muslim religious figure B such as caliphate (17), radical (13), migrate (*hijrah*) (13), terror (10), terror (7), and caliph (7) which the keywords in 1-gram. The difference percentage of between both Muslim Religious Figure A and B for positive sentiment are 12.56%, neutral sentiment are 17.77%, and negative sentiment are 5.2%. Total Muslim Religious Figure A Tweets is 999, while the total Muslim Religious Figure B Tweets is 1274. The use of morphological vocabulary from Muslim religious figure B about accusations of radicalism have a relation with Muslim religious figure B tweets. For the next research, the analysis is not only limited based on the morphology of the word but based on the intent and implicit meaning of the sentence.

Keywords
Visualizing Tweets, Indonesian Muslim Religious Figures, Moderate and Accusations of Radicalism, Sentiment Analysis, Word Clouds

1. Introduction

Sentiment analysis is the partial scope of Natural Language Processing (NLP) which has the aim of analyzing text to extract, explore, also classify sentiments and opinions from the text information (Hearst et al., 2020)(Chaturvedi et al., 2018). Furthermore, sentiment analysis starts from the basic analytical techniques which applied to combine various information from the text and the other form such as visualizing data (Sánchez-Rada & Iglesias, 2019). The principal task of sentiment analysis can be categorized as a text classification problem because the process can include several operations that conclude and classify the text to determine whether it has positive or negative sentiments. Generally, sentiment analysis may seem an easy process, but in fact, it is necessary to consider many NLP subtasks

like sarcasm, radicalism, moderatism, and other words that are supposedly good or bad. Nowadays, sentiment analysis is becoming an additional insight, not only for researchers, in the other side such as companies, governments, also organizations have involved a lot of sentiment analysis to make a decision (Valdivia et al., 2018). The aforementioned situation is supported by the increasing number of internet users who build web and applications to more active spread broadcast information quickly by bringing important information. Millions of Indonesian social media users often express their opinions through forums, blogs, social media, and other web resources.

In addition, based on the scientific perspective of data analysis, it can be explained if the opinion and sentiment need analyzing the data if that used to observe public opinion and decision making (Ramírez-Tinoco et al., 2018). For example, the religious figures in Indonesia have boosted issues that potentially raised in a post, and become a major concern for people on Twitter, therefore the issue point can create the people's perspectives to judge religious figures as moderate, or radical. The phenomenon in Twitter social media delivers opinions significantly and potentially brings a piece hoax information (Pakpahan, 2017). According to the existing problems, we use a sentiment analysis approach use utilizing and retrieving data from Twitter. To make our research refined and easier, we also use a sentiment analysis approach combining with data visualization using a word cloud to blend and present positive and negative words.

The sentiment analysis approach on Twitter faces distinctive challenges because there are many error words such as idiomatic expressions, abbreviations or other unstructured data. The data preprocessing performs an important section in sentiment analysis for processing the necessary information and conducting analysis to investigate whether the word consisting positive or negative information. Moreover, the preprocessing data techniques will select, sort, and classify inconsistent and redundant data. Therefore, preprocessing data can be the main idea in sentiment analysis to evaluate words with carefully and precisely, hence we can obtain the classification data accurately. Therefore, preprocessing approach is very helpful to bringing up and perfecting words that often present with visualization techniques using word clouds (Venkatesh & Kalaivani, 2019). We also need to identify if word cloud the visualization technique can help researchers quickly present data from Twitter by highlighting the most common words and presenting the data in a form and make easier for everyone to understand.

Accordingly, the main purpose of this research is to determine and solve the problems based main category data about moderate and accusations of radicalism to Muslim religious figures in Indonesia which are retrieved from Twitter data source in the form of text data by conducting sentiment analysis to present positive and negative words using the word cloud visualization technique.

1.1 Objectives

Indonesia is an archipelagic country that has a diversity of cultures, races, ethnicities, beliefs, religions, and languages. Almost 87.2% of Indonesian citizens are Muslim, 9.90% are Protestant or Catholic, 1.69% Hindu, 0.72% Buddhist and 0.05% Confucian (Hefner, 2021). Many Indonesia people who embraced Islam, they need a role model such as a Muslim religious figure who can teach religious studies, be moderate, tolerance, respect differences of opinion and belief, be sociable, be fair, and not be extreme. Moderate Islam is defined as a group of people who have moderate character and attitude in dealing with confrontation and conflict over differences (Rozi, 2014). Otherwise, radicalism opposes democracy and wants a return to the caliphate, rejects freedom of religion and apostasy from Islam, there is a desire for Islam to rule, and supports jihadist groups (Denial, 2017). The problem is that currently the notions of extremism, radicalism, and terrorism are spread by religious figures through social media where the target is mostly young people to adults (Huda et al., 2021). Therefore, this study proposes to explore tweets uploaded by Indonesian Muslim religious figures or certain communities on Twitter that lead to moderate or radicalism. Text visualization analysis and word cloud were used to reveal all the things hidden behind the tweet.

2. Literature Review

The fast flows of information in the era of globalization can spread terrorism and radicalism. (Suciati & Erzad, 2018) found that the media and information technology that developed had a role in radicalism and terrorism. Radical groups convey their messages that what they did is correct and seek support from people. Therefore, the detection of texts containing radicals and terrorists can help prevent the spread of unwanted information.

(Alvianda & Adikara, 2019) use Support Vector Machine method to classify sentiment analysis against text radicals on Twitter. Twitter is a social media that stores and disseminates personal opinions. The categories obtained are negative and positive radical content. Accuracy gained in the use of SVM is up to 70%. (Miranda et al., 2020) also

use SVM to detect radicalism content in Twitter. TF-IDF is used in the feature selection process. There are two classes of classifier, radicalism and no-radicalism. High accuracy of 83% obtained with the model.

(Purwiantono & Aditya, 2020) classify radical, hoax, and *SARA* (ethnicity, religion, race, and inter-group relations) texts using naive bayes multinomial. The data source taken from Twitter. The categories applied to the classification model are normal, radical, *SARA*, and hoax. The four categories have been confirmed to have a balanced probability in the dataset. The accuracy result of the research reached 99.61%.

(Subhan et al., 2017) applied feature selection to handle radical dataset in Indonesia which obtained from Indonesian Ministry of Information and Communication, especially regarding jihad. The Ministry blocked the website that contain several categories such as radicalism, SARA, pornography, and etc. Feature selection is applied to reduce dimensions to gain the classification algorithm performance works well. DF-Threshold and human brain is adopted as feature selection method in the research. The accuracy result reached 66.37% with kNN method which *k* value is 7.

(Idris et al., 2019) used LSTM algorithm (Long Short Term Memory) to classify radical dataset in Indonesia and collected from Indonesian Ministry of Information and Communication. The categories of the class label are radical and non-radical. LSTM algorithm given the best performance reached 81.6% compared with k-NN and SVM methods.

3. Methods

In this section it will be explained some methods used in the proposed research. The most important thing is how to collect tweet through API access provided by Twitter (Chaudhary & Niveditha, 2021) related to Indonesian religious figures. After the tweet data is collected, the next step is to translate it from Indonesian to English using Google Translate API (Singh & Saravanan, 2021). The text is pre-processed to remove duplicate tweets and remove punctuation and characters. After that, the text is analyzed into polarity to be categorized as positive sentiment, negative sentiment and neutral sentiment including the percentage using VADER (Valence Aware Dictionary for Sentiment Reasoning) sentiment analyzer (Hutto & Gilbert, 2014). Each category can be visualized into a word cloud to display a representation of what words are dominant. The output is that we can find out which polarity of each Indonesian Muslim religious figure will lead to which sentiment category and how the word cloud is represented (Sasikala P & Sheela, 2018). Finally, analyzing of word frequency leads on moderate and radicalism in each polarity category. The proposed system design of the research is illustrated in Figure 1.

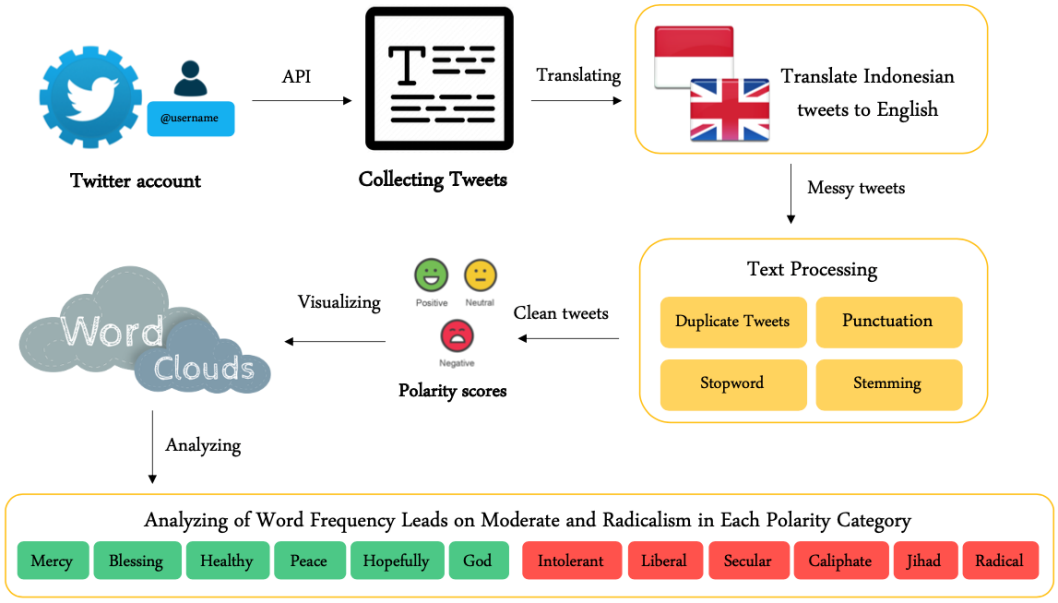


Figure 1. The Proposed System Design of Visualizing Tweets Text Analysis

3.1 Indonesian Religious Figures

Indonesia is a country that has a large Muslim population. The country has many religions, such as Islam, Christianity, Buddhism, Confucianism, and Hindi, but Muslims are the majority here. Therefore, there are many religious figures living in this country who can give a great influence once they make a few statements. For example,

a famous religious figure on social media uploads something, followers will be loyal to him and follow his instructions. There are two major Islamic organizations whose followers are in Indonesia and are moderate, namely *Nahdhatul Ulama* and *Muhammadiyah* (Zarkasyi, 2019). On the other hand, Indonesia has disbanded community organizations such as *Hizbut Tahrir Indonesia* which tried to change Pancasila as a state ideology into a caliphate system (Burhani, 2017). However, currently public sentiment in Indonesia is more sensitive than before. Social media is now a place for exchanging information, news and events in the form of text, images and videos. Behind the freedom of social media users to share something results in slander, conflict, confrontation and division, bad influence, and mobilization. Therefore, this study tries to explore the public's perception of a religious figure through visualization and sentiment analysis whether there is a tendency to radicalism or moderate.

3.2 Text Pre-processing

When the tweet data is collected in the data frame, the tweet text is still in Indonesian. We try to translate it into English to make it easier when it will be processed into polarization analysis and word cloud visualization. However, before that, it is necessary to do text preprocessing.

- **Delete duplicate tweets**
  When tweet data is collected, usually re-tweets are also collected so it is suspected whether the tweets are from humans or from bots. In order not to bias the analysis, it is necessary to remove duplicate tweets.
- **Remove punctuation**
  Tweets usually contain punctuation marks, symbols, numbers, certain notations that will interfere with the analysis process. Therefore, it is necessary to remove it for a perfect cleansing.
- **Tokenization**
  Tokenization changes from tweets in the form of sentences into tokens or individual words to facilitate further analysis. Basically, computers are not able to directly read complete sentences like humans, so they need to change from text to quantitative form.
- **Stop word**
  Sentences usually contain affixes *in-*, *at-*, *which-, where-, to-,* and others. These affixes have no meaning for the purposes of analysis. Therefore, the token in the sentence recorded on the stop word will be deleted.
- **Stemming**
  The token from the sentence is usually not returned to the base word, stemmer is useful for getting the base word from affixes that contain prefixes or suffixes.

3.3 Sentiment Analysis using Polarity Scores

Every sentence that comes out of our fingers or mouth always contains emotion. Emotions are divided into 27, such as happy, disgusted, sad, angry, and others. It is often classified into three sentiments such as negative, positive, and neutral in sentiment analysis. Each word has a score so that the sentence can be classified into one of them. In English, there is a sentiment dictionary where each word has a score using VADER. Illustration of the calculation, first we have a sentence, for example the sentence contains 10 words. Then each word we look for how much the score. For positive words, the value is close to +4, while for negative words close to -4. For neutral words, the value is always close to zero. However, in this case, we want to classify the sentiments of those tweets and then visualize each of those sentiments into a word cloud. Then we reanalyze the word cloud results whether there are certain words that lead to radicalism such as caliphate, jihad, radical and others. Figure 2 show the sentiment analysis based on polarity such as positive, neutral, and negative.

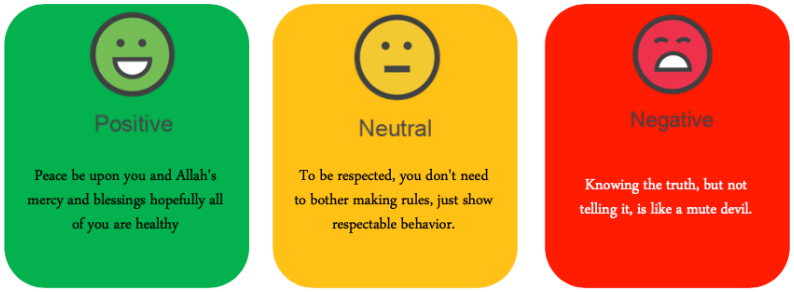


Figure 2. The Sentiment Analysis based on Polarity

After each sentence gets a positive, neutral, or negative using polarity scores using VADER's sentiment intensity analyzer, then the values between polarities are compared using certain conditions. If the negative polarity score is greater than the positive polarity score, the sentence will be included in the negative sentence category. On the other hand, if the positive polarity score is greater than the negative polarity score, the sentence will be included in the positive sentence category. Another condition, if the negative polarity score is equal to the positive polarity score, then the sentence will be included in the neutral sentence category. Figure 3 show the pseudocode of the sentiment analysis using polarity score.

```
Input:score ← polarity score of text using Sentiment Intensity Analyzer
      neg  ← score of negative sentiment
      neu  ← score of neutral sentiment
      pos  ← score of positive sentiment
      negative ← flag increment by 1 for counting negative sentiment
      neutral  ← flag increment by 1 for counting neutral sentiment
      positive ← flag increment by 1 for counting positive sentiment
      negative_list ← append element to list of negative sentiment
      neutral_list. ← append element to list of neutral sentiment
      positive_list ← append element to list of positive sentiment
Output: the sentiment category based on polarity score
for each tweet in tweets
   score ← SentimentIntensityAnalyzer().polarity_scores(tweet.text)
   neg ← score['neg']
   neu ← score['neu']
   pos ← score['pos']
   if neg is greater than pos then
         negative_list.append(tweet.text)
         negative ← negative + 1
   elif pos is greater than neg then
         positive_list.append(tweet.text)
         positive ← positive + 1
   elif pos is equal than neg then
         neutral_list.append(tweet.text)
         neutral ← neutral + 1
end for
```

Figure 3. The Pseudo-code of the Sentiment Analysis using Polarity Score

3.4 Word Cloud

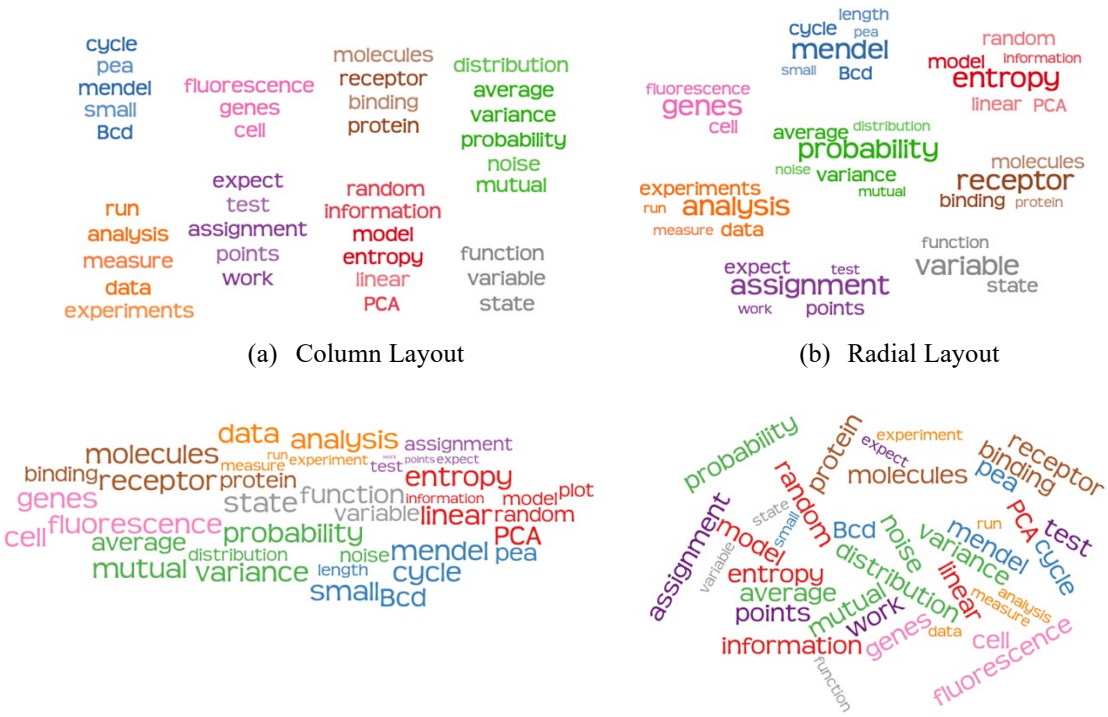
Word Cloud is like a bar plot of a set of words, but it is presented and visualized in a different form. The set of words is processed and the weight is calculated and then stored into a structured dataset which can be seen in Table 1. The weight represents the total number of words that show in the whole data. To create the word cloud, first, we do the preprocessing that is explained in the previous section. The higher weight is the bigger word on the word cloud, and vice versa. The word cloud is represented in Figure 2 which presents the distribution of the word using the count for the word that is obtained in the tweets from Twitter, post from Facebook, or text from document.

Table 1. The Structured Dataset for Word Cloud

Word	Weight
Word-1	w_1
Word-2	w_2
⋮	⋮
Word-n	w_n

The word cloud makes it easier to understand how the word frequency condition actually is. If previously using a simple way of calculating the frequency of words and presented in a table, word cloud presents it in a visual form. A word that appears in large size in the word cloud indicates that the word does have a high weight and can be considered an important word. This needs to be further analyzed what is behind the word, whether it is related to certain issues,

topics, and interests. Figure 4 show the several design of word cloud in (a) column layout, (b) radial layout, (c) BSC layout an, (d) davies layout (Hearst et al., 2020).



(a) Column Layout (b) Radial Layout

(c) BSC Layout (d) Davies Layout

Figure 4. The Design of Word Cloud in Column, Radial, BSC, and Davies Layout

4. Data Collection

The dataset used in this study was obtained from tweets from two Muslim religious figures who, according to public perception and news, both have different ways of conveying religious messages. Muslim religious figure A conveys religious messages with polite words, prays, conveys greetings, gives hope and motivation, and advices. In contrast to Muslim religious figure A, Muslim religious figure B expresses more religious messages that adopt words such as accusations of radicals, caliphate, and pro against Islamic organizations which are prohibited by the government lead even though morphologically the words are positive sentences.

The data collection is obtained through Twitter's API (Application Programming Interface) authentication using the Consumer Key, Consumer Secret, access token, and access token secret. Then, enter the username based on screen name of Muslim religious figures to get tweets on their respective timelines. The authentication of client application is registered by Tweepy library through creating the OAuthHandler instance and access token. After the account is verified by Twitter, the client gets a response to the request for tweet data. Figure 5 show the authentication of Twitter's API.

```
# Authentication
consumerKey = ''
consumerSecret = ''
accessToken = ''
accessTokenSecret = ''
auth = tweepy.OAuthHandler(consumerKey, consumerSecret)
auth.set_access_token(accessToken, accessTokenSecret)
api = tweepy.API(auth)
```

Figure 5. The Authentication of Twitter's API

5. Results and Discussion

In this section, some of the numerical and graphical results, proposed improvement, and validation in the proposed research will be described.

5.1 Numerical Results

The tabular dataset of experiment result consists of columns and rows including Indonesian (language) tweets, English (language) tweets, polarity, subjectivity, and sentiment labels. Table 3 shows the tabular dataset of experiment result from Muslim religious figures tweets. As humans, we know that sentences with positive polarization contain a frequency of positive words, and vice versa. However, if you look at the tabular dataset in the second column, there are several sentences that show positive, neutral and negative sentences. There are even positive sentences that use the vocabulary of "radical", "extreme", "caliphate", and "terror", but we cannot judge that this is related to an open understanding of radicalism. In addition, there are several sentences that contain implicit meanings and the possibility of the system recognizing them morphologically, not by meaning. Therefore, it is necessary to be very careful in analyzing the tweet data because it relates to various fields of science regarding religion, ideology, social, and politics.

The polarity scores of each tweet are applied using VADER sentiment analyzer to get the polarity, subjectivity, and sentiment labels. The polarity score represents the sentiment score of positive, neutral, and negative. The subjectivity represents whether a sentence has a high objectivity or not. The higher the subjectivity means less objective. On the other hand, lower subjectivity means more objective. The sentiment label represents the classification of positive, neutral, and negative. Table 3 show the polarity, subjectivity, and sentiment labels of Muslim religious figures tweets. For example, the first-row tweet, it is a positive tweet that conveys greetings and wishes good things to someone by showing a polarity value of 0.649999, subjectivity of 0.538194, and the label is categorized as positive sentiment. The third-row tweet, it is a neutral tweet that conveys there is someone who can speak Arabic and English by showing a polarity value of 1.302083, subjectivity of 0.260417, and the label is categorized as neutral sentiment. The sixth-row tweet, it is a negative tweet that conveys there were individuals and parties who linked the terrorist bombing in the city of Surabaya with Muslims, especially those considered "radical" such as "bearers of the caliphate". He also said that it is very heartbreaking for someone who uses the bomb tragedy for negative framing and really has no feelings by showing a polarity value of 0.0388889, subjectivity of 0.5388889, and the label is categorized as negative sentiment.

Table 3. The Polarity, Subjectivity, and Sentiment Labels of Muslim Religious Figures Tweets

No.	Indonesian (language) Tweets Type: Text	English (language) Tweets Type: Text	Polarity Type: Continuous	Subjectivity Type: Continuous	Sentiment Labels Type: Nominal
1	*Assalamualaikum warahmatullahi wabarakatuh. Selamat pagi. Semoga hari-hari Anda senantiasa dipenuhi rahmat dan berkah Allah*	Peace be upon you, and Allah's mercy and blessings. Good morning. May your days always be filled with Allah's grace and blessings	0.649999	0.538194	Positive
2	*Ada yang mempunyai hanya ideologi buatan akal manusia, tapi yakin dan berkata, "inilah sistem yang paling baik", dia bangga walau sesat*	There are those who have only ideologies made by human reason, but believe and say, "this is the best system", they are proud even though they are misguided	0.459999	0.6577778	Positive
3	*Beliau bisa bahasa Arab (aslinya kan Lebanon) maupun Inggris. Dengan saya menggunakan bahasa Arab; dengan tamunya yang dari Amerika menggunakan bahasa Inggris*	He can speak Arabic (originally Lebanese) and English. With me in Arabic; with the guest from America speaking English	1.302083	0.260417	Neutral

No.	Indonesian (language) Tweets Type: Text	English (language) Tweets Type: Text	Polarity Type: Continuous	Subjectivity Type: Continuous	Sentiment Labels Type: Nominal
4	*Tindakan Rasulullah sangat ekstrim, dan mungkin sangat radikal, tapi ini yang diperlukan pada kondisi itu. Karena bila tindakan itu tidak dilakukan, mudharatnya bakal lebih lagi*	Acts of Prophet is very extreme, and may be very radical, but this is what is needed in that condition. Because if the action is not carried out, the harm will be more	0.159375	0.475000	Neutral
5	*Penghinaan orang terhadap koruptor dianggap meringankan hukumannya, apakah ini benar atau hoax?*	People's insults against corruptors are considered to lighten their sentences, is this true or a hoax?	0.285714	0.535714	Negative
6	*Seperti sudah diduga, mulai ada oknum dan pihak yang mengaitkan aksi teror bom surabaya, dengan kaum Muslim, khususnya yang dianggap "radikal", seperti "pengusung khilafah" kata mereka. Teganya memanfaatkan tragedi bom untuk framing negatif, benar-benar tak punya perasaan*	As expected, there have been individuals and parties who have been linked with the Surabaya bombing terror attack, with Muslims, especially those who are considered "radicals", such as "bearers of the caliphate," he said. How dare to use the tragedy of the bomb to frame the negative, really have no feelings	0.0388889	0.5388889	Negative
	...	…	…	…	…

Table 4 show the word frequency of Muslim religious figure A tweets which conveys the word such as bless (370), Allah (250), good (237), and upon (204), while the word frequency of Muslim religious figure B tweets which conveys the word such as God (102), Islam (48), Muslim (41), and love (36). The words are conveyed by the two Muslim religious figures with the highest frequency of words were quite positive.

Table 4. The Word Frequency of Muslim Religious Figure A and B Tweets

No.	The Word Frequency of Muslim Religious Figure A		The Word Frequency of Muslim Religious Figure B	
	Word	Frequency	Word	Frequency
1	Bless	370	God	102
2	Allah	250	Islam	48
3	Good	237	Muslim	41
4	Upon	204	Love	36
5	Hope	83	Hope	27
6	Love	30	Good	27
7	Heart	23	Pray	20
…	…	….	…	….

However, we found some use of vocabulary related to radicalism as found in Muslim religious figure B such as caliphate (17), radical (13), migrate (*hijrah*) (13), terror (10), terror (7), and caliph (7) which the keywords in 1-gram. While, the keywords in 2-gram such as {‘caliph’, ‘khilafah’}, {‘radical’, ‘islam’}, {‘radical’, ‘threatening’}, and {‘hijrah’, ‘start’}. Table 5 show some use of vocabulary related to radicalism as found in Muslim Religious Figure B tweets. There are differences in how the Muslim religious figure A and B discusses a topic in conveying religious messages through tweets, but uses vocabulary that leads to radicalism in morphology of words.

Table 5. Some use of Vocabulary related to Radicalism as found in Muslim Religious Figure B Tweets

No.	The Vocabulary related to Radicalism based on Frequency of Muslim Religious Figure B			
	1-gram	Frequency	2-gram	Frequency
1	Caliphate (Khilafah)	17	{'caliph', 'khilafah'}	2
2	Radical	13	{'radical', 'islam'}	2
3	Migrate (*Hijrah*)	13	{'radical', 'threatening'}	2
4	Terror	10	{'hijrah', 'start'}	2
5	Caliph (Khalifah)	7	{'*hijrah*', '*ustadz* (islam teacher)'}	2
6	Hizbut Tahrir Indonesia (some keywords related to)	7	{'terrorism', 'religion'}	2
7	Liberalist	2	{'anti', 'pancasila'}	2
8	Terrorist	3	{'extreme', 'radical'}	1
9	Islamophobia	3	{'radical', 'needed'}	1
10	Anti-pancasila (Ideology of Indonesia)	2	-	-
11	Communism	1	-	-
12	Secular	1	-	-
Total		79	Total	16

5.2 Graphical Results

To get a percentage of the classification of positive, neutral, and negative sentiments, we present them in the form of a pie chart. The pie chart can represent the proportion of a category in percentage and makes it easier to understand the condition of a certain category through color. Figure 6 show the proportion of sentiment in the pie chart which positive sentiment is green, neutral sentiment is blue, and negative sentiment is red. Both Muslim religious figure tweets generally convey religious messages through tweets with positive messages, but it seems that Muslim religious figure B may have a slightly larger percentage of negative sentiment than Muslim religious figure A.

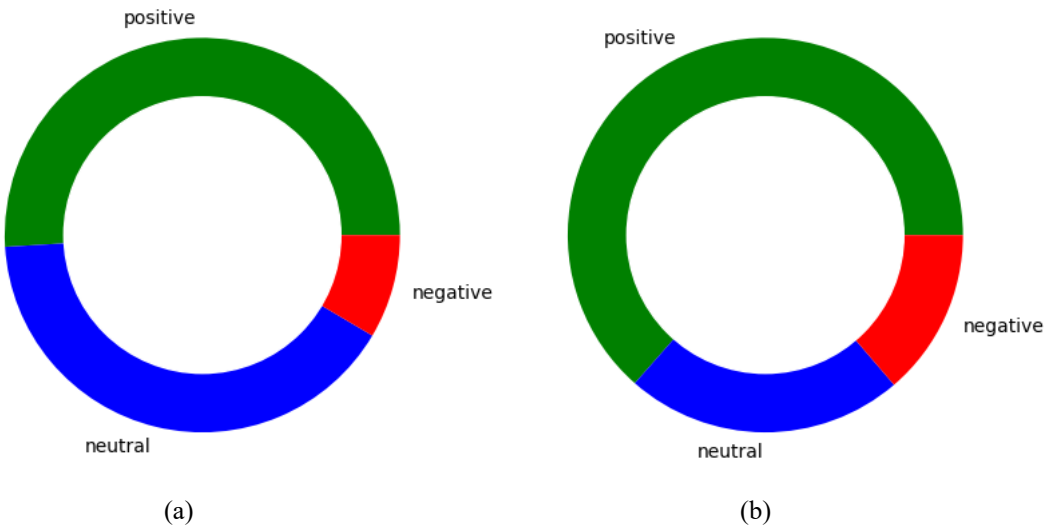


Figure 6. The Pie Chart of Sentiment Labels:
(a) Muslim Religious Figure A Tweets, (b) Muslim Religious Figure B Tweets

Table 6 show the percentage of sentiment classification based on Muslim religious figure A and B tweets. The difference percentage of between both Muslim Religious Figures for positive sentiment are 12.56%, neutral sentiment

are 17.77%, and negative sentiment are 5.2%. Total Muslim Religious Figure A Tweets is 999, while the total Muslim Religious Figure B Tweets is 1274.

Table 6. The Percentage of Sentiment Classification based on Muslim Religious Figure A and B Tweets

Sentiment labels	Muslim Religious Figure A Tweets		Muslim Religious Figure B Tweets	
	Total	Percentage	Total	Percentage
Positive	509	50.95%	792	63.51%
Neutral	405	40.54%	284	22.77%
Negative	85	8.51%	171	13.71%
Total	999	100%	1274	100%

5.3 Proposed Improvements

After getting the sentiment classification, we get how the percentage of positive, neutral and negative sentiment. However, we need to see how the word cloud visualization looks like on each Muslim religious figure tweets. Figure 7 show the word cloud visualization based on Muslim religious figure tweet A and B. There are some differences in words that stand out in both of them in using positive and negative vocabulary. The word 'peace' looks big in the word cloud on Muslim Religious Figure A tweets, while the word 'flow' looks big in the word cloud on Muslim Religious Figure B tweets but it means about religious sect.

(a) (b)

Figure 7. The Word Cloud Visualization:
(a) Muslim Religious Figure A Tweets, (b) Muslim Religious Figure B Tweets

5.4 Validation

The experiment result show that there are differences in the choice of vocabulary used by both Muslim Religious Figures. Both convey positive messages about religion, but Muslim religious figure B morphologically conveys religious messages related to radicalism such as caliphate, caliph, radical, terror, liberal, communism, secular, Islamophobia, anti-pancasila (ideology of Indonesia), and even discussed banned organizations such as *Hizbut Tahrir Indonesia.* It is suspected that the analysis results of the use of morphological vocabulary from Muslim religious figure B about accusations of radicalism have a relation with Muslim religious figure B tweets which are indicated by 1-gram as many as 79 words and 2-gram as many as 16 words.

6. Conclusion

The research problems of the proposed research about moderate and accusations of radicalism to Muslim religious figures in Indonesia which are retrieved from Twitter by conducting sentiment analysis to present positive, neutral, and negative words using the word cloud visualization. The data collection is obtained through Twitter's API (Application Programming Interface) authentication using the Consumer Key, Consumer Secret, access token, and access token secret. The polarity scores of each tweet are applied using VADER sentiment analyzer to get the polarity, subjectivity, and sentiment labels. The words are conveyed by the two Muslim religious figures with the highest frequency of words were quite positive. The word frequency of Muslim religious figure A and B tweets which conveys the word such as bless (370), Allah (250), good (237), and upon (204), God (102), Islam (48), Muslim (41), and love (36). However, we found some use of vocabulary related to radicalism as found in Muslim religious figure B such as caliphate (17), radical (13), migrate (*hijrah*) (13), terror (10), terror (7), and caliph (7) which the keywords in 1-gram. The difference percentage of between both Muslim Religious Figure A and B for positive sentiment are 12.56%, neutral sentiment are 17.77%, and negative sentiment are 5.2%. Total Muslim Religious Figure A Tweets is 999, while

the total Muslim Religious Figure B Tweets is 1274. The use of morphological vocabulary from Muslim religious figure B about accusations of radicalism have a relation with Muslim religious figure B tweets. For the next research, the analysis is not only limited based on the morphology of the word but based on the intent and implicit meaning of the sentence.

References

Alvianda, F., & Adikara, P. P. (2019). Analisis Sentimen Konten Radikal Di Media Sosial Twitter Menggunakan Metode Support Vector Machine (SVM). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, *3*(1), 241–246.

Burhani, A. N. (2017). The Banning of Hizbut Tahrir and the Consolidation of Democracy in Indonesia. *ISEAS Perspective*, *71*(2017), 1–10.

Chaturvedi, I., Cambria, E., Welsch, R. E., & Herrera, F. (2018). Distinguishing between facts and opinions for sentiment analysis: Survey and challenges. *Information Fusion*, *44*, 65–77. https://doi.org/10.1016/j.inffus.2017.12.006

Chaudhary, J., & Niveditha, S. (2021). Twitter Sentiment Analysis using Deep Learning. *International Research Journal of Engineering and Technology (IRJET)*, *8*(4), 4512–4516. https://doi.org/10.2139/ssrn.3885137

Denial, B. (2017). Moderate Muslims and Islamist Terrorism : Between Denial and About the Author. *The International Centre for Counter-Terrorism - The Hague*, *8*(August). https://doi.org/10.19165/2017.1.09

Hearst, M. A., Pedersen, E., Patil, L., Lee, E., Laskowski, P., & Franconeri, S. (2020). An Evaluation of Semantically Grouped Word Cloud Designs. *IEEE Transactions on Visualization and Computer Graphics*, *26*(9), 2748–2761. https://doi.org/10.1109/TVCG.2019.2904683

Hefner, R. W. (2021). Islam and institutional religious freedom in indonesia. *Religions*, *12*(6), 1–20. https://doi.org/10.3390/rel12060415

Huda, A. Z., Runturambi, A. J. S., & Syauqillah, M. (2021). Social Media as An Incubator of Youth Terrorism In Indonesia : Hybrid Threat and Warfare. *Jurnal Indo-Islamika*, *11*(1), 21–40. https://doi.org/10.15408/jii.v11i1.20362

Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Eighth International AAAI Conference on Weblogs and Social Media*, *8*(1), 216–225.

Idris, N. O., Widyawan, & Adji, T. B. (2019). Classification of Radicalism Content from Twitter Written in Indonesian Language using Long Short Term Memory. *ICICOS 2019 - 3rd International Conference on Informatics and Computational Sciences: Accelerating Informatics and Computational Research for Smarter Society in The Era of Industry 4.0, Proceedings*, 1–5. https://doi.org/10.1109/ICICoS48119.2019.8982498

Miranda, E., Aryuni, M., Fernando, Y., & Kibtiah, T. M. (2020). A study of radicalism contents detection in twitter: Insights from support vector machine technique. *Proceedings of 2020 International Conference on Information Management and Technology, ICIMTech 2020*, *August*, 549–554. https://doi.org/10.1109/ICIMTech50083.2020.9211229

Pakpahan, R. (2017). Analisis Fenomena Hoax Diberbagai Media. *Konferensi Nasional Ilmu Sosial & Teknologi (KNiST)*, 479–484.

Purwiantono, F. E., & Aditya, A. (2020). Klasifikasi Sentimen Sara, Hoaks Dan Radikal Pada Postingan Media Sosial Menggunakan Algoritma Naive Bayes Multinomial Text. *Jurnal Tekno Kompak*, *14*(2), 68–73. https://doi.org/https://doi.org/10.33365/jtk.v14i2.709

Ramírez-Tinoco, F. J., Alor-Hernández, G., Sánchez-Cervantes, J. L., Olivares-Zepahua, B. A., & Rodríguez-Mazahua, L. (2018). A Brief Review on the Use of Sentiment Analysis Approaches in Social Networks. In J. Mejia, M. Muñoz, Á. Rocha, Y. Quiñonez, & J. Calvo-Manzano (Eds.), *Trends and Applications in Software Engineering* (pp. 263–273). Springer International Publishing.

Rozi, S. (2014). Tracking on moderate muslim in indonesia. *Turãst: Jurnal Penelitian & Pengabdian*, *2*(1), 63–81. https://doi.org/10.15548/turast.v2i1.393

Sánchez-Rada, J. F., & Iglesias, C. A. (2019). Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison. *Information Fusion*, *52*, 344–356. https://doi.org/10.1016/j.inffus.2019.05.003

Sasikala P, & Sheela, L. M. I. (2018). Sentiment Analysis of Online Food Reviews using Customer Ratings. *International Journal of Pure and Applied Mathematics*, *119*(15), 3509–3514.

Subhan, M., Sudarsono, A., & Barakbah, A. (2017). Preprocessing of radicalism dataset to predict radical content in Indonesia. In *Proceedings - International Electronics Symposium on Knowledge Creation and Intelligent Computing, IES-KCIC 2017* (pp. 270–275). IEEE. https://doi.org/10.1109/KCIC.2017.8228598

Suciati, & Erzad, A. M. (2018). The existence of Kudus Islamic local culture to prevent radicalism in globalization era. *Qudus International Journal of Islamic Studies*, *6*(1), 39–56. https://doi.org/10.21043/qijis.v1i1.3460

Valdivia, A., Luzón, M. V., Cambria, E., & Herrera, F. (2018). Consensus vote models for detecting and filtering neutrality in sentiment analysis. *Information Fusion*, *44*, 126–135. https://doi.org/10.1016/j.inffus.2018.03.007

Venkatesh, N., & Kalaivani, A. (2019). Word cloud for online mobile phone tweets towards sentiment analysis. *International Journal of Engineering and Advanced Technology*, *8*(6), 683–689. https://doi.org/10.35940/ijeat.F7964.088619

Zarkasyi, H. F. (2019). Appraising the Moderation Indonesian Muslims with Special Reference to Muhammadiyah and Nahdlatul Ulama. *Addin*, *12*(1), 1–30. https://doi.org/10.21043/addin.v12i1.4179

Acknowledgements

The author would like to thank Universitas Pembangunan Nasional "Veteran" Jawa Timur, Ministry of Education, Culture, Research, and Technology which has provided internal funding for the publication of this scientific article.

Biography / Biographies

Tresna Maulana Fahrudin is a Full-time Lecturer of Data Science in Faculty of Computer Science, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Indonesia. He earned Applied Bachelor and Master of Engineering in Department of Informatics and Computer Engineering from Politeknik Elektronika Negeri Surabaya, Indonesia. He has published journal and conference papers. He got the Best Paper and Presenter Award in several National and International conference such as KCIC 2016, SEMNASIF 2020, ICPM 2020, and ITIS 2021, especially in Data Mining fields in the last research. Actively participates in training and workshops, such the Digital Talent Scholarship Training 2019 organized by Ministry of Communication and Information Technology of the Republic of Indonesia and Microsoft in Data Science and Data Analytic program. Currently, pursuing research in computational linguistic in text mining, data mining, edit-distance in text similarity, AI chatbot, and social network analysis.

Amri Muhaimin is a Full-time Lecturer of Data Science in Faculty of Computer Science, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Indonesia. He earned Bachelor of Engineering in Department of Statistics from Institut Teknologi Sepuluh Nopember, Indonesia. His master's degree is coming from joint degree between Institut Teknologi Sepuluh Nopember and National Chiao Tung University Taiwan. His background is statistics, but for now, he interested in deep learning. He already published some paper in Google Scholar, Majority of his research is about forecasting and machine learning. Right now, he is focusing on intermittent demand data, which is an interesting data to explore. He is member of Data Science Indonesia East Java region as committee in advocate division. The task is about to advocate some university when they want to create Data Science subject in their campus. He also works at Narasiodata company, which is a company to provide online course about data science, data analyst, and data engineering. His perspective about data is cliché, good data provide good result, trash raw material provide bad result. He likes to talk about basic fundamental about data. For now, he is busy to write book about unstructured data, swimming, and preparing about his Ph.D. soon.

Prismahardi Aji Riyantoko is a Full-time Lecturer in Department of Data Science, Faculty of Computer Science, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Indonesia. He earned Bachelor (BSc-Computational Mathematics) in Department of Mathematics from Institut Teknologi Sepuluh Nopember, Masters (MSc-Applied Mathematics) in Institut Teknologi Sepuluh Nopember. He has published journal and conferences papers. He has taught courses in mathematics for data science, discrete mathematics, calculus multivariable, introduction of statistics, statistic and probability. He has been recognized as a professional IT consultant with around one year of experience in working at Viseo with salesforce practice business partner. He manages International Journal of Data Science, Engineering, and Analytics as an editorial team in Department of Data Science with around one year. His research interest includes partial differential equation, optimal control, modelling system, and data analytics.

Kartika Maulida Hindrayani is a Full-time Lecturer in Department of Data Science, Faculty of Computer Science, Universitas Pembangunan Nasional "Veteran" Jawa Timur, Indonesia. She earned Undergraduate and Master Degree in Information Systems from Institut Teknologi Sepuluh Nopember. She started publishing journal and conference paper. Her research interest include optimization, scheduling, and business intelligence.