

Original Article

# Sentiment Analysis using Multi-Criteria Against Public Figures

Rina Candra Noor Santi<sup>1</sup>, Kristophorus Hadiono<sup>2</sup>, Sri Eniyati<sup>3</sup>, Sugiyamta<sup>4</sup>

<sup>1,2,3,4</sup> Faculty of Information Technology and Industry, Stikubank University, Indonesia

Received: 12 April 2022

Revised: 01 June 2022

Accepted: 18 June 2022

Published: 30 June 2022

**Abstract** - This study tries to classify the sentiments or opinions expressed through social media Twitter tweeted by a public figure and grouped based on several criteria of performance, character, and popularity. For performance criteria, the measurement takes tweets related to the words service and *giat*. This study analyzes positive, negative, and neutral sentiments from public figures' tweets on Twitter based on performance, character, and popularity criteria. The tweet data obtained from Twitter is 800 tweets where the performance criteria are sorted by the words service and active; a total of 334 tweets were obtained. For the character criteria, the data is sorted by good and firm words, and data is obtained from as many as 232 tweets, while for the popularity criteria, the data sorting is carried out Based on the words like and know; 234 tweets were obtained. The final data is displayed in graphical form. From the graph, it can be explained that the sentiment performance criteria generated are neutral, the sentiment character criteria generated are neutral, and the sentiment population criteria generated are neutral. So it can be concluded that each of these criteria has neutral sentiments.

**Keywords** - Sentiment analysis, Social network, Twitter, Performance, Character, Popularity.

## 1. Introduction

Data on internet users in 2016 reached 132.7 million, and from that figure, 97.4 percent of users use the internet to access social networking sites. Indonesian internet users frequently visit social networking sites besides Facebook, one of which is Twitter, where Twitter occupies the 5th position with users of 7.2 million (APJII, 2016).

When an incident that has the potential to become viral and attract enough attention is tweeted on Twitter, many Twitter users will give their opinion. Moreover, if it is tweeted by a public figure who has many followers or the public figure comments on the incident, many opinions will also be formed.

Various kinds of opinions sent via Twitter will have a good (positive) or bad (negative) impression, depending on the user's thoughts. Through sentiment analysis techniques or opinion mining, various opinions on Twitter issued by these public figures and their followers can be analyzed against various criteria. Many studies using sentiment analysis have been conducted, but they tend to compare the performance between the two methods and use a supervised machine learning approach.

This study will try to classify by utilizing sentiment analysis on tweets or opinions expressed through social media Twitter from public figures and classify them based on several criteria such as character, performance, and

popularity. This research results in the conclusion of positive, negative, and neutral sentiment on public figures' character based on performance, character, and popularity criteria by using R language.

The benefit of this research is knowing the tweets from these public figures in a positive, negative, or neutral direction based on the performance criteria, character, and popularity of public figures.

## 2. Background and Related Works

### 2.1. Sentiment Analysis

Sentiment analysis or opinion mining is a multi-disciplinary field that utilizes computing capabilities to research opinions, behavior, and human emotions on products, services, organizations, individuals, current issues, and so on in a text (Medhat et al. 2014; Liu 2012).

In general, sentiment analysis can be divided into 3 parts: sentiment analysis of documents, sentiment analysis of sentences, and sentiment analysis of aspects of an entity, where sentiment analysis for documents aims to classify documents into positive or negative sentiments. Sentiment analysis of sentences is similar to sentiment analysis of documents, aiming to classify positive or negative sentiments. However, what is analyzed is a sentence, not a document that usually discusses a topic.



Sentiment analysis on aspects aims to classify or classify sentiments specifically for certain aspects of an entity. The important thing in sentiment analysis is the data set used, and usually, the data sources that are often used are micro-social networks such as Twitter and websites (Medhat et al. 2014).

According to Liu (2008), sentiment analysis (sentiment analysis) or often referred to as opinion mining (opinion mining), is a computational study to recognize and express opinions, sentiments, evaluations, attitudes, emotions, subjectivity, judgments, or views contained in a text. The sources of sentiment analysis data can be obtained from:

Opinions on blogs, microblogs, or forums. For example, Twitter, Wordpress, Kaskus, comments on articles, topics, issues, or reviews, and posting on social media sites such as Facebook, LinkedIn, and email.

## 2.2. Text Mining

Text mining is the discovery of new and previously unknown information by computers, which automatically extract information from different unstructured text sources. The key to this process is to combine information that has been successfully extracted from various sources (Eko, 2011).

There are four main process stages in text mining: initial processing of text (text preprocessing), text transformation (text transformation), feature selection, and pattern discovery.

## 2.3. Preliminary Research

Many kinds of research on sentiment analysis on Indonesian-language tweets have been carried out. Several sentiment analysis studies on Indonesian-language tweets use the Naive Bayes or SVM classification method. However, the research conducted is more inclined to compare the performance of the two methods. One example is a comparative study of performance applied to two languages, namely English and Indonesian, to find the best between the two methods. (Saraswati 2011).

Hidayatullah & Azhari (2014) measured the results of sentiment classification (positive and negative) and categories (capability, integrity, and acceptability). Using the naive Bayes classification method and SVM against two different languages. The object used to measure the performance of the two methods with these two features is the tweets of public figures on Twitter in Indonesian.

Sentiment analysis research using the lexicon to conduct sentiment analysis was also conducted to detect emotions

such as happiness, love, anger, afraid, and sadness. In this study, researchers developed a lexicon consisting of seed words and an expanded lexicon. The selection of seed words was based on the types of emotions happy, love, anger, fear, and sadness. For expanding the lexicon, the researcher used an Indonesian thesaurus, and each word in the lexicon was given a binary weight of 1 and 0. The total words for the emotional lexicon obtained from the study were 1165 words (Bata, J. et al., 2015).

Other research still within the scope of sentiment analysis was carried out by Cheng et al. (2016) on opinion polls for public figures in China through micro-blogging text (Twitter and Sina Weibo). In this study, the researcher proposes a new method which is an adaptation of the recurrent neural network (RNN) method and is combined with the sequence labeling method, where this method is a sentiment parsing method.

## 3. Materials and Methods

The method of data collection is done utilizing observation. Observations were made by taking data on Twitter social media, with the sample being a public figure, using the R language. Besides that, a literature study was also used to obtain relevant literature on the research object.

### 3.1. The methods used for analysis and application design include

#### 3.1.1. Analysis

At this stage, the most appropriate problem-solving alternatives are sought to overcome the existing problems. Make the selection of modules according to system requirements.

#### 3.1.2. Design

Designing a troubleshooting system to determine operating steps and procedures and making a comprehensive system design that includes databases and system interfaces.

#### 3.1.3. Implementation

Implementation of the system that has been made, following the specifications specified in the system design.

#### 3.1.4. Testing

After the program is completed, the next step is to test the application.

### 3.2. Research Step

The research was conducted to get the opinion of Twitter users towards public figures' character. To get the desired result, do the steps shown in Figure 1.

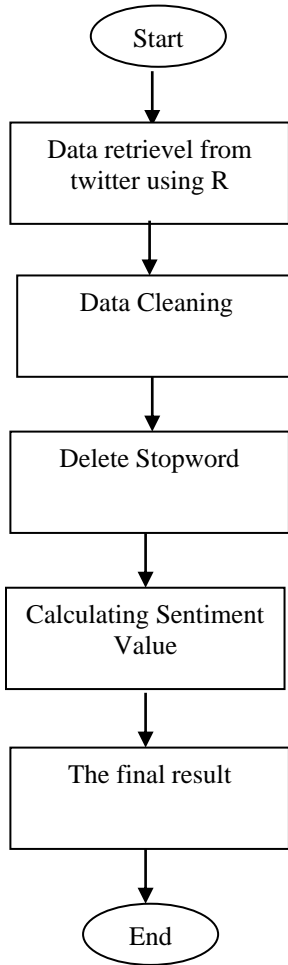


Fig. 1 Research Step

## 4. Results and Discussion

### 4.1. Twitter Data Retrieval

In this study, the public figure who will be the research object is Ridwan Kamil. The criteria to be studied are performance, character, and popularity. To get the data been determined some words to search. For performance criteria, the words used to query, e.g., service and enterprising. For the character criteria, the words used for the query are good and firm, while for the popularity criteria, the words used for the query are like and know.

The data retrieval uses the R language. R has provided a package to access the Twitter API, which allows us to crawl (data retrieval) Twitter data, namely the *twitteR* library. To use the Twitter library, you must have an *api\_key*, *api\_secret*, *access\_token*, and *access\_token\_secret*, which can be obtained by registering with the Twitter web developer (<https://dev.twitter.com/>). A piece of the program is as follows:

```

api_key <- "kode api key"
api_secret <- "kode api secret"
access_token <- "kode access token"
access_token_secret <- "kode access token secret"
setup_twitter_oauth(api_key,api_secret,access_token,access_token_secret)
    
```

Twitter data was obtained after crawling the Twitter timeline with the key of public figure Ridwan Kamil and service criteria and active for performance criteria, good and firm for character criteria, and likes and acquaintances for popularity criteria, as many as 800 tweets consisting of 232 tweet data for character criteria, 334 for performance criteria and 234 for popularity criteria.

### 4.2. Data Cleaning

Data cleaning phase is a process to remove/delete data from punctuation, spaces and connectors in order to get the desired data. A piece of the program is as follows:

```

# removes https://
tweet = gsub('https://', '', tweet)
# removes http://
tweet = gsub('http://', '', tweet)
# removes graphic characters like emoticons
tweet = gsub('[^\p{graph:}]', ' ', tweet)
# removes punctuation
tweet = gsub('[[:punct:]]', ' ', tweet)
# removes control characters
tweet = gsub('[[:cntrl:]]', ' ', tweet)
# removes numbers
tweet = gsub('\d+', ' ', tweet)
tweet = str_replace_all(tweet, "[^\p{graph:}]", " ")
    
```

In addition to removing punctuation marks, this study removed links in tweets, numbers, retweets, and emoticon characters, which were considered not to give meaning to sentiment analysis towards public figures.

### 4.3. Stopword

Stopword is removing/removing words that have no meaning, for example: and, or, there are, are, there are, as for, somewhat, and so on that have no meaning. Stopword uses TALA's proposed word list.

To delete stopwords, the R language uses the *tm* library. A piece of the program is as follows:

```

custom_stopwords <- read.table("id-stopwords.txt")
custom_stopwords <- as.character(custom_stopwords$V1)
#ubah df jadi corpus
documents <- Corpus(VectorSource(tweet_txt))
documents <- tm_map(documents,
content_transformer(tolower))
documents <- tm_map(documents, removeWords,
custom_stopwords)
    
```

```
documents <- tm_map(documents, stripWhitespace)
documents <- tm_map(documents, PlainTextDocument)
tweet_td=documents[[1]]$content
```

**4.4. Sentiment Calculation**

Calculating the sentiment value is based on the words in the list of positive and negative words. Positive words, for example, are eternal, able, true, exciting, abundant, hastening, accepting, recognized, etc.

Calculating the sentiment score is based on the difference between the number of positive and negative words. Positive and negative words play an important role in calculating sentiment. Therefore the completeness of the word list will determine the sentiment. A piece of the program is as follows:

```
neg = scan("kata_negatif.txt", what="character",
comment.char=";")
pos = scan("kata_positif.txt", what="character",
comment.char=";")

score.sentiment = function(tweets, pos.words, neg.words){
require(plyr)
require(stringr)
scores = laply(tweets, function(tweet, pos.words,
neg.words) {
# splits the tweets by word in a list
word.list = str_split(tweet, '\\s+')
# turns the list into vector
words = unlist(word.list)
## returns matching values for words from list
pos.matches = match(words, pos.words)
neg.matches = match(words, neg.words)
## converts matching values to true or false
pos.matches = !is.na(pos.matches)
neg.matches = !is.na(neg.matches)
score = sum(pos.matches) - sum(neg.matches)
return(score)
}, pos.words, neg.words )
scores.df = data.frame(score=scores, text=tweets)
return(scores.df)
}
analysis = score.sentiment(tweet_td, pos, neg)
```

**4.5. Sentiment Calculation Result Graph**

The final data obtained from the previous step will be readable better when displayed in graphical form. The snippet of the program to present the final data in the form of a histogram graph using R is as follows:

```
#histogram
hist(analysis$score, main="Grafik Sentimen Popularitas
Untuk Ridwan Kamil",xlab="Score Tweet", border="blue",
col="green")
```

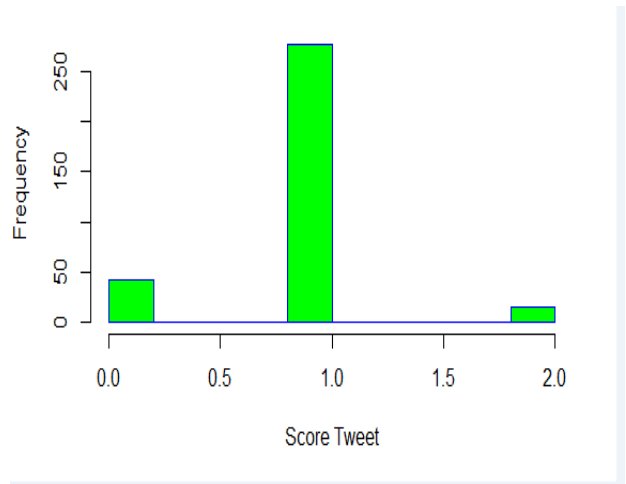


Fig.2 Popularity Graph

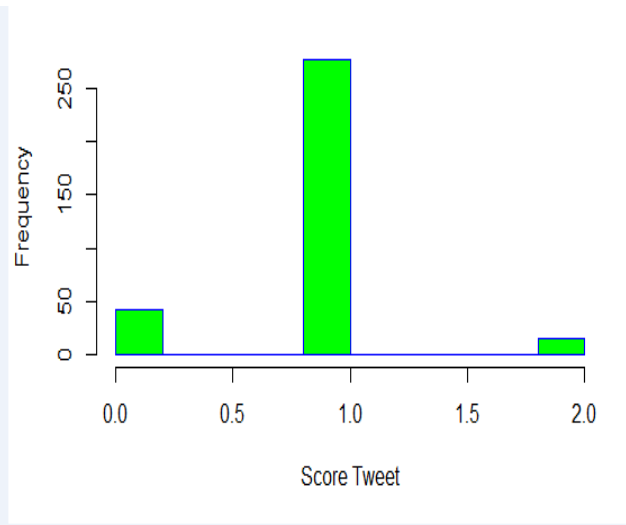


Fig.3 Performance Sentiment Graph

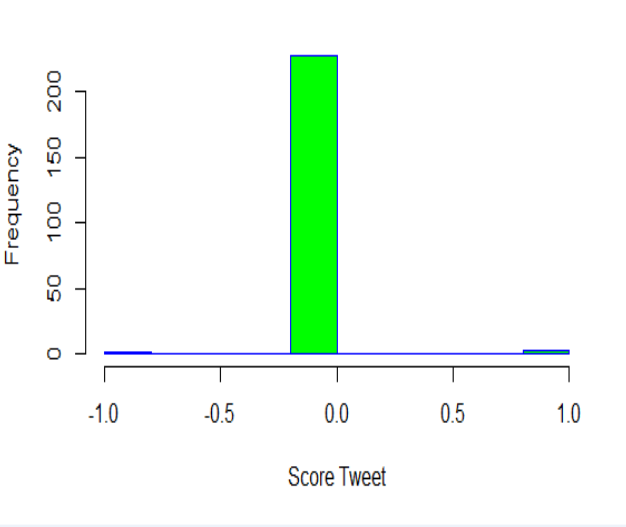


Fig. 4. Character Sentiment Graph

## 5. Conclusion

Based on the research made, the following conclusions can be drawn:

- Many Twitter users give positive, negative, and neutral opinions to public figures, one of which is Ridwan Kamil.
- In this research, sentiment analysis has been carried out through several stages: data collection, cleaning, removing stopwords, calculating, and final sentiment results.
- This research results in the conclusion of positive, negative, and neutral sentiments on Ridwan Kamil's character based on performance, character, and popularity criteria.
- The results of 800 tweets consist of performance criteria, where the measurement is carried out by taking tweets related to the word service and active with 334 tweet data results. For character criteria, tweets related to good and firm words get 232 tweet data results; for popularity criteria, tweets related to words like and know to get 234 tweet data results.

- The results from the graph form show that the sentiment performance criteria generated are neutral, the sentiment character criteria generated are neutral, and the sentiment population criteria generated are neutral. So it can be concluded that there are sentiments for each of these criteria. All neutral means that public figures with multi-criteria sentiment analysis tend to give neutral tweets.

And the suggestions that can be given for future research include:

- It Can be used for further research with different public figures with more criteria.
- The language used is not only Indonesian but can use regional or foreign languages such as English and other foreign languages.

For future research, it is hoped that an application system can be made that can be used to classify sentiments towards political figures or other products.

## References

- [1] APJII, 2016, Survey Penetrasi dan Perilaku Pengguna Internet Indonesia 2016.
- [2] Bata, J. et al., 2015. Leksikon untuk deteksi emosi dari teks bahasa indonesia. , 2015(November), pp.195–202.
- [3] Bollen, J., Mao, H. & Pepe, A., 2011. Modeling Public Mood and Emotion : Twitter Sentiment and Socio-Economic Phenomena. In Proceeding of the Fifth International AAAI Conference on Weblogs and Social Media. pp. 450–453.
- [4] Cheng, J. et al., 2016. Exploring sentiment parsing of microblogging texts for opinion polling on Chinese public figures. Applied Intelligence, 45(2), pp.429–442.
- [5] Hidayat, A.N., 2015. Analisis Sentimen Terhadap Wacana Politik Pada Media Masa Online Menggunakan Algoritma Support Vector Machine Dan Naive Bayes. *Jurnal Elektronik Sistim Informasi Dan Komputer (Jesik)*, 1(1), pp.1–7.
- [6] Hidayatullah, A.F. & Azhari, A.S.N., 2014. Analisis Sentimen Dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter. In *Seminar Nasional Informatika (SEMNASIF)*.
- [7] Liu, B., 2012. *Sentiment Analysis and Opinion Mining*.
- [8] Medhat, W., Hassan, A. & Korashy, H., 2014. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), pp.1093–1113. Available at: <http://dx.doi.org/10.1016/j.asej.2014.04.011>.
- [9] Saraswati, N.W.S., 2011. *Text mining dengan metode naive bayes classifier dan support vector machines untuk sentiment analysis*. Universitas Udayana.
- [10] Forest Stewardship Council (FSC). 1994. *Forest Stewardship Principles and Criteria for Natural Forest Management*. Oaxaca, Mexico.
- [11] Golden, B., F. Harker dan E. Wasil. 1989. *The Analytic hierarchy Process: Applications and Studies*. Springer Verlag, the Netherlands.
- [12] International Timber Trade Organization, 1992. *Criteria for the measurement of sustainable tropical forest management*, Yokohama, Japan.
- [13] Landres, F.B. 1992. Ecological indicators: Panacea of liability? Dalam: McKenzie, D.H., Hyatt, D.E. dan McDonald, J.E. (ed.) *Ecological Indicators*, Volume 2, hal. 1295/1319. Elsevier Applied Science, London.
- [14] Wendoza, G.A. 1997, *Introduction to the Analytic Hierarchy Process: Theory and Application to Natural Resources Management*. Proceedings: Joint Annual Meeting of the American Congress on Surveying and Mapping (ACSM) & American Association of Photogrammetry and Remote Sensing (ASFRS), and Resources Technology Institute (RTI). April 5/10. Seattle, WA.
- [15] Wendoza, G.A. 1997, *A GIS-based Multicriteria approaches to land suitability assessment and allocation*. Proceedings: Seventh International Symposium on Systems Analysis in Forest Resources. Way 28/31. Traverse City, Michigan.
- [16] Wendoza, G.A. dan R. Prabhu. 1998, *Multiple Criteria Decision Making Approaches to Assessing Forest Sustainability Using Criteria and Indicators: A Case Study / Faft I*. CIFOR, Bogor, Indonesia.
- [17] Wendoza, G.A. dan R. Prabhu. 1998, *Multiple Criteria Analysis for Assessing Criteria and Indicators in Sustainable Forest Management: A Case Study on Participatory Decision Making / Faft II*. CIFOR, Bogor, Indonesia.

- [18] Oxford Dictionary of Current English. 1987. Oxford University Press, New York.
- [19] Prabhu, R., C.J.F. Colfer, F. Venkateswafu, L.C. Tan, R. Soekmadi, dan E. Wollenbefg. 1996. Testing Criteria and Indicators for the sustainable management of forests. Phase I. Final Report. CIFOR Special Publication. CIFOR, Bogor, Indonesia.
- [20] Prabhu, R., C.J.F. Colfer, dan R.G. Dudley. 1999. Guidelines for Developing, Testing and Selecting Criteria and Indicators for Sustainable Forest Management. Criteria and Indicators Toolbox Series No. 1. CIFOR, Bogor, Indonesia.
- [21] Saaty, T. 1995. Decision making for leaders: The Analytic Hierarchy Process in a complex world. RWS Publications, Pittsburgh, PA.
- [22] Scientific Certification Systems (SCS). 1994. The Forest Conservation Program: Programme Description and Operations Manual. SCS, California.
- [23] SGS Forestfy. 1994. Assessors Handbook, Policy Document, and Procedures Manual.
- [24] Tropenbos Foundation. 1997. Hierarchical Framework for the formulation of sustainable forest management standard. Tropenbos, The Netherlands.
- [25] Vafgas, L. dan F. Zahedi. 1993. Special Issue. Analytic Hierarchy Process and its applications. Mathematical and Computer Modeling. Vol 17