Sentiment Analysis Of Internet Service Providers

Heribertus Yulianton[1,a)], Felix Andreas Sutanto[2,b)], Kristophorus Hadiono[3,c)]

[a)]heri@edu.unisbank.ac.id
[b)]Corressponding author: felix@edu.unisbank.ac.id
[c)]kristophorus.hadiono@edu.unisbank.ac.id

ABSTRACT: The public's impression of an internet service provider can be analyzed through their responses. One media that can be used to get this impression is Twitter social media. Social media can be used to get an honest response because usually, people will not hesitate to express their feelings indirectly.This study will analyze public sentiment towards internet service providers by considering several aspects such as price, speed, and service. The method consists of three activities; the first is taking tweets from Twitter users that contain words related to internet service providers. The second activity is cleaning data and stopwords. The third is predicting whether the statements written by Twitter users are positive, negative, or neutral. Opinion mining obtained from Twitter users will provide an overview of the percentage of Twitter user comments on internet service providers.The results of this study are in the form of visualizing the sentiments of Twitter users towards three internet service providers. This data can be used as a decision support for internet users in choosing a good internet service provider.

Keywords: sentiment, mining, Twitter, provider, Internet.

INTRODUCTION

The Internet has become a necessity for most people. Not only people from certain groups but all levels of society feel its development. All world information is presented through various media in the form of writing, pictures, and videos. Currently, there are many internet service providers in Indonesia whose existence is beneficial to the community's needs in getting internet access. Nonetheless, the increasing need for public access also requires internet service providers to provide excellent service.

The public's impression of an internet service provider can be analyzed through their responses. One media that can be used to get this impression is Twitter social media. Social media can be used to get an honest response because usually, people will not hesitate to express their feelings indirectly.

This study will analyze public sentiment toward Internet service providers. The method consists of two activities; the first is to take tweets from Twitter users that contain words related to internet service providers. The second process analyzes whether the grouped tweets are categorized as positive, negative, or neutral.

From the details of the description above, the formulation of the problem to be discussed in this study includes:

1. Get tweets from Twitter and classify them into predetermined aspects.
2. Get the Indonesian stopword.
3. Analyze sentiment towards internet service providers.

Twitter is a website owned and operated by Twitter, Inc., which offers a social network through a microblog and social media. It is called a microblog because this site allows users to send and read blog messages as usual, but limited to only 140 characters displayed on the user's profile page. Twitter has unique characteristics and writing formats with special symbols or rules. Messages on Twitter are known as tweets. Twitter has also provided an API for developers who need data. (L. Zhang, 2011).

Sentiment analysis, or opinion mining, is a branch of data mining that aims to analyze, understand, process, and extract textual data in the form of opinions on entities such as products, services, organizations, individuals, and specific topics [1]. Sentiment analysis is a method for analyzing partial data to discover human emotions. Sentiment analysis can be categorized into three tasks: informative text detection, information extraction, and

sentiment interestingness classification (emotional, polarity identification). Sentiment classification (negative or positive) predicts polarity sentiment based on user sentiment data [2], [3].

A tweet has content positive and negative sentiments. Sentiment analysis can be used to identify sentiment content in tweets. [4] did a study on sentiment lexicon-based analysis. The sentiment parameters used in system analysis are very positive, positive, somewhat positive, neutral, somewhat negative, negative, and very negative. Stages system analysis built sentiment _ that is data collection, preprocessing, tokenization, POS tagging, class attribute determination, load dictionary, extract and weighting feature with use Double Propagation method.

METHOD

The data collection for this research is done by abservations that were made at Stikubank University Semarang to obtain data on social media Twitter. In this research, the internet service providers examined are IndiHome, BiznetHome, and MyRepublic. Data regarding user impressions of the three providers was obtained from social media Twitter. As a data sample, a data search was conducted on 1 – 7 July 2017. For each service provider, 500 tweets were randomly taken, and the data was stored in Microsoft Excel for easy presentation of information.

The analysis method of this study is consist 3 steps which is:
1. Analysis
   At this stage, the most practical problem-solving alternatives are sought to overcome the existing problems.
2. Design
   This stage is designing problem-solving to determine operating steps and procedures and creating an overall design includeing data, programming procedures and algorithms.
3. Implementation
   The last stage are implementing of the algorithm made to the specifications specified in the design.

RESULTS AND DISCUSSION

This research will analyze the tweets or comments of Twitter users on the internet service providers they use or something they have heard of before. The steps to be carried out in this study are divided into four parts (figure 1). The first step is to search for data from Twitter and save it. Twitter was chosen because internet users widely use this social media; it provides an API that software developers can access.

After getting the data from Twitter, it is followed by preprocessing, namely data cleaning and stopwords. The third step is determining positive, negative, or neutral sentiments. The final step is to provide a clearer picture of the recap of tweets obtained from Twitter, and an easy-to-understand graph is made.

Figure 1. Research Steps

Data Collection

In this research, the internet service providers examined are IndiHome, BiznetHome, and MyRepublic. The R language is used for processing, which has provided a library to get data from Twitter, namely the twitteR library. Its use is shown in Figure 2.

```
Library (twitteR)

api_key<- "kode api"

api_secret<- "kode api secret"

access_token<- "kode token"

access_token_secret<- "kode token secret"

setup_twitter_oauth(api_key,api_secret,access_token,access_token_secret)


#Pencarian Data

indihome = searchTwitter("indihome". n=500)
```

Figure 2. Twitter Data Search

This research was carried out in two steps to get data on Twitter user comments. The first step is to get the codes to access data from Twitter. You must have a Twitter account first, then get the codes through the Twitter web developer. The second step is to use the searchTwitter() function to find the desired data. The search was carried out using the search method for tweets containing the word isp in this study, namely indihome, biznet, and myrepublic.

Pre Processing

This stage cleans unnecessary text and lists of meaningless words (stopwords). When searching, the twitteR library retrieves data from comments and usernames, replies, retweets, creation dates, locations, and so on. At the same time, the thing that is needed is just a text field. Therefore, in this study, the first thing to do at this stage is to take only the text.

Feature extraction is needed so that the words whose sentiment will be predicted are adequate. In Indonesian, words that have no meaning or are not essential are also known, generally called stopwords. Stopword removal is done based on the list made by Tala with the addition of non-standard words in Indonesian, which are also insignificant.

SENTIMENT PREDICTIONS

The third step in this research is predicting the sentiment of Twitter users towards internet service providers. The process is carried out using the dictionary method. First, make a list of positive and negative words. In this research, the data was obtained from the words given by Bing Liu and translated into Indonesian. The way the prediction process will work is to compare tweets with positive and negative word lists.

To predict positive and negative words by adopting the sentiment algorithm, as shown in Figure 3.

```
score.sentiment = function(sentences, pos.words, neg.words, .progress='none')

{

 require(plyr)

 require(stringr)


 scores = laply(sentences, function(sentence, pos.words, neg.words) {


  word.list = str_split(sentence, '\\s+')

  words = unlist(word.list)


  pos.matches = match(words, pos.words)

  neg.matches = match(words, neg.words)


  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, .progress=.progress)


 scores.df = data.frame(score=scores, x=sentences)

 return(scores.df)

}
```

FIGURE 3. Sentiment function

This function will look for positive and negative word lists in tweet sentences. Then the number will be calculated, and the number of positive words will be subtracted from the number of negative words. The sentence can be classified as a positive sentiment if the result is a positive number. Meanwhile, if the number is the opposite, the sentence tweet is classified as a negative sentiment. If the result shows the number 0, then the sentence is categorized as neutral. The results of predictions can be given examples as **TABLE 1**.

TABLE 1. Predictions Result Example.

Sentence	Results
sobat indihome paket ramadhan lho juli segera berlangganan langsung	neutral
tan paket indihome perbulan berapa ya	neutral
alesadewi pake speedy cari ganti smpt ganti indihome lihat gini tidak hehe	negative
alvionitarani speedytrs indihome speedy versi pake fiber optic hebat	positive
saksikan aksi udin uda kompetisi indihome grass roots festival jumat juli minggu juli	positive
indihome ihhh malesin bgt	negative
rudiantorudy telkomcare selamat siang bapak indihome dikecepatan mbps bapak	positive
kalo provider masuk jogja suami secepat kilat pindah ganti indihome sih	negative
alesadewi akhirnya si bojo temennya janjian komplein langsung tatap muka mata indihome mereka berencana marah marah	negative
duhh tidak ganti indihome deh kalo gitu	negative

The results of the sentiment analysis for internet service providers can be shown in the following figures (**FIGURE 4** to **FIGURE 6**).
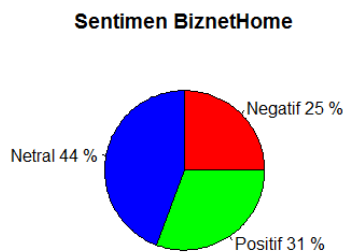

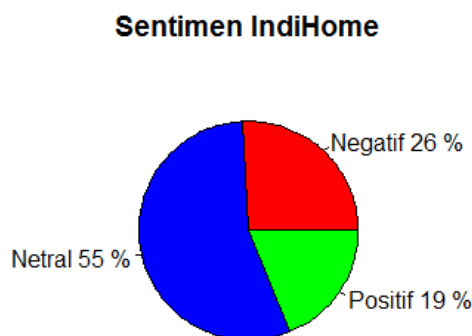**FIGURE 4**. BiznetHome Sentiment Percentage
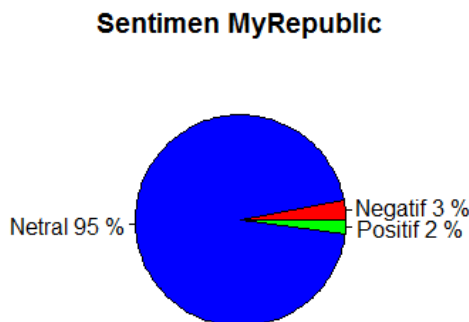

FIGURE 5. IndiHome Sentiment Percentage

FIGURE 6. MyRepublic Sentiment Percentage

From the results obtained from 500 tweets for each internet service provider, it can be seen that IndiHome and BiznetHome have almost the same negative sentiment. However, BiznetHome has more positive sentiments than others. The conclusion is that the two internet providers have almost the same strong position. Meanwhile, MyRepublic has yet to comment much because the tweet data is mainly in the form of promotions, and the company does much of it.

CONCLUSIONS

The conclusions of this study are:
1. Opinions of internet service users can be obtained from social media Twitter.
2. Twitter users determine the contents of the tweet query results; they cannot be conditioned according to the researcher's wishes.
3. Stopwords need to add non-standard words.
4. A dictionary system can be used for mining opinions, but the accuracy is highly dependent on the dictionary's contents.

Suggestions that can be given for further research are:
1. Complete a non-standard vocabulary dictionary because Twitter users may abbreviate or use terms in local languages.
2. Use machine learning as an alternative in predicting Twitter user sentiment towards internet service providers.

REFERENCES

[1]   L. Zhang and B. Liu, "Sentiment Analysis and Opinion Mining," in *Encyclopedia of Machine Learning and Data Mining*, 2016. doi: 10.1007/978-1-4899-7502-7_907-1.

[2]   S. J. Pan, X. Ni, J. T. Sun, Q. Yang, and Z. Chen, "Cross-domain sentiment classification via spectral feature alignment," in *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, 2010. doi: 10.1145/1772690.1772767.

[3]   E. Indrayuni, "Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization," *Jurnal Evolusi Volume 4 Nomor 2 - 2016*, vol. 4, no. 2, 2016.

[4]   G. A. Buntoro, T. B. Adji, and A. E. Purnamasari, "Sentiment Analysis Twitter dengan Kombinasi Lexicon Based dan Double Propagation," *Citee*, no. June, 2014.