OPEN ACCESS
J DATA SCI APPL, VOL. 2, NO. 1, PP.38-48, JANUARY 2019
E-ISSN 2614-7408
DOI: 10.21108/JDSA.2019.2.20

JOURNAL OF DATA SCIENCE AND ITS APPLICATIONS

Sentiment Analysis of Cyberbullying on Instagram User Comments

Alhamda Adisoka Bimantara[1], Afiatari Larasati[2], Ezar Mega Risondang[3], Muhammad Zidny Naf'an[4], Novanda Alim Setya Nugraha[5]

Informatic Department, Institut Teknologi Telkom Purwokerto
Jalan D.I. Panjaitan, No. 128. Purwokerto. Indonesia
[1] 15102007@st3telkom.ac.id
[2] 15102083@st3telkom.ac.id
[3] 15102016@st3telkom.ac.id
[4] zidny@ittelkom-pwt.ac.id
[5] novanda@ittelkom-pwt.ac.id

Received on 12-02-2019, revised on 19-02-2019, accepted on 09-04-2019

Abstract

Instagram is a social media for sharing images, photos and videos. Having many active users from all walks of life, beginning from general users, artists, public figures to senior officials make Instagram becoming as the number one most popular social media based on photo in the world. In addition, to sharing submissions, Instagram users can also give likes and comments to other users' posts. However, the comment feature is often misused, for example used for cyberbullying. Though the actions of cyberbullying include acts that are against the law. But until now, Instagram still does not provide a feature to detect cyberbullying. Therefore, this study aims to create a system that can classify comments whether they contain elements of cyberbullying or not. The results of the classification will be used to detect cyberbullying comments. The algorithm used for classification is Naïve Bayes Classifier. Then for each comment will pass the preprocessing and feature extraction stages with the TF-IDF method. For evaluation and testing using the K-Fold Cross Validation method. The experiment is divided into two, namely using stemming and without stemming. The training data used is 455 data. The best experimental results obtained an accuracy of 83,53% from experiment with stemming process.

Keywords: Cyberbullying, Instagram, Classification, Naïve Bayes Classifier

I. INTRODUCTION

Today, the development of social media is increasing rapidly. With the presence of social media, it allows users to always connect and interact with other social media users directly without being limited by space and time. Until now, there have been many social media that can be used freely, one of which is photo or image based social media, Instagram. Instagram is one of the most popular social media in the world that allows every user to share photos, images and videos. Instagram currently has many active users from all walks of life, ranging from ordinary users, artists, public figures, to state officials. Every Instagram user can also give likes and comments to posts that have been shared.

However, if it is associated with the phenomenon of cyberbullying, Instagram is also ranked first as the most widely used social media for cyberbullying, as a survey conducted by Ditch the Label in 2017 with

participants surveying 10,000 teenagers aged 12 to 20 years who are domiciled in the UK. The survey results show that 42% of cyberbullying victims occur on Instagram. Meanwhile, 37% were victims of cyberbullying on Facebook, and 31% were on Snapchat. Cyberbullying is a form of intimidation, harassment or verbal abuse done continuously in cyberspace [1, 14]. The impact of cyberbullying is more painful than physical bullying. Victims of cyberbullying will experience severe pressure, depression, and even take more extreme actions, namely suicide. While until now, Instagram still does not provide a feature to detect and delete comments that contain elements of cyberbullying automatically. The feature that has been provided by Instagram to minimize the act of cyberbullying is limited to the feature of reporting a report to Instagram, the feature "hide inappropriate comments" to hide comments based on keywords that have been provided by Instagram, and finally a feature to limit or disable comments on each shipment. Where these features are still considered to be less effective in reducing acts of cyberbullying, so there are still many comments that contain elements of cyberbullying sent by users.

Sentiment analysis or opinion mining is the process of understanding, extracting and processing textual data automatically to get sentiment information contained in an opinion sentence [2, 10, 11, 12]. In this study, sentiment analysis was conducted to identify a comment sent by Instagram users to other users, whether the comments contained elements of cyberbullying. Therefore, an algorithm is needed that can classify these comments into positive classes and negative classes. Negative class means comments that contain elements of cyberbullying, and vice versa positive class means that the comments do not contain elements of cyberbullying. There are various algorithms that can be used for classification, such as SVM (Support Vector Machine), NBC (Naïve Bayes Classifier), C45, K-Nearest Neighbours, and many other algorithms.

In this study, the algorithm used for classification is Naïve Bayes Classifier. When viewed by its complexity, Naïve Bayes Classifier is simpler and conventional than other algorithms [13]. So that the computational time needed in the classification process will certainly be shorter. Such research has been conducted by Sentiaji and Bachtiar on [3], that the Naïve Bayes Classifier algorithm can classify an opinion in the form of tweets into two classes namely positive and negative accurately. Then in [4] using the same algorithm, also get a high accuracy of 90%. These results were obtained from testing 100 random data that have been manually classified polarity using 1400 training data. That way, the Naïve Bayes Classifier algorithm looks more suitable to be applied in this study. In addition to the high level of accuracy, the results of the analysis of this study will be displayed in the form of a website-based application so that a short classification time is needed so that the processing of web pages displayed to users is faster and lighter.

II. RELATED WORKS

Previous research on sentiment analysis was conducted by Nurhuda [4], where the study tried to analyze the public sentiment towards the 2014 Indonesian President candidates based on opinions from Twitter. The classification method used is Naïve Bayes Classifier. In this study, the feature extraction stage has been applied with the method used is rule based. This stage is done to get tweets containing opinions or sentiments that have been done before. By testing 100 random data that have been manually classified polarity with 1400 training data, the results of the accuracy are 90%.

Still with the same method namely Naïve Bayes Classifier, Sentiaji [3] analyzes sentiment towards television programs based on public opinion. The data used in the form of tweets taken directly using the Twitter API. Testing is done using a percentage split algorithm by obtaining an accuracy of around 90%. It is assisted by a preprocessing process that aims to delete unnecessary parts and also change the form of documents in the form of tweets to standard forms so that the classification done by Naïve Bayes Classifier becomes more accurate.

Muthia [5] in his research tried to integrate the method of selecting the Genetic Algorithm feature in analyzing sentiment review of a restaurant using the Naïve Bayes Algorithm. The data used is a data review obtained from the zomato.com site, consisting of 100 positive reviews and 100 negative reviews. Before combining the two features, the accuracy obtained was 86.5%. Then after merging the two Genetic Algorithm and Naïve Bayes features get an accuracy of 90.5% where an accuracy increase of 4% occurs. The use of methods other than Naïve Bayes Classifier has been carried out by Nurjanah [6] where the researcher determines the analysis of public opinion on a television show with the method used is K-Nearest Neighbour. The data used is public opinion on television shows on Twitter as many as 400 tweets. In this study textual and non-textual weighting methods when combined can improve system accuracy so that the positive and negative classifications are clearly seen. The results of the accuracy obtained were 80.83% with

the optimal K classification of KNN is k = 3. Then precision reaches 72.28%, recall reaches 100%, and f-measure reaches 83.91%.

Then Buntoro [7] in his research conducted a comparison of the classification methods between Naïve Bayes Classifier and SVM (Support Vector Machine) to analyze sentiments towards candidates for the Governor of DKI Jakarta. By using a dataset in the form of tweets in Indonesian with AHY, Ahok, and Anies keywords as many as 300 tweets, the highest accuracy is obtained when using the Naïve Bayes Classifier method, which is an average value of 95%. While the highest accuracy results when using the SVM (Support Vector Machine) method are 90%.

III. RESEARCH METHOD

A. Data Collection

At this stage data collection will be carried out as needed in this research. The data in question is a collection of comments from various Instagram user accounts. The data collection process is done by utilizing the Instagram API library.

B. Data Labelling

The data that has been taken previously will be labeled. Data will be classified into 2 (two) classes namely cyberbullying and not cyberbullying. The labeling of cyberbullying is given to comments that contain elements of cyberbullying. Then for comments that do not contain elements of cyberbullying, the label will not be labeled as cyberbullying.

C. *Preprocessing*

At this stage, both the training data and the testing data will be made a preprocessing process to be converted into data that is ready for use. Each data will pass through several stages in this preprocessing process, including:

a) Folding Case

In this folding case stage, comments will be changed entirely to lowercase or lowercase letters.

b) Cleansing

Then the cleaning process will be carried out in the comments, namely:

- Eliminate various symbol characters namely punctuation and numbers.

- Eliminating typical text components on Instagram comments, namely usernames, hashtags, e-mail, and URLs (Uniform Resource Locator)

c) Tokenizing

At this stage the tokenizing process will be carried out, i.e. comments will be broken down into word units based on spaces or their constituent words.

d) Word Replacer

Word replacer is used to change the word abbreviation or a word that has a typo (typo) in a comment based on the dictionary that has been prepared.

e) Stop words Removal

This stage serves to eliminate words that are not important and have no meaning in the classification process, for example, is the word to, but, which, with, and so on.

f) Stemming

A comment will be made on the stemming process, which functions to convert the words that are mixed into basic words. The stemming method used in this study is the Nazief-Adriani Algorithm. Stemming is useful for reducing variations of actual words derived from the same basic words.

a) *Case Folding*
In this folding case stage, comments will be changed entirely to lowercase or lowercase letters.
b) *Cleansing*
Then the cleaning process will be carried out in the comments, namely:
- Eliminate various symbol characters namely punctuation and numbers.
- Eliminating typical text components on Instagram comments, namely usernames, hashtags, e-mail, and URLs (Uniform Resource Locator)
c) *Tokenizing*
At this stage the tokenizing process will be carried out, i.e. comments will be broken down into word units based on spaces or their constituent words.
d) *Word Replacer*
Word replacer is used to change the word abbreviation or a word that has a typo (typo) in a comment based on the dictionary that has been prepared.
e) *Stop Word Removal*
This stage serves to eliminate words that are not important and have no meaning in the classification process, for example, is the word to, but, which, with, and so on.

f) *Stemming*
A comment will be made on the stemming process, which functions to convert the words that are mixed into basic words. The stemming method used in this study is the Nazief-Adriani Algorithm. Stemming is useful for reducing variations of actual words derived from the same basic words.

D. Feature Extraction

At this stage the weighting of each term will be carried out based on the level of importance of the term in a set of input documents. The method that will be used for weighting in this study is TF-IDF (Term Frequency - Inverse Document Frequency). TF-IDF will provide word weighting based on statistical values that indicate the frequency of occurrence of a word in the document. The greater the emergence of a word (term) it will provide a greater value of conformity. Because the frequency of occurrence of words (term frequency) is an indication of the extent to which the term represents the contents of the document [8]. Following is the formula for calculating TF-IDF:

$$\text{TF-IDF}(w,d)=\text{TF}(w,d) * \left(\log\left(\frac{N}{DF(w)} \right) \right)$$

Information:

TF-IDF(w,d)	: the weight of a word in the entire document
w	: word
d	: document
TF(w,d)	: frequency of occurrence of a word w in the document d
IDF(w)	: inverse DF of the word w
N	: total number of documents
DF(w)	: the number of documents containing the word w

E. Classification

At this stage the data will be classified using the Naïve Bayes Classifier Algorithm. Naïve Bayes Classifier is a classification of statistics that can be used to predict the probability of membership of a class. At the testing / classification stage, the category value of a document is determined based on the terms that appear in the classified document.

Assumed to have a document collection D = {in | I = 1,2, ... | D |} = {d1, d2, ... d | D |} and collection of categories V = {vj | j = 1,2, ... | V |} = {v1, v2, ... vn | V}. The Naïve Bayes classification is done by finding the probability P (A = vj | D = di), which is the probability of the xj category if known. Documents are seen as tuples of words in the document, namely <a1, a2, ..., an>, whose frequency of occurrence is assumed to be a random variable with a probability distribution. Next, the document classification is to find the maximum value from:

$$V_{MAP} = \underset{vj \in V}{argrmax} P(v_j | a1, a2, \ldots . a_n) \qquad (1)$$

Bayes's theorem about conditional probabilities states:

$$P(B|A) = \frac{P(B|A)P(B)}{P(A)} \qquad (2)$$

Information :

A = Data with unknown classes
B = The data hypothesis A is a specific class.
P(B|A) = The probability of hypothesis B is based on condition A (conditional / posterior probability)
P(B) = Probability of hypothesis B (*prior probability*)
P(A|B) = Probability A is based on conditions in hypothesis B
P(A) = Probability A
Argmax = function that returns an index of the maximum value and a set of data sets

Applying the Bayes theorem to equation (1) can be written:

$$V_{MAP} = \underset{vj \in V}{argmax} \frac{P(a1, a2, \ldots . an|v_j)P(v_j)}{P(a1,a2,\ldots .an)} \qquad (3)$$

The value of P (a1, a2, ..., an) for all vj is the same, the value can be ignored, so equation (3) becomes:

$$v_{MAP=} \underset{vJ \in V}{argmax} P(a1, a2, \ldots . an|v_j)P(v_j) \qquad (4)$$

Assuming that every word in <a1, a2, ..., an is independent, then P (a1, a2, ... an | vj) in equation (4) can be written as:

$$P(a1, a2, \ldots ., an|vj) = \prod_i P(a1|vj) \qquad (5)$$

So equation (4) can be written:

$$V_{MAP} = \underset{vj \in V}{argmax} P(vj) \prod_i P(a_i|v_j) \qquad (6)$$

he value of P (vj) is determined at the time of training, whose value is approached by:

$$P(vj) = \frac{|docj|}{|Example|} \qquad (7)$$

On | docj | is the number of documents that have a j category in training, while | Example | the number of documents in the example used for training. The value of P (wk | vj), that is, the probability of the word wk in category j is determined by:

$$P(wk|vj) = \frac{nk + 1}{n + |vocabulary|} \qquad (8)$$

nk is the frequency of the appearance of the word wk in a document that is categorized as vj, while the value of n is the number of all words in the document categorized as vj, and | vocabulary | is the number of words in the training example.

F. Evaluation and Testing

After obtaining the comment class from the previous classification process, the next step is to conduct evaluation and validation. Evaluation and validation were carried out to determine the results of the accuracy of the experiments that had been done before. The method used is K-Fold Cross Validation.

In this technique, the dataset is randomly divided into a number of K-pieces. Then a number of experiments were conducted, where each experiment used the K partition data as data testing and utilized the rest of the other partitions as training data [9].

G. System Implementation

System implementation is done by making a website-based application using the Python language and microframework Flask. The system will be provided with a feature to detect comments containing elements of cyberbullying on Instagram users by applying all the previous stages in this study, namely preprocessing, feature extraction using TF-IDF, classification with Naïve Bayes Algorithm, and validation using K-Fold Cross Validation.

IV. RESULTS AND DISCUSSION

A. Data Collection

Data collected are comments sent by various Instagram users. The data will be labeled manually and used as training data. The following are examples of data used:

TABLE 1
EXAMPLE OF DATA *TRAINING*

No.	Text
1	Semangat terus kak ayu @aytutingting92, sukses utk karir dan keluarga *(Keep your spirit on kak ayu @aytutingting92, for a successful career and family)*
2	Orang bodoh seperti ini kok bisa terkenal *(How can a fool like this be famous)*
3	Cantik banget jadi makin fans aku sama kak ayu :D *(So beautiful, so I'm getting more and more fans with kak ayu :D)*
4	Dasar cewek alay, sok cantik jijik aku *(Tacky girl, feeling beautiful, I'm disgusted)*
5	Jadi orang jangan bodoh kali mbak *(Being a person, don't be foolish, mbak)*

6	Cantik boleh, bodoh jangan, dasar alay *(Be pretty, stupid not to, you are tacky)*

B. Data Labelling

After the data is collected, the data will be labeled manually from the data that has been obtained. The author gets a total of 455 comments data that have been labeled, consisting of 171 comments containing elements of cyberbullying, and 284 comments that do not contain elements of cyberbullying. The training data is stored in the MySQL database.

C. *Preprocessing*

Both training data and testing data will pass through the preprocessing stage before the classification process is carried out so that the results obtained are more optimal. In table 4 shows the preprocessing results from table 2 data.

TABEL II
PREPROCESSING RESULTS

No.	**Texts**
1	semangat terus kakak ayu sukses karir keluarga *(keep your spirit on kakak ayu, success career family)*
2	orang bodoh kok terkenal *(fool person be famous)*
3	cantik banget jadi makin fans aku sama kak ayu *(so beautiful, so I'm getting more fans with kak ayu)*
4	dasar cewek alay sok cantik jijik aku *(tacky girl, feeling beautiful, I'm disgusted)*
5	jadi orang jangan bodoh kali mbak *(being a person, don't be foolish, mbak)*
6	cantik bodoh jangan dasar alay *(be pretty, stupid not to, you are tacky)*

D. Feature Extraction

Each training data and data testing will be weighted using the TF-IDF technique. The weighting process using TF-IDF starts with finding the number of terms in each document (TF). Then calculate the value of DF, namely the number of documents that have a term. Then the IDF calculation, and finally the TF-IDF calculation where the TF value is multiplied by the IDF value. The following is an example of calculating the word "beautiful" weight in 3 documents using the TF-IDF technique:

TABEL III
DATA EXAMPLE

No	**Texts**
1	cantik banget jadi makin fans aku sama kak ayu *(so beautiful, so I'm getting more fans with kak ayu)*
2	dasar cewek alay sok cantik jijik aku *(tacky girl, feeling beautiful, I'm disgusted)*
3	jadi orang jangan bodoh kali mbak *(being a person, don't be foolish, mbak)*

TABEL IV
RESULTS OF CALCULATION TF-IDF

Word	TF			DF	IDF	TF-IDF		
	D1	D2	D3			D1	D2	D3
cantik	1	1	0	2	0.176	0.176	0.176	0

E. Classification

After passing the feature extraction stage using the TF-IDF technique, the next step is to do the learning process from the training data which will be used for the next process, namely the testing process. Then for testing data that has passed the preprocessing and feature extraction stages, it will be used for the classification process using the Naïve Bayes Classifier algorithm. the next step is the classification process, which is to determine a comment entered in the cyberbullying class or class rather than cyberbullying based on a greater probability calculation value. If the probability results in a comment for the cyberbullying class are greater than the probability results for the class instead of cyberbullying, then the comments enter the cyberbullying class, and vice versa.

F. Evaluation and Testing

After the prediction results from the classification stage are obtained, then the test is performed using the K-Fold Cross Validation technique to obtain the value of accuracy. Before the testing process, first manual labeling of testing data was carried out.

TABEL V
TESTING RESULT

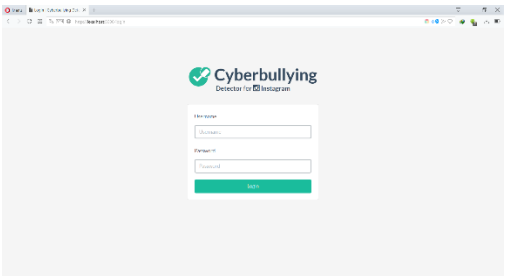
User	**Total Data Testing**	**Accuration of Cross Validation with Stemming process**	**Accuration of Cross Validation without Stemming**
Rizky Darmawan	100	83%	83%
Ayu Ting Ting	100	84%	83%
Presiden Joko Widodo	100	83.6%	83%
Average		**83.53%**	**83%**

G. System Implementation

In this study the program produced is a website-based application system in the Python language and microframework Flask. The program has implemented all the steps that have been done before, starting from data retrieval, preprocessing, feature extraction with TF-IDF, classification using the Naïve Bayes Classifier algorithm, and validation with the K-Fold Cross Validation method. Data retrieval is done in realtime using the InstagramAPI library. Next is the display of each website page::

a) Login Page

Users are required to log in by entering their Instagram account username and password, because InstagramAPI requires access to an Instagram account.



Pict. 1. Login Page

b) Preprocessing Page

The preprocessing page is used to make settings around preprocessing, and the number of comments taken. Users can activate or deactivate the stages in preprocessing.

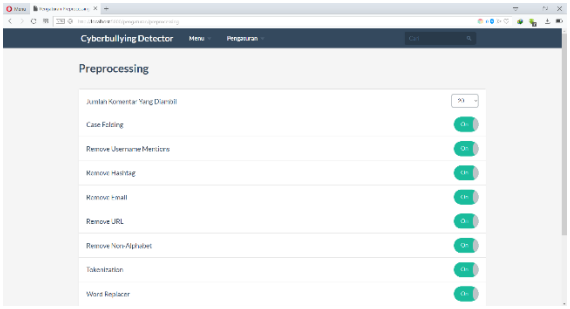


Figure 2. Preprocessing Settings page

c) Training Data Page

Users can also add training data manually on this page. Every comment added will be automatically preprocessed.

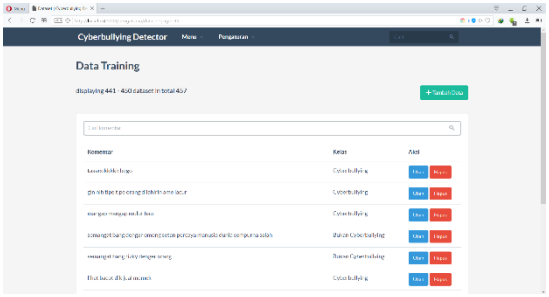


Figure 3. Training Data Page

d) Detection Results Page

This page will display information from the results of detection, such as the total comments detected as cyberbullying or not cyberbullying.

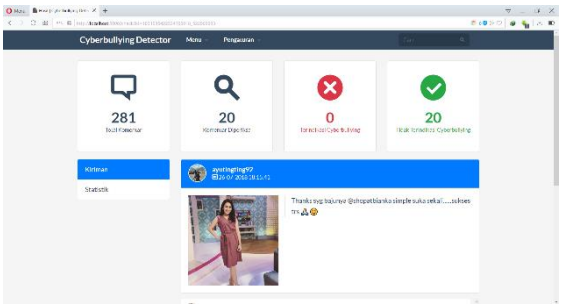


Figure 4. Detection Results Page

e) Statistics page

This page will display statistics from the results of detection in the form of a graph or chart.

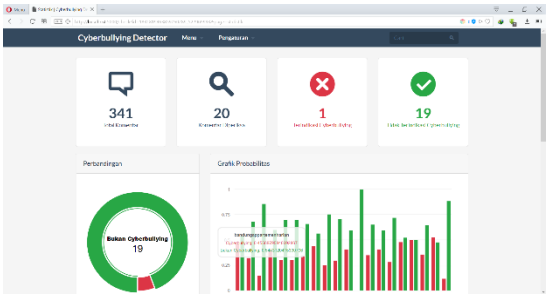


Figure 5. Statistics page

V. CONCLUSION

Based on the research that has been done, it can be concluded that the Naïve Bayes Classifier algorithm can classify comments into cyberbullying classes and not cyberbullying properly. Both using stemming and stemming at the preprocessing stage, both of them get the same best accuracy results, which is equal to 84%. But the use of stemming is quite influential in the number of detection, where the results of experiments conducted on the account @ayutingting92 get the number of comments cyberbullying as many as 12 comments with stemming, and as many as 8 comments cyberbullying with no stemming.

REFERENCES

[1] F. Rohman, "Analisis Meningkatnya Kejahatan Cyberbullying Dan Hatespeech Menggunakan Berbagai Media Sosial," *SNIPTEK*, pp. 382–387, 2016.

[2] S. A. F. Alvi Pranandha Syah, Adiwijaya, "Analisis Sentimen Pada Data Ulasan Produk Toko Online Dengan Metode Maximum Entropy Sentiment Analysis on Online Store Product Reviews With Maximum," *e-Proceeding Eng.*, vol. 4, no. 3, pp. 4632–4640, 2017.

[3] A. M. B. Aditia Rakhmat Sentiaji, "Analisis Sentimen Terhadap Acara Televisi Berdasarkan Opini Publik," *J. Ilm. Komput. dan Inform.*, vol. 2, no. 1, pp. 55–60, 2014.

[4] F. Nurhuda, S. W. Sihwi, and A. Doewas, "Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier," *ITSmart J. Ilm. Teknol. dan Inf.*, vol. 2, no. 2, pp. 35–42, 2014.

[5] D. A. Muthia, "ANALISIS SENTIMEN PADA REVIEW RESTORAN DENGAN TEKS BAHASA INDONESIA MENGUNAKAN," *J. Ilmu Pengetah. dan Teknol. Komput.*, vol. 2, no. 2, pp. 39–45, 2017.

[6] W. E. Nurjanah, R. S. Perdana, and M. A. Fauzi, "Analisis Sentimen Terhadap Tayangan Televisi Berdasarkan Opini Masyarakat pada Media Sosial Twitter menggunakan Metode K-Nearest Neighbor dan Pembobotan Jumlah Retweet," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 12, pp. 1750–1757, 2017.

[7] G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," *Integer J.*, vol. 2, no. 1, pp. 32–41, 2017.

[8] H. A. S. Vipy Wahyu Perdana, "Analisis Penerapan Algoritma Naive Bayes dalam Pengklasifikasian Konten Berita Bahasa Indonesia," *Univ. Dian Nuswantoro Semarang*, no. 5, pp. 5–6, 2014.

[9] D. P. Rendra Dwi Lingga P, Chastine Fatichah, "Deteksi Gempa Berdasarkan Data Twitter," *J. Tek. ITS*, vol. 6, no. 1, pp. 159–162, 2017.

[10] Asriyanti Indah Pratiwi, Adiwijaya. "On the Feature Selection and Classification Based on Information Gain for Document Sentiment Analysis." Applied Computational Intelligence and Soft Computing, 2018.

[11] Al Faraby, S., Jasin, E.R.R. and Kusumaningrum, A., "Classification of hadith into positive suggestion, negative suggestion, and information." In *Journal of Physics: Conference Series* (Vol. 971, No. 1, p. 012046). IOP Publishing. 2018

[12] Mubarok, M.S., Adiwijaya and Aldhi, M.D., "Aspect-based sentiment analysis to review products using Naïve Bayes". In AIP Conference Proceedings (Vol. 1867, No. 1, p. 020060). AIP Publishing. 2017.

[13] Manuel B., Tricahyono D., "Classifying Electronic Word of Mouth and Competitive Position in Online Game Industry", Journal of Data Science and Its Applications (JDSA), 1(1), pp.20-27. 2018

[14] Alamsyah, A. and Syawiluna, M., Mapping Organization Knowledge Network and Social Media Based Reputation Management". *Journal of Data Science and Its Applications*, *1*(1), pp.39-48. 2018.