EMITTER International Journal of Engineering Technology
ISSN: 2443-1168, Vol. 8, No. 1, June 2020, pp. 200~220
DOI : 10.24003/emitter.v8i1.477

Unsupervised Twitter Sentiment Analysis on The Revision of Indonesian Code Law and the Anti-Corruption Law using Combination Method of Lexicon Based and Agglomerative Hierarchical Clustering

Nur Restu Prayoga[1], Tresna Maulana Fahrudin[2], Made Kamisutara[3], Angga Rahagiyanto[4], Tahegga Primananda Alfath[5], Latipah[6], Slamet Winardi[7], Kunto Eko Susilo[8]

[1,2,3,6,7,8]Faculty of Computer Science, Universitas Narotama
[4]Department of Medical Records, Politeknik Negeri Jember
[5]Faculty of Law, Universitas Narotama
Email: [1]nurrestuprayoga@gmail.com, {[2]tresna.maulana, [3]made.kamisutara, [5]tahegga.primananda, [6]latifah.rifani, [7]slamet.winardi, [8]kunto.eko.susilo}@narotama.ac.id, [4]rahagiyanto@polije.ac.id

Received January 28, 2020; Revised April 19, 2020; Accepted May 16, 2020

Abstract

The rejection on ratification of the revision of Indonesian Code Law or known as RKUHP and Corruption Law raises several opinions from various perspectives in social media. Twitter as one of many platforms affected, has more than 19.5 million users in Indonesia. Twitter is one of many social media in Indonesia where people can share their views, arguments, information, and opinions from all points of view. Since Twitter has a great diversity of users, it needs a system which is designed to determine the opinion tendency towards the problems or objects. The purpose of this study is to analyze the sentiment of Twitter users' tweets to reject the revision of the Law whether they have positive or negative sentiments using the Agglomerative Hierarchical Clustering method. The data that being used in this study were obtained from the results of crawling tweets based on hashtag (#) (#ReformasiDikorupsi). The next stage is pre-processing which consists of case folding, tokenizing, cleansing, sanitizing, and stemming. The extraction features Lexicon Based and Term Frequency (TF) which performs the process automatically. In the clustering stage, two clusters use three approaches; single linkage, complete linkage and average linkage. In the accuracy calculation phase, the writer uses the error ratio, confusion matrix, and silhouette coefficient. Therefore, the results are quite good. From 2408 tweets, the highest accuracy results are 61.6%.

Keywords: Tweet, Law Revision, Sentiment Analysis, Clustering, Agglomerative Hierarchical Clustering.

1. INTRODUCTION

The revision of the Corruption Eradication Commission or it is known as KPK and of Indonesian Code Law or known as RKUHP which was an initiative of the Publics' Representative Council (DPR) is gained a polemic because it contained a number of regulations that were considered to weaken the Corruption Eradication Commission (KPK). A lot of students from various universities have conveyed a motion of no confidence to the Publics' Representative Council (DPR) in a demonstration or even in a social media and one of the examples of the social media is Twitter. The formed of the message is delivered throughout the Twitter which is done by the students that related with the revision of the bill (RUU) of Corruption Eradication Commission (KPK) has become a polemic and make a hashtag (#) #ReformasiDikorupsi. Those tweets raise many opinions from Twitter users. Currently, public discussion on social media is one of the interesting things to study. From the topic of discussion, it produced comments that mostly contained sentiment opinions [1].

That is why it needs a system which is able to consider the opinions or opinions tendencies on an issue or an object using analysis of opinions or sentiment (opinion analysis or sentiment analysis). Sentiment analysis or opinion mining is a computational study of people's opinions, the sentiment through entities and attributes that are expressed in text form in sentences or documents in order to find out the opinions which expressed in those sentences or documents whether they are positive, negative or neutral.

This research conducts sentiment analysis by classifying Indonesian Twitter data. Sentiment analysis is very necessary in filtering comments on social media. Sentiment analysis on comments is done to find out negative comments and positive comments [2]. The data will be processed by text mining to avoid lacking data and then combining the tweet data based on a hash tag #ReformasiDiKorupsi using the Agglomerative Hierarchical Clustering algorithm. The data which used in this study were taken as many as 2408 Indonesian tweets that containing the hash tag #ReformasiDiKorupsi and the exraction feature using lexicon based and term frequency (TF) and clustered with a single linkage, complete linkage, and average linkage approach. Furthermore, the accuracy stage uses the error ratio, confusion matrix and silhouette coefficient.

2. RELATED WORKS

Muqtar Unnisa, et al [3] from Darussalam Hyderabad TS. Based on the results of research that has been done, the first research is measuring public opinion with a reasonable assessment in film reviews with a machine learning algorithm based on spectral clustering for sentiment analysis. The results of testing on two thousand tweets indicate that it is not specific to film reviews and can easily be applied to other domains with a fairly large corpus.

Bo Wang, et al [4] from The University of Warwick. Based on the results of research that has been done, the second research is to measure interesting topics in a tweet because that tweet is only a few characters and is noisy. The results of this test show a two-stage hierarchical topic modelling system,

named GSDMM + LFLDA, which utilizes a structured Twitter topic model, a topic models with embedding with words entered and steps to merge tweets without the use of metadata in any form. The results showed an approach that outperformed other methods and other grouping-based topic models, both in topic classification.

Andrii Yu, et al [5] from the Institute of NASU-NSAU. Based on the results of research that has been done, the third research is software for crawling data that allows users to analyze big databases to solve business decision problems. Crawling data in some ways, is an extension of statistics, with some artificial intelligence and machine learning. This study shows the clustering method can support investors' decisions to choose investment stocks. So, identify the group of companies from the shares given by the market.

Annisa, et al [6] based on the results of research that has been done, the fourth research is an analysis to create an automatic summary (text summarization) for multi tweet based on Twitter's trending topic. Text summarization is a process that automatically generates summary information that is useful for the user. The results of testing the tweet data are the results in the form of clusters and evaluation results. Evaluation results will be analyzed to draw a conclusion.

3. ORIGINALITY

Twitter users are not only limited to social media for friendship, but Twitter is also used as a promotional and campaign tool [7]. Sentiment analysis of Twitter data became a research tread in the last decade. Among popular social network portals, Twitter had been the point to fascination to several researchers in important areas like the prediction of democratic events, customer brands, movie box-office, stock market, the reputation of personalities etc [8]. Social media like Twitter is widely used to pour out the hearts of its users so that the data can describe sentiment [9]. The advantage of this research is that the data clustering process is carried out using two methods, the first is feature extraction using lexicon based. Lexicon based are divided into two data, namely positive word data and negative word data that has been labelled and grouped manually by humans. Second, by using feature extraction with term frequency which automatically has an automatic word dataset that can set the number of words that represent the topic of rejection of this RUU. The two feature extractions are then categorized into positive sentiment and negative sentiment by evaluating using an error ratio with label data that was previously labelled sentiment 1 and -1 for comparative evaluation data because the error ratio requirements must have classified data. And the second evaluation using confusion matrix to find out is used to determine the distribution of data accuracy of each label after clustering. With the confusion matrix, we can find out the accuracy of each positive and negative label with the recall method [19]. And the third is by using Silhouette Coefficient to see how well the clusters are formed. Values range from -1 and 1. The closer to zero means the cluster density is not good. Conversely the more away from zero clusters the better. All of these processes will be visualized on a dendrogram and there is a predict feature to find out which

tweets have positive or negative sentiment values. And this study uses a combination of term frequency where the results will be integrated with agglomerative hierarchical clustering automatically and the results will be compared using data that has been obtained on a lexicon based that is processed automatically with output in the **form** of sentiment comparisons and evaluations in the form of visual results. This combination process makes this research better without having to sort one by one with human logic because the combination of term frequency, agglomerative hierarchical clustering and lexicon based automatically processes the system with good results.

4. SYSTEM DESIGN

The proposed system consists of 5 phases: (1) Data collection, (2) Data preprocessing, (3) Features extraction, (4) Clustering using Agglomerative Hierarchical Clustering method and evaluation, (5) Visual result. The whole system design is shown in Figure 1. Each phase on system design will be explained in part 4.1-4.5.

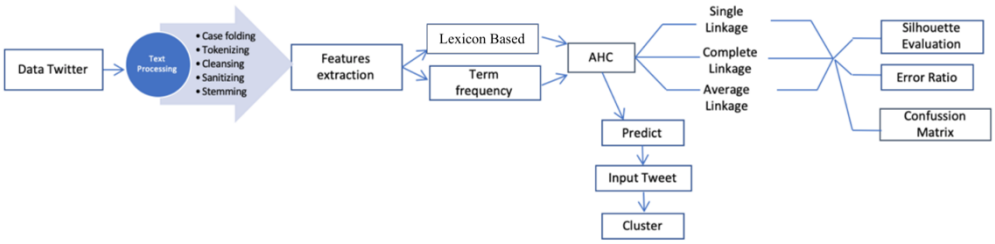


Figure1. The system design of proposed research

4.1 Sentiment Analysis

Sentiment analysis or opinion mining is a process of understanding, extracting, and textual data processing to acquire opinions regarding the sentiment. These data play an important role as feedback product, services, and other topics [6]. Sentiment analysis also called opinion mining, which is the process of extracting an opinion or opinion from a document for a particular topic.

4.2 Feature Extraction

Feature extraction is used to obtain the attributes that represent the topics using the centre of the cluster to choose attributes of the data. Sentiment analysis or opinion mining is a process of understanding, extracting and processing textual data automatically to get the sentiment information contained in an opinion sentence [10]. This attribute has most contribution in distinguishing clusters. The feature extraction in this study uses two methods; lexicon based and term frequency (TF). Lexicon based method has a database in which it list of positive and negative words manually. Thus, this database has become an automatic table dictionary. On the other hand, the term frequency is an algorithm which is used to calculate the content of each word that has been extracted. TF is the total emergence of each word on the

document, the more words emerge in each document, the more the TF value will appear.

4.3 Agglomerative Hierarchical Clustering

Agglomerative Hierarchical Clustering is a hierarchical clustering method that uses an approach called a bottom-up approach. The process of grouping started from each data as one cluster, then recursively looking for the closest cluster as a pair to join as one larger cluster. The process is repeated so that it moved up and formed into a hierarchy [5]. There are three techniques that can be used to calculate the approach between two clusters, it is called as single-linkage (the closest distance), complete linkage (the furthest distance), and average linkage (the average). The result of the linkage clustering is presented in a graphical form of a dendrogram or tree diagram. The branches on the diagram represent the cluster itself. Then, the branches merge into a node which has the same length as the axis distance where the merging has occurred.

4.4 Data Collection

The process of data collection comes from a tweet related to the polemic of the law. This process carried out to explore, process and manage information and to analyze the textual relationship of structured and unstructured data [11]. The data which has been collected is taken are tweets that used the hashtag (#) #ReformasiDikorupsi. The data is obtained by creating a crawling program using the Python Programming Language with the library tweepy and use the Twitter API that was obtained previously at http://developer.twitter.com. Tweepy is an API that helps to authenticate a user and allows the user to access Live Twitter Data by just creating a Twitter application and retrieving your API access keys and tokens [12].

Twitter API consists of consumer key, consumer secret, access token, and access token secret. The Twitter Search API emphasizes the search function of the past while the Twitter Streaming API emphasizes the search function of the future [13]. Crawling data was conducted step by step from September 25th, 2019 to September 28th, 2019. As a result, there is 2409 tweets obtained and stored in the excel CSV (Comma-Separated Value) format. Data crawling is conducted by entering the query '#ReformasiDikorupsi'. This query contains all tweets which consist of #ReformasiDiKorupsi that is relevant to the topic of rejection of the law and use an additional filter no-retweet. For more details, it can be considered in Table 1.

4.5 Data Preprocessing

The data that has been collected is still the raw data of this analysis. Therefore, the purpose of preprocessing is done by selecting the text and removing the parts that are not needed. Preprocessing is used to analyze the text segmentation so that the characteristics of the text can be assessed, analyzed, and classified [6]. The preprocessing steps taken are:

1. Case folding, in this stage, there will be some changes for all documents into lowercase fonts from a-z.

2. Tokenizing, in this stage, there will be some cutting process on a document which is called a token. Moreover, there will be a removal of whitespace.
3. Normalization, in this stage there is a process of correcting non-formal words such as slangs, abbreviations, and misspellings. This normalization process is done by matching tokens of tokenizing results with a standard word dictionary that has been made by the author. This process is done by matching each word in the training data document and test data with the words in the non-standard language dictionary [14].
4. Cleansing, this stage is conducted in order to clean up the features that are not needed, or it is called as cleaning or deleting all documents that contain url (http: //), username (@), hash sign (#), delimiters such as comma (,) and period (.) and other punctuation marks. The maximum limit of characters in a tweet is 140 characters so that most tweets contain words in abbreviations [15].
5. Sanitizing is conducted by removing unnecessary words. Common words will be deleted to reduce the number of words saved and it will be processed later. If the word is discarded, it will not change or eliminate the information contained in the sentence. For instance, conjunctions like a will, in, on, and others. Data cleaning is necessary for data pre-processing because not all the components are useful for the sentiment analysis task. Normally, the noise phrases, stopwords and meaningless symbols are removed [16].
6. Stemming is conducted to find out the basic word of stem from the results of stopword which removed the affix of the word. The added affixes that are omitted consist of the prefix, suffix, insert (infix), and the combined prefix (confix). In this study, it used the literary python library for the stemming process.

Table 1. Data preprocessing

Input	Case Folding	Tokenizing	Normalization	Sanitizing	Stemming
Negeri ini hebat karena masih banyak pemuda-pemudi yg peduli dengan masa depan Bangsa Indonesia. MERDEKA!! #STMmelawan #DennySiregarPenyebarHoax #MahasiswaBergerak #stmmahasiswabersatu'	negeri ini hebat karena masih banyak pemuda-pemudi yg peduli dengan masa depan bangsa indonesia. merdeka!! #stmmelawan #dennysiregarpenyebarhoax #mahasiswabergerak	[“perkuat”, “kaum”, “elite”, “tumpul”, “kaum”, “marjinal”, “#stmmelawan”, “#mahasiswabergerak”, “#massageplusplus”, “#mahasiswapelajaranarkis”, “//t.co/c5uwosdolg	[“negeri”, “ini”, “hebat”, “karena”, “masih”, “banyak”, “pemuda-pemudi”, “yang”, “peduli”, “dengan”, “masa”, “depan”, “bangsa”, “indonesia”, “merdeka”]	[“negeri”, “hebat”, “masih”, “banyak”, “pemuda”, ”pemudi”, “peduli”, “masa”, “depan”, “bangsa”, “indonesia”, “merdeka”]	[“negeri”, “hebat”, “masih”, “banyak”, “pemuda”, ”pemudi”, “peduli”, “masa”, “depan”, “bangsa”, “indonesia”, “merdeka”]

4.6 Feature Extraction

A sentence is represented as an object and the words that make it up are represented as features [15]. Feature extraction uses two different methods; lexicon based and term frequency. It aims to find and compare the best feature extraction for Twitter polemic against the rejection of the Law (RUU). Textual data will form with as many objects as there are and the number of features is different [15].

4.6.1 Lexicon based

In this extraction feature, firstly, the lists of Lexicon Based need to be defined. The list of Lexicon Based consists of two classes that have been created manually, i.e. what words will be included in the category of words that represent positive and negative. The sentence has a score> 0 will be classified in the positive class, if the sentence has a score = 0 will be classified in the neutral class, whereas if the sentence has a score <0 classified in the negative class [17]. Figure 2 is an example of 13 positive words out of 1182

positive words and 13 negative words out of 2402 negative words that reflect on the topic of the rejection of the bill.

	A	B
1	Positive words	Negative words
2	acungan jempol	abnormal
3	adaptif	absurd
4	adil	acak
5	afinitas	acak-acakan
6	afirmasi	acuh
7	agilely	acuh tak acuh
8	agung	adiktif
9	ahli	agresi
10	ahlinya	agresif
11	ajaib	agresor
12	aklamasi	aib

Figure 2. The illustration words of lexicon based

And the result of clustering with lexicon based on Table 2.

Table 2. Clustering with lexicon based

Tweet	(+)	(-)	cluster
suara rakyat tidak rakyat mahasiswa ideologi lawan ketidakberesan sistem	0	2	-1
pas udah lumayan asep gandeng sengaja lepas blm sempet liat orgnya blm sempet bilang makasih thanks for everything abang sehat abang semangat indonesia	4	0	1

4.6.2 Term Frequency (TF)

The process of valuing the words aims to calculate the value of each word to be used as a feature. The more documents being processed, the more features will appear. TF is the amount of the emergence of each word in a document. Thus, the more words appear in each document, the more value will be on the TF.

4.7 The Grouping Method of Agglomerative Hierarchical Clustering

To analysis the sentiment automatically and more reasonably, unsupervised machine learning methods have drawn wide attention, for example, the clustering methods [16]. Agglomerative Hierarchical Clustering is a hierarchical clustering method where each data will be grouped based on its proximity characteristics. Clustering is the process of discovering homogeneous groups among a set of objects [18]. This grouping process uses three methods; single linkage, complete linkage, and average linkage.

1. Single Linkage (the closest distance)

$$d_{uv} = \min\{d_{uv}\}, d_{uv} \in D \quad (1)$$

2. Complete Linkage (the furthest distance)

$$d_{uv} = \max\{d_{uv}\}, d_{uv} \in D \quad (2)$$

3. Average Linkage (the average distance)

$$d_{uv} = average\{d_{uv}\}, d_{uv} \in D \quad (3)$$

Agglomerative Hierarchical Clustering bottom-up procedure initializes a trivial partition composed of singletons then, iteratively merges the two closest clusters until all items are grouped together [18].

4.8 The Performance Evaluation of Clustering

The accuracy testing is conducted in order to be acquainted with the accuracy of the results of the grouping. In testing the accuracy, there will be three aspects that are used; error ratio, confusion matrix, and silhouette coefficient and visualization to dendrogram. A dendrogram is more informative than a single partition because it provides more insights about the relationships between objects and clusters [18].

1. Error ratio is used to find out how big is the label of accuracy clustering to manual labels. Previous tweets have been manually labelled, then compared to the results of clustering labels. The use of a combination method for each label clustering occurred because it does not know whether the clustering label represents positive or negative.
2. Confusion matrix is used to find out how big the success of the system. Confusion matrix was chosen as a measure of evaluation because the data used in this study already had a label.
3. Silhouette coefficient, will perceive how well the cluster is formed. The values are ranged from -1 and 1. The closest number to zero means the cluster density is not good and vice versa.

5. EXPERIMENT AND ANALYSIS

The data collections in this study were taken from one of social media named Twitter which uses the Twitter API (Application Programming Interface). This study was conducted periodically during 23-28 September 2019. Tweets that had been taken only contain the hashtag #ReformasiDiKorupsi because the hashtag is relevant to what happened to during the demonstration. Through the crawling process, there were as many as 2408 data tweets obtained, consisting of random data containing positive tweets supporting the draft law and negative tweets that did not support the draft law. Furthermore, preprocessing data is done to reduce noise and missing value through several stages, namely case folding, tokenizing, normalization, sanitizing and stemming. After preprocessing, feature extraction will then be performed using two different methods, namely lexicon based and term frequency. It aims to find and compare which feature extractions are best for research studies of bill opinions on social media Twitter.

All tweets with a total of 2408 tweets will be extracted using the lexicon based feature extraction algorithm. The use of lexicon based dictionaries makes the data possible to find out, especially about the distribution of

negative and positive words in the total of 2408 tweets.

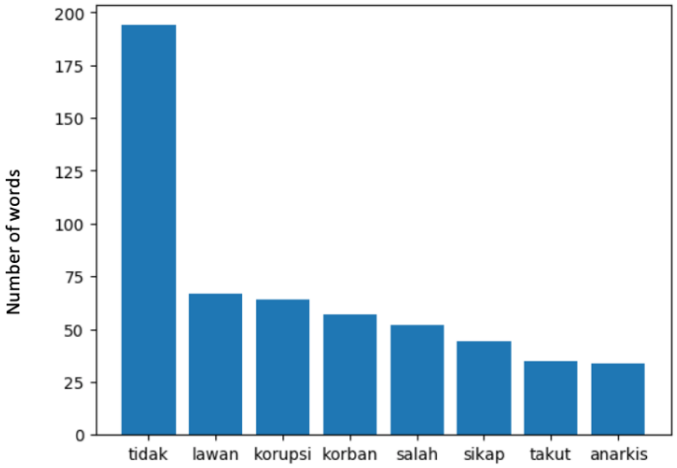


Figure 3. Word frequency on negative clustering

Figure 3 shows that the word "no" is the most frequently used by Twitter users in expressing their opinions about the bill (RUU), which is close to 200 words. Meanwhile, the word that is least used by Twitter users in expressing opinions about the bill is the word "anarchist".

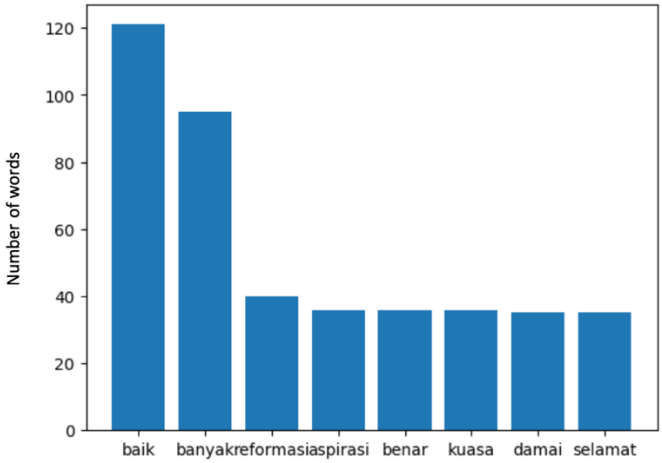


Figure 4. Word frequency on positive clustering

Figure 4 shows that the words "good" and "many" are dominating the tweets about the bill (RUU) with 120 and 90 words.

The next step is performing the extraction feature using the term frequency (TF) algorithm by comparing the frequency of the maximum value term of the whole or a set of term frequencies in a document. However, the TF number of features is used in such a way as to get maximum results that are automatically obtained from existing tweets.

Table 3. Word feature

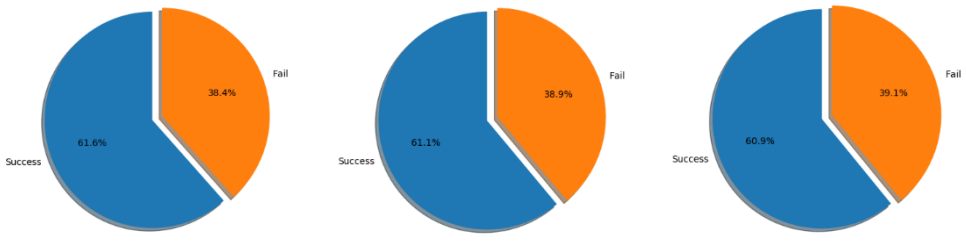
Word Feature	
ada	mahasiswa
aja	negara
aksi	orang
baik	perppu
banyak	polisi
demo	rakyat
dpr	revisi
indonesia	ruu
jalan	september
jangan	suara
kpk	tidak
lebih	tolak

Table 3 is an automatic word feature used with 24 words that best represent tweets that are relevant to the topic. Word feature converts words into features for use in TF processing. Evaluating each word is a simple method it is determined by the number of occurrences of words in a tweet containing #ReformasiDiKorupsi. The method of assessing words using TF is superior to TF-IDF. Therefore, the feature suitable for use is term frequency, because it is more sufficient to represent the tweet. Whereas when using the inverse document frequency (TF-IDF) term, the feature value will represent the tweet of the entire tweet so that it requires more computational calculations than just counting the word frequency on the tweet, so in the case study this study is sufficient to only use the term frequency.

The representation of the number of positive and negative words will greatly affect the clustering process, whether these two positive and negative features can represent each tweet or not.

Next step is the clustering stage, in which it uses the Agglomerative Hierarchical Clustering method. That method uses three combinations of variations between AHC clusters, called single linkage, complete linkage, and average linkage with three types of performance evaluation methods, called error ratio, confusion matrix, and silhouette coefficient.

5.1 Clustering Evaluation with Lexicon based as a Feature

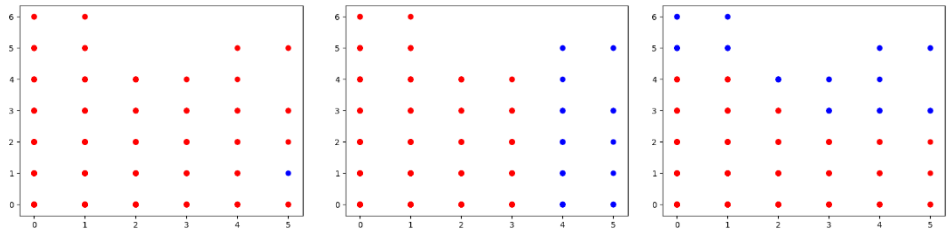


(i) (ii) (iii)

Figure 5. Circle diagram of proportion results (i) Single Linkage, (ii)Complete Linkage, (iii) Average Linkage on Lexicon based

On Figure 5 (i) it shows that the error ratio of a single linkage shows good results with an accuracy rate of 61.58637873754152%, on Figure 5 (ii) also shows that the error ratio of the complete linkage shows satisfactory results with an accuracy rate of 61.088039867109636%, while in Figure 5 (iii) shows that the error ratio of the average linkage shows quite satisfactory results with an accuracy rate of 60.92192691029901%.

Furthermore. The next step is to plot the result by using scatterplot matplotlib as shown in Figure 6.



(i) (ii) (iii)

Figure 6. Plot (i) Single Linkage, (ii) Complete Linkage, (iii) Average Linkage

Figure 6 shows that the red plot indicates a negative cluster while the blue plot indicates a positive cluster. Figure 6 (i) shows a pretty bad result since it showed that in a single linkage it only has one positive cluster in blue. Figure 6 (ii) shows that complete linkage results in a fairly good cluster visualization. Figure 6 (iii) shows that the average linkage results in a fairly good visualization of the cluster even though some plots are still wrong. The ideal cluster division plot is shown in Figure 7 where the cluster is ideally divided into two parts as if divided by a blue line.

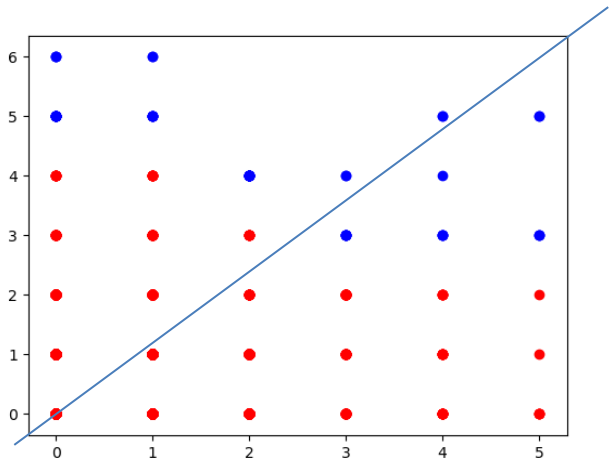


Figure 7. The ideal plot cluster

Confusion matrix table is needed to calculate the accuracy result and the classification result.

Table 3a. The prediction of confusion matrix of single linkage

Label	-1 (predicted)	1 (predicted)
-1 (actual)	TN (1483)	FP (1)
1 (actual)	FN (924)	TP (0)

Table 3a explains that as many as 1483 tweets were correctly detected negatively in clustering while one tweet was incorrectly predicted by the AHC single linkage clustering algorithm. As for the positive label (1) there are no tweets that are correctly predicted by the single linkage of AHC clustering algorithm. In the confusion matrix, it is clear that the results of the AHC single linkage clustering algorithm are not good enough to be used as a benchmark because it is very good for negative tweets but very bad for positive tweets.

Table 3b. The Prediction of confusion matrix of complete linkage

Label	-1 (predicted)	1 (predicted)
-1 (actual)	TN (1456)	FP (28)
1 (actual)	FN (909)	TP (19)

Table 3b explains that as many as 1456 tweets were correctly detected negatively in clustering while 28 tweets were incorrectly predicted by the AHC complete linkage clustering algorithm. As for the positive label (1) there are 19 tweets that were correctly predicted by the AHC complete linkage clustering algorithm. In the confusion matrix, it is clear that the results of the AHC complete linkage clustering algorithm are not good enough to be used as a Patoka because it is very good for negative tweets but very bad for positive tweets. However, this result is better than the single linkage algorithm.

Tabel 3c. The Prediction of confusion matrix of average linkage

Label	-1 (predicted)	1 (predicted)
-1 (actual)	TN (1447)	FP (37)
1 (actual)	FN (904)	TP (20)

Table 3c explains that as many as 1447 tweets were correctly detected negatively in clustering while 37 tweets were incorrectly predicted by the AHC average linkage clustering algorithm. As for the positive label (1) there are 904 tweets that are correctly predicted by the AHC average linkage clustering algorithm. In the confusion matrix, it is clear that the results of the AHC average linkage clustering algorithm are not good enough to be used as a benchmark because it has more negative tweets but less for positive tweets. However, this result is better than the single linkage algorithm.

The silhouette coefficient performance algorithm only known whether the clusters are formed or not.

(i)
```
In [69]: labels = clustering.labels_
         metrics.silhouette_score(matriks, labels, metric='euclidean')
Out[69]: 0.6134522816367602
```

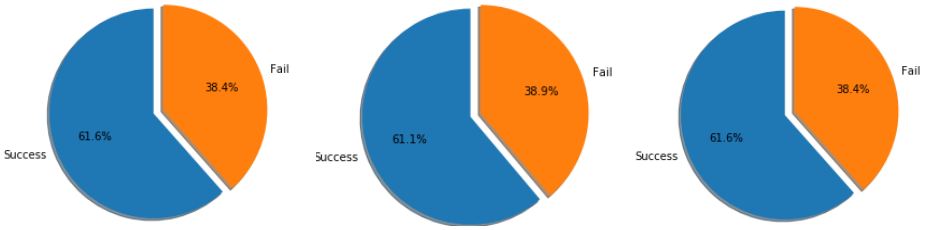
(ii)
```
In [35]: labels = clustering.labels_
         metrics.silhouette_score(matriks, labels, metric='euclidean')
Out[35]: 0.6144829830850764
```

(iii)
```
In [126]: labels = clustering.labels_
          metrics.silhouette_score(matriks, labels, metric='euclidean')
Out[126]: 0.6507060201725828
```

Figure 8. The result of silhouette coefficient (i) Single Linkage, (ii) Complete Linkage, (iii) Average Linkage

Figure 8 (i) shows that the cluster was formed in a single linkage with quite well of the accuracy of 0.6134522816367602 %. The distance between clusters and other clusters is good enough. In Figure 8 (ii) shows that the cluster has formed at complete linkage which is quite good with an accuracy of 0.6507060201725828. The distance between clusters and other clusters is also quite good. Figure 8 (iii) shows that the cluster has formed quite well with an accuracy of 0.6507060201725828 %. The distance between clusters and other clusters is good enough.

5.2 Clustering Evaluation with Term Frequency as a Feature

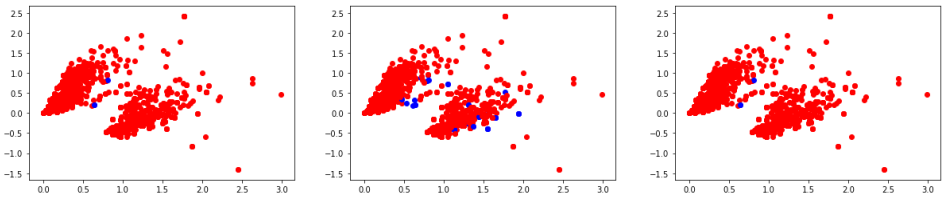


(i) (ii) (iii)

Figure 9. Circle diagram of proportion results (i) Single Linkage, (ii) Complete Linkage, (iii) Average Linkage on Term Frequency

Figure 9 shows the error ratio, which is the percentage of success on each tweet. The results of the label tweet from the clustering process are compared with the manual label tweet that has been done before. Figure 9 (i) shows that the error ratio of a single linkage shows satisfactory results with an accuracy level of 61.58637873754152%, in Figure 9 (ii) it appears that the error ratio of the complete linkage shows satisfactory results with an accuracy rate of 61, 12956810631229%, while in Figure 9 (iii) it appears that the error ratio of the average linkage shows satisfactory results with an accuracy rate of 61,58637873754152%.

Next, plotting the data distribution on the frequency term which showed in Figure 10.



(i) (ii) (iii)

Figure 10. Plot of data distribution on term frequency (i) Single Linkage, (ii) Complete Linkage, (iii) Average Linkage

Figure 10 shows that the red plot indicates a negative cluster and the blue plot indicates a positive cluster. In addition, it showed that the distribution of the data is not arranged as in trials with lexicon based as a feature. That is because many TF features are forced to be reduced to just two features so that two-dimensional visualization can be done. Feature reduction

uses the SVD feature decomposition truncate method. So, the plot still does not represent the distribution of tweet clusters.

Confusion matrix table is needed to calculate the result of the accuracy and the result of the classification.

Table 4a. The prediction of confusion matrix of single linkage

Label	-1 (predicted)	1 (predicted)
-1 (actual)	TN (1482)	FP (2)
1 (actual)	FN (923)	TP (1)

Table 4a shows that as many as 1482 tweets were correctly detected as negative in clustering while 2 tweets were incorrectly predicted by the single linkage AHC clustering algorithm. As for the positive label (1) there is 1 tweet that is correctly predicted by the AHC single linkage clustering algorithm. The confusion matrix clearly shows that the results of the AHC single linkage clustering algorithm are not good enough to be used as a benchmark because it is very good for negative tweets, but very bad for positive tweets.

Table 5a. Confusion matrix of single linkage

	Precission	Recall	F1-score	support
-1	0,62	1,00	0,76	1484
1	0,33	0,00	0,00	924
Accuracy			0,62	2408
Macro avg	0,47	0,50	0,38	2408
Weighted avg	0,51	0,62	0,47	2408

Table 5a explains that the recall value of the first label is 1, which indicates that almost all negative sentiment tweet labels are detected correctly by the clustering algorithm in accordance with the manual tweet label. However, far compared to the value of a recall on a positive label that is worth 0, only one tweet is detected correctly as a positive label.

Table 4b. The prediction of confusion matrix of complete linkage

Label	-1 (predicted)	1 (predicted)
-1 (actual)	TN (1459)	FP (25)
1 (actual)	FN (911)	TP (13)

Table 4b explains that as many as 1459 tweets were correctly detected negatively in clustering while 25 tweets were incorrectly predicted by the AHC complete linkage clustering algorithm. As for the positive label (1) there are 13 tweets that were correctly predicted by the AHC complete linkage clustering algorithm. In the confusion matrix, it can be seen more clearly that the results of the AHC clustering algorithm are very good for negative tweets, but very bad for positive tweets.

Table 5b. Confusion matrix of complete linkage

	Precission	Recall	F1-score	support
-1	0,62	0,98	0,76	1484
1	0,34	0,01	0,03	924
Accuracy			0,61	2408
Macro avg	0,48	0,50	0,39	2408
Weighted avg	0,51	0,61	0,48	2408

Table 5b explains that recall value from the first table is 0.98 which marks that almost all labels are correctly predicted to consist of negative tweet sentiment in accordance to the original tweet. However, the recall values on positive words were the opposite. With recall value only about 0,01, there was only one tweet that is correctly predicted by the algorithm.

Table 4c. The prediction of confusion matrix of average linkage

Label	-1 (predicted)	1 (predicted)
-1 (actual)	TN (1482)	FP (2)
1 (actual)	FN (923)	TP (1)

Table 4c explains that as many as 1482 tweets were correctly detected negatively in clustering while 2 tweets were incorrectly predicted by the AHC average linkage clustering algorithm. As for the positive label (1) there is 1 tweet that is correctly predicted by the AHC average linkage clustering algorithm. In the confusion matrix, it is clear that the results of the AHC average linkage clustering algorithm are not good enough to be used as a benchmark because it is very good for negative tweets but very bad for positive tweets.

Table 5c. Confusion matrix Average Linkage

	Precission	Recall	F1-score	support
-1	0,62	1,00	0,76	1484
1	0,33	0,00	0,00	924
Accuracy			0,62	2408
Macro avg	0,47	0,50	0,38	2408
Weighted avg	0,51	0,62	0,47	2408

Table 5c explains that rhe recall value of the first label is 1 which indicates that almost all negative sentiment tweet labels are detected correctly by the clustering algorithm in accordance with the manual tweet label. However, far compared to the value of a recall on a positive label that is worth 0, only one tweet is detected correctly as a positive label.

The performed silhouette coefficient algorithm is only known for the good or bad of the form of the cluster.

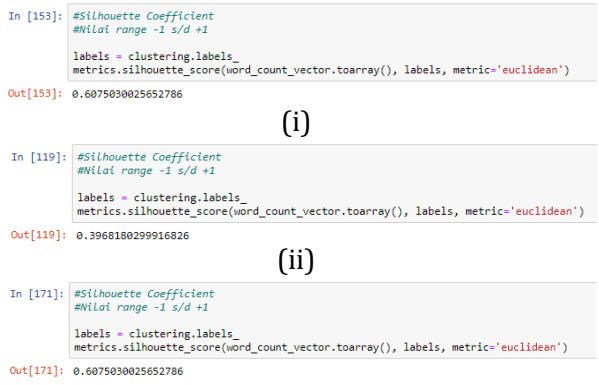


Figure 11. The result of silhouette coefficient (i) Single Linkage, (ii) Complete Linkage, (iii) Average Linkage

Figure 11 (i) shows that the cluster has formed on a single linkage quite well. The error ratio results show satisfactory results with an accuracy level of 61.12956810631229%. Figure 11 (ii) shows that the clusters that have been formed are less good than the single linkage algorithm. The distance between clusters with other clusters is still too far because in pl there is a blue cluster that is quite far apart from the other members. Figure 11 (iii) shows that the cluster has formed quite well. The distance between clusters and other clusters is good enough.

The examples of tweets that were used in predictions are:

- "UU bagus, hanya saja kurang relevan. jangan terlalu sadis mengomentari",
- "tidak ada yang buruk. padahal semua baik baik saja",
- "begitulah, bagus bagus saja sih",
- "bisa tidak sih pemerintah tidak jahat. UU itu tidak relevan, kurang profesional nih"

These sentences will be passed to the stages that were previously clustered through the stages of text mining, feature extraction, then predicted into clustering.

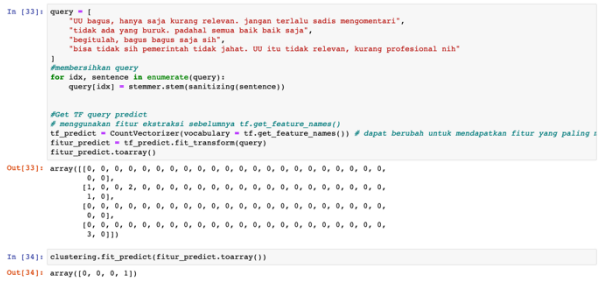
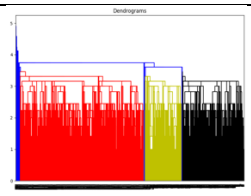
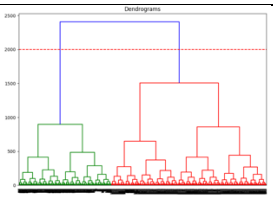
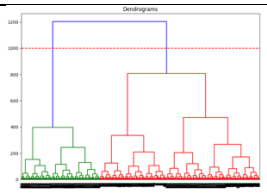
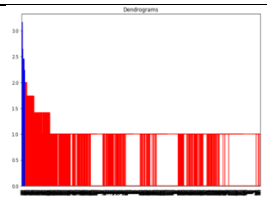
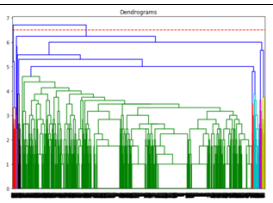
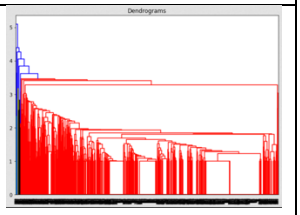


Figure 12. The cluster label prediction of clustering

Based on the prediction results in Figure 12, the sentiment label cannot be considered whether positive or negative because the cluster label indeed cannot be considered before using one of the evaluation methods such as error ratio. However, it can be seen that negative tweets are properly assembled with negative tweets.

Based on the dendrogram result, it can be seen that the distribution of data is unbalanced as indicated by the presence of clusters that have more members than other clusters. This is in accordance with the trials on the term frequency (TF) feature. The picture below is a visualization of dendrogram obtained by AHC clustering by using the word feature and term frequency.

Table 6. The comparison of dendrogram between lexicon based and term frequency

	Single linkage	Complete linkage	Average linkage
Lexicon based	 (a)	 (b)	 (c)
Term Frequency	 (d)	 (e)	 (f)

Based on Table 6 the visualization of data distribution of term frequency shown in figures (e) and (f), it produces different patterns because of the distribution of data is obtained in a form of a hierarchy of data groups based on the closeness that has been determined. Figures (b) and (c) show the results of clustering that were obtained and were visualized well because they formed two regular hierarchies.

After the visualization results are obtained, a comparison will be prepared between the three AHC methods with different features.

Table 7. The comparison of AHC method

Method	Accuracy	Silhouette
Lexicon based Single Linkage	61,58637873754152%	0,6134522816367602
Lexicon based Complete Linkage	61,088039867109636%	0,6507060201725828
Lexicon based Average Linkage	60,92192691029901%	0,650706201725828
TF Single Linkage	61,58637873754152%	0,6075030025652786
TF Complete Linkage	61,12956810631229%	0,3968180299916826
TF Average Linkage	61,58637873754152%	0,6075030025652786

According to Table 7, it can be seen that when using the TF extraction feature, it produces better and more consistent results compared to when using the lexicon based extraction feature. This happens because the features in TF extraction are more and more suitable from the topic used, known as the rejection of the draft law.

Table 8. The extraction comparison between lexicon based and term frequency

Comparison	Extraction of Lexicon based	Extraction of TF
Accuration	Good enough	Better and consistent
Process	Longer	Faster
Automatic	No	Yes

According to Table 8, it is known that the TF extraction process is faster due to the fact that there is no need to prepare opinion through the dictionary and give positive negative labels to the data one by one. Simply prepare a model that is already available from the scikit-learn library for TF feature extraction. Moreover, TF extraction is automatic without human intervention. The number of features used can also be adjusted according to need.

After the results are obtained, the visualized results will then be displayed on the web that contains the results of the lexicon based feature extraction, plotting, term frequency feature extraction, and dendrogram results on one page shown in Figure 14.

Figure 14. Web Display

5 CONCLUSION

Based on the research, we conclude that:

1. Whatever the type of extraction feature, the cluster results are similar to each other. This means that the feature could represent the tweets between lexicon based and term frequency which is related to the topic used for research.

2. Tweets which are being represented are only those that have negative labels, while positive tweets are not represented by both extraction features, both Lexicon Based and term frequency. Only negative tweets have succeeded in having a high recall on the evaluation of confusion matrix due to the fact that the features represent positive tweets were not as suitable as for the other one. Positive tweets have fewer numbers than negative tweets, and positive tweets have less representative words or words that really represent positive tweets.
3. Clusters are formed in the variation of two feature extractions and three AHC methods which are equally good. This means that the AHC algorithm is running well, but needs further development for feature extraction since it only represents the negative tweets.
4. The TF feature extraction which uses the scikit-learn library is better than the lexicon based method.

The suggestions for further research are expected to use other AHC methods and in addition to single linkage, average linkage, and complete linkage, and for subsequent sentiment analysis an accuracy calculation can be done with calculations other than confusion matrix. In addition, it is expected to use more data from Twitter so that it can include more vocabularies and approaches to the recent topic and it becomes more accurate.

REFERENCES

[1] Anggraini, N., & Suroyo, H, **Comparison of Sentiment Analysis against Digital Payment "T-cash and Go-pay" in Social Media Using Orange Data Mining**, *Journal of Information Systems and Informatics*, https://doi.org/10.33557/journalisi.v1i2.21, 2019.

[2] Luqyana, W. A., Cholissodin, I., & Perdana, R. S, **Analisis Sentimen Cyberbullying Pada Komentar Instagram dengan Metode Klasifikasi Support Vector Machine**, *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2018.

[3] M. Unnisa, A. Ameen, and S. Raziuddin, **Opinion Mining on Twitter Data using Unsupervised Learning Technique**, Int. J. Comput. Appl., vol. 148, no. 12, pp. 12–19, 2016.

[4] Wang, B., Liakata, M., Zubiaga, A., & Procter, R, **A Hierarchical Topic Modelling Approach For Tweet Clustering**, *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017.

[5] Prasetyo, E, **Data Mining: Konsep dan Aplikasi Menggunakan Matlab**, *Andi (Yogyakarta),* 2012.

[6] Y. Y. Yang dan F. Zhon, **Microblog Sentiment Analysis Algorithm Research and Implementation**, 14th International Symposium on Distributed Computing and Applications for Business Engineering and Science, pp. 288-291, 2015.

[7] G. A. Buntoro, **Sentiment Analysis Candidates of Indonesian Presiden 2014 with Five Class Attribute**, *International Journal of Computer Applications (0975 –8887)*, Volume 136 –No.2, 2016.
[8] Desai, R. D., **Sentiment Analysis of Twitter Data**, *Proceedings of the 2nd International Conference on Intelligent Computing and Control Systems, ICICCS 2018*, https://doi.org/10.1109/ICCONS.2018.8662942, 2019.
[9] Rustiana, D., & Rahayu, N, **Analisis Sentimen Pasar Otomotif Mobil**, *Jurnal SIMETRIS*, *8*(1), pp. 113–120, 2017.
[10] Buntoro, G. A, **Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter**, *Integer Journal Maret*, 2017.
[11] Nugroho, G. A. P., **Analisis Sentimen Data Twitter Menggunakan K-Means Clustering**, 2016.
[12] Bakshi, R. K., Kaur, N., Kaur, R., & Kaur, G, **Opinion Mining And Sentiment Analysis**, *Proceedings of the 10th INDIACom; 2016 3rd International Conference on Computing for Sustainable Global Development, INDIACom 2016*, https://doi.org/10.1561/1500000011, 2016.
[13] Cahyo Ryan Dwi, et al, **Deteksi dan Validasi Informasi Gempa Secara Real-Time Berbasis Social Sensor dengan *Twitter***, *JURNAL* TEKNIK POMITS Vol. 2, No. 1, 2014.
[14] Darma, I. M. B. S., **Penerapan Sentimen Analisis Acara Televisi Pada Twitter Menggunakan Support Vector Machine dan Algoritma Genetika sebagai Metode Seleksi Fitur**, 2017.
[15] Indraloka, D. S., & Santosa, B, **Penerapan Text Mining untuk Melakukan Clustering Data Tweet Shopee Indonesia**, *Jurnal Sains Dan Seni ITS*, https://doi.org/10.12962/j23373520.v6i2.24419, 2017.
[16] Ma, B., Yuan, H., & Wu, Y, **Exploring Performance Of Clustering Methods On Document Sentiment Analysis**, *Journal of Information Science*, 2017.
[17] E. W. Pamungkas and D. G. P. Putri, **An experimental study of lexicon-based sentiment analysis on Bahasa Indonesia**, *Proc. - 2016 6th Int. Annu. Eng. Semin. Ina. 2016*, pp. 28–31, 2017.
[18] Bouguettaya, A., Yu, Q., Liu, X., Zhou, X., & Song, A, **Efficient Agglomerative Hierarchical Clustering**, *Expert Systems with Applications*, https://doi.org/10.1016/j.eswa.2014.09.054, 2015.
[19] T. M. Fahrudin, I. Syarif, and A. R. Barakbah, **Data Mining Approach for Breast Cancer Patient Recovery**, *Emit. Int. J. Eng. Technol.,* vol. 5, no. 1, pp. 36–71, 2017.