# The Effect of Brill Tagger on The Classification Results of Sentiment Analysis Using Multinomial Naïve Bayes Algorithm

R. Astero Nandito [a,1], Abdiansah Abdiansah [b,2], Novi Yusliani [c,3]

[a, b, c] Computer Science Faculty, Universitas Sriwijaya Indonesia
[1] astero.nandito@gmail.com; [2] abdiansah.unsri@gmail.com; [3] noviyusliani@gmail.com;

ARTICLE INFO

**Article history**
Received
Revised
Published

**Keywords**
Sentiment Analysis
Parf of Speech Tagging
Brill Tagger
Multinomial Naïve Bayes
Weighting Part of Speech

ABSTRACT

Twitter is a good indicator for influence in research, the problem that arises in research in the field of sentiment analysis is the large number of factors such as the use of informal or colloquial language and other factors that can affect the results of sentiment classification. To improve the results of sentiment classification, an information extraction process can be carried out. One part of the information extraction feature is a part of speech tagging, which is the giving of word classes automatically. The results of part of speech tagging are used for weighting words based on part of speech. This study examines the effect of Part of Speech Tagging with the method Brill Tagger in sentiment analysis using the Naive Bayes Multinomial algorithm. Testing were carried out on 500 twitter tweet texts and obtained the results of the sentiment classification with implementing part of speech tagging precision by 73,2%, recall by 63,2%, f-measure by 67,6%, accuracy by 60,7% and without implementing part of speech tagging precision by 65,2%, recall by 60,6%, f-measure by 62,4% accuracy by 53,3%. From the results of the accuracy obtained, it shows that the application of part of speech tagging in sentiment analysis using the Multinomial Naïve Bayes algorithm has an effect with an increase in classification performance.

## 1. Introduction

Twitter is a social media that is often used by researchers as an object of research to conduct sentiment analysis. Twitter is also a good indicator for influencing research, but there are still not many applications and sentiment analysis methods developed for the Indonesian [8]. The problem that often arises in research in the field of sentiment analysis is the number of factors such as the use of informal or colloquial language and other factors that can affect the results of sentiment classification.

One thing that can be done to improve sentiment classification results is to extract information. One of the information extraction features is POS-Tagging. POS-Tagging is a general process that is usually required by natural language-based systems. POS-Tagging aims to determine the word class of each word in a sentence. Manual labeling or word classes will take a long time and require a lot of resources. Research implement POS-Tagging in a system that can automatically determine the word class of an Indonesian sentence, this system makes it easy to know the class of words that are in a sentence [5]. Furthermore, The research regarding the effect of POS-Tagging in identifying a document whether the document contains an opinion or not with the Hidden Markov Model method. Various approaches to the tagging process have been developed. Some of them use probabilistic, statistical, and rule-based calculations [3].

POS-Tagging research for the Indonesian language has been carried out, such as the research of regarding the effect of POS-Tagging on the classification of sentiment analysis. This study shows that the effect of POS-Tagging on the classification results is higher than without implementing POS-Tagging with an accuracy of 62.4%, while without POS-Tagging the accuracy obtained is 52.4% [2]. Comparison of the POS-Tagging method using the Hidden Markov Model and Brill Tagger in Indonesian-language text. This study shows that the accuracy obtained using the Brill Tagger method is better than using the HMM method using 30,000 sentences [1]. The accuracy obtained by Brill Tagger was 76.78% while the HMM was 62.69%. However, there has been no previous research that discusses POS-Tagging using the Brill Tagger method for classification cases in Indonesian language sentiment analysis. Brill Tagger is a rule-based method. The rule-based approach was chosen with the consideration of the high level of accuracy that was achieved using a rule-based approach in the tagging process [6].

In the process of sentiment analysis, a classification method is needed. One of the most commonly used classification methods is Naïve Bayes. Naïve Bayes has two event models, namely Multi-Variate Bernouli and Multinomial Naïve Bayes. Multinomial Naïve Bayes was chosen to classify sentiment analysis because its performance is better than Multi-Variate Bernouli in the case of a large number of words or terms. Multinomial Naïve Bayes can reduce errors in classification with an average value of 27% and even 50% of the trials against Multi-Variate Bernouli [4]. This study will examine the effect of Brill Tagger on the accuracy of sentiment classification using the Naïve Bayes Multinomial algorithm.

## 2. Related work

Research conducted by Amrullah et al., (2017) which compared some part of-speech tagging techniques for Bahasa Indonesia experiments using statistical approach (Unigram, Hidden Markov Models) and Brill's tagger. Used supervised POS Tagging approach requiring a large number of annotatedtraining corpuses to tag properly and some resource annotation corpus of Bahasa. Those corpuses were implemented with POS Tagging techniques, then compares and analyzes the results. Compared the accuracy and highlighted some advantages and disadvantages for every technique used. Unigram showeda higher accuracy compared to HMM and Brill tagger with 88,37% on a tagged corpus [1].

Muljono et al., (2017) this research shows that the Indonesian POS Tagger system could be appliedto assist learners of Indonesian language with independent learning. The system is designed to help assigntags to a sentence with 31 POS tags applied by the system. During the evaluation phase, objective and subjective measurements are used. The objective measurement is used to assess the accuracy level of POS tagging in the method used by the system. The objective measurement results show that HMM trigram and Morphology Analyzer (MA) methods applied here show higher accuracy than other methods being tested. In the subjective measurement, 3 evaluation criteria are tested to participants, namely user interface design, user-friendliness of the system, and properness of the system. The results of subjective evaluation from 24 participants indicate quite promising results [5].

Hamzah and Widyastuti (2015) through their research used a technique of two order HMM Based Part-of-Speech (POS) Tagger to identfy opinions from the text. This technique was also used to determine the target opinion in an opinion sentence. Both document subjectivity and target opinion was implemented using rule-based techniques. The results that subjectivity documents (opinion) can be identified with precision 0.95 and recall 0.92 using corpus II from student comments, and opinion targets can be identified with precision 0.91 and recall 0.89. The tests using different type of corpus shows difference performance for rule-set, whether for document subjectivity or target detection. It can be concluded that we always need to improve rule-set accoding the type of corpus. It can be argued because the addition with just one or two rule-set will effect to the performance of the test [3].

## 3. Research Method

Data obtained from Aprillia's (2018) research on the Palembang South Sumatra LRT. The data has been grouped into positive and negative sentiments. The amount of data obtained is 500 texts, 293 positive sentiments and 207 negative sentiments. List of Part of Speech weight values can be seen in

**Table 1.** Weight Part of Speech

| Tag | Part Of Speech | Weight |
|---|---|---|
| JJ | Adjective | 4 |
| VB | Verb | 3 |
| NN | Noun | 2 |
| RB | Adverb | 1 |

The research was carried out in stages according to the framework in Figure I for the software stages process with POS-Tagging.

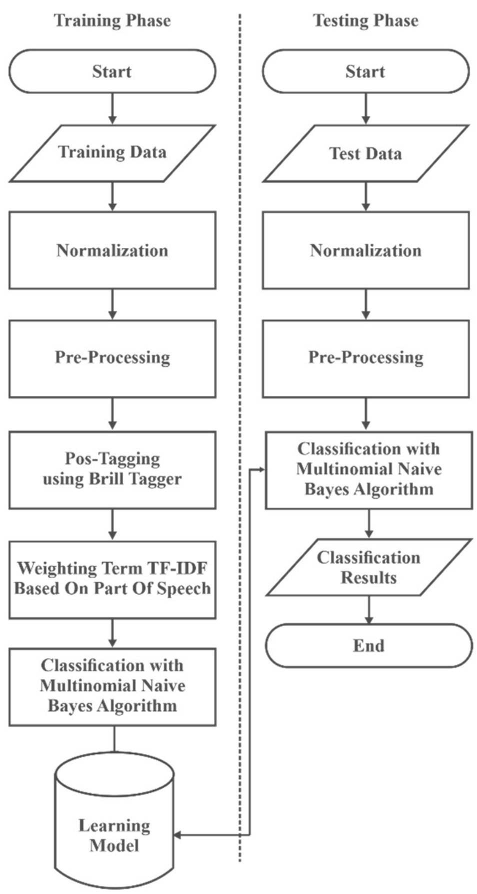


**Fig. 1.**Process Staged Software with POS-Tagging

The process of software stages without POS-Tagging can be seen in figure 2.

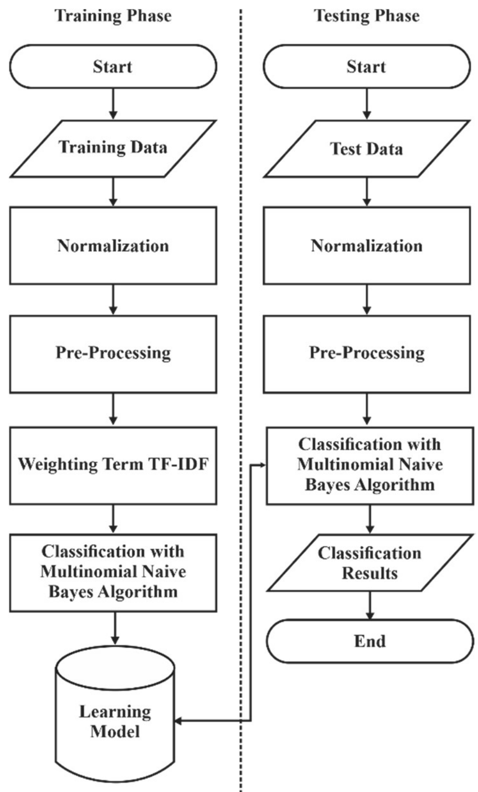


**Fig. 2.**Process Staged Software without POS-Tagging

A. Normalization

The normalization stage is to change non-standard words into standard words and remove unicode such as RT or Retweet, quotes marked with a “ ” symbol, hashtags, user names, urls, punctuation marks, symbols, and emoticons in the text.

B. Pre-Processing

The pre-processing steps are casefolding, tokenizing, stopword removal and stemming, between the tokenizing process and the stopword removal process, the POS-Tagging process will be applied. The case folding process is changing all letters in the data to lowercase. Furthermore, the data or text will bebroken down into word for word which is called the tokenizing process. Words that have been broken down and have Part of Speech go through a stop-word removal process to eliminate words that often appear in a collection of words and have no meaning. The last stage is stemming, which functions to change words that have affixes into their basic forms.

C. POS-Tagging

After the pre-processing process is carried out, the labeling process for each word is based on the tagset. This stage aims to get words with part of speech that will be used for word weighting. The method used is Brill Tagger. In this study, the categories of parts of speech which are indicators of sentiment for use in word weighting are nouns, verbs, adjectives, adverbs.

Brill Tagger is a rule-based algorithm and is often referred to as Transformation-Based Error-Driven Learning, commonly abbreviated as TEL. The term TEL means that Brill Tagger works based on rule transformation, where the transformation process learns from errors detected during the learning process [6]. Figure III shows the Brill Tagger work process.

*D.* Weighting TF-IDF Based On The *Part of Speech*

After each word has a word class or part of speech. The next stage is to give weight to each term in the text which is useful for the sentiment classification process. Weighting is done using a combined method of term frequency and inverse document frequency, known as TF-IDF. Each term is given a different weight based on the part of speech he said, then the frequency of terms that appear in a document and the length of a document are calculated by the weight of the part of speech. Weighting of parts of speech using the TF-IDF method is based on research [9], that is:

1) The TF (Term Frequency) value is calculated using equation:

$$tf_{pos}(t,d) = \frac{c(t,d) \times w_{pos}(t)}{\sum_i c(t_i,d) \times w_{pos(ti)}}$$

*t*: term
*d*: document
*c(t,d)*: number of terms in the document *d*
*c(ti,d)*: number of certain terms in the document *d*
$^{w}pos(t)$: weight term that has been determined in each category of part of speech
$^{w}pos(t)$: weight of a certain term that has been determined in each category of part of speech

2) To calculate the IDF (Inverse Document Frequency) value using an equation:

$$idf(t) = 1 + \log\left(\frac{n}{df(t)}\right)$$

3) TF-IDF weighting based on part of speech is calculated by equations:

$$w_{pos}(t,d) = tf_{pos}(t,d) \times idf(t)$$

E. Classification with Multinomial Naïve Bayes

The classification process stage with Multinomial Naïve Bayes is the stage for classifying previously processed sentiments into 2 classifications (positive and negative). Text data that has passedthe normalization process, pre-processing, POS-tagging, weighting of TF-IDF words based on the part of speech, and finally the grouping process is carried out using the Naïve Bayes Multinomial classifier. Multinomial text classification Naïve Bayes [7], that is:

*1)* Calculating the prior probability *P(c)*

$$P(c) = \frac{\sum_{j=1}^{n} \delta(c_{j,}c) + 1}{n + l}$$

*n*: number of training document
*l*: number of classes
$\delta(cj,c)$: number of class *c* in the training document

2) Calculating the conditional probability $P(wi|c)$

$$P(w_i|c) = \frac{\sum_{j=1}^{n} f_{ji}\delta\left(c_j, c\right) + 1}{\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ji}\delta(c_{j,}c) + 1}$$

$\sum_{j=1}^{n} f_{ji}\delta(c_{j,}c)$: the number of occurrences of *a* particular word in *a* class training document *c*

$\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ji}\ \delta(c_{j,}c)$: the total number of words in a training document from class *c*.

3) Classifying the *d* test document indicated by word vectors <$w1$, $w2$, ..., $wm$> using equations:

$$c(d) = \arg max\ [\log P(c) + \sum_{i=1}^{m} f_i \log P(w_i|c)]$$

P($wi|c$): probability of a document's conditionals on the class
*P(c)*: prior probability of a document in class *c*
$wi$ (I = 1,2,…, m): words that appear in the
document $dfi$ (I = 1,2,…, m): $wi$ frequency in
the document *d*

4) Determine the document class by selecting the highest probability value.

## 4. Result and Discussion

Testing on the software was carried out using 500 texts of pre-prepared test data divided into two classes, namely positive and negative classes, 293 positive class data and 207 negative class data. The testing process is carried out in accordance with the research framework, in Figure I the testing phase for the Multinomial Naïve Bayes classification with POS-Tagging and Figure II the testing phase for the classification without POS-Tagging. The results of both classifications will then be compared and summarized in the confusion matrix table. Next, the recall, precision, f-measure and accuracy values will be calculated based on the confusion matrix table that has been created. Table Confusion Matrix Classification Test Results can be seen in Table II.

**Table 2.** Confusion Matrix Classification Test Results

| Percobaan ke- | MNB | | | | MNB + *POS-Tagging* | | | |
|---|---|---|---|---|---|---|---|---|
| | Prediksi | | | | Prediksi | | | |
| | Positif | | Negatif | | Positif | | Negatif | |
| | TP | FP | TN | FN | TP | FP | TN | FN |
| *Fold* 1 | 18 | 13 | 5 | 13 | 22 | 10 | 8 | 9 |
| *Fold* 2 | 23 | 9 | 9 | 7 | 16 | 7 | 11 | 14 |
| *Fold* 3 | 19 | 12 | 3 | 15 | 18 | 5 | 12 | 12 |
| *Fold* 4 | 23 | 16 | 2 | 9 | 24 | 8 | 8 | 9 |
| *Fold* 5 | 18 | 7 | 10 | 15 | 20 | 5 | 13 | 11 |
| *Fold* 6 | 19 | 8 | 5 | 16 | 19 | 6 | 12 | 12 |
| *Fold* 7 | 22 | 13 | 4 | 10 | 19 | 7 | 10 | 13 |
| *Fold* 8 | 22 | 8 | 9 | 11 | 19 | 15 | 2 | 14 |
| *Fold* 9 | 22 | 12 | 5 | 11 | 22 | 7 | 10 | 9 |
| *Fold* 10 | 19 | 15 | 2 | 13 | 19 | 5 | 12 | 12 |

Based on the confusion matrix in Table I, a classification test result table can be created as an evaluation of the Multinomial Naïve Bayes classification method without POS-Tagging and classificationusing Multinomial Naïve Bayes with POS-Tagging in classifying, evaluation is done by calculating the recall value, precision , f-measure and accuracy. Each evaluation value can be seen in Table III.

**Table 3.** Classification Test Results

| Experiment To- | MNB | | | | MNB + *POS-Tagging* | | | |
|---|---|---|---|---|---|---|---|---|
| | *Precision* | *Recall* | *F-Measure* | *Accuracy* | *Precision* | *Recall* | *F-Measure* | *Accuracy* |
| *Fold* 1 | 0.581 | 0.581 | 0.581 | 0.469 | 0.688 | 0.710 | 0.698 | 0.612 |
| *Fold* 2 | 0.719 | 0.767 | 0.742 | 0.667 | 0.696 | 0.533 | 0.604 | 0.562 |
| *Fold* 3 | 0.612 | 0.559 | 0.584 | 0.449 | 0.783 | 0.600 | 0.679 | 0.653 |
| *Fold* 4 | 0.590 | 0.719 | 0.648 | 0.500 | 0.750 | 0.727 | 0.739 | 0.627 |
| *Fold* 5 | 0.720 | 0.545 | 0.621 | 0.560 | 0.800 | 0.645 | 0.714 | 0.673 |
| *Fold* 6 | 0.655 | 0.543 | 0.594 | 0.500 | 0.760 | 0.613 | 0.679 | 0.632 |
| *Fold* 7 | 0.704 | 0.667 | 0.698 | 0.620 | 0.731 | 0.594 | 0.655 | 0.592 |
| *Fold* 8 | 0.733 | 0.423 | 0.536 | 0.595 | 0.559 | 0.576 | 0.567 | 0.420 |

| Experime nt To- | MNB | | | | MNB + *POS-Tagging* | | | |
|---|---|---|---|---|---|---|---|---|
| | *Precision* | *Recall* | *Precision* | *Recall* | *Precision* | *Recall* | *Precision* | *Recall* |
| *Fold* 9 | 0.647 | 0.667 | 0.657 | 0.540 | 0.759 | 0.710 | 0.734 | 0.653 |
| *Fold* 10 | 0.559 | 0.593 | 0.575 | 0.428 | 0.792 | 0.613 | 0.691 | 0.646 |
| Average | 0.652 | 0.606 | 0.624 | 0.533 | 0.732 | 0.632 | 0.676 | 0.607 |

Based on the test results obtained, it can be seen the comparison of the classification results using Multinomial Naïve Bayes without POS-Tagging and the results of the Multinomial Naïve Bayes classification with POS-Tagging in the form of a bar graph which is depicted in Figure III.

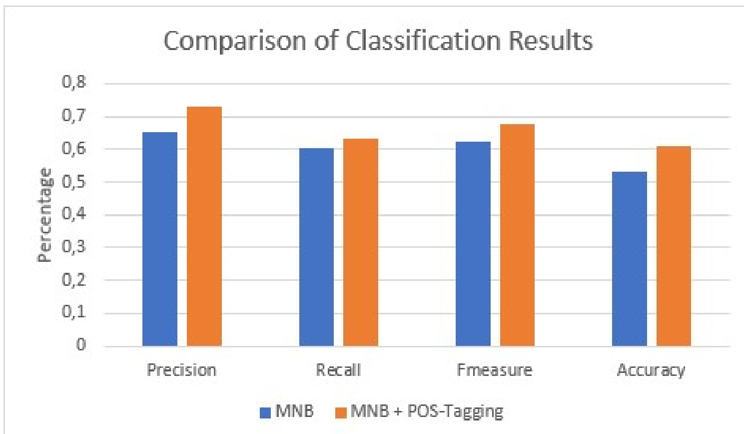


**Fig. 3.**Comparison of Classification Results

The classification results using the Multinomial Naïve Bayes with POS-Tagging get better results than the classification without POS-Tagging, this can be seen from the precision, recall, f-measure and accuracy values obtained, namely for classification with part of speech tagging the precision obtained 73.2%, recall of 63.2%, f-measure of 67.6%, accuracy of 60.7% and for classifications without part of speech tagging the precision obtained was 65.2%, recall of 60.6 %, f-measure of 62.4%, accuracy of 53.3%.

The Naive Bayes Multinomial classification method with POS-Tagging has better results than the Naive Bayes Mulitnomial classification without POS-Tagging because the Naive Bayes Multinomial classification method with POS-Tagging has a stopword process that only takes words that have the JJ word class (words characteristic), RB (adverb), VB (verb) and NN (noun) where each class of words has a part of speech weight, this weight will affect the value of the prior probability and conditional probability. It is possible that if there is test data that does not have the word class JJ, RB, VB or NN so that the data is empty, the blank test data is then deleted. Thus, the test data to be tested is automatically reduced, with a reduction in the number of test data, it can reduce the number of errors in classification.

Based on Figure III, it can be seen that the precision value in the Multinomial Naïve Bayes classification method with POS-Tagging has better results than the Naïve Bayes Multinomial classification method without POS-Tagging, this is because the number of test data documents that willbe used in testing is less in the classification. using Multinomial Naïve Bayes with POS-Tagging instead of the Multinomial Naïve Bayes classification without POS-Tagging which resulted in reduced classification errors. For example, there are 50 test data, from the 50 data, the confusion matrix results obtained in the classification without POS-Tagging are TP = 20, FP = 7, TN = 13 and FN = 10, the total confusion matrix is 50 because there are 50 data If tested, the confusion matrix results obtained in the POS-Tagging classification are TP = 21, FP = 4, TN = 13 and FN = 10, the total confusion matrix is 48 because there is data that is deleted based on the explanation in the previous paragraph. From the two results of confusion matrix, to calculate the precision value is to divide the TP value by the sum of TP and FP values, in the classification without POS-Tagging the sum of TP and FP values is greater than the classification with POS-Tagging. This causes the precision value in the POS-Tagging classification to be better or higher than the classification without POS-Tagging.

The recall value in the Naïve Bayes Multinomial classification method with POS-Tagging has better results than the Naïve Bayes Multinomial classification method without POS-Tagging, which is thesame as the explanation in the previous paragraph and the example used is the same as the example forthe precision value, from both confusion results. The matrix to calculate the recall value is to divide theTP value by the sum of TP and FN values, in the classification without POS-Tagging the sum of TP and FN values is greater than the classification with POS-Tagging. This is why the recall value in the POS- Tagging classification is better or higher than the classification without POS-Tagging.

The f-measure value in the Multinomial Naïve Bayes classification method with POS-Tagging has better results than the Multinomial Naïve Bayes classification method without POS-Tagging, this is because to calculate the f-measure is to look at the values of precision and recall obtained, at classification with POS-Tagging, the value of precision and recall obtained is greater than the classification without POS-Tagging. This is why the f-measure value of the POS-Tagging classification is better or higher than the classification without POS-Tagging.

The accuracy value in the Naïve Bayes Multinomial classification method with POS-Tagging has better results than the Naïve Bayes Multinomial classification method without POS-Tagging, is the same as the explanation for recall and precision values and the example used is the same as the example for the precision and recall values. of the two results of confusion matrix to calculate the accuracy value is to divide the sum of the TP and TN values with the sum of the confusion matrix values, in a classification without POS-Tagging the sum of the confusion matrix values is greater than the classification with POS- Tagging. This is what causes the accuracy value of the POS-Tagging classification to be better or higher than the classification without POS-Tagging.

The percentage of accuracy results obtained in the classification using the Multinomial Naïve Bayes algorithm without POS-Tagging and with POS-Tagging is still relatively low, namely 53.3% for classification accuracy results without POS-Tagging and 60.7% accuracy results with POS-Tagging. In research, the accuracy results obtained are still relatively low, namely 50.4% for classification accuracy results without POS-Tagging and 60.2% accuracy results with POS-Tagging. This is because the vocabulary size is an indicator that affects the performance of the Multinomial Naïve Bayes, the distribution of each category in the training data also affects the performance of the Multinomial Naïve Bayes because the data is randomized before being divided using the k-fold cross validation technique. The distribution of terms that have the word class "NN" in the positive and negative class categories does not describe the polarity that tends to one of the categories, including: routes, cities, communities, and transportation. Therefore, the accuracy results obtained are still relatively small compared to the accuracy results on other data [2].

So it can be concluded that the classification using Mulitnomial Naive Bayes with POS-Tagging using the Brill Tagger method can affect the classification results..

## 5. Conclusion

The effect of Brill Tagger on the classification results of sentiment analysis using the Naïve Bayes Multinomial algorithm is the classification results by applying POS-Tagging, the precision obtained is 73.2%, recall is 63.2%, f-measure is 67.6%, accuracy is 60, 7% and without applying POS-Tagging the precision obtained is 65.2%, recall is 60.6%, f-measure is 62.4%, accuracy is 53.3%. So, with the POS-Tagging process using the Brill Tagger method it can provide better results than without the POS-Tagging process. In future research it is hoped that:

1. Comparing the Naïve Bayes Multinomial classification by implementing the POS-Tagging process using the Brill Tagger method and without carrying out the POS-Tagging process for different datasets.
2. Using other POS-Tagging methods.
3. Using other classification methods.

# References

[1] Amrullah, A. Z., Hartanto, R., & Mustika, I. W. (2017). A comparison of different part-of-speech tagging technique for text in Bahasa Indonesia. Proceedings - 2017 7th International Annual Engineering Seminar, InAES 2017.

[2] Aprillia (2018). “Pengaruh Parf Of Speech Tagging dalam Analisis Sentimen menggunakan Algoritma Multinomial Naïve Bayes.” Skripsi Jurusan Teknik Informatika, Fak. Ilmu Komputer, UNSRI – Tidak diterbitkan.

[3] Hamzah, A., & Widyastuti, N. (2015). Document subjectivity and target detection in opinion mining using HMM POS-Tagger. Proceedings of 2015 International Conference on Information and Communication Technology and Systems, ICTS 2015, September 2015, 83–87..

[4] McCallum, A., & Nigam, K. (1998). A Comparison of Event Models for Naive Bayes Text Classification. 58(8), 582–591.

[5] Muljono, Afini, U., Supriyanto, C., & Nugroho, R. A. (2017). The development of Indonesian POS- Tagging system for computer-aided independent language learning. International Journal of Emerging Technologies in Learning, 12(11), 138–150.6.

[6] Setyaningsih, E. R. (2017). Part of Speech Tagger Untuk Bahasa Indonesia Dengan Menggunakan Modifikasi Brill. Dinamika Teknologi, 9(1), 37–42.

[7] Song, J., Kim, K. T., Lee, B., Kim, S., & Youn, H. Y. (2017). A novel classification approach based on Naïve Bayes for Twitter sentiment analysis. KSII Transactions on Internet and Information Systems, 11(6), 2996–3011.

[8] Winarko, E., & Putrianti, N. D. (2014). Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine. IJCCS (Indonesian Journal of Computing and Cybernetics Systems), 8(1), 91–100.

[9] Xu, R. (2014). POS weighted TF-IDF algorithm and its application for an MOOC search engine. ICALIP 2014-2014 International Conference on Audio, Language and Image Processing, Proceedings, 868–873.