Implementation of negation handling techniques using modified syntactic rule in Indonesian sentiment analysis

T G Prahasiwi, R Kusumaningrum
Department of Informatics, Universitas Diponegoro, Indonesia
Jl. Prof. Soedharto, SH, Tembalang, Semarang 50275, Indonesia
Email: retno@live.undip.ac.id

Abstract. Hotels are one of the supporting facilities for tourism in Indonesia. The development of hotels is an essential part of the tourism sector. Hotel development can be performed by knowing public opinion obtained from user reviews. However, direct comment analysis, by reading reviews one by one, becomes difficult due to the vast numbers involved. This problem can be overcome by implementing sentiment analysis. In practice, sentiment analysis has critical issues that can affect performance, i.e., several words have different meanings in different sentence contexts, and the existence of the negation words can invert the meaning of an entire text. Therefore, this research proposes the implementation of negation handling techniques using modified syntactic rules in Naive Bayes based sentiment analysis. The proposed modified syntactic rules of negation handling techniques have been shown to improve sentiment analysis performance by 3.3%; compared to sentiment analysis without negation handling.

1. Introduction
Hotels are one of the supporting facilities for tourism in Indonesia. Hotel development has become an important factor for increasing the tourism sector. Increased tourism creates opportunities for companies to build hotel ticket reservations businesses; one of which is Traveloka. Traveloka users provide information and reviews about hotels that contain comments. Comments are important because they can be used for user decision-making in booking a hotel from positive comments and for hotel development from negative comments. With the growing process of hotel development, the quality of services will increase. So this will support Indonesian tourism by increasing one of the tourism facilities. Also, there are positive comments that show that the hotel is of high quality and this will attract interest from tourists both from abroad and from Indonesia itself. Travelers will feel comfortable and interested in doing tours if the existing facilities are guaranteed quality.

However, comment analysis is sometimes difficult due to a large number of comments involved. It creates the need for technology to manage comment data information contents automatically. This technology is known as sentiment analysis. Research on sentiment analysis has been done extensively; both in English [1-4] and in Indonesian [5-10]. However, the application of sentiment analysis still has many problems; one of which is the inappropriate classification of comments if a word has different meanings in different sentence contexts. It is happening because word negation appears in a sentence and reverses the polarity of an influencing word. Research of negation handling has been done by [1][11][2][7]. Research by [18] discusses to determine the scope of negation. The scope of negation is detected by various methods, such as a static window, punctuation mark [1-2] and Part of Speech (POS) [11].

Negative handling research in the Indonesian language is also done by various methods, namely static window and punctuation mark [12][5][7]. Research by [13] handles negation with negation
lexicon-based. Also, research by [14] takes care of negation using the adaptation rule of syntactic negation, which determines the scope of negation using POS. However, in that double study negation is not handled, the existing rule is incomplete, and there is no rule for non-standard informal negation. Therefore, we propose a method called Modified Syntactic Rule, which handles double negation and adds a rule for some negation words from the rule of negation handling used in previous research [14]. This paper is organized as follows. Section 2 describes negation handling in general terms and its implementation in Indonesian sentiment analysis. Section 3 describes the experimental setup, experimental scenario, proposed method approach, and experimental results. Conclusion and future works are described in Section 4.

2. Negation

2.1. Negation in General
Negation is the process or construction of denial, which expresses contention of the meaning of a sentence. In English, the word negation has many forms. All forms of negation words are categorized into 3 classes; namely syntactic, diminisher and morphological negation. Morphological negation follows a specific structure where the base word is changed by a prefix (such as in-, non-, un-, etc.) or suffix (like -less). Syntactic negation is a negation signal used to reverse a word or word order in full. Examples of syntactic negations are not, no, instead, none, never, do not, does, cannot, and others. Meanwhile, diminisher negation is a reducer that reduces the polarity of an affected word but does not reverse its polarity [1].

Negation can change the state where it affects the text polarity. It causes negation to be considered. Negative handling is one method of improving the accuracy of the classifier [11]. The main scope of negation handling is to determine the order of words in sentences influenced by the words of negation [15]. Sentences that are influenced by the words of negation are also called the scope of the negation. The scope of the negation can be determined by several methods, including static window, punctuation mark, and POS. In the static window, some words that remain after negation are considered as the scope of the negation, and the polarities of only these words are reversed. Punctuation mark handles the negation by adding artificial words, i.e., if word x is preceded by the negation word, the new NOT_x feature is created. It is done by repeating every word until the next punctuation. Meanwhile, the POS method is applied by creating a new feature after the word negation is done, until a certain word or phrase is encountered, and runs to the end.

2.2. Negation in Indonesian
In Indonesian, experts have different opinions in determining the constituent negation. Negation constituents include “tidak”, “tak”, “bukan”, “jangan”, “tanpa”, and “belum” [16]. A negation constituent is also called denial or a negation word.

Indonesian word negation generally comes in the form of syntactic negation; which reverses the word in full. Previous research on negation was performed by [14][12][5][7] using the static window, punctuation mark, and POS methods. However, static window and punctuation mark methods caused an increase in the number of unnecessary features, because the words added to the new feature were not restricted. Words are added without considering whether they are affected by the negation or not [17]. Therefore, in this study, the determination of the scope of negation is done using POS.

POS is the process of tagging word class labels. Part of Speech (POS) of a word is a class of words which are categories of words based on the grammatical and morphological structure of a particular word. POS tagging is the main process for identifying words that support NLP applications. Negation handling, using POS, uses a rule that considers the structure of the negation word. Rule handling negation for the Indonesian language was performed by [14]. The rule of negation handling is shown in Table 1.
### Table 1. Synthetic Rule of Negation Handling

| Negation word | Tag       |
|---------------|-----------|
| “tidak”       | JJ – Adjective |
|               | VB – Verb   |
|               | NN – Noun   |
|               | CD – Cardinal Number |
|               | RB – Adverb |
|               | MD – Modal  |
| “bukan”       | JJ – Adjective |
|               | CD – Cardinal Number |
| “belum”       | JJ – Adjective |
|               | VB – Verb   |
|               | RB – Adverb |
|               | MD – Modal  |
| “jangan”      | JJ – Adjective |
|               | VB – Verb   |
|               | RB – Adverb |
|               | MD – Modal  |
| “tetapi, tapi”| JJ – Adjective |
|               | NN – Noun   |
|               | RB – Adverb |

### 3. Experiment and Result

#### 3.1. Experiment Setup
The research data used, derived from 1000 hotel reviews, contained an equal division of 500 positive data and 500 negative data. In this study, we apply 10-fold cross-validation with 10% for data testing and 90% for test data. Each fold is searched for accuracy values to determine the performance of the model, which will then be searched for the average value of accuracy. For the classifier used in this study, we implemented a Multinomial Naive Bayes with term frequency method for its vectorization.

#### 3.2. Experiment Scenarios
The aim of this research scenario is to identify the influence of negation handling using a Modified Syntactic rule. The scenario used is divided into 3 scenarios. Scenario 1 aims to reveal the performance of sentiment analysis without handling negation using a Modified Syntactic rule. The research data will be processed without handling negation, namely by applying preprocessing sentiment analysis in general. Scenario 2 aims to reveal the performance of sentiment analysis by applying negation handling using a Modified Syntactic rule. For scenario 2, the research data will be processed by applying negation handling with the Modified Synthetic rule where the preprocessing usually carried out will be added to the negation handling process that applies the rule. To compare the first 2 scenarios, scenario 3 is used. In scenario 3, scenarios 1 and 2 are searched per fold performance to get the average performance of the model. So, we can compare the results of both to find out how effective the implementation of negation handling with the Modified Syntactic rule.

#### 3.3. Proposed Method Approach
Negative handling is one of the challenges of sentiment analysis; as studied by many researchers. In the Indonesian language, the implementation of the handling of negations has been done in various ways. The implementation of syntactic negation handling used the structure of the word negation rule, in the Indonesian language, with the POS method. Research on sentiment analysis that applies negation handling with the POS method in Indonesian has been done. Previous research [14] applied rules of negation handling that was still incomplete and also had not implemented double negation handling. Double negation occurs when there are negation words that appear in sequence. If not handled, this will be wrong in determining the polarity of a word. So, to improve the optimality of negation handling, this paper proposed a Modified Synthetic rule that adds new rules to the existing rules of negation handling that have been applied before. Also, this Modified Syntactic Rule also deals with double negation cases. Handling of the negation with the Modified Synthetic rule can be seen in Figure 1.

![Diagram](image1.png)

**Figure 1.** The modified syntactic rule is handling the flow.

Implementation of the Modified Syntactic Rule is performed using a syntactic rule with a modification rule (as shown in Table 2).

| Negation Word | Tag          |
|---------------|--------------|
| “tak”, “ga”, “engga”, “enggak” | JJ – Adjective |
|               | VB – Verb    |
|               | NN – Noun    |
|               | CD – Cardinal Number |
|               | RB – Adverb  |
|               | MD – Modal   |
| “bukan”       | VB – Verb    |
| “tanpa”       | NN – Noun    |
|               | MD – Modal   |

3.4. Experimental Result and Analysis
Scenario 1 is performed by managing comment data using pre-processing and the steps shown in Figure 2.

![Diagram](image2.png)

**Figure 2.** The pre-processing steps of Scenario 1
Scenario 1 is performed without using the Modified Syntactic Rule using the results achieved (as shown in Figure 2). This scenario shows an average accuracy of 84.2% with the performance of the model being considered good enough to perform classification. However, in this scenario, there is still a classification error; as there is a negation word in the comment. Furthermore, prediction errors occurred, because the test data contained many classy words of reverse sentiment; thus minimizing the probability of a true class of sentiments. This is one of the gaps of the Naive Bayes Classifier.

![Accuracy of Scenario 1](image)

*Figure 3. The results of Scenario 1.*

Scenario 2 is performed by managing comment data using pre-processing using the steps shown in Figure 4.

![Pre-processing steps of Scenario 2](image)

*Figure 4. The pre-processing steps of Scenario 2*

Scenario 2 is performed using the Modified Syntactic Rule and the results achieved (as shown in Figure 5). This scenario shows an average accuracy of 87.5%, which indicates that the negation handling is suitable to be applied with the Naive Bayes Classifier. However, in this scenario, there is still a misclassification; as there is a negation word in diminished form, such as the word "less." Furthermore, the data handling negations are less precise because words are not listed in the POS dictionary, and therefore are not affected by negation.

![Accuracy of Scenario 2](image)

*Figure 5. The results of Scenario 2.*
The comparison of Scenarios 1 and 2 yielded an average difference of 3.3% accuracy (as shown in Figure 6). In Scenario 2, the accuracy value increases due to misclassification in already handled Scenario 1. The classification error occurs because a negation word appears in the sentence. The modified Syntactic Rule applied has been shown to increase the performance of sentiment analysis with the Naive Bayes Classifier. The Modified Syntactic Rule is applied using a POS negation scope determination method, which has both advantages and disadvantages. This method has the advantage of reducing features that are not important for classification. However, it also has disadvantages of high dependence on the rule and the method or corpus POS tagging.

![Comparison of Skenario 1 and Skenario 2](image)

**Figure 6.** Comparison between the result of scenario 1 and scenario 2

4. Conclusion and Future Work

We applied negation handling using a Modified Syntactic Rule on hotel review data with a Naive Bayes Classifier. After completing the research, the effect of handling negation using the Modified Syntactic Rule showed good results by increasing the average accuracy value by 3.3%. It proves that the handling of negation using Modified Syntactic Rule can improve the performance of sentiment analysis with a Naive Bayes Classifier. With this Modified Syntactic rule, it can optimize the performance of sentiment analysis to identify the class of each comment correctly. So that it can be used to help make decisions in choosing hotels for hotel customers, and to become a reference in the development of hotels in the future.

For future work, research in this area can be developed by handling negations with different negation forms, such as diminisher and morphological, in Indonesian document reviews.

References

[1] Farooq U, Mansoor H, gaillard A, Ouzrout Y, and Qadir M 2016 *Journal Of Computer* 12
[2] Narayanan V, Arora I, and Bhatia A, 2013 *Intelligent Data Engineering and Automated Learning (IDEAL)* 194
[3] Tripathy A, Agrawal A and Rath S 2016 *Expert Systems with Applications* 57 117
[4] Dinakaramani A, Rashel F, Luthfi A, and Manurung R, 2014 *International Conference on Asian Language Processing*
[5] Franky, Bojar O, and Veselovská K 2015 *The Prague Bulletin of Mathematical Linguistics* 21
[6] Naradhipa A and Purwarianti A 2012 *International Conference on Cloud Computing and Social Networking*
[7] Wisudawati J, Adiwijaya, and Faraby S 2017 *e-Proceeding of Engineering* 4 2987
[8] Putri I and Kusumaningrum R 2017 *IOP: Journal of Physics* 801
[9] Bashri M and Kusumaningrum R 2017 *International Conference on Information and Communication Technology (ICoIC7)*, 2017.

[10] Usop E, Isnanto R, and Kusumaningrum R 2017 *International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, 2017.

[11] Garg S and Meher R, 2015 Naive Bayes model with improved negation handling and n-gram method for Sentiment classification *National Institute of Technology Roukel*

[12] Franky and Ruli Manurung, 2008 Machine Learning-based Sentiment Analysis of Automatic Indonesian Translations of English Movie Reviews

[13] Buntoro G, Adji T, and Permanasari A 2014 *Conference on Information Technology and Electrical Engineering(CITEE)* (Gadjah Mada University: Yogyakarta)

[14] Amalia R, Bijaksana M, and Darmantoro D 2018 JOP: Conference Series 971 *International Conference on Data and Information Science* (Telkom University: Indonesia)

[15] Sharif W, Samsudin N, Deris M, and Naseem R 2016 *Sixth International Conference on Innovative Computing Technology*

[16] Sudaryono, *Negasi dalam bahasa Indonesia : Suatu Tinjauan Sintaksis Semantik*. Jakarta: Depdiknas, 1992.

[17] Wiegand M, Balahur A, Roth B, Klakow D, and Montoyo A 2010 *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing* 60

[18] Hogenboom A, Iterson P, Heerschop B, Frasincar F, and Kaymak U 2011 *IEEE International Conference on Systems, Man, and Cybernetics*. 