An Experimental Study of Supervised Sentiment Analysis Using Gaussian Naïve Bayes

Unggul Widodo Wijayanto
Informatics Department, Faculty of Technology Information and Communication
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
unggul15@mhs.if.its.ac.id

Riyanarto Sarno
Informatics Department, Faculty of Technology Information and Communication
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
riyanarto@if.its.ac.id

***Abstract*— In this millennial generation everyone using technology at higher rates than people from other generations. It means the millennial generation is aware of evolving technology. Many companies are taking chances by receiving customer reviews through applications. In this study, we use customer reviews from Yelp (foods), IMDb (movies) and Amazon (products). The reviews received by the company are numerous. Product management does not have much time to read customer reviews one by one. So, to speed up the reading of customer reviews we were using sentiment analysis. There are many methods that used in sentiment analysis such as supervised sentiment analysis. We used TF-IDF to convert word to features implements the supervised method. Performance of the supervised method depends on the data training quality. So, to improve the accuracy of the results by improving data training quality. The methods used to improve the data training quality in this paper using CHI2 Features Selection and Stopwords. In this study, we use K-folds Cross-validation to get valid results. This study proves the use of Context-based Stopwords can improve the results. Context-based Stopwords enrich the number of Stopwords that removing bias features.**

Keywords—supervised, sentiment analysis, data training, text mining, cross-validation, features selection, stopwords.

I. INTRODUCTION

In this era, people can do anything on the internet. The Internet World Stats state that there are 4.1 billion internet users in the world on December 31, 2017 [1]. Based on internet world stats on December 31, 2017 internet user growth 1,052% for 18 years [1]. Besides using the internet to search for information they can also review. Internet users often review food, movies, and products. There are many websites that provide user reviews, like Yelp.com, IMDb.com, and Amazon.com.

Product reviews can help a company's decision-making. For the example is decision making in a cinema company. They can predict a good movie that will be of interest to many people using reviews from IMDb.com. When good reviews received massively in the first week. Cinemas can open more theaters. If the opposite happens, the cinema may reduce the duration of the theater. We can classify good and bad review by using sentiment analysis.

Some applications from Natural Language Processing (NLP) iFs opinion mining and sentiment analysis, text analysis to find out and extract the meaning of the information obtained from the review document, blogs or any source materials. And positive or negative can be the pole of a sentence [2]. Example, “This is definitely a cult classic well worth viewing and sharing with others.” it means the audience has a positive sentiment to the movie. Another example, “Otherwise, don't even waste your time on this.” It means the audience has a negative sentiment when watching the movie.

Sentiment analysis or usually called opinion mining had been studied many times. With the development of connectivity technology that provides many ways to interpret and process the user opinion. Some use machine learning methods such as Naïve Bayes, Maximum Entropy and Support Vector Machines [3]. In this study, we use Gaussian Naive Bayes.

This experimental study used supervised methods. Supervised methods have a good accuracy, but it depends on the data train that gives in the preprocessing. The methods used to improve the quality of data training in this paper using CHI2 features selection and Stopwords. The CHI2 algorithm is based on the χ^2 statistic to select the minimum number of attributes. Stopwords by definition are meaningless words that have low discrimination power (Lo et al., 2005) [4]. Stopwords are deleting meaningless words to avoid ambiguous results.

In this paper, we compare the correlations between the 3 datasets with the number of features selected. In addition, we also compared the effect of using General Stopwords and Context-based Stopwords. General Stopword is obtained from the function on TF-IDF. While Context-based Stopwords is obtained from each dataset.

This paper is organized as follows: Section II contains similar research has ever done; Section III explains the methods used in this study; Section IV consists of an explanation of the results; Section V consists of conclusions and future research.

II. RELATED WORKS

Research on sentiment analysis by classifying customer reviews using machine learning or lexicon-based has been widely practiced.

S. Rani and P. Kumar [5] conducting research on sentiment analysis system which aims to improve teaching and learning. This research uses natural language processing and machine learning to analyze student feedbacks.

Y. M. Aye and S. S. Aung [6] research sentiment analysis customer review on a restaurant in Myanmar in the form of text by using Lexicon-based.

III. Approach

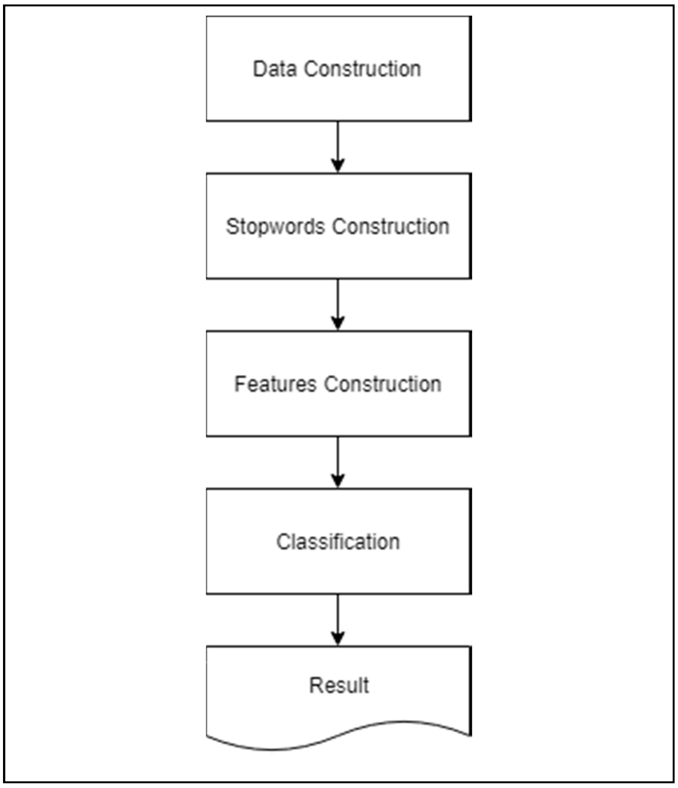

Fig. 1. Research Overview Diagram

Fig. 1 shows the flow from data construction to sentiment classification. The flow is divided by 4 part, the first part data construction, the second part is Stopwords Construction, and the third part features construction divided by 2 different processes that we want to compare using TF-IDF Stopwords and using Context-based Stopwords. The results of each algorithm used will be sorted and compared based on which has the highest scores on the F-score and accuracy.

A. Data Construction

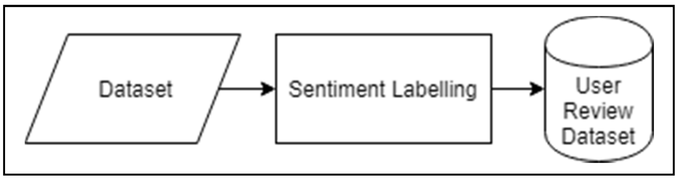

Fig. 2. Data Construction Diagram

There are 3 datasets used in this study. The dataset is a customer review of 3 websites that have different topics. The first topic is the food reviews obtained from the yelp.com website. The second topic is film reviews obtained from IMDb.com website. The third topic is the product reviews obtained from the amazon.com website. 3 datasets are obtained from The UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/datasets/Sentiment+Labelled+Sentences) [9]. The dataset consists of customer reviews and labels sentiments. Each dataset has been assigned a positive (1) or negative (0) label based on the sentiments of each review. Each dataset consists of 500 positives and 500 negatives. Fig. 2 shows how the dataset is formed.

B. Stopwords Construction

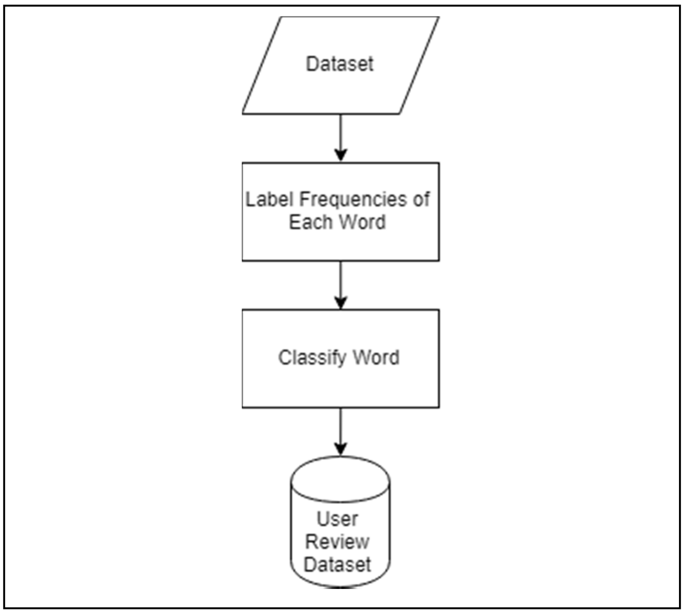

Fig. 3. Stopwords Construction Diagram

Stopwords is a meaningless word [7]. The presence of Stopwords makes the classification has a lower accuracy. By removing Stopwords can improve the final result. In this study, we compare the two Stopwords. That is the General Stopwords and the Context-based Stopwords. General Stopwords are obtained from existing TF-IDF functions. While Context-based Stopwords are obtained by searching for positive and negative ratios within the dataset.

Fig. 3 explains how Stopwords are formed. First, each dataset calculates the labeled frequency for each word. For example, can be seen in Table 1. Column words contain word list, column positives contain positive label frequency, column negative contains negative label frequency, column total contains the frequency of the word appears in the dataset.

TABLE I
DATASET LABEL FREQUENCY

No	Words	Positive	Negative	Total
0	Wow	2	0	2
1	...	20	19	39
2	Loved	2	0	2
3	this	48	48	96
4	place	56	49	105
5	.	387	434	821
6	Crust	0	1	1
7	is	105	69	174
8	not	17	89	106

Second, classify the word is it a stopword or not. One way that can be used is by finding the ratio of positive frequencies to the total. Ratio positive is the result of the positive frequency divided by total frequency as can be seen in Equation (1) below. Then we can conclude the word stopword or not in the Equation (2) below. For example, the word "is". The word "is" cannot be used to distinguish a sentiment. It is Actually more difficult because it has an ambiguous context.

$$Ratio_{positive} = \frac{Positive}{Total} \quad (1)$$

$$Stopwords \{IF\ Ratio_{positive} \geq 0.4\ AND\ Ratio_{positive} \leq 0.6\} \quad (2)$$

Lastly, Stopwords are saved into the array according to the initial dataset.

C. Features Construction

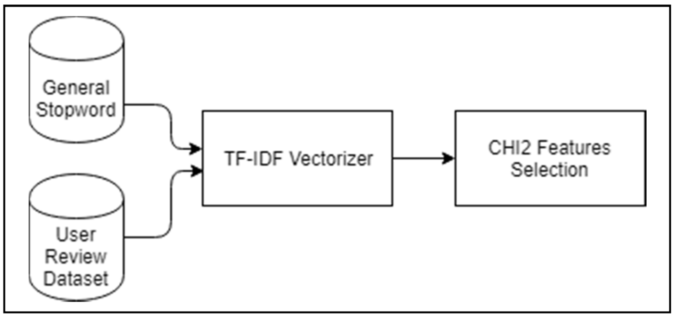

Fig. 4. Features Construction Diagram with General Stopwords

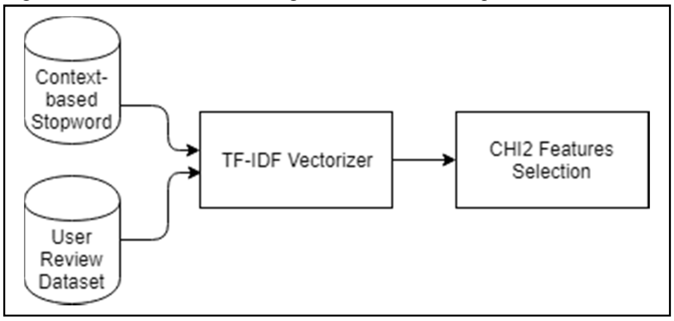

Fig. 5. Features Construction Diagram with Context-based Stopwords

In this section, we focus on generating features from a text and optimizing it. We used TF-IDF to generate features and we use Stopwords and CHI2 to optimize the feature. In this section, we use 2 different methods to compare. There is the use of Stopwords on TF-IDF and Context-based Stopwords.

TF-IDF or called term frequency-inverse document frequency is an efficient approach for determining the weight for each feature [12]. We used the TF-IDF Vectorizer Sklearn method in python to makes features in this study. The result of the features can be huge and may contain noisy and irrelevant features.

In this study, to make the feature filtered from the noise and irrelevant features we use 2 methods. There are Stopwords removal and CHI2 features selection. There are two Stopwords that will be compared. General Stopwords from TF-IDF (Fig. 4) and Context-based Stopwords (Fig. 5). The feature that contains Stopwords will be removed.

Besides using Stopwords Removal. In this study, we also use CHI2 features selection. Which is popular for sentiment analysis features selection. The CHI2 value represents the association between the text feature and its associated sentiment class [4], [13]. We used the CHI2 Features Selection Sklearn method in python to makes features in this study. We also compare the number of features used to determine the optimal number of features.

D. Classification

Fig. 6. Classification Diagram

In this section, we receive input in the form of features that have been fixed from the previous process and produce output in the form of predictable sentiment label from data test. That can be look in Fig.6. Machine learning that we use in doing data train and data test is Gaussian Naive Bayes. We used K-folds Cross-validation in order to obtain valid results.

Cross-validation is a statistical method that can be used to evaluate the performance of a model or algorithm in which data is separated into two subsets data train and data test. Data train is used for data training to build the model and the data test is used for data testing for validating the data. One of the Cross-validation is K-fold. K-fold is usually used because it can reduce computational time while maintaining the accuracy of the estimation. 10-fold Cross-validation is the best model selection, it tends to provide a less biased estimation of the accuracy [10].

Machine learning algorithm that we used for data training and data testing is Gaussian Naïve Bayes. The ratio of data train and data test in this paper is 9:1. Naïve Bayes is a popular machine learning for classifying. Many domains have minim misclassification rate [11].

E. Comparing Results

In this study, we compared the results of each dataset used. In addition, we compared the results from the minimal number features of CHI2 feature selections and compared the use of Stopwords on TF-IDF with Stopwords based on the context generated from the dataset. The results to be compared are measured using accuracy, precision, recall, and f1-score. Precision is the ability of a classification model to identify only the relevant data points [8], see Equation (3). Recall is the ability of a model to find all the relevant cases within a dataset [8], see Equation (4). F1-score is the harmonic mean of Precision and Recall [8], taking both metrics into account in the Equation (5). Accuracy is the quality or state of being correct or precise [8], see Equation (6).

$$Precision = \frac{TP}{(TP+FP)} \quad (3)$$

$$Recall = \frac{TP}{(TP+FN)} \quad (4)$$

$$F1Score = 2\,x \frac{(Precision\ x\ Recall)}{(Precision + Recall)} \quad (5)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

IV. DISCUSSION AND RESULT

I build a python program that has an input CSV file that contains the dataset of the review. The dataset contains comment and sentiment for each review. We took user review from UCI machine learning repository (https://archive.ics.uci.edu/ml/datasets/Sentiment+Labelled+Sentences) [9]. The dataset contains 500 positive and 500 negative sentences and sentence score is either 1(for positive) or 1(for negative). The result of the experimental study can be seen in Table 2.

TABLE II. EXPERIMENTAL RESULT

Dataset	Number of Features		Number of features best (CHI2)	Precision			Recall			accuracy			f1-scores		
	general stopwords	context-based Stopwords		general stopwords	context-based Stopwords	Diffrence	general stopwords	context-based Stopwords	Diffrence	general stopwords	context-based Stopwords	Diffrence	general stopwords	context-based Stopwords	Diffrence
Yelp.com	884	1078	5%	0.6	0.66	0.06	0.99	0.93	-0.06	0.68	0.72	0.04	0.72	0.73	0.01
			10%	0.64	0.66	0.02	0.98	0.98	0	0.75	0.77	0.02	0.75	0.77	0.02
			15%	0.67	0.7	0.03	0.99	0.98	-0.01	0.78	0.81	0.03	0.77	0.79	0.02
			20%	0.69	0.72	0.03	0.99	0.98	-0.01	0.8	0.83	0.03	0.79	0.8	0.01
			25%	0.71	0.77	0.06	0.99	0.97	-0.02	0.82	0.87	0.05	0.8	0.83	0.03
			30%	0.73	0.77	0.04	0.99	0.97	-0.02	0.85	0.87	0.02	0.82	0.84	0.02
			35%	0.75	0.79	0.04	0.99	0.97	-0.02	0.86	0.88	0.02	0.83	0.85	0.02
			40%	0.76	0.8	0.04	0.99	0.98	-0.01	0.88	0.89	0.01	0.84	0.86	0.02
			45%	0.77	0.8	0.03	0.99	0.98	-0.01	0.88	0.9	0.02	0.85	0.86	0.01
			50%	0.78	0.8	0.02	0.99	0.99	0	0.89	0.9	0.01	0.86	0.86	0
			55%	0.82	0.8	-0.02	0.99	0.99	0	0.91	0.9	-0.01	0.89	0.86	-0.03
			60%	0.78	0.8	0.02	0.98	0.99	0.01	0.88	0.9	0.02	0.85	0.86	0.01
			65%	0.78	0.8	0.02	0.97	0.98	0.01	0.88	0.9	0.02	0.85	0.86	0.01
			70%	0.77	0.8	0.03	0.94	0.98	0.04	0.86	0.9	0.04	0.84	0.86	0.02
			75%	0.76	0.79	0.03	0.92	0.97	0.05	0.84	0.89	0.05	0.82	0.85	0.03
			80%	0.75	0.79	0.04	0.89	0.95	0.06	0.82	0.88	0.06	0.8	0.84	0.04
			85%	0.72	0.78	0.06	0.84	0.94	0.1	0.79	0.87	0.08	0.76	0.83	0.07
			90%	0.69	0.77	0.08	0.81	0.92	0.11	0.76	0.85	0.09	0.73	0.81	0.08
			95%	0.68	0.76	0.08	0.79	0.88	0.09	0.74	0.82	0.08	0.71	0.79	0.08
			100%	0.66	0.74	0.08	0.77	0.85	0.08	0.71	0.79	0.08	0.69	0.76	0.07
Imdb.com	1186	1298	5%	0.6	0.71	0.11	0.99	0.85	-0.14	0.68	0.74	0.06	0.74	0.74	0
			10%	0.65	0.86	0.21	0.98	0.71	-0.27	0.75	0.8	0.05	0.78	0.77	-0.01
			15%	0.71	0.71	0	0.94	0.97	0.03	0.78	0.8	0.02	0.8	0.81	0.01
			20%	0.71	0.76	0.05	0.97	0.96	-0.01	0.8	0.84	0.04	0.82	0.84	0.02
			25%	0.74	0.96	0.22	0.97	0.79	-0.18	0.82	0.88	0.06	0.84	0.87	0.03
			30%	0.75	0.96	0.21	0.98	0.78	-0.2	0.85	0.87	0.02	0.85	0.86	0.01
			35%	0.77	0.96	0.19	0.98	0.83	-0.15	0.86	0.9	0.04	0.86	0.89	0.03
			40%	0.79	0.97	0.18	0.98	0.81	-0.17	0.88	0.89	0.01	0.87	0.88	0.01
			45%	0.8	0.98	0.18	0.98	0.81	-0.17	0.88	0.89	0.01	0.88	0.88	0
			50%	0.82	0.97	0.15	0.98	0.82	-0.16	0.89	0.9	0.01	0.89	0.89	0
			55%	0.82	0.98	0.16	0.98	0.83	-0.15	0.91	0.9	-0.01	0.89	0.9	0.01
			60%	0.83	0.98	0.15	0.98	0.84	-0.14	0.88	0.91	0.03	0.89	0.9	0.01
			65%	0.82	0.98	0.16	0.97	0.84	-0.13	0.88	0.91	0.03	0.89	0.9	0.01
			70%	0.81	0.97	0.16	0.97	0.82	-0.15	0.86	0.9	0.04	0.88	0.89	0.01
			75%	0.81	0.97	0.16	0.94	0.82	-0.12	0.84	0.9	0.06	0.87	0.89	0.02
			80%	0.79	0.94	0.15	0.91	0.81	-0.1	0.82	0.88	0.06	0.85	0.87	0.02
			85%	0.77	0.91	0.14	0.89	0.8	-0.09	0.79	0.87	0.08	0.83	0.85	0.02
			90%	0.76	0.88	0.12	0.87	0.8	-0.07	0.76	0.85	0.09	0.81	0.83	0.02
			95%	0.75	0.85	0.1	0.85	0.78	-0.07	0.74	0.83	0.09	0.79	0.81	0.02
			100%	0.73	0.82	0.09	0.82	0.77	-0.05	0.71	0.81	0.1	0.77	0.79	0.02
Amazon.com	826	1023	5%	0.6	0.64	0.04	0.99	0.97	-0.02	0.66	0.71	0.05	0.74	0.77	0.03
			10%	0.65	0.71	0.06	0.98	0.96	-0.02	0.73	0.79	0.06	0.78	0.82	0.04
			15%	0.71	0.85	0.14	0.94	0.88	-0.06	0.77	0.86	0.09	0.8	0.86	0.06
			20%	0.71	0.78	0.07	0.97	0.95	-0.02	0.79	0.85	0.06	0.82	0.86	0.04
			25%	0.74	0.8	0.06	0.97	0.96	-0.01	0.81	0.86	0.05	0.84	0.87	0.03
			30%	0.75	0.81	0.06	0.98	0.97	-0.01	0.83	0.87	0.04	0.85	0.88	0.03
			35%	0.77	0.82	0.05	0.98	0.97	-0.01	0.85	0.88	0.03	0.86	0.88	0.02
			40%	0.79	0.83	0.04	0.98	0.97	-0.01	0.86	0.88	0.02	0.87	0.89	0.02
			45%	0.8	0.84	0.04	0.98	0.97	-0.01	0.87	0.9	0.03	0.88	0.9	0.02
			50%	0.82	0.86	0.04	0.98	0.97	-0.01	0.88	0.91	0.03	0.89	0.91	0.02
			55%	0.82	0.87	0.05	0.98	0.97	-0.01	0.88	0.92	0.04	0.89	0.92	0.03
			60%	0.83	0.88	0.05	0.98	0.98	0	0.89	0.92	0.03	0.89	0.92	0.03
			65%	0.82	0.89	0.07	0.97	0.98	0.01	0.88	0.93	0.05	0.89	0.93	0.04
			70%	0.81	0.87	0.06	0.97	0.98	0.01	0.87	0.92	0.05	0.88	0.92	0.04
			75%	0.81	0.86	0.05	0.94	0.97	0.03	0.86	0.91	0.05	0.87	0.91	0.04
			80%	0.79	0.86	0.07	0.91	0.95	0.04	0.84	0.9	0.06	0.85	0.9	0.05
			85%	0.77	0.84	0.07	0.89	0.93	0.04	0.82	0.88	0.06	0.83	0.88	0.05
			90%	0.76	0.82	0.06	0.87	0.9	0.03	0.8	0.85	0.05	0.81	0.86	0.05
			95%	0.75	0.79	0.04	0.85	0.87	0.02	0.78	0.83	0.05	0.79	0.83	0.04
			100%	0.73	0.78	0.05	0.82	0.85	0.03	0.76	0.81	0.05	0.77	0.81	0.04

TABLE III.
A COMPARED TO CONTEXT-BASED STOPWORDS

Dataset	Number of Features	
	General Stopwords	*Context-based Stopwords*
Yelp	884	1078
IMDb	1186	1298
Amazon	826	1023

Table 3 shows about the number of features for each dataset. General Stopwords produces fewer features than those of Context-based Stopwords. General Stopwords remove more features than those of Context-based Stopwords.

TABLE IV
THE COMPARISON BETWEEN THIS STUDY AND ANOTHER STUDY

Dataset	Average Difference			
	Precision	*Recall*	*F1-Score*	*Accuracy*
Yelp	0.040	0.020	0.025	0.038
IMDb	0.145	-0.125	0.013	0.045
Amazon	0.059	0.001	0.036	0.048
Average	**0.081**	**-0.035**	**0.025**	**0.043**

Context-based Stopwords improves precision 8%, F1-score 2.5%, accuracy 4.3%, and decreases recall 3.5%. The results of the accuracy, precision, recall, and f1 scores are pretty good. In this experiment appears some problems like slang word, dataset number, and number in the sentence cannot be the feature. This result can be seen in Table 4.

However, in the case of IMDb dataset decreased recall score up to 12% and increased precision up to 14%. But, increase the f1-score up to 1.3% and accuracy 4.5%. This means, even though the recall reduces, the accuracy still increases. Table 4 shows the greatest difference average is in the precision score up to 8%.

TABLE V
THE COMPARISON BETWEEN THIS STUDY AND ANOTHER STUDY

Dataset		Methods	
		Context-based Stopwords	*GICF with Embeddings on Sentences [14]*
Accuracy	Yelp	90%	92%
	IMDb	91%	88%
	Amazon	93%	88%

Table 5 shows a comparison between Context-based Stopwords method in this study and GICF with embeddings on Sentence in “From Group to Individual Labels using Deep Features” study by K. Dimitrios, D. Misha, D.F. Nando, and S. Padhraic [14]. The comparison using the same dataset. GICF method using Convolutional Neural Network Deep Learning. Context-based Stopwords methods better than GICF in IMDb and Amazon dataset. In Yelp dataset, GICF has a better result than Context-based Stopwords.

There are three things that can improve accuracy in this study:

A. Slang Word

User review usually did not use standard sentences. For example, “I recommend this for EVERYONE who loves film, movies, anything...A Work of Art!” the word in “, anything...A” cannot be transformed to the exact word. So, to optimize the result we must build a slang word method. That can transform the sentence to “I recommend this for everyone who loves film, movies, anything a work of art”.

B. Dataset Number

In this study has 1000 sentences for each dataset. The features that produce from this dataset is not more than 1300. The bigger the dataset and the balance of the dataset can make better results. The bigger dataset can represent all the possibility.

C. The Number in The Sentence Cannot be a Feature

Some of the users put a rating in their review. For example, “1/10 - and only because there is no setting for 0/10.” It can be a key of the classifying. “0/10” or “1/10” it means a negative sentiment. If the TF-IDF can detect this to transform it into the vector it will help the results.

V. CONCLUSIONS AND FUTURE WORKS

According to the result of the approach from the dataset, we can conclude that the Supervised Method can be optimized by using CHI2 features selection and Stopwords Removal. The use of Context-based Stopwords enriched the number of Stopwords which removed biased features.

There are three parts to optimize the approach in this paper. The first part makes a function to detect and clean the slang word. The second part increases the dataset number that can represent all the possibility. The third part makes a function to detect user rating in the user review.

REFERENCES

[1] Internetworldstats.com. (2018). World Internet Users Statistics and 2018 World Population Stats. [online] Available at: https://www.internetworldstats.com/stats.htm [Accessed 6 May 2018].

[2] R. Rohini S and M. Emmanuel, “Feature Selection for Sentiment Analysis by using SVM,” Vol.84(5), December 2013, pp. 21-25.

[3] S. ChandraKala1 and C. Sindhu2, “OPINION MINING AND SENTIMENT CLASSIFICATION: A SURVEY,” Vol.3(1), October 2012, pp. 420-427.

[4] H. Liu and R. Setiono, “Chi2: Feature selection and discretization of numeric attributes,” IEEE Artificial Intelligence, 2017. doi:10.1109/tai.1995.479783

[5] S. Rani and P. Kumar, "A Sentiment Analysis System to Improve Teaching and Learning," in Computer, vol. 50(5), pp. 36-43, 2017. doi:10.1109/MC.2017.133

[6] Y. M. Aye and S. S. Aung, "Sentiment analysis for reviews of restaurants in Myanmar text," IEEE Software Engineering. Artificial Intelligence. Networking and Parallel/Distributed Computing (SNPD). Japan, pp. 321-326, 2017. doi:10.1109/SNPD.2017.8022740

[7] H. Saif, M. Fernandez, Y. He, and H. Alani, “SentiCircles for Contextual and Conceptual Semantic Sentiment Analysis of Twitter, ” Lecture Notes

in Computer Science The Semantic Web: Trends and Challenges, 2014, pp. 83-98. doi:10.1007/978-3-319-07443-6_7

[8] M. Jupri and R. Sarno, "Taxpayer compliance classification using C4.5, SVM, KNN, Naive Bayes and MLP," in 2018 International Conference on Information and Communications Technology (ICOIACT), 2018, pp. 297–303. DOI: 10.1109/ICOIACT.2018.8350710

[9] D. Dua and E.T. Karra. (2017). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

[10] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," In Proceedings of International Joint Conference on AI (IJCAI), 1995, pp. 1137–1145.

[11] P. Domingos and M. Pazzani, Machine Learning (1997) 29: 103. DOI:10.1023/A:1007413511361

[12] B.Y. Pratama and R. Sarno, "Personality classification based on Twitter text using Naive Bayes, KNN and SVM," International Conference on Data and Software Engineering (ICoDSE), 2015, pp. 170 – 174. DOI: 10.1109/ICODSE.2015.7436992.

[13] L. Yang, B. Jian-Wu, and F. Zhi-Ping. (2017). Multi-class sentiment classification: The experimental comparisons of feature selection and machine learning algorithms. Expert Systems with Applications. 80. 10.1016/j.eswa.2017.03.0

[14] K. Dimitrios, D. Misha, D.F. Nando, and S. Padhraic, "From Group to Individual Labels using Deep Features," in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015, pp.597-606.

[15] B.S. Rintyarna, R. Sarno, and C. Fatichah, "Enhancing the performance of sentiment analysis task on product reviews by handling both local and global context," International Journal of Information and Decision Sciences. 2018 Vol. 11.

[16] F.H. Rachman, R. Sarno, and C. Fatichah, "Music Emotion Classification based on Lyrics-Audio using Corpus Based Emotion," International Journal of Electrical and Computer Engineering (IJECE), vol. 8, no. 3, pp. 1720-1730, Juni 2018. DOI: 10.11591/ijece.v8i3.pp1720-1730

[17] E.W. Pamungkas, R. Sarno, and A. Munif, "B-BabelNet: Business-Specific Lexical Database for Improving Semantic Analysis of Business Process Models," Telecommunication, Computing, Electronics and Control (TELKOMNIKA), Vol 15, No 1, 2017. DOI: http://dx.doi.org/10.12928/telkomnika.v15i1.3176

[18] F.H. Rachman, R. Sarno, and C. Fatichah. "CBE : Corpus-Based off Emotion for Emotion Detection in Text Document," The 3rd International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE), Pp. 331 – 335, 2016. DOI: 10.1109/ICITACEE.2016.7892466

[19] B. Rintyarna and R. Sarno, "Adapted Weighted Graph for Word Sense Disambiguation," The 4th International Conference on Information and Communication Technology (ICoICT), 2016. DOI: 10.1109/ICoICT.2016.7571884