Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 4, October 2022
DOI : https://doi.org/10.33395/sinkron.v7i4.11706
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Comparison of Feature Extraction Methods on Sentiment Analysis in Hotel Reviews

Arie Satia Dharma [1)*], **Yosua Giat Raja Saragih** [2)],
[1,2)] Institut Teknologi Del, Laguboti, Indonesia
[1)]ariesatia@del.ac.id, [2)]yosuagrs@gmail.com

Submitted : Aug 20, 2022 | **Accepted** : Sep 26, 2022 | **Published** : Oct 3, 2022

Abstract: The development of technology causes things that done through meet in person or coming to a place can now be done by viewing information through gadgets or websites. Nowadays, to find out information about a place that provides accommodation for a vacation or a business visit, it can be done by accessing social media to see reviews from visitors who have visited the place, example, a hotel. Reviews given by hotel visitors are seen as more credible than information obtained from advertisements but the problem is that there are many reviews circulating on social media and it takes a time to analyze them. This study aims to analyze hotel reviews using the sentiment analysis method with the Support Vector Machine (SVM) approach. Sentiment analysis can be used to analyze the opinions of a large number of hotel visitors where it usually focuses on opinions that positive, negative and neutral. Before being analyzed with the support vector machine algorithm, 3 feature extraction methods will be used, namely Bag Of Words, TF-IDF and improvement TF-IDF to get the value of each word weight. The selection of these three methods is carried out by considering the influence of the presence of the same word feature in each review. In this comparison method, TF-IDF was found to be the best feature extraction method with 71.75% accuracy, 78.66% precision, 71.91% recall and 70.08% f1-score. The results obtained indicate that there are influence of features of the word in the hotel review data.

Keywords: Bag Of Words, Improvement TF-IDF, Sentiment analysis, Hotel , Support Vector machine, TF-IDF

INTRODUCTION

Technological Developments Making things that used to be done directly through through meet in person or coming to a place can now be done by viewing information through gadgets or websites where technological differences can affect the global economy (Zhu, 2021). Nowadays to find out information about a place to be visited for a vacation or a business visit such as a hotel, it can be done by accessing social media to see reviews from customers. In the past, customers found information about a hotel through brochures, but now customer reviews have changed the way goals by developing product descriptions from previous customer reviews (Silaa, Masui, & Ptaszynski, 2022). One of the places that got a review is the Hotel. Visitors who come to a hotel can provide comments that uploaded on social media. Tourists who have just visited will find it difficult to choose a hotel where there are many hotels chosen, one way for visitors to know the quality of a hotel is by looking at comments from previous hotel visitors (Sarudin, 2021).

Reviews given by hotel visitors are more credible and important than information obtained from advertisements (Lo & Yao, 2019) but the reviews circulating on social media are very large because reviews can be taken from several websites that provide reviews so that the amount of information available on the internet presents a challenge to find and draw conclusions from the many comments that exist. One technique to extract useful information from many reviews is to perform sentiment analysis (Wankhade, Rao, & Kulkarni, 2022). Sentiment analysis can be used to analyze the opinions of a large number of hotel visitors where sentiment analysis usually focuses on opinions that express positive, negative and neutral. Along with the development of social media such as discussion forums, blogs and internet sites, visitors began to use comments on social media to make decisions about which place to visit. Comments usually consist of several sentences and in general can be in the form of comments that are positive, neutral and negative (Himawan, Kaswidjanti, Sentimen, Sosial, & Based, 2018).

Two methods that can be used in sentiment analysis are lexicon-based and supervised learning or also using machine learning algorithms. Comparison of two sentiment analysis methods, namely lexicon-based and

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 4, October 2022
DOI : https://doi.org/10.33395/sinkron.v7i4.11706
e-ISSN : 2541-2019
p-ISSN : 2541-044X

machine learning, has been done several times with several methods. The combination of machine learning methods using the nave Bayes algorithm produces results in the form of better accuracy, precision and recall in the machine learning method (Kurniawan, Indriarti, & Adinugroho, 2019). Comparison between lexicon-based and machine learning is also done using the SVM algorithm as a machine learning algorithm and produces better accuracy and precision in the lexicon-based method but recall is better in the SVM algorithm because lexicon-based will work better along with the number of dictionaries that are used (Himawan et al., 2018). Comparison using SVM also gives better results of accuracy, precision and recall caused by the dictionary than the lexicon-based method which is less adequate (Najib, Irsyad, Qandi, & Rakhmawati, 2019).

Extending the method can make the classification even better. To perform the classification, data in the form of text must be taken for word features by using the feature extraction method. One of the feature extraction methods that can be used is Bag Of Words where Bag Of Words is the most common method used for text and object categorization (Qader, Ameen, & Ahmed, 2019). Using data from twitter, one study compared the feature extraction method using n-gram and TF-IDF with six types of classification algorithms obtained by TF-IDF giving better results about 3-4% compared to n-gram (Ahuja, Chug, Kohli, Gupta, & Ahuja, 2019). One of the studies using data in the form of news texts regarding finance, sports, military and entertainment has developed the weight of the features produced by TF-IDF to try to develop the TF-IDF feature extraction algorithm (Guo & Yang, 2016) which in this study will be referred to as improvements TF-IDF.

This study compares three feature methods, namely Bag Of Words, TF-IDF and also TF-IDF improvement using the SVM algorithm to see the results of accuracy, precision, recall and f1-score so that the results can show the influence of the presence of the same word feature in hotel reviews.

METHOD

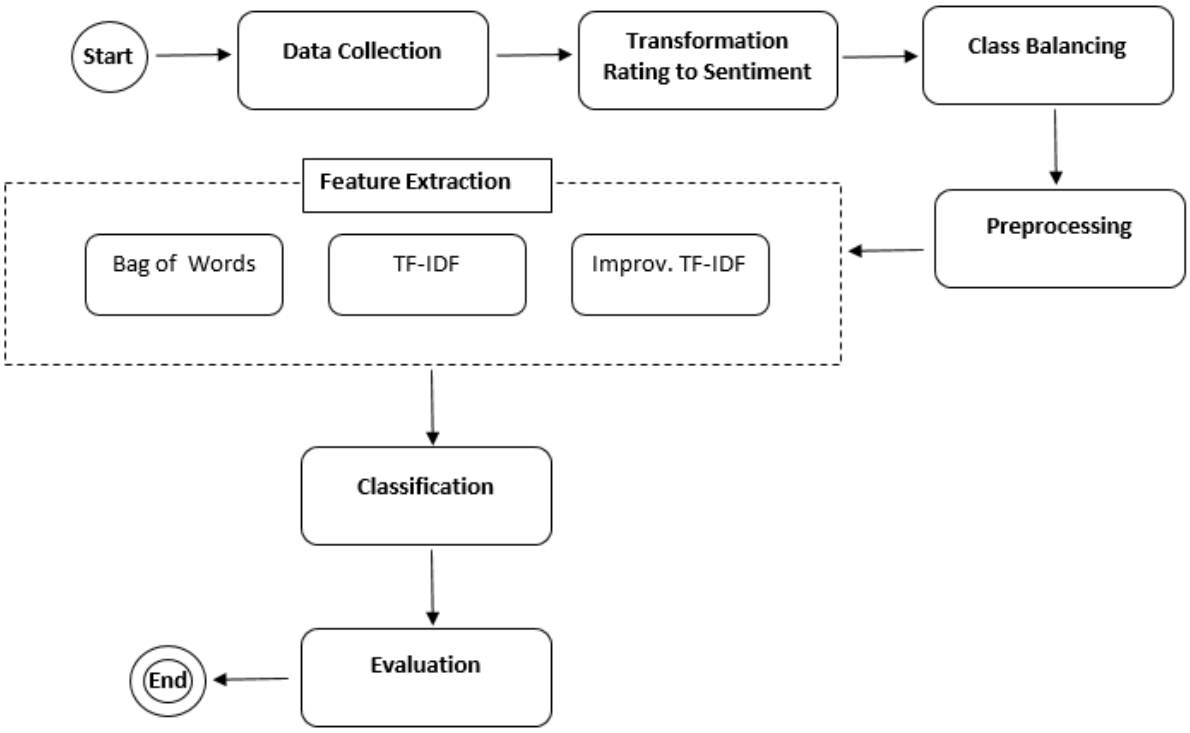

Fig 1. Flow of Research

This research begins by collecting data, transforming ratings into sentiments, balancing classes, preprocessing text, feature extraction using the three selected methods, then doing the classification process using SVM. The results of the classification are evaluated to determine which feature extraction method is the best. The flow of this research as shown in Figure 1 above consists of:

Data collection

Data from Tripadvisor of tabo cottages review obtained using the web scraping technique using web Harvy's tools in the form of ratings and reviews where ratings are still in the form of numbers so they must be changed from numbers to positive, negative and neutral sentiments.

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Tansformation of rating into Sentiment

The rating obtained in the form of numbers still worth 10 to 50 so it must be changed from numbers to positive, negative and neutral sentiments. Ratings with a value of 10 which means 'very bad' and 20 which means 'bad' have comments containing negative reviews so that 10 and 20 rating will be converted into negative sentiment. For a rating with a value of 30 which means 'average' has comments that contain negative and positive reviews so 30 rating will be changed to neutral sentiment. For ratings with a value of 40 which means 'very good' and 50 which means 'extraordinary' have comments that contain positive reviews so that 40 and 50 rating will be converted into positive sentiment

Class balancing

When the class is not balanced, the classification will usually be biased towards the majority class (Padurariu & Breaban, 2019) so it is necessary to balance the class of the data. The data obtained through web scrapping are 311 comments consist of 254 positive class, 16 negative class and 41 neutral class so that two methods will be carried out namely random oversampling by duplicating 14 negative class data so that there are 32 comments with negative class where duplication on data is only carried out on 16 data or not more than once because less varied data will cause overfitting (Ying, 2019). Furthermore, undersampling with positive and neutral class data totaling 254 and 41 so that the positive and neutral class data is the same as the negative class, which is 32. then, 96 data will be obtained with 32 positive classes, 32 neutral classes and 32 negative classes.

Preprocessing

Preprocessing has a significant effect on sentiment analysis (Pecar, Simko, & Bielikova, 2018), the data collected from tripadvisor still raw and have components that are not needed. The data still has to be processed to get good data for feature extraction, while the preprocessing stages are case folding, remove punctuation, remove stop words and lemmatizing.

Feature Extraction

To perform text classification, features are needed from text where the role of features are crucial because feature affecting classification accuracy (Liang, Sun, Sun, & Gao, 2017). The algorithm used to perform feature extraction in this study is TF-IDF, Bag Of Words, and improvement TF-IDF. The three methods will be used to perform feature extraction on each review that has been collected where later the extracted features will be used for classification where feature extraction will convert text data into word features and weights for classification.

Classification

SVM will be used to classify data into three classes, namely positive, negative and negative to obtain the evaluation results of the three feature extraction methods used. To be able to perform SVM classification requires two types of input, namely word features and target class where word features are obtained through the feature extraction method and the target class is the sentiment of each review.

Classification is carried out using k-fold cross validation which is one of the most widely used data resampling methods to estimate the actual model prediction error (Berrar, 2018). The value of k in k-fold cross validation is 4 so it is also called 4-fold cross validation, this is because there are 96 data so that the distribution of data for each experiment is 24 test data and also 72 train data with divisions for three classes, namely positive, negative and neutral so that there are 8 data for each class in the test data and 24 data for each class in the train data

Evaluation

From the classification results, accuracy, precision, recall and f1-score will be obtained. Experiments with cross validation will give four results for each feature extraction method so that the total of 12 experiments are obtained. The results of each experiment will be compared according to the average of accuracy, precision, recall and f1-score for each feature extraction method.

RESULT

From the web scraping technique, the number of comments obtained are 311 comments consist of 254 positive class, 16 negative class and 41 neutral class that have been obtained using the web scraping in the form of ratings and reviews where ratings. After the data is in the form of reviews and sentiments, the unbalanced data in each class is balanced into 32 data for each class and then preprocessing is successfully carried out. which is then converted into the form of word features according to the feature extraction algorithm used. Using k-fold cross validation and SVM from the sklearn library, the results obtained in the form of accuracy, precision, recall and f1-score.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 4, October 2022
DOI : https://doi.org/10.33395/sinkron.v7i4.11706
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The score of each feature extraction method is obtained from the results of accuracy, precision, recall and f1-score. The results of precision, recall and f1-score from the cross validation method will be averaged so that results are obtained based on each class, namely positive, negative and neutral which can be seen in Table 1.

Table 1.

Table of Precision, Recall and F1-score for each class

Feature Extraction	Negative			Positive			Neutral		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
Bag Of Words	97.25	100	98.5	54	71.75	61.25	62	37.5	44.5
TF-IDF	97.25	93.75	95	56	84.5	67	82.75	37.5	48.25
Improvement TF-IDF	90	100	94.5	54	53	52.75	57.5	50	52.75

Based on the positive class precision, recall and the highest f1-score obtained from the TF-IDF method. Based on the highest negative precision class obtained from the Bag Of Words and TF-IDF methods, the highest recall was obtained from the Bag Of Words method and the highest improvement TF-IDF and f1-score was obtained from the Bag Of Words method. Based on the neutral class, the highest precision was obtained by the TF-IDF method, the highest recall was obtained from the TF-IDF improvement method and the highest f1-score was obtained from the TF-IDF improvement method.

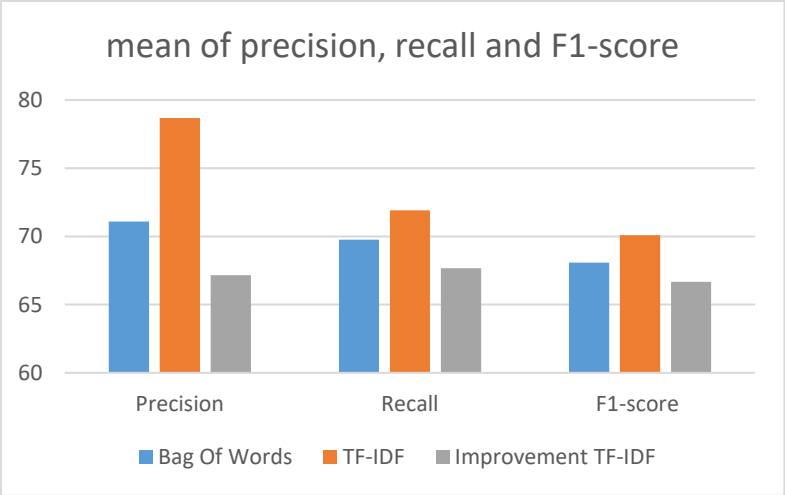

Fig 2. Average value of precision, recall and F1-score for all class

Based on Fig 2 we compared Bag Of words, TF-IDF and Improvement TF-IDF precision, recall and f1-score by the average from each class and the higest average value of precision, recall and f1-score is TF-IDF where TF-IDF obtain 78.66% precision, 71.91% recall and 70.08% f1-score while Bag Of Words obtain 71.08% precision, 69.75% recall and 68.08% f1-score and Improvement TF-IDF obtain 67.16% precision, 67.66% recall and 66.66% f1-score

Table 2

Table of Accuracy

No	***Bag Of Words***	**TF-IDF**	***Improvement* TF-IDF**
1	71	75	71
2	71	62	67
3	67	71	62
4	71	79	71
average	70	71.75	67.75

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 6, Number 4, October 2022
DOI : https://doi.org/10.33395/sinkron.v7i4.11706
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 2 shows the accuracy results of 4 test using cross validation, For the first, third and fourth test TF-IDF got the best accuracy results, in the second test Bag Of Words got the best accuracy and for the average accuracy we got TF-IDF which has the highest value of 71.75%.

DISCUSSIONS

Table 2 above shows the experimental results of the feature extraction method, namely Bag Of Words, TF-IDF and Improvement TF-IDF where each experiment produces different precision, recall and F1-scores depending on the class, namely negative, positive and neutral classes. The experiment was carried out 4 times according to the k-fold cross validation rule and classified using the SVM method. The results of classification give the value of each experiment carried out from each class and the total average is the average of the evaluation results of all experiments and classes.

The results obtained from the experiments that have been done showed that the TF-IDF feature extraction method obtains the highest score through accuracy, precision, recall, and F-1 score. Both TF-IDF and Improvement TF-IDF have the ability to calculate the number of feature words that appear more than once when compared to the Bag of Words. The use of TF-IDF will not reduce the weight of the word feature that only appears once. Based on the characteristics of the dataset itself, which has word features that appear more than once in several reviews, it can be understood why the highest score was obtained by TF-IDF, then Bag of Words, and last Improvement TF-IDF.

Therefor we can state that TF-IDF is suitable for feature extraction from data that has word feature occurrences that only appear once in each review but these features can appear repeatedly in the entire review as the features shown in this Tabo Cottage hotel review data.

CONCLUSION

Based on the results that have been obtained, the conclusions of this study for the best feature extraction method by comparing the results of accurary, precision, recall and f1-score from the Bag Of Words, TF-IDF and TF-IDF improvement methods is TF-IDF where TF-IDF obtains the highest evaluation results from every aspect, namely accuracy with 71.75% result, precision 78.66%, recall 71.91% and F1-score 70.08%. Based on the results obtained and the characteristics of the dataset used, we can conclude that there is an influence from the presence of the same word feature in hotel reviews data.

REFERENCES

Ahuja, R., Chug, A., Kohli, S., Gupta, S., & Ahuja, P. (2019). The impact of features extraction on the sentiment analysis. *Procedia Computer Science*, *152*. https://doi.org/10.1016/j.procs.2019.05.008

Berrar, D. (2018). Cross-validation. In *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics* (Vol. 1–3). https://doi.org/10.1016/B978-0-12-809633-8.20349-X

Guo, A., & Yang, T. (2016). Research and improvement of feature words weight based on TFIDF algorithm. *Proceedings of 2016 IEEE Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2016*. https://doi.org/10.1109/ITNEC.2016.7560393

Himawan, H., Kaswidjanti, W., Sentimen, A., Sosial, M., & Based, L. (2018). Metode Lexicon Based dan Support Vector Machine untuk Menganalisis Sentimen pada Media Sosial sebagai Rekomendasi Oleh-Oleh Favorit. *Seminar Nasional Informatika*, *2018*(November).

Kurniawan, A., Indriarti, & Adinugroho, S. (2019). Analisis Sentimen Opini Film Menggunakan Metode Naïve Bayes dan Lexicon Based Features. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(9).

Liang, H., Sun, X., Sun, Y., & Gao, Y. (2017). Text feature extraction based on deep learning: a review. *Eurasip Journal on Wireless Communications and Networking*, Vol. 2017. https://doi.org/10.1186/s13638-017-0993-1

Lo, A. S., & Yao, S. S. (2019). What makes hotel online reviews credible?: An investigation of the roles of reviewer expertise, review rating consistency and review valence. *International Journal of Contemporary Hospitality Management*, *31*(1). https://doi.org/10.1108/IJCHM-10-2017-0671

Najib, A. C., Irsyad, A., Qandi, G. A., & Rakhmawati, N. A. (2019). Perbandingan Metode Lexicon-based dan SVM untuk Analisis Sentimen Berbasis Ontologi pada Kampanye Pilpres Indonesia Tahun 2019 di Twitter. *Fountain of Informatics Journal*, *4*(2). https://doi.org/10.21111/fij.v4i2.3573

Padurariu, C., & Breaban, M. E. (2019). Dealing with data imbalance in text classification. *Procedia Computer Science*, *159*. https://doi.org/10.1016/j.procs.2019.09.229

Pecar, S., Simko, M., & Bielikova, M. (2018). Sentiment analysis of customer reviews: Impact of text pre-processing. *DISA 2018 - IEEE World Symposium on Digital Intelligence for Systems and Machines, Proceedings*. https://doi.org/10.1109/DISA.2018.8490619

Qader, W. A., Ameen, M. M., & Ahmed, B. I. (2019). An Overview of Bag of Words;Importance,

*name of corresponding author

e-ISSN : 2541-2019
p-ISSN : 2541-044X

Implementation, Applications, and Challenges. *Proceedings of the 5th International Engineering Conference, IEC 2019*. https://doi.org/10.1109/IEC47844.2019.8950616

Sarudin, R. (2021). ANALISIS ONLINE REVIEW TRIPADVISOR.COM TERHADAP MINAT PEMBELIAN PRODUK JASA AKOMODASI DI HOTEL MANHATTAN. *Jurnal Hospitality Dan Pariwisata*, *7*(1). https://doi.org/10.30813/jhp.v7i1.2634

Silaa, V., Masui, F., & Ptaszynski, M. (2022). A Method of Supplementing Reviews to Less-Known Tourist Spots Using Geotagged Tweets. *Applied Sciences (Switzerland)*, *12*(5). https://doi.org/10.3390/app12052321

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*. https://doi.org/10.1007/s10462-022-10144-1

Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, *1168*(2). https://doi.org/10.1088/1742-6596/1168/2/022022

Zhu, F. (2021). The Impact of High Technology on the Economy. *Proceedings - 2021 5th International Conference on Data Science and Business Analytics, ICDSBA 2021*. https://doi.org/10.1109/ICDSBA53075.2021.00069

*name of corresponding author