Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer
Vol. 22, No. 1, November 2022, pp. 25∼36
ISSN: 2476-9843, accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: 10.30812/matrik.v22i1.1848

Recognize The Polarity of Hotel Reviews using Support Vector Machine

Ni Wayan Sumartini Saraswati[1], I Gusti Ayu Agung Diatri Indradewi[2]
[1]Institut Bisnis dan Teknologi, Bali, Indonesia
[2]Universitas Pendidikan Ganesha, Bali, Indonesia

Article Info

Article history:

Received March 13, 2022
Revised October 26, 2022
Accepted November 10, 2022

Keywords:

Hotel Reviews
K-Fold Cross Validation
Support Vector Machines
Text Classification
TripAdvisor Review

ABSTRACT

A brand is very dependent on consumer perceptions of the product or services. In assessing consumer perceptions of products and services, companies are often faced with data analysis problems. One of the data that is very useful to produce a picture of consumer perceptions of the products and services is review data. So that the company's ability to process review data means that the company has a picture of the strength of the brand it has. Some of the most popular machine learning algorithms for creating text classification models include the naive Bayes family of algorithms, support vector machines (SVM) and deep learning algorithms. In this research, SVM has been proven to be a reliable method in pattern recognition. In particular, this study aims to produce a model that can be used to classify the polarity of hotel reviews automatically. The experimental data comes from review data on hotels in Europe sourced from TripAdvisor with a total of 38000 reviews. We also measure the quality of the classification engine model. The test results of the SVM model built from hotel review data are quite good. The average accuracy of the classification engine is 92.48%. Because the recall and precision values are balanced, the accuracy value is considered sufficient to describe the quality of the classification. .

Corresponding Author:

Ni Wayan Sumartini Saraswati,
Teknik Informatika,
Institut Bisnis dan Teknologi, Bali, Indonesia,
Email: sumartini.saraswati@stiki-indonesia.ac.id

How to Cite: N. W. Saraswati and I. G. A. Diatri Indradewi, Recognize The Polarity of Hotel Reviews Using Support Vector Machine, MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, vol. 22, no. 1, pp. 25-36, Nov. 2022.

1. INTRODUCTION

A hotel brand of course is very dependent on consumer perceptions of the services. In assessing consumer perceptions of services, hotel are often faced with data analysis problems. One of the data that is very useful to produce a picture of consumer perceptions of the service is review data. So that the hotel's ability to process review data means that the hotel has a picture of the strength of the brand it has.

In marketing, we know the term word of mouth (WoM). Where this WoM greatly affects how the decisions of new consumers on the product or service to be consumed. WoM, which used to run locally from the immediate environment of consumers, has now developed into electronic word of mouth (e-WoM) which touches the community very broadly. Especially in the tourism industry, tourists want a travel experience with a failure rate of 0%, where e-WoM is one of the factors that consumers will pay attention to.

As we know, we are currently in an era of data flooding but lack of information. Every second millions of consumers provide a review for a product or service. Moreover, the use of mobile phones will make this activity even easier. Knowing that the review given is a positive review or a negative review without reading one by one as a whole is something that is really needed to provide an overview of consumer perceptions. While the activity of manually labeling reviews available on the internet is a very expensive job and tends to be impossible to do. So to answer these needs, an automatic classification process is needed for consumer reviews on the internet. In this case, the text mining method in machine learning is one of the best solutions that can be used as shown by research [1].

To provide a comprehensive picture of a hotel's image, for example, what is the conclusion of a negative review of a hotel's services? whether consumers mostly mention small rooms, or expensive breakfasts is a advance analysis whose information is very useful not only for hotels to improve their image but also considerations for potential customers. Of course, before doing the analysis, we must first determine the polarity of the existing review.

This research aims to produce a model that is able to classify hotel reviews as positive reviews and negative reviews and measuring the quality of classifier machines and models in classifying hotel reviews as positive reviews and negative reviews using Support Vector Machines (SVM).

SVM has proven to be a reliable method for pattern recognition. Research [2, 3] shows that the SVM method is able to pattern student retention in the second semester of students based on several criteria, so the SVM model developed is able to predict the probability that students will drop out with an accuracy of 97.46%. In this study [4, 5], image classification of normal lungs and lungs of patients with COVID-19 was carried out. SVM provides performance with an average accuracy of 93.91%. Especially for text classification, SVM also provides good performance as was done in research [3, 6]. Meanwhile, other studies use the Nave Bayes Classification [7–9] and KNN [10, 11] methods in building a text classification model. Based on studies from research [12, 3, 13] that compares the performance of several text classification methods, SVM is reported to have the best model performance.

Considering the success of some of the studies mentioned above, to identify the polarity of hotel reviews in this study, we used the SVM method. In contrast to previous studies, we used SVM light which was developed by Thorsten Joachims with the consideration that this model is light and reliable for classifying large amounts of data.

2. RESEARCH METHOD

Text classification is one of the most popular processes in the text mining domain. Text mining is the process of exploring and analyzing large amounts of unstructured text data assisted by software that can identify concepts, patterns, topics, keywords, and other attributes in the data [14]. Text mining is an artificial intelligence (AI) technology that allows users to quickly convert the core content of a text document into quantitative data. The quantitative data can later be used or followed up according to the wishes of the users [2].

As part of AI processing, text classification also uses AI methods, one of which is machine learning methods. Some of the most popular machine learning algorithms for creating text classification models include the naive Bayes family of algorithms, support vector machines (SVM) and deep learning algorithms. In this study SVM will be used to detect hotel consumer review patterns in training data and classify review data automatically in trial data. In this study, we will also measure the quality of the classification engine model formed by SVM for hotel review data. In particular, this study aims to produce a model that can be used to classify the polarity of hotel reviews automatically.

Basically, Support Vector Machine (SVM) is a classification algorithm for linear and non-linear data. SVM uses non-linear mapping to transform the initial training data to a higher dimension. SVM technique is used to obtain the optimal hyperplane function to separate observations that have different target variable values. This hyperplane can be a line in two dimensions and can be a flat plane in multiple dimensions.

The Support Vector Machine method has several advantages, namely: (1) Generalization: Generalization is defined as the ability of a method to classify a pattern, which does not include the data used in the learning phase of the method. (2) Curse of dimensionality: Curse of dimensionality is defined as the problem faced by a pattern recognition method in estimating parameters due to the relatively small number of data samples compared to the dimensions of the vector space. (3) Feasibility: SVM can be implemented relatively easily, because the process of determining the support vector can be formulated in the Quadratic Programming (QP) problem.

The following research is included in the category of software development research, where the stages consist of collecting review data, compiling training data and test data, building SVM models and testing models. Text Preprocessing is the stage where the application selects the data to be processed in each document. This preprocessing process includes (1) case folding, (2) tokenizing, (3) filtering, and (4) stemming.

Not all text documents are consistent in the use of capital letters. Therefore, the role of Case Folding is needed in converting the entire text in the document into a standard form (usually lowercase). Tokenizing stage is the stage of cutting the input string based on each word that composes it. Tokenization broadly breaks down a set of characters in a text into word units, how to distinguish certain characters that can be treated as word separators or not. Filtering stage is the stage of taking important words from the token results. Can use stoplist algorithm (remove less important words) or wordlist (save important words). Stoplists/stopwords are non-descriptive words that can be discarded in the bag-of-words approach. Words like from, which, at, and to are some examples of high-frequency words that can be found in almost every document (referred to as stopwords). Removing this stopword can reduce index size and processing time. In addition, it can also reduce the noise level.

Index creation is done because a document cannot be recognized directly by the Information Retrieval System (IRS). Therefore, the document first needs to be mapped into a representation using the text in it. Stemming technique is needed in addition to reducing the number of different indexes of a document, also to group other words that have the same basic word and meaning but have a different form because they get different affixes. However, like stopping, stemming performance also varies and often depends on the language domain used. The stemming process in Indonesian texts is different from stemming in English texts. In English text, the only process needed is removing the suffix. Meanwhile, in Indonesian texts, all affixes, both suffixes and prefixes, are also omitted.

Some of basic processes that was used in this research are:

2.1. Cross Validation

Cross validation is a common method for evaluating the performance of text classifiers. It consists in randomly splitting the training data set into sample sets of equal length (eg 4 sets with 25% data). For each set, the text classifier is trained with the remaining sample (eg 75% of the sample). Next, the classifier makes predictions on the respective sets and the results are compared with the human annotated tags. This makes it possible to find out when the prediction was correct (true positive and true negative) and when it made an error (false positive, false negative).

With these results, researchers can create useful performance metrics for a quick assessment of how well the classifier is performing, including:

1. Accuracy : percentage of predicted text with correct tags.
2. Precision : the percentage of samples that the classifier gets from the total number of samples predicted for a particular tag.
3. Recall : the percentage of examples predicted by the classifier for a particular tag from the total number of examples that should be predicted for that tag.
4. F1 Score : average precision and harmonious gain.

2.2. Support Vector Machines

Pattern recognition is a field in computer science that maps data into certain predefined concepts. This particular concept is called a class or category. Various methods are known in pattern recognition, such as linear discriminant analysis, hidden Markov models, to artificial intelligence methods such as artificial neural networks. One method that has received much attention as a state of the art in pattern recognition is the Support Vector Machine.

Support Vector Machines (SVM) is a set of guided learning methods that analyze data and recognize patterns, used for classification and regression analysis. The original SVM algorithm was created by Vladimir Vapnik and the current standard derivative (soft margin) was proposed by Corinna Cortes and Vapnik Vladimir. Standard SVM takes a set of input data, and predicts, for any given input, the probability that the input is a member of one of the two classes, which makes an SVM a binary linear non-probabilistic classifier. Since an SVM is a classifier, then assigned a training set, each marked as belonging to one of two categories, an SVM training algorithm constructs a model that predicts whether the new data falls into one category or another. In this research we use SVM light engine by joachims that is an implementation of Vapnik's Support Vector Machine.

The main idea of the SVM method is the concept of a maximal hyperplane margin. By finding the maximum hyperplane margin, the vector will divide the data into the most optimum classification form. Some examples of hyperplanes that may appear to classify data are shown in Figure 1.

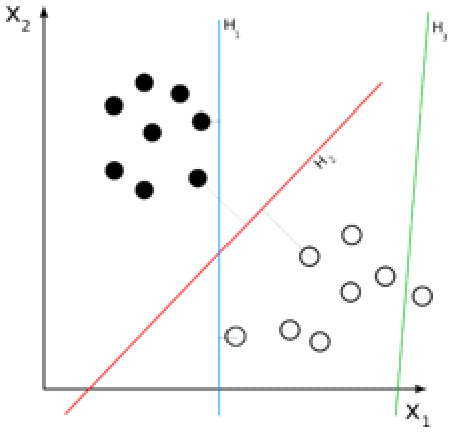

Figure 1. Examples of multiple hyperlanes

From Figure 1, it can be seen that the line H3 (green) does not separate the two classes. The H1 (blue) line separates, with a small margin and the H2 (red) line with the maximum margin. Classifying data is a common task in machine learning. Suppose some given data points each belong to one of two classes, and the goal is to determine the class of a new data point to be tested. In the case of SVM, the data points are viewed as p-dimensional vectors (a list of p sums), and we want to know whether we can separate these points with a (p - 1) dimensional hyperplane. This is called a linear classifier. There are many possible hyperplanes to classify data. One normal choice as the best hyperplane is the one that represents the separation with the greatest margin, between the two classes. So we choose a hyperplane so that the distance to and from the nearest data point on each side is maximized. If such a hyperplane exists, it is known as the maximum margin hyperplane and the linear classifier it defines is known as the maximum margin classifier. The illustration of the maximum margin hyperplane is shown in Figure 2 below. The maximum-margin hyperplane and margin for an SVM were trained with samples from the two classes. The sample at the margin is called the support vector.

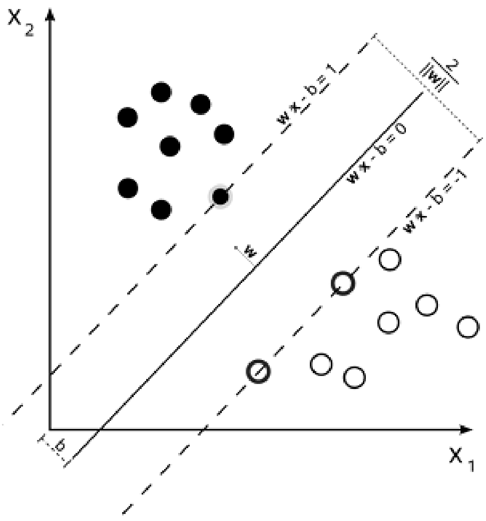

Figure 2. Hyperplane maximum margin

Figure 3 below describes the stages of building the SVM model in classifying text. Starting with tokenization to build a bag of words, the system indexes the words that appear and calculates their frequency. So that the resulting vector format in weighting by calculating (term frequency- Inverse Document Frequency) tf-idf of the words that appear. Tf-idf is an algorithm that can be used to analyze the relationship between a phrase/sentence and a set of documents. The next step is to eliminate stopwords (words that are not meaningful). Then the training for training data is carried out in developing an SVM model that recognizes patterns. The process is then continued with the classification carried out for the trial data whether it includes positive or negative reviews. The results of the classification of the test data are measured for their performance.

To maintain the validity of the performance, cross validation is carried out on the testing process. As explained above, cross validation exchanges training data and test data in the data sample domain continuously.

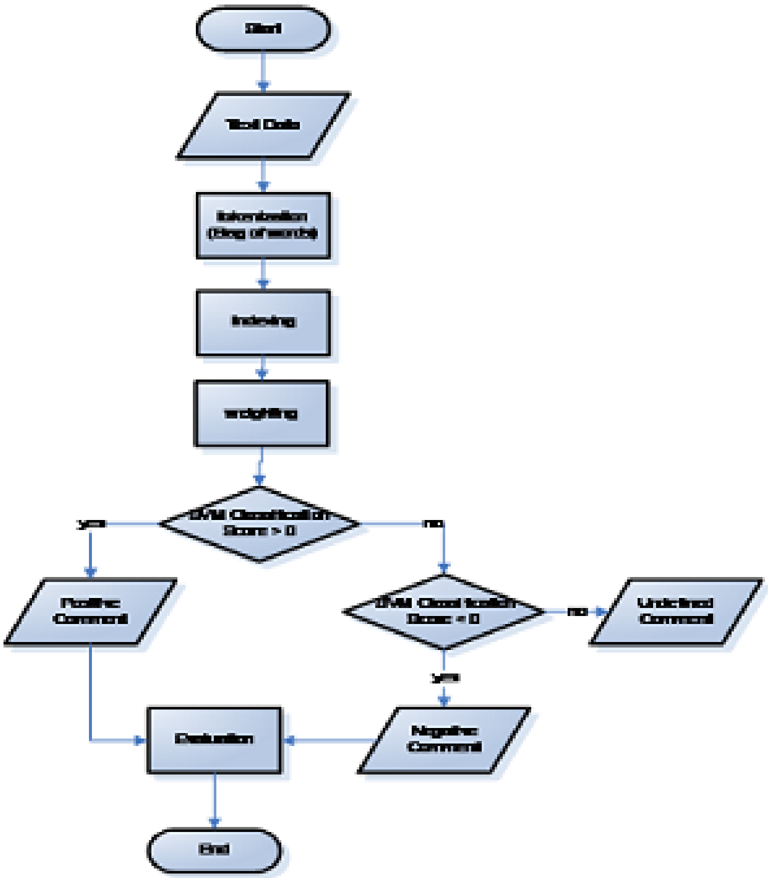

Figure 3. SVM Classification Process Flowchart

3. RESULTS AND ANALYSIS

Data collection is done by studying literature on online and offline media. The hotel review data source is downloaded from the site https://www.kaggle.com/jiashenliu/515k-hotel-reviews-data-in-europe in English format. The following figure 4 shows a snippet of the review data. It is a complete sentences of hotel review in English. The review has been labeled as positive reviews and negative reviews.

Table 1. Review data snippet

Only the park outside of the hotel was beautiful
No real complaints the hotel was great great location surroundings rooms amenities and service Two recommendations however firstly the staff upon check in are very confusing regarding deposit payments and the staff offer you upon checkout to refund your original payment and you can make a new one Bit confusing Secondly the on site restaurant is a bit lacking very well thought out and excellent quality food for anyone of a vegetarian or vegan background but even a wrap or toasted sandwich option would be great Aside from those minor minor things fantastic spot and will be back when i return to Amsterdam
Location was good and staff were ok It is cute hotel the breakfast range is nice Will go back
Great location in nice surroundings the bar and restaurant are nice and have a lovely outdoor area The building also has quite some character
Amazing location and building Romantic setting
Good restaurant with modern design great chill out place Great park nearby the hotel and awesome main stairs
The room is spacious and bright The hotel is located in a quiet and beautiful park
Good location Set in a lovely park friendly staff Food high quality We Oth enjoyed the breakfast
The room was big enough and the bed is good The breakfast food and service on the hotel is good outside the hotel there is a big park which is very good for walk in the morning and evening Many people are having picnics and do some bicycling
Rooms were stunningly decorated and really spacious in the top of the building Pictures are of room 300 The true beauty of the building has been kept but modernised brilliantly Also the bath was lovely and big and inviting Great more for couples Restaurant menu was a bit pricey but there were loads of little eatery places nearby within walking distance and the tram stop into the centre was about a 6 minute walk away and only about 3 or 4 stops from the centre of Amsterdam Would recommend this hotel to anyone it s unbelievably well priced too
Style location rooms
Comfy bed good location
This hotel is being renovated with great care and with an appreciation for its unique structure and location My spacious and comfortable room had a large double paned glass window onto the lush greenery of the park The breakfast selection was spectacular All considered this was a great hotel for the price and I plan to return
It was very good very historic building that s why I chose it
This hotel is awesome I took it sincirely because a bit cheaper but the structure seem in an hold church close to one awesome park Arrive in the city are like 10 minutes by tram and is super easy The hotel inside is awesome and really cool and the room is incredible nice with two floor and up one super big comfortable room I ll come back for sure there The staff very gentle one Spanish man really really good
Great onsite cafe Amazing building Park location Amazing Bobby Gin and Tonic
We loved the location of this hotel The fact that it is set in a Park away from the busy centre of dam square was great The tram system was brilliant and easy to handle The hotel is lovely and the bed was comfy Staff were very friendly and helpful and familiarized themselves with us when they realized we travelled from Ireland
Public areas are lovely and the room was nice but the window was broken and the drains in the bathroom smelt Its an old building and clearly has old building issues ...

The data source is as shown in Table 1 above, then we remove the stopwords. So that the data as shown in Table 2 below.

Table 2. Data snippet after stop word removal and stemming

Only park beautiful
No real complaints great great location surroundings rooms amenities service Two recommendations firstly staff check confusing deposit payments staff offer checkout to refund original payment Bit confusing Secondly site restaurant bit lacking well thought excellent quality food vegetarian vegan background wrap toasted sandwich option great Aside minor minor things fantastic spot will i return to Amsterdam
Location good staff It cute breakfast range nice Will back
Great location nice surroundings bar restaurant nice lovely outdoor area The building character
Amazing location building Romantic setting
Good restaurant modern design great chill place Great park nearby awesome main stairs
The room spacious bright The located quiet beautiful park
Good location Set lovely park friendly staff Food high quality We Oth enjoyed breakfast
The room big bed good The breakfast food service good big park good walk morning evening Many people picnics bicycling
Rooms stunningly decorated spacious top building Pictures room 300 The true beauty building modernised brilliantly Also bath lovely big inviting Great couples Restaurant menu bit pricey loads little eatery places nearby walking distance tram centre 6 minute walk 3 4 stops centre Amsterdam Would recommend to s unbelievably well priced too
Style location rooms
Comfy bed good location
This renovated great care appreciation unique structure location My spacious comfortable room large double paned glass window lush greenery park The breakfast selection spectacular All considered great price I plan to return
It good historic building s I chose it
This awesome I sincirely bit cheaper structure hold church close to awesome park Arrive city 10 minutes tram super easy The inside awesome cool room incredible nice floor super big comfortable room I ll The staff gentle Spanish man good
Great onsite cafe Amazing building Park location Amazing Bobby Gin Tonic
We loved location The fact set Park busy centre dam square great The tram system brilliant easy to handle The lovely bed comfy Staff friendly helpful familiarized realized travelled Ireland
Public areas lovely room nice window broken drains bathroom smelt Its building clearly building issues
...

The data as shown in Table 2 then we process with the tfidf method to make it as vector data. The processing results are as shown in Figure 4 below.

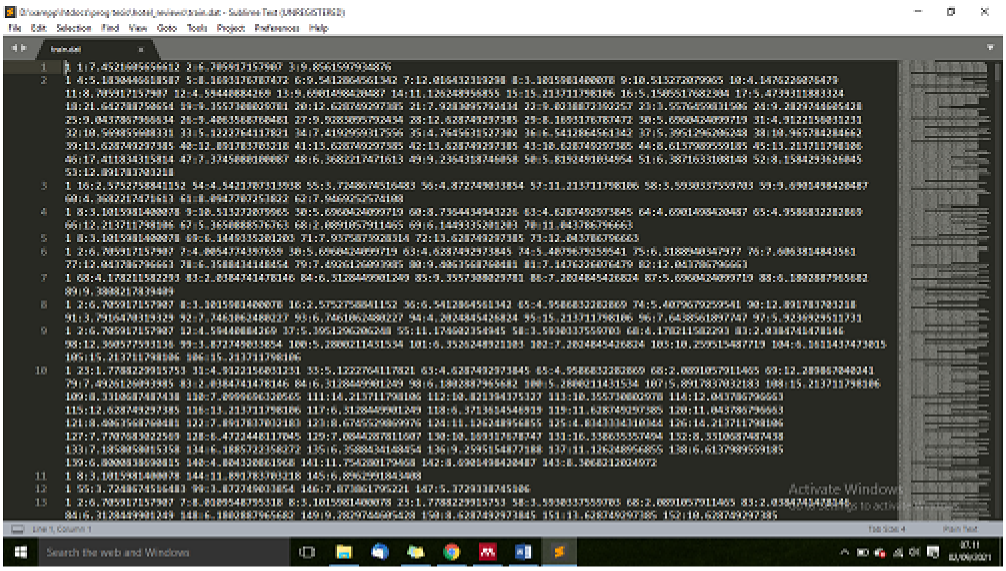

Figure 4. Vector Data

We then fold the vector data into 10 cross folds where the resulting pieces are shown in Figure 5 below. After the data folding is ready, then training and classification is carried out one by one according to the folding data.

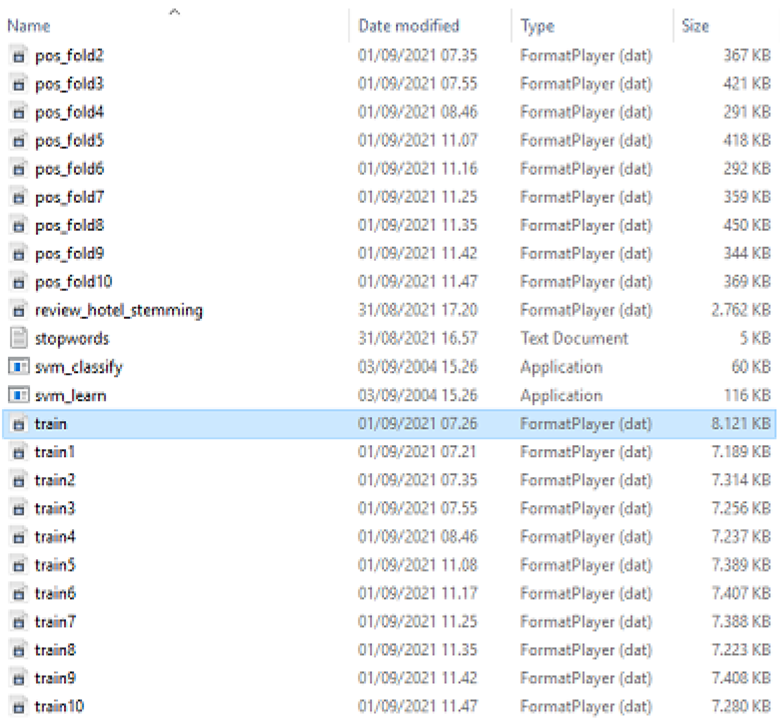

Figure 5. Results of folding

The measurement of the validity of the model is carried out after the classification process produces a confusion matrix as shown in the Table 3. For each experiment in the 10-fold cross validation we will know the number of positive reviews classified as positive reviews (TP), positive reviews classified as negative reviews (FN), negative reviews classified as negative reviews (TN), and negative reviews classified as negative reviews. classified as a positive review (FP).

The classification results are measured in units of accuracy, recall, specificity, precision, F1 score and error rate. We calculate the value of each measure with equation (1) to equation (5) where P is the number of positive reviews and N is the number of negative reviews.

Table 3. Confusion Matrix

	Actual Values Positive	Actual Values Negative
Predicted Values Positive	TP	FP
Predicted Values Negative	FN	TN

$$Accuracy = \frac{TP + TN}{P + N} \tag{1}$$

$$Recall = \frac{TP}{P} \tag{2}$$

$$Recall = \frac{TN}{N} \tag{3}$$

$$F1Score = \frac{2 \times precision \times recall}{precision + recall} \tag{4}$$

$$Accuracy = \frac{FP + FN}{P + N} \tag{5}$$

The Table 4 shows all the test results from each fold of data, from fold 1 to fold 10. In the last line, the average is done as the final result of the measurement that can be used in drawing conclusions.

Table 4. Classification Results

Fold	Accuracy	Recall	Specificity	Precision	F1 score	Error Rate
1	94,68%	94,47%	94,89%	94,87%	94,67%	5,32%
2	92,00%	89,42%	94,58%	94,28%	91,79%	8,00%
3	92,58%	92,11%	93,05%	92,99%	92,54%	7,42%
4	90,92%	88,89%	92,95%	92,65%	90,73%	9,08%
5	91,18%	88,95%	93,42%	93,11%	90,98%	8,82%
6	91,47%	88,21%	94,74%	94,37%	91,19%	8,53%
7	91,87%	90,68%	93,05%	92,88%	91,77%	8,13%
8	93,39%	93,84%	92,95%	93,01%	93,42%	6,61%
9	92,89%	91,63%	94,16%	94,01%	92,80%	7,11%
10	93,79%	93,32%	94,26%	94,21%	93,76%	6,21%
Average	**92,48%**	**91,15%**	**93,81%**	**93,64%**	**92,37%**	**7,52%**

The model shows good performance with a classification accuracy of 92.48%. This shows that from 100 classification test data, the machine is able to classify 92 data correctly. The tendency of introduction to positive reviews represented by recall is 91.15% and recognition of negative reviews represented by precision of 93.64% is also quite balanced, so that accuracy can represent the quality of the classification well.

The measurement chart from Table 1 is presented in two forms as shown in Figures 6 and Figure 7 below. From figures 6 we see imbalance between recall and specificity values, but this gap is not too large. The poor performance is shown by the data in the 2nd, 4th, 5th, 6th, 7th and 9th folds where there is an imbalance between the recall and precision values but not too far away. A small recall value indicates the model's weakness in recognizing positive tuples, which in this case is recognizing positive reviews.

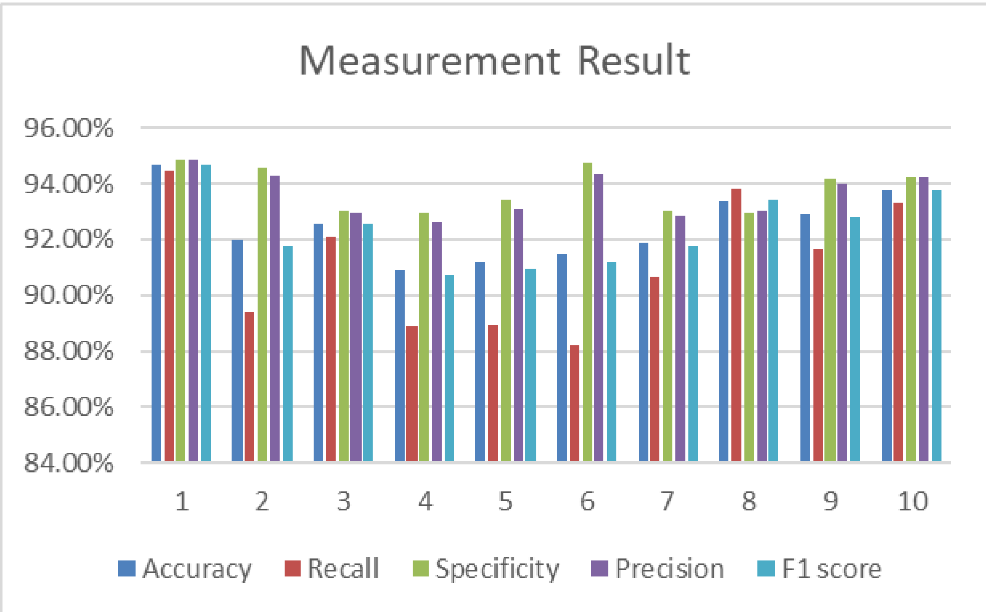

Figure 6. Graph of measurement results based on folding

From figures 7 we can see that the variation in the results between trials of 10 cross validation folding did not show a big difference. So we can say that the model is quite stable with a wide range of experimental data. We can also analyze that the resulting model is quite good with graph stability between recall, specificity and precision.

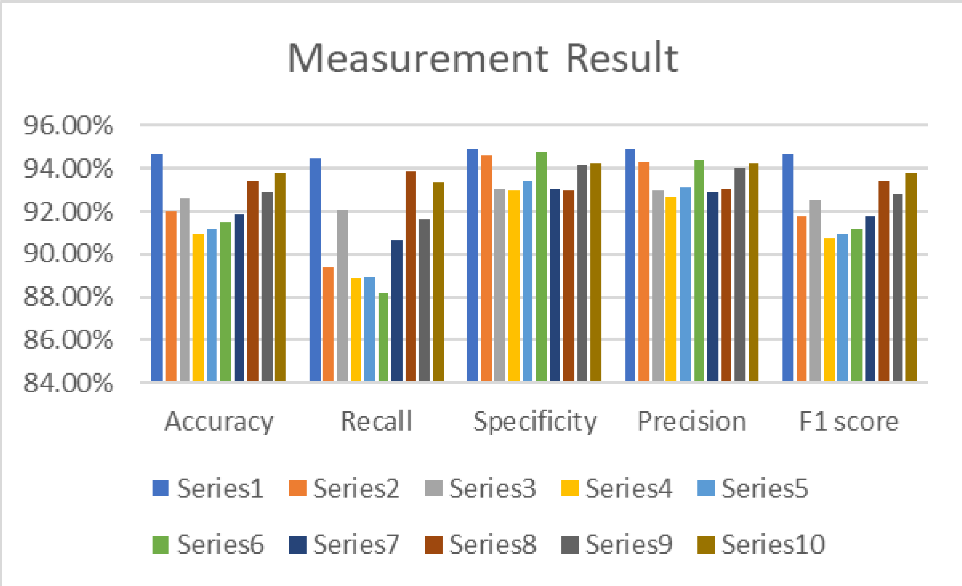

Figure 7. Graph of measurement results by unit of measurement

The success of the SVM method to determine the polarity of hotel reviews was also confirmed by the report on the success of the SVM method for a similar case where [7] reported that SVM provided 81% accuracy to classify the TripAdvisor Hotel Reviews dataset which is already on the Kaggle, and [15] stated that SVM contributed 89.86% accuracy for classifying hotels in Indonesia-language review.

4. CONCLUSION

The stages in classifying hotel review texts are eliminating stopwords, stemming, tokenization, tf-idf weighting to vector data, folding, training and classification.The classification results show an accuracy of 92.48%, recall 91.15%, specificity 93.81%, precision 93.64% and F1 score 92.37%. Balance performance of recall dan precision show that the model is good for classification and accuracy is enough to show the quality of model classification.

This study contributes that the use of SVM light for hotel review classification provides very high performance with light and fast computing for large amounts of data. So that SVM Light can be recommended for text classification in future research.

Further research development can be done by comparing the results of this study with different text classification methods, such as Random Forest and Naive Bayes Classification.

5. ACKNOWLEDGEMENTS

We would like to thank LPPM INSTIKI for the financial support for this research. We also extend our appreciation to previous researchers who contributed to the framework for this research.

6. DECLARATIONS

AUTHOR CONTIBUTION

The first author contributed in designing the research outline and building the SVM model, while the second author contributed in carrying out data preprocessing.

FUNDING STATEMENT

This research was fully funded from a PDM grant organized by LPPM STMIK STIKOM Indonesia in 2021.

COMPETING INTEREST

The next research that becomes the competing interest of this research is how to build a clustering topic model from the data.

REFERENCES

REFERENCES

[1] H. Irawan, G. Akmalia, and R. A. Masrury, "Mining tourist's perception toward Indonesia tourism destination using sentiment analysis and topic modelling," in *ACM International Conference Proceeding Series*, no. 1, 2019, pp. 7–12.

[2] N. W. S. Saraswati, K. K. Widiartha, and L. P. A. Prapitasari, "Vector machine to predict student retention: A computerized approach," *Journal of Physics: Conference Series*, vol. 1469, no. 1, p. 012045, feb 2020. [Online]. Available: https://iopscience.iop.org/article/10.1088/1742-6596/1469/1/012045

[3] F. Fatmawati and M. Affandes, "Klasifikasi Keluhan Menggunakan Metode Support Vector Machine (SVM) Pada Akun Facebook Group iRaise Helpdesk," *Jurnal CoreIT: Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*, vol. 3, no. 1, p. 24, 2018.

[4] I. G. A. A. D. Indradewi, N. W. S. Saraswati, and N. W. Wardani, "COVID-19 Chest X-Ray Detection Performance Through Variations of Wavelets Basis Function," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 1, pp. 31–42, 2021.

[5] N. W. S. Saraswati, N. W. Wardani, and I. G. A. A. D. Indradewi, "Detection of Covid Chest X-Ray using Wavelet and Support Vector Machines," *International Journal of Engineering and Emerging Technology*, vol. 5, no. 2, pp. 116–121, 2020.

[6] A. Apriani, H. Zakiyudin, and K. Marzuki, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta," *Jurnal Bumigora Information Technology (BITe)*, vol. 3, no. 1, pp. 19–27, 2021.

[7] I. P. A. M. Utama, S. S. Prasetyowati, and Y. Sibaroni, "Multi-Aspect Sentiment Analysis Hotel Review Using RF, SVM, and Naïve Bayes based Hybrid Classifier," *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, pp. 630–639, apr 2021. [Online]. Available: https://ejurnal.stmik-budidarma.ac.id/index.php/mib/article/view/2959

[8] K. Suppala and N. Rao, "Sentiment analysis using naïve bayes classifier," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 8, pp. 264–269, 2019.

[9] P. P. Surya and B. Subbulakshmi, "Sentimental Analysis using Naive Bayes Classifier," in *2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN)*. IEEE, mar 2019, pp. 1–5. [Online]. Available: https://ieeexplore.ieee.org/document/8899618/

[10] J. Harsono, R. M. No, P. Minggu, and J. S. Jakarta, "Klasifikasi Teks Berbahasa Indonesia Pada Artikel Berita Menggunakan Metode K-Nearest Neighbor Dengan Fungsi Squared Euclidean Distance Classification of Indonesian Text on News Articles Using K-Nearest Neighbor Method With Squared," *BRITech (Jurnal Ilmiah Ilmu Komputer, Sains dan Teknologi Terapan)*, vol. 1, no. 2, pp. 60–65, 2020.

[11] C. F. Suharno, M. A. Fauzi, and R. S. Perdana, "Klasifikasi Teks Bahasa Indonesia Pada Dokumen Pengaduan Sambat Online Menggunakan Metode K-Nearest Neighbors Dan Chi-square," *Systemic: Information System and Informatics Journal*, vol. 3, no. 1, pp. 25–32, 2017.

[12] B.-M. Hsu, "Comparison of Supervised Classification Models on Textual Data," *Mathematics*, vol. 8, no. 5, pp. 1–16, may 2020. [Online]. Available: https://www.mdpi.com/2227-7390/8/5/851

[13] S. Efendi and P. Sihombing, "Sentiment Analysis of Food Order Tweets to Find Out Demographic Customer Profile Using SVM," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 3, pp. 583–594, jul 2022. [Online]. Available: https://journal.universitasbumigora.ac.id/index.php/matrik/article/view/1898

[14] H. Hairani, A. S. Suweleh, and D. Susilowaty, "Penanganan Ketidak Seimbangan Kelas Menggunakan Pendekatan Level Data," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 20, no. 1, pp. 109–116, 2020.

[15] A. Taufik, "Komparasi Algoritma Klasifikasi Text Mining Untuk Analisis Sentimen Pada Review Restoran," *Jurnal Teknik Komputer AMIK BSI*, vol. 4, no. 2, pp. 112–118, 2018.