JURNAL INFORMATIKA, Vol. 9 No. 2 Oktober 2022, Halaman 180~186
ISSN: 2355-6579 | E-ISSN**:** 2528-2247

Classification of Gimbal Stabilizer Products Using Naive Bayes Classification

Artika Surniandari[1], Hilda Rachmi[2], Suparni[3], Lisda Widiastuti[4], Haryani[5]

[1,2,3,4,5] Universitas Bina Sarana Informatika
Jl. Kramat Raya No 98 Jakarta Pusat, Indonesia

e-mail: [1]artika.ats@bsi.ac.id, [2]hilda.hlr@bsi.ac.id, [3]suparni.spn@bsi.ac.id, [4]lisda.ltt@bsi.ac.id, [5]haryani.hyi@bsi.ac.id

Informasi Artikel	Diterima: 03-10-2022	Direvisi: 18-10-2022	Disetujui: 26-10-2022

Abstrak

Menjadi videografer adalah hobi yang populer di masa pandemi ini karena berkreasi dalam bentuk video dan konten di YouTube menjadi alternatif selain sekedar mengisi waktu luang atau menghasilkan uang. Untuk mendukung kamera diperlukan perangkat pendukung, dalam hal ini seiring berjalannya waktu kamera yang mumpuni juga bisa didapatkan dari perangkat smartphone, teknologi perangkat tersebut harus diimbangi dengan kemampuan pengguna dalam mengoperasikannya. Gimbal Stabilizer salah satu jawabannya, menggunakan gimbal stabilizer menjadi salah satu alternatif karena dapat meredam getaran sehingga gambar yang dihasilkan lebih maksimal. Banyak website memberikan informasi tentang produk gimbal stabilizer dengan memberikan banyak informasi dalam gambar dan ulasan pengguna. Oleh karena itu, analisis sentimen merupakan solusi untuk masalah pengelompokan opini atau review menjadi opini positif atau negatif secara otomatis berdasarkan hal ini untuk mendapatkan penilaian penggunaan gimbal berdasarkan analisis sentimen yang diberikan melalui review produk, kami akan mencoba menguji parameter untuk menghasilkan n gram pada tahap pre-processing, k-fold pada cross validation dan penerapan particle swarm optimization untuk meningkatkan akurasi menggunakan metode Naive Bayes. Hasil dari tester ini menghasilkan akurasi sebesar 84,42.

Kata Kunci: Klasifikasi; Stabilizer Gimbal; Naïve Bayes

Abstract

Becoming a videographer is a popular hobby during this pandemic because creating works in the form of videos and content on YouTube is an alternative to just filling your spare time or making money. To support the camera, the supporting deviceis needed, in this case, as time goes by, a capable camera can also be obtained from smartphone devices, the technology of the device must be balanced with the user's ability to operate it. Gimbal Stabilizer is one of the answers, using a gimbal stabilizer is an alternative because it can reduce vibrations so that the resulting image is maximized. Many websites provide information aboutgimbal stabilizer products by providing a lot of information in images and user reviews. Therefore, sentiment analysis is a solutionto the problem of grouping opinions or reviews into positive or negative opinions automatically based on this to get an assessmentof the use of gimbals based on the sentiment analysis provided through product reviews, we will try to test the parameters to produce n grams at the pre-processing stage, k-fold on cross validation and the application of particle swarm optimization to increase accuracy using the Naive Bayes method. The results of this tester produce an accuracy of 84.42

Keywords*: Classification; Gimbal Stabilizer; Naïve Bayes*

1. Introduction

The world of photography and videography grow rapidly, creativity that supported by technology can produce ex- traordinary works. A suitable device makes the idea we have can be realized into a satisfying work even generate a lot of income. The use of cameras is still a favorite thing for professionals to produce a good quality of images and videos. During the current pandemic, many ordinary people turn to be a content creator, just to fill their spare time orto earn some incomes.

To become a content creator, we do not have to use high-tech cameras, many of its use cell phones to get pictures or videos and make use of additional applications that provided by application

providers to be able to produce photo or video quality that is almost the same as those produced by pro cameras. The current development of smartphones has made this communication tool equipped by very adequate audio-visual features. The quality of smartphone cameras is getting more sophisticated, making the dependence on DSLR cameras are diverted, anyone can be a photographer,anyone can be a video maker just with a smartphone on hand. The problem in shooting using a smartphone is the vibration that generated from shooting a moving object. The wrong motion will cause vibrations in the image taken,even some people have a deficiency that is often called tremor which causes the hands to not move smoothly and always vibrates. With all the factors that cause the video vibrate or hard to focus and blur photo, a tool is needed likea gimbal stabilizer, the gimbal stabilizer is a device used to maintain the stability of the image taken by keeping a smartphone or camera stable.

In 2021, the number of Indonesian populations increase by 1.1% from the previous year into 274.9 million. 202.6 million are internet users. The increase of number internet users when compared to 2020 has significantly increased by 16% (Kemp, 2021). Currently, the world, including Indonesia, is being hit by the Covid-19 pandemic, which has an impact on changing transaction patterns in daily needs. The public has shown a positive response to the transition of buying and selling transactions through e-commerce. In Indonesia, E-commerce has been around since the 2000s, the use ofE-commerce was only in demand by the public in 2014 but can be seen through start-up companies in Indonesia, suchas Tokopedia, Bukalapak, Blibli, Shopee, and others (Permana et al., 2021). E-commerce provides review facilities for visitors. Reviewsare usually short descriptions of feedback about services or goods that have been purchased. This review can be used as a measuring tool for the seller, while for customers, reviews can be taken into consideration in making decisions tomake purchases and measuring the quality of goods and services that sold (Lutfi et al., 2018).

Sentiment analysis is widely used in the case of review analysis. Based on research by (Ganesan & Zhou, 2016), praise sentences usually show a positive subset consisting of adjectives, words that reinforce the meaning of other expressions and show emphasis, and others. Meanwhile, the complaint sentence shows a negative subset. In cyberspace, sentences of praise and complaints are conveyed in more complex forms. Therefore, it is necessary to have a sentiment analysis that canhelp solve problems in the review analysis (Prananda & Thalib, 2020). The machine learning approach builds sentiment classification using a selection feature selected with the help of labels. The selection features that are widely used include: Information Gain (IG), Document Frequency (DF), CHI Statistics, and Gain Ratio. Classification methods that are often used are Support Vector Machines (SVM), Naïve Bayes (NB), Decision Tree (DT), K-Nearest Network (K-NN), Artificial Neural Network (ANN), Random Forest, Linear Regression, Logistic Regression, and others (R & J, 2018). This study aims todetermine the accuracy of the sentiment analysis review of the gimbal stabilizer using the Naïve Bayes model and theselection of selection features designed for text categorization.

In 2017 (Nurfalah & Suryani, 2017) conducted a research using a Lexicon-based approach to see whether the comments on social media related to Pasti Pas of Pertamina services that is used in Bahasa Indonesia has negative or positive sentiments. This study produces an accuracy of 66%. Other research related to sentiment analysis was carried out to evaluate sales based on sales reviews by applying the Support Vector Machine algorithm. The accuracy that produced in this study was 93.65% 3. Sentiment analysis is also used to identify business intelligence analysis in GO-JEK using several classification algorithms such as Decision Tree, Naïve Bayes, Support Vector Machine and Neural Network. Based on the research results it can be concluded that the Decision Tree is the best algorithm 5. Research using the Particle Swarm Optimization selection feature on the Naïve Bayes algorithm has been carried out to analyze public sentimentregarding the presidential election in Indonesia. The results of the study (Hayatin et al., 2020) showed an increase in accuracy of 4.12%with the use of Particle Swarm Optimization so that an accuracy was obtained in 90.74%.

A. Text Mining

Text mining is the process of developing data then analyzed by the help of software, so that it can be identified the concepts, patterns, topics, keywords and other attributes contained in the data. According to (Deepa et al., 2013) data mining is a seriesof activities that used to find new hidden patterns or unexpected patterns in the previous data. Through the opinion that was found by (Sadiku, Matthew N. O; Shadare & M., 2015), data mining is a way to find meaningful patterns in large

amounts of data. On another occasion (Fayyad et al., 1999) argued that data mining is the application of a specific algorithm to extract patterns from data. Patterns that are generated from data mining can be used to predict new data based on these patterns. The pattern is represented in a structure that can be analyzed, can be easily understood and can be used in making decision (Witten et al., 2016).

B. Pre-Processing

Pre-Processing aims to prepare sentences before keyword extraction and sentiment determination 7. At this stage the data is filtered by removing irrelevant data, inconsistent data, and noisy data. The tasks that must be done at the pre-processing stage are removing special characters, URL, numbers and punctuation marks, removing stop words, stemming, and tokenization (Mittal & Patidar, 2019).

C. Naive bayes Classification

The Naive Bayes classifier algorithm is an algorithm that is used to find the highest probability value to classify thetest data in the most appropriate category (Feldman & Sanger, 2007). Naïve Bayes is one of the extensively studied classification methods for categorizing text. Usually, Naïve Bayes adopts the assumption that the value of a feature does not depend on the value of other features. Naïve Bayes assumes the probability that each word appearing in the document does not depend onthe appearance of other words in the same document (Deng et al., 2018). In this study, the test data is the buyer's comments. Thereare two stages to the classification of comments. The first stage is testing using training data against data with known categories. Then the second stage is the testing process with testing data.

Bayes method is an expert system method, Bayes method is useful for determining the probability value of expert hypotheses and the value of the evidence obtained from facts obtained from the object being diagnosed (Junaidi et al, 2020) .

D. Particle Swarm Optimization

Particle Swarm Optimization is a feature selection method to obtain optimization values (Desai, 2018). Its application in NaïveBayes can improve the accuracy results obtained in the Naïve Bayes method 8. The way PSO works by selecting the subset that produces the best accuracy data can be controlled using the Particle Swarm Optimization (PSO) algorithm. From the available dataset, PSO is used to find the maximum and best features.

E. N-Grams

N-grams is a contiguous n-item sequence of sorted text. Items can be phonemes, syllables, letters, words or base pairs (Devika et al., 2016).

F. Evaluation

Evaluation in sentiment analysis is the final result that is analyzed to make a decision whether we should choose it or not. The final result is shown in the form of bar graphs, pie charts and line charts (Mittal & Patidar, 2019).

2. Research Methods

The method that is used in this research consists of the stage of data collection, preliminary data processing, experimentation with the proposed method, testing methods, evaluation and validation of test results. Figure 1 shows the steps carried out in the proposed sentiment analysis.

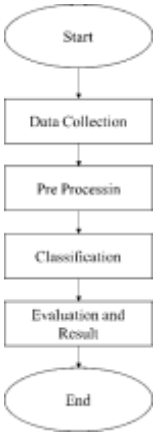

Figure 1. Research Stages

A. Data Collecting

The data that the author uses in this study is in the comments from users who buy gimbal stabilizers through the marketplace. The comments data are collected then divided into 2 categories; positive and negative comments. Comment collection is done by filtering out positive or negative comments and saving them in a .txt file.

B. Initial Data Processing

Initial data processing was carried out through the preprocessing stage. There are many preprocessing stages that can be used for research related to text mining. In this study, the

preprocessing stages were used, including:

a. Transform Cases/Case Folding

Transform cases or it can be called case folding is the process of homogenizing letters into uppercase or lowercase letters all in 1 comment sentence. The transform case process can be seen in the table below:

Table 1. Transform Cases Process

Input	Oiutput
Gimbal ini sangat saya suka	gimbal ini sangat saya suka
Ngebantu Banget buat Lebih Stabilkan HP P20 Saya	ngebantu banget buat lebih stabilkan hp p20 pro saya
Mode 2nya juga OK smua	mode 2nya juga ok semua

b. Tokenization

Tokenization is the process of breaking or cutting a sentence into several words. The result of this cutting word is called token. This process also eliminates punctuation, symbols, and other shapes that are not letters. The tokenization process can be seen in the table below:

Table 2. Tokenization Process

Input	Oiutput
gimbal ini sangat saya suka.	gimbal ini sangat saya suka
ngebantu banget buat lebih stabil kan hp p20 pro saya	ngebantu banget buat lebih stabil kan hp p20 pro saya
mode 2nya juga ok semua	mode 2nya juga ok semua

c. Filter Stopwords

The stopwords filter is the process of removing words that appear frequently but have no meaning. The set of stopwords used is taken from Tala's research.

Table 3. List Stopwords in Bahasa

ada	agaknya	agar	akan	ataukah
awal	bahwa	banget	berikut	bertanya
boleh	bukanlah	caranya	cukup	cuma
darip ada	dengan	demikian	enggak	entah

In addition to the stages mentioned above, this study also uses feature selection. Feature selection is the process of selecting the features that contribute the most to the desired results. Feature selection is applied to eliminate noisy, less informative, and repetitive features that can reduce space to be manageable. Feature selection can improve the efficiency and accuracy of the classification used [15]. In sentiment analysis, there are many feature selections are used, one of them is Generate n-Grams. Generate n-Grams can combine frequent adjectives to show a sentiment. This research uses unigram, bigram, trigram, and quadgram tokens. From the results of using this N-gram, the highest accuracy results will be taken.

C. Experiment with the proposed method

This study proposes the Naïve Bayes method with the additional of feature selection for the classification of com- ments using gimbal stabilizer. Testing methods in this study using Rapidminer to get accuracy values.

D. Evaluation and validation of test results

After the sentiment classification process, the results obtained are evaluated. At this stage, the calculation will be tested with accuracy, precision, and recall parameters. The evaluation model and accuracy are measured using the Area Under Cover (AUC) Curve. Accuracy (A) is the percentage level of conformity for document grouping. Calculating the accuracy value is done with the equation:

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \times 100\% \quad (1)$$

Precision is the percentage of relevance processing to the information sought. Precision is also defined as a True Positive classification and all data are predicted as positive. To calculate precision using the equation:

$$Precision = \frac{TP}{(TP+FP)} \times 100\% \quad (2)$$

Recall is the number of relevant documents in the collection that generated by the system. Fapat is also defined as the number of documents that have a True Positive classification of all documents that are truly positive (including False Negatives). Calculating the recall value using the equation:

$$Recall = \frac{TP}{(TP+FN)} \times 100\% \quad (3)$$

Confusion Matrix is used to get TP, FP, TN, and FN variables. The confusion matrix table can be seen below:

Table 4. Confusion Matrix

	Positive Prediction	Negative Prediction
True Positive	TP	FN
True Negative	FP	TN

After the data is collected, the data is divided into training and testing data. Data sharing is done by cross validation N-fold to eliminate word bias. Cross validation N-fold divides the document into n parts.

3. Result and Discussion

Sentiment analysis about the gimbal stabilizer based on sales reviews using the Naïve Bayes algorithm is carried out by entering the .txt dataset, where the dataset will go through the pre-processing stage, then validated with the cross-validation feature, and the results are seen in the apply model. The data used in this study comes from several marketplaces in Indonesia. This data consists of 200 reviews consisting of 100 positive reviews and 100 negative reviews.

The main process in the rapidminer application can be seen in the image below. The "Process Documents" operator is used to read data in a text file. Set Role operators are used so that labels are not counted in data categorization and the results obtained do not change. In the "Process Documents" operator there are steps for preprocessing which consist of transform cases operator, tokenize, filter stopwords using Bahasa Indonesia and Generate n-Grams. For classification, this study conducted two processes with Naïve Bayes without the help of optimization of the selection and process features using Particle Swarm Optimization represented by the "Optimize Weight" operator to optimize the results obtained by the selected algorithm.

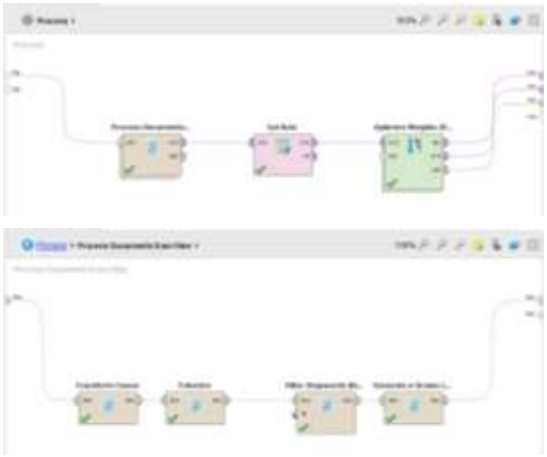

Figure 2. Pre-Processing and Process with Particle Swarm Optimization in Rapidminer

The following is the image of a test using the K-Fold Cross Validation method by applying the Naïve Bayes algorithm. Naïve Bayes Algorithm will class each comment to be positive and negative based on probability. A comment will be classified as positive if the probability value for the positive class is higher than negative. Vice versa. The operator "Cross Validation" is used for evaluating of sentiment analysis with experiment that is done ten times (k = 10).

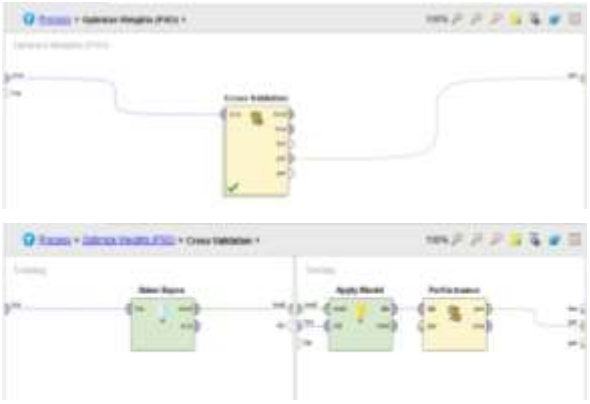

Figure 3. Use of Validation and Naïve Bayes on Rapidminer

The results of the implementation of the Naïve Bayes method on the Rapid Miner tool obtained an accuracy value of 71.79% with a precision of 68.75%, a recall of 81.67%, and the displayed AUC value of 0.515 while the implementation of the Naïve Bayes method and Particle Swarm Optimization obtained an

accuracy value of 84, 42%, Figure 4 shows the accuracy results of Naïve Bayes and Particle Swarm Optimization which show greater accuracy in rapidminer applications. From this process, an AUC value of 0.731 can be seen which can be seen in Figure 5 below:

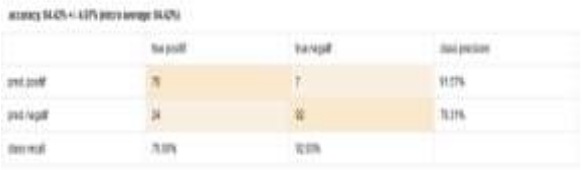

Figure 4. Accuracy Value of Naïve Bayes Algorithm and Particle Swarm Optimization

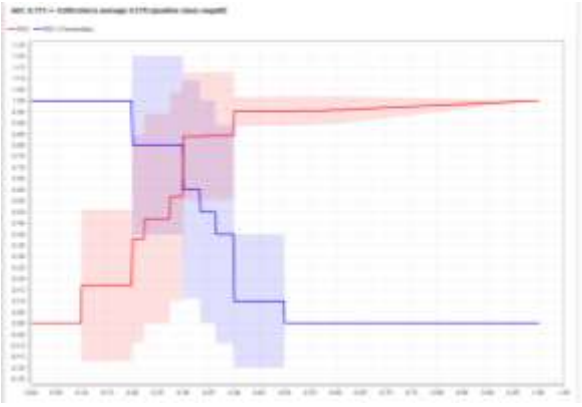

Figure 5. AUC of Naïve Bayes Algorithm and Particle Swarm Optimization

Testing using the Naïve Bayes method gets the confusion matrix value as follows:
True Positive (TP) = 76 record
True Negative (TN) = 92 record
False Positive (FP) = 7 record
False Negative (FN) = 24 record

4. Conclusion

In this research, sentiment analysis classification from comments on sales review of gimbal stabilizer was carried out. The implementation of the Naïve Bayes and Naive Bayes-PSO algorithms on gimbal reviews has been successfully done by using Rapidminer. From the results of testing and analysis, it is concluded that the classification of sentiment analysis can be done by the Naïve Bayes algorithm using 10-fold cross validation in the distribution of the dataset. The results show an accuracy of 71.79%, a precision of 68.75%, and a recall of 81.67%, while the result of the work of the Naive Bayes-PSO shows an accuracy of 84.42%. From the result of the experiment, it can be seen that the application of PSO can increase the accuracy by 12.63% and the AUC increases by 0.216. Thus, it is proved that the application of the Naive Bayes-PSO increases the accuracy value so that it can provide videographers with a more accurate decision making in selecting the gimbal stabilizer.

References

Deepa, V. K., Rexy, J., & Geetha, R. (2013). Rapid development of applications in data mining. *2013 International Conference on Green High Performance Computing, ICGHPC 2013*, 1–4. https://doi.org/10.1109/ICGHPC.2013.6533916

Deng, X., Li, Y., Weng, J., & Zhang, J. (2018). Feature Selection for Text Classification: A review. *Muiltimed Tools Appl*, 257–276. https://doi.org/10.1007/s11042-018-6083-5

Desai, R. D. (2018). Semantic Sentiment Analysis of Twitter Data. *Proceedings of the Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018) IEEE Xplore Compliant Part Number: CFP18K74-ART; ISBN:978-1-5386-2842-3 Sentiment*, *ICICCS*, 114–117.

Devika, M. D., Sunitha, C., & Ganesh, A. (2016). Sentiment Analysis: A Comparative Study on Different Approaches. *Procedia Computer Science*, *87*, 44–49. https://doi.org/10.1016/j.procs.2016.05.124

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1999). Knowledge discovery and data mining. *American Scientist*, *87*(1), 54–61. https://doi.org/10.1511/1999.16.807

Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook : Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.

Ganesan, K., & Zhou, G. (2016). Linguistic Understanding of Complaints and Praises in User Reviews. *Proceedings OfNAACL-HLT*, 109–114. https://doi.org/10.18653/v1/w16-0418

Hayatin, N., Marthasari, G. I., & Nuraini, L. (2020). Optimization of Sentiment Analysis for Indonesian Presidential Election using Naïve Bayes and Particle Swarm Optimization. *JOIN (Jurnal Online Informatika)*, *5*(1), 81–88. https://doi.org/10.15575/join.v5i1.558

Junaidi, A., Dewi, N., Baidawi, T., Agustiani, S.,

Arifin, Y. T., & Sihotang, H. T. (2020, November). Expert System Of Syzygium Aqueum Disease Diagnose Using Bayes Method. In *Journal of Physics: Conference Series* (Vol. 1641, No. 1, p. 012097). IOP Publishing.

Kemp, S. (2021). *Digital 2021: Indonesia*. https://datareportal.com/reports/digital-2021-indonesia

Lutfi, A. A., Permanasari, A. E., & Fauziati, S. (2018). Sentiment Analysis in the Sales Review of Indonesian Marketplace by Utilizing Support Vector Machine. *Journal of Information Systems Engineering and Business Intelligence*, *4*(2), 169. https://doi.org/10.20473/jisebi.4.2.169

Mittal, A., & Patidar, S. (2019). Sentiment analysis on twitter data: A survey. *ACM International Conference Proceeding Series*, 91–95. https://doi.org/10.1145/3348445.3348466

Nurfalah, A., & Suryani, A. A. (2017). Analisis Sentimen Berbahasa Indonesia dengan Pendekatan Lexicon-Based Pada Media Sosial. In *Jurnal Masyarakat Informatika Indonesia* (Vol. 2, Issue 1, pp. 1–8).

Permana, A. E., Reyhan, A. M., Rafli, H., & Rakhmawati, N. A. (2021). Analisa Transaksi Belanja Online Pada Masa Pandemi Covid-19. *Jurnal Teknoinfo*, *15*(1), 32. https://doi.org/10.33365/jti.v15i1.868

Prananda, A. R., & Thalib, I. (2020). Sentiment Analysis for Customer Review: Case Study of GO-JEK Expansion. *Journal of Information Systems Engineering and Business Intelligence*, *6*(1), 1. https://doi.org/10.20473/jisebi.6.1.1-8

R, P. K. M., & J, P. (2018). Role of sentiment classification in sentiment analysis : a survey. *Annals of Library and Information Studies*, *65*(September), 196–209.

Sadiku, Matthew N. O; Shadare, A. E. M., & M., S. (2015). A Brief Introduction to Data Mining. *European Scientific Journal*, *11*(21), 1–3.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical Machine Learning Tools and Techniques. In M. Kaufmann (Ed.), *Data Mining: Practical Machine Learning Tools and Techniques*. https://doi.org/10.1016/c2009-0-19715-5