Computer and Information Science.

INDONESIAN REVIEW RATING PREDICTION USING SUPPORT VECTOR MACHINES AND N-GRAM

M. Ali Fauzi
Faculty of Computer Science, Brawijaya University
Jalan Veteran 8, Malang, Indonesia

Abstract

Consumers tend to learn how others like or dislike a product by reading online reviews before deciding which product to be purchased. Some popular websites like Amazon and Female Daily make user able to give star-rating representing a quick overview of the product quality to know not only whether the product is good, but also how good the product is. However, many reviews are not equipped with a scale rating system, so it takes a long time to oppose different products in order to finally make a selection between them. Therefore, automatic product star-rating prediction from the text review is significantly required. In this study, Support Vector Machines were employed to classify the reviews rating with three N-Gram features combination to construct six different models. The experimental result shows that the system generally has poor performance because reviews having different but close ratings include very similar words with the best accuracy of only 46.1%. However, the wrong rating prediction results are mostly close to the actual rating. It can be seen from the accuracy results with 1 tolerance measurement which reach 82.8%. The experimental result also shows that combination of features has better performance compared to the individual features alone. The combination of unigram, bigram, and trigram produces the best performance with 46.1% accuracy on 0 tolerance and 82.8% on 1 tolerance. Meanwhile, bigram has the best performance among individual features with an accuracy of 45.9% in 0 tolerance and 78.6% in 1 tolerance.

Keywords: rating prediction, n-gram, support vector machine, female daily, multi-class classification

摘要 在决定购买哪种产品之前，消费者倾向于通过阅读在线评论来了解其他人喜欢或不喜欢产品的方式。一些流行的网站，如亚马逊和 Female Daily，使用户能够给出代表产品质量的快速概述的星级评定，不仅知道产品是否好，而且知道产品有多好。然而，许多评论没有配备规模评级系统，因此需要很长时间才能反对不同的产品，以便最终在它们之间进行选择。因此，非常需要从文本评论中自动进行产品星级预测。在这项研究中，支持向量机被用来对评论评级进行分类，用三个 N-Gram 特征组合来构建六个不同的模型。实验结果表明，该系统通常性能较差，因为具有不同但接近评级的评论具有非常相似的单词，最佳准确度仅为 46.1%。但是，错误的评级预测结果大多接近实际评级。从 1 个公差测量的精度结果可以看出，精度为 82.8%。实验结果还表明，与单独的特征相比，特征组合具有更好的性能。單字組，兩字 和卦 的组合产生最佳性能，0 容差为 46.1%，1 容差为 82.8%。同时，与其他单个功能相比，兩字 具有最佳性能，0 容差精度为 45.9%，1 容差精度为 78.6%。

I. INTRODUCTION

Nowadays, the number of user generated contents in the Internet has been increasing rapidly. Instead of only become information consumers, more and more people are becoming information producers [1]. Along with the trend, there is a lot of shopping websites allowing people to express their opinions, experiences, views, feeling, and complaints about their products in the form of text reviews [2]. This information plays a very important role in decision making for both producers and consumers. Before deciding which product to be purchased, consumers tend to learn how others like or dislike a product by reading online reviews about the product [3], [4]. Online reviews have become a more trustworthy reference for consumers [5]. For producer, online reviews can become valuable information to predict public acceptance of their products and can be the reference for product improvement and marketing strategies [6]. Therefore, analyzing sentiment from user generated content such as text reviews has become very important for both producers and consumers. Nevertheless, searching and comparing text reviews about some particular products to make a purchase decision can be very frustrating because of the sheer amount of text reviews available. Hence, the automatic way is preferred.

Sentiment analysis is a task of analyzing people's opinion or sentiment from a piece of text, product review text for instance. Generally, some previous studies about sentiment analysis such as [7], [8], and [9] only classify the review text into positive and negative category. This binary classification leaves a problem for consumers. To decide which product to buy, consumers not only require knowing whether the product is good, but also how good the product is [10]. Therefore, some popular website like Amazon and Female Daily make user able to give star-rating i.e. stars from 1-5 representing a quick overview of the product quality. This numerical rating can be very useful for consumers to rank products before deciding to buy. However, many product reviews (from other websites than Amazon and Female Daily) are not equipped by a scale rating system so it takes a long time to oppose different products in order to finally make a selection between them. Therefore, automatic product star-rating prediction from the text review is significantly required [11].

Nevertheless, there are some challenges in this task. Firstly, the text review and the star rating correlation is not clear [12]. In addition, users have diverse standards in giving rating for a product [13]. For example, two users with similar comment may give different star rating. Meanwhile, two users who rate product the same star may have very different comments. Therefore, constructing an effective model able to extract good features from the product text reviews and then measure their relationship to the rating became prominent challenge in this task.

In this study, we consider the review rating prediction task as a Machine Learning based multi-class classification problem where the star ratings are the category labels. We employ Support Vector Machines (SVM) as the classifier due to its satisfactory performance in text classification task [14], [15], [16]. Besides, we combine three N-Gram feature extraction methods, namely unigrams, bigrams, and trigrams to build six different models. These models will be trained and evaluated using reviews dataset from Female Daily website.

II. RESEARCH METHOD

The main flowchart of the review rating prediction system in this study is illustrated in Figure 1. There are four main stages in this system: a) preprocessing, b) extracting N-Gram features, c) term weighting using TF-IDF, and d) rating prediction using SVM. Each review will be classified into 1-5 star ratings.
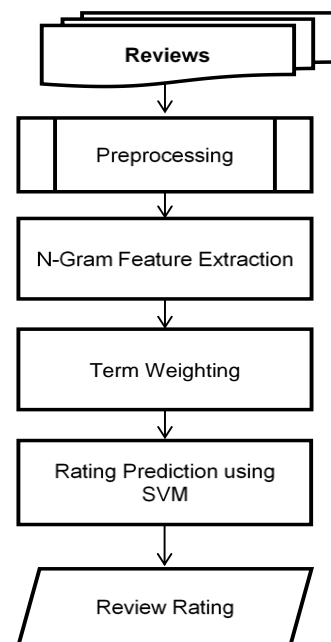

Figure 1.System Main Flowchart

A. Preprocessing

Text preprocessing is a text cleaning and preparation process before classification conducted. Online text usually contains a lot of noise and non-informative parts such as HTML tags, scripts, and advertisements [17]. Some steps conducted during preprocessing are tokenization, filtering, and stemming. After some characters that are recognized as punctuation were excluded, each review texts were split into tokens in the tokenization step. Next, some uninformative words, like "dan", "atau", "dari", and etc., were discarded in filtering step using stop list by Tala [18]. Finally, in stemming, each token or word was transformed into its root words using Sastrawi Stemmer.

B. N-Gram Feature Extraction

Generally, N-gram is a slice of N-character obtained from a string [19]. In this study, N-gram used is a slice of N-word taken from a sentence. The n can vary from 1 (unigram), 2 (bigram), 3 (trigram), 4, and so on [20]. Unigram, bigram, trigram and combination of them were employed in this study. For instance, a document contains a sentence "lipstick ini sangat bagus (the lipstick is very good)" can be extracted into N-Gram features like showed in Table 1.

Table 1.
N-Gram Feature Extraction Result

Features Extraction Method	Features List
Unigram	[lipstik, ini, sangat, bagus]
Bigram	[lipstik_ini, ini_sangat, sangat_bagus]
Trigram	[lipstik_ini_sangat, ini_sangat_bagus]
Unigram and Bigram Combination	[lipstik, ini, sangat, bagus, lipstik_ini, ini_sangat, sangat_bagus]
Combination of All	[lipstik, ini, sangat, bagus, lipstik_ini, ini_sangat, sangat_bagus, lipstik_ini_sangat, ini_sangat_bagus]

C. Term Weighting

Each word or term from the preprocessing results will be assigned a value based on its importance to the document. In this study, Term Frequency-Inverse Document Frequency (TF-IDF) is used as the term weighting method because its popularity and satisfactory performance in many text classification tasks.

TF-IDF is the multiplication of Term Frequency (TF) values and Inverse Document Frequency (IDF) of a term in a document. TF value of a term in a document is based on number of occurrences of the term in the document, while IDF value is inversely proportional to the number of documents containing the term. The TF-IDF value of term t in document d can be calculated as follows [21]:

$$TF - IDF(t,d) = (1 + \log tf(t,d)) * (\frac{N}{df_t}) \quad (1)$$

where $tf(t,d)$ is the number of occurrences of term t in the document d, N is the number of documents in corpus, and df_t is the number of documents containing term t.

D. Rating Prediction using Support Vector Machine

Rating prediction in this study was considered as classification problem. The ratings will be the classes and each reviews will be classified into one of these classes. The output of this step is the rating of each review.

Support Vector Machine (SVM) is employed to handle this classification problem due to its terrific empirical successes in many fields. SVM is a classifier that finds a hyperplane between two classes of data in a high dimensional space [22]. The hyper-plane accomplishes a good separation when it has the maximal margin or largest distance to the nearest data points of those two classes. Linear SVM should works well for text classification task because it can handle some characteristics of text including its high dimensional input space and vector sparsity. Moreover, most text classification problems are linearly separable [23].

III. RESULTS AND ANALYSIS

Dataset used for the experiment is text reviews taken from Female Daily website. All of the reviews are in the Indonesian language. The dataset contains 940 reviews that are evenly distributed into five rating categories i.e. 1, 2, 3, 4, and 5 star ratings. The performance of the classification system was evaluated using accuracy in five-fold cross validation scenario. Accuracy on test data was measured at two tolerances. Accuracy with 0 tolerance means the classification system result will be considered correct when the predicted star rating exactly equals the actual rating. Meanwhile, accuracy with 1 tolerance means the classification system result are still considered correct if the prediction rating is only one difference with the actual rating. For example, if the system classifies a document whose actual rating is 2 into star rating 2, then the prediction result will be considered correct by both measurements. If the system classifies a document whose actual rating is 2

into star rating 1 or 3, then the prediction result will be considered wrong by accuracy with 0 tolerance measurement. However, it will still be considered correct by accuracy with 1 tolerance measurement. Several values of N-Gram and its combination will be tested for this task. The

effect of preprocessing use will also be tested in this experiment. The experiment is conducted using Scikit-learn library [24]. Finally, the experimental result is shown in Table 2.

Table 2.
Experimental result

Features	Accuracy with 0 tolerance (%)		Accuracy with 1 tolerance (%)	
	Using Preprocessing	Without Using Preprocessing	Using Preprocessing	Without Using Preprocessing
Unigram	39.8	43.5	73.9	78.2
Bigram	45.9	45.1	78.0	78.6
Trigram	42.2	43.0	72.6	73.8
Unigram and Bigram Combination	45.0	45.7	79.9	81.5
Combination of All	**46.1**	45.5	81.2	**82.8**

Table 2 shows that the system tends to have poor performance with the best accuracy of only 46.1%. This is because reviews that have different but close ratings (e.g. reviews with ratings 1 and 2 or 4 and 5) have very similar words. Therefore, the classifier experiences difficulties in classifying the reviews correctly. For instance, some reviews with a rating of 5 have the words "enak banget (very satisfactory)". These words also appear a lot in reviews with rating 4. Another example, some reviews with ratings 1 and 2 also contain the same words, for instance "kurang cocok (less suitable)". However, in fact, the wrong rating prediction results are mostly having slight differences from the actual rating. This can be seen in the results of accuracy with 1 tolerance which has quite high accuracy with the best accuracy of 82.8%.

Table 2 also shows that the combination of features has better performance compared to individual features. The combination of unigram, bigram, and trigram produces the best performance with 46.1% accuracy on 0 tolerance and 82.8% on 1 tolerance. Meanwhile, bigram has the best performance compared to other individual features with an accuracy of 45.9% in 0 tolerance and 78.6% in 1 tolerance.

Unigram is the most favorite feature used in text classification. This feature is also often referred to Bag of Words. This unigram is quite simple to implement and in many cases, can provide high accuracy. However, unigram has a weakness in its inability to grasp the relationship between two words because it treats each word separately. Unigram is not able to capture phrases like "sangat bagus (very good)" in a 5-star review as a single feature. It will treat the phrase as two features, which are "sangat (very)" and "bagus (good)" instead. This phrase should be a differentiator with the word "bagus (good)" only

in a 4-star review. This drawback also makes unigram unable to catch negations like the word "tidak bagus (not good)". This word will also be broken down into two features namely "tidak (not)" and "bagus (good)" so that it will have similarities with the word "bagus (good)" only. Therefore, this weakness makes unigram unable to provide optimal performance with an accuracy of only 39.8% with preprocessing and 43.5% without preprocessing on 0 tolerance. Because the wrong rating predictions are actually not much different from the actual rating, the accuracy is still quite high if measured using 1 tolerance with an accuracy of 73.9% with preprocessing and 78.2% without preprocessing.

Bigram, on the other hand, is capable of capturing two-word phrases like "sangat bagus (very good)" and handle negations like "tidak bagus (not good)". These phrases often play an important role in rating classification. Therefore, bigram provided the best accuracy compared to other individual features. In the contrary, the last individual feature, trigram, have the lowest accuracy. Trigram is indeed able to capture three-word phrases like "sangat bagus sekali (very very good)". However, the same trigram would rarely appear across different reviews because every person has unique way of expressing their opinions and rarely uses three-word phrases. Therefore, the sparsity of the representation vector is very high and the system accuracy becomes not optimal.

Feature combinations make the features used get richer and are able to capture many important features. Therefore, the accuracy of this kind of features surpasses the accuracy of individual features. The combination of unigram, bigram, and trigram can capture more important feature than the combination of unigram and bigram only so that the three features combination has higher

accuracy than the combination of two. However, the different is the difference is very tight because trigram, which is the differentiator between the two combinations, only contributes a few important features.

The use of preprocessing also has some effects on the experimental results. The classification system tends to have higher accuracy when not using preprocessing. This is mainly because there is filtering step in the preprocessing. Filtering used in this study is based on stop list for Indonesian language by Tala [18]. The stop list contains several important words such as "tidak (no)", "sangat (very)", "kurang (less)", etc. This causes many important features to be lost. Therefore, the use of preprocessing declines the system accuracy.

IV. CONCLUSION

In this study, the review rating prediction task was regarded as a multi-class classification problem. Support Vector Machine was employed as the classifier with three N-Gram features combination to construct six different models. These models were trained and evaluated using Indonesian reviews dataset from Female Daily website. The experimental result shows that the system generally has poor performance with the best accuracy of only 46.1%. This is because reviews that have different but close ratings include very similar words. However, the wrong rating prediction results are mostly close to the actual rating. Therefore, the results of accuracy with 1 tolerance measurement have quite high accuracy with the best accuracy of 82.8%. The experimental result also shows that the combination of features has better performance compared to the individual features alone. The combination of unigram, bigram, and trigram produces the best performance with 46.1% accuracy on 0 tolerance and 82.8% on 1 tolerance. Meanwhile, bigram has the best performance compared to other individual features with an accuracy of 45.9% in 0 tolerance and 78.6% in 1 tolerance. Besides, the use of preprocessing tends to decline the system accuracy because of the filtering step that removes several important words.

Future studies should use feature selection because the features combination make feature dimensions even higher. In addition, a dedicated stop list for Indonesian reviews rating prediction is needed instead of using Tala stop list which is suitable actually for general classification task.

REFERENCES

[1] DANG, Y., ZHANG, Y. and CHEN, H. (2010) A lexicon-enhanced method for sentiment classification: An experiment on online product reviews. *IEEE Intelligent Systems*, *25*(4), pp.46-53. doi**:** 10.1109/MIS.2009.105

[2] BAILEY, A.A. (2004) Thiscompanysucks. com: the use of the Internet in negative consumer-to-consumer articulations. *Journal of Marketing Communications*, *10*(3), pp.169-182. doi: 10.1080/1352726042000186634

[3] HU, N., BOSE, I., KOH, N.S. and LIU, L. (2012) Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision Support Systems*, *52*(3), pp.674-684. doi: 10.1016/J.DSS.2011.11.002

[4] CUI, H., MITTAL, V. and DATAR, M. (2006) Comparative experiments on sentiment classification for online product reviews. Proceedings of the Twenty-First National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference, July 16-20, 2006, Boston, Massachusetts, USA. (*AAAI)*, 6, pp. 1265-1270. https://static.googleusercontent.com/media/research.google.com/ru//pubs/archive/4.pdf

[5] BICKART, B. and SCHINDLER, R.M. (2001) Internet forums as influential sources of consumer information. *Journal of Interactive Marketing*, *15*(3), pp.31-40. doi: 10.1002/DIR.1014

[6] BASUROY, S., CHATTERJEE, S. and RAVID, S.A. (2003) How critical are critical reviews? The box office effects of film critics, star power, and budgets. *Journal of Marketing*, *67*(4), pp.103-117. doi: 10.1509/jmkg.67.4.103.18692

[7] FAUZI, M.A., FIRMANSYAH, R.I.F.N. and AFIRIANTO, T. (2018) Improving Sentiment Analysis of Short Informal Indonesian Product Reviews using Synonym Based Feature Expansion. *Telkomnika*, *16*(3), pp. 1345-1350. doi: 10.12928/TELKOMNIKA.v16i3.7751

[8] FAUZI, M.A. (2018) Random Forest Approach for Sentiment Analysis in Indonesian Language. *Indonesian Journal of Electrical Engineering and Computer Science*, *12*(1), 46-50. doi: 10.11591/ijeecs.v12.i1.pp%25p

[9] GUNAWAN, F., FAUZI, M.A., ADIKARA, P.P. (2017) Analisis Sentimen Pada Ulasan Aplikasi Mobile Menggunakan Naive Bayes dan Normalisasi Kata Berbasis Levenshtein Distance (Studi Kasus Aplikasi BCA Mobile). *Systemic: Information System and Informatics Journal*, 3(2), pp.1-6. doi: 10.29080/Systemic.V3i2.234

[10] RATHI, M.S.M. and RAMTEKE, P.L., (2017) An Overview: Rating Prediction System From Textual Review. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 6*(1), pp. 73-76.

[11] BACCIANELLA, S., ESULI, A. and SEBASTIANI, F. (2009) Multi-facet rating of product reviews. In: Boughanem M., Berrut C., Mothe J., Soule-Dupuy C. (Eds.) Advances in Information Retrieval. ECIR Lecture Notes in Computer Science, vol 5478, (pp. 461-472). Springer, Berlin, Heidelberg. doi: 10.1007/978-3-642-00958-7_41

[12] ASGHAR, N. (2016) Yelp Dataset Challenge: Review Rating Prediction. https://arxiv.org/pdf/1605.05362.pdf

[13] LIU, D., CHAI, Y., ZHENG, C. and ZHANG, Y., (2017) Rating Prediction Based on TripAdvisor Reviews. https://pdfs.semanticscholar.org/51b7/a2cdfd350 b0d5d8e3fd71e2941d495fda2b8.pdf

[14] ROFIQOH, U., PERDANA, R.S., and FAUZI, MA. (2017) Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler Indonesia Pada Twitter Dengan Metode Support Vector Machine dan Lexicon Based Features. *Jurnal Pengembangan Teknologi Informasidan Ilmu Komputer, 1*(12), pp.1725-1732. e-ISSN: 2548-964X.

[15] PANG, B., LEE, L. and VAITHYANATHAN, S. (2002) Thumbs up?: sentiment classification using machine learning techniques. *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10* (pp. 79-86). Association for Computational Linguistics.

[16] MULLEN, T. and COLLIER, N. (2004) Sentiment analysis using support vector machines with diverse information sources. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*. http://www.aclweb.org/anthology/W04-3253

[17] HADDI, E., LIU, X. and SHI, Y. (2013) The role of text pre-processing in sentiment analysis. *Procedia Computer Science*, *17*, pp.26-32. doi: 10.1016/j.procs.2013.05.005

[18] TALA, F.Z. (2003) A study of stemming effects on information retrieval in Bahasa Indonesia. *Institute for Logic, Language and Computation, Universiteit van Amsterdam, The Netherlands*.

[19] CAVNAR, W.B. and TRENKLE, J.M. (1994) N-gram-based text categorization. *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval,* pp.161-175.

[20] PRASANTI, A.A., FAUZI, M.A. and FURQON, M.T. (2018) Neighbor Weighted K-Nearest Neighbor for Sambat Online Classification. *Indonesian Journal of Electrical Engineering and Computer Science*, *12*(1), pp. 150-160. doi: 10.11591/ijeecs.v12.i1.pp%25p

[21] MANNING, Ch.D., RAGHAVAN P., and SCHÜTZE, H. (2008) Introduction to information retrieval. *An Introduction To Information Retrieval*, *151*(177), p.5. Cambridge University Press.

[22] TONG, S. and KOLLER, D. (2001) Support vector machine active learning with applications to text classification. *Journal of machine learning research*, *2*(Nov), pp.45-66. doi: 10.1162/153244302760185243

[23] JOACHIMS, T. (1998) Text categorization with support vector machines: Learning with many relevant features. *Proceedings European Conference on Machine Learning* (pp. 137-142). Springer, Berlin, Heidelberg. doi: 10.1007/BFb0026683

[24] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A., MICHEL, V., THIRION, B., GRISEL, O., BLONDEL, M., PRETTENHOFER, P., WEISS, R., DUBOURG, V. and VANDERPLAS, J. (2011) Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*(Oct), pp.2825-2830.

引用

[1] Dang，Y.，Zhang，Y。和Chen，H。（2010）用于情感分类的词典增强方法：在线产品评论实验。IEEE智能係統，25（4），pp.46-53。doi：10.1109 / MIS.2009.105

[2] Bailey，A.A。（2004）這家公司很糟糕。：互联网在负面的消费者对消费者关系中的使用。營銷傳播雜誌，10（3），pp.169-182。doi：10.1080 / 1352726042000186634

[3] Hu，N.，Bose，I.，Koh，N.S。和Liu，L。（2012）操纵在线评论：对评级，可读性和情绪的分析。决策支持系统，52（3），pp.674-684。doi：10.1016 / J.DSS.2011.11.002

[4] Cui，H.，Mittal，V。和Datar，M。（2006）在线产品评论的情感分类比较实验。2006年7月16日至20日，美国马萨诸塞州波士顿举行的第二十一届全国人工智能会议和人工智能第十八次创新应用会议录。（AAAI），6，pp.1265-1270

https://static.googleusercontent.com/media/research.google.com/ru//pubs/archive/4.pdf

[5] Bickart，B。和Schindler，R.M。（2001）互联网论坛作为消费者信息的有影响力的来源。交互式营销杂志，15（3），pp.31-40。doi：10.1002 / DIR.1014

[6] Basuroy，S.，Chatterjee，S。和Ravid，S.A。（2003）关键评论有多重要？电影评论家，明星力量和预算的票房效应。市場營銷雜誌，67（4），pp.103-117。doi：10.1509 / jmkg.67.4.103.18692

[7] Fauzi，M.A.，Firmansyah，R.I.F.N。和Afirianto，T。（2018）使用基于同义词的特征扩展改善短期非正式印尼产品评论的情感分析。Telkomnika，16（3），pp.1345-1350。doi：10.12928 / TELKOMNIKA.v16i3.7751

[8] Fauzi，M.A。（2018）随机森林方法用于印尼语的情感分析。印度尼西亚电气工程与计算机科学杂志，12（1），46-50。doi：10.11591 / ijeecs.v12.i1.pp%25p

[9] Gunawan F，Fauzi MA，Adikara PP。（2017）Analisis Sentimen Pada Ulasan Aplikasi Mobile Menggunakan Naive Bayes dan Normalisasi Kata Berbasis Levenshtein Distance（Studi Kasus Aplikasi BCA Mobile）。系统性：信息系统与信息学期刊，3（2），pp.1-6。doi：10.29080 / Systemic.V3i2.234

[10] Rathi，M.S.M。和Ramteke，P.L.，（2017）概述：评级预测系统从文本审查。国际计算机工程与技术高级研究期刊（IJARCET），6（1），pp.73-76。

[11] Baccianella，S.，Esuli，A。和Sebastiani，F。（2009）产品评论的多方面评级。In：Boughanem M.，Berrut C.，Mothe J.，Soule-Dupuy C.（Eds。）信息檢索的進展。计算机科学讲义，第5478卷，（第461-472页）。斯普林格，柏林，海德堡。doi：10.1007 / 978-3-642-00958-7_41

[12] Asghar，N。（2016）Yelp数据集挑战：回顾评级预测。https://arxiv.org/pdf/1605.05362.pdf

[13] Liu，D.，Chai，Y.，Zheng，C。和Zhang，Y.，（2017）评级预测基于TripAdvisor的评论。https://pdfs.semanticscholar.org/51b7/a2cdfd350b0d5d8e3fd71e2941d495fda2b8.pdf

[14] Rofiqoh U，Perdana RS，Fauzi MA。（2017）Analisis Sentimen Tingkat Kepuasan Pengguna Penyedia Layanan Telekomunikasi Seluler印度尼西亚Pada Twitter DenganMetode支持向量机 dan Lexicon Based Features。Jurnal Pengembangan Teknologi Informasidan Ilmu Komputer，1（12），pp.1725-1732。e-ISSN：2548-964X。

[15] Pang，B.，Lee，L。和Vaithyanathan，S。（2002）竖起大拇指?:使用机器学习技术的情感分类。关于自然语言处理中经验方法的ACL-02会议论文集 - 第10卷（第79-86页）。计算语言学协会。

[16] Mullen，T。和Collier，N。（2004）使用具有不同信息源的支持向量机的情感分析。2004年自然语言处理经验方法会议记录。http://www.aclweb.org/anthology/W04-3253

[17] Haddi，E.，Liu，X。和Shi，Y。（2013）文本预处理在情绪分析中的作用。程序计算机科學，17，pp.26-32。doi：10.1016 / j.procs.2013.05.005

[18] Tala，F.Z。（2003）印度尼西亚语对信息检索的干扰效应研究。荷兰阿姆斯特丹大学逻辑，语言和计算研究所。

[19] Cavnar，W.B。和Trenkle，J.M。（1994）基于N-gram的文本分类。SDAIR-94会议记录，第三届文献分析与信息检索年度研讨会，第161-175页。

[20] Prasanti，A.A.，Fauzi，M.A。和Furqon，M.T。（2018）Sambat在线分类的邻居加权K-最近邻。印度尼西亚电气工程与计算机科学杂志，12（1），pp.150-160。doi：10.11591 / ijeecs.v12.i1.pp%25p

[21] Manning，Ch.D.，Raghavan P.和Schütze，H。（2008）信息检索简介。信息检索导论，151（177），第5页。剑桥大学出版社。

[22] Tong，S。和Koller，D。（2001）支持向量机主动学习，应用于文本分类