Effects of kernels and the proportion of training data on the accuracy of SVM sentiment analysis in lecturer evaluation

Daniel Febrian Sengkey[1], Agustinus Jacobus[2], Fabian Johanes Manoppo[3]
[1,2]Department of Electrical Engineering, Faculty of Engineering, Universitas Sam Ratulangi, Manado, Indonesia
[3]Department of Civil Engineering, Faculty of Engineering, Universitas Sam Ratulangi, Manado, Indonesia

Article Info

ABSTRACT

Support vector machine (SVM) is a known method for supervised learning in sentiment analysis and there are many studies about the use of SVM in classifying the sentiments in lecturer evaluation. SVM has various parameters that can be tuned and kernels that can be chosen to improve the classifier accuracy. However, not all options have been explored. Therefore, in this study we compared the four SVM kernels: radial, linear, polynomial, and sigmoid, to discover how each kernel influences the accuracy of the classifier. To make a proper assessment, we used our labeled dataset of students' evaluations toward the lecturer. The dataset was split, one for training the classifier, and another one for testing the model. As an addition, we also used several different ratios of the training:testing dataset. The split ratios are 0.5 to 0.95, with the increment factor of 0.05. The dataset was split randomly, hence the splitting-training-testing processes were repeated 1,000 times for each kernel and splitting ratio. Therefore, at the end of the experiment, we got 40,000 accuracy data. Later, we applied statistical methods to see whether the differences are significant. Based on the statistical test, we found that in this particular case, the linear kernel significantly has higher accuracy compared to the other kernels. However, there is a tradeoff, where the results are getting more varied with a higher proportion of data used for training.

Corresponding Author:

Daniel Febrian Sengkey
Department of Electrical Engineering
Faculty of Engineering, Universitas Sam Ratulangi
Jl. Kampus UNSRAT, Bahu-Manado, 95115, Indonesia
Email: danielsengkey@unsrat.ac.id

1. INTRODUCTION

Since introduced in 1995 [1], support vector machine (SVM) has become a popular supervised method in machine learning (ML). The rise of ML as a potential tool for the future [2], leads to the emerging use of SVM. For example, SVM regression was used in telecommunication to estimate channel in multiple input multiple output (MIMO) orthogonal frequency division multiplexing (OFDM) by using the interpolation mechanism [3]. The study in [4] employed SVM, to extract buildings object from very high-resolution satellite images of Tetuan, Morocco, by using spatial and spectral radius with 83.76% accuracy. Other utilization of SVM in image processing could also be found in [5], where the multi-class SVM is used to classify three types of rice grain with 92.22% accuracy. SVM was used in [6], to forecast wind speed, based on the direction, the former data (historical), pressure, moisture, and heat, although it is bested by the other methods in the same study. Recently, SVM was also considered in a study of flood risk prediction [7].

In spite of its accuracy issue in a some scopes of applications, as found in [6], SVM is still a preferable method in sentiment and/or emotion analysis [8]. To improve its accuracy, there are studies about the kernel tricks. For instance, Amari and Wu [9], found that modifying the gaussian radial basis function (RBF) kernel brought remarkable improvement. In another study, the kernels are compared to find the best one for emotion recognition based on the processing of physiological signals and it is found that the linear kernel gave the highest accuracy rate [10]. On the other hand, the work of Asraf and Shah Rizam [11], compared the SVM kernels to classify nutrient disease in oil palm leaves showed that the polynomial kernel performed best compared with the other kernels. Similarly, the kernel comparison for termite detection also found the polynomial kernel as the best one [12]. Shantini *et al* in [13], compared SVM kernels for fault classification in the analog filter circuit. Therefore, based on the former studies [9-13], each kernel may bring different impacts on the accuracy of the classifier.

Educational sentiment analysis is a field of SVM application. This field commonly addresses the students' evaluations of the lecturer [14-19]. In one of our former work [14], we have applied support vector machine (SVM) to classify the sentiments in lecturer evaluation by students. In [15], Altrabsheh et al proposed the sentiment analysis for education (SA-E), which is a system that will collect students' feedback by using the social network and then processed them by using SVM and/or naive bayes. The results in [16], showed that when used to classify sentiments within this particular field, SVM came as the best method. Similarly, the study in [17] also proposed the use of Twitter for their data collection and in sentiment analysis, with SVM to classify the feedbacks, for the improvement of the teaching-learning process. Later, this study is extended in [18]. In the Indonesia context, another study on student feedbacks toward the lecturer was reported in [19].

There are pieces of evidence which found that different kernel may bring different effects on the accuracy of the classifier, yet the study on SVM kernels for sentiment analysis in education is rare. Therefore, in this study we compare the accuracy rates yielded by various SVM kernels. In the comparison, we also include several different ratios of the training:testing data so we may identify the impact of the kernels in different sizes of training data. The rest of this paper is organized as follows: In Section 2 we present the methodologies used in this study, such as the data source, and the steps/stages in the research. Then we present the results in Section 3, and finally in Section 4 we conclude this paper and present some plausible extensions to this study.

2. RESEARCH METHOD

The course of this study is presented in Figure 1. There are 4 major phases, which are data collection, pre-processing, processing, and post-processing. Generally, it can be seen as a mixture of the steps used in [18, 20-21] with the exceptions on the data collection and post-processing phases, where the works in [18] and [20] used crawled data from Twitter, while the work in [21] used crawled data from news portals. The final phase, post-processing, refers to a set of statistical tests where the accuracy data are compared.
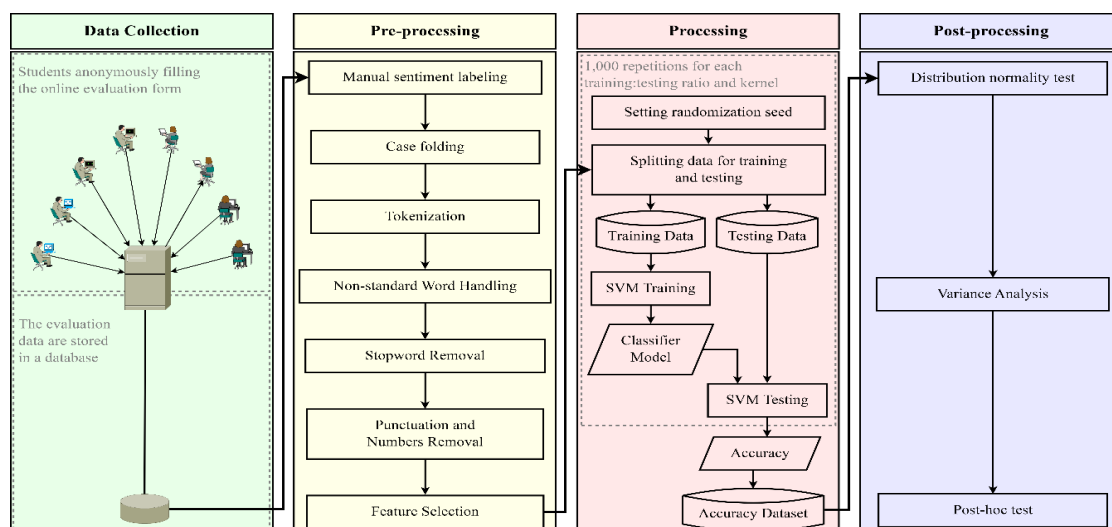

Figure 1. Course of research

2.1. Data collection

The data for this study is acquired from the lecturer evaluation/student feedbacks of the spring semester 2019. The same data was used in [14]. In May to mid-June, students were asked to fill an anonymous online evaluation form for each lecturer they had during the semester that was about to end. It means students might have made several submissions. As presented in Figure 1, these feedbacks are stored in a database. Collected data consists of 636 feedbacks with 430, 109, and 97 of positive, negative, and neutral sentiments, respectively. It is found that the students used several languages when giving their feedback. Mostly use standard Bahasa Indonesia, some used English, and some used Bahasa Indonesia mixed with local diction.

2.2. Pre-processing

The pre-processing phase, along with the processing phase discussed in the next subsection are the common phases in natural language processing (NLP). In this study, the pre-processing phase was done in several steps: First, each feedback was manually labeled. In this step, the collected feedbacks, or usually referred to as documents in the field of sentiment analysis, were manually analyzed and labeled according to its sentiment. From the 636 collected documents, there are 430, 109, and 97 of positive, negative, and neutral sentiments, respectively. The percentages of the sentiments are shown in Figure 2.
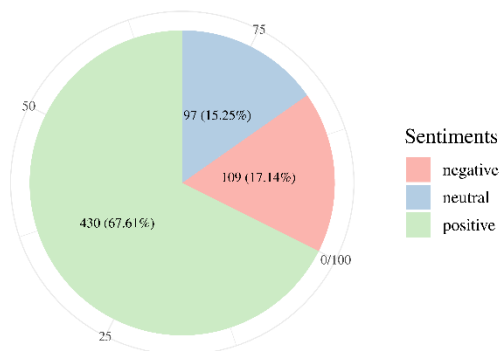

Figure 2. Sentiments composition in the dataset

After all, feedbacks labeled, the letter case in all documents is folded to small caps. Later, each document was tokenized, which in this case each token is a word. The non-standard words (NSW) handling is a step that closely related to the carelessly abbreviated and/or mistyped words, which is found as a characteristic of people in Indonesia [22]. In this step, a customized dictionary containing pairs of non-standard: standard word is used. When the NSWs were handled, then the stopwords were removed. As described earlier in Subsection 2.1, since the students used various languages, then several stopword dictionaries were used in this step:
1.   The English stopwords from the Tidytext packet for R [23];
2.   The Bahasa Indonesia stopwords dictionary, as available on-line in [24-25]; and
3.   We built our own list of local stopword dictionary, based on the words used by the students.

After the documents cleared from stopwords, all numbers and punctuation are removed. The final step in this phase is the Feature selection, which in this study it is based on the word frequency. All words that appeared less than 10 times were deemed as insignificant and so removed.

Although stemming is commonly used in the pre-processing phase, in this study this step was omitted, as was done in [14]. The reason behind this is that stemming has been identified to bring no significant impact to the accuracy when applied to documents with Bahasa Indonesia [26].

The steps within this phase brought a great impact on the data. First, as shown in Figure 3(a), the stopwords were frequently used. The rank of the most frequent words then drastically changed after the nonstandard word handling and stopword removal, as shown in Figure 3(b). This change also reflected in the number of words, as shown in Figure 4. Prior to the stopword removal, there were 8,446 and after that, only 1,049 words are left. It means, there were 7,397 stopword and non-standard words in the collected data that had been removed, or approximately 87.58%.
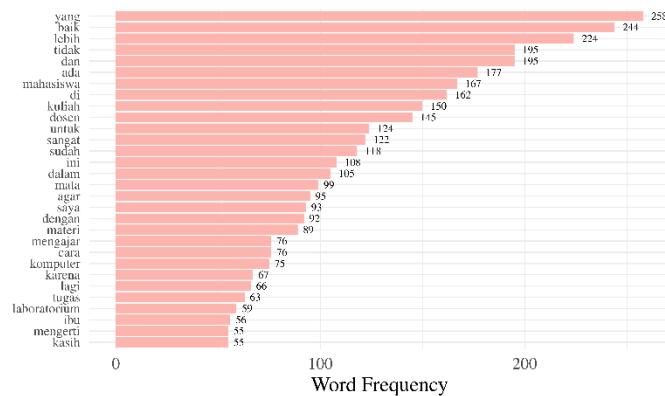
(a) Post tokenization
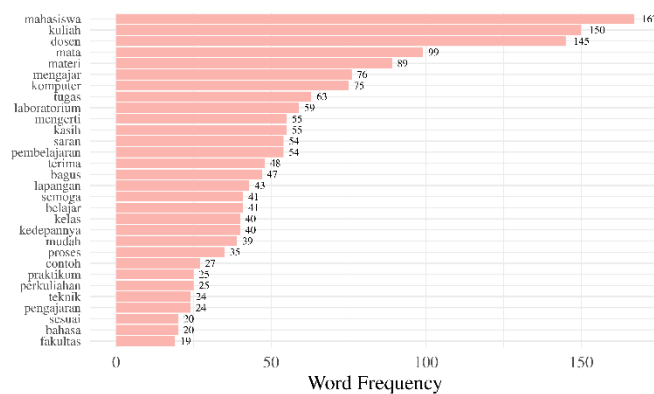
(b) Post stopword removal

Figure 3. The comparison of the 30 most frequent words after tokenization/before NSW handling, and after stopword removal
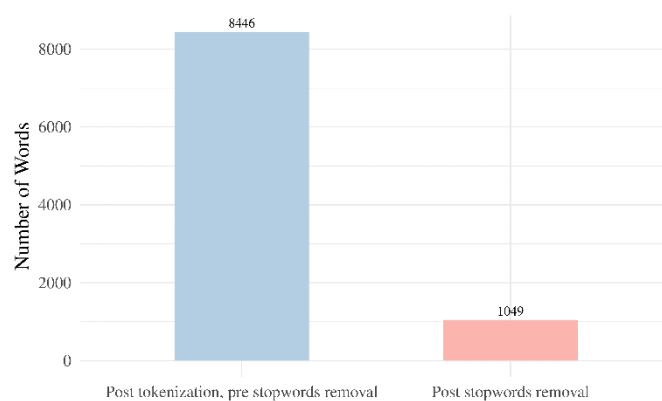
Figure 4. Number of words before and after stopword removal

2.3. Processing

As the objective of this study is to make comparisons between the SVM kernels and also to compare the effect of the training:testing dataset ratio, then a major part of this phase is designed to be done several times, to achieve a large number of samples for each parameter tested so that the law of large number could be satisfied [27]. Therefore, for each pair of the kernel and training:testing ratio, the following processes were repeated for 1000 times:
1. Setting up the randomization seed with the iterator, so for all i-th repetition in all pairs of kernel training:testing ratio, the training datasets were always identical, and so with the testing dataset.

2. Split the dataset for training and testing, with the percentage of training dataset started from 50% toward 95%, uniformly increased by 5%, or in other words 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, parts from the collected data. These data were randomly split by using the createDataPartition() function, available in the caret package for GNU R [28]. Later, each dataset was transformed into a document term matrix (DTM) by using the tm package for GNU R [29].

3. The training dataset then used for training the SVM classifier by using the e1071 package for R [30]. This step produced an SVM classifier (model), based on the particular training dataset.

4. To test the accuracy of the model, it was used with the testing dataset. The model classified each document in the testing dataset into one of the sentiments. After that, the output of the model were compared with the real sentiment label of the documents. This process then yielded the accuracy of the model in percent.

After all repetitions for each kernel training:testing ratio finished, the accuracy data are collected in a dataset. As there were 4 kernels, 10 training:testing ratios and 1,000 repetitions, then this dataset contains 40,000 rows of data.

2.4. Post-processing

The term post-processing was chosen to represent the analysis phase done after the SVM processing. Mainly, this phase covers the statistical procedures used to make the comparison between tested parameters, which are kernels and ratio between training and testing datasets. The main step within this phase is the Variance Analysis step by using the multivariate test. In this step, the data were compared to find any significant variance due to the effect of the parameters. However, to decide which multivariate method to be used, the distribution normality must be identified first. If the variance analysis indicates any significant variance, then this continued to the post-hoc test. The results within this phase are comprehensively discussed in Section 3. For all statistical tests, $\alpha = 0.05$ is adopted. The GNU R statistical software [31] was used to conduct these tests.

3. RESULTS AND DISCUSSION

To provide a general overview, we plotted the accuracy of each experiment as a scatter plot, grouped by the proportion (ratio) of the training data, as shown in Figure 5. For each kernel type, there is a line that connects the mean of the accuracy of each particular training ratio. It is clear that in every ratio of the training data to the testing data, the linear kernel has higher average accuracy. However, there are several points of the linear kernel accuracy that located too far, although this is a common phenomenon in all kernel types when the ratio of training data increased.
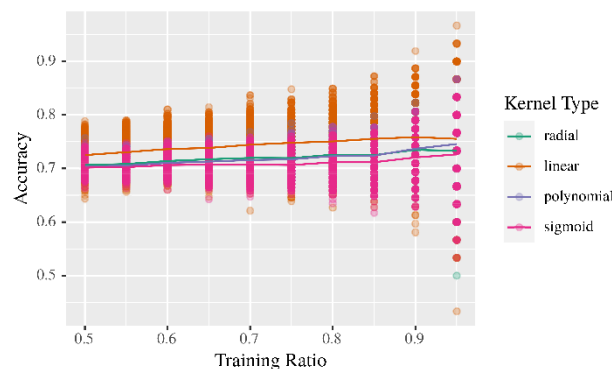
Figure 5. The accuracy of each experiment. Each line represents and connects the mean of the accuracy, for each ratio of the training data

The boxplots in Figure 6 confirm that the linear kernel tends to yield higher accuracy, judged by the vertical positions of the box in Figure 6(a). When the boxplots of the accuracy grouped by the kernel, as shown in Figure 6(b), the indication of the higher variation due to the higher ratio of the training data can be visually inspected clearer. Although the linear kernel tends to have higher variation, yet the shape of the box and the location of the medians are commonly balanced, in which the median line located around the center, meanwhile for the other kernel, the median sometimes located near one of the quartiles (i.e polynomial kernel

at training ratio 0.8 and 0.95, while at training ratio 0.85 the median is closer to the first quartile; radial kernel at training ration 0.6, 0.8, 0.85 and 0.9).

(a) Grouped by the ratio of training data
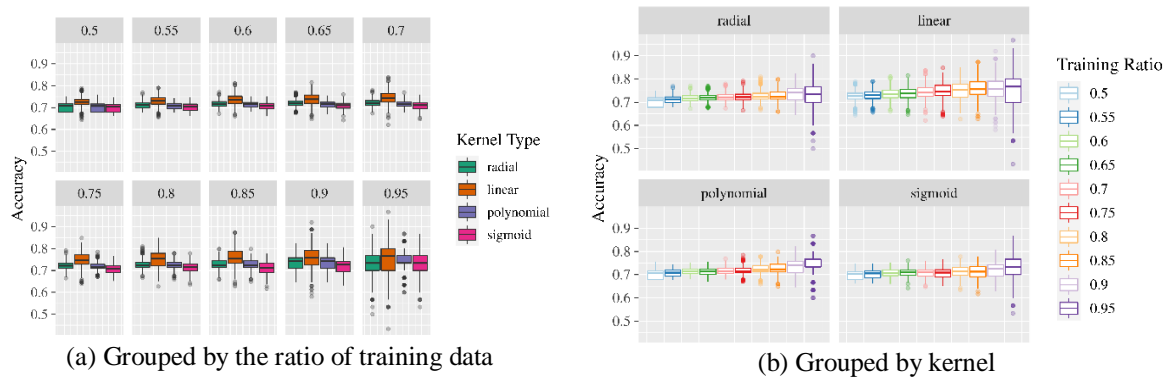
(b) Grouped by kernel

Figure 6. Boxplot of the accuracy from each experiment.

Figure 7 gives a visual comparison of the standard deviation, where Figure 7(a) and Figure 7(b) show grouped comparisons by the ratio of training data and by kernel, respectively. It is very clear that the accuracy of the linear kernel always deviates higher than the other kernels, followed by the radial kernel. The standard deviation of the polynomial and the sigmoid kernels tend to coincide in lower training ratio. However, with the higher proportion of the data used for training, the polynomial kernels deviate lower than the sigmoid kernel, as shown in Figure 7(b).

(a) Grouped by the ratio of training data

(b) Grouped by kernel

Figure 7. Boxplot of the accuracy from each experiment

3.1. Distribution normality

The distribution normality test was done to each kernel training:testing ratio by using the Shapiro-Wilk method [32]. The results, in p-value, are presented in Table 1. As described in Subsection 2.3., the threshold for the p-value is 0.05, hence the H_0 in most cases was rejected. As not all data normally distributed, then the subsequent tests must use the non-parametric methods.

Table 1. Results from the Shapiro-Wilk test

Kernel Type	P-value for each ratio of the training data				
	0.5	0.55	0.6	0.65	0.7
radial	38.01×10^{-27}	1.092×10^{-24}	794.4×10^{-24}	1.497×10^{-18}	15.33×10^{-15}
linear	1.685×10^{-3}	6.788×10^{-3}	45.06×10^{-3}	3.802×10^{-3}	2.429×10^{-3}
polynomial	326.1×10^{-27}	5.852×10^{-24}	1.753×10^{-21}	58.37×10^{-18}	30.10×10^{-15}
sigmoid	452.4×10^{-24}	17.29×10^{-21}	128.4×10^{-18}	542.5×10^{-15}	20.28×10^{-9}
	0.75	0.8	0.85	0.9	0.95
radial	11.83×10^{-12}	477.1×10^{-12}	140.5×10^{-12}	7.598×10^{-12}	120.5×10^{-18}
linear	56.04×10^{-6}	451.0×10^{-6}	158.0×10^{-6}	6.702×10^{-6}	27.98×10^{-12}
polynomial	843.2×10^{-15}	157.0×10^{-12}	79.26×10^{-12}	589.4×10^{-15}	2.831×10^{-18}
sigmoid	13.29×10^{-9}	43.61×10^{-9}	323.7×10^{-12}	6.470×10^{-12}	5.778×10^{-18}

3.2. Multivariate analysis

To gain knowledge regarding the effects of the kernel and the effects of training data ratio, each scenario was tested separately. First, we compared the ratios of training data in the same kernel; and second, we compared the kernels in each ratio of the training data. As the data are not distributed normally, then we applied the two-tail Kruskal-Wallis test in this step [33].

3.2.1. The ratio of the training data

Based on the p-values in Table 2, the results are significant since all values are less than 0.05. The p-value for the polynomial kernel is not actually equal to 0. However, since the value is too small, it is way beyond the limits of the computer to display. Therefore, for this step, it can be said that in every kernel tested, at least one training ratio gave significantly different accuracy.

Table 2. Results of the Kruskal-Wallis test for the ratio of training data comparison in each kernel

Kernel Type	p-value
radial	585.4×10^{-237}
linear	567.5×10^{-189}
polynomial	< 0.0001
sigmoid	8.806×10^{-135}

3.2.2. Kernels

Similar to the comparison of the ratios of training data in the same kernel, when the kernels in each ratio put together to be compared, the differences are also significant, as shown in Table 3. Hence, for each proportion of the dataset used for training, at least one kernel had significantly different accuracy.

Table 3. Results of the Kruskal-Wallis test for the kernel comparison of each ratio of training data

Ratio of Training Data	p-value
0.5	415.0×10^{-162}
0.5	8.39×10^{-198}
0.6	1.22×10^{-204}
0.65	91.80×10^{-198}
0.7	648.0×10^{-228}
0.75	1.32×10^{-240}
0.8	1.20×10^{-174}
0.85	916.0×10^{-174}
0.9	17.30×10^{-93}
0.95	143.0×10^{-36}

3.3. Post-hoc analysis

The post-hoc analysis was done to compare pairs of experiments in the same group. Based on the pairwise tests, it possible to find the pairs that significantly different. The method used in this step was a two-tail Dunn test with the Benjamini-Hochberg adjustment. Since there are too many combinations tested, then to save space, then in only the results with insignificant differences (p-value > 0.05) are shown in Table 4.

Table 4. Selected results from the Dunn test for the impact of the ratio of the training dataset, on the accuracy for every kernel type

Training Ratio	Comparison	Z	P.unadj	P.adj
0.80	polynomial - radial	-1.89	0.06	0.06
0.85	polynomial - radial	-0.75	0.45	0.45
0.90	polynomial - radial	1.61	0.11	0.11

3.3.1. Kernel

The results of the Dunn test for the comparison of the kernels are shown in Table 5. Based on the adjusted p-values, only the polynomial and radial kernels that have several insignificant differences. Therefore, it can be said that most kernels give significantly different accuracy. Moreover, the cases where the differences are not significant, happened when the ratios are in very close range.

Table 5. Selected results from the Dunn test for comparison in each kernel the kernel impact on the accuracy, for every ratio of the training dataset

Kernel Type	Comparison	Z	P.unadj	P.adj
radial	0.5 - 0.55	-1.24	0.22	0.23
radial	0.65 - 0.75	-1.57	0.12	0.13
radial	0.7 - 0.75	0.97	0.33	0.34
radial	0.8 - 0.85	0.53	0.59	0.59
linear	0.6 - 0.65	-1.51	0.13	0.14
linear	0.75 - 0.8	-1.00	0.32	0.33
linear	0.85 - 0.9	-0.49	0.62	0.62
linear	0.75 - 0.95	-1.58	0.11	0.13
linear	0.8 - 0.95	-0.58	0.56	0.57
linear	0.85 - 0.95	1.66	0.10	0.11
polynomial	0.5 - 0.55	-1.23	0.22	0.23
polynomial	0.6 - 0.65	-1.98	0.05	0.05
polynomial	0.7 - 0.75	-1.04	0.30	0.31
polynomial	0.8 - 0.85	-0.27	0.79	0.79
polynomial	0.9 - 0.95	-1.64	0.10	0.11
sigmoid	0.5 - 0.55	-0.06	0.95	0.97
sigmoid	0.6 - 0.65	-0.86	0.39	0.44
sigmoid	0.6 - 0.7	-0.85	0.40	0.44
sigmoid	0.65 - 0.7	0.02	0.99	0.99
sigmoid	0.6 - 0.75	0.35	0.72	0.78
sigmoid	0.65 - 0.75	1.22	0.22	0.27
sigmoid	0.7 - 0.75	1.20	0.23	0.27
sigmoid	0.8 - 0.85	-0.25	0.80	0.84
sigmoid	0.9 - 0.95	-1.73	0.08	0.10

3.3.2. The ratio of the training data

In total, for each kernel type, there are 45 comparisons of the training ratio which when accumulated, the result of the Dunn test consists of 186 rows. Therefore, only the comparisons with insignificant differences are shown in Table 3. The table confirms that when the ratio too close together, such as 0.5 and 0.55, then the results will not be significantly different. The sigmoid and then the linear kernel are the 2 least affected by the proportion of the data used for training. Contrarily, the radial kernel, in this case, is the most affected by the ratio, since only 4 of 45 experiments that are not significantly different.

4. CONCLUSION AND FUTURE WORK

The sentiment analysis with the SVM classifier has been studied as a solution for the tiresome works of manual sentiment classification. The student evaluation toward the lecturer is a field where it has been applied. SVM has several parameters that can be tuned to increase its accuracy, such as using different kernels. Yet, in sentiment analysis of the students' evaluation, this option is rarely explored. In this study, we

have specifically evaluated the use of various kernels and the proportions of data used for training to the accuracy of the classification. Similar to the findings of the previous studies about SVM kernels, this study also found that the utilization of different kernels affects the accuracy of the SVM classifier, which in this case specifically trained for the sentiment analysis of the lecturer feedbacks from students. Only the polynomial and radial kernels that at a certain proportion of the dataset used for training ratio have no significantly different impact the accuracy. On the other hand, we also studied the impact of the dataset proportion used for SVM training. Based on the boxplot, a higher proportion of data used for training gives higher variation to the accuracy of the model. It means the model would be unreliable when the proportion of the training dataset increased. However, in the post-hoc test, the differences are not always significant. The sigmoid and then the linear are the 2 kernels with the less significant variance between close-range training ratio. This study already evaluated the use of kernels and the proportion of datasets used for training in SVM-based sentiment analysis of the students' feedbacks toward the lecturer. Future work must address the issue of the better kernel, as well as the right proportion of the training dataset, that should be used in this particular case.

REFERENCES

[1] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/bf00994018.
[2] V. Siddesh Padala, K. Gandhi, and D. V. Pushpalatha, "Machine learning: The new language for applications," *IAES Int. J. Artif. Intell.*, vol. 8, no. 4, pp. 411–421, Dec. 2019, doi: 10.11591/ijai.v8.i4.pp411-421.
[3] A. Charrada, "Support Vector Machines Regression for MIMO-OFDM Channel Estimation," *IAES Int. J. Artif. Intell.*, vol. 1, no. 4, pp. 214-224, Dec. 2012, doi: 10.11591/ij-ai.v1i4.1832.
[4] O. Benarchid and N. Raissouni, "Support Vector Machines for Object Based Building Extraction in Suburban Area using Very High Resolution Satellite Images, a Case Study: Tetuan, Morocco," *IAES Int. J. Artif. Intell.*, vol. 2, no. 1, pp. 43–50, Mar. 2013, doi: 10.11591/ij-ai.v2i1.1781.
[5] S. Ibrahim, N. A. Zulkifli, N. Sabri, A. A. Shari, and M. R. M. Noordin, "Rice grain classification using multi-class support vector machine (SVM)," *IAES Int. J. Artif. Intell.*, vol. 8, no. 3, pp. 215–220, Dec. 2019, doi: 10.11591/ijai.v8.i3.pp215-220.
[6] S. Barhmi and O. El Fatni, "Hourly wind speed forecasting based on support vector machine and artificial neural networks," *IAES Int. J. Artif. Intell.*, vol. 8, no. 3, pp. 286-291, Sep. 2019, doi: 10.11591/ijai.v8.i3.pp286-291.
[7] N. Razali, S. Ismail, and A. Mustapha, "Machine learning approach for flood risks prediction," *IAES Int. J. Artif. Intell.*, vol. 9, no. 1, pp. 73-80, Mar. 2020, doi: 10.11591/ijai.v9.i1.pp73-80.
[8] M. Mohaiminul and N. Sultana, "Comparative Study on Machine Learning Algorithms for Sentiment Classification," *Int. J. Comput. Appl.*, vol. 182, no. 21, pp. 1-7, Oct. 2018, doi: 10.5120/ijca2018917961.
[9] S. Amari and S. Wu, "Improving support vector machine classifiers by modifying kernel functions," *Neural Networks*, vol. 12, no. 6, pp. 783-789, Jul. 1999, doi: 10.1016/S0893-6080(99)00032-5.
[10] C. Maaoui and A. Pruski, "A comparative study of SVM kernel applied to emotion recognition from physiological signals," in *2008 5th International Multi-Conference on Systems, Signals and Devices*, Jul. 2008, pp. 1-6, doi: 10.1109/SSD.2008.4632891.
[11] H. M. Asraf, M. T. Nooritawati, and M. S. B. S. Rizam, "A Comparative Study in Kernel-Based Support Vector Machine of Oil Palm Leaves Nutrient Disease," *Procedia Eng.*, vol. 41, pp. 1353-1359, 2012, doi: 10.1016/j.proeng.2012.07.321.
[12] M. Achirul Nanda, K. Boro Seminar, D. Nandika, and A. Maddu, "A Comparison Study of Kernel Functions in the Support Vector Machine and Its Application for Termite Detection," *Information*, vol. 9, no. 1, p. 5, Jan. 2018, doi: 10.3390/info9010005.
[13] D. Shanthini, M. Shanthi, and D. M. C. Bhuvaneswari, "A Comparative Study of SVM Kernel Functions Based on Polynomial Coefficients and V-Transform Coefficients," *Int. J. Eng. Comput. Sci.*, vol. 6, no. 3, pp. 20765-20769, 2017, doi: 10.18535/ijecs/v6i3.65.
[14] D. F. Sengkey, A. Jacobus, and F. J. Manoppo, "Implementing Support Vector Machine Sentiment Analysis to Students' Opinion toward Lecturer in an Indonesian Public University," *J. Sustain. Eng. Proc. Ser.*, vol. 1, no. 2, pp. 194–198, Sep. 2019, doi: 10.35793/joseps.v1i2.27.
[15] N. Altrabsheh, M. M. Gaber, and M. Cocea, "SA-E: Sentiment analysis for education," in *Frontiers in Artificial Intelligence and Applications*, 2013, vol. 255, pp. 353–362, doi: 10.3233/978-1-61499-264-6-353.
[16] M. A. Ullah, "Sentiment analysis of students feedback: A study towards optimal tools," in *2016 International Workshop on Computational Intelligence (IWCI)*, Dec. 2016, pp. 175-180, doi: 10.1109/IWCI.2016.7860361.
[17] G. G. Esparza, A. P. Díaz, J. Canul-Reich, C. A. De-Luna, and J. Ponce, "Proposal of a Sentiment Analysis Model in Tweets for Improvement of the Teaching - Learning Process in the Classroom Using a Corpus of Subjectivity," *Int. J. Comb. Optim. Probl. Informatics*, vol. 7, no. 2, pp. 22-34, May 2016, Accessed: Jul. 13, 2019. [Online]. Available: https://ijcopi.org/index.php/ojs/article/view/37.
[18] G. G. Esparza *et al.*, "A sentiment analysis model to analyze students reviews of teacher performance using support vector machines," in *14th International Conference, Advances in Intelligent Systems and Computing*, vol. 620, pp. 157-164, 2018, doi: 10.1007/978-3-319-62410-5_19.
[19] V. I. Santoso, G. Virginia, and Y. Lukito, "Penerapan Sentiment Analysis pada Hasil Evaluasi Dosen dengan

Metode Support Vector Machine [The Implementation of Support Vector Machine for Sentiment Analysis of Lecturer Evaluation Results]," *J. Transform.*, vol. 14, no. 2, p. 72, Jan. 2017, doi: 10.26623/transformatika.v14i2.439.

[20] A. F. Hidayatullah and M. R. Ma'arif, "Pre-processing Tasks in Indonesian Twitter Messages," *J. Phys. Conf. Ser.*, vol. 801, p. 012072, Jan. 2017, doi: 10.1088/1742-6596/801/1/012072.

[21] S. K. Lidya, O. S. Sitompul, and S. Efendi, "Sentiment Analysis Pada Teks Bahasa Indonesia Menggunakan Support Vector Machine (SVM) dan K - Nearest Neighbour (K-NN) [Sentiment Analysis for Text in Bahasa Indonesia with Support Vector Machine (SVM) and K-Nearest Neighbour (K-NN)]," in *Seminar NasionalTeknologi Informasi dan Komunikasi 2015 (SENTIKA 2015)*, 2015, vol., no., pp. 1-8, [Online]. Available: https://fti.uajy.ac.id/sentika/publikasi/makalah/2015/1.pdf.

[22] A. F. Hidayatullah, "Language tweet characteristics of Indonesian citizens," in *2015 International Conference on Science and Technology (TICST)*, Nov. 2015, pp. 397-401, doi: 10.1109/TICST.2015.7369393.

[23] J. Silge and D. Robinson, "tidytext: Text Mining and Analysis Using Tidy Data Principles in R," *JOSS*, vol. 1, no. 3, 2016, doi: 10.21105/joss.00037.

[24] Devid Haryalesmana Wahid, "masdevid/ID-Stopwords: Stopwords collection of Bahasa Indonesia collected from many sources." 2016, Accessed: Jul. 15, 2019. [Online]. Available: https://github.com/masdevid/ID-Stopwords/.

[25] F. Z. Tala, "A Study of Stemming Effect on Information Retrieval in Bahasa Indonesia," Universiteit van Amsterdam, 2003.

[26] A. F. Hidayatullah, C. I. Ratnasari, and S. Wisnugroho, "Analysis of Stemming Influence on Indonesian Tweet Classification," *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 14, no. 2, pp. 665-673, 2016, doi: 10.12928/telkomnika.v14i2.3113.

[27] P. S. Mann, *Introductory Statistics*, 9th ed. Wiley, 2016.

[28] M. Kuhn, "caret: Classification and Regression Training." 2020, [Online]. Available: https://cran.r-project.org/package=caret.

[29] I. Feinerer, K. Hornik, and D. Meyer, "Text Mining Infrastructure in R," *J. Stat. Softw.*, vol. 25, no. 5, pp. 1-54, Mar. 2008, [Online]. Available: http://www.jstatsoft.org/v25/i05/.

[30] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch, "e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien." 2019, [Online]. Available: https://cran.r-project.org/package=e1071.

[31] R Core Team, "R: A Language and Environment for Statistical Computing." Vienna, Austria, 2020, [Online]. Available: https://www.r-project.org/.

[32] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3-4, pp. 591-611, Dec. 1965, doi: 10.1093/biomet/52.3-4.591.

[33] W. H. Kruskal and W. A. Wallis, "Use of Ranks in One-Criterion Variance Analysis," *J. Am. Stat. Assoc.*, vol. 47, no. 260, pp. 583–621, Dec. 1952, doi: 10.1080/01621459.1952.10483441.

BIOGRAPHIES OF AUTHORS

Daniel Febrian Sengkey was graduated from the Department of Electrical Engineering, Universitas Sam Ratulangi, Indonesia in 2012; and from the Department of Electrical Engineering and Information Technology, Universitas Gadjah Mada, Indonesia in 2015 for his undergraduate and master studies, respectively. In 2016 he joined the Electrical Engineering Department, Universitas Sam Ratulangi as a lecturer. His research interests are data communications networks, intelligent transportation systems, IT-assisted learning, and machine learning applications. Further info on the faculty directory: https://fatek.unsrat.ac.id/portal/index.php/public/dosen/id/199002232018031001

Agustinus Jacobus is a tenured lecturer in the Undergraduate Program of Informatics, Universitas Sam Ratulangi. He holds a bachelor's degree in Electrical Engineering from Universitas Sam Ratulangi and a master's degree in Computer Science from Universitas Gadjah Mada. His research interests are Data Science and Artificial Intelligence. Further info on the faculty directory: https://fatek.unsrat.ac.id/portal/index.php/public/dosen/id/197908042014041001

Fabian Johanes Manoppo is a Full Professor in Geotechnical Engineering and the current Dean of the Faculty of Engineering, Universitas Sam Ratulangi. Despite his Geotechnical background, he is very fluent in computer programming, hence in his term as dean many applications are developed for the faculty internal use, such as the student evaluation toward the lecturer system, in which its data is used in this study. Further info on the faculty directory: https://fatek.unsrat.ac.id/portal/index.php/public/dosen/id/196210141992031001