TECHNIUM

Romanian Journal of Applied Sciences and Technology

ISSN: 2668-778X

vol. 17 / 2023

The 7th International Conference on Science Technology

organized by
Faculty of Social Science and
Law Universitas Negeri Manado and
Consortium of International Conference
on Science and Technology

The Innovation Breakthrough in Digital and Disruptive Era

PLUS COMMUNICATION

International Communication & PR

Technium: Romanian Journal of
Applied Sciences and Technology
Vol. 17, 427-434, November, 2023
ISSN: 2668-778X
www.techniumscience.com

Classification of Public Opinion on Online Learning Policies using Various Support Vector Machine's Kernel

Husni[1*], *Arif* Muntasa[1], and *Mochamad Dani* Hartanto[1]

[1]Informatics Engineering Department, University of Trunojoyo Madura, Jl. Telang Raya Kamal, Bangkalan, East Java, Indonesia

Abstract. The COVID-19 pandemic has resulted in significant changes in the education sector. The government issued a policy so that learning must be carried out online from home. This policy became a polemic for teachers and students so that pro and con opinions emerged on social media, especially Twitter. Sentiment analysis of public opinion is an interesting study. Standard classification algorithms such as k-Nearest neighbours, naïve bayes, decision tree, random forest, and support vector machine (SVM) can categorize these opinions in a short time with good accuracy. Many studies show that SVM is more accurate than all other classification methods. SVM works using kernels, including Linear, Polynomial and Radial Basis Functions (RBF) where each kernel requires different parameters. The linear kernel only requires one parameter, namely c (Cost). The RBF kernel requires 2 parameters, c and ɣ (gamma) while the Polynomial kernel uses 2 parameters, c and degrees. SVM does not have default values for these parameters and are based on experience and experimentation. The wider the range of parameters, the more likely the classifier obtains the optimal values. This study tries some parameters values of SVM kernels for text classification based on sentiment. Testing using 5-fold cross validation and confusion matrix show that SVM with a linear kernel provides the best performance with an accuracy of above 84%.

* Corresponding author: husni@trunojoyo.ac.id

1 Introduction

The COVID-19 pandemic is a health crisis that has a profound impact on life. Many countries have stopped the spread of this corona virus by stopping all gathering activities, including the teaching and learning. In the next stage, the Government oblige the implementation of online learning activities from home [1] which rely heavily on notebook devices, mobile phones, tablets, Internet connections and video communication applications [2]. Some people immediately accepted this policy, some others did not accept it and brought the discussion to social media such as Twitter [3]. Netizens bring their own sentiments when discussing and it becomes an interesting study to find out the polarity of these social media users. A supervised machine learning approach for text classification can be used. The results of a survey by the Association of Indonesian Internet Service Providers (APJII) show that there are around 171,176,716.8 active internet service users in Indonesia and 1.7% or around 2.9 million are active visitors to social media Twitter [4] motivated us to do a sentiment analysis of this topic.

There are many machine learning methods that have been considered standard, outside of deep learning approaches. Windasari [5] used the Support Vector Machine (SVM) method for sentiment classification and obtained an average accuracy of 86% where every document with positive sentiment was able to be classified perfectly while for negative documents there was an error of 67.44%. Other methods provide average accuracy under SVM, including lexicon-based methods [6], naïve bayes [7][8], and clustering methods [9][10].

The SVM method works with kernel functions, including linear, polynomial, and Radial Basis Function (RBF) kernels where each kernel has certain mandatory parameters. The default kernel of SVM is RBF with parameters c = 1.0, degree = 3, and gamma = 'scale' as used in [11], [12], and [13] which shows that RBF is superior to other classification methods. However, until now it cannot be concluded that RBF always provides better accuracy than other kernels, for example [14] shows the results of classification with linear kernels are more accurate. The same thing happens in setting values for parameters from the kernel. The best kernels and parameters are obtained from experience and experimentation and behave differently on different datasets.

Several other studies using SVM and its kernel are Pratama [15] which uses the default SVM for classification of complaint text with an accuracy of 89%, Nurajijah [16] which uses SVM, naïve bayes and decision trees for calcification of financing approvals where SVM excels with an accuracy of 90%. , naïve Bayes 77% and decision tree 89%, Pratyo [17] uses SVM with a polynomial kernel for classification of public opinion regarding Covid-19 which provides 82% accuracy, Wirasati [18] uses 3 SVM kernels for thalassemia classification which obtains almost the same accuracy for the three kernels, namely 99.63% for RBF, 98.23% for linear, and 97.9% for polynomial. Finally, Agustina [19] uses SVM to classify public sentiment regarding online sales such as marketplaces.

This paper reports the results of research using the three SVM kernels above to classify public opinion on Twitter regarding online learning policies. In the next section, the SVM method will be explained, the architecture of the proposed approach, the testing and evaluation approach, the discussion of research results and closed by drawing conclusions.

2 Support Vector Machines

Support Vector Machine (SVM) is a linear classification method that is implemented to find the best hyperplane capable of separating the two categories in the input space [20]. The main task of the SVM method is to find the best classifier function that accurately separates the dataset into two different categories [21].

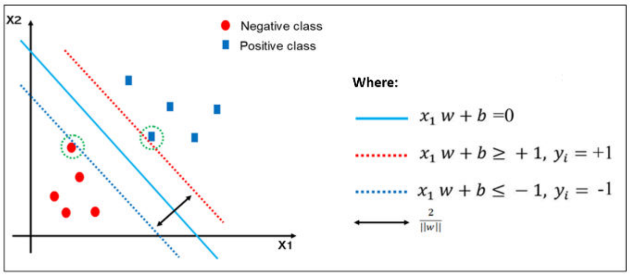

Fig. 1. Hyperplane (classifier) in SVM classification

The information contained within the boundary is known as a support vector. Separation of the two classes is carried out with a pair of the same fields. The top bounding field is for limiting the first class, while the bottom bounding field is for limiting the second class. The best separator field is one that is able to isolate all data in separate classes as shown in Fig. 1 and the distance between the two lines (margins) is getting bigger [22]. Equations (1) and (2) show the formulas for the two closest dividing lines for each class.

$$x_1\ w + b\ \geq +1\ , y_i\ =\ +1 \tag{1}$$
$$x_1\ w + b\ \leq -1\ , y_i\ =\ -1 \tag{2}$$

Variable w is the sector or normal area between the separation surface and the coordinate center, and variable b is the plane view relative to the coordinate center. The value of w and b can be obtained using equations (3) and (4).

$$w = \sum_{i=1}^{N} \alpha_i y_i x_i \tag{3}$$

$$b = -\frac{1}{2}(w.x^{+} + w.x^{-}) \tag{4}$$

The distance between the dividing line and the nearest point where the maximum margin is obtained is maximized by the function $\frac{1}{||w||}$ or minimized by $||w||^2$. The two dividing planes called hyperplanes are described in equation (5).

Technium: Romanian Journal of
Applied Sciences and Technology
Vol. 17, 427-434, November, 2023
ISSN: 2668-778X
www.techniumscience.com

$$y_i\ (w\ .x_i + b\ ) \geq 1, i = \ 1,2, \dots \dots, n \tag{5}$$

By finding the minimum point, the equation above can be expressed as a quadratic equation. The optimal hyperplane from the largest margin value can be formulated as a finite optimization problem with equation (6).

$$Min = \frac{1}{2}||w||^2 \tag{6}$$

After the b value is obtained, it can then be used to determine the support vector value of the new data. The optimal classification of the test data is obtained using equation (7).

$$f(x) = w.x + b \tag{7}$$

where:
x = data
y = class of data
w = weight value
b = bias value
N = total amount of data
α = weight values in each data point

The SVM classification process relies on a kernel function that takes data as input and transforms the data into the desired form based on the given parameters. Here are some of the types of kernel functions in SVM [23]:

- Linear kernels, known as soft margins, try to find hyperplanes that are straight lines, but can also tolerate one or more data misclassifications. This kernel tries to find the line that maximizes margins and minimizes errors. The amount of error tolerance greatly affects the accuracy of the hyperplane.
- Radial Basis Function (RBF) is a kernel function that is required when the data is distributed unevenly. When training RBF requires two parameter values cost (c) and ɣ (gamma). Parameter c determines how much error must be avoided in the classification of training data. The greater the value of c, the smaller the training data classification error. The parameter ɣ determines the effect of a training data sample, the smaller the value of ɣ, the farther the distance from the data point to be calculated.
- Polynomial kernels are used when data cannot be separated by a straight line. This kernel can generate non-linear decision boundaries, generating new features by applying polynomial combinations of existing features.

The linear kernel only uses one parameter that needs to be optimized, namely c. The RBF kernel requires 2 parameters c and ɣ. The polynomial kernel takes 2 parameters c and degree. The default kernel for SVM is RBF with a value of c = 1.0, degree = 3, and gamma = 'scale' [13]. Such kernel configurations and parameters do not always provide the best classification accuracy. Repeated experiences and experiments serve as the basis for determining these values. The wider the range of parameters, the more likely the classification process will find the best combination of parameters.

As with other methods of text classification, SVM accepts input data that is ready to be categorized. Therefore, before it can be used by SVM, the dataset must be preprocessed which includes case folding tasks to convert text into lowercase, cleaning to clean non-text elements such as punctuation, numbers, binary codes, and emoji, tokenization to extract words (tokens) from sentences (every word, number, punctuation and entity can be considered as a token), filtering to select the token that is important in computing, stopword removal to remove tokens that have no effect in computing, and stemming to convert each token to the form its roots [25].

An important step that must be carried out before entering into the classification process is giving weight to each word collected at the preprocessing stage. The de-facto weighting approach is TF-IDF (term frequency and inverse document frequency). TF is the frequency of occurrence of a word in a certain document, while IDF is the inverse value of the number of documents containing a certain term. The TF-IDF formula of a term is shown in equation (8) [26].

$$TF.IDF(ti, dj) = \ tf(ti, dj) * log\frac{N+1}{ni} + 1 \tag{8}$$

where:
TF-IDF (ti,dj) : The weight of the term i in document j.
tf(ti,dj) : frequency of words or terms i in documents j.
N : The total number of documents in the dataset.
n_i : The number of documents where the term i appears.

The last task before classification is normalization which aims to reduce term redundancies and ensure dependencies on equation (8). This normalization is carried out by L2 with equation (9) [27].

$$V_{norm} = \frac{\text{v}}{\sqrt{v_1^2 + v_2^2 + \dots\dots + v_n^2}} \tag{9}$$

where:
Vnorm: normalization results.
v: calculation results in equation (8).

3 Cross Validation and Accuracy

Cross Validation is a data splitting technique designed to obtain maximum accuracy when the data is devided into two parts (training and testing data). The most widely used type of Cross Validation is the K-Fold Cross Validation which evaluates the classification performance of the model. Cross Validation works by dividing the sample data randomly and grouping the data into K values in the K-Fold [28]. Training and testing data must cross each other in successive iterations so that each piece of data gets the opportunity to be validated with the same size. A working illustration of k-fold cross validation with a value of k=5 is shown in Fig. 2.

Technium: Romanian Journal of
Applied Sciences and Technology
Vol. 17, 427-434, November, 2023
ISSN: 2668-778X
www.techniumscience.com

The accuracy of the classification method can be measured using a confusion matrix diagram in the form of a table consisting of 4 combinations of classification results. An illustration of the confusion matrix is shown in Fig. 3.

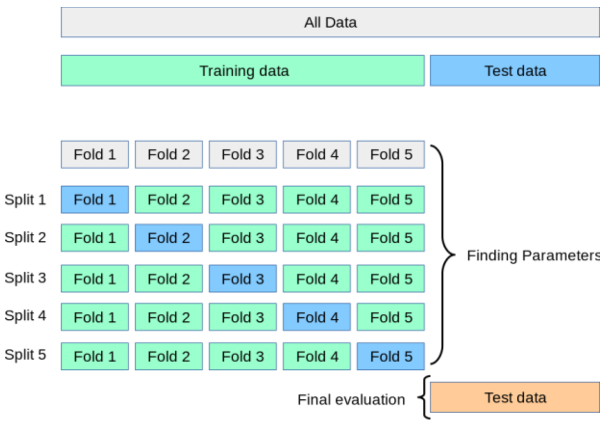

Fig. 2. Data splitting and cross validation in 5-fold

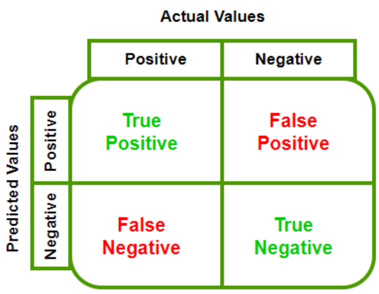

Fig. 3. Confusion matrix to compute accuracy

The following is the definition of the results of the classification [21]:

- *True Postive (TP)*: the predicted result in the positive class is the same as the actual positive label.
- *True Negative (TN)*: the predicted result in the negative class is the same as the actual negative label.
- *False Positive (FP)*: the predicted results in the positive class are different from the actual negative labels.
- *False Negative (FN)*: the predicted results in the negative class are different from the actual positive labels.

The accuracy of the classification is calculated using the formula (10).

$$Accuracy=(TP+TN)/(TP+TN+FP+FN) \qquad (10)$$

4 Proposed Method

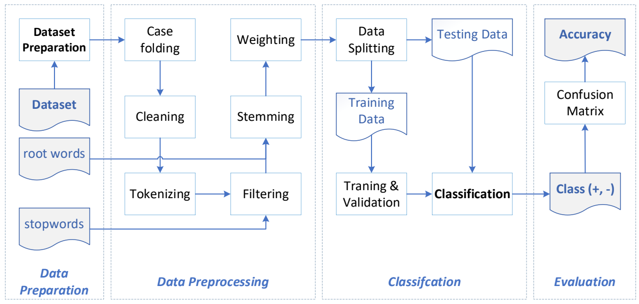

Fig. 4. Architecture of proposed classification system

Broadly speaking, the classification process with SVM is divided into 4 stages as shown in Fig. 4, that is:

1. **Data Preparation**. Here all the necessary data is prepared, including removing unnecessary attributes or features in determining the polarity of tweets. The most important work at this stage is the labeling of each tweet and its validation by the Indonesian language experts involved.
2. **Data Preprocessing**. Here data proprocessing works through 6 processes, namely case folding, cleaning, tokenizing, filtering which requires a list of stopwords and NLTK libraries, stemming specifically for Indonesian using the Literature library, and finally weighting with TF-IDF.
3. **Classification** consisting of 3 tasks, namely dataset splitting and training using 5-fold cross validation and classification with SVM.
4. **Evaluation**. At this stage, a review is carried out on the results of the SVM classification and compared with the labels that have been set at the Preparation stage. This evaluation is summarized using a confusion matrix to calculate accuracy. For all three kernels used.

Evaluation of the SVM classifier with three kernels follows the following steps:

a. Data initialization results of data processing.
b. SVM kernel type determination.
c. Determination of the values of the SVM kernel parameters that will be applied in the classification process: c = (1, 2, 3), ɣ = (1, 2, 3) and degree = (4, 5, 6).
d. Dividing training data and testing data with 5-fold cross validation.

Determination of the value of c, ɣ, and degree in the testing scenario refers to research that has been done previously by Azies et al, using c = (1, 2, 3), then for the value of ɣ = (1, 2, 3), and the value of degree = (4, 5, 6)[24]. Linear kernel only uses c parameter, RBF kernel uses c and ɣ parameters, and Polynomial kernel uses c and degree. Whereas in k-fold cross validation k = 5 is used according to the discussion in [25]. Details of this testing scenario are shown in Table 1.

Table 1 shows that there are three kernels with each parameter value. Previous research conducted a study on the RBF kernel, linear and polynomial, found that the RBF kernel function has the highest performance [29]. The use of appropriate parameters in

Technium: Romanian Journal of
Applied Sciences and Technology
Vol. 17, 427-434, November, 2023
ISSN: 2668-778X
www.techniumscience.com

the RBF kernel is able to provide high accuracy values compared to linear and polynomial kernels, with a combination of parameter values c = 1 and ɣ = 1, a classification accuracy of 98.1% is obtained, which is the best classification accuracy for the RBF kernel [24]. Other studies have compared the performance of linear and polynomial kernels, it was found that the best performance was given by the polynomial kernel function with an accuracy value of 51.2% while for the Linear kernel with an accuracy value of 43.7% [30].

Table 1 Evaluation scenarios

#	Kernel	Fold	Parameter		
			C	**gamma**	**degree**
1	Linear	5	1, 2, 3	-	-
2	RBF	5	1, 2, 3	1, 2, 3	-
3	Polynomial	5	1, 2, 3	-	4,5, 6

The dataset in this study is the same as one used in the research by Musfiroh, et al [6] who used the Lexicon InSet to analyze sentiment for online lectures in Indonesia from the Twitter dataset. The tweet data used is 5393 tweets in csv format. Dataset collection utilizes Twitter's API (Application Programming Interface) with the keyword used in the scraping process being "online lectures" with tweet data limits from June 17 to December 19, 2020.

Table 2 The confusion matrix for scenario 1

#	c	k	TP	FP	FN	TN
1	1	1	720	51	126	182
2		2	724	36	133	186
3		3	698	37	114	230
4		4	691	51	134	202
5		5	697	49	125	207
6	2	1	707	64	113	195
7		2	711	49	113	206
8		3	677	58	96	248
9		4	673	69	108	228
10		5	679	67	105	227
11	3	1	702	69	113	195
12		2	702	58	106	213
13		3	668	67	91	253
14		4	664	78	98	238
15		5	673	73	97	235

5 Discussion

The evaluation phase is carried out by entering testing data and recalling the model that has been previously trained to determine the performance of the model by calculating accuracy, recall, precision, and f1_score.

5.1 Evaluation Scenario 1

In the first scenario, it was carried out to determine the performance produced by the model using a linear kernel. The prediction results from the model are formed in the Confusion Matrix as presented in Table 2. The experiment was carried out 15 times with a value of c = (1,2, 3). The performance of the linear kernel is shown in Table 3.

Technium: Romanian Journal of Applied Sciences and Technology
Vol. 17, 427-434, November, 2023
ISSN: 2668-778X
www.techniumscience.com

Table 3 The performance of scenario 1

#	c	k	accuracy	recall	precision	f1 score
1	1	*1*	83,60	85,11	93,39	89,05
2		*2*	84,34	84,48	95,26	89,55
3		*3*	86,01	85,96	94,97	90,24
4		*4*	82,84	83,76	93,13	88,19
5		*5*	83,86	84,79	93,43	88,90
6	2	*1*	83,60	86,22	91,70	88,87
7		*2*	84,99	86,29	93,55	89,77
8		*3*	85,73	87,58	92,11	89,79
9		*4*	83,58	86,17	90,70	88,38
10		*5*	84,04	86,61	91,02	88,76
11	3	*1*	83,13	86,13	91,05	88,52
12		*2*	84,80	86,88	92,37	89,54
13		*3*	85,36	88,01	90,88	89,42
14		*4*	83,67	87,14	89,49	88,30
15		*5*	84,23	87,40	90,21	88,79
Std. Deviation			0,94	1,20	1,71	0,62
Average			84,25	86,17	92,22	89,07
Maximum			86,01	88,01	95,26	90,24
Minimum			82,84	83,76	89,49	88,19

5.2 Evaluation Scenario 2

The Scenario 2 were conducted to determine the performance of the SVM model with the RBF kernel. Based on this trained model, trials and evaluations of the prediction results with test data are carried out. The prediction results from the model are made in the Confusion Matrix as presented in Table 4 and the performance of the RBF kernel is shown in Table 5.

Table 4 The confusion matrix for scenario 2

#	c	γ	k	TP	FP	FN	TN
1	1	1	*1*	735	16	169	159
2			*2*	743	29	164	143
3			*3*	715	17	192	155
4			*4*	726	24	163	165
5			*5*	720	29	186	143
6	2	2	*1*	724	12	214	129
7			*2*	719	14	245	101
8			*3*	752	9	203	115
9			*4*	730	17	233	98
10			*5*	762	15	195	106
11	3	3	*1*	756	3	287	33
12			*2*	762	1	300	16
13			*3*	760	1	286	32
14			*4*	753	0	293	32
15			*5*	716	2	334	26

5.3 Evaluation Scenario 3

The Scenario 3 were conducted to determine the performance produced by the model using the Polynomial kernel. Based on the classification model that has been trained, evaluation of the prediction results with test data are carried out. The prediction results from the model are formed in the Confusion Matrix as presented in Table 6 and the summary of the performance of the kernel function is presented in Table 7.

Table 5 The performance of scenario 2

#	c	γ	k	accuracy	recall	precision	f1 score
1	1	1	*1*	82,85	81,31	97,87	88,82
2			*2*	82,11	81,92	96,24	88,51
3			*3*	80,63	78,83	97,68	87,25
4			*4*	82,65	81,66	96,80	88,59
5			*5*	80,06	79,47	96,13	87,01
6	2	2	*1*	79,05	77,19	98,37	86,50
7			*2*	76,00	74,59	98,09	84,74
8			*3*	80,35	78,74	98,82	87,65
9			*4*	76,81	75,80	97,72	85,38
10			*5*	80,52	79,62	98,07	87,89
11	3	3	*1*	73,12	72,48	99,60	83,91
12			*2*	72,10	71,75	99,87	83,51
13			*3*	73,40	72,66	99,87	84,12
14			*4*	72,82	71,99	100,00	83,71
15			*5*	68,83	68,19	99,72	81,00
Std. Deviation				4,45	4,27	1,31	2,33
Average				77,42	76,41	98,32	85,90
Maximum				82,85	81,92	100,00	88,82
Minimum				68,83	68,19	96,13	81,00

Table 6 The confusion matrix for scenario 3

#	C	degree	k	TP	FP	FN	TN
1	1	4	*1*	770	1	289	19
2			*2*	759	1	303	16
3			*3*	735	0	321	23
4			*4*	741	1	317	19
5			*5*	744	2	319	13
6	2	5	*1*	770	1	289	19
7			*2*	760	0	304	15
8			*3*	735	0	323	21
9			*4*	739	3	319	17
10			*5*	744	2	320	12
11	3	6	*1*	770	1	290	18

12			2	760	0	307	12
13			3	735	0	325	19
14			4	740	2	320	16
15			5	744	2	320	12

Table 7 The performance of scenario 3

#	c	degree	k	accuracy	recall	precision	f1 score
1	1	1	1	73,12	72,71	99,87	84,15
2			2	71,83	71,47	99,87	83,32
3			3	70,25	69,60	100,00	82,08
4			4	70,50	70,04	99,87	82,33
5			5	70,22	69,99	99,73	82,26
6	2	2	1	73,12	72,71	99,87	84,15
7			2	71,83	71,43	100,00	83,33
8			3	70,06	69,47	100,00	81,99
9			4	70,13	69,85	99,60	82,11
10			5	70,13	69,92	99,73	82,21
11	3	3	1	73,03	72,64	99,87	84,11
12			2	71,55	71,23	100,00	83,20
13			3	69,88	69,34	100,00	81,89
14			4	70,13	69,81	99,73	82,13
15			5	70,13	69,92	99,73	82,21
Std. Deviation				1,23	1,24	0,13	0,85
Average				71,06	70,68	99,86	82,76
Maximum				73,12	72,71	100,00	84,15
Minimum				69,88	69,34	99,60	81,89

Tables 2 to 7 show the test results for each kernel in detail. The summary of the evaluation in the form of average accuracy can be seen in Table 8, Table 9 and Table 10.

Table 8 The Performance of linear kernel

Accuracy				Parameter
Average	Min	Max	Std. Deviation	*cost*
84.12	82.84	86.01	1.18	1
84.38	83.58	85.73	0.94	2
84.23	83.13	85.3	0.88	3

Table 9 The Performance of RBF kernel

Accuracy				Parameters	
Average	Min	Max	Std. Deviation	*cost*	*gamma*
81.66	80.06	82.85	1.25	1	1
78.54	76.00	80.52	2.05	2	2
72.05	68.83	73.40	1.86	3	3

Table 10 The Performance of Polynomial kernel

Accuracy				Parameters	
Average	Min	Max	Std. Deviation	*cost*	*degree*
71.18	70.22	73.11	1.27	1	4
71.05	70.06	73.12	1.38	2	5
70.09	69.88	73.03	1.34	3	6

The summary in Table 8, Table 9, and Table 10 shows that in the first scenario the SVM model with a linear kernel gives an average accuracy value of 84.38% sat parameter c = 2. In the second scenario, the RBF kernel function is able to provide an accuracy of 81.66% when using parameters c=1 and ɣ = 1. In the third scenario, the Polynomial kernel function provides an average accuracy of 71.18% when applying parameters c=1 and degree = 4. Fig. 5 summarizes the performance (accuracy) comparison of the three scenarios and kernel functions. The tests that have been carried out lead us to an agreement that the SVM classifier with a linear kernel function is able to provide the best accuracy in the case of tweet text classification regarding public opinion regarding online learning policies after the Covid-19 Pandemic.

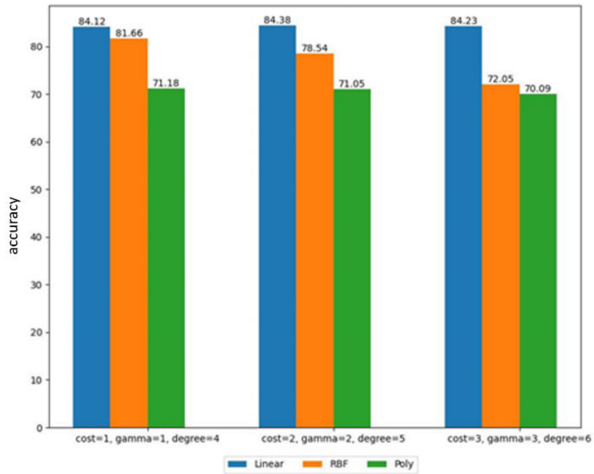

Fig. 5 Visualization of the average accuracy of each scenario

6 Conclusion

This research has applied the SVM classification method to conduct a sentiment analysis of public opinion on social media related to government policies for implementing online learning at home. Three kernel functions in SVM (linear, RBF, and polynomial) have been tested and the results show that the linear kernel function is able to provide better accuracy (84%) than other kernal functions when a value of c=1 is applied. RBF which is the default kernel function of SVM provides accuracy that competes with linear kernels (81%). This study, based on the data used, can conclude that the kernel liner and RBF functions are recommended for use with the SVM classifier, especially in sentiment-based text classification on social media..

References

1. A. Purwanto et al., “Studi Eksploratif Dampak Pandemi COVID-19 Terhadap Proses Pembelajaran Online di

Sekolah Dasar," EduPsyCouns J. Educ. Psychol. Couns. **2** (2020)

2. N. H. Zhafira, Y. Ertika, and Chairiyaton, "Persepsi Mahasiswa Terhadap Perkuliahan Daring Sebagai Sarana Pembelajaran Selama Masa Karantina Covid-19," J. Bisnis dan Kaji. Strateg. Manaj. **4** (2020)
3. I. H. Oktafia and S. W. Siti, "Pembelajaran Daring Sebagai Upaya Study From House (SFH) selama Pandemi Covid 19," J. Pendidik. Adm. Perkantoran **8** (2020)
4. APJII, "Penetrasi & Profil Perilaku Pengguna Internet Indonesia Tahun 2018," APPJI Report (2019)
5. I. P. Windasari, F. N. Uzzi, and K. I. Satoto, "Sentiment analysis on Twitter posts: An analysis of positive or negative opinion on GoJek," in Proceedings of International Conference on Information Technology, Computer, and Electrical Engineering ICITACEE (2017)
6. D. Musfiroh, U. Khaira, P. Eko, P. Utomo, and T. Suratno, "Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon," MALCOM Indones. J. Mach. Learn. Comput. Sci. J.**1** (2021)
7. U. Verawardina, F. Edi, and R. Watrianthos, "Analisis Sentimen Pembelajaran Daring Pada Twitter di Masa Pandemi COVID-19 Menggunakan Metode Naïve Bayes," J. Media Inform. Budidarma **5** (2021)
8. G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," INTEGER J. Inf. Technol. **2** (2017)
9. W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," Ain Shams Eng. J. **5** (2014)
10. H. Tuhuteru, "Analisis Sentimen Masyarakat Terhadap Pembatasan Sosial Berksala Besar Menggunakan Algoritma Support Vector Machine," Inf. Syst. Dev. **5** (2020)
11. C. Zoltan, "SVM and Kernel SVM", URL: https://towardsdatascience.com/svm-and-kernel-svm-fed02bef1200, accessed August 20 (2023)
12. Kaggle, "Understanding Parameters of SVM", URL: https://www.kaggle.com/code/gorkemgunay /understanding-parameters-of-svm, accessed August 20 (2023)
13. Albert, "Analisis Topik dan Perbandingan Klasifikasi pada Kolom Komentar Video Youtube Edukasi Indonesia Menggunakan Pendekatan Latent Dirichlet Allocation" J. on Education **05**, 03 (2023)
14. G. Abdurraman, "Klasifikasi Kanker Payudara Menggunakan Algoritma SVM dengan Kernel RBF, Linier, dan Sigmoid" JUSTIFY: J. Sistem Informasi Ibrahimy **2**, 1 (2023)
15. P. H. Prastyo, A. S. Sumi, A. W. Dian, and A. E. Permanasari, "Tweets Responding to the Indonesian Government's Handling of COVID-19: Sentiment Analysis Using SVM with Normalized Poly Kernel," J. Inf. Syst. Eng. Bus. Intell. **6** (2020)
16. M. J. Lavin, Z. Leblanc, and Q. Dombrowski, "The Programming Historian Analyzing Documents with TF-IDF," in Programming Historian (2020)
17. J. Liu, H. Yan, and Y. Du, "Application of Text Analysis Technology in Aviation Safety Information Analysis Application of Text Analysis Technology in Aviation Safety Information Analysis," in 2nd International Conference on Computer Modeling, Simulation and Algorithm, (2020)
18. L. Muflikhah, N. Hidayat, and D. J. Hariyanto, "Prediction of hypertention drug therapy response using K-NN imputation and SVM algorithm," Indones. J. Electr. Eng. Comput. Sci. **15** (2019)
19. M. Ahmad, S. Aftab, and I. Ali, "Sentiment Analysis of Tweets using SVM," Int. J. Comput. Appl. **177** (2017)
20. M. R. A. Nasution and M. Hayaty, "Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter," J. Inform. **6** (2019)
21. I. M. Parapat, M. T. Furqon, and Sutrisno, "Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak," J. Pengemb. Teknol. Inf. dan Ilmu Komput. **2** (2018)
22. H. Al Azies, D. Trishnanti, and E. Mustikawati P.H, "Comparison of Kernel Support Vector Machine (SVM) in Classification of Human Development Index (HDI)," in IPTEK Journal of Proceedings Series **6** (2019)
23. G. Balita, S. Dengan, and P. K. Cross, "Aplikasi naïve bayes classifier (nbc) pada klasifikasi status gizi balita stunting dengan pengujian k-fold cross validation," J. GAUSSIAN **11** (2022)
24. E. E. Pratama and B. R. Trilaksono, "Klasifikasi Topik Keluhan Pelanggan Berdasarkan Tweet dengan Menggunakan Penggabungan Feature Hasil Ekstraksi pada Metode Support Vector Machine (SVM)," J. Edukasi dan Penelit. Inform. **1** (2015)
25. N. Nurajijah and D. Riana, "Algoritma Naïve Bayes, Decision Tree, dan SVM untuk Klasifikasi Persetujuan Pembiayaan Nasabah Koperasi Syariah," J. Teknol. dan Sist. Komput. **7** (2019)
26. I. Wirasati, Z. Rustam, J. E. Aurelia, S. Hartini, and G. S. Saragih, "Comparison some of kernel functions with support vector machines classifier for thalassemia dataset," IAES Int. J. Artif. Intell. **10** (2021)
27. D. A. Agustina, S. Subanti, and E. Zukhronah, "Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Marketplace di Indonesia Menggunakan Algoritma Support Vector Machine," Indones. J. Appl. Stat. **3** (2021)
28. D. A. Agustina, S. Subanti, and E. Zukhronah, "Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Marketplace di Indonesia Menggunakan Algoritma Support Vector Machine," Indones. J. Appl. Stat. **3** (2021)
29. M. T. Furqon, Indriati, and A. Hutapea, "Penerapan Algoritma Modified K-Nearest Neighbour Pada Pengklasifikasian Penyakit Kejiwaan Skizofrenia," J. Pengemb. Teknol. Inf. dan Ilmu Komput. **2** (2018)
30. R. Mukarramah, D. Atmajaya, and L. B. Ilmawan, "Performance comparison of support vector machine (SVM) with linear kernel and polynomial kernel for multiclass sentiment analysis on twitter," Ilk. J. Ilm. **13** (2021)