TELKOMNIKA Telecommunication Computing Electronics and Control
Vol. 22, No. 2, April 2024, pp. 329~339
ISSN: 1693-6930, DOI: 10.12928/TELKOMNIKA.v22i2.25437

Optimization of support vector machine with cubic kernel function to detect cyberbullying in social networks

Al-Khowarizmi[1], Indah Purnama Sari[1], Halim Maulana[2]
[1]Department of Information Technology, Faculty of Computer Science and Information Technology, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia
[2]Department of Information System, Faculty of Computer Science and Information Technology, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia

Article Info

Article history:

Received Jun 16, 2023
Revised Dec 9, 2023
Accepted Jan 12, 2024

Keywords:

Cubic kernel
Cyberbullying
Linier kernel function
Social network
Support vector machine

ABSTRACT

Social networking is a place where humans can interact using the internet network to be able to disseminate information, discuss, exchange ideas, pour out their hearts, and share activities. Many social networks are popularly used, one of which is Twitter. Information can be received quickly using Twitter. In addition, various government agencies also use Twitter to be able to interact directly with the community so that every government policy is disseminated through this social network. Every government policy neglects to reap the pros and cons of society, both collectively and individually. As a result of the pros and cons, a trial called cyberbullying was recorded. Cyberbullying in various studies has been carried out to change a person's raw material so that with the application of information technology, identifying cyberbullying needs to be carried out further. The problem of cyberbullying is generally detected using the support vector machine (SVM) method. Cyberbullying detection is conducted in dealing with government policy data such as "cipta kerja" by using the SVM method which is optimized using the cubic kernel function. The accuracy value achieved in SVM uses a linear kernel function of 92.3% while using a cubic linear function of 90%.

Corresponding Author:

Al-Khowarizmi
Department of Information Technology, Faculty of Computer Science and Information Technology
Universitas Muhammadiyah Sumatera Utara
Jl. Kapten Mukhtar Basri No 3 Medan, 20238, Sumatera Utara, Indonesia
Email: alkhowarizmi@umsu.ac.id

1. INTRODUCTION

Social networking is a forum where human beings can disseminate information, discuss, exchange ideas, pour out their hearts and share activities carried out in cyberspace by being delivered via the internet network [1], [2]. Various social networks have their respective advantages, such as uploading photos, text, videos and voice recording that are able to interact completely and personally [3]. All ages and groups can conduct discussions on social networks without differences in human degrees [4]. The government can use the advantages and benefits of social networking to interact with the community in conveying information and receiving complaints from the public [5].

As a result of freedom of opinion in social networks, every government in various countries makes policies so that there are no unpleasant actions for social network users [6]. In addition, the government in making policies for a country can be broadcast and disseminated through social networks so that delivery can be done quickly [7]. With this, we can reap the pros and cons of policies made by the government because

they are networked by the social community, they can give opinions without any restrictions by using social networks.

With the pros and cons of government policies made, social networking cannot resist the accelerated rate of giving opinions made on social networks [8]. So that arises encouragement from various groups and individuals on social networking [9]. This accelerates the delivery of bladder to become cyberbullying, where cyberbullying is an abuse of the use of the internet network by harassing, humiliating, threatening, and insulting others with digital trace recording [10].

Cyberbullying in various literacy which is embodied by LaFrancis and Putnam [11] explained that cyberbullying in the teenage group, the role of parents is needed to avoid raw failure in the real world. Meanwhile, according to Jiang *et al.* [12] it is necessary to have a government policy in determining the regulations for using cyberbullying. From the literacy, cyberbullying needs to be limited and detected using computer science methods. In accordance with the science of computer science, there are methods for classifying and clustering with results to detect certain events [13]–[15]. This knowledge is contained in data mining in combination with artificial neural networks [16], [17].

The speed of increasing data on cyberbullying on social networks cannot be stopped manually [18], [19]. So, the need for a method to be able to detect the outpouring of cyberbullying on social networks [20]. Potha and Maragoudakis [21] performs cyberbullying to detect and search for cyber predators using the time series method where each predator question is manually explained in terms of severity using a numeric label and applies the dynamic time warping algorithm, so that similarities in the signals are evident, providing a direct indicator for the severity of cyber bullying in this particular dialogue. Meanwhile, Nahar *et al.* [22] with a data mining approach in detecting cyberbullying with stochastic data and explaining that support vector machine (SVM) is an algorithm with supervised learning where the process must first have a label with fuzzy so that it is continued with the SVM algorithm so that the data becomes complex and multi-dimensional and this method is called fuzzy SVM.

Andriansyah *et al*. [23] classified the comment column on Instagram on artist accounts which are often referred to as celebrities in detecting cyberbullying where there were 1053 comments and 34 inspection documents containing cyberbullying of 79.412%. Meanwhile, Noviantho *et al*. [24] identify cyberbullying conversations with a combination of SVM and Naïve Bayes with an accuracy of 92.81% but the algorithm is modified with a poly kernel and achieves optimization at an accuracy of 97.11%. From various studies before doing cyberbullying, they often use SVM so that in this paper a study is designed to detect cyberbullying with data that is tested for cyberbullying on government policies using SVM which is focused on the method, namely using the kernel function to get optimal results using the cubic kernel function.

2. MATERIAL AND METHOD

2.1. Dataset

The dataset in this paper uses trending data on twitter on government policies regarding the determination of "cipta kerja" which uses the pros and cons so that cyberbullying is identified in every tweet posted by the Twitter social network user. Through keywords using "cipta kerja", the activities are sorted so that the data is crawling and then cyberbullying is detected in government policy. The amount of data achieved is 2400 tweets.

2.2. General architecture

This paper is the result of research that focuses on the SVM algorithm that was developed. Where the kernel used is a linear kernel but compared to the use of other kernels such as cubic kernels. So, the contribution in this paper is to produce optimal accuracy values in detecting. So, a general architecture was formed in this research so that it does not spread. The general architecture in this paper can be illustrated in Figure 1. The explanation in Figure 1 is loaded according to the following steps:

- Perform data search with the keyword "cipta kerja" on twitter.
- Crawling data.
- Preprocessing data.
- Conduct detection training.
- Classifying with SVM using the linear kernel function.
- Classifying with SVM using the cubic kernel function.
- Analysis of the success in obtaining optimization from detection using SVM.

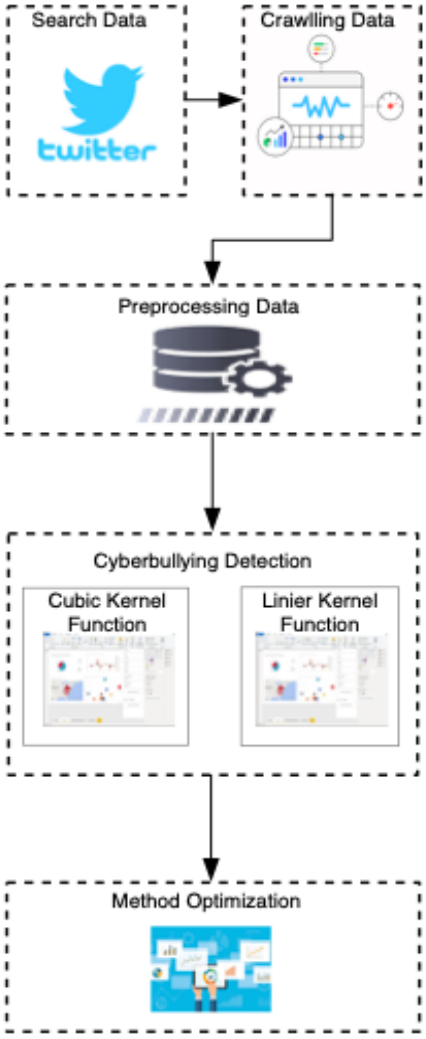

Figure 1. General architecture

2.3. Optimization of SVM

SVM is an algorithm that is relatively new and is able to solve regression and classification problems on very large amounts of data [25], [26]. In carrying out learning, of course SVM uses supervised learning techniques with supervision and then enters the testing stage [27], [28]. In the process, SVM also performs a linear projection from the feature space to the kernel and creates separate and linear classes [29], [30].

The SVM algorithm also has a very significant advantage in optimization, namely by achieving a minimum value in global and local problems. So that SVM is often used in cases to classify, detect, estimate, and predict [27]. In its history, SVM was discovered by Vapnik in 1963 which was used in doing classification problems [31], [32]. SVM can also basically classify data with high dimensional non-linearity [33]. With strong learning, SVM is widely applied and applied to the real world with accurate computational techniques [29], [34]. SSVM uses supervised learning techniques which are often found in biophotonics, image detection, classification, estimation, and prediction problems [35]. SVM is calculated based on the hyperplane as in the (1) [36], [37]:

$$W\ x\ \phi(x) + b = 0 \quad (1)$$

Where W is the normal value of the hyperplane, $\phi(x)$ is a function of the input vector, and b is the bias value. Then from these calculations an optimization is carried out in order to get the minimum value based on (2) [35]:

$$\min_{\alpha}\psi(\alpha) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} y_i y_j K(x_i, x_j)\ \alpha_i\alpha_j - \sum_{i=1}^{N}\alpha_i \quad (2)$$

Where x_i is the input vector value, y_i is the n class correspondence value, (α) is the lagrange multiplier value, and K is the linear kernel and cubic kernel function. In the combination of parameters must be determined in advance in getting it. Thus, SVM can be grouped into linear and nonlinear. As seen in (3):

$$(x) = \sum_{i=0}^{N} \alpha_i \, y_i x_i^T . x + \beta_0 \quad (3)$$

Where x_i is the value on the label y_i, α is the value to be multiplied by lagrange and β_0 is the bias value of 0. In non-linear problems the equation is changed based on (4) [38].

$$f(x) = \sum_{i=0}^{N} \alpha_i y_i K(x_i . x) . x + \beta_0 \quad (4)$$

Where the value of N is the value of the number of input vectors.

3. RESULTS AND DISCUSSION

At this stage, the results of applying sentiment analysis to data originating from social media with the keyword "cipta kerja" will be presented using the SVM algorithm with kernel cubic and linear kernel. This study will use as much as 2400 data originating from social media twitter. The data is obtained from the crawling process on twitter social media then the text preprocessing stage is carried out to change unstructured data into structured data which will then be carried out manually labeling, in data labeling there will be positive, negative and neutral labels which will then be carried out sentiment analysis with an SVM algorithm. Following are the results of the sentiment analysis process using the SVM algorithm with cubic kernels and linear kernels.

3.1. Text preprocessing results

The result of the text preprocessing process is tweet data with the keyword "cipta kerja" which has been changed and cleaned of irrelevant characters. In the text preprocessing there are 4 stages including case folding, tokenization, stopword and stemming. However, in this study only case folding, tokenization and stopword stages were carried out because the resulting data does not require a stemming process. The following data is the result of text preprocessing for each stage.

a. Case folding

Case folding is a text preprocessing process that changes the entire contents of tweets to lowercase with the aim of making text data analysis able. Where this research uses text data originating from Twitter and case folding will be carried out. Following are the results of case folding on the dataset contained in Table 1.

Table 1. The case folding process

No	Tweet	Result of case folding
1	RT @YLBHI: Rezim pemerintahan saat ini banyak memunculkan kebijakan yang menyengsarakan rakyat dan hanya mementingkan penguasa. Hukum diper…	Rezim pemerintahan saat ini banyak memunculkan kebijakan yang menyengsarakan rakyat dan hanya mementingkan diper
2	RT @ChaUnk_VR1: UU Cipta kerja telah ditolak oleh berbagai lapisan masyarakat sejak belum disahkan dan mengakibatkan gelombang PHK yang uga…	UU cipta kerja telah ditolak oleh berbagai lapisan masyarakat sejak belum disahkan dan mengakibatkan gelombang phk yang uga
….	….	……
….	……	……
2399	RT @atr_bpn: Halo #SobATRBPN, Kementerian ATR/BPN melalui PERPU Nomor 2 Tahun 2022 tentang Cipta kerja, akan mengatur pemanfaatan Hak Peng…	Halo sobatrbpn kementerian atrbpn melalui perpu nomor 2 tahun 2022 tentang cipta kerja akan mengatur pemanfaatan hak peng
2400	RT @vita_AVP: Agar tercapai pemahaman yang sinkron antar pemangku kepentingan, Kominfo terus melakukan sosialisasi UU No.2/2020 Tentang Cip…	agar tercapai pemahaman yang sinkron antar pemangku kepentingan kominfo terus melakukan sosialisasi uu no22020 tentang cip

b. Tokenization

Tokenization is a process carried out to separate tweet data into separate words or tokens. In this case, the data uses text originating from Twitter and has been casefolded and then tokenized. Following are the results of the tokenization of the dataset contained in Table 2.

Table 2. The tokenization process

No	Tweet	Tokenization results
1	Rezim pemerintahan saat ini banyak memunculkan kebijakan yang menyengsarakan rakyat dan hanya mementingkan diper	'rezim', 'pemerintahan', 'saat', 'ini', 'banyak', 'memunculkan', 'kebijakan', 'yang', 'menyengsarakan', 'rakyat', 'dan', 'hanya', 'mementingkan', 'diper'
2	UU cipta kerja telah ditolak oleh berbagai lapisan masyarakat sejak belum disahkan dan mengakibatkan gelombang phk yang uga	'uu', 'cipta', 'kerja', 'telah', 'ditolak', 'oleh', 'berbagai', 'lapisan', 'masyarakat', 'sejak', 'belum', 'disahkan', 'dan', 'mengakibatkan', 'gelombang', 'phk', 'yang', 'uga'
…	…….	……
…	……	……
2399	Halo sobatrbpn kementerian atrbpn melalui perpu nomor 2 tahun 2022 tentang cipta kerja akan mengatur pemanfaatan hak peng	'halo', 'sobatrbpn', 'kementerian', 'atrbpn', 'melalui', 'perpu', 'nomor', 'tahun', 'tentang', 'cipta', 'kerja', 'akan', 'mengatur', 'pemanfaatan', 'hak', 'peng'
2400	Agar tercapai pemahaman yang sinkron antar pemangku kepentingan kominfo terus melakukan sosialisasi uu no22020 tentang cip	'agar', 'tercapai', 'pemahaman', 'yang', 'sinkron', 'antar', 'pemangku', 'kepentingan', 'kominfo', 'terus', 'melakukan', 'sosialisasi', 'uu', 'no', 'tentang', 'cip'

c. Stopword

Stopword is the process of removing words that have no meaning contained in the tweet data. By previously carrying out case folding and tokenization and then doing stop words. Following are the results of stopwords in the dataset contained in Table 3.

Table 3. The stopword process

No	Tweet	Stopword result
1	Rezim pemerintahan saat ini banyak memunculkan kebijakan yang menyengsarakan rakyat dan hanya mementingkan diper	Rezim pemerintahan saat ini banyak memunculkan kebijakan yang menyengsarakan rakyat dan hanya mementingkan
2	UU cipta kerja telah ditolak oleh berbagai lapisan masyarakat sejak belum disahkan dan mengakibatkan gelombang phk yang uga	uu cipta kerja telah ditolak oleh berbagai lapisan masyarakat sejak belum disahkan dan mengakibatkan gelombang phk
…	….	….
…	……	……
2399	Halo sobatrbpn kementerian atrbpn melalui perpu nomor 2 tahun 2022 tentang cipta kerja akan mengatur pemanfaatan hak peng	Halo sobat kementerian atrbpn melalui perpu nomor 2 tahun 2022 tentang cipta kerja akan mengatur pemanfaatan hak
2400	Agar tercapai pemahaman yang sinkron antar pemangku kepentingan kominfo terus melakukan sosialisasi uu no22020 tentang cip	Agar tercapai pemahaman yang sinkron antar pemangku kepentingan kominfo terus melakukan sosialisasi uu no22020 tentang

3.2. Sentiment distribution

Sentiment distribution on data to identify and map tweet data on positive, negative and neutral sentiments, the purpose of the distribution is to see opinions contained in tweet data originating from social media with the keyword "cipta kerja". By distributing sentiment, you can find out the opinion patterns given by the community. The following is the sentiment distribution contained in Figure 2.

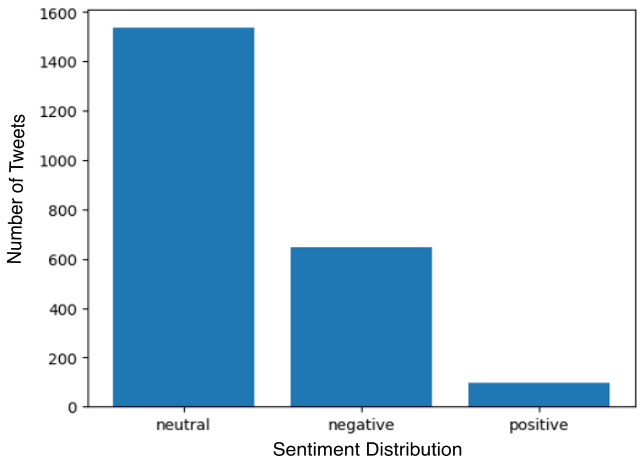

Figure 2. Sentiment distribution

Where Figure 2 is the distribution sentiment from a data warehouse in the form of data from Twitter. Information in Figure 2 contains data with neutral sentiment labels with a total of 1550, negative sentiment labels totaling 700 data and positive sentiment labels totaling 150 data out of the total data used totaling 2400. From the distributed data, training and testing can then be carried out with the SVM algorithm.

3.3. Test data sharing

The division of test data is the process of breaking data into two stages, namely data training and data testing, data training will be used to train the sentiment analysis model using the SVM algorithm while data testing will be used to test the model that has been formed. The following is a visualization of the distribution of test data in Figure 3. Figure 3 explains that in the process of building a sentiment analysis model using the SVM algorithm, there is a division of training and testing data, in the training data the percentage is 70% of all data and 30% for data testing of all data.

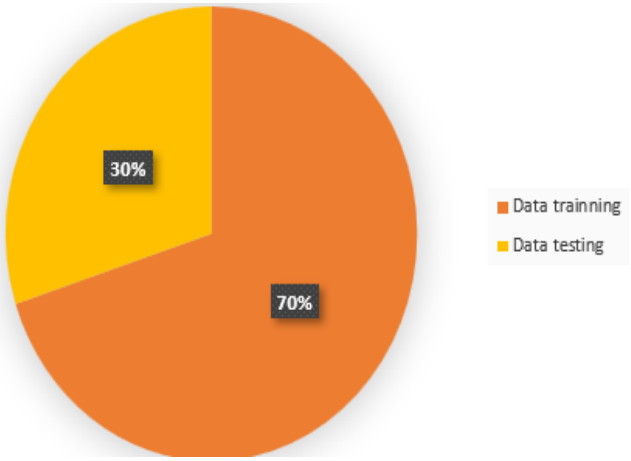

Figure 3. Distribution of data sharing

3.4. The results of the SVM model use the linear kernel

In the results of applying sentiment analysis to the SVM model using a linear kernal, it can be seen that the sentiment analysis model produces performance evaluations such as accuracy, precision, recall and f1 score. Following are the results of the SVM model using the linear kernel which can be seen from the results of the following confusion matrix in Figure 4. Description of Figure 4 shows the distribution of predictive data with a confusion matrix with positive, negative, and neutral values. The following is a Table 4 of manual confusion matrix calculations to find accuracy, recall, precision, and F1 score values.

Table 4. Manual confusion matrix calculation table with linier kernel

Actual	Positive predictions	Negative predictions	Neutral prediction
Positive	16	3	9
Negatif	0	166	28
Neutral	3	6	452

Based on Table 4, the calculation of accuracy, precision, recall and F1 score will be carried out as follows:

a. Accuracy

$$\text{Accuracy}=\frac{(TP+TN+TN)}{(TP+TN+FP+FN+TN+FP+FN+TN+TP)}x100\% = \frac{(16+166+425)}{(16+3+9+0+166+28+3+6+425)}x100\% = 92.3\%$$

b. Precision

$$\text{Positive}= \frac{TP}{(TP+FP)} = \frac{16}{(16+3)} = 0.842$$

$$\text{Negative}= \frac{TP}{(TN+FN)} = \frac{166}{(166+6)} = 0.965$$

$$\text{Neutral}= \frac{TP}{(TP+FP)} = \frac{452}{(452+28)} = 0.924$$

c. Recall

$$\text{Positive} = \frac{TP}{(TP+FN)} = \frac{16}{(16+9)} = 0.640$$

$$\text{Negative} = \frac{TP}{(TN+FP)} = \frac{166}{(166+28)} = 0.855$$

$$\text{Neutral} = \frac{TP}{(TP+FN)} = \frac{16}{(452+3+6)} = 0.984$$

d. F1 score

$$\text{Positive} = 2\ x\ \frac{(0.842 * 0.640)}{(0.842 + 0.640)} = 0.727$$

$$\text{Negative} = 2\ x\ \frac{(0.965 * 0.855)}{(0.965 + 0.855)} = 0.907$$

$$\text{Neutral} = 2\ x\ \frac{(0.924 * 0.984)}{(0.924 + 0.984)} = 0.953$$

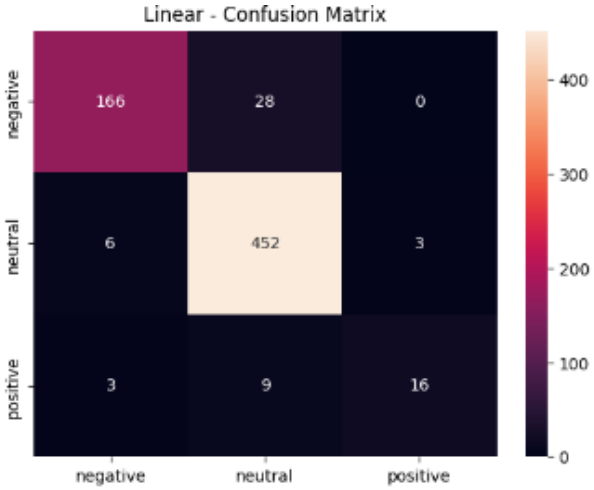

Figure 4. Confusion matrix

The receiver operating characteristic (ROC) curve aims to see the performance of the sentiment analysis model on tweet data with the keyword “cipta kerja”. The results of the ROC curve will produce a comparison graph between the positive class and other classes. Here are the results of the ROC curve contained in Figure 5.

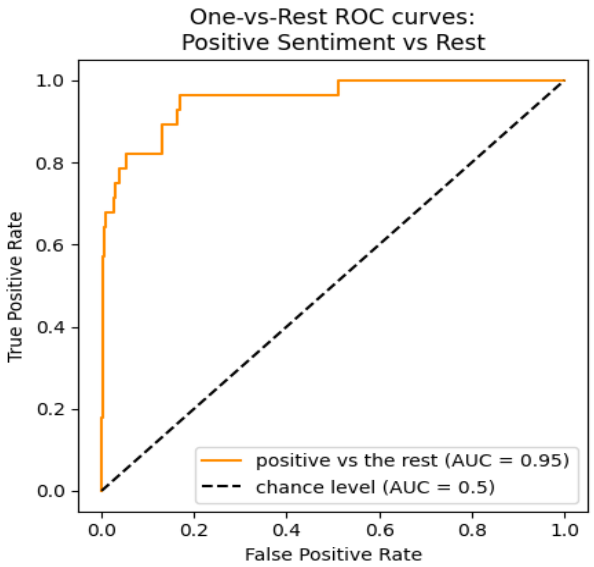

Figure 5. Result of the ROC curve SVM model with a linear kernel

3.5. The output of the SVM Model uses a cubic kernel

In the results of applying sentiment analysis to the SVM model using cubic kernel, it can be seen that the sentiment analysis model produces performance evaluations such as accuracy, precision, recall and f1 score. The SVM algorithm process certainly does not go through the calculations available from (1)–(4). The following are the results of the SVM model using cubic kernels which can be seen from the results of the following confusion matrix in Figure 6. The description of Figure 6 shows the distribution of predictive data with a confusion matrix with positive, negative and neutral values. The following is a Table 5 of manual confusion matrix calculations to find accuracy, recall, precision and F1 score values.

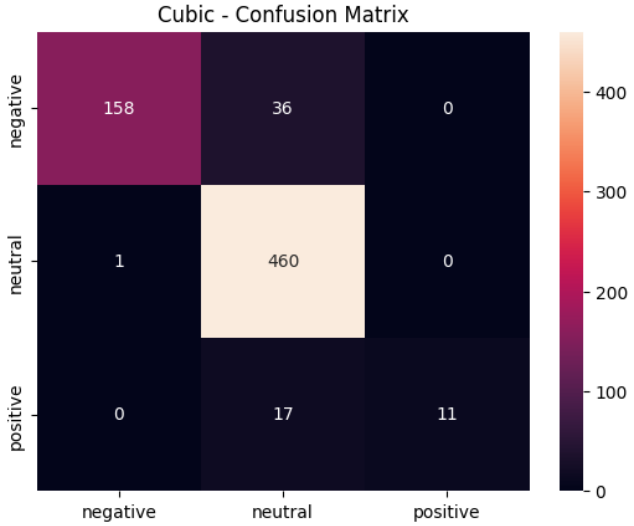

Figure 6. Confusion matrix kernel cubic

Table 5. Manual confusion matrix calculation table with cubic kernel

Actual	Positive predictions	Negative predictions	Neutral prediction
Positive	11	0	17
Negative	0	158	36
Neutral	0	1	460

Based on Table 5, the calculation of accuracy, precision, recall and F1 score will be carried out as follows:

a. Accuracy

$$\text{Accuracy}= \frac{(TP+TN+TN)}{(TP+TN+FP+FN+TN+FP+FN+TN+TP)} x100\% = \frac{(11+158+460)}{(11+0+17+0+158+36+0+1+460)} 100\% = 90\%$$

b. Precision

$$\text{Positive}= \frac{TP}{(TP+FP)} = \frac{11}{(11+0)} = 1$$

$$\text{Negative}= \frac{TP}{(TN+FN)} = \frac{158}{(158+1)} = 0.99$$

$$\text{Neutral}= \frac{TP}{(TP+FP)} = \frac{460}{(460+53)} = 0.89$$

c. Recall

$$\text{Positive}= \frac{TP}{(TP+FN)} = \frac{11}{(11+17)} = 0.39$$

$$\text{Negative}= \frac{TP}{(TN+FP)} = \frac{12}{(12+36)} = 0.25$$

$$\text{Neutral}= \frac{TP}{(TP+FN)} = \frac{194}{(194+1)} = 0.99$$

d. F1 score

$$\text{Positive}= 2\ x\ \frac{(1.0 * 0.3939)}{(1.0 + 0.3939)} = 0.56$$

$$\text{Negative}= 2\ x\ \frac{(0.9937 * 0.25)}{(0.9937 + 0.25)} = 0.39$$

$$\text{Neutral}= 2\ x\ \frac{(0.8961 * 0.9949)}{(0.8961 + 0.9949)} = 0.94$$

The ROC curve aims to see the performance of the sentiment analysis model on tweet data with the keyword job creation. The results of the ROC curve will produce a comparison graph between the positive class and other classes. Here are the results of the ROC curve in Figure 7.

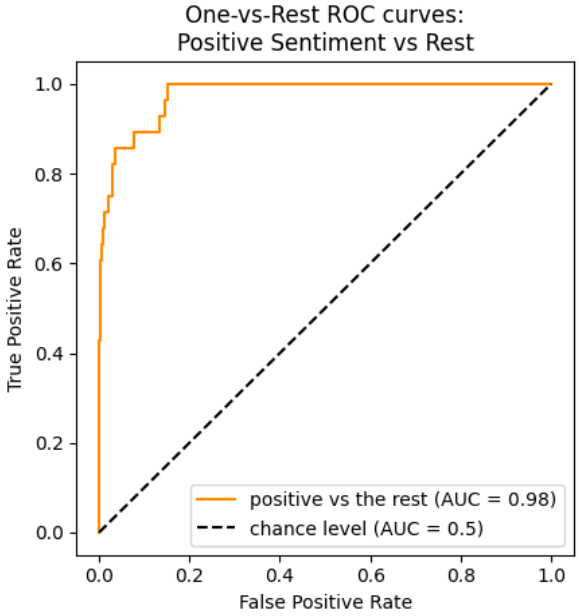

Figure 7. Result of the ROC curve SVM model with a cubic kernel

4. CONCLUSION

In this paper, cyberbullying has been detected on tweets from the word government policy in tweets. From the example of the word "cipta kerja" which fills comments from twitter users, it causes bullying so that cyberbullying is formed. However, the method used in detecting using SVM has been optimized on kernel cubic. The results show that the accuracy of the linear kernel is 92.3% while the cubic kernel function optimization gets an accuracy of 90%. ham shows optimal because the resulting accuracy is more optimal.

ACKNOWLEDGEMENTS

Thankful to the Rector of Universitas Muhammadiyah Sumatera Utara, Prof. Dr. Agussani, M.AP. and the LP2M UMSU with contract number 100/II.3-AU/UMSU-LP2M/C/2023.

REFERENCES

[1] X. Wang, B. Fang, H. Zhang, and X. Wang, "A dynamic model on news popularity prediction in online social networks," in *Proceedings of 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference, ITNEC 2019*, IEEE, Mar. 2019, pp. 847–851, doi: 10.1109/ITNEC.2019.8729161.

[2] K. S. K. Chung, M. Piraveenan, and S. Uddin, "Community evolution and engagement through assortative mixing in online social networks," in *Proceedings of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2012*, IEEE, Aug. 2012, pp. 724–725, doi: 10.1109/ASONAM.2012.130.

[3] G. Yavetz and N. Aharony, "Social media in government offices: usage and strategies," *Aslib Journal of Information Management*, vol. 72, no. 4, pp. 445–462, Apr. 2020, doi: 10.1108/AJIM-11-2019-0313.

[4] D. J. Haw, R. Pung, J. M. Read, and S. Riley, "Strong spatial embedding of social networks generates nonstandard epidemic dynamics independent of degree distribution and clustering," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 38, pp. 23636–23642, Sep. 2020, doi: 10.1073/pnas.1910181117.

[5] F. Darmawan, D. Fardiah, and R. Rinawati, "Utilization of Social Media in Developing Community-Based Internet on Village BT - 2nd Social and Humaniora Research Symposium (SoRes 2019)," *Atlantis Press*, 2020, pp. 277–282, doi: 10.2991/assehr.k.200225.058.

[6] V. Botti-Cebriá, E. d. Val, and A. García-Fornes, "Automatic Detection of Sensitive Information in Educative Social Networks BT- 13th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2020)," Á. Herrero, C. Cambra, D. Urda, J. Sedano, H. Quintián, and E. Corchado, Eds., Cham: Springer International Publishing, 2021, pp. 184–194.

[7] L. Wang, X. R. Luo, and M. P. Jurkat, "Understanding Inconsistent Corruption Control through E-government Participation: Updated Evidence from a Cross-Country Investigation," *Electronic Commerce Research*, vol. 22, no. 3, pp. 979–1006, Sep. 2022, doi: 10.1007/s10660-020-09444-x.

[8] L. Studen and V. Tiberius, "Social media, quo vadis? Prospective development and implications," *Future Internet*, vol. 12, no. 9, p. 146, Aug. 2020, doi: 10.3390/FI12090146.

[9] J. A. Obar and A. Oeldorf-Hirsch, "The biggest lie on the Internet: ignoring the privacy policies and terms of service policies of

social networking services," *Information Communication and Society*, vol. 23, no. 1, pp. 128–147, Jan. 2020, doi: 10.1080/1369118X.2018.1486870.
[10] T. I. Sari, Z. N. Ardilla, N. Hayatin, and R. Maskat, "Abusive comment identification on Indonesian social media data using hybrid deep learning," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 3, pp. 895–904, Sep. 2022, doi: 10.11591/ijai.v11.i3.pp895-904.
[11] K. C. LaFrancis and J. Putnam, "Cyberbullying and the Middle School Adolescent: Educator Perceptions," *Dissertation abstracts international. A, The humanities and social sciences*, vol. 81, no. 10-A, p. 223, 2019.
[12] Q. Jiang *et al.*, "Difficulties in Emotion Regulation and Cyberbullying Among Chinese Adolescents: A Mediation Model of Loneliness and Depression," *Journal of Interpersonal Violence*, vol. 37, no. 1–2, pp. NP1105–NP1124, Jan. 2022, doi: 10.1177/0886260520917517.
[13] A. Khowarizmi, Akhm, M. Lubis, and A. R. Lubis, "Classification of Tajweed Al-Qur'an on Images Applied Varying Normalized Distance Formulas," in *ACM International Conference Proceeding Series*, New York, NY, USA: ACM, Apr. 2020, pp. 21–25, doi: 10.1145/3396730.3396739.
[14] A. R. Lubis, M. Lubis, Al-Khowarizmi, and D. Listriani, "Big Data Forecasting Applied Nearest Neighbor Method," in *ICSECC 2019 - International Conference on Sustainable Engineering and Creative Computing: New Idea, New Innovation, Proceedings*, IEEE, Aug. 2019, pp. 116–120, doi: 10.1109/ICSECC.2019.8907010.
[15] A. Al-Khowarizmi, O. S. Sitompul, S. Suherman, and E. B. Nababan, "Measuring the Accuracy of Simple Evolving Connectionist System with Varying Distance Formulas," *Journal of Physics: Conference Series*, vol. 930, no. 1, p. 012004, Dec. 2017, doi: 10.1088/1742-6596/930/1/012004.
[16] Al-Khowarizmi, I. R. Nasution, M. Lubis, and A. R. Lubis, "The effect of a secos in crude palm oil forecasting to improve business intelligence," *Bulletin of Electrical Engineering and Informatics*, vol. 9, no. 4, pp. 1604–1611, Aug. 2020, doi: 10.11591/eei.v9i4.2388.
[17] S. Prayudani, A. Hizriadi, Y. Y. Lase, Y. Fatmi, and Al-Khowarizmi, "Analysis Accuracy of Forecasting Measurement Technique on Random K-Nearest Neighbor (RKNN) Using MAPE and MSE," *Journal of Physics: Conference Series*, vol. 1361, no. 1, 2019, doi: 10.1088/1742-6596/1361/1/012089.
[18] H. Nurrahmi and D. Nurjanah, "Indonesian Twitter Cyberbullying Detection using Text Classification and User Credibility," in *2018 International Conference on Information and Communications Technology, ICOIACT 2018*, IEEE, Mar. 2018, pp. 543–548, doi: 10.1109/ICOIACT.2018.8350758.
[19] D. Duchar, L. Costa, L. Dipippo, and L. Ha, "SVM Constraint Discovery using KNN applied to the Identification of Cyberbullying," *Proceedings of the 13th International Conference on Data Mining DMIN'17*, pp. 111–117.
[20] M. T. Ahmed, M. Rahman, S. Nur, A. Z. M. T. Islam, and D. Das, "Natural language processing and machine learning based cyberbullying detection for Bangla and Romanized Bangla texts," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 20, no. 1, pp. 89–97, Feb. 2022, doi: 10.12928/TELKOMNIKA.v20i1.18630.
[21] N. Potha and M. Maragoudakis, "Cyberbullying detection using time series modeling," in *IEEE International Conference on Data Mining Workshops, ICDMW*, IEEE, Dec. 2015, pp. 373–382, doi: 10.1109/ICDMW.2014.170.
[22] V. Nahar, S. Al-Maskari, X. Li, and C. Pang, "Semi-supervised Learning for Cyberbullying Detection in Social Networks BT - Databases Theory and Applications," H. Wang and M. A. Sharaf, Eds., Cham: Springer International Publishing, 2014, pp. 160–171.
[23] M. Andriansyah *et al.*, "Cyberbullying comment classification on Indonesian Selebgram using support vector machine method," in *Proceedings of the 2nd International Conference on Informatics and Computing, ICIC 2017*, IEEE, Nov. 2017, pp. 1–5, doi: 10.1109/IAC.2017.8280617.
[24] Noviantho, S. M. Isa, and L. Ashianti, "Cyberbullying classification using text mining," in *Proceedings - 2017 1st International Conference on Informatics and Computational Sciences, ICICoS 2017*, IEEE, Nov. 2017, pp. 241–245, doi: 10.1109/ICICOS.2017.8276369.
[25] S. Widodo, R. N. Rohmah, B. Handaga, and L. D. D. Arini, "Lung diseases detection caused by smoking using support vector machine," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 17, no. 3, pp. 1256–1266, Jun. 2019, doi: 10.12928/TELKOMNIKA.V17I3.9799.
[26] A. M. Abd and S. M. Abd, "Modelling the strength of lightweight foamed concrete using support vector machine (SVM)," *Case Studies in Construction Materials*, vol. 6, pp. 8–15, Jun. 2017, doi: 10.1016/j.cscm.2016.11.002.
[27] M. Wang and H. Chen, "Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis," *Applied Soft Computing Journal*, vol. 88, p. 105946, Mar. 2020, doi: 10.1016/j.asoc.2019.105946.
[28] B. Xiang, J. Yao, X. Lu, L. Li, and R. Xie, "Segmentation-Based classification for 3D urban point clouds," in *2016 IEEE International Conference on Information and Automation, IEEE ICIA 2016*, IEEE, Aug. 2017, pp. 172–177, doi: 10.1109/ICInfA.2016.7831817.
[29] S. Huang, C. A. I. Nianguang, P. P. Pacheco, S. Narandes, Y. Wang, and X. U. Wayne, "Applications of support vector machine (SVM) learning in cancer genomics," *Cancer Genomics and Proteomics*, vol. 15, no. 1, pp. 41–51, Jan. 2018, doi: 10.21873/cgp.20063.
[30] V. S. S. S. S. V. Ranganathan, "Fire Detection Using Support Vector Machines (SVM)," *International Journal of Science and Research (IJSR)*, vol. 6, no. 5, pp. 1607–1618, 2017.
[31] I. Ahmad, M. Basheri, M. J. Iqbal, and A. Rahim, "Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection," *IEEE Access*, vol. 6, pp. 33789–33795, 2018, doi: 10.1109/ACCESS.2018.2841987.
[32] A. F. Agarap, "An Architecture Combining Convolutional Neural Network (CNN) and Support Vector Machine (SVM) for Image Classification," pp. 5–8, 2017.
[33] R. Varatharajan, G. Manogaran, and M. K. Priyan, "A big data classification approach using LDA with an enhanced SVM method for ECG signals in cloud computing," *Multimedia Tools and Applications*, vol. 77, no. 8, pp. 10195–10215, Apr. 2018, doi: 10.1007/s11042-017-5318-1.
[34] I. Aljarah, A. M. Al-Zoubi, H. Faris, M. A. Hassonah, S. Mirjalili, and H. Saadeh, "Simultaneous Feature Selection and Support Vector Machine Optimization Using the Grasshopper Optimization Algorithm," *Cognitive Computation*, vol. 10, no. 3, pp. 478–495, Jun. 2018, doi: 10.1007/s12559-017-9542-9.
[35] M. A. Elgammal, H. Mostafa, K. N. Salama, and A. Nader Mohieldin, "A Comparison of Artificial Neural Network(ANN) and Support Vector Machine(SVM) Classifiers for Neural Seizure Detection," in *Midwest Symposium on Circuits and Systems*, IEEE, Aug. 2019, pp. 646–649, doi: 10.1109/MWSCAS.2019.8884989.
[36] S. Khan, R. Ullah, A. Khan, N. Wahab, M. Bilal, and M. Ahmed, "Analysis of dengue infection based on Raman spectroscopy and support vector machine (SVM)," *Biomedical Optics Express*, vol. 7, no. 6, p. 2249, Jun. 2016, doi: 10.1364/boe.7.002249.
[37] E. Tuba and Z. Stanimirovic, "Elephant herding optimization algorithm for support vector machine parameters tuning," in

Proceedings of the 9th International Conference on Electronics, Computers and Artificial Intelligence, ECAI 2017, IEEE, Jun. 2017, pp. 1–4, doi: 10.1109/ECAI.2017.8166464.
[38] A. Ambarwari, Q. J. Adrian, Y. Herdiyeni, and I. Hermadi, "Plant species identification based on leaf venation features using SVM," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 2, pp. 726–732, Apr. 2020, doi: 10.12928/TELKOMNIKA.V18I2.14062.

BIOGRAPHIES OF AUTHORS

Al-Khowarizmi was born in Medan, Indonesia, in 1992. He is a Dean in Faculty of Computer Science and Information Technology at Universitas Muhammadiyah Sumatera Utara (UMSU). He got doctoral degree from Universitas Sumatera Utara in 2023 field Computer Science. His main research interest is data science, big data, machine learning, neural network, artificial intelligence, and business intelligence. He can be contacted at email: alkhowarizmi@umsu.ac.id.

Indah Purnama Sari received a master's degree in Computer Science, University of North Sumatra (USU) in 2019. Currently she is a lecturer in the Information Technology, Faculty of Computer Science and Information Technology, Universitas Muhammadiyah Sumatera Utara. She is a lecturer and teaches software engineering courses. Actively writes articles and becomes a speaker in international and national conference seminars. She can be contacted at email: indahpurnama@umsu.ac.id.

Halim Maulana was born in Medan, Indonesia, in 1991. He is a Vice Dean in Faculty of Computer Science and Information Technology at Universitas Muhammadiyah Sumatera Utara (UMSU). He got master from Universitas Sumatera Utara in 2016 field Computer Science. His main research interest is networking, data science, big data, machine learning, neural network, and artificial intelligence. He can be contacted at email: halimmaulana@umsu.ac.id.