Comparison of SVM and LIWC for Sentiment Analysis of SARA

AAIN Eka Karyawati*[1], Prasetyo Adi Utomo[2], I Gede Arta Wibawa[3]
[1,2,3]Informatics Studi Program, FMIPA, Universitas Udayana, Bali, Indonesia
e-mail: ***[1]eka.karyawati@unud.ac.id**, [2]pras.au404@gmail.com, [3]gede.arta@unud.ac.id

Abstrak

SARA adalah isu sensitif berdasarkan sentimen tentang identitas diri tentang keturunan, agama, kebangsaan atau etnis. Dampak dari isu SARA adalah konflik antar kelompok yang berujung pada kebencian dan perpecahan. Isu SARA banyak tersebar melalui media sosial, khususnya Twitter. Untuk mengatasi masalah SARA, perlu dikembangkan metode yang efektif untuk menyaring SARA negatif. Penelitian ini bertujuan untuk menganalisis tweet berbahasa Indonesia dan menentukan apakah tweet tersebut mengandung SARA positif atau negatif atau tidak mengandung SARA (netral). Metode berbasis pembelajaran mesin (yaitu, SVM) dan metode berbasis leksikon (yaitu, LIWC) dibandingkan berdasarkan 450 data tweet untuk menentukan pendekatan terbaik untuk setiap sentimen (positif, negatif, dan netral). Hasil evaluasi terbaik ditunjukkan pada klasifikasi SARA negatif menggunakan SVM dengan $\lambda = 3$ dan $\gamma = 0.1$, dimana Precision = 0.9, Recall = 0.6, dan F1-Score = 0.72. Hasil terbaik dari klasifikasi SARA positif ditunjukkan pada metode LIWC, dimana Precision = 0.6, Recall = 0.8, dan F1-Score = 0.69. Hasil evaluasi terbaik untuk klasifikasi netral ditunjukkan pada SVM dengan $\lambda = 3$ dan $\gamma = 0.1$, dengan Precision = 0.52, Recall = 0.87, dan F1-Score = 0.65.

Kata kunci—*SVM, LIWC, Sentiment Analysis, SARA*

Abstract

SARA is a sensitive issue based on sentiments about self-identity regarding ancestry, religion, nationality or ethnicity. The impact of the issue of SARA is conflict between groups that leads to hatred and division. SARA issues are widely spread through social media, especially Twitter. To overcome the problem of SARA, it is necessary to develop an effective method to filter negative SARA. This study aims to analyze Indonesian-language tweets and determine whether the tweet contains positive or negative SARA or does not contain SARA (neutral). Machine learning (i.e., SVM) and lexicon-based method (i.e., LIWC) were compared based on 450 tweet data to determine the best approach for each sentiment (positive, negative, and neutral). The best evaluation results are shown in the negative SARA classification using SVM with $\lambda = 3$ and $\gamma = 0.1$, where Precision = 0.9, Recall = 0.6, and F1-Score = 0.72. The best results from the positive SARA classification were shown in the LIWC method, where Precision = 0.6, Recall = 0.8, and F1-Score = 0.69. The best evaluation results for neutral classification are shown in SVM with $\lambda = 3$ and $\gamma = 0.1$, with Precision = 0.52, Recall = 0.87, and F1-Score = 0.65.

Keywords—*SVM, LIWC, Sentiment Analysis, SARA*

1. INTRODUCTION

SARA are the sensitive issues based on sentiments about self-identity regarding descent, religion, nationality or ethnicity and class. The impact of the issue of SARA is conflict between groups that leads to hatred and division. SARA issues are widely spread through social media, especially Twitter. To overcome SARA problems, it is necessary to develop an effective method to filter out negative SARA.

SARA can be divided into three categories: 1) individual SARA with actions that offend, harass, discriminate, or insult other groups; 2) Institutional SARA with actions through rules or policies that are discriminatory against a group; and 3) Culture SARA with spreading discriminatory traditions or ideas between groups.

This research was based on tweet data about individual SARA. There are several approaches to analyzing SARA issues of Indonesian tweets [1-3]. [1] used LIWC to analyze three classes of SARA sentiments (i.e., negative, neutral, and positive). The results showed that the average values of *Precision, Recall,* and *F1-Score* were 69.62%, 70%, and 69.81%, respectively. [2] analyzed the two classes of SARA sentiments (i.e., negative, and positive) using the Improved-KNN method. This study shows good results in term of *Precision, Recall* and *F1-Score* of 0.976422, 1, and 0.987944444, respectively. SARA sentiment analysis was also carried out by [3] using a combination of k-means and SVM. SVM was used to classify two classes of SARA sentiments. The results obtained based on k-means include 118 positive SARA tweets and 83 negative SARA tweets where the results showed *Precision* and *Recall* values of 64.18% and 63.68%, respectively.

Quite a lot of research on sentiment analysis using the Support Vector Machine (SVM) method [3-6]. [4] compared the SVM and Naive Bayes methods to analyze three classes of Covid-19 sentiment (negative, neutral, and positive) of Indonesian tweets. This study showed that SVM results were better than Naive-Bayes method, where the average *F1-Scores* were 93% and 92%, respectively. [5] analyzed two classes of sentiments about radical content (i.e., positive and negative) of Indonesian tweets using SVM method with a polynomial degree 2 kernel. The best accuracy value from this study was 70% with $\lambda = 0.1$ and $\gamma = 0.1$. [6] analyzed the performance of the SVM method for sentiment analysis. This study uses two tweet datasets (about self-driving cars and about apple products). The results showed that the average *Precision, Recall* and *F1-Score* values were 55.8%, 59.9% and 57.2%, respectively, for the first dataset, and 70.2%, 71.2% and 69.9, respectively, for the second dataset.

There were very few studies using Linguistic Inquiry and Word Count (LIWC) for sentiment analysis [1, 7, 8]. [7] applied LIWC as a booster feature for improving machine learning performance for analyzing sentiment of hate speech in the English and Spanish. The result showed improvement of the baseline machine learning method without LIWC feature, with the average of accuracy are 0.7 and 0.8 for the English and Spanish corpora, respectively. [8] compared the LIWC and some machine learning methods on different social media datasets of positive and negative sentiment reviews. The results showed that LIWC performed poorly compared with machine learning.

However, to the best of my knowledge, comparative studies between machine learning and lexicon-based methods in sentiment analysis especially for SARA (i.e., ancestry, religion, nationality or ethnicity) were not yet available. Most of the comparative studies conducted for sentiment analysis but not for SARA were comparisons between several machine learning-based methods [9, 10, 11]. Moreover, most of the sentiment analysis studies for SARA only classify sentiment into two classes, namely positive and negative [2, 3].

In this research, we compared the machine learning (i.e., SVM) and the lexicon-based method (i.e., LIWC) for analyzing three classes of SARA sentiment (i.e., negative, neutral, and positive) of Indonesian tweets. Some experiments have been conducted to select the best SVM model. Moreover, some new lexicon has been added to improve the performance of LIWC method.

2. METHODS

2.1 Collecting and Labeling Data

2.1.1 Collecting Data

      The data collection process was carried out by collecting data on Indonesian-language tweets. Tweet data used is primary data obtained by using the TwitterAPI library 'tweepy' in Python. Tweet data retrieval is done using specific keywords of ethnicity and religion. The tweet data collected are those that contain SARA issues, negative and positive sentiments and those that do not contain SARA issues. Examples of tweet data collected are shown in Table 1, Table 2, and Table 3.

Table 1 Examples of Positive SARA Tweet Data

No	Tweets
1	*Umat Kristen diajarkan berbuat baik kepada siapa pun sebab dasar ajaran Kristen adalah KASIH, bahkan harus mengasihi 'musuh'! Jika takut karena perbuatan baik berupa bantuan, maka bantulah umatmu, jangan biarkan miskin* Christians are taught to do good to anyone because the basis of Christian teaching is LOVE, even to love the 'enemy'! If you are afraid because of a good deed in the form of help, then help your ummah, don't let it be poor
2	*Lagi heboh keluarga beda agama. Well.. gue kristen, adek muslim, kakak kristen terus mualaf... kami saling mengasihi, tetapi gue nggak bisa mengatakan itu adalah keindahan, terlebih sesuatu yang mesti dicapai. Tanpa mengurangi rasa sayang, menurut gue iman adalah dasar family.* There's a lot of excitement about families of different religions. Well.. I'm a Christian, my younger brother is a Muslim, my older brother is a Christian and a convert... we love each other, but I can't say it's beauty, especially something that must be achieved. Without reducing affection, in my opinion faith is the basis of family.

Table 2 Examples of Negative SARA Tweet Data

No	Data Tweet SARA Negatif
1	*woiii.goblok.... kau islam apa manusia penyembah berhala kau njing ??? kau hina bendera rasulku, sama saja kau meludahi wajah umat islam.. dasar kristen anjing kau babi...sempat kau disini, udah kubunuh kau njing !!!!* wow. stupid.... are you a muslim or an idol worshiper are you a dog??? you insult the flag of my apostle, it's the same as spitting on the faces of Muslims.
2	*Dasar bahasa madura tak bermazhab.* Indeed, the Madurese language has no sect.

Table 3 Examples of Non-SARA (Neutral) Tweet Data

No	Data Tweet Bukan SARA
1	*Entah ide pemerintah atau bukan kyk makin hari makin Menggila gila* Tweet Data Not SARA Whether it's the government's idea or not, it seems to be getting more and more crazy
2	*Jagung Boseh, makanan olahahan berbahan dasar Jagung sbg makanan pokok pengganti nasi di Kupang. Cc: @Info_Kupang* Maize Boseh, processed food made from corn as a staple food substitute for rice in Kupang. Cc: @Info_Kupang

2.1.2 Labeling Data

The data collected is 450 tweets with a composition of 150 data for each class. Data collection was carried out for four weeks starting from 25-02-2021 to 25-03-2021. The data obtained were selected and stored in a file with .csv extension with 2 columns, the first column is to store the sentiment class (positive, neutral, and negative) and the second column is to store the tweet text. The results of the tweets that have been obtained were labeled into three classes by two labelers. To validate and determine data consistency, Cohen's Kappa test was performed [12]. From the tests that have been carried out by 2 labelers, a Kappa value of 0.8 was obtained that means the level of approval of data labeling was said to be "strong".

About 90% of the total data, which is 405, is used in the process of selecting the best model in the SVM classification method using 5-fold cross validation, and 45 data or 10% of the total data is used in the process of evaluating the best model of SVM and the LIWC method.

2.2 Feature extraction

This study uses the Bag-of-Word model, using vocabulary terms from the tweet data set. LIWC only uses the word feature. On the other hand, SVM uses TF-IDF weights as a feature.

2.2.1 Preprocessing

The preprocessing stages carried out in this study were tokenization, case folding, data cleaning, stopword removal, and ended with stemming [13].

2.2.2 TF-IDF weighting

Based on [13], the term weighting ($Wf_{t,d}$) began by calculating the term frequency (*tf*). To reduce the value of terms that often appear in many documents, the inverse-document frequency (*idf*) value was calculated, where N is the total number of documents. Furthermore, the term weighting was carried out by calculating the TF-IDF. The *tf*, *idf*, and $Wf_{t,d}$ are calculated as follows [13],

$$Wf_{t,d} = \begin{cases} 1 + \log tf_{t,d}, & jika\ f_{t,d} > 0 \\ 0, & jika\ f_{t,d} \leq 0 \end{cases} \tag{1}$$

$$idf_t = \log \frac{N}{df_t} \tag{2}$$

$$Wtf\text{-}idf_{t,d} = Wf_{t,d} * idf_t \tag{3}$$

2.3 SVM

SVM is a linear classification method. The main principle of SVM in classifying is to determine a separator in the search space that can separate different classes which is commonly called a hyperplane. One of the advantages of the SVM method is that it is quite good at classifying high-dimensional data because this method tries to determine the optimal direction of discrimination in the feature space by checking the right combination of features [14]. Classification with SVM goes by specifying the document class. In this research, SARA class documents are classified and if the value is -1 then it is a negative class, if it is 1 then it is a positive class, and if it is 0 it is a neutral class.

2.3.1 Training

In the training process, the combination of model parameters, namely *C* (slack variable), *d* (degree polynomial), *γ* (learning rate), and *λ* (error control) is determined manually, and adjusted to get the optimal value (the best SVM model) via *k*-fold validation. The SVM kernel used in this study is a polynomial kernel with the following equation [15].

$$K\ (x_i,\ x_j) = \left((x_i, x_j) + C\right)^d, \gamma > 0 \tag{4}$$

With K$(x_i,\ x_j)$ is kernel fuction, x_i is *i*-th data, x_j is *j*-th data. The steps of the SVM method are based on equation (5) – equation (10), as follows [15].

a. Initiation of parameters used.
b. Calculate the Hessian matrix.

$$D_{ij} = y_i\ y_j\left(K(x_i, x_j) + \lambda^2\right) \tag{5}$$

With D$_{ij}$ is Hessian matrix value, y_i is i-th class, y_j is *j*-th class.

c. Starting from the 1st data to the nth data, do the calculation iterations.

$$\varepsilon_i = \sum_{j=1}^{n} \alpha_j D_{ij} \tag{6}$$

$$\delta\alpha_i = min\{max[\gamma(1 - \varepsilon_i), -\alpha_i], C - \alpha_i\} \tag{7}$$

$$\alpha_i = \alpha_i + \delta\alpha_i \tag{8}$$

With ε is error value and α_i is support vector.

d. From the previous calculation, the largest value of α_i is sought and calculations are carried out to determine the bias.

$$b = -\frac{1}{2}\left[\left(\sum_{i=1}^{n}\alpha_i y_i K(x_i, x^-)\right) + \left(\sum_{i=1}^{n}\alpha_i y_i K(x_i, x^+)\right)\right]$$

with *b* is for bias value.

e. Finally, the classification model is defined as follow,

$$f(x) = \sum_{i=0}^{n}\alpha_i\ y_i K(x_i, x) + b \tag{10}$$

with *f(x)* is the predicted label.

2.3.1 Testing

The testing process is carried out to validate the classification model. In this study, validation was carried out by 5-fold cross validation using 405 datasets. The actual label is compared with the predicted label shown in equation (10). The best model was selected based on the *F1-Score* value. The best SVM model is the highest *F1-Score* among some combinations of *d* (degree polynomial), *γ* (learning rate), and *λ* (error control).

The *F1-Measure* is calculated based on the coincidence matrix [16]. Table 3 shows the coincidence matrix for the classification of the three classes.

Tabel 3 Coincidence Matrix

	Actual Label 1	Actual Label 2	Actual label 3
Predicted Label 1	TK11	FK12	FK13
Predicted Label 2	FK21	TK22	FK23
Predicted Label 3	FK31	FK32	TK33

The *F1-Measure* formula based on Table 2.1 given by equation (13) [16].

$$Precision = \frac{\frac{TK1}{TK11+FK12+FK13}+\frac{TK22}{FK21+TK22+FK23}+\frac{TK33}{FK31+FK32+TK33}}{3} \quad (11)$$

$$Recall = \frac{\frac{TK11}{TK11+FK21+FK21}+\frac{TK22}{FK12+TK22+FK32}+\frac{TK33}{FK13+FK23+TK33}}{3} \quad (12)$$

$$F\text{-}Measure = \frac{2}{\frac{1}{Precision}+\frac{1}{Recall}} \quad (13)$$

2.4 LIWC

LIWC is a text analysis application developed with the aim of analyzing the emotional, cognitive, and structural components of a text [17]. LIWC works by searching for each word in the text and matching it to a word in the lexicon. The lexicon contains words that fall into categories that reflect the word linguistically, psychologically, and socially, such as pronouns, positive emotions, social processes, and so on. LIWC adds a word category percentage value if a category is found [17]. If all the words in the text document have been categorized, the results will be displayed in the form of a table containing the category percentage values for the text document [8]. The steps taken to perform sentiment analysis using LIWC were as follows [17].

1. Read all terms in the document
2. Count the frequency of terms that are part of each class label defined in the dictionary (positive and negative)
3. Calculate the class ratio with equation (14).

$$R_k = \frac{\sum n_{t,k}}{N_d} * 100\% \quad (14)$$

With R_k is the class ratio k (positive or negative), $n_{t,k}$ is frequency of term t in document and term t is a member of class k in dictionary, and N_d is frequency of the total term in the document.

4. Determine the class label, if class ratio positive > negative then class = 1, if class ratio negative > positive then class = -1, and if class ratio positive = negative then class = 0.

In this research, the LIWC lexicon used was obtained from the translation of Liu's opinion word list by Wahyu & Azhari [18,19], which contains 1182 positive words and 2402 negative words. The dictionary was then modified by adding new words so that it contained 1256 positive words and 2463 negative words.

2.5 Evaluation

Some experiments were conducted to evaluate the performance of the SVM and LIWC method using three evaluation metrics which are *Precision, Recall,* and *F1-Measure* measured by equation (11), (12), and (13), respectively.

3. RESULTS AND DISCUSSION

3.1 Model Selection of the SVM Method

The model selection was done by doing a 5-fold cross validation. The process started by using *d* (polynomial degree) = 1 and *γ* (learning rate) = 0.1. Table 5 presents the *F-1 Score* for each fold and its average value based on changes in *λ* (error control). As can be seen, the highest average F1-Score (= 0.6477) was obtained at *λ* = 3.

Table 5 The Effects of *λ* Changes to F1-Score Values

λ	F1-Score of Fold 1	F1-Score of Fold 2	F1-Score of Fold 3	F1-Score of Fold 4	F1-Score of Fold 5	AVG of F1-Score
0.01	0.6265	0.6556	0.4862	0.5582	0.5873	0.5828
0.1	0.6198	0.6742	0.5109	0.5582	0.5873	0.5901
1	0.6319	0.6377	0.5459	0.607	0.671	0.6187
2	0.6804	0.6591	0.57	0.6094	0.6828	0.6404
3	**0.7179**	**0.6599**	**0.5816**	**0.5846**	**0.6943**	**0.6477**
4	0.7181	0.6952	0.5515	0.5856	0.6711	0.6443
5	0.7181	0.6881	0.5377	0.5592	0.6837	0.6374
10	0.4442	0.4089	0.5344	0.4062	0.4868	0.4561

Furthermore, the process was continued by using *d* = 1 and *λ* = 3, to see the effect of changes in the value of *γ*. Table 6 shows the F-1 score for each fold and its average value. The highest average *F1-Score* (=0.6477) was obtained when *γ* = 0.1.

Table 6 The Effects of *γ* Changes to F1-Score Values

γ	F1-Score of Fold 1	F1-Score of Fold 1	F1-Score of Fold 1	F1-Score of Fold 1	F1-Score of Fold 1	AVG of F1-Score
0.05	0.665	0.6715	0.5137	0.577	0.6571	0.6169
0.1	**0.7179**	**0.6599**	**0.5816**	**0.5846**	**0.6943**	**0.6477**
0.2	0.6690	0.6503	0.5702	0.5974	0.6588	0.6291
0.3	0.6318	0.6376	0.5459	0.6070	0.6709	0.6186
0.4	0.6691	0.6508	0.5347	0.6556	0.6583	0.6337
0.5	0.6798	0.6627	0.5459	0.6569	0.6583	0.6407
0.6	0.6549	0.6739	0.5314	0.6697	0.6591	0.6378
0.7	0.6433	0.6739	0.5314	0.6697	0.6472	0.6331
0.8	0.6433	0.6739	0.5314	0.6446	0.6472	0.6281
0.9	0.6433	0.6739	0.5422	0.6449	0.6472	0.6303
1	0.6433	0.6862	0.5437	0.6320	0.6370	0.6284

However, experiments conducted for the SVM with degree 2 polynomial kernel (*d* = 2) resulted in poor performance. The best degree 2 polynomial SVM model was obtained using *λ* = 3 and *γ* = 0.1, which resulted *F1-Score* of 0.5381.

3.2 The SVM Evaluation Results

The best SVM model was SVM with *d* = 1, *γ* = 0.1 and *λ* = 3, evaluated based on 45 test datasets. Table 7 presents the evaluation results of *Precision, Recall,* and *F1-Score* for each class label.

Table 7 The evaluation results of the SVM

Evaluation Metrics	Positive	Neutral	Negative	Average
Precision	0.80	0.52	0.90	0.74
Recall	0.53	0.87	0.60	0.67
F1-Score	0.64	0.65	0.72	0.67

As can be seen, the negative class *Precision* was the highest value (= 0.9), meaning that about 90% of the negative SARA tweets obtained were correct. However, the negative class *Recall* was low (= 0.6), meaning that only 60% of the negative SARA tweets were taken. High neutral class *Recall* (= 0.87), meaning that it obtained about 87% non-SARA tweets, but low *Precision* (0.52), this indicates that only 52% of non-SARA tweets obtained were correct. In addition, the *Precision* of positive class was also high (0.8), meaning that 80% of the positive SARA tweets obtained were correct, but only 53% positive SARA tweets were captured (*Recall* = 0.53).

The overall results of SVM's performance in the sentiment analysis of Indonesian-language tweets contained SARA issues were not good in terms of *F1-Score*. The average *F1-Score* values for positive, neutral, and negative sentiment were 0.64, 0.65, and 0.72, respectively. Poor SVM performance is likely due to the data features used in the SVM method. In this study, text data was only represented by the document terms from the TF-IDF weight matrix. It might be better to take a semantic representation approach, for example by incorporating the LIWC lexicon as a feature.

3.3 Lexicon Modification of the LIWC Method

To determine the effect of the lexicon on the LIWC classification, a comparison of the results of the classification between LIWC with the lexicon before modification and with the lexicon after modification was carried out. Table 8 and Table 9 respectively show the results of the evaluation before and after the modification.

Table 8 The evaluation results before the new LIWC lexicon added

Evaluation Metrics	Positif	Netral	Negatif	Rata-Rata
Precision	0.5931	0.4787	0.5857	0.5525
Recall	0.5733	0.5266	0.5466	0.5488
F1-Score	0.5830	0.5015	0.5655	0.5501

Table 9 The evaluation results after the new LIWC lexicon added

Evaluation Metrics	Positif	Netral	Negatif	Rata-Rata
Precision	0.60	0.67	0.69	0.65
Recall	0.80	0.53	0.60	0.64
F1-Score	0.69	0.59	0.64	0.64

As can be seen, there were some improvements in terms of *F1-Score* for each class label. The highest increase occurred in the positive class. *F1-Score* increased from 0.5830 to 0.69 or increased by 18.35%. *F1-Score* for the neutral class showed an increase of 17.65%. For the negative class, there was also an increase of 13.17% in term of *F1-Score*.

Similar to the performance of the SVM method, the overall results of LIWC showed poor performance, where the average value of F1-Score was 0.64. This value was worse than the SVM result.

Improving the performance of the LIWC method by adding several new lexicon provides a possible way how to improve the performance of sentiment analysis by progressively adding new relevant lexicon for each sentiment class.

4. CONCLUSIONS

In this study, a comparison between machine learning-based methods (i.e., SVM) and lexicon-based methods (i.e., LIWC) for analyzing the sentiment of Indonesian-language tweets showed that machine learning was better, but both approaches perform poorly overall. The best evaluation results were shown in the negative SARA classification with polynomials degree 1 SVM, $\lambda = 3$ and $\gamma = 0.1$, where Precision = 0.9, Recall = 0.6, and F1-Score = 0.72. The best results from the positive SARA classification test were shown in the LIWC method, where Precision = 0.6, Recall = 0.8, and F1-Score = 0.69. The best evaluation results for neutral classification are shown on polynomials degree 1 SVM, $\lambda = 3$ and $\gamma = 0.1$, with Precision = 0.52, Recall = 0.87, and F1-Score = 0.65.

Poor SVM performance was likely due to the data features used in the SVM method. In this study, text data was only represented by the document terms from the TF-IDF weight matrix. It might be better to take a semantic representation approach, for example by incorporating the LIWC lexicon as a feature.

On the other hand, improving the performance of the LIWC method by adding several new lexicon provides a possible way how to improve the performance of sentiment analysis by progressively adding new relevant lexicon for each sentiment class.

REFERENCES

[1]   P. A. Utomo and AAIN E. Karyawati, "Sentiment Analysis Of Tribal, Religion, And Race With LIWC" JELIKU (*Jurnal Elektronik Ilmu Komputer Udayana*) Vol. 9, No. 3. February 2021 [Online].
Available: https://ojs.unud.ac.id/index.php/JLK/article/view/64361/38478

[2]   F. R. Saputra, Indriati, and Sutrisno, "Klasifikasi Isu Suku, Antar Golongan, Ras, Agama (SARA) pada Twitter Berbahasa Indonesia menggunakan Metode Improved K-Nearest Neighbor (K-NN)" JPTIIK (*Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*),Vol. 4, No. 1, pp. 373-380, Januari 2020 [Online].
Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/6917/3359

[3]   W. M. Baihaqi, M. Pinilih, and M. Rohmah, "Kombinasi k-Means dan Support Vector Machine (SVM) untuk Memprediksi Unsur SARA pada Tweet", JPTIIK (*Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*), Vol. 7, No. 3, pp. 501-510, Juni 2018 [Online]. Available: https://jtiik.ub.ac.id/index.php/jtiik/article/view/2126/pdf

[4]   F. J. Damanik1, and D. B. Setyohadi, "Analysis Of Public Sentiment About Covid-19 In Indonesia On Twitter Using Multinomial Naive Bayes And Support Vector Machine", IOP Conf. Ser.: Earth Environ. Sci. 704, 2021 [Online].
Available: https://iopscience.iop.org/article/10.1088/1755-1315/704/1/012027/pdf

[5]   F. Alvianda and P. P. Adikara, "Analisis Sentimen Konten Radikal Di Media Sosial Twitter Menggunakan Metode Support Vector Machine (SVM)," JPTIIK (*Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*), Vol. 3, No. 1, pp. 241–246, 2019 [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4084/1877

[6]   M. Ahmad, S. Aftab, and I. Ali, "Sentiment Analysis of Tweets using SVM", IJCA (*International Journal of Computer Applications*), Vol. 177, No.5, November 2017 [Online]. Available: https://www.ijcaonline.org/archives/volume177/number5/ahmad-2017-ijca-915758.pdf

[7]   R. Cervero, "Use of Lexical and Psycho-Emotional Information to Detect Hate Speech Spreaders on Twitter", CLEF 2021 – Conference and Labs of the Evaluation Forum, 2021 [Online]. Available: http://ceur-ws.org/Vol-2936/paper-161.pdf

[8]   J. Hartmann, J. Huppertz, C. Schamp, and M. Heitmann, "Comparing Automated Text classification Methods", IJRM (*International Journal of Research in Marketing*), Vol 36 pp. 20–38, 2019 [Online].
        Available: https://www.sciencedirect.com/science/article/pii/S0167811618300545

[9]   S. Qaiser, N. Yusoff, R. Ali, M.A. Remli, and H.K. Adli, "A Comparison of Machine Learning Techniques for Sentiment Analysis",*Turkish Journal of Computer and Mathematics Education*, Vol. 12, No. 3, pp. 1738-1744, 2021 [Online].
        Available: https://turcomat.org/index.php/turkbilmat/article/view/999/788

[10]   M.T.H.K. Tusar and M.T. Islam, "A Comparative Study of Sentiment Analysis Using NLP and Different Machine Learning Techniques on US Airline Twitter Data" Proceeding of the International Conference on Electronics, Communications and Information Technology (ICECIT), 14-16 September 2021 [Online].
        Available: https://arxiv.org/pdf/2110.00859.pdf

[11]   P. Sudhir and V.D. Suresh, "Comparative study of various approaches, applications and classifiers for sentiment analysis", Global Transitions Proceedings Vol. 2, pp. 205–211, 2021 [Online].
        Available: https://www.sciencedirect.com/science/article/pii/S2666285X21000327

[12]   J. Wang and B. Xia, "Relationships of Cohen's Kappa, Sensitity, and Specificity for Unbiased Annotations", Proceedings of the 2019 4th International Conference on Biomedical Signal and Image Processing (ICBIP 2019), pp. 98–101, 2019 [Online].
        Available: https://dl.acm.org/doi/pdf/10.1145/3354031.3354040

[13]   C. Manning, P. Raghavan, and H. Schütze, "An Introduction to Information Retrieval", Cambridge University Press, Cambridge, 2009.

[14]   C.C. Aggarwal and C. Zhai, "Mining Text Data", Kluwer Academic Publishers, 2012, doi:10.1007/978-1-4614-3223-4.

[15]   P.-N. Tan, M. Steinbach, and V. Kumar, "Introduction to Data Mining", Person Education Limited, 2014.

[16]   D.L. Olson and D. Delen, "Advanced Data Mining Techniques", Springer-Verlag Berlin Heidelberg, 2008.

[17]   J. W., Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of LIWC2015". Technical report, Austin, TX: University of Texas at Austin, 2015 [Online].
        Available:
        https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf?sequence=3&isAllowed=y

[18]   Liu, Bing, Hu, Minqing, and Cheng, Junsheng, "Opinion Observer: Analyzing and Comparing Opinions on the Web." Proceedings of the 14th International World Wide Web Conference (WWW-2005), May 10-14, Chiba, Japan, 2005.

[19]   D.H. Wahid and S.N. Azhari, "Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity", IJCCS (*Indonesian Journal of Computing and Cybernetics Systems*), Vol. 10, No. 2, pp. 207-21, 2016 [Online]
        Available: https://jurnal.ugm.ac.id/ijccs/article/view/16625/11694