Development and Comparison of Multiple Emotion Classification Models in Indonesia Text Using Machine Learning

Ahmad Zamsuri [1,*], Sarjon Defit [2], and Gunadi Widi Nurcahyo [2]

[1] Department of Informatics Engineering, Faculty of Computer Science, University of Lancang Kuning, Pekanbaru, Indonesia

[2] Department of Information Technology, Faculty of Computer Science, University of Putra Indonesia YPTK, Padang, Indonesia

Email: ahmadzamsuri@unilak.ac.id (A.Z.); Sarjon_defit@upiyptk.ac.id (S.D.); gunadiwidi@yahoo.co.id (G.W.N.)

*Corresponding author

***Abstract*—Emotion is an individual's response to an event or situation. Research related to emotions in the field of data science falls under sentiment analysis, where sentiment analysis mostly focuses on determining the positive or negative emotional tone. This study attempts to classify emotions into labels such as Happy, Love, Surprise, Anger, Fear, and Sadness. Then, these six emotion labels are further categorized into positive and negative groups. The dataset for this research comprises tweets from Twitter related to the 2024 presidential election in Indonesia. Several machine learning algorithms are employed in this study, including Naïve Bayes (Multinomial Bayes, Bernoulli Bayes, Complement Bayes), K-Nearest Neighbors (KNN), and Support Vector Machines (SVM), comparing two feature extraction methods: Term Frequency-Inverse Document Frequency (TF-IDF) and Bag of Words (BoW). The results show that the use of SVM does not yield better accuracy, whether using TF-IDF or BoW. Therefore, this study improves the accuracy of the SVM algorithm by combining kernels available in SVM (Polynomial, Radial Basis Function (RBF), and Linear). From the conducted experiments, it is evident that SVM with a combined kernel, referred to as SVM Polynomial, RBF, Linier (PoRLi) in this study, achieves better accuracy compared to a single kernel. This is reflected in the classification accuracy with 6 labels, reaching 62%, indicating a 2% increase from SVM Linear, which obtained the highest accuracy using a single kernel. Furthermore, in the classification of 2 labels, there is also a 1% increase when using SVM PoRLi. It can be concluded that SVM PoRLi can enhance accuracy across various labels.**

***Keywords*—Term Frequency-Inverse Document Frequency (TF-IDF), Bag of Words (BoW), machine learning, multiple emotion, Support Vector Machines (SVM), Radial Basis Function (RBF), Polynomial, RBF, Linier (PoRLi), SVM PoRLi**

I. Introduction

Communication in modern society was observed to have been influenced by the rapid development of the internet. This is evident in the freedom to express opinions through social networks such as Twitter which is quite popular among internet users. The number of users have reached approximately 328 million, an increase of 6% or 9 million active users, in the first quarter of 2017 [1]. Each Twitter user is free to express opinions through comments, known as tweets, with a character limit of 140 [2, 3]. Twitter data is widely used for experimentation in data science, such as classification tasks [3, 4].

Classification is a method of grouping objects or words based on their characteristics and this can be achieved through a variety of methods, either manually or using technology. Manual classification is usually conducted by humans without the help of intelligent computer algorithms while ethnology-assisted classification uses several algorithms such as Naïve Bayes Algorithm, Support Vector Machine, Decision Tree, Fuzzy Logic, and Artificial Neural Networks [5, 6]. Previous research has also discussed various data for classification using the Indonesian language [7].

Indonesian is the official language normally used for communication purposes in Indonesia. Its sentence structure consists of a subject, predicate, object, and adjunct [8], as well as is further classified as simple or compound based on the number of clauses or ideas [9]. The simple structure only has one clause while the compound type usually has more than one clause connected by a conjunction [10]. Moreover, these compound-structured sentences have the potential to exhibit complex emotions due to the existence of multiple ideas [11].

Emotional data can be extracted from Indonesian sentences using machine learning processes such as text classification. For example, categories such as joy, love, fear, anger, surprise, and sadness can be generated from a statement using text classification based on six emotional classes [12]. This can be achieved using several algorithms such as K-Nearest Neighbors (KNN) [13], Naïve Bayes [14], Support Vector Machines (SVM) [15], and Neural Networks [16]. Some of these algorithms, including the naïve Bayes family (multinomial, Bernoulli,

Manuscript received October 3, 2023; revised November 11, 2023; accepted December 13, 2023; published April 24, 2024.

complement), KNN, and SVM were compared in this study.

A previous study compared the performance of SVM, KNN, NB, RF, and NN in analyzing the emotions in the comments of netizens on Facebook concerning artists and the results showed that SVM had the highest algorithm at 62% [17]. Furthermore, the sentiment towards bullying on Facebook was analyzed using Naïve Bayes, J48, SVM, and KNN algorithms, as well as the highest accuracy was also recorded in the SVM algorithm at 95% [18]. A similar observation was made in the comparison of SVM and Naïve Bayes on two different datasets including IMDB and Twitter, and the findings showed that SVM performed better with 83% accuracy on IMDB and 82% on Twitter [19]. Moreover, the sentiments of people towards Moroccan universities were assessed using RF, MNB, LR DT, SVM, and XGboost, as well as the results indicated that the highest accuracy at 98% [20]. Another sentiment analysis was also conducted on a marketplace dataset using naïve Bayes and random forest and the highest accuracy was achieved by Naïve Bayes at 87% [21]. Nine algorithms were also analyzed using five datasets and the most accurate result of 85.1% was recorded in Naïve Bayes, RF, and LR [22]. Similarly, a dataset extracted from live Twitter streams in several languages was evaluated using six algorithms, and the best accuracy was obtained by the SVM at 95% [23].

The information obtained from these previous studies was used as the basis to gather data from dea.uii.ac.id and Twitter with a particular focus on the 2024 presidential election for this study. The Twitter data was labelled based on the wheel of emotions defined as mental states influencing individual behavior, physiological changes, and thoughts [24]. It is possible to transmit emotions from one individual to another as can be observed from the anger and fear triggered in an individual during a fight or moment of anger [25]. Likewise, seeing someone smile can also elicit emotions of contentment [26]. The data collected were labeled and verified by language and psychology experts in line with the theories in previous study. This was achieved through two steps including emotion labeling and categorization into positive and negative labels. The positive category included love and joy while the negative was anger, surprise, fear, and sadness. The approach was observed to be different from previous studies that only focused on the accuracy of the emotion [13, 15, 27] or sentiment using positive and negative labels [28, 29].

The first stage was to prepare the data obtained before measuring the accuracy. This was achieved by transforming the data into a format that is easier and more effective for machine learning and artificial intelligence to produce accurate results [30]. The preprocessing stage included data cleaning, case folding, tokenizing, filtering, stemming, and data transformation [31]. This was followed by weighting which was the term for the assignment of weights to the words in the documents in order to measure their significance and contribution in determining the class or category [32]. The two methods used were Bag of Words (BoW) and TF-IDF.

BoW is one of the simplest methods to convert text data into computer-understandable vectors due to its ability to calculate the frequency of word occurrences across all documents [33]. Meanwhile, TF-IDF stands for Term Frequency-Inverse Document Frequency and combines the processes of Term Frequency (TF) as well as Inverse Document Frequency (IDF) to transform text data into vectors while considering the informativeness of the words [34]. These two methods were compared to determine the best accuracy among the algorithms. This was in line with the practice in previous study that compared different term weighting methods such as Word2Vec and GloVe [35], TF-IDF and Count Vectorizer (CV) [36, 37], as well as Word2Vec, GloVe, and fastText [37].

The next stage after term weighting was the development of the model using Multinomial Bayes, Bernoulli Bayes, Complement Bayes, KNN, and SVM. The process was repeated 20 times with the first experiment conducted using TF-IDF with six labels (five attempts), the second using BoW with six labels (five attempts), the third using TF-IDF with two labels (five attempts), and the fourth using BoW with two labels (five attempts). The aim was to observe the improvement or decline in the accuracy of the algorithms used. Receiver Operating Characteristics (ROC) which is a method usually used to depict, organize, and classify multiple predefined categories in a statistical model based on its performance was also applied to the six-label experiments [38].

This study was conducted to enhance the accuracy of the Support Vector Machine (SVM) algorithm considered to have unsatisfactory performance compared to its application in other studies. The limitation was overcome by proposing a novel approach designed by combining three kernels including Polynomial, Radial Basis Function (RBF), and Linear simultaneously and referred to as PoRLi. This combination has been proven to have the capacity to significantly improve the accuracy of an algorithm.

II. Materials and Methods

The motivation, framework, stages, materials, initial data processing, tools, methods, experiments, tests, evaluation, and validation associated with this study were discussed in this section. The methodology employed is presented in the following Fig. 1.

Study framework is a diagram that provides an overview of the logical flow of the study. Fig. 1 shows that this study was broadly divided into three components including the input, process, and output. The input was the sequential data, particularly Multiple Text Datasets, the process was the adoption of the Naïve Bayes, Support Vector Machine, and Neural Network algorithms for text classification and outlier detection in the multiple text dataset, and the output included the deviant news or information from the multiple-text dataset and the classification of different emotions in Indonesian text.

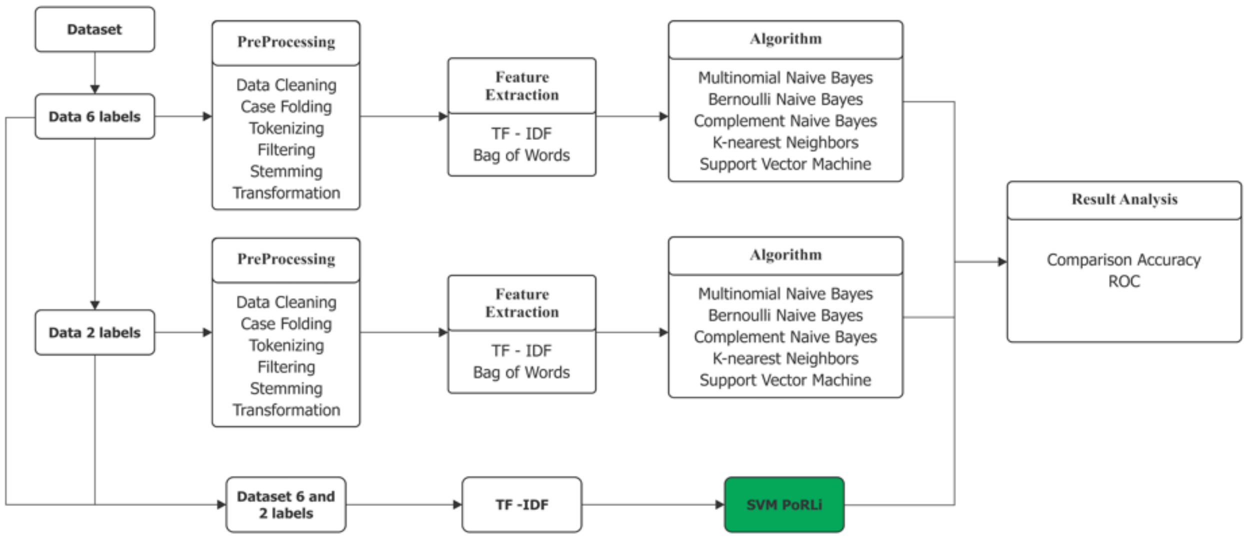


Fig. 1. Methodology flow.

A. Dataset

The data for this study was collected from tweets on Twitter. The tweet data was retrieved using the Drone Emprit Academy portal (https://dea.uii.ac.id/). The collected data is from a project named 'pemilu' (election) spanning from January to April. A total of 5,000 tweets were initially retrieved, and then filtering was applied. Retweets were excluded from the study because a retweet is a feature where another account shares someone else's post as their own tweet. After removing these retweets, the total number of tweet data was reduced to 1,648. This set of tweet data will undergo further processing

B. Data Labelling

This study focuses on individuals' emotions in expressing their opinions on the elections in Indonesia. A total of 1648 tweets were then labeled based on Plutchik's wheel of emotions. From this emotional wheel, six emotion labels were obtained: sadness, fear, surprise, anger, joy, and love. The labeling process was also validated by an expert in Indonesian language and a psychologist, who is an active lecturer at Universitas Lancang Kuning and Universitas Islam Riau. After the data labeling process into six emotion labels, this research further categorized these emotion labels into positive and negative based on expert advice used in this study. Emotion labels of love and joy were classified as positive, while sadness, fear, surprise, and anger were classified as negative. This categorization aims to assess the reliability of the model to be improved, examining whether it can enhance accuracy for both 2 and 6 labels.

C. Pre-Processing

The pre-processing stated was based on the following six stages depicted in Fig. 2.

1) Data cleaning

The data cleaning process was used to sanitize the data by removing mentions, hashtags, retweets, symbols, links, numbers, converting line breaks into spaces, and trimming leading as well as trailing spaces from the text [31].

2) Case folding

The data obtained were unstructured in terms of capitalization and this led to the application of case folding to convert all the words to lowercase.

3) Tokenizing

The sentences in the data were broken down or separated into words in this stage to aid the analysis process.

4) Filtering

The tokenization was followed by the filtering stage to remove the words considered unimportant.

5) Stemming

The stemming process was used to remove the affixes and endings in each word to have only the base form and this was achieved using the Sastrawi library.

6) Transformation

The final stage in the preprocessing was the transformation which focused on converting the previous word-based labels into numeric form automatically using Label Encoder.

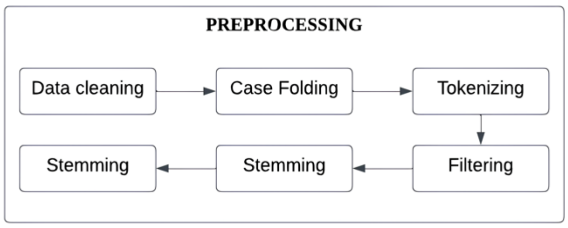


Fig. 2. Pre-processing.

D. Feature Extraction

Feature extraction is a technique normally used to select certain characteristics from a given form and analyze the contents produced for further processing. Meanwhile, classification can be explained as the process of assigning an object to one of the predefined categories. This study used two feature extraction techniques, TF-IDF and Bag of Words (BoW), which have also been used in previous studies related to machine learning.

The reason for choosing TF-IDF and BoW as feature extraction methods is based on the primary characteristics of text data. TF-IDF provides a simple representation of terms, taking into account term significance, reducing the weight of common terms, and scalability. Moreover, TF-IDF, as indicated by several studies, has been shown to improve accuracy compared to other feature extraction methods when applied to machine learning algorithms [36, 39]. Furthermore, BoW is chosen because it represents words based on their frequency of occurrence within a sentence in textual data [40].

E. Modeling

The selection of algorithms in this study is based on several previous research studies that indicate these algorithms achieve relatively high accuracy. Table I presents the findings of previous research studies, which affirm the high accuracy of the chosen algorithms.

Based on previous studies, testing was conducted using the mentioned algorithms. The experiments were performed 20 times, as outlined in Table II. Additionally, this research will also compare the TF-IDF and BoW feature extraction methods.

TABLE I. REASONS FOR SELECTING THE ALGORITHM

No	Researcher	Algorithm	Dataset	Accuracy
1	Ressan and Hassan [41]	Multinomial Bayes	Twitter (3,057 Tweet)	91.6%
2	Rahman and Hossen [42]	Bernoulli Bayes	Review Movie (2,000)	87.50%
3	Seref and Bostanci [43]	Complement Bayes	Amazon Movie Review (1,3576)	75.63%
4	Isnain *et al.* [44]	KNN	Twitter (1,825 Tweet)	85%
5	Gopi *et al.* [45]	SVM with kernel RBF	Twitter (25,000 Tweet)	98.8%
6	Imamah *et al.* [46]	SVM with kernel Linear	Tourist Reviews in bangkalan regency (1,394 reviews)	70.22%
7	Muflikhah *et al.* [47]	SVM with kernel Polynomial	400 comments from Tokopedia.com	88.75%

TABLE II. ALGORITHM TESTS ON 2 AND 6 LAB

Labels	Feature Extraction	Algorithm	Researcher
Fine-Grained	TF-IDF	Multinomial Bayes	Imelda and Kurnianto [48]
		Bernoulli Bayes	Sani *et al*. [49]
		Complement Bayes	Cahyani *et al*. [50]
		KNN	Zamsuri *et al*. [51]
		SVM	Mohammed and Omar [52]
	BoW	Multinomial Bayes	Chingmuankim and Jindal [53]
		Bernoulli Bayes	Nishadi [54]
		Complement Bayes	Ibrahim *et al*. [55]
		KNN	Zamsuri *et al*. [51]
		SVM	–
More than 4 labels	TF-IDF	Multinomial Bayes	Ilic *et al*. [56]
		Bernoulli Bayes	Wongso *et al*. [57]
		Complement Bayes	–
		KNN	Chen *et al*. [58]
		SVM	Siddiqua *et al*. [59]
	BoW	Multinomial Bayes	Mocherla *et al*. [60]
		Bernoulli Bayes	–
		Complement Bayes	–
		KNN	Zamsuri *et al*. [51]
		SVM	–

From several experiments, it is observed that there are 4 experiments that have not been conducted by previous researchers. Therefore, this study will attempt to examine the accuracy generated by the algorithms used, using both 6 labels and 2 labels. In addition to accuracy, this research also examines the results of the Receiver Operating Characteristic (ROC). The ROC is a curve that illustrates the classification results of true positive and false positive classes.

F. SVM Model Development

Only a single algorithm, SVM, was developed in this study but it offers several kernels such as polynomial, RBF, Sigmoid, and Linear [61, 62] as indicated in the following Fig. 3.

Fig. 3 shows that the SVM developed was based on three different kernels and this is the novelty of this study compared to the previous ones. This was because most past studies used only SVM kernels and compared the modeling accuracy [63, 64] while this study included additional kernels.

Improvement in the SVM algorithm is implemented because the algorithm in the previous study achieved reasonably good accuracy but occasionally obtained relatively low accuracy. This can be observed in Table III.

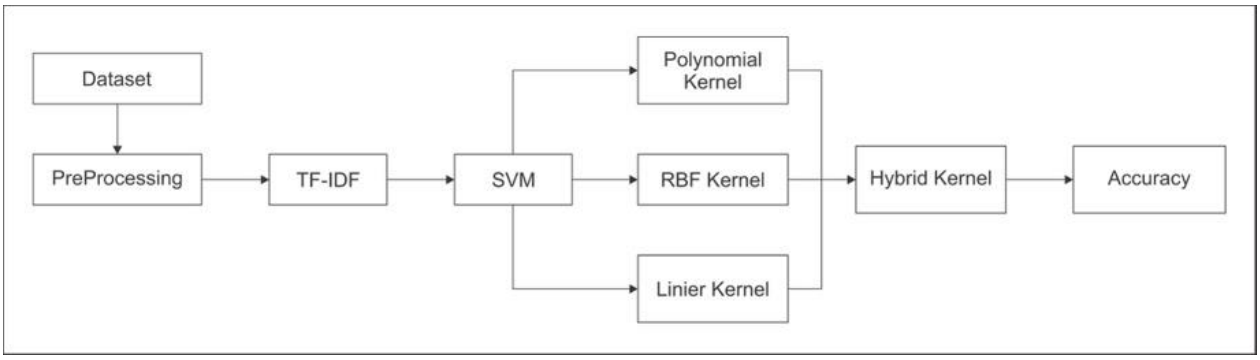


Fig. 3. Hybrid kernel SVM PoRLi multiple emotion.

TABLE III. SVM INCONSISTENCY

Researcher	Accuracy
Aljwari [65]	63%
Muis *et al*. [66]	70%
Imamah *et al*. [46]	98.8%

Performance of SVM by employing a hybrid kernel approach, specifically utilizing Polynomial, RBF, and Linear kernels (PoRLi). This initiative aims to bolster the model's capacity to capture diverse patterns and relationships within the data. Furthermore, it seeks to capitalize on the unique strengths inherent in each type of kernel. The Polynomial kernel is chosen to capture non-linear relationships, the RBF kernel is employed to address complex and non-convex patterns, and the Linear kernel is utilized to handle linearly separable patterns.

The rationale behind selecting this combination lies in the complementary nature of these kernels, allowing the model to exhibit versatility across various data patterns. By integrating these kernels, the expectation is that the model becomes more robust and adaptable to the intricacies present in the dataset. This strategic combination is anticipated to empower the model, enabling it to effectively navigate and adapt to the complexity inherent in the dataset.

III. RESULT AND DISCUSSION

The results from the 20 experiments conducted using 5 algorithms including Multinomial Bayes, Bernoulli Bayes, Complement Bayes, KNN, and SVM are discussed in this chapter. The experiments were based on 2 feature extraction techniques including Term Frequency-Inverse Document Frequency (TF-IDF) and Bag of Words (BoW). The results obtained based on the 6 labels previously identified are presented in the following Fig. 4.

1) Term weighting

Term weighting is the process of calculating the weight of each term searched for in each document so that the availability and similarity of a term in the document can be determined [67]. The following is the weighting term used in this research.

2) Term Frequency-Inverse Document Frequency (TF-IDF)

TF-IDF was used to measure the importance of a word in a document within the context of the larger collection. The main purpose was to assign weights to words having a high frequency in a particular document but rarely in others within the collection.

3) Bag of Word (BoW)

Bag of Words (BoW) was used to represent text or documents in the form of numeric vectors in machine learning. The approach treated each document as a collection of words contained, regardless of their order. The goal was to measure the frequency of word occurrences in the document to be used as features in analysis or machine learning modelling.

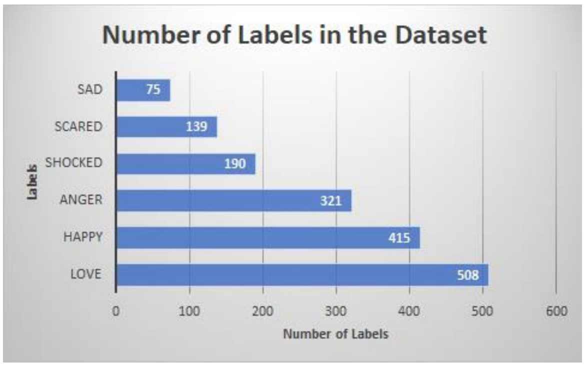


Fig. 4. Number of per label data.

A. *Model 6 Label Performance*

Performance results on the model using 6 labels with TF-IDF feature extraction show that the SVM method has maximum accuracy with a score of 60% while Bernoulli Bayes has a minimum accuracy of 53%. In KNN, the elbow method is used to determine the optimal K value based on model performance on validation or test data, and the accuracy is found to be 57%. then if we use BoW feature extraction, the SVM algorithm's accuracy is reduced to 58% while Complement Bayes increases from 59% to 60% with BoW. The modeling process is visualized using a bar diagram as presented in Fig. 5.

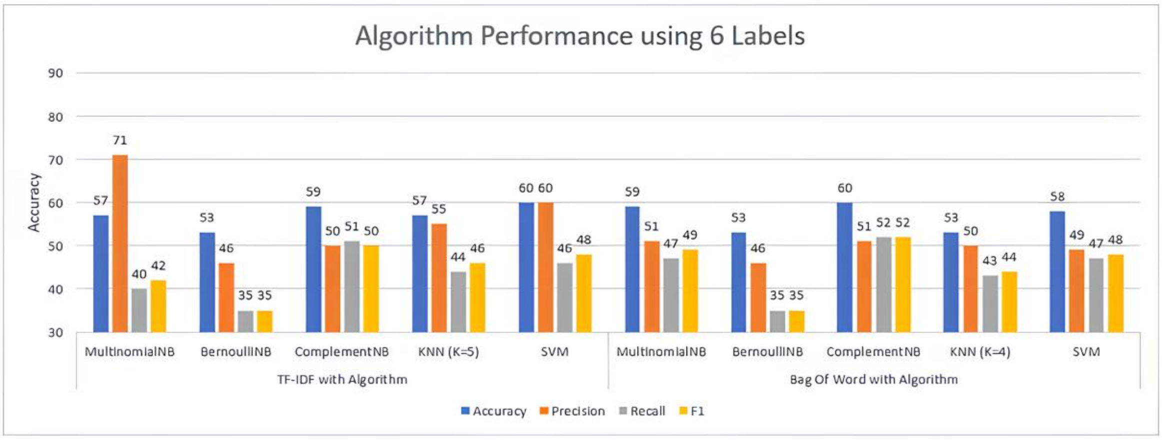


Fig. 5. Algorithm comparison accuracy with 6-labels.

The information presented in Fig. 5 showed that 6 labels used produced an accuracy below 70% with the highest recorded to be 60%. Moreover, the highest precision score was obtained with the Multinomial algorithm at 71% while the lowest was found in BoW KNN, Bernoulli Bayes, and TF-IDF Bernoulli Bayes to be 53%.

B. *Model 2 Label Performance*

The highest accuracy in TF-IDF was found to be 82% and it was achieved by Complement Bayes. Meanwhile, the highest for the 6-label classification was recorded to be 79% for SVM. This simply showed that the highest-performing algorithm in one dataset might not necessarily be the best in another, thereby indicating the influence of different label sets on the accuracy of models. In the BoW experiment, it was discovered that the Multinomial Bayes improved from 80% to 81% while Complement Bayes reduced from 82% to 80%. The findings further noted that KNN had the lowest accuracy with only 75% as indicated in the comparison graph presented in Fig. 6.

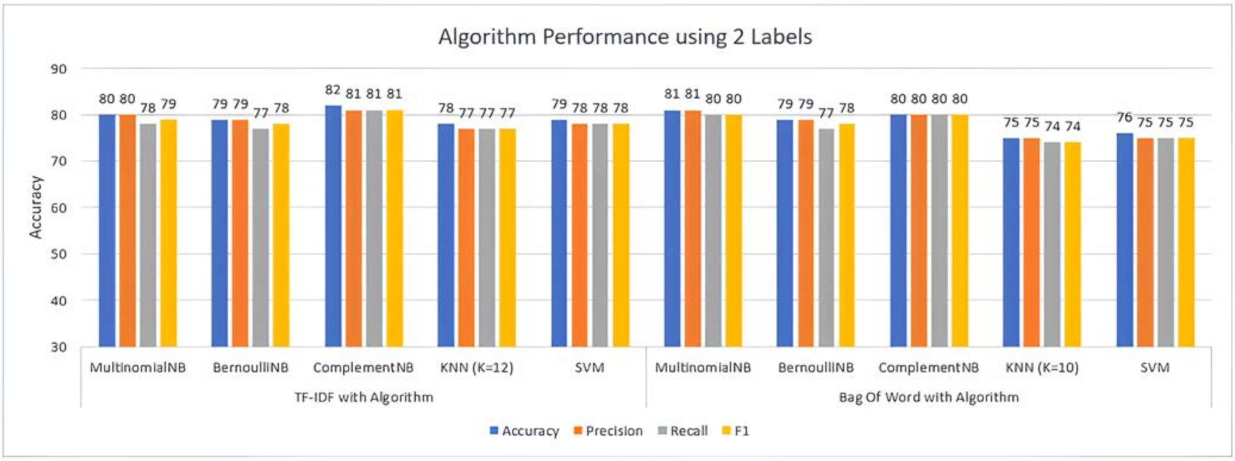


Fig. 6. Accuracy comparison of algorithms with 2-labels.

The overall results were quite good and showed a significant improvement compared to using 6 labels as indicated in Fig. 6. The highest accuracy recorded in TF-IDF was found with Complement Bayes while the lowest in BoW was with the KNN algorithm.

From the conducted experiments, it is evident that all models tested on the 6-label dataset yielded suboptimal results. Multinomial Naïve Bayes in this study achieved a relatively high accuracy using TF-IDF feature extraction, but exhibited a decrease of approximately 12% when utilizing BoW. Similar trends were observed with other algorithms, nearly all experiencing a decline in performance when employing BoW for feature extraction. Notably, one algorithm, Bernoulli Naïve Bayes, maintained its accuracy.

From the aforementioned explanations, it can be concluded that BoW may not be suitable for textual data with only 1648 instances and 6 labels. Furthermore, while TF-IDF produced suboptimal accuracy results, each algorithm can be enhanced through various methods. In particular, this study should address data imbalance before accurately assessing performance, with data balancing achievable through techniques such as SMOTE [7].

On the other hand, the dataset with 2 labels exhibited commendable accuracy results. This is attributed to the balanced nature of data within each label. Both BoW and TF-IDF feature extraction methods demonstrated equally good accuracy values.

C. Hybrid Kernel SVM

The accuracy of different kernels used in SVM was observed to have been compared in previous studies. For example, the sentiment analysis conducted on social distancing showed that the polynomial kernel achieved the highest accuracy compared to linear and RBF [58]. Another study also compared four kernels in SVM and found that the Radial Basis Function (RBF) yielded the best F1-Score of 96.36% compared to the Linear, Sigmoid, and Polynomial [61]. Meanwhile, in contrast to these two studies, this study employed a different approach by using a hybrid kernel SVM which combined the Radial Basis Function (RBF), Linear, and Polynomial kernels.

The hybrid or multiple kernel SVM was used to combine the strengths of different kernel functions to improve the performance of the SVM model. Some of its advantages are presented as follows:

- Improved Classification Accuracy: It was able to capture a wider range of patterns and relationships in the data, thereby enhancing the classification accuracy compared to using a single kernel function.
- Flexibility in Modelling Complex Data: The method excelled at capturing various types of patterns in data, thereby allowing the capability to model complex data with diverse patterns and dependencies.
- Enhanced Generalization: The method allowed effective generalization to unseen data by combining different kernel functions as well as ensured adaptation to different data distributions and capturing both local and global patterns, thereby improving generalization performance.
- Increased Robustness: The method was more robust to noise and outliers in the data and effectively handled data points that may be poorly classified by a single kernel function.
- Improved Interpretability: In comparison to other complex machine learning models, hybrid kernel SVM presented findings that are easier to understand. It also provided the opportunity to learn more about the most important features or patterns in the classification task by analyzing the contributions of each kernel function.
- Adaptive Learning: The method enabled adaptive learning by dynamically adjusting the weights assigned to each kernel function based on the data characteristics, thereby providing optimal learning capability.
- Higher Flexibility in Kernel Selection: The method offered flexibility in selecting and combining different kernel functions based on the specific problem at hand. This customized kernel selection enabled the model to adapt to the specific characteristics of the data.

Hybrid kernel SVM was generally observed to offer a robust framework to model complex data, improve classification accuracy, as well as enhance the generalization and robustness of SVM models. This was due to its ability to provide a versatile and flexible approach to machine learning tasks, thereby delivering improved performance across various applications. The study was conducted on different numbers of labels, specifically 2 and 6, and the findings from the first experiment test to compare three SVM kernels on the 2 labels are presented in Table IV.

TABLE IV. SINGLE KERNEL 2 LABEL SVM

Kernel	Accuracy	Precision	Recall	F1-Score
Linear	79%	78%	78%	78%
RBF	76%	77%	74%	75%
Polynomial	58%	29%	50%	37%

The information presented in Table IV showed that the linear kernel performed better than the others in line with the results of earlier studies that examined these three kernels [68]. This was confirmed by the 79% accuracy recorded for linear compared to 76% for RBF and the lowest, 58%, for the polynomial.

In this study, the main focus is on hybridizing the SVM kernel. After the hybrid process, there is an improvement, although not significant. This is observed in the confusion matrix in Fig. 7. The confusion matrix is one of the predictive analytical tools that displays and compares actual values or ground truth values with the predicted values of the model. It can be used to generate evaluation metrics such as Accuracy, Precision, Recall, and F1-Score or F-Measure.

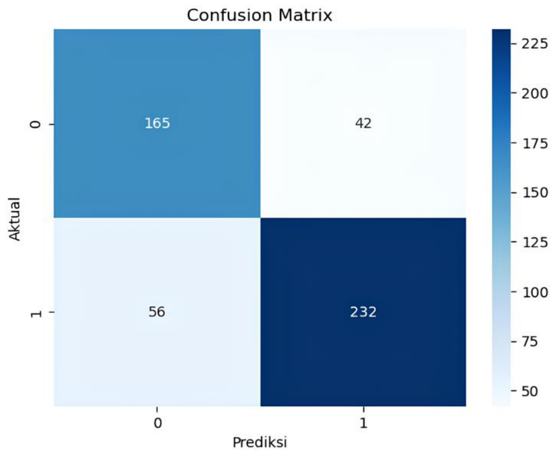


Fig. 7. Confusion Matrix model SVM PoRLi with 2-labels.

From Fig. 7, measurements were taken for accuracy, precision, recall, and F1-Score. Accuracy measures how well your model correctly predicts the labels in the dataset. Accuracy can be calculated using Eq. (1).

$$A = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

where, A = Accuracy; TP = True Positive; TN = True Negative; FP = False Positive; FN = False Negative.

By using the SVM PoRLi algorithm, the accuracy achieved is 62%, meaning that 62% of all model predictions are correct. Then precision measures how well your model successfully identifies a specific class. Precision is measured using Eq. (2).

$$P = \frac{tp}{tp + fp} \quad (2)$$

where P = Precision.

In this study, there is precision for labels 0 and 1. The following is the labeling that has been replaced with numbers, as presented in Table V.

TABLE V. NAME OF 2 LABELS

Labels	Number
Negative	0
Positive	1

Then recall is also calculated using Eq. (3).

$$R = \frac{tp}{tp + fn} \tag{3}$$

where R = Recall.

Recall measures how well your model manages to find all instances of a specific class. Recall also has each label ranging from 0 to 5. Finally, calculate the F1-Score value using Eq. (4).

$$F = 2 \times \frac{P \times R}{P + R} \tag{4}$$

where F = F1-Score.

F1-Score is a combined metric that integrates precision and recall. After measuring using Eqs. (1)–(4), Table VI represents all the calculations based on the confusion matrix.

TABLE VI. CALCULATION RESULTS 2 LABELS

Label	Precision	Recall	F1-Score
0	75%	80%	77%
1	85%	81%	83%
Accuracy		80%	

Then, a classification report was generated from the results of the confusion matrix using a Python application, as seen in Fig. 8.

```
              precision    recall  f1-score   support

           0       0.75      0.80      0.77       207
           1       0.85      0.81      0.83       288

    accuracy                           0.80       495
   macro avg       0.80      0.80      0.80       495
weighted avg       0.80      0.80      0.80       495
```

Fig. 8. Report classification SVM PoRLi with 2-labels.

Fig. 8 shows there is an increase in accuracy using the hybrid model when compared to the single kernel as indicated by the 80% recorded. The value was a 1% increase from the highest achieved by the linear model in the previous test. This was followed by the application of the models to 6-labels and the results obtained using a single SVM kernel are shown in Table VII.

TABLE VII. SINGLE KERNEL SVM 6 LABELS

Kernel	Accuracy	Precision	Recall	F1-Score
Linear	60%	60%	46%	48%
RBF	48%	47%	29%	27%
Polynomial	32%	5%	17%	8%

The results presented in Table VII also showed that linear kernel had the highest accuracy with 60% while RBF had 48%, and Polynomial had the lowest with 32%. Then modelling was conducted using SVM PoRLi, and the figures in the confusion matrix are presented in Fig. 9.

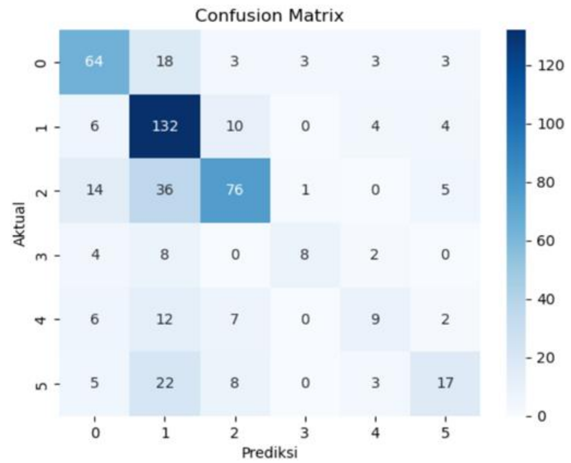
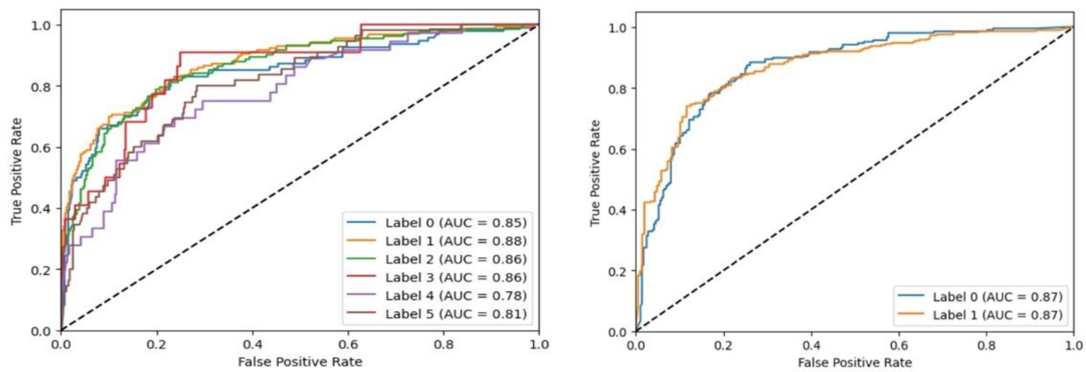


Fig. 9. Confusion matrix model SVM PoRLi with 6-labels.

The labels, which were previously emotion labels, have been converted into numbers 0 and 5, as shown in Table VIII.

TABLE VIII. NAMES OF 6 LABELS

Label	Number
Anger	0
Love	1
Happy	2
Sad	3
Afraid	4
Surprised	5



ROC Curve: 0.8403 ROC Curve: 0.8703

6 Labels 2 Labels

Fig. 10. Comparison of ROC model SVM PoRLi 6 Labels and 2 label.

Furthermore, the ROC results obtained for both 2-label and 6-label datasets are presented in the following Fig. 10.

From Fig. 10, the calculation results for accuracy, precision, recall, and F1-Score can be observed in Table IX.

TABLE IX. CALCULATION RESULTS 6 LABELS

Label	Precision	Recall	F1-Score
0	65%	68%	66%
1	58%	85%	69%
2	73%	58%	64%
3	67%	36%	47%
4	43%	25%	32%
5	55%	31%	40%
Accuracy		62%	

Afterwards, a classification report was conducted as seen in Fig. 11.

Fig. 11 shows an improvement in accuracy from the single-model testing with the linear kernel reported to have achieved an accuracy of 60% while the hybrid model had 62%. Moreover, all the kernels used for both 2 and 6 labels were compared and the results are presented graphically in the following Fig. 12.

```
              precision    recall  f1-score   support

           0       0.65      0.68      0.66        94
           1       0.58      0.85      0.69       156
           2       0.73      0.58      0.64       132
           3       0.67      0.36      0.47        22
           4       0.43      0.25      0.32        36
           5       0.55      0.31      0.40        55

    accuracy                           0.62       495
   macro avg       0.60      0.50      0.53       495
weighted avg       0.62      0.62      0.60       495
```

Fig. 11. Classification report hybrid kernel SVM with 6-labels.

The comparison made using the graph in Fig. 12 showed that the polynomial kernel had the lowest accuracy compared to the others. It was also discovered that the hybrid kernel SVM was highly effective as indicated by its ability to improve accuracy across different datasets with varying labels.

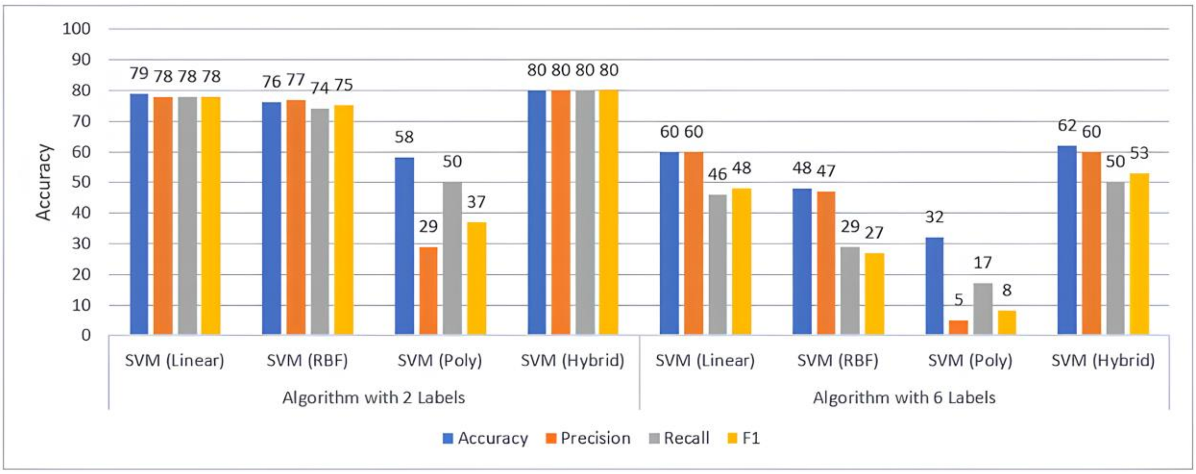


Fig. 12. Comparison graph of hybrid and single models with 2-label and 6-label datasets.

The results obtained through the Hybrid Kernel SVM (PoRLi) have shown improvement, albeit not significantly, reaching 62%. This marks a 2% increase compared to a single kernel approach. When compared to previous studies, our research demonstrates superiority. In a study with emotions categorized into 5 labels [69], accuracies were reported as 49.6% with decision tree, 55.3% with KNN, and 56.5% with RFC. Another study utilizing KNN alone achieved 58% accuracy [70], and the application of KNN in yet another study yielded a 58% accuracy as well [51]. In this study, even with the use of a single SVM kernel, the highest accuracy attained was 60%, specifically with the SVM Linear kernel.

The Receiver Operating Characteristic (ROC) Analysis is a useful method for evaluating the performance of a classification model with multiple labels [71]. In ROC analysis, the Area Under the Curve (AUC) is one of the main metrics used to measure how well the model can distinguish between positive and negative classes [72]. The higher the AUC value, the better the model's ability to differentiate between these classes. Fig. 10 shows the ROC results in this research

In the 6-label ROC, Label 0 has an AUC of around 0.85, indicating that the model has a good ability to distinguish between positive and negative classes for this label. The closer the AUC value is to 1, the better the model's performance. Label 1 has a higher AUC, namely 0.88. This indicates that the model is excellent at distinguishing between positive and negative classes for Label 1. This is a very good result. Label 2 also has a high AUC, namely 0.86. The model has a good ability to distinguish between positive and negative classes for Label 2. Label 3 has the same AUC as Label 2, which is 0.86. This model also has a good ability to distinguish between positive and negative classes for Label 3. However, Label 4 has a slightly lower AUC, namely 0.78. This indicates that the model has a less effective ability to distinguish between positive and negative classes for Label 4 compared to other labels. Label 5 has an AUC of around 0.81, indicating that the model has a good ability to distinguish between positive and negative classes for Label 5.

In general, high AUC results indicate that your model is effective in classifying positive and negative classes for the given labels. However, other factors such as precision, recall, and specific application context should also be considered when evaluating the overall performance of the model. Then, in the 2-label scenario, the SVM PoRLi model consistently performs well in classifying both labels, and the high AUC value (0.87) indicates that the model has a good ability to distinguish between positive and negative classes, both for Label 0 and Label 1.

This study did not conduct data balancing on the utilized dataset. Consequently, a mismatch between the ROC results and the SVM PoRLi 6-label modeling occurred. While ROC achieved favorable outcomes, the SVM PoRLi 6-label modeling exhibited low confusion metric results. In Fig. 4, it can be observed that each label has a different quantity, with the highest being 508 and the lowest being 75.

Future research is expected to address dataset-related issues by implementing data balancing techniques such as SMOTE and Adasyn. Additionally, the enhancement of algorithms through alternative methods like hyperparameter tuning and fusion techniques (boosting, voting, or ensemble methods) is crucial for further improvements.

D. Discussion

Based on the conducted experiments, it was found that machine learning algorithms still have weaknesses in classifying a larger number of labels. This is evident from the accuracy results, where the highest accuracy for the 6 labels was 71%, obtained from the multinomial naïve Bayes algorithm with TF-IDF feature extraction. Generally, the use of TF-IDF feature extraction in modeling machine learning is more effective compared to Bag of Words (BoW) feature extraction. In this study, BoW was unable to achieve an accuracy improvement of more than 70% for the 6 labels. Furthermore, for the 2-label scenario, the overall results were quite good, whether using TF-IDF or BoW feature extraction.

Other studies examining machine learning models with 6 labels or more also achieved accuracies below 70%. In a prior study [73], accuracies of 50% for KNN with TF-IDF, 66% for SVM with TF-IDF, and 55% for Decision Tree with TF-ICF were reported. Another research [74] found accuracies of 61% for naïve Bayes, logistic regression, SVM, 64% for Decision Tree, and 57% for Adaboost. From these results, it can be concluded that word weighting with feature extraction, whether using TF-IDF or BoW, has not been optimized for modeling in machine learning. In the future, improvements need to be explored using other methods such as boosting or ensemble techniques with different machine learning algorithms.

This research also aimed to enhance the SVM algorithm. Although SVM demonstrated relatively high accuracy, it occasionally fell below 70%. The improvement of SVM involved combining three kernels: Polynomial, RBF, and Linear. This fusion showed potential to increase accuracy, albeit insignificantly. The following is an explanation of Fig. 12, the accuracy results of SVM with 6 labels and 2 labels are presented in Table X.

TABLE X. ACCURACY COMPARISON

Label	**Model**	**Evaluation Metrics**			
		Accuracy	**Precision**	**Recall**	**F1-Score**
Algorithm with 2 Labels	SVM (Linear)	79%	78%	78%	78%
	SVM (RBF)	76%	77%	74%	75%
	SVM (Poly)	58%	29%	50%	37%
	SVM (Hybrid)	80%	80%	80%	80%
Algorithm with 6 Labels	SVM (Linear)	60%	60%	46%	48%
	SVM (RBF)	48%	47%	29%	27%
	SVM (Poly)	32%	5%	17%	8%
	SVM (Hybrid)	62%	60%	50%	53%

SVM Linear achieved 60%, SVM RBF 48%, and SVM Polynomial 32%. After obtaining accuracy with a single kernel, this study implemented a hybrid of the three kernels, resulting in a 62% accuracy. In the case of 2 labels, SVM Linear achieved 79% accuracy, SVM RBF 76%, and SVM Polynomial 58%, while the hybrid kernel (PoRLi) achieved an accuracy of 80%, indicating a 1% improvement. With the improved accuracy from SVM PoRLi, it suggests potential use for detecting varying label quantities in the future. However, further research is necessary to assess the algorithm's performance with diverse and larger datasets. Additionally, variations in data splitting can impact accuracy, thus warranting further comparative studies by future researchers.

The detection results obtained in this study can serve as an evaluation for the organizers of the 2024 elections in Indonesia. Upon analyzing the emotional identification of election-related tweets, it is evident that many express negative emotions such as sadness, fear, surprise, and anger. There is a need for socialization efforts targeted at the millennial and Generation Z demographics, who play a significant role in electing the President and Vice President of Indonesia, to prevent voter abstention. It is hoped that this research can provide insights for stakeholders to devise strategies for the 2024 election socialization in Indonesia.

IV. CONCLUSION

In conclusion, the highest accuracy in sentiment emotion analysis among the 6 labels was recorded at 71% with Multinomial Bayes using TF-IDF for feature extraction. For 2 labels, the best-performing algorithm was Complement Bayes with TF-IDF, achieving an accuracy of 82% and a ROC score of 87%. These results indicate a significant accuracy difference of 22% between the use of 6 labels and 2 labels. The study also focused on improving the SVM algorithm, which demonstrated suboptimal performance. The best accuracy for 2 labels was 80%, and for 6 labels, it was only 60% using TF-IDF for feature extraction. Subsequent enhancement with SVM PoRLi, a combination of three kernels, resulted in an improvement, albeit not significant.

The limitations observed in this study were expected to be solved in future studies. The 60% accuracy in the 6-label scenario can be improved further using boosting algorithms such as Gradient Boosting, Adaboost, and others. Moreover, only Machine Learning techniques were used and experiments were conducted on only a single dataset. It is recommended that future studies explore the model using Deep Learning and test it on different datasets to determine its strengths and weaknesses. Furthermore, hybrid testing needs to be implemented on the algorithms used as demonstrated in previous studies such as SVM with Naïve Bayes [59], SVM with KNN [60], and others.

Conflict of Interest

The authors declare no conflict of interest

Author Contributions

Sarjon Defit and Gunadi Widi Nurcahyo contributed by providing inputs at each stage, starting from the background to the conclusion. Sarjon Defit. provided relevant ideas while Gunadi Widi Nurcahyo ensured the validation of the labels used in the dataset; all authors had approved the final version.

References

[1] S. Ranginwala and A. J. Towbin, "Use of social media in radiology education," *Journal of the American College of Radiology*, vol. 15, no. 1, pp. 190–200, Jan. 2018. doi: 10.1016/j.jacr.2017.09.010

[2] D. G. Serfass and R. A. Sherman, "Situations in 140 characters: Assessing real-world situations on twitter," *PLoS One*, vol. 10, no. 11, Nov. 2015. doi: 10.1371/journal.pone.0143051

[3] A. Chenna, B. Srinivas, and S. Nagaraju, "Emotion and sentiment analysis from twitter text," *Turkish Journal of Computer and Mathematics Education*, vol. 12, no. 12, pp. 4614–4620, 2021.

[4] K. Sailunaz and R. Alhajj, "Emotion and sentiment analysis from Twitter text," *J Comput Sci*, vol. 36, 101003, 2019. doi: 10.1016/j.jocs.2019.05.009

[5] Z. Pezeshki and S. M. Mazinani, "Comparison of artificial neural networks, fuzzy logic and neuro fuzzy for predicting optimization of building thermal consumption: A survey," *Artif. Intell. Rev.*, vol. 52, no. 1, pp. 495–525, Jun. 2019. doi: 10.1007/s10462-018-9630-6

[6] P. Sangita and B. R. Deshmukh, "Use of support vector machine, decision tree and Naive Bayesian techniques for wind speed classification," in *Proc. International Conference on Power and Energy Systems*, IEEE, 2011, pp. 1–8. doi: 10.1109/ICPES.2011.6156687

[7] M. K. Anam *et al.*, "Sentiment analysis for online learning using the lexicon-based method and the support vector machine algorithm," *ILKOM Jurnal. Ilmiah.*, vol. 15, no. 2, pp. 290–302, 2023. doi: 10.33096/ilkom.v15i2.1590.290-302

[8] W. Nainggolan, B. Sinurat, and B. S. Hutauruk, "Accuracy of the translations of English noun phrases into Indonesian using google translate," *Journal of English Teaching as a Foreign Language*, vol. 6, no. 2, pp. 66–87, 2020.

[9] Darto, "Types of simple sentence used on the Joko Widodo's speech at Bali," *JournE*, vol. 1, no. 1, pp. 52–61, 2019.

[10] T. N. Fitria, "Non-EFL students' ability in writing simple Sentences of descriptive paragraphs," *JETAL: Journal of English Teaching and Applied Linguistics*, vol. 4, no. 1, pp. 1–9, 2022. doi: 10.36655/jetal.v4i1.762

[11] X. Yu, "A study on the role of sentence structure analysis in English learning," *OAlib*, vol. 8, no. 9, pp. 1–14, 2021. doi: 10.4236/oalib.1107859

[12] R. Plutchik, "A psych evolutionary theory of emotions," *Social Science Information*, vol. 21, no. 4–5, pp. 529–553, 1982. doi: 10.1177/053901882021004003

[13] B. Fernandes, J. C. Bhargavi, S. Arshad, S. Kumar, and G. Sandeep, "Emotion recognition in speech signals using optimization based multi-SVNN classifier," *International Journal of Scientific and Technology Research*, vol. 9, no. 1, pp. 3998–4001, Feb. 2020.

[14] A. Ahammed, B. Harangi, and A. Hajdu, "Hybrid adaboost and naïve bayes classifier for supervised learning," in *Proc. Conference on Information Technology and Data Science*, 2020, pp. 1–18.

[15] A. N. Sutranggono and E. M. Imah, "Tweets emotions analysis of community activities restriction as COVID-19 policy in Indonesia using support vector machine," *Communication and Information Technology*, vol. 17, no. 1, 2023.

[16] E. Batbaatar, M. Li, and K. H. Ryu, "Semantic-emotion neural network for emotion recognition from text," *IEEE Access*, vol. 7, pp. 111866–111878, 2019. doi: 10.1109/ACCESS.2019.2934529

[17] M. S. S. Khan, S. R. Rafa, A. E. H. Abir, and A. K. Das, "Sentiment analysis on Bengali Facebook comments to predict Fan's emotions towards a celebrity," *Journal of Engineering Advancements*, pp. 118–124, Jul. 2021. doi: 10.38032/jea.2021.03.001

[18] A. Al-Mamun and S. Akhter, "Social media bullying detection using machine learning on Bangla text," in *Proc. International Conference on Electrical and Computer Engineering*, Academy and Industry Research Collaboration Center (AIRCC), Sep. 2018, pp. 385–388. doi: 110.1109/ICECE.2018.8636797

[19] M. S. Basarslan and F. Kayaalp, "Sentiment analysis with machine learning methods on social media," *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, vol. 9, no. 3, pp. 5–15, Sep. 2020. doi: 10.14201/adcaij202093515

[20] I. Lasri, A. Riadsolh, and M. Elbelkacemi, "Real-time twitter sentiment analysis for Moroccan universities using machine learning and big data technologies," *International Journal of Emerging Technologies in Learning*, vol. 18, no. 05, pp. 42–61, Mar. 2023. doi: 10.3991/ijet.v18i05.35959

[21] W. Wisnalmawati, A. S. Aribowo, and Y. Herawati, "Semi-supervised learning models for sentiment analysis on marketplace dataset," *International Journal of Artificial Intelligence & Robotics (IJAIR)*, vol. 4, no. 2, pp. 78–85, Dec. 2022. doi: 10.25139/ijair.v4i2.5267

[22] I. Lazrig and S. L. Humpherys, "Using machine learning sentiment analysis to evaluate learning impact," *Information Systems Education Journal*, vol. 20, no. 1, 2022.

[23] K. Arun and A. Srinagesh, "Multi-lingual twitter sentiment analysis using machine learning," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 6, pp. 5992–6000, Dec. 2020. doi: 10.11591/ijece.v10i6.pp5992-6000

[24] C. E. Izard, "Emotion theory and research: Highlights, unanswered questions, and emerging issues," *Annual Review of Psychology*, vol. 60, pp. 1–25, Jan. 2009. doi: 10.1146/annurev.psych.60.110707.163539

[25] D. Wollebæk, R. Karlsen, K. S. Johnsen, and B. Enjolras, "Anger, fear, and echo chambers: The emotional basis for online behavior," *Social Media and Society*, vol. 5, no. 2, 2019. doi: 10.1177/2056305119829859

[26] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak, "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements," *Psychological Science in the Public Interest*, vol. 20, no. 1, pp. 1–68, Jul. 2019. doi: 10.1177/1529100619832930

[27] M. A. Ullah, S. M. Marium, S. A. Begum, and N. S. Dipa, "An algorithm and method for sentiment analysis using the text and emoticon," *ICT Express*, vol. 6, no. 4, pp. 357–360, Dec. 2020. doi: 10.1016/j.icte.2020.07.003

[28] B. Yoo and J. T. Rayz, "Understanding emojis for sentiment analysis," in *Proc. International FLAIRS Conference*, 2021, pp. 1–4. doi: 10.32473/flairs.v34i1.128562

[29] H. Alhuzali, T. Zhang, and S. Ananiadou, "Emotions and topics expressed on twitter during the COVID-19 pandemic in the United Kingdom: Comparative geolocation and text mining analysis," *J. Med. Internet. Res.*, vol. 24, no. 10, Oct. 2022. doi: 10.2196/40323

[30] K. Maharana, S. Mondal, and B. Nemade, "A review: Data preprocessing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, Jun. 2022. doi: 10.1016/j.gltp.2022.04.020

[31] M. K. Anam, M. I. Mahendra, W. Agustin, Rahmaddeni, and Nurjayadi, "Framework for analyzing netizen opinions on BPJS using sentiment analysis and Social Network Analysis (SNA)," *Intensif*, vol. 6, no. 1, pp. 2549–6824, 2022. doi: 10.29407/intensif.v6i1.15870

[32] R. N. Rathi and A. Mustafi, "The importance of term weighting in semantic understanding of text: A review of techniques," *Multimed. Tools Appl.*, vol. 82, no. 7, pp. 9761–9783, Mar. 2023. doi: 10.1007/s11042-022-12538-3

[33] S. Gite *et al.*, "Textual feature extraction using ant colony optimization for hate speech classification," *Big Data and Cognitive Computing*, vol. 7, no. 1, p. 45, Mar. 2023. doi: 10.3390/bdcc7010045

[34] T. Peng, L. Liu, and W. Zuo, "PU text classification enhanced by term frequency-inverse document frequency-improved weighting," *Concurr. Comput.*, vol. 26, no. 3, pp. 728–741, Mar. 2014. doi: 10.1002/cpe.3040.

[35] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Social Network Analysis and Mining*, vol. 11, no. 1. Springer, Dec. 01, 2021. doi: 10.1007/s13278-021-00776-6

[36] R. Haque, N. Islam, M. Tasneem, and A. K. Das, "Multi-class sentiment classification on Bengali social media comments using machine learning," *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 21–35, Jun. 2023. doi: 10.1016/j.ijcce.2023.01.001

[37] M. R. Karim, B. R. Chakravarthi, J. P. M. Crae, and M. Cochez, "Classification benchmarks for under-resourced Bengali language based on multichannel convolutional-LSTM network," in *Proc. 2020 IEEE 7th International Conference on Data Science and Advanced Analytics, DSAA 2020*, Institute of Electrical and Electronics Engineers Inc., Oct. 2020, pp. 390–399. doi: 10.1109/DSAA49011.2020.00053

[38] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett*, vol. 27, no. 8, pp. 861–874, Jun. 2006. doi: 10.1016/j.patrec.2005.10.010

[39] A. Al-Mamun and S. Akhter, "Social media bullying detection using machine learning on Bangla text," in *Proc. International Conference on Electrical and Computer Engineering*, Academy and Industry Research Collaboration Center (AIRCC), Sep. 2018, pp. 385–388. doi: 10.1109/ICECE.2018.8636797

[40] I. R. Nkechi, U. Chidiebere, and A. Tunde, "Analysis and representation of Igbo text document for a text-based system," *International Journal of Data Mining Techniques and Applications*, vol. 6, no. 1, pp. 26–32, 2017. doi: 10.48550/arXiv.2009.06376

[41] M. B. Ressan and R. F. Hassan, "Naïve-bayes family for sentiment analysis during COVID-19 pandemic and classification tweets," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 28, no. 1, pp. 375–383, Oct. 2022. doi: 10.11591/ijeecs.v28.i1.pp375-383

[42] A. Rahman and M. S. Hossen, "Sentiment analysis on movie review data using machine learning approach," in *Proc. 2019 International Conference on Bangla Speech and Language Processing, ICBSLP 2019*, Institute of Electrical and Electronics Engineers Inc., Sep. 2019. doi: 10.1109/ICBSLP47725.2019.201470

[43] B. Seref and E. Bostanci, "Sentiment analysis using naive bayes and complement naive bayes classifier algorithms on hadoop framework," in *Proc. International Symposium on Multidisciplinary Studies and Innovative Technologies: ISMSIT*, 2018, pp. 1–7. doi: 10.1109/ISMSIT.2018.8567243

[44] A. R. Isnain, J. Supriyanto, and M. P. Kharisma, "Implementation of k-nearest neighbor (K-NN) algorithm for public sentiment analysis of online learning," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 2, 121, Apr. 2021. doi: 10.22146/ijccs.65176

[45] A. P. Gopi, R. N. S. Jyothi, V. L. Narayana, and K. S. Sandeep, "Classification of tweets data based on polarity using improved RBF kernel of SVM," *International Journal of Information Technology (Singapore)*, vol. 15, no. 2, pp. 965–980, Feb. 2023. doi: 10.1007/s41870-019-00409-4.

[46] H. Imamah, E. M. Rachman, I. O. Suzanti, and F. A. Mufarroha, "Text mining and support vector machine for sentiment analysis of tourist reviews in Bangkalan regency," *Journal of Physics: Conference Series*, pp. 1–6, 2020. doi: 10.1088/1742-6596/1477/2/022023

[47] L. Muflikhah, D. J. Haryanto, A. A. Soebroto, and E. Santoso, "High performance of polynomial kernel at SVM algorithm for sentiment analysis," *Journal of Information Technology and Computer Science*, vol. 3, no. 2, pp. 194–201, 2018. doi: 10.25126/jitecs.20183260

[48] I. Imelda and A. R. Kurnianto, "Naïve bayes and TF-IDF for sentiment analysis of the Covid-19 booster vaccine," *Jurnal RESTI*, vol. 7, no. 1, pp. 1–6, Jan. 2023. doi: 10.29207/resti.v7i1.4467

[49] R. R. Sani, Y. A. Pratiwi, S. Winarno, D. E. Udayanti, and F. A. Zami, "Comparative analysis of the naive bayes classifier algorithm and support vector machine for hoax classification in Indonesian online news," *Jurnal Masyarakat Informatika*, vol. 13, no. 2, pp. 85–98, 2022. doi: 10.14710/jmasif.13.2.47983

[50] T. I. Cahyani, W. Gata, D. D. Saputra, H. B. Novitasari, and Hernawati, "Sentiment analysis toward isp Telkomsel and Xl on twitter using machine learning classification," *Journal of Information Technology and Computer Science (INTECOMS)*, vol. 6, no. 1, pp. 265–273, 2023. doi: 10.31539/intecoms.v6i1.5765

[51] A. Zamsuri, S. Defit, and G. W. Nurcahyo, "Classification of multiple emotions in Indonesian text using the k-nearest neighbor method," *Journal of Applied Engineering and Technological Science*, vol. 4, no. 2, pp. 1012–1021, 2023. doi: 10.37385/jaets.v4i2.1964

[52] M. Mohammed and N. Omar, "Question classification based on bloom's taxonomy using enhanced TF-IDF," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 4, pp. 1679–1685, 2018. doi: 10.18517/ijaseit.8.4-2.6835

[53] Chingmuankim and R. Jindal, "A comparative study of Naive Bayes classifiers with improved technique on text classification," *TechRxiv.*, pp. 1–8, 2022. doi: 10.36227/techrxiv.19918360.v1

[54] A. S. T. Nishadi, "Text analysis: Naïve bayes algorithm using python Jupyterlab," *International Journal of Scientific and Research Publications (IJSRP)*, vol. 9, no. 11, p9515, Nov. 2019. doi: 10.29322/ijsrp.9.11.2019.p9515

[55] M. F. Ibrahim, M. A. Alhakeem, and N. A. Fadhil, "Evaluation of naïve bayes classification in Arabic short text classification," *Al-Mustansiriyah Journal of Science*, vol. 32, no. 4, pp. 42–50, Nov. 2021. doi: 10.23851/mjs.v32i4.994

[56] E. Ilic, M. G. Martinez, and M. S. Pastor, "A review of Text classification models from Bayesian to transformers," in *Proc. CEUR Workshop*, 2023, pp. 1–5.

[57] R. Wongso, F. A. Luwinda, B. C. Trisnajaya, O. Rusli, and Rudy, "News article text classification in Indonesian language," *Procedia Computer Science*, 2017, pp. 137–143. doi: 10.1016/j.procs.2017.10.039

[58] Z. Chen, L. J. Zhou, X. Da Li, J. N. Zhang, and W. J. Huo, "The Lao text classification method based on KNN," *Procedia Computer Science*, pp. 523–528, 2020. doi: 10.1016/j.procs.2020.02.053

[59] U. A. Siddiqua, T. Ahsan, and A. N. Chy, "Combining a rule-based Classifier with ensemble of feature sets and machine learning techniques for sentiment analysis on microblog," in *Proc. International Conference on Computer and Information Technology*, Dhaka, 2016, pp. 304–309. doi: 10.1109/ICCITECHN.2016.7860214

[60] S. Mocherla, A. Danehy, and C. Impey, "Evaluation of naive bayes and support vector machines for Wikipedia," *Applied Artificial Intelligence*, vol. 31, no. 9–10, pp. 733–744, Nov. 2017. doi: 10.1080/08839514.2018.1440907

[61] N. Kadek, M. Budayani, I. Slamet, and S. S. Handajani, "A comparison of SVM kernel functions for sentiment analysis of UU TPKS," in *Proc. International Conference on Religion, Science and Education*, 2023, pp. 761–765.

[62] A. N. Ulfah, M. K. Anam, N. Y. S. Munti, S. Yaakub, and M. B. Firdaus, "Sentiment analysis of the convict assimilation program on handling COVID-19," *JUITA: Jurnal Informatika*, vol. 10, no. 2, pp. 209–216, 2022. doi: 10.30595/juita.v10i2.12308

[63] H. A. Azies, D. Trishnanti, and E. Mustikawati, "Comparison of kernel Support Vector Machine (SVM) in classification of Human Development Index (HDI)," in *Proc. International Conference on Global Development*, 2019. doi: 10.12962/j23546026.y2019i6.6394

[64] S. Panja, A. Chatterjee, and G. Yasmin, "Kernel functions of SVM: A comparison and optimal solution," *Communications in Computer and Information Science*, 2019, pp. 88–97. doi: 10.1007/978-981-13-3140-4_9
[65] F. K. Aljwari, "Emotion detection in Arabic text Using machine learning methods," *International Journal of Information System and Computer Science*, vol. 6, no. 5, pp. 175–185, 2022. doi: 10.56327/ijiscs.v6i3.1322
[66] A. Muis, A. Mubarak, A. M. Mamonto, and S. D. Surya, "Support Vector Machine (SVM) algorithm for student sentiment analysis of online lectures," *Jurnal Informatika dan Komputer*, vol. 6, no. 1, pp. 43–51, Apr. 2023. doi: 10.33387/jiko.v6i1.5836
[67] T. Sabbah *et al.*, "Modified frequency-based term weighting schemes for text classification," *Appl. Soft. Comput.*, vol. 58, pp. 193–206, Sep. 2017. doi: 10.1016/j.asoc.2017.04.069
[68] K. D. Y. Wijaya and A. E. Karyawati, "The effects of different kernels in SVM sentiment analysis on mass social distancing," *Jurnal Elektronik Ilmu Komputer Udayana*, vol. 9, no. 2, pp. 162–168, 2020. doi: 10.24843/JLK.2020.v09.i02.p01
[69] M. Płaza, S. Trusz, J. Kęczkowska, E. Boksa, S. Sadowski, and Z. Koruba, "Machine learning algorithms for detection and classifications of emotions in contact center applications," *Sensors*, vol. 22, no. 14, Jul. 2022. doi: 10.3390/s22145311
[70] G. S. Kumar, A. Arun, N. Sampathila, and R. Vinoth, "Machine learning models for classification of human emotions using multivariate brain signals," *Computers*, vol. 11, no. 10, Oct. 2022. doi: 10.3390/computers11100152
[71] T. A. Lasko, J. G. Bhagwat, K. H. Zou, and L. Ohno-Machado, "The use of receiver operating characteristic curves in biomedical informatics," *Journal of Biomedical Informatics*, vol. 38, no. 5. pp. 404–415, Oct. 2005. doi: 10.1016/j.jbi.2005.02.008
[72] D. Brzezinski and J. Stefanowski, "Prequential AUC: Properties of the area under the ROC curve for data streams with concept drift," *Knowl. Inf. Syst.*, vol. 52, no. 2, pp. 531–562, Aug. 2017. doi: 10.1007/s10115-017-1022-8
[73] K. S. Nugroho, F. A. Bachtiar, and W. F. Mahmudy, "Detecting emotion in Indonesian tweets: A term-weighting scheme study," *Journal of Information Systems Engineering and Business Intelligence*, vol. 8, no. 1, pp. 61–70, Apr. 2022. doi: 10.20473/jisebi.8.1.61-70
[74] M. Kastrati, M. Biba, A. S. Imran, and Z. Kastrati, "Sentiment polarity and emotion detection from tweets using distant supervision and deep learning models," *Lecture Notes in Computer Science*, 2022, pp. 13–23. doi: 10.1007/978-3-031-16564-1_2