Public Response to the Legalization of The Criminal Code Bill with Twitter Data Sentiment Analysis

Deny Irawan[1], Dana Indra Sensuse[2], Prasetyo Adi Wibowo Putro[3], Aji Prasetyo[4]
Faculty of Computer Science, Universitas Indonesia, Jakarta, Indonesia[1, 2, 3, 4]

***Abstract*—The Criminal Code Bill, also known as Rancangan Kitab Undang-undang Hukum Pidana (RKUHP), passed in the House of Representatives (DPR) on December 6, 2022, is being debated because several issues need to be fixed. Therefore, research was conducted to determine the public's reaction to the ratification of the Criminal Code Bill by analyzing Twitter data. This study aims to obtain a general response to the legalized RKUHP. We use sentiment analysis, a text-processing method, to get data from the public. To do this, we used N-grams (unigrams, bigrams, and trigrams) along with three algorithms: Naïve Bayes, Classification and Regression Tree (CART), and Support Vector Machine (SVM). The result of sentiment analysis found that 51% of tweets were positive about the ratification of the RKUHP, and 49% were negative. In addition, it was also found that SVM has the best accuracy compared to other algorithms, with an accuracy value of 0.81 on the unigram combination.**

Keywords—Sentiment analysis; RKUHP; support vector Machine (SVM); Naïve Bayes; classification and regression tree (CART)

I. Introduction

The Criminal Code Bill, also known as RKUHP, was signed into law on December 6, 2022. This is a momentous occasion, as this law would replace the existing Criminal Code (KUHP) [1]. Because the current Criminal Code is a legacy of Dutch colonialism, it is a version of the *Wetboek van Strafrecht voor Nederlandsch-Indie*. Changes are needed because the old Criminal Code is not keeping up with the times [2]. In addition, the revision of the previous RKUHP was carried out in partly making new laws related to the KUHP, which made regulations run wild, had no system or pattern, was inconsistent, made problematic laws, and even damaged the basic building system old KUHP [3]. RKUHP will be valid for three years from the date of promulgation [4]. However, the RKUHP is considered to have problematic articles still. Nurina S. (2022), in an interview with The Guardian, stated that at least 88 reports contain broad provisions that can be exploited and misconstrued by both the government and the general public to punish anyone and suppress the freedom of expression [5].

The response to ratifying RKUHP still has pros and cons. Because responses are essential determinants of every human action, interesting to see the public's response to RKUHP; when deciding, we need others' opinions. Companies or governments must know how the public feels about their products and services. The public sometimes utilizes Facebook and Twitter to engage socially online. Web-based social networks gradually engage the public [6]. This is consistent with research conducted in the United States about the public's reaction to the Chicago Department of Public Health's laws on electronic cigarettes, which examined the public's response on Twitter. The data can help organizations predict, recognize, and respond to how the community will react by finding patterns in how people have responded to this policy [7]. In addition, according to research conducted in Mexico, governments frequently use Twitter to interact with their citizens. As a result, it has emerged as a valuable source of information for studying how governments interact with their constituents and how those citizens respond to those communications. These insights about how people interact with the government can be used to help make public policies and understand how the public sees those policies [8].

The ratification of RKUHP has the same context. It will be fascinating to watch how Twitter data is utilized to gauge public opinion toward the ratification of RUU KUHP because these messages on Twitter are openly accessible. Consequently, it can be viewed as raw data primarily for the extraction of opinions and for the analysis of policy by analyzing the sentiment [9]. This will aid the government's ability to forecast, detect, and respond to the public's reaction to the dissemination of information before it is completely implemented. Sentiment analysis is another term for "opinion mining" or "emotion Artificial Intelligence". It refers to applying natural language processing (NLP), Using text mining, computational linguistics, and biometrics to carefully identify, extract, assess, and look into people's emotions and personal data [6].

This research aims to identify the sentiments surrounding the ratification of the RKUHP. The analysis results are reprocessed to determine what aspects of RKUHP concern the public. The use of public sentiment will assist the government in gauging the public's reaction to the ratification of the RKUHP and can be utilized as input for the planned socialization. In addition, by using multiple algorithm models, this research will identify the optimal categorization model that might be used by the government when trying to determine public responses with data from twitter.

This research consists of five sections. The introduction, which contains the research's context and objectives, is the first section. The second section is a review of previous research and the theoretical framework. The study's research methodology is described in the third section. The fourth section is the results and discussion, which includes the findings from the research. The conclusion is the concluding section of the study.

II. Literature Review

A. Previous Research

This section contains considerable research that employs various methods for sentiment analysis. The first research authors use one methodology for measurement sentiment analysis, shown in Table I. Authors in [10] have investigated the Naïve Bayes algorithm's capacity to classify public mood under COVID-19's new normal. From the 2807 tweets that have been processed, the test results show that Naïve Bayes has done an excellent job, with an accuracy of 83% and an F1-score of 84%. The author in [11] researched sentiment analysis using the Support Vector Machine (SVM) with Weka (Waikato Environment for Knowledge Analysis) method and tested it on three different data sets with various labels. Because of this, the data set with the highest f1-score is the third one, which only has two titles: positive and negative.

Further research uses two methodologies for measurement sentiment analysis. In a study [12], sentiment analysis tests on comments on YouTube using Naïve Bayes and Support Vector Machines (SVM). Results when using a data scale of 7:3, with 70% of the data used for training and 30% for testing, show that the combination of Naïve Bayes and SVM results in higher accuracy and superior performance. In a study [13], researchers compare Naïve Bayes and SVM to evaluate the classification results that each method produces. Twitter data is used in this study for Tokopedia services. The outcomes demonstrated that, with an accuracy of 83.34%, the SVM linear kernel technique surpassed the Naïve Bayes technique. In a study [14] Using Twitter data, researchers assess the sentiment analysis of the COVID-19 virus infection on Indonesian public transportation. In this study, the authors used two comparison methods: Naïve Bayes and decision trees. The result is that Naïve Bayes outperforms the Decision Tree with an accuracy of 73.59%.

The third research uses more than two methodologies for measurement sentiment analysis. In a study [15], Researchers researched the sentiment analysis of tourists in Thailand during the COVID-19 pandemic. This study used three methods: SVM, Classification and Regression Tree (CART), and random forest. Consequently, SVM could identify the attitudes and intentions of the English-language tweets that included Phuket and Chiang Mai the best. Still, for tweets mentioning Bangkok, CART is the most accurate, with accuracies of 94.3%. Bangkok has more data tweets than others. Subsequent research, customer reviews of Amazon products. Researchers in this study [16] used four sentiment analysis methods: Naïve Bayes, SVM, Decision Tree, and K-Nearest Neighbor. In addition, this research also added TF-IDF and N-gram to its processing. The results of the TF-IDF method with N-grams show unigrams with SVM were the maximum accuracy results for Amazon product customer reviews. This study also found that comments on Amazon products influence potential consumers' purchasing decisions. The two studies [15][16] were conducted to determine the differences and accuracy of the sentiment analysis method.

TABLE I. Previous Research

No	Ref	Algorithm and Method	Sentiment Analysis and Objective	Result
1	[10]	Naïve Bayes	Provides a sentiment analysis of how well society is accepting the new normal with data from Twitter and investigates the Naïve Bayes algorithm's capacity to classify public mood under COVID-19's new normal.	During the COVID-19 pandemic, the majority of people were able to adjust to their new everyday normal. The test results show that Naïve Bayes has done an excellent job.
2	[11]	SVM with Weka (Waikato Environment for Knowledge Analysis)	This study didn't specifically look at sentiment analysis. Instead, it used three different data sets, two from Twitter and one from the Internet Movie Database, to test how well the SVM algorithm worked (IMDB).	The highest-scoring data set is the third (IMDB data), with only two titles: positive and negative.
3	[12]	Naïve Bayes and SVM	Provides a sentiment analysis of positive and negative YouTube comments and evaluates the combination of two algorithms, naïve Bayes and SVM.	The combination of Naïve Bayes and SVM results in higher accuracy and superior performance for seeing sentiment in YouTube comments.
4	[13]	SVM and Naïve Bayes	Sentiment analysis for Tokopedia service with data from Twitter and evaluation of the performance of Naïve Bayes and SVM.	The data do not specify the sentiment analysis results for the Tokopedia service; they evaluate that SVM is more accurate than Naïve Bayes.
5	[14]	SVM and Decision Tree	Sentiment analysis to determine what commuter line riders think about how the Covid-19 pandemic could spread on public transportation—and Comparison Accuracy from algorithm naïve Bayes and decision tree.	Most people in the community have a positive outlook that includes a plea and a call to stop the COVID-19 outbreak and get it under control. With an accuracy of 73.59%, Naive Bayes is better than the Decision Tree.
6	[15]	decision tree, random forest, and SVM with TF-IDF and combination ngram (unigram, bigram, and trigram)	Sentiment analysis to find out the expression of tourists about tourist attractions, events, festivals, and experiences from July to December 2020 whit data from Twitter.	The results showed the top 10 words for each type of feeling, which can be looked at to learn more and give the right advice.
7	[16]	SVM with a combination of term weighting and ngram	This study aims to assess the impact of sentiment (positive, negative, and neutral) and Amazon product reviews on sales performance. Also, to	The Result found that comments on Amazon products influence potential consumers' purchasing decisions. In

			identify the optimal combination of SVM, TF-IDF, and ngram.	addition, the TF-IDF method with N-grams shows unigrams with SVM were the maximum accuracy results for Amazon product customer reviews.

Based on previous research, researchers will use the Naïve Bayes [10][13][14], SVM [11][12] [13][15][16], and CART [15] in evaluating sentiment analysis. In addition, the N-gram and TF-IDF methods will be used because they are proven to increase accuracy [16]. The study used positive and negative labels because it was established in research [11] that they have the highest accuracy compared to data using more than two labels.

B. Sentiment Analysis

According to Pang et al. (2002), opinion mining and sentiment analysis are two terms that refer to the same process. Sentiment analysis automatically analyzes, extracts, and textually processes material to derive the sentiment information in a single opinion sentence. An individual's perspective, or their predisposition to have a positive or negative view or opinion about a particular issue or object, can be determined using a technique known as "sentiment analysis" [17] [12].

C. Data Preprocessing

Data Preprocessing involves converting raw data into a format the user may understand. Frequently, the data must be more structured and consistent, lack specific behaviors or patterns, and contain missing values, all of which contribute to many errors. Consequently, it needs to be cleaned, integrated, altered and decreased. The noise is eliminated, and missing values are filled in when cleaning is performed [18][19].

D. N-Gram

The word n-gram feature counts sets of sequential N words in each tweet, where N can range from 1 to N. [20]. N-grams can be more informative. There could be t^2 bigrams containing t different words. In practice, only some characteristics are generated because terms can't follow each other. Usually, n-grams are more distinct than words. A more extensive, less common feature space is an n-gram. A larger n increases information and computational expense [21]. In this research, we combine the unigram, the bigram, and the trigram forms of the n-gram.

E. Term Frequency - Inverse Document Frequency (TF-IDF)

According to Jones (1972), Inverse Document Frequency (IDF) is a technique that can be combined with term frequency to lessen the influence of implicitly famous words in the corpus. This is how IDF is meant to be used. IDF gives greater weight to terms that appear more frequently in the document, regardless of whether those words are used often or infrequently [22][23]. TF-IDF is now the most popular text classification and document categorization scheme [24][21].

F. Naïve Bayes Algorithm

This categorization method is based on Bayes' Theorem and makes strong (naive) assumptions about feature independence. A Naïve Bayes classifier makes the following assumptions: that the proximity of one feature (element) within a class is unrelated to the proximity of other items. The Naïve Bayes algorithm is often used to divide texts into different groups, and it was recently used to separate data from sentiment analysis into groups [6].

The algorithm relies on Bayes' theorem and presumes that the class variable's value provides information for all variables independently. It is simple to program the Naïve Bayes classification algorithm to perform exceptionally well in supervised learning, and it can also be used in difficult real-world situations. The Naïve Bayes method is simple to grasp, needs an education dataset to figure out how to calculate its variables, doesn't care about things that have nothing to do with the problem, and works well with correct data from a single source [25][10].

G. Support Vector Machine (SVM) Algorithm

According to Han et al. (2012), the Support Vector Machine (SVM) algorithm's goal is to locate the Maximum Marginal Hyperplane (MMH) by utilizing margins and support vectors. The MMH hyperplane is the best one available since it has the most significant margin distance and can be used to accurately and maximally segregate data for each class. Suppose both margins are in a position that is parallel to the hyperplane. In that case, the margin is defined as the point at which the shortest distance from a hyperplane to one side equals the distance from the hyperplane to the other side of the margin [26][24].

H. Classification and Regression Tree (CART)

The classification and regression trees (CART) method is a systematic technique that was developed by Breiman et al. (1984) [27][28]. For the construction of decision trees, CART employs historical data. The dependent variable decides whether a classification tree (for categorical categories) or a regression tree (for variables with continuous categories) will be formed. The newly discovered observations can then be predicted (using a regression tree) and classified (using a classification tree) using the constructed tree. Contrary to classification trees, regression trees do not have any pre-determined classes. On the other hand, classification trees allow the user to select or calculate dependent variable types based on an external criterion. [27][29][30][28]. The CART approach consists of three steps: (1) the creation of the entire tree; (2) the selection of the ideal tree size; and (3) the evaluation of the results. (3) using a built tree to organize data or generate new information[28].

III. Methodology

The research consisted of several stages, including the collection of data, the creation of data sets, the labeling of data, the processing of data, the grouping of words using n-grams and term weighting (TF-IDF), classification modeling, the evaluation of classification modeling, and, finally, the output of sentiment results and recommendations. This is shown in **Error! Reference source not found.**.

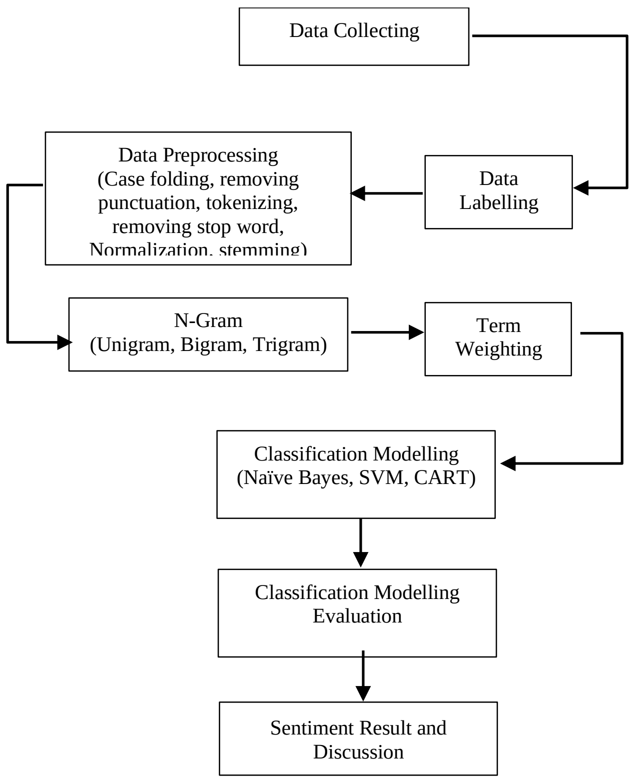

Fig. 1. Methodology for research

A. Data Collecting

Python and the twitter-snscrape library package are used to harvest Twitter tweet data at this step. Data was gathered using a search for the phrase "RKUHP" tweeted between December 6, 2022, and December 31, 2022. Tweets taken are in Indonesian, and identical tweets will be deleted. Related Tweets that only use the RKHUP hashtag and only contain ads will also be disqualified. Tweets are not converted into English due to possible differences in meaning in processing. All words resulting from sentiment will use the Indonesian language.

B. Data Labelling

In this phase, the training data is labeled manually whether tweets are positive (pro) or negative (con) with the ratification of the RKUHP. In this phase, irrelevant tweets are also deleted.

C. Data Preprocessing

In the preprocessing of tweet data, a series of operations are performed so that machine learning algorithms can read the tweet's standards and patterns in Table II. The method is as follows:

1) Case Folding turns all capital characters in tweets into lowercase letters.

2) Remove Punctuation and eliminates punctuation, URL links, numbers, hashtags, and emoticons from tweets.

3) Tokenization is the process of breaking sentences into separate words.

4) Stop Word is the process of removing words that don't add any meaning.

5) Normalization is the process of uniforming words with the same meaning but different spellings.

6) Stemming is changing words that have affixes into essential words.

TABLE II. PREPROCESSING PROCESS

Process	Tweet
User Tweet	Dukung pengesahan RKUHP untuk supremasi hukum https://t.co/MLmR6BHBIS #DukungPengesahanKUHP
Case folding	dukung pengesahan rkuhp untuk supremasi hukum https://t.co/mlmr6bhbis #dukungpengesahankuhp
Remove Punctuation	dukung pengesahan rkuhp untuk supremasi hukum
Tokenizing	['dukung', 'pengesahan', 'rkuhp', 'untuk', 'supremasi', 'hukum']
Stop Word	['dukung', 'pengesahan', 'rkuhp', 'supremasi', 'hukum']
Normalization	['dukung', 'pengesahan', 'rkuhp', 'supremasi', 'hukum']
Stemming	['dukung', 'kesah', 'rkuhp', 'supremasi', 'hukum']

D. N-Gram

In this phase, word separation is carried out; we combine the unigram, the bigram, and the trigram forms of the n-gram. Words are created using unigrams (one word), bigrams (two words), and trigrams (three words). Tweets that have at most three words will be deleted.

E. Term Weighting

The next step was word feature extraction using the term frequency-inverse document frequency (TF–IDF). The word weight in a given document is typically calculated using the TF-IDF technique. The term frequency describes the often appearing; that often appears in a manuscript (TF). Frequently occurring terms will obstruct the search for uncommon words. The inverse document frequency (IDF), which lessens the weight of often-appearing words, can gauge how significant a word's meaning is in a document [31].

F. Classification Modelling

In this step, classification modeling is applied to the test data using three machine learning algorithms: Nave Bayes, SVM, and CART. Modeling is done separately to produce accurate results. Each algorithm tests the words formed in the ngram process, and the term weighting process has been carried out. This classifier uses the sklearn library in Python. This study used 80% training data and 20% testing data. This is so that machine learning algorithms can perform better, according to research by Pham et al. in Nguyen et al. research, when training data is raised from 30% to 80%. However, when it is increased from 80% to 90%, the opposite occurs[32] [33].

G. Classification Modelling Evaluation

In this phase, the performance of each machine learning algorithm in the previous step will be evaluated. Evaluation is conducted using a confusion matrix by looking at the value of the accuracy of each algorithm. Accuracy, precision, and recall are the evaluative test parameters whose computations are derived from the confusion matrix table [13].

H. Sentiment Result and Discussion

This is the final stage in producing sentiment words for the word cloud. Which, according to the N-gram phase, consists of one word, two words, and three terms and is derived from the sentiment with the maximum accuracy. Then, the discussion will be made in light of these findings.

IV. Result and Discussion

A. Results of Classification Modeling Evaluation

The number of tweets extracted using the snscrape library is 17,107. After cleaning the same tweets, the number of tweets increases to 10,763. Then, label each tweet manually. Then, preprocessing process the tweet and generate it again to yield 9,079 tweets. The tweet then executed the classification algorithms and the n-gram combination method. After preprocessing, the dataset is split into training and test sets. 80% of the dataset is used for training, and the remaining 20% is used for testing. The dataset's features are produced using an n-gram mix of unigrams, bigrams, and trigrams. The created words will then be weighted using term analysis. Different data are made when n-grams and term weighting are combined. The results are presented in Tables III, IV, and V.

Using the confusion metrics, we have calculated the performance of each algorithm here. The confusion matrix, which measures the classification overlap, is an effective tool for performance evaluation. The multi-label classification task must establish the confusion matrix because each instance may be assigned to multiple classes [34]. The performance evaluation of the multi-label classifier is based on computing performance averages, including precision, recall, and F1-score [34]. Precision measures how accurate a class's predictions are relative to all the predictions included in the course. Recall is the percentage of a class's total number of categorized facts that can be predicted accurately. The f1 score calculation was then utilized to mix the precision and recall [35] [12].

For each n-gram combination used, precision, recall, and f1-scores for the CART algorithm are displayed in table III. The findings of CART do not differ much when unigrams, bigrams, or trigrams are used. In the bigram findings, for example, the precision value is 0.73 for negative and 0.75 for positive, and the recall value is 0.70 for negative and 0.75 for positive. The f1 values for positive and negative are then 0.72 and 0.74, respectively. As shown in Fig. 2, out of the 852 negatively judged tweets, 624 were true negatives (TN), and 228 were false negatives. In contrast, out of 964 positive tweets, 263 were false positives (FP), and 701 were true positives (TP).

TABLE III. Precision, Recall, and F1-Score Cart

	N-Gram	Precision	Recall	F1-score
Negative	**Unigram**	0.74	0.70	0.72
	Bigram	0.73	0.70	0.72
	Trigram	**0.75**	0.69	0.72
Positive	**Unigram**	0.72	0.76	0.74
	Bigram	0.73	0.75	0.74
	Trigram	**0.71**	**0.77**	0.74

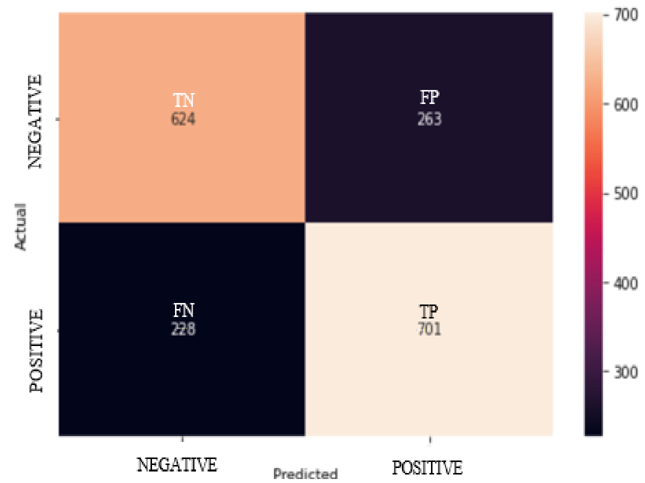

Fig. 2. The confusion matrix bigram CART

Table IV shows the SVM algorithm's precision, recall, and f1-score for each n-gram combination. The unigram test had the best average outcomes, with precision values of 0.81 for negative and 0.80 for positive groups and recalled values of 0.76 for negative and 0.82 for positive. The f1-score is 0.78 for the negative and 0.81 for the positive. As shown in **Error! Reference source not found.**, 711 of the 877 tweets that received a negative evaluation were true negatives, and 166 were false negatives. Comparatively, out of 939 positive tweets, 186 were false positives, and 753 were true positives.

TABLE IV. Precision, Recall, and F1-Score SVM

	N-Gram	precision	recall	f1-score
Negative	**Unigram**	**0.81**	**0.79**	**0.80**
	Bigram	0.79	**0.78**	0.79
	Trigram	0.81	0.76	0.78
Positive	**Unigram**	**0.80**	**0.82**	**0.81**
	Bigram	0.79	0.81	0.80
	Trigram	0.78	0.82	0.80

Fig. 3. The confusion matrix unigram SVM

The Naïve Bayes algorithm is presented in Table V for each possible combination of n-grams. The trigram test had the most outstanding results overall, with accuracy scores of 0.78 for negative repones and 0.79 for positive ones, recall scores of 0.79 for negative answers and 0.77 for positive reactions, and an f1-score of 0.78 for negative and positive responses. As shown in **Error! Reference source not found.**, of the 926 tweets that received a negative rating, 718 were true negatives, and the remaining 208 were false negatives. The 890 positive tweets, in contrast, contained 186 false positives and 704 true positives.

TABLE V. PRECISION, RECALL AND F1-SCORE NAÏVE BAYES

	N-Gram	precision	recall	f1-score
Negative	**Unigram**	0.75	**0.77**	0.76
	Bigram	0.75	0.79	0.77
	Trigram	**0.78**	0.79	**0.78**
Positive	**Unigram**	0.77	0.75	0.76
	Bigram	0.79	0.75	0.77
	Trigram	**0.79**	**0.77**	**0.78**

Fig. 4. The confusion matrix trigram naïve bayes

Calculating the accuracy of each method is another function of the confusion matrix, which can be seen in Table VI. It has been demonstrated that the SVM constructed using the unigram has the maximum accuracy, equal to 0.81. In addition, bigram and trigram SVM continues to have the highest accuracy compared to other algorithms, with respective values of 0.79 and 0.78. When utilizing trigram combinations, Naïve Bayes on 0.78 achieves a higher level of accuracy. CART has the same accuracy in all ngram combinations.

TABLE VI. ACCURACY CART, SVM AND NAÏVE BAYES

Algorithm	NGRAM Accuracy		
	Unigram	Bigram	Trigram
CART	0.73	0.73	0.73
SVM	**0.81**	0.79	0.79
Naïve Bayes	0.76	0.77	0.78

The analysis results in Table VI are consistent with research in [13] and [15] that shows that SVM has higher accuracy when compared to Naïve Bayes and CART. The SVM study achieved an accuracy of 83.34% and a Naïve Bayes of 75%. According to research [15], the amount of data used by the random forest and CART algorithms determines the soundness of multiple decision trees, the complexity of the trees, and thus the algorithm's accuracy. This explains why CART has the same accuracy because it has the same number of data sets.

In line with the results of this investigation, a study in [16] discovered that SVM with the unigram combination had the highest accuracy when compared to the other ngram combinations. This is likely due to the ease with which SVM can map words weighted with TF-IDF rather than utilizing multiple words to infer sentiment. This compares favorably with naïve Bayes, where the more word combinations in ngrams, the higher the accuracy. Many ngram combinations raise the level of accuracy in Naïve Bayes. Therefore, based on this, it was found that the combination would affect the accuracy of each algorithm. SVM is preferred over algorithms, Naïve Bayes, and CART because of its high accuracy. For Naïve Bayes, a higher gram would be preferable. Because CART is affected by a large amount of data, vast amounts of data will affect its accuracy.

B. Content Sentiment Analysis

Nine thousand seventy-nine tweets were included in the data obtained after being processed using Python and Microsoft Excel programming languages. Several duplicate and irrelevant tweets have been removed from the message. The result is that 51% of tweets, or 4.623 of them, favor the ratification of the RKUHP, while 49% of tweets, or as many as 4.455 of them, are in opposition to it, as can say be seen in **Error! Reference source not found.**.

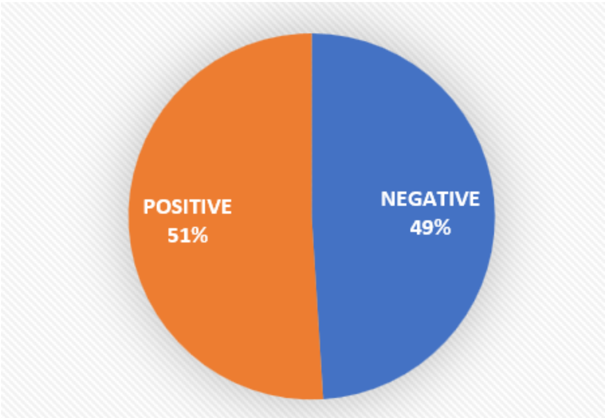

Fig. 5. Result of the sentiment of the ratification of RKUHP

The word cloud for a negative sentiment is displayed in **Error! Reference source not found.**. Negative sentiment is associated with a wide variety of topics and concepts, including ‘tolak’, ‘kontroversi’, ‘kritik’, ’hina’, ’koruptor’, ’penting kuasa’ The most common words are 'sah,' 'tolak,' 'rakyat,' and 'negara' in that order. The RKUHP received a negative response because it was thought to contain several articles that could be construed as contentious. Based on word cloud sentiments such as the words ‘kritik’,’hina’ and ‘demokrasi’, some articles are considered to silence criticism, specifically regarding insulting the president. Then the words ‘koruptor’ and ‘korupsi’ articles regarding corruption, with a minimum reduction in prison for corruption. Then there's the phrase ‘penting kuasa’ and ‘rakyat’, because some people believe that a lot of the new RKUHP articles were written more for the authorities’ interests than for the people’s interests themselves.

Fig. 6. Word cloud of negative sentiment

Besides having negative sentiments, there are also words representing positive sentiments in the word cloud. Word like ‘sah’, ‘baru’, ‘hukum pidana’, ‘baru’ dan ’tuju’ support the approval of the ratification of this RKUHP. The RKUHP is significant because it strengthens Indonesia's current criminal code. In Indonesia, criminal law that has undergone patchwork is no longer regarded as complying with legal criteria. The positive word Cloud is shown in **Error! Reference source not found.**.

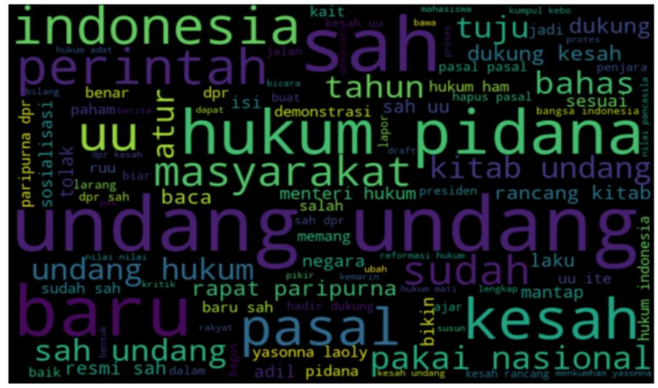

Fig. 7. Word cloud of positive sentiment

We may deduce what words are at the center of people's conversations based on the outcomes of positive and negative sentiments in **Error! Reference source not found.** and **Error! Reference source not found.**. The words that arise may serve as a first reflection for the organization of positive and negative things that are the community's response. This can be used as a resource for organizations to improve their understanding of the policies they issue. This is in line with the findings of a study [8] on how the Mexican government uses Twitter to connect with the people.

As a result, the outcomes of these attitudes can be employed by the government as a foundation for socialization. Because there is still the problem of the pessimism of RKHUP, there are still drawbacks to the ratification of the RKUHP, which is still relatively high and reaches 49% of the population. To find a solution to this problem, the government needs to engage in more social activities and listen to people's perspectives on matters that are regarded as contentious. This is done to ensure that both the adoption of the RKUHP in 2025 and its passage into the Criminal Code happen smoothly. Words that elicit negative responses might be utilized as the primary focus of socialization. This will help in mitigating the public's adverse reaction. Mitigation of this rejection will be better if there is a grouping of tweets based on topic, as in Research [36]. In this study, we used BERTaopic to classify the tweets. BERTopik will help categorize tweets and make it easier for the government to bring up specific RKUHP-related topics. Throughout the processing, it was discovered that the steaming method had limitations since particular terms, such as “pengesahan”, were mistakenly converted into the root word “kesah”. Hence the potential limits of the world cloud’s word interpretation.

V. Conclusion

Examining Twitter sentiment, this study identifies responses to the ratification of the RKUHP. The RKUHP ratification drew 51% positive and 49% negative comments on Twitter. This demonstrates that, even though the positive is superior, the value is just 2%. According to the negative comments, the problem of controversial articles is related to the article about insulting the president, the post about cutting punishments for corrupt officials, and the piece about not representing the people. This must be the emphasis of the government's efforts to socialize the RKUHP.

The evaluation of the three tested algorithms—CART, SVM, and Naïve Bayes —found that SVM had the highest accuracy and was the most reliable even when the n-gram combination was used. SVM produces an accuracy value of 0.81 on the unigram, 0.79 on the bigram, and 0.79 on the trigram.

This research is limited to grouping tweets that have yet to be grouped into specific topics and imperfections in the steaming process. It is hoped that future research can categorize recent tweets based on grouping relevant issues related to the RKUHP so that they are not only the results of grouping terms from the Word Cloud. It can also add more data which makes the topic even better. In addition, it can improve the algorithm steaming process to make it better.

ACKNOWLEDGMENT

Thanks to the Indonesian Ministry of Communication and Information for supporting, assisting, and funding this research. As well as support from the Faculty of Computer Science, University of Indonesia.

REFERENCES

[1] Humas dan Protokol BPHN, "RUU KUHP Disahkan menjadi Undang-undang," 2022. https://bphn.go.id/publikasi/berita/202212061210189/ruu-kuhp-disahkan-menjadi-undang-undang (accessed Dec. 16, 2022).

[2] Y. Y. WEDHA and E. NURCAHYO, "Criminal Law Reform Toward Deprivation of Property Resulting from Corruption Criminal Acts," PRIZREN Soc. Sci. J., vol. 5, no. 1, pp. 97–103, Apr. 2021, doi: 10.32936/pssj.v5i1.207.

[3] M. F. Butar-butar, "Exemptions from Liability in Indonesian Criminal Law Reform," Ann. R.S.C.B, vol. 25, no. 5, pp. 5528–5533, 2021.

[4] "UU KUHP Telah Akomodir Seluruh Aspirasi Masyarakat Indonesia," 2022. https://www.dpr.go.id/berita/detail/id/42227/t/UU+KUHP+Telah+Akomodir+Seluruh+Aspirasi+Masyarakat+Indonesia (accessed Dec. 16, 2022).

[5] S. Strangio, "Indonesia Set to Pass Controversial New Criminal Code This Month," The Diplomat, Dec. 2022. [Online]. Available: https://www.proquest.com/magazines/indonesia-set-pass-controversial-new-criminal/docview/2745669907/se-2.

[6] A. Alsaeedi and M. Z. Khan, "A study on sentiment analysis techniques of Twitter data," Int. J. Adv. Comput. Sci. Appl., vol. 10, no. 2, pp. 361–374, 2019, doi: 10.14569/ijacsa.2019.0100248.

[7] J. K. Harris, S. Moreland-Russell, B. Choucair, R. Mansour, M. Staub, and K. Simmons, "Tweeting for and against public health policy: Response to the Chicago Department of Public Health's electronic cigarette Twitter campaign," J. Med. Internet Res., vol. 16, no. 10, p. e238, 2014, doi: 10.2196/jmir.3622.

[8] R. B. Hubert, E. Estevez, A. Maguitman, and T. Janowski, "Analyzing and Visualizing Government-Citizen Interactions on Twitter to Support Public Policy-making," Digit. Gov. Res. Pract., vol. 1, no. 2, pp. 1–20, 2020, doi: 10.1145/3360001.

[9] M. A. Kausar, A. Soosaimanickam, and M. Nasar, "Public Sentiment Analysis on Twitter Data during COVID-19 Outbreak," Int. J. Adv. Comput. Sci. Appl., vol. 12, no. 2, pp. 415–422, 2021, doi: 10.14569/IJACSA.2021.0120252.

[10] S. H. A. Samsudin, N. M. Sabri, N. Isa, and U. F. M. Bahrin, "Sentiment Analysis on Acceptance of New Normal in COVID-19 Pandemic using Naïve Bayes Algorithm," Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 9, pp. 581–588, 2022, doi: 10.14569/IJACSA.2022.0130968.

[11] M. Ahmad, S. Aftab, M. S. Bashir, N. Hameed, I. Ali, and Z. Nawaz, "SVM optimization for sentiment analysis," Int. J. Adv. Comput. Sci. Appl., vol. 9, no. 4, pp. 393–398, 2018, doi: 10.14569/IJACSA.2018.090455.

[12] A. N. Muhammad, S. Bukhori, and P. Pandunata, "Sentiment Analysis of Positive and Negative of YouTube Comments Using Naïve Bayes-Support Vector Machine (NBSVM) Classifier," Proc. - 2019 Int. Conf. Comput. Sci. Inf. Technol. Electr. Eng. ICOMITEE 2019, pp. 199–205, 2019, doi: 10.1109/ICOMITEE.2019.8920923.

[13] R. Kusumawati, A. D'Arofah, and P. A. Pramana, "Comparison Performance of Naive Bayes Classifier and Support Vector Machine Algorithm for Twitter's Classification of Tokopedia Services," J. Phys. Conf. Ser., vol. 1320, no. 1, 2019, doi: 10.1088/1742-6596/1320/1/012016.

[14] I. C. Sari and Y. Ruldeviyani, "Sentiment Analysis of the Covid-19 Virus Infection in Indonesian Public Transportation on Twitter Data: A Case Study of Commuter Line Passengers," 2020 Int. Work. Big Data Inf. Secur. IWBIS 2020, pp. 23–28, 2020, doi: 10.1109/IWBIS50925.2020.9255531.

[15] N. Leelawat, S. Jariyapongpaiboon, A. Promjun, and S. Boonyarak, "Heliyon Twitter data sentiment analysis of tourism in Thailand during the COVID-19 pandemic using machine learning," Heliyon, vol. 8, no. September, p. e10894, 2022, doi: 10.1016/j.heliyon.2022.e10894.

[16] T. Sinnasamy and N. N. A. Sjaif, "Sentiment Analysis using Term based Method for Customers' Reviews in Amazon Product," Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 7, pp. 685–691, 2022, doi: 10.14569/IJACSA.2022.0130780.

[17] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?," in Proceedings of the ACL-02 conference on Empirical methods in natural language processing - EMNLP '02, 2002, vol. 10, pp. 79–86. doi: 10.3115/1118693.1118704.

[18] S. B. Kotsiantis and D. Kanellopoulos, "Data preprocessing for supervised leaning," Int. J. …, vol. 1, no. 2, pp. 1–7, 2006, doi: 10.1080/02331931003692557.

[19] P. Yerpude and V. Gudur, "Predictive Modelling of Crime Dataset Using Data Mining," Int. J. Data Min. Knowl. Manag. Process, vol. 7, no. 4, pp. 43–58, 2017, doi: 10.5121/ijdkp.2017.7404.

[20] O. Oriola and E. Kotze, "Evaluating Machine Learning Techniques for Detecting Offensive and Hate Speech in South African Tweets," IEEE Access, vol. 8, pp. 21496–21509, 2020, doi: 10.1109/ACCESS.2020.2968173.

[21] K. S. Nugroho et al., "Detecting Emotion in Indonesian Tweets: A Term-Weighting Scheme Study," Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 1, pp. 61–70, 2022, doi: 10.20473/jisebi.8.1.61-70.

[22] K. SPARCK JONES, "A STATISTICAL INTERPRETATION OF TERM SPECIFICITY AND ITS APPLICATION IN RETRIEVAL," J. Doc., vol. 28, no. 1, pp. 11–21, Jan. 1972, doi: 10.1108/eb026526.

[23] Kowsari, Jafari Meimandi, Heidarysafa, Mendu, Barnes, and Brown, "Text Classification Algorithms: A Survey," Information, vol. 10, no. 4, p. 150, Apr. 2019, doi: 10.3390/info10040150.

[24] F. A. Bachtiar, W. Paulina, and A. N. Rusydi, "Text Mining for Aspect Based Sentiment Analysis on Customer Review : a Case Study in the Hotel Industry," 5th Int. Work. Innov. Inf. Commun. Sci. Technol., no. March, 2020.

[25] F. Razaque et al., "Using naïve bayes algorithm to students' bachelor academic performances analysis," in 2017 4th IEEE International Conference on Engineering Technologies and Applied Sciences (ICETAS), Nov. 2017, pp. 1–5. doi: 10.1109/ICETAS.2017.8277884.

[26] J. Han, M. Kamber, and J. Pei, "Data Mining: Concepts and Techniques," Third., Boston: Elsevier, 2012, pp. 408–415. doi: 10.1016/B978-0-12-381479-1.00027-7.

[27] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, Classification and Regression Trees. Routledge, 1984. doi: 10.1201/9781315139470.

[28] B. Choubin, G. Zehtabian, A. Azareh, E. Rafiei-Sardooi, F. Sajedi-Hosseini, and Ö. Kişi, "Precipitation forecasting using classification and regression trees (CART) model: a comparative study of different approaches," Environ. Earth Sci., vol. 77, no. 8, pp. 1–13, 2018, doi: 10.1007/s12665-018-7498-z.

[29] R. Singh, T. Wagener, R. Crane, M. E. Mann, and L. Ning, "A vulnerability driven approach to identify adverse climate and land use change combinations for critical hydrologic indicator thresholds: Application to a watershed in Pennsylvania, USA," Water Resour. Res., vol. 50, no. 4, pp. 3409–3427, Apr. 2014, doi: 10.1002/2013WR014988.

[30] B. Choubin, H. Darabi, O. Rahmati, F. Sajedi-Hosseini, and B. Kløve, "River suspended sediment modelling using the CART model: A comparative study of machine learning techniques," Sci. Total Environ., vol. 615, pp. 272–281, Feb. 2018, doi: 10.1016/j.scitotenv.2017.09.293.

[31] Raksaka Indra Alhaqq, I Made Kurniawan Putra, and Yova Ruldeviyani, "Analisis Sentimen terhadap Penggunaan Aplikasi MySAPK BKN di Google Play Store," J. Nas. Tek. Elektro dan Teknol. Inf., vol. 11, no. 2, pp. 105–113, 2022, doi: 10.22146/jnteti.v11i2.3528.

[32] B. T. Pham et al., "A Novel Hybrid Soft Computing Model Using Random Forest and Particle Swarm Optimization for Estimation of Undrained Shear Strength of Soil," Sustainability, vol. 12, no. 6, p. 2218, Mar. 2020, doi: 10.3390/su12062218.

[33] Q. H. Nguyen et al., "Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil," Math. Probl. Eng., vol. 2021, pp. 1–15, Feb. 2021, doi: 10.1155/2021/4832864.

[34] M. Heydarian, T. E. Doyle, and R. Samavi, "MLCM: Multi-Label Confusion Matrix," IEEE Access, vol. 10, pp. 19083–19095, 2022, doi: 10.1109/ACCESS.2022.3151048.

[35] D. H. Wahid and A. SN, "Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity," IJCCS (Indonesian J. Comput. Cybern. Syst., vol. 10, no. 2, p. 207, Jul. 2016, doi: 10.22146/ijccs.16625.

[36] F. Alhaj, A. Al-Haj, A. Sharieh, and R. Jabri, "Improving Arabic Cognitive Distortion Classification in Twitter using BERTopic," Int. J. Adv. Comput. Sci. Appl., vol. 13, no. 1, pp. 854–860, 2022, doi: 10.14569/IJACSA.2022.0130199.