Science in Information Technology Letters
Vol. 3., No. 2, November 2022, pp. 105-115
ISSN 2722-4139
http://pubs2.ascee.org/index.php/sitech

Sentiment analysis of wayang climen using naive bayes method

Fitriana Kurniawati [a,1], Aji Prasetya Wibawa [a,2,*], Agung Bella Putra Utama [a,3], Urvi Ghosh [b,4]

[a] Department of Electrical Engineering and Informatics, Faculty of Engineering, Universitas Negeri Malang
[b] Monash University, Australia
[1] fitriana.kurniawati.1705356@students.um.ac.id; [2] aji.prasetya.ft@um.ac.id; [3] agungbpu02@gmail.com; [4] ugho0001@student.monash.edu
* Corresponding Author

ARTICLE INFO

Article history
Received 10 October, 2022
Revised 26 October, 2022
Accepted 12 November, 2022

Keywords
Sentiment analysis
Wayang climen
Social media
Naïve bayes

ABSTRACT

This research focuses on sentiment analysis of Wayang Climen performances in Indonesia using the Naïve Bayes algorithm. Wayang, a traditional puppet show, holds cultural significance and has persisted alongside modern entertainment options. The study collected public comments from Dalang Seno and Ki Seno Nugroho's YouTube channels, classified them into positive, negative, and neutral sentiments, and employed a translation process to align comments with program language objectives. Preprocessing steps included case folding, removing punctuation, tokenizing, stopword removal, and post-tagging. To address data class imbalances, resampling was performed using the Synthetic Minority Oversampling Technique (SMOTE). The Naïve Bayes algorithm was utilized for data classification, exploring various translation scenarios. Evaluation involved the confusion matrix method and metrics like accuracy, precision, recall, and f-measure. Results demonstrated that the Dalang Seno train data scenario outperformed Ki Seno Nugroho's, with higher precision, recall, accuracy, and f-measure values. Additionally, the translation scenario from Indonesian to English yielded the most effective results. In conclusion, this study highlights the suitability of the Naïve Bayes algorithm for sentiment analysis in the context of Wayang Climen performances, with practical implications for understanding public sentiment in the digital age.

1. Introduction

Indonesia is a country rich in culture. One of the cultures that Indonesia has is a puppet show or *wayang*. *Wayang* is a cultural relic from ancient ancestors estimated to have existed since +-1500 BC [1]. *Wayang* is a performance that tells an event by depicting human characters [2]. These performances basically have their own function, including as a medium for conveying moral messages to society [2], as respect and a request for blessings from the spirits of ancestors [1], as a medium of education for the public [2], as a medium for da'wah during the time of Sunan Kalijaga [3], as the communication medium, as the character building of people according to the *wayang*'s moral value [2]. The development of *wayang* still happens in this era even though there are some entertainment such as television, social media, etc. It means that *wayang* has been rooted in society for a long time [4].

Social media is online content that is accessible to people [5]. The society uses it to share the knowledge and information. People actively participate in social media to give their opinions and comments [6]. During the pandemic, the government did the mobility distancing by implementing a large-scale social distancing (PSBB). People are advised to stay at home, not in crowds [7]. Due to government regulations, Dalang Seno initiated an innovative wayang climen performance that was streamed on YouTube to cure people's longing for wayang performances. He presented *wayang climen* that deviates from the standard *wayang* presentations. It was designed with a straightforward and minimalistic idea. YouTube can provide public access to provide opinions on Dalang Seno's *wayang climen* performances. To find out how the public comments on *wayang climen*, a sentiment analysis was carried out on the public's comments. It is the process of managing textual data into important sentiment information contained in a sentence [8]. In this study, researchers propose to conduct sentiment analysis using the Naïve Bayes method to determine the public's response to the *wayang climen* performance. The Naïve Bayes method is a simple method to apply [9]. Previous research conducted by Ghulam, comparing the results of the Naïve Bayes, Lexicon Based, and Support Vector Machine algorithms, concluded that the highest accuracy value was the Naïve Bayes algorithm with an accuracy value of 95%. In his research, Ghulam compared algorithms on the 2017 DKI Jakarta Governor Candidate dataset on Twitter social media, totaling 300 data. Ghulam suggested that in future research, it is necessary to add real-time and more data [10]. That reference is used as rationality by researchers to apply the Naïve Bayes method to *wayang climen* research.

2. Method

There are several methods in research on sentiment analysis of wayang climen using the Naïve Bayes algorithm. The following is the method that the author used.

2.1. Dataset Collection

The dataset used in this research is public comments on the YouTube channels of Dalang Seno and Ki Seno Nugroho for three months (July 2020 to September 2020). From public comments on Dalang Seno's YouTube channel, 2126 lines of data were obtained. Meanwhile, on Ki Seno Nugroho's YouTube channel, 794 lines of data were obtained. Comments from the two YouTube channels were not combined into one dataset for the reason that a classification scenario was carried out related to the amount of data train and data test.

2.2. Labeling

Researchers carried out data labeling manually, which the language expert later validated, Prof. Dr. Suyono, M.Pd, from the Department of Indonesian Literature, Universitas Negeri Malang. In the labeling process, researchers grouped the data into three sentiment classes: positive, negative, and neutral. Validation by an expert needs to be done to increase the accuracy of the labeling process since an expert has more knowledge and understanding of any text written in the data [9]. Furthermore, it groups sentiments based on adjectives and verbs in comment sentences. If an adjective or verb with a good meaning has the highest percentage of occurrences, it means positive.

Meanwhile, if an adjective or verb means bad, it has the highest percentage of occurrences, which means negative. Further, if an adjective with a good meaning has the same percentage of occurrence as the percentage of occurrence of a word with a bad meaning, it means neutral. Interrogative sentences and sentences that do not contain adjectives or verbs are also included in the neutral sentiment category. An example of the data labeling process can be seen in Table 1.

Table 1. Labeling Processing

No	Comment	Adjective and Verb			Sentiment
		Positive	*Negative*	*Neutral*	
1	*mbelisa cantik sekali suarane bagus*	*cantik, bagus* (pretty, good)	-	-	positive
2	*terlambat, gak masalah nggih ki seno... tetap terhibur...*	*terhibur* (entertained)	*terlambat* (late)	*tetap* (still)	neutral
3	*mas soundnya pecah mas, ground pecah*	-	pecah (broken)	-	negative
4	*latamahusadi itu artinya apa ya lur?*	-	-	-	neutral

From the data labeling process, a data distribution graph is obtained containing positive, negative and neutral sentiment classes. The data distribution in the Dalang Seno and Ki Seno Nugroho datasets can be seen in Fig. 1.

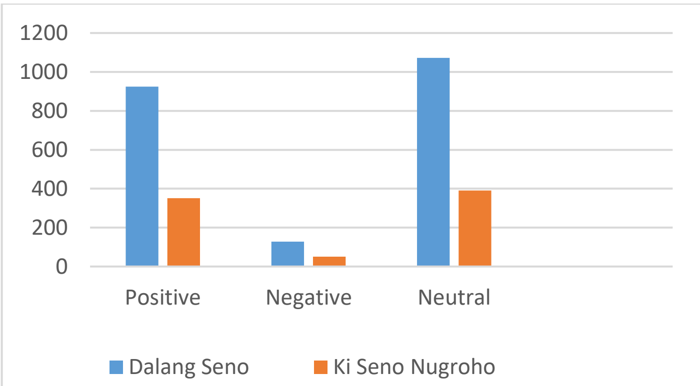

Fig. 1. Data Distribution Graphic

2.3. Text Translation

The translation process is carried out to align the language objectives applied to the program, considering that the comment data obtained contains more than one language. Translation is done automatically using the library of Python google_translate_new. The translation process is shown in the pseudocode in Fig. 2. In this research, 5 translation scenarios were carried out, namely Javanese to Indonesian, Javanese to English, Indonesian to English, Javanese to Indonesian to English, and Indonesian to Javanese to English.

```
Input: Comment Text
Output: Translated text
Process:
Begin
    1) Enter the comment text in the program
    2) Translate the comment text to the language specified by the program
    3) Output the translated text
End
```

Fig. 2. Pseudocode Process of Translation

2.4. Pre-processing Data

The input data is in the form of text containing sentences, words, language, and various symbols. To prepare the data to be ready for processing, the input data needs to be pre-processed first. Pre-processing is a crucial step because it removes irrelevant data and/or transforms data into a format that the system can understand [11]. The pre-processing data step is displayed on Fig. 3.

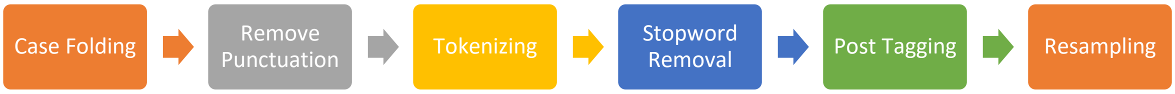

Fig. 3. Step of Pre-processing Data

Case Folding is a process of pre-processing data that is used to change the capital letters to the small ones [12]. It needs to avoid data redundancy; the "A" capital letter has a different meaning from the "a" small letter. The pseudocode of the case folding process is shown in Fig. 4.

Input: Translated text
Output: Text with lowercase letters
Process:
Begin
1) Input the translated text to the program
2) Convert the translated text containing capital letters to lowercase letters
3) Output the translated text with lowercase letters

End

Fig. 4. Pseudocode of the Case Folding Process

The next step is Remove Punctuation. It is a data pre-processing process where punctuation marks and symbols in the text are removed [13]. The pseudocode of the Remove Punctuation process can be seen in Fig. 5.

Input: Text with lowercase letters
Output: Text without punctuation and symbols
Process:
Begin
1) Enter text with lowercase letters in the program
2) Remove punctuation marks and symbols from the input text
3) Produces output text without punctuation and symbols

End

Fig. 5. Pseudocode of the Remove Punctuation Process

The third step in data pre-processing is Tokenizing. It is the process of breaking up sentences into single words quoted with single quotation marks [14]. The pseudocode of the Tokenizing process is shown in Fig. 6.

Input: Text without punctuation and symbols
Output: Single word set
Process:
Begin
1) Enter text without punctuation and symbols
2) Calls text that is a sentence into a collection of single words
3) Eliminate the output of single word computation

End

Fig. 6. Pseudocode of the Tokenizing Process

After the tokenizing process, the data is then subjected to a stopword removal process. Stopword removal is the process of removing words that contain stopwords [15]. Stopwords contain a collection of words that do not affect sentiment, such as connecting words, auxiliary words, and others. To simplify the tagging process, a stopword removal process is implemented. The pseudocode of the stopword removal process is shown in Fig. 7.

```
Input: A collection of single words
Output: Single word set without pronouns and conjunctions
Process:
Begin
    1) Enter the set of single words
    2) Remove conjunctions and pronouns
    3) Output a set of single words without pronouns and conjunctions
End
```

Fig. 7. Pseudocode of the Stopword Removal Process

The fifth step in the data pre-processing stage is Post-tagging. It is the process of labeling types of words [16]. The Pseudocode of the Post-tagging process is given in Fig. 8.

```
Input: Set of single words without pronouns and conjunctions
Output: Set of single words with tags
Process:
Begin
    1) Input a set of single words without pronouns and conjunctions to the program
    2) Add tags to each single word according to its word type
    3) Produce output in the form of a collection of single words with tags
End
```

Fig. 8. Pseudocode of the Post-Tagging Process

The final step is Resampling. It is the process of balancing the uneven distribution of classes, which consists of 2 methods, namely undersampling and oversampling [17]. The distribution of data in the *wayang climen* dataset can be said to be unbalanced because the value ranges for the three classes are quite far. Based on the spread graph, the neutral and positive sentiment classes have a very high level of spread, in contrast to the negative sentiment class, which is very low. To overcome this problem, a resampling process was carried out to balance the data distribution.

The resampling process was done by using SMOTE. Systemic Minority Oversampling Technique (SMOTE) is one of the oversampling techniques to balance a data class by producing a new synthesis from the oversampling method. The new synthesis is generated from the distance between the minority sample and the nearest neighbor of the minority sample [18]. Table 2 illustrates the variation in the percentage of data distribution before and after the resampling procedure.

Table 2. Percentage Table Before and After the Oversampling Process

Dataset	Class	Before Oversampling			After Oversampling		
		Total	*Percentage (%)*	*Amount*	*Total*	*Percentage (%)*	*Amount*
Dalang Seno	Positive	925	43.50	2126	1073	33.33	3219
	Negative	128	6.02		1073	33.33	
	Neutral	1073	50.47		1073	33.33	
Ki Seno Nugroho	Positive	352	44.33	794	391	33.33	1173
	Negative	51	6.42		391	33.33	
	Neutral	391	49.24		391	33.33	

2.5. Data Classification

This research uses the Naïve Bayes algorithm at the classification stage. Naïve Bayes is a supervised learning algorithm that requires training data for the classification process to form a model [19]. The principle of Naïve Bayes is that attributes in a data set do not influence each other. This principle is usually called class conditional confidence. The Naïve Bayes algorithm classification process analyzes document samples and determines category values [20]. The advantages of this algorithm include that it can handle discrete and continuous data, it is easy to implement and is able to provide good results from various types of cases, it can carry out document classification automatically effectively [21], and the implementation is quite simple as well as short in the processing time [9]. As for the weaknesses, it cannot model the relationship between variables because Naïve Bayes has the principle that the variables in the data do not affect each other [21] and requires many records to produce good accuracy values [22]. Bayes' theorem is a theory based on conditional probability. In general, Bayes theory can be seen in equation 1 [11].

$$P\ (A\ |B) = \frac{P\ (A|B)P(A)}{P(B)} \tag{1}$$

Where A is a label of data and B is a feature used in it. At the data classification stage, it was designed by applying ten combinations of scenarios to see the accuracy and performance of the classification algorithm for each scenario. This combination can be seen in Table 3.

Table 3. Combination of Scenario Classification Tabel

No	Scenario	
1	Train = Dataset Dalang Seno Test = Dataset Ki Seno Nugroho	Javanese to English
	Train = Dataset Dalang Seno Test = Dataset Ki Seno Nugroho	Indonesian to English
3	Train = Dataset Dalang Seno Test = Dataset Ki Seno Nugroho	Javanese to Indonesian
4	Train = Dataset Dalang Seno Test = Dataset Ki Seno Nugroho	Javanese to Indonesian to English
5	Train = Dataset Dalang Seno Test = Dataset Ki Seno Nugroho	Indonesian to Javanese to English
6	Train = Dataset Ki Seno Nugroho Test = Dataset Dalang Seno	Javanese to English
7	Train = Dataset Dalang Seno Test = Dataset Ki Seno Nugroho	Indonesian to English
8	Train = Dataset Dalang Seno Test = Dataset Ki Seno Nugroho	Javanese to Indonesian
9	Train = Dataset Dalang Seno Test = Dataset Ki Seno Nugroho	Javanese to Indonesian to English
10	Train = Dataset Dalang Seno Test = Dataset Ki Seno Nugroho	Indonesian to Javanese to English

In Table 2, scenarios 1 to 5 use Dalang Seno's data train with five kinds of translation processes, while scenarios 6 to 10 use Ki Seno Nugroho's data train with the same five kinds of translation scenarios. From the same translation scenarios, the average value is taken to be evaluated to find the best translation scenario and the best data train application scenario.

2.6. Evaluation of Classification Results

The final stage that is carried out after obtaining the classification results is evaluating the performance results of the algorithm provided for each scenario. In this research, evaluation was carried out using the confusion matrix method, and then the recall, precision, accuracy, and f-measure values were obtained. Recall is the percentage of true cases that are identified as true [23]. Precision is a comparison of true cases identified as true with all predicted positive results [23]. Accuracy is the proportion of correct prediction [24]. F-measure is a combination value of precision and recall that becomes the overall algorithm performance value [25]. The shape of the confusion matrix table can be seen in Table 4.

Table 4. Confusion Matrix Table

		Prediction		
		Positive	*Negative*	*Neutral*
Actual	*Positive*	True Positive (A)	False Positive (B)	False Positive (C)
	Negative	False Negative (D)	True Negative (E)	False Negative (F)
	Neutral	False Neutral (G)	False Neutral (H)	True Netral (I)

From Table 4, the accuracy, precision, recall, and f-measure values can be found as follows :

$$Accuracy = \frac{\text{A+E+I}}{\text{A+B+C+D+E+F+G+H+I}}\ x\ 100 \quad (2)$$

$$Positive\ Recall = \frac{\text{A}}{\text{A+B+C}}\ x\ 100\% \quad (3)$$

$$Negative\ Recall = \frac{\text{E}}{\text{D+E+F}}\ x\ 100\% \quad (4)$$

$$Neutral\ Recall = \frac{\text{I}}{\text{G+H+I}}\ x\ 100\% \quad (5)$$

$$Recall = \frac{\text{Positive Recall +Negative Recall +Neutral Recall}}{3} \quad (6)$$

$$Positive\ Precision = \frac{\text{A}}{\text{A+D+G}}\ x\ 100\% \quad (7)$$

$$Negative\ Precision = \frac{\text{E}}{\text{B+E+H}}\ x\ 100\% \quad (8)$$

$$Neutral\ Precision = \frac{\text{I}}{\text{C+F+I}}\ x\ 100\% \quad (9)$$

$$Precision \frac{\text{Positive Precision +Negative Precision +Neutral Precision}}{3} \quad (10)$$

$$F - measure = 2\ \text{x}\ \frac{\text{Precision x Recall}}{\text{Precision+Recall}}\ x\ 100\% \quad (11)$$

3. Results and Discussion

A confusion matrix table is created during the classification process using a variety of scenarios, and this table is used to assess the algorithm's performance. Recall, precision, accuracy, and f-measure values have been derived using the confusion matrix table. Table 5 displays the scenario results with Dalang Seno train data, and Table 6 displays the scenario results utilizing Ki Seno Nugroho train data.

Table 5. Scenario Results Using Dalang Seno Train Data and Ki Seno Nugroho Test Data

Translation Process	Precision	Recall	F-measure	Accuracy
Javanese to English	53,93%	57,15%	53,97%	63,73%
Javanese to Indonesian	52,12%	50,39%	50,84%	63,22%
Indonesian to English	55,81%	59,39%	56,41%	65,99%
Javanese to Indonesian to English	52,21%	53,83%	52,35%	63,10%
Indonesian to Javanese to English	52,48%	55,01%	52,02%	61,71%
Average	53,31%	55,15%	53,12%	63,5%

Table 6. Scenario Results Using Ki Seno Nugroho Train Data and Dalang Seno Test Data

Translation Process *Proses Translate*	Precision *Presisi*	Recall *Recall*	F-measure *F-measure*	Accuracy *Akurasi*
Javanese to English	49,52%	50,22%	48,35%	59,69%
Javanese to Indonesian	47,70%	47,58%	47,37%	59,64%
Indonesian to English	51,29%	51,45%	50,24%	62,70%
Javanese to Indonesian to English	49,16%	49,80%	48,47%	60,21%
Indonesian to Javanese to English	49,04%	48,89%	47,87%	60,40%
Average	49,34%	49,59%	48,46%	60,53%

From Table 5 and Table 6, it can be seen that the highest confusion matrix value is in the Dalang Seno train data with a precision value of 53.31%, a recall value of 55.15%, an accuracy value of 63.5%, and an f-measure value of 53 .12%. Meanwhile, the Ki Seno Nugroho train data produces a precision value of 49.34%, a recall value of 49.59%, an accuracy value of 60.53%, and an f-measure value of 48.46%.

Apart from comparing the confusion matrix results of Dalang Seno and Ki Seno Nugroho train data, a comparison of the confusion matrix results of the translated scenario was also carried out in Table 7.

Table 7. Average Score of Translation Scenario

Translation Process	Precision	Recall	F-measure	Accuracy
Javanese to English	51,59%	54,14%	49,68%	57,87%
Javanese to Indonesian	49,78%	48,79%	48,94%	61,17%
Indonesian to English	53,62%	55,48%	53,61%	64,77%
Javanese to Indonesian to English	50,88%	52,43%	50,65%	61,20%
Indonesian to Javanese to English	49,94%	51,30%	49,49%	60,54%

Table 7 shows that the translation scenario that produces the highest confusion matrix value is Indonesian to English, with a recall value of 55.48%, precision value of 53.62%, accuracy value of 64.77%, and f-value measure of 53.61%. The confusion matrix results of the translated scenarios are related to the feature results produced by each scenario at the Post-tagging stage, shown in Fig. 9.

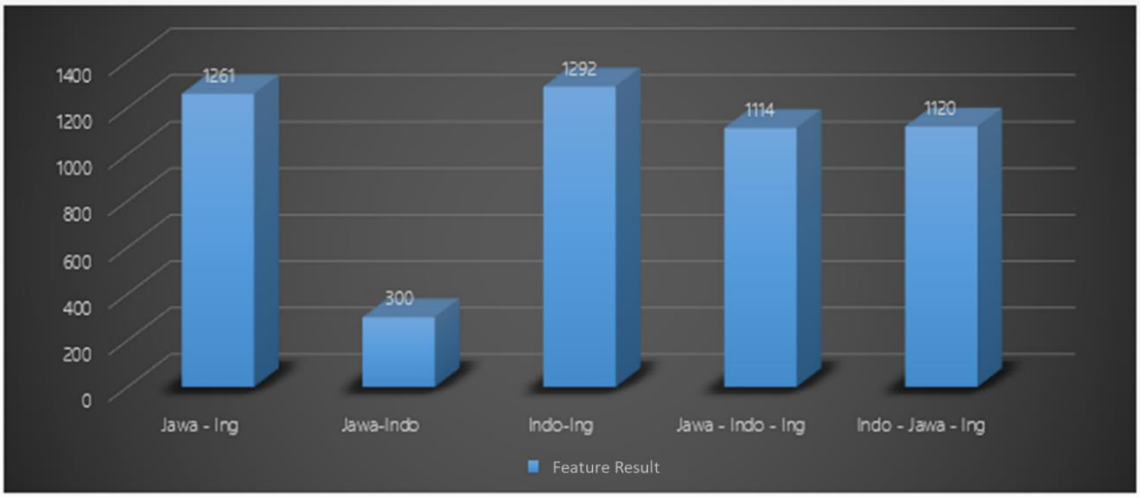

Fig. 9. Translation Scenario Feature Results

Fig. 9 illustrates that the scenario involving the translation from Indonesian to English yields the greatest number of features (1292). The scenarios involving the translation from Javanese to English follow with 1261 features, the translation from Indonesian to Javanese to English with 1120 features, the translation from Javanese to Indonesian to English with 1114 features, and the final scenario involving the translation from Javanese to Indonesian with 300 features. The number of adjectives that are successfully identified at the post-tagging step is known as features.

4. Conclusion

Based on the research results, it can be concluded that the Naïve Bayes algorithm is able to carry out a sentiment analysis on data from wayang climen performances. The research results show that the Dalang Seno train data scenario produces the best confusion matrix values compared to the scenario with Ki Seno Nugroho train data. The scenario with Dalang Seno train data produces an average precision value of 53.31%, recall value of 55.15%, accuracy value of 63.5%, and f-measure value of 53.12%.

In addition, in the comparison of the results of the confusion matrix of translation scenarios, the best scenario is the Indonesian to English translation scenario with a recall value of 55.48%, a precision value of 53.62%, an accuracy value of 64.77%, and an f-measure value of 53.61%. The confusion matrix's findings are connected to the features produced during the post-tagging phase, where the translation scenario from Indonesian to English yields the greatest amount of items—1292 features.

References

[1] A. Arafahan, Saidin, and F. Y. Sitepu, "Copyright Protection Of Wayang Kulit As A Traditional Indonesian Culture," *POLICY, LAW, NOTARY Regul. ISSUES*, vol. 1, no. 4, pp. 1–8, Oct. 2022, doi: 10.55047/polri.v1i4.413.

[2] D. R. Indah, "The symbolic meaning of 'Punakawan Javanese Wayang '(a value imaging study in character education at the character education course in STKIP Bina Insan Mandiri Surabaya)," *SELL Journal, e-ISSN*, pp. 2580–8400, 2019.[Online]. Available at: http://publikasi.stkippgri-bkl.ac.id/index.php/SL/article/view/93.

[3] M. Muslih, A. Rohman, A. Ahmad, and A. Saifullah, "Sunan Kalijaga's Da'wah Strategy In Suluk Linglung And Its Implication To Indonesian Radicalism Movement," *El-HARAKAH (TERAKREDITASI)*, vol. 23, no. 1, pp. 1–19, Jun. 2021, doi: 10.18860/eh.v23i1.11672.

[4] A. W. Pudjastawa, Y. Perdananto, and H. Bastian, "Wayang Pegon: The Way Malang People Embrace Differences," in *1st International Conference Of Education, Social And Humanities (INCESH 2021)*, 2021, pp. 62–68, doi: 10.33005/irofonic.v1i1.12.

[5] R. Rogers, "Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media," *Eur. J. Commun.*, vol. 35, no. 3, pp. 213–229, 2020, doi: 10.1177/0267323120922066.

[6] Y. A. Ahmed, M. N. Ahmad, N. Ahmad, and N. H. Zakaria, "Social media for knowledge-sharing: A systematic literature review," *Telemat. Informatics*, vol. 37, pp. 72–112, Apr. 2019, doi: 10.1016/j.tele.2018.01.015.

[7] N. Ivanka, "Large-scale Social Restrictions: What's Next?," *Indones. J. Int. Clin. Leg. Educ.*, vol. 2, no. 2, pp. 201–214, 2020, doi: 10.15294/ijicle.v2i2.38324.

[8] G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu, "Sentiment analysis of comment texts based on BiLSTM," *Ieee Access*, vol. 7, pp. 51522–51532, 2019, doi: 10.1109/ACCESS.2019.2909919.

[9] M. Abbas, S. Memon, K. A. Memon, A. A. Jamali, and A. Ahmed, "Multinomial Naive Bayes Classification Model for Sentiment Analysis," *Int. J. Comput. Sci. Netw. Secur.*, vol. 19, no. 3, pp. 62–67, 2019, doi: 10.13140/RG.2.2.30021.40169.

[10] H. F. Nweke, Y. W. Teh, G. Mujtaba, and M. A. Al-Garadi, "Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions," *Inf. Fusion*, vol. 46, pp. 147–170, 2019, doi: 10.1016/j.inffus.2018.06.002.

[11] P. Oliveri, C. Malegori, R. Simonetti, and M. Casale, "The impact of signal pre-processing on the final interpretation of analytical outcomes–A tutorial," *Anal. Chim. Acta*, vol. 1058, pp. 9–17, 2019, doi: 10.1016/j.aca.2018.10.055.

[12] U. Naseem, I. Razzak, and P. W. Eklund, "A survey of pre-processing techniques to improve short-text quality: a case study on hate speech detection on twitter," *Multimed. Tools Appl.*, vol. 80, no. 28–29, pp. 35239–35266, Nov. 2021, doi: 10.1007/s11042-020-10082-6.

[13] A. Kathuria, A. Gupta, and R. K. Singla, "A review of tools and techniques for preprocessing of textual data," *Comput. Methods Data Eng. Proc. ICMDE 2020, Vol. 1*, pp. 407–422, 2021, doi: 10.1007/978-981-15-6876-3_31.

[14] S. Y. Senanayake, K. Kariyawasam, and P. S. Haddela, "Enhanced tokenizer for sinhala language," in *2019 National Information Technology Conference (NITC)*, 2019, pp. 84–89, doi: 10.1109/NITC48475.2019.9114420.

[15] D. J. Ladani and N. P. Desai, "Stopword identification and removal techniques on tc and ir applications: A survey," in *2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2020, pp. 466–472, doi: 10.1109/ICACCS48705.2020.9074166.

[16] S. Mohammed, "Using machine learning to build POS tagger for under-resourced language: the case of Somali," *Int. J. Inf. Technol.*, vol. 12, no. 3, pp. 717–729, 2020, doi: 10.1007/s41870-020-00480-2.

[17] M. Khushi *et al.*, "A comparative performance analysis of data resampling methods on imbalance medical data," *IEEE Access*, vol. 9, pp. 109960–109975, 2021, doi: 10.1109/ACCESS.2021.3102399.

[18] G. A. Pradipta, R. Wardoyo, A. Musdholifah, and I. N. H. Sanjaya, "Radius-SMOTE: a new oversampling technique of minority samples based on radius distance for learning from imbalanced data," *IEEE Access*, vol. 9, pp. 74763–74777, 2021, doi: 10.1109/ACCESS.2021.3080316.

[19] P. C. Sen, M. Hajra, and M. Ghosh, "Supervised Classification Algorithms in Machine Learning: A Survey and Review," in *Advances in Intelligent Systems and Computing*, 2020, pp. 99–111, doi: 10.1007/978-981-13-7403-6_11.

[20] M. Wongkar and A. Angdresey, “Sentiment analysis using Naive Bayes Algorithm of the data crawler: Twitter,” in *2019 Fourth International Conference on Informatics and Computing (ICIC)*, 2019, pp. 1–5, doi: 10.1109/ICIC47613.2019.8985884.

[21] S. S. Bafjaish, “Comparative analysis of Naive Bayesian techniques in health-related for classification task,” *J. Soft Comput. Data Min.*, vol. 1, no. 2, pp. 1–10, 2020.[Online]. Available at: https://publisher.uthm.edu.my/ojs/index.php/jscdm/article/view/7144.

[22] E. Manino, L. Tran-Thanh, and N. R. Jennings, “On the efficiency of data collection for multiple Naïve Bayes classifiers,” *Artif. Intell.*, vol. 275, pp. 356–378, 2019, doi: 10.1016/j.artint.2019.06.010.

[23] H. R. Sofaer, J. A. Hoeting, and C. S. Jarnevich, “The area under the precision-recall curve as a performance metric for rare binary events,” *Methods Ecol. Evol.*, vol. 10, no. 4, pp. 565–577, 2019, doi: 10.1111/2041-210X.13140.

[24] M. VerMilyea *et al.*, “Development of an artificial intelligence-based assessment model for prediction of embryo viability using static images captured by optical light microscopy during IVF,” *Hum. Reprod.*, vol. 35, no. 4, pp. 770–784, 2020, doi: 10.1093/humrep/deaa013.

[25] R. Soleymani, E. Granger, and G. Fumera, “F-measure curves: A tool to visualize classifier performance under imbalance,” *Pattern Recognit.*, vol. 100, p. 107146, 2020, doi: 10.1016/j.patcog.2019.107146.