Institut Riset dan Publikasi Indonesia (IRPI)
IJATIS: Indonesian Journal of Applied Technology and Innovation Science
Journal Homepage: https://journal.irpi.or.id/index.php/ijatis
Vol. 1 Iss. 2 August 2024, pp: 85-92
ISSN(P): 3032-7466 | ISSN(E): 3032-7474

Sentiment Analysis of Public Opinion on Films Taylor Swift Eras Tour on the Twitter Platform Using the Machine Learning

Nova Idriani R[1], Annisa Dahlia[2], Pratiwi Ningum[3]

[1,2]Department of Information Systems, Faculty of Science and Technology,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia
[3]Mesa Community College, America

E-Mail: [1]12050323216@student.uin-suska.ac.id,
[2]12050323694@student.uin-suska.ac.id, [3]pra2133191@maricopa.edu

Received Dec 27th 2023; Revised Jun 16th 2024; Accepted Jul 21th 2024
Corresponding Author: Nova Idriani R.

Abstract

Sentiment analysis is the understanding of opinions, feelings, or attitudes conveyed in texts, such as tweets, reviews, or other forms. The film “Taylor Swift: The Eras Tour” is trending among teenager, narrating Taylor Swift’s journey in the “Eras Tour” concert across various countries, encapsulated in a music-filled film. This has prompted research on sentiment analysis of netizens’ tweets about this film, considering the possibility of negative reviews. Three algorithms Naïve Bayes, Decision Tree, and Random Forest were used with an 80:20 data ratio and the SMOTE oversampling method, which is a unique in this research to ensure data sizes for the three sentiments: positive, negative, and neutral. The final result of this research is a word cloud for each sentiment towards the film, with the Decision Tree algorithm achieving the highest accuracy at 91%. The hope for future research is to implement and focus on the emotional aspect in conducting sentiment analysis.

Keyword: Decision Tree, Naïve Bayes, Random Forest, Sentiment Analysis, Taylor Swift

1. INTRODUCTION

The development of information and communication technology is inseparable from increasing sophistication of existing technologies, particulary in the film industry. With hundreds to thousands of film produced each year, competition in the film industry is becoming more intense.

The film industry is experiencing rapid global growth. According to Thompson Bordwell (2013), the success of a film depends on its technical aspects and its ability to evoke emotions from the audience. Currently, the film “Taylor Swift Eras Tour”, released on October 13, 2023, has become popular among fans who missed out of tickets to Taylor Swift’s concerts in various countries. This raises questions about netizens’ reponses to the film, whether they are positive or negative. Currently, watching movies has become a common activity to spend leisure time, and the increasing number of films produced both domestically and internationally provides numerous options for viewers. Viewers subsequently seek information and read opinions about these films, which have been increasing annually by 80%, especially from unstructured data in the form of text generated from social media platforms like Twitter or X [1][2].

Twitter or X is a social media platform that allows users to send and read text messages up to 140 characters, known as “tweets” [3]. This platform serves as a venue for thousands or ever millions of opinions in the form of tweets, including public sentiment towards films. The sentiments expressed by viewers towards a film greatly influence its commercial success and critical reception. Sentiment analysis involves natural language programming, text information extraction, and artificial intelligence to qualitatively identify and categorize text data [4]. It is an automated method to express opinions or feeling from text, used to assess whether opinions towards a subject tend to be positive or negative, utilizing emotion-based approaches in data processing [5].

Emotion-based approaches are categorized into emotions such a joy, sadness, anxiety, and admiration, revealing how a viewer’s experience of a film influences them psychologically and emotionally. Emotion-focused approaches are utilized to understand individual assessments of the film “Taylor Swift: The Eras Tour” using Naïve Bayes, Decision Tree and Support Vector Machine. In a series of previous studies: [2] sentiment analysis on Korean Drama using Naïve Bayes achieved an accuracy of 69%, precision of 73%, recall of 69%, and F1 Score of 69%. [6]using Random Forest, to handle diverse data, the feature selection method used is Mutual Information. With these two methods, the system achieved an accuracy of 79% and F1-Score of 75%. [7] hypertension prediction using Random Forest and Decision Tree reached 100% accuracy. [8] sentiment analysis of anti-LGBT campaigns with Naïve Bayes achieved an accuracy of 86.43%, higher than the Decision Tree and Random Forest (82.91%). [9] comparison of Decision Tree and Support Vector Machine on Instagram showed that TF-IDF Support Vector Machine achieved an accuracy of 94.36%, recall of 94.30%, and F1-score of 95.53%. [10] Random Forest had the highest accuracy at 97.16% (AUC 0.996), followed by Support Vector Machine at 96.01% (AUC 0.543), and Naïve Bayes at 94.16% (AUC 0.999).

The researcher selected three algorithms with the highest accuracy from previous studies: Naïve Bayes, Decision Tree, and Random Forest. This research aims to compare the accuracy of these three algorithms. The data used pertains to the Taylor Swift The Eras Tour film issue sourced from Twitter. Therefore, this research can provide a clear comparison of accuracy among these algorithms and determine which yealds the best results.

2. MATERIAL AND METHOD

In this research methodology, the initial step is data collection, followed by the data preprocessing stage, which include tokenization and the use of SMOTE. Subsequently, the research involves the implementation of three primary models as the main focus. The evaluation phase is then conducted to assess the performance of the models. The research methodology flow is illustrated in Figure 1.

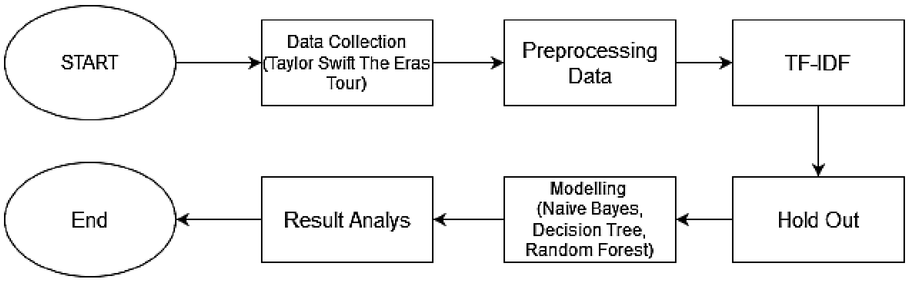

Figure 1. Research Methodology

Based on Figure 1, this research begins by collecting data from Twitter or X from August 1, 2023, until November 11, 2023, using keyword ‘taylorswifterastour’. The data was collected in English because gathering data in Indonesian yieled only 129 entries. Data preprocessing included cleaning the text by removing emojis, symbols, characters, and abbreviations in each tweet, followed by further data cleaning. SMOTE was used to address data imbalance. The algorithms used in this research were Naïve Bayes, Decision Tree, and Random Forest. The data was split using the Holdout technique with 80% for training and 20% for testing. A compound score of 0.05 or higher indicated positive sentiment, -0.05 or lower indicated negative sentiment, and scores in between indicated neutral sentiment.

2.1 Sentiment Analysis

Sentiment analysis is the process of identifying and quantifying the sentiment of text or audio using natural language processing, text analysis, computational linguistics and other techniques, with some nuances related to the film, such as “grieving” referring to a gloomy or dissapointing film, or “crying” having a positive sentiment in a tragedy but a negative one in comedy [11] [12]. In analyzing specific topics or opinions, sentiment analysis related to data mining plays an important role in delivering results. Research in sentiment analysis includes creating summaries based on emotions, extracting feelings, or thoughts with NLP to monitor People's emotions and views on specific topics, products, or services [13]

2.2. Naïve Bayes

The Naïve Bayes algorithm assumes that the presence of a particular feature in a class is independent of other features [14]. Similarly, according to [15] Naïve Bayes assumes that the presence of a specific feature in a class is unrelated to the presence of other feature classes, even though these features may depend on each other, which affects the probability. The Naïve Bayes algorithm is suitable for small datasets because it typical yields good result. However, its effectiveness with small datasets depends on the specific conditions and characteristics of the dataset. The general form of the Naïve Bayes theorem is as equation 1.

$$P(C|X) = (P(X|C).P(C))/(P(X)) \quad (1)$$

Description:

$P(C|X)$: Probability of a class C given the attribute X
$P(X|C)$: Probability of attribute X appearing in class C
$P(C)$: Prior probability of class C
$P(X)$: Prior probability of attribute X

2.3. SMOTE

SMOTE, which stands for Synthetic Minority Over-Sampling Technique, is one of the commonly used oversampling methods to address imbalanced datasets by creating synthetic data in the minority class along the line connecting some or all of the randomly selected neighbors from the samples [16] [17]. The (simple) steps of SMOTE are as follows:

1. Choose the number k of nearest neighbors to be used and how many times $l = N/100$ oversampling will be conducted.
2. For each data point x in the class, the key points are:
   a. Randomly select l data points from the k nearest neighbors.
   b. For each data point, create synthetic data by taking a randomly observed data point parallel to the existing point x for each observed l.
3. The result of SMOTE is the initial data plus the oversampling result performed.

Before SMOTE was applied, the difference in data can be observed as table 1.

Table 1. Amount of Data Before SMOTE

Class	Amount of data
0	81
1	175
2	445

Based on Table 1, class 0 represents negative sentiment with 81 data points, class 1 represents positive sentiment with 175 data points, and class 2 represents neutral sentiment with the highest number of data points, totaling 445. Due to the imbalance in the number of data points among these classes, SMOT was applied. The result of the SMOTE processfor the three classes can be seen in Table 2.

Table 2. Amount of Data after SMOTE

Class	Amount of Data
0	445
1	445
2	445

After applying SMOTE, all three classes now have an equal number of data points, which is 445. Thus, the obtained data is now balanced and no longer suffers from imbalance.

2.4. Decision Tree

The Decision Tree is an algorithm method that utilizes a tree structure to determine the outcome of an event. Essentially, a test node predicts an outcome by evaluating the attribute values of a particular instance, where each possible outcome is associated with one branch of a subtree [18]. Decision Trees rely on different criteria for making predictions, such as the Gini Index and Information Gain, with the Gini Index being the most commonly used [19]. The basic formula for a Decision Tree is equation 2.

$$\text{Gini} = 1 - \sum_{i=1}(P_i)^2 \tag{2}$$

In the situation where p_i, for the probability of an element being classified into a particular class, the equation of a Decision Tree is as equation 3 and 4.

$$\text{Entropy}(S) = \sum_{i=1}^{n} - pi * \log_2 pi \tag{3}$$

and

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \tag{4}$$

Where:

p_i : proportion of S_i to S
A : attribute
I : a possible value for attribute A
$|S_i|$: number of samples for value i
$|S|$: total number of data samples

2.5. Random Forest

Random Forests is a learning technique used for classification and regression. When handling a new case, this method sends the case to each tree to perform classification and determine its class [20]. In essence, Random Forest is a combination of many decision trees. The process involves Bagging (Bootstrap Aggregating), Decision Trees, and prediction. Random Forest constructs various training datasets to enchance the diversity among the combined models using a sequence of classification models $\{h_1(X), h_2\ (X) \ldots\ h_k(X)$ [21].

2.6. TF-IDF

TF-IDF is a step in the feature extraction process where each word in a tweet is assigned a weight value based on its important in the tweet that has been processed previously [22]. The formula for TF-IDF is as equation 5.

$$\mathrm{TF}-\mathrm{IDF}_{(t,d,D)} = \mathrm{TF}_{(d,t)} \text{ x } \mathrm{IDF}_{(t,D)} \tag{5}$$

where

$$\mathrm{TF}_{(d,t)} = \frac{\text{the frequency of term t in d document}}{\text{the total number of term in document d}} \tag{6}$$

$$\mathrm{IDF}_{(D)} = \log(\frac{\text{the total number of documents in collection D}}{1+\text{the number of documents containing term t}}) \tag{7}$$

and

t : term being counted for its frequency
d : document currently being processed
D : collection of documents

2.7. Word Clouds

Word Clouds visually depict text-based information by sizing words based on their frequency in a document. They are commonly used for analyzing website content and documents, showcasing significant words like titles, jounal entries, or tags in a rectangular format. The size and color of each word indicate its frecuency and importance [23]. By examining word frequency, intructors can identify patterns or absences of specific words and phases in text data [24]. However, word clouds alone may not yield actionable insights, necessitating deeper analysis [25].

3. RESULTS AND DISCUSSION

This research studies algorithms that analyze opinion sentiment towards films based on twitter data. The data was processed using Google Colab tools, including Naïve Bayes, Random Forest, and Decision Tree algorithms.

3.1. Naïve Bayes Algorithm

The accuracy of Naive Bayes is 93.1%, as detailed in the confusion matrix with categories negative, positive, and neutral, and in the classfication report, which includes precision, recall, F1-score, and support.

1. Confusion Matrix
   Based on Table 3, illustrates the performance of the Naïve Bayes algorithm across three categories: negative, positive, and neutral. It correctly classified 86 instances as negative, 79 instances as positive, and 64 instances as neutral. Classification errors occurred with 8 instances of positive being incorrectly classified as negative, 2 instances of positive being incorrectly classified as neutral. 18 instances of neutral being incorrectly classified as negative, and 10 instances of neutral being incorrectly classified as positive. The illustrates the algorithm's effectiveness in classifying data, with the majority of classfications being correct.

Table 3. Naïve Bayes Confusion Matrix

	Negative	Positive	Neutral
Negative	86	0	0
Positive	8	79	2
Neutral	18	10	64

2. Classification Report Naive Bayes
   By examining the Naive Bayes classification report, classification can be performed for each class and overall. Precision, recall, and F1-score offer insights into the model's predictive performance for each class, while accuracy provides an overall assessment of the model's ability to make correct predictions.

Table 4. Classification Report Naïve Bayes

	Precision	Recall	F1-Score	Support
0	0,77	1,00	0,87	86
1	0,89	0,89	0,89	89
2	0,97	0,70	0,81	92
Accuracy			0,86	267
Macro Aug	0,88	0,86	0,86	267
Weighted Aug	0,88	0,86	0,85	267

Based on Table 4, the accuracy achieved with Naive Bayes is 86%, while the highest accuracy is 91%. The recall for class 0 is 1.00, indicating that the model can perfectly identify all instances of class 0.

3.2. Decision Tree Algorithm

The accuracy achieved with the Decision Tree algorithm is 88.9%. Below you can see the Confusion matrix for the decision tree, depicting the positiv, negative and neutral class.

1. Cofusion Decision Tree
   Based on Tabel 4, the confusion matrix can be interpreted as a decision tree with predicted categories as the first branch and actual categories as the second branch. Out of 91 prediction of negative, 83 instances were truly negative, 2 were incorrectly classified as positive, and 1 was incorrectly classified as neutral. This illustrates the classification flow and the errors that occur with each predicted category compared to the actual categories.

Table 5. Confusion Matrix Decision Tree

	Negative	Positive	Neutral
Negative	83	2	1
Positive	1	88	0
Neutral	7	12	73

2. Clasification Decision Tree
   Observing the decision tree algorithm's classification report, Precision, recall, F1-Score were evaluated.

Table 6. Classification Report Decision Tree

	Precision	Recall	F1-Score	Support
0	0,91	0,97	0,94	86
1	0,86	0,99	0,92	89
2	0,99	0,79	0,88	92
Accuracy			0,91	267
Macro Aug	0,92	0,92	0,91	267
Weighted Aug	0,92	0,91	0,91	267

Meanwhile, when comparing the Decision Tree and Random Forest algorithms, the Decision Tree achieved the highest accuracy at 91%. The best precision is in class 2, with a percentage of 99%. The best recall is 99%, and the best F1-Score is 94%. In this algorithm, class 1 performs the best is compared to class 0 and class 3.

3.3. Random Forest Algorithm

The accuracy obtained was 88%, the best precision was found in class 0 and 2, while the lowest was in class 1, with a percentage of 75%. Class 1 had nearly perfect recall at 99%, and the highest F1-Score was in class 0, with a percentage of 95%. The following is a table for Random Forest classification.

Table 7. Classification Report Random Forest

	Precision	Recall	F1-Score	Support
0	0,99	0,92	0,95	86
1	0,75	0,99	0,85	89
2	0,99	0,75	0,85	92
Accuracy			0,88	267
Macro Aug	0,91	0,89	0,89	267
Weighted Aug	0,91	0,88	0,88	267

Based on these three tables, after testing three different algorithms with a 20:80 data split, the accuracy obtained from each algorithm is high. There is not a significant difference in accuracy among the three algorithms: Naive Bayes at 86%, Random Forest at 88%, and the highest, Decision Tree, at 91%. Therefore, the best algorithm for analyzing opinion sentiment regarding the Taylor Swift: The Eras Tour film is Decision Tree. Figure 2 is a comparison graph of three algorithms tested.

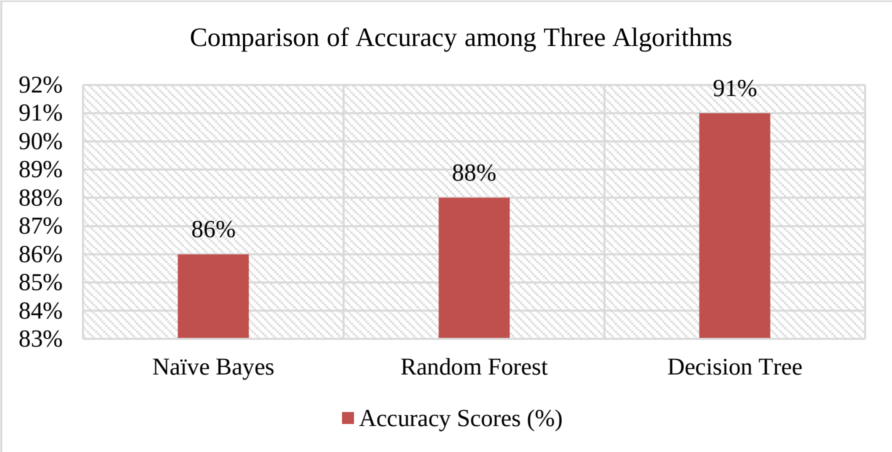

Figure 2. Accuracy Comparison of Three Algorithms

Based on Figure 1, the comparison among the three algorithms can be seen, with a notable difference between each algorithm, especially between Decision Tree and Naïve Bayes. Thus, after evaluation, it was that Decision Tree is the best algorithm with an accuracy of 91%.

3.4 Data Visualization

Wordcloud is a visualization of a series of words where frequently occurring words are shown in larger sizes. This provides an instant visual representation of how often these words appear in a text or a collection of documents.

1. Visualization of the Most Frequent Words
   The following is a visualization of the opinions obtained after data cleaning, text cleaning, TF-IDF, and SMOTE. Based on the Figure 3, the frequently mentioned words include Swift, Era, Taylor, Tour, taylorswift, taylorswifterastour, ticket, eras tour, and friendship. This indicates that these words appear frequently in tweets, suggesting a focus on these topics.

Figure 3. Visualization of the word The Most Word

2. Visualization of Positive Wordcloud Opinion
   Based on Figure 4, in the Positive feature aspect of opinion towards the Taylor Swift's film The Eras Tour, the words frequently appear in each tweet are 'swift', 'taylor', 'era', 'taylorswifterastour', 'tour',

'like', 'concert', 'friendship', 'interest', 'ticket', 'chief' , 'hope'. This indicates that the positive opinion of the film have resonated with audiences, especially regarding the recap of Taylor Swift's concert recordings across multiple countries and the friendships cultivated backstage during her concert tour.

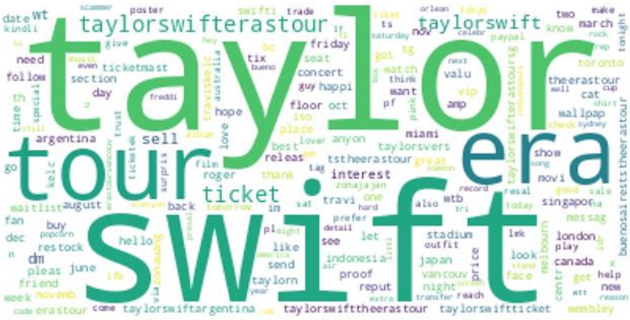

Figure 4. Wordcloud Positive Opinion Visualization

3. Visualizatin of Negative Wordcloud Opinion
   Based on Figure 5, in the negative aspects of opinions regarding Taylor Swift's film The Eras Tour, the words that frequently appear are 'taylorswifterastour', 'taylor', 'swift', 'ticket', 'swift', 'sell', 'price', 'wrong', 'proof', 'scam', 'gross', 'cinema', 'hate', 'broke', 'freak', 'scream', 'bad'. This indicates that some viewers who watched the film complained about the ticket prices, their dislike for the film's visualization, and scams that occurred during the film's screening in cinemas. These negative opinions reflect the audience's experiences and perceptions while watchingTaylor Swift's The Eras Tour.

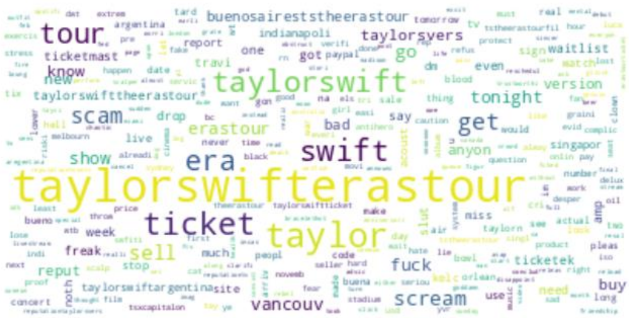

Figure 5. Visualization of Negative Wordcloud Opinions

4. CONCLUSION

After conducting various series of research, it can be concluded that the Decision Tree algorithm achieved the highest accuracy rate at 91% in the sentiment analysis study of the 'Taylor Swift: The Eras Tour' film. The frequently identified words mainly include the name of the film itself. We also compared it with the KNN algorithm, but the accuracy obtained was far below expectations, at 64%. We hope that in future research, emotional sentiment can be incorporated to categorize each word tweeted or appearing in sentiment analysis, rather than solely relying on positive, negative, and neutral categorizations.

REFERENCES

[1] Y. Nurtikasari, Syariful Alam, and Teguh Iman Hermanto, "Analisis Sentimen Opini Masyarakat Terhadap Film Pada Platform Twitter Menggunakan Algoritma Naive Bayes," *INSOLOGI: Jurnal Sains dan Teknologi*, vol. 1, no. 4, pp. 411–423, Aug. 2022, doi: 10.55123/insologi.v1i4.770.

[2] R. Amelia, D. Darmansah, N. S. Prastiwi, and M. E. Purbaya, "Impementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Masyarakat Indonesia Mengenai Drama Korea Pada Twitter," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 2, p. 338, Apr. 2022, doi: 10.30865/jurikom.v9i2.3895.

[3] R. Fajar, S. Program, P. Rekayasa, N. Lunak, and R. Bengkalis, "Implementasi Algoritma Naive Bayes Terhadap Analisis Sentimen Opini Film Pada Twitter," vol. 3, no. 1.

[4] A. D. Dwi, S. Hafidz, S. W. H. Aji, Y. A. Almi, and U. Ema, "Perbandingan Metode Random Forest dan KNN pada," 2023.

[5] O. I. Gifari, M. Adha, I. Rifky Hendrawan, F. Freddy, and S. Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," *JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY*, vol. 2, no. 1, 2022, doi: 10.46229/jifotech.v2i1.330.

[6] A. Firdausi Nuzula Zamzami and Mahendra Dwifebri P, “Perbandingan Metode Random Forest dan KNN pada,” 2021. doi: DOI 10.30865/mib.v5i2.2835.

[7] R. A. M. Fahrul, L. A. Nuril, and I. Uce, “Prediksi Penyakit Hipertensi Menggunakan Metode Decison Tree dan Random Forest,” *Jurnal Ilmiah Komputasi*, vol. 23, no. 1, Mar. 2024, doi: 10.32409/jikstik.23.1.3503.

[8] V. A. Fitri, R. Andreswari, and M. A. Hasibuan, “Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm,” in *Procedia Computer Science*, Elsevier B.V., 2019, pp. 765–772. doi: 10.1016/j.procs.2019.11.181.

[9] M. F. Asshiddiqi and K. M. Lhaksmana, “Perbandingan Metode Decision Tree dan Support Vector Machine untuk Analisis Sentimen pada Instagram Mengenai Kinerja PSSI,” 2020.

[10] E. Fitri, Y. Yuliani, S. Rosyida, and W. Gata, “Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine,” *TRANSFORMTIKA*, vol. 18, no. 1, pp. 71–80, 2020, [Online]. Available: www.nusamandiri.ac.id,

[11] Y. Wang, G. Shen, and L. Hu, “Importance evaluation of movie aspects: Aspect-based sentiment analysis,” *Proceedings - 2020 5th International Conference on Mechanical, Control and Computer Engineering, ICMCCE 2020*, pp. 2444–2448, 2020, doi: 10.1109/ICMCCE51767.2020.00527.

[12] M. Wankhade, A. C. S. Rao, and C. Kulkarni, *A survey on sentiment analysis methods, applications, and challenges*, vol. 55, no. 7. Springer Netherlands, 2022. doi: 10.1007/s10462-022-10144-1.

[13] P. Mehta and S. Pandya, “A review on sentiment analysis methodologies, practices and applications,” *International Journal of Scientific and Technology Research*, vol. 9, no. 2, pp. 601–609, 2020.

[14] V. Jackins, S. Vimal, M. Kaliappan, and M. Y. Lee, “AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes,” *Journal of Supercomputing*, vol. 77, no. 5, pp. 5198–5219, 2021, doi: 10.1007/s11227-020-03481-x.

[15] I. A. P. Banlawe, J. C. D. Cruz, J. C. P. Gaspar, and E. J. I. Gutierrez, “Optimal Frequency Characterization of Mango Pulp Weevil Mating Activity using Naïve Bayes Classifier Algorithm,” *Proceeding - 2021 IEEE 17th International Colloquium on Signal Processing and Its Applications, CSPA 2021*, no. March, pp. 116–120, 2021, doi: 10.1109/CSPA52141.2021.9377277.

[16] R. Das, S. K. Biswas, D. Devi, and B. Sarma, “An Oversampling Technique by Integrating Reverse Nearest Neighbor in SMOTE: Reverse-SMOTE,” *Proceedings - International Conference on Smart Electronics and Communication, ICOSEC 2020*, no. Icosec, pp. 1239–1244, 2020, doi: 10.1109/ICOSEC49089.2020.9215387.

[17] W. Satriaji and R. Kusumaningrum, “Effect of Synthetic Minority Oversampling Technique (SMOTE), Feature Representation, and Classification Algorithm on Imbalanced Sentiment Analysis,” *2018 2nd International Conference on Informatics and Computational Sciences, ICICoS 2018*, pp. 99–103, 2018, doi: 10.1109/ICICOS.2018.8621648.

[18] M. Guia and R. R. Silva, “Comparison of Naïve Bayes , Support Vector Machine , Decision Trees and Random Forest on Sentiment Analysis,” no. Ic3k, pp. 525–531, 2019, doi: 10.5220/0008364105250531.

[19] T. D. Alhafidh, P. H. Gunawan, and B. A. Wahyudi, “The Sentiment Analysis of Spider-Man : No Way Home Film Based on,” vol. 5, no. 158, pp. 177–182, 2022.

[20] P. Baid, “Sentiment Analysis of Movie Reviews using Machine Learning Techniques,” vol. 179, no. 7, pp. 45–49, 2017.

[21] D. Yuan, J. Huang, X. Yang, and J. Cui, “Improved random forest classification approach based on hybrid clustering selection,” *Proceedings - 2020 Chinese Automation Congress, CAC 2020*, pp. 1559–1563, 2020, doi: 10.1109/CAC51589.2020.9326711.

[22] M. F. Fayyad and D. T. Savra, “Sentiment Analysis of Towards Electric Cars using Naive Bayes Classifier and Support Vector Machine Algorithm,” vol. 1, no. July, pp. 1–9, 2023.

[23] P. Susan Ramlo, “Using Word Clouds to Visually Present Q Methodology Data and Findings,” *Human Subjectivity*, vol. 9, no. 2, pp. 99–111, 2011, doi: 10.16194/j.cnki.31-1059/g4.2011.07.016.

[24] C. A. DePaolo and K. Wilkinson, “Get Your Head into the Clouds: Using Word Clouds for Analyzing Qualitative Assessment Data,” *TechTrends*, vol. 58, no. 3, pp. 38–44, 2014, doi: 10.1007/s11528-014-0750-9.

[25] S. Dharaiya, B. Soneji, D. Kakkad, and N. Tada, “Generating Positive and Negative Sentiment Word Clouds from E-Commerce Product Reviews,” *2020 International Conference on Computational Performance Evaluation, ComPE 2020*, pp. 459–463, 2020, doi: 10.1109/ComPE49325.2020.9200056.