Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th June 2025. Vol. 4. No. 3; e-ISSN: 2808-4519

Classification of Youtube User sentiment on 5G Technology Videos with Naïve Bayes

Taupik Hidayat[1]*, Agus Bahtiar[2], Kaslani[3]

[1]*Informatics Engginering, STMIK IKMI Cirebon, Indonesia*
[2]*Information System, STMIK IKMI Cirebon, Indonesia*
[3]*Computerized Accounting, STMIK IKMI Cirebon, Indonesia*
zzgabriela044@gmail.com[1], agusbahtiar038@gmail.com[2], kaslani@ikmi.ac.id[3]*

Abstract

The rapid development of 5G technology has triggered various reactions from the public on social media platforms such as YouTube. User sentiment towards videos discussing 5G technology varies, from positive to negative. This research aims to improve the sentiment classification model of YouTube user reviews of videos about 5G technology with the Naïve Bayes algorithm, which is known to be able to handle large text data and sentiment variations. This research goes through the stages of collecting review data from YouTube, data preprocessing including tokenization, stop word removal, and stemming, and sentiment classification into positive, neutral, and negative categories using Naïve Bayes. The bag-of-words (BOW) technique is used to improve the algorithm's performance. The results showed a sentiment distribution of 1,581 neutral, 1,165 positive, and 517 negative. The proposed model achieved 98% accuracy, with macro average precision 0.99, recall 0.98, and F1-score 0.98. Weighted average resulted in precision 0.98, recall 0.98, and F1-score 0.98. These results show the model performs very well in sentiment classification. This research is expected to make a significant contribution in understanding public perception of 5G technology.

Keywords: *Analytics Sentiment, Naïve Bayes, 5G Technology, YouTube, TF-IDF*

1. Introduction

The development of information and communication technology has progressed rapidly. These advancements have had a significant impact on various sectors, including industry, education, health, and social. One of the most important developments in technology is the advent of 5G networks, which are expected to improve data access speeds, network capacity, and overall communication efficiency. 5G technology promises significantly higher data rates, lower latency, and the ability to support millions of connected devices simultaneously, making it critical for advanced applications such as the Internet of Things (IoT), autonomous vehicles, virtual reality, and smart cities.[1]

Youtube is a popular video sharing platform where users can interact, watch, comment and share videos for free. There are many types of videos that users can upload on the Youtube platform. One example of this type of video is a video that discusses a particular product. The many products that are discussed or reviewed cause irrelevant comments, namely different public sentiments.[2]. Youtube is also a means of expressing opinions or opinions from each individual, both through video uploads and from user comments. Users can freely provide opinions through the comments column regarding the video content they watch. [3]

Term Frequency-Inverse Document Frequency (TFG-IDF) is a method to weight words against documents. This method calculates the weight based on how often the word appears in the document and how often the word appears in the document set. If the word appears frequently in the document, then the weight will be high.[4]

In addition to the TF-IDF method, the Bag of Words (BOW) method is one of the methods used to retrieve features from sentences or documents. The BOW model converts text data into a vector of predetermined length through counting the occurrences of words. The use of BOW involves two stages, namely determining the words present in the entire sentence or document, and determining the scoring method for each occurrence of the previously identified words.[5]

Naive Bayes is a classification based on probability theory created by Thomas Bayes, an English scientist, which provides estimates for potential likelihoods based on past performance, known as Bayes' Theorem.[6]

The problem formulation of this research is how to apply the Naïve Bayes algorithm in classifying the sentiment of YouTube user comments on videos that discuss 5G technology. This research also aims to evaluate the performance of the Naïve Bayes algorithm in the classification of comment sentiment into positive, negative, and neutral categories, and determine the distribution of the comment sentiment level. The results of this research are expected to provide a deeper understanding of public perceptions of 5G technology.

2. Literature Review

2.1. Data Mining

Data mining is the process of collecting and processing data to extract important information from the data. The information obtained can be in the form of numbers, or information that can be used for various purposes. The process of collecting and extracting this information can be done using software with the help of statistical calculations, mathematics or the latest using Artificial Intelligence (AI).[7].

2.2. Naïve Bayes

Previous research by (Febriyanti & Syofiani, 2023) discusses the analysis of public sentiment towards the concept of childfree based on comments taken from the YouTube platform. This research uses a data collection method through crawling comments with the help of the YouTube API, followed by a preprocessing process that includes tokenization, transformation, and classification using the Naïve Bayes algorithm. The analysis results show that the majority of public reactions to childfree tend to be negative, with an accuracy rate of 97%. In addition, this research also presents data visualization in the form of word clouds to describe words that often appear in comments, both positive and negative. Recommendations for future research include taking data from international comments and using other sentiment analysis algorithms.

To evaluate the performance of the classification model in this study, several evaluation metrics that are common in machine learning are used, namely accuracy, precision, recall, and F1-score. These metrics aim to measure how well the model classifies the available review data. The following is an explanation and formula for each evaluation metric:

$$Accuracy = \frac{number\ of\ correct\ predictions}{total\ amount\ of\ data}$$

$$Precision = \frac{True\ Positive}{True\ positive\ \ + False\ Positive}$$

$$Recall = \frac{True\ Positive}{True\ positive\ \ + False\ Negative}$$

3. Research Method

The methodology used in this research, namely Knowledge Discovery in Database (KDD) itself is a technique for obtaining information or knowledge from selected data [8]. There are 5 stages as follows Data Selection, Preprocessing, Transformation, Data Mining, Evaluation. Figure 1 is the research methodology that will be carried out.

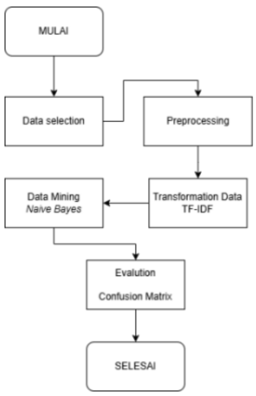

Fig. 1: Research Method

Table 1: Activity Description Research Method

Stages	Activity	Activity Description
Data Selection	Data Collection	The data used in this research consists of user comments on YouTube videos discussing 5G technology. The data was collected using the YouTube API.
Preprocessing	Data Cleaning & Labeling	This step includes removing special characters, numbers, and URLs, as well as assigning sentiment labels (positive, neutral, negative) based on the comment content.

Transformation Data	TF-IDF Weighting	Cleaned comments are transformed into numerical vectors using the *Term Frequency–Inverse Document Frequency* (TF-IDF) method.
Data Mining	Naïve Bayes Classification	The Naïve Bayes algorithm is applied to classify comments into three sentiment classes based on features generated from the TF-IDF process.
Evaluation	Model Performance Testing	Model performance is evaluated using metrics such as *Accuracy*, *Precision*, *Recall*, and *F1-Score* to determine the effectiveness of the classification.

3.1. Data Selection

In this initial stage is data collection and data labeling. Data retrieved using web scrapping techniques using phyton programming [9] Crawling data, comment collection (comment crawling) is taken by crawling data on YouTube social media with the help of the YouTube API (Application Programming Interface) [10] In this data crawling process, it is done by pulling comments from one of the YouTube content creators with the topic of 5G Technology.

3.2. Preprocessing

Preprocessing is used to clean and prepare the dataset before further analysis. It processes the removal of irrelevant data such as URLs, hastags, as well as the removal of non-alphabetic character symbols. Furthermore, text data is converted into lowercase letters, tokenized, stopwords, and stemming. Through these preprocessing steps, text data can be processed into a more structured form, and there are no more irrelevant text characters, so the data is ready for further processing.

3.3. Transformation Data

Data transformation aims to get the required document representation. In this transformation stage, the author performs feature extraction with TF-IDF. [9].

3.4. Data Mining

Data Mining is carried out for the sentiment classification process on review data that has gone through the transformation stage using the Naïve Bayes algorithm.

3.5 Evalution

In the evaluation stage, an assessment of the performance of the Naïve Bayes model that has been trained using user comment data is carried out. The evaluation is conducted on test data to measure the extent to which the model can classify sentiment into three categories Negative, Neutral, and Positive.

4. Result and Discussion

4.1. Research Result

This research applies the Knowledge Discovery in Database (KDD) method through a quantitative approach to analyze YouTube user sentiment towards videos discussing 5G technology, using the Naïve Bayes algorithm.

4.1.1. Data Selection Results

The data used for the sentiment analysis process comes from YouTube user comments on 5G technology discussion videos totaling 3326 comments, then crawling comment data from YouTube, researchers do Data Selection first because it only focuses on video comments on the topic of 5G Technology. The results of the data crawlng process are depicted in Figure 2.

	publishedAt	authorDisplayName	textDisplay	likeCount
0	2024-10-12T12:00:13Z	@ArrisYeYeKurniawan	Indonesia..?? 😅😂🤣 Yang jauh lebih penti...	0
1	2024-10-02T05:31:53Z	@AntaGonisme99	Kalo beli hp yg 5g apakah masih bisa dipakai d...	0
2	2024-09-24T03:53:58Z	@SetiaSetia-b9l	Selamat pagi🌞 semua nya salam kenal budaya bor...	1
3	2024-09-19T01:52:55Z	@username-7699	Boro2 mau 5G sedangkan 4G aja masih amburadul 😂😂	0
4	2024-09-08T08:08:07Z	@ohhnobb	Saya tinggal di daerah yang bukan prioritas 5G...	0
...	...	...	...	...
3332	2022-07-16T04:32:33Z	@alfathaja6771	First	0
3333	2022-07-16T04:32:31Z	@fajarrudin4544	Iya	0
3334	2022-07-16T04:32:30Z	@danukusumo3298	Like dan komentar pertama	0
3335	2022-07-16T04:32:29Z	@maggakurnia7237	first ye	0
3336	2022-07-16T04:32:27Z	@tacpahdza	Iyah	2

3337 rows × 4 columns

Fig. 2: Data Crawling Results

4.1.2. Preprocessing

Preprocessing is the initial stage that must be done before analyzing data. Data preprocessing aims to turn unstructured data into structured data to facilitate the analysis process. The data preprocessing process involves several stages such as cleaning, transformation, tokenizing, stopword removal, stemming, translation, and labeling.

	textDisplay	cleaned_text	case_folding	tokenize	Filtering/stopword removal	stemming_data	Compound_Score	Sentiments
0	Indonesia..?? 😂😂😂 Yang jauh lebih penti...	Indonesia Yang jauh lebih penting bukan 5G ata...	indonesia yang jauh lebih penting bukan 5g ata...	['indonesia', 'yang', 'jauh', 'lebih', 'pentin...	['indonesia', '5g', '4g', 'membasmi', 'otak', ...	Indonesia 5G 4G eradicate the brains of corrup...	-0.2263	Negatif
1	Kalo beli hp yg 5g apakah masih bisa dipakai d...	Kalo beli hp yg 5g apakah masih bisa dipakai d...	kalo beli hp yg 5g apakah masih bisa dipakai d...	['kalo', 'beli', 'hp', 'yg', '5g', 'apakah', '...	['kalo', 'beli', 'hp', 'yg', '5g', 'dipakai', ...	If you buy a 5G cellphone, use an area with 5G...	0.0000	Netral
2	**Selamat pagi😊 semua nya salam kenal budaya bor...**			[]	[]	in	0.0000	Netral
3	Boro2 mau 5G sedangkan 4G aja masih amburadul 😂😂	Boro2 mau 5G sedangkan 4G aja masih amburadul	boro2 mau 5g sedangkan 4g aja masih amburadul	['boro2', 'mau', '5g', 'sedangkan', '4g', 'aja...	['boro2', '5g', '4g', 'aja', 'amburadul']	boro2 5g 4g just a mess	-0.3612	Negatif
4	Saya tinggal di daerah yang bukan prioritas 5G...	Saya tinggal di daerah yang bukan prioritas 5G...	saya tinggal di daerah yang bukan prioritas 5g...	['saya', 'tinggal', 'di', 'daerah', 'yang', 'b...	['tinggal', 'daerah', 'prioritas', '5g', 'nggk...	Just a 5G priority area without a 5G cellphone	0.0000	Netral

Fig. 3: Preprocessing Result

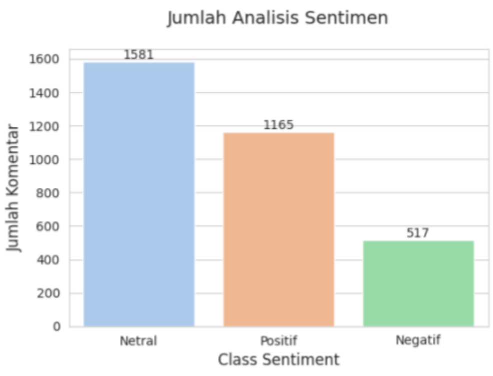

Fig. 4: Labeling Result

It can be seen that the sentiment of YouTube user comments on 5G network technology is dominated by neutral sentiment of 48.5% or 1,581 data out of a total of 3,263 comment data.

4.1.3. Transformation Data

In the process of data mining, a numeric dataset is needed because the current dataset is still in the form of a string, so it is necessary to change it into word form. This stage goes through several processes, namely changing the data type to numeric and weighting words using the TF- IDF algorithm. The calculation of word weight is done by first determining the Term Frequency (TF) value [11]. The following is the weighting of words on test and training data.

```
Pembobotan Kata Data Latih:
      0000  0000 bangun   05  05 mbps   07  07 mbps  0ms   10  10 20  10 5g  \
0      0.0           0.0  0.0     0.0  0.0      0.0  0.0  0.0    0.0    0.0
1      0.0           0.0  0.0     0.0  0.0      0.0  0.0  0.0    0.0    0.0
2      0.0           0.0  0.0     0.0  0.0      0.0  0.0  0.0    0.0    0.0
3      0.0           0.0  0.0     0.0  0.0      0.0  0.0  0.0    0.0    0.0
4      0.0           0.0  0.0     0.0  0.0      0.0  1.0  0.0    0.0    0.0
...    ...           ...  ...     ...  ...      ...  ...  ...    ...    ...
2602   0.0           0.0  0.0     0.0  0.0      0.0  0.0  0.0    0.0    0.0
2603   0.0           0.0  0.0     0.0  0.0      0.0  0.0  0.0    0.0    0.0
2604   0.0           0.0  0.0     0.0  0.0      0.0  0.0  0.0    0.0    0.0
2605   0.0           0.0  0.0     0.0  0.0      0.0  0.0  0.0    0.0    0.0
2606   0.0           0.0  0.0     0.0  0.0      0.0  0.0  0.0    0.0    0.0

      ...  zeperti  zeus  zeus did  zona  zona hijau  zona merah  zona nyoman
0     ...      0.0   0.0       0.0   0.0         0.0         0.0          0.0
1     ...      0.0   0.0       0.0   0.0         0.0         0.0          0.0
2     ...      0.0   0.0       0.0   0.0         0.0         0.0          0.0
3     ...      0.0   0.0       0.0   0.0         0.0         0.0          0.0
4     ...      0.0   0.0       0.0   0.0         0.0         0.0          0.0
...   ...      ...   ...       ...   ...         ...         ...          ...
2602  ...      0.0   0.0       0.0   0.0         0.0         0.0          0.0
2603  ...      0.0   0.0       0.0   0.0         0.0         0.0          0.0
2604  ...      0.0   0.0       0.0   0.0         0.0         0.0          0.0
2605  ...      0.0   0.0       0.0   0.0         0.0         0.0          0.0
2606  ...      0.0   0.0       0.0   0.0         0.0         0.0          0.0
```

Fig. 3 :Weighting words in the training data

```
Pembobotan Kata untuk Data Uji:
     000  000 space  0000  0000 in   05  05 mbps   10  10 20  10 5g  \
0    0.0        0.0   0.0      0.0  0.0      0.0  0.0    0.0    0.0
1    0.0        0.0   0.0      0.0  0.0      0.0  0.0    0.0    0.0
2    0.0        0.0   0.0      0.0  0.0      0.0  0.0    0.0    0.0
3    0.0        0.0   0.0      0.0  0.0      0.0  0.0    0.0    0.0
4    0.0        0.0   0.0      0.0  0.0      0.0  0.0    0.0    0.0
..   ...        ...   ...      ...  ...      ...  ...    ...    ...
648  0.0        0.0   0.0      0.0  0.0      0.0  0.0    0.0    0.0
649  0.0        0.0   0.0      0.0  0.0      0.0  0.0    0.0    0.0
650  0.0        0.0   0.0      0.0  0.0      0.0  0.0    0.0    0.0
651  0.0        0.0   0.0      0.0  0.0      0.0  0.0    0.0    0.0
652  0.0        0.0   0.0      0.0  0.0      0.0  0.0    0.0    0.0

     10 countries  ...  zone area  zone bro  zone is  zone tower  zones  \
0             0.0  ...        0.0       0.0      0.0         0.0    0.0
1             0.0  ...        0.0       0.0      0.0         0.0    0.0
2             0.0  ...        0.0       0.0      0.0         0.0    0.0
3             0.0  ...        0.0       0.0      0.0         0.0    0.0
4             0.0  ...        0.0       0.0      0.0         0.0    0.0
..            ...  ...        ...       ...      ...         ...    ...
648           0.0  ...        0.0       0.0      0.0         0.0    0.0
649           0.0  ...        0.0       0.0      0.0         0.0    0.0
650           0.0  ...        0.0       0.0      0.0         0.0    0.0
651           0.0  ...        0.0       0.0      0.0         0.0    0.0
652           0.0  ...        0.0       0.0      0.0         0.0    0.0
```

Fig. 4 : Word weighting on test data

Splitting Data; Data splitting is an important process in developing machine learning models. This process aims to divide the dataset into two main parts, one for training the model and the other for testing the trained model with a ratio of 80:20. The main goal is to avoid overfitting and ensure that the model built can generalize well to new data that has never been seen before.

4.1.4. Data Mining

Data Mining is the process of extracting knowledge or information from large unstructured data sets with the aim of discovering useful patterns or relationships. In the context of KDD, the data mining process involves the use of statistical techniques, algorithms, and machine learning to extract relevant information from a prepared dataset. Here are the results of the data mining process:

```
Akurasi Data latih: 0.9521072796934866
Confusion Matrix data latih:
         Negatif  Netral  Positif
Negatif      389       6       38
Netral         0    1184       77
Positif        1       3      912
```

Fig. 7: Training Data Evaluation Results

The classification model performed very well on the training data with an accuracy of 95.21%. This shows that the model is able to predict the class with a high success rate. Based on the confusion matrix, for the Negative class, there were 389 correct predictions, with 6 data incorrectly predicted as Neutral and 38 data incorrectly predicted as Positive. For the Neutral class, there were 1184 correct predictions, while 77 data were incorrectly predicted as Positive. In the Positive class, the model successfully predicted 912 data correctly, but there was 1 data incorrectly predicted as Negative and 3 data incorrectly predicted as Neutral.

```
Akurasi Data Uji: 0.9831546707503829
Confusion Matrix data Uji:
         Negatif  Netral  Positif
Negatif       82       1        2
Netral         0     316        8
Positif        0       0      244
```

Fig. 8: Test data evaluation results

The model showed excellent performance on the test data with an accuracy of 98.32%, indicating a high ability to correctly predict the class. Based on the confusion matrix, for the Negative class, there were 82 correct predictions, with 1 data incorrectly predicted as Neutral and 2 data incorrectly predicted as Positive. For the Neutral class, the model successfully predicted 316 data correctly, but there were 8 data incorrectly predicted as Positive. Meanwhile, for the Positive class, the model gave highly accurate results with 244 correct predictions without error. This result confirms that the model performs very well, especially in distinguishing between Neutral and Positive classes on the test data, with only a few errors occurring on the Negative class.

4.1.5. Evaluation

In the evaluation stage, an assessment of the performance of the Naïve Bayes model that has been trained using user comment data is carried out. Evaluation is carried out on test data to measure the extent to which the model can classify sentiment into three categories Negative, Neutral, and Positive. The following are the results of the evaluation carried out on Naïve Bayes performance

```
Hasil Evaluasi Model Naïve Bayes pada Data Uji
==================================================
Akurasi           : 0.9816
Precision per kelas: Negatif=1.0000, Netral=0.9968, Positif=0.9570
Recall per kelas   : Negatif=0.9535, Netral=0.9752, Positif=1.0000
F1-score per kelas : Negatif=0.9762, Netral=0.9859, Positif=0.9780

Precision keseluruhan: 0.9823
Recall keseluruhan   : 0.9816
F1-score keseluruhan : 0.9817

Laporan Klasifikasi:
              precision    recall  f1-score   support

     Negatif       1.00      0.95      0.98        86
      Netral       1.00      0.98      0.99       322
     Positif       0.96      1.00      0.98       245

    accuracy                           0.98       653
   macro avg       0.98      0.98      0.98       653
weighted avg       0.98      0.98      0.98       653
```

Fig. 9: Naive Bayes model evaluation results on test data

The model performed very well in the classification of the three classes of negative, neutral, and positive, with an overall accuracy of 98%. For the negative class, precision reached 1.00, indicating all predictions for this class were correct. However, recall was slightly lower at 0.96, resulting in an F1-score of 0.98 with a support of 85 samples. Performance for the neutral class was excellent, with a precision of 1.00 and recall of 0.98, resulting in an F1-score of 0.99 from 324 samples. The positive class showed a precision of 0.96, with a perfect recall of 1.00, resulting in an F1-score of 0.98 from 244 samples. Overall, precision was recorded at 0.98, recall at 0.98, and F1-score at 0.98, showing a very stable performance in classifying all classes. The macro averages for precision, recall, and F1-score were 0.99, 0.98, and 0.98 respectively, indicating that the model worked consistently across all classes. The weighted average, which considers the distribution of the number of samples, shows precision, recall, and F1-score values of 0.98. With these results, the model can be concluded to have consistent and excellent performance in classifying the data, despite the difference in the number of samples in each class.

4.2. Discussion

The results obtained from both training and testing phases indicate that the Naïve Bayes classification model performs with high effectiveness in categorizing user sentiments into Negative, Neutral, and Positive classes on YouTube comments related to 5G technology. During the training phase, the model achieved a high accuracy of 95.21%, demonstrating a strong ability to correctly classify sentiment categories. An analysis of the confusion matrix reveals that for the Negative class, the model made 389 correct predictions, with 6 instances misclassified as Neutral and 38 as Positive. In the Neutral class, the model correctly classified 1,184 instances, while 77 were incorrectly predicted as Positive. For the Positive class, 912 instances were correctly classified, with 1 misclassified as Negative and 3 as Neutral. These results indicate that while the model performs exceptionally well on the training data, there is a small degree of confusion, particularly between Neutral and Positive classes. The performance on the testing data further reinforces the robustness of the model, achieving an even higher accuracy of 98.32%. For the Negative class, the model correctly predicted 82 instances, with only 1 instance misclassified as Neutral and 2 as Positive. The Neutral class had 316 correct predictions, with 8 instances misclassified as Positive. The Positive class achieved flawless results with 244 correct predictions and no misclassification. These results indicate a highly reliable classification capability, particularly in distinguishing between Neutral and Positive sentiments, with minimal confusion across classes.

Evaluation metrics such as Precision, Recall, and F1-Score further affirm the model's strong performance. The Negative class achieved a precision of 1.00, indicating that all predictions labeled as Negative were indeed correct. The recall for this class was slightly lower at 0.96, resulting in an F1-score of 0.98. For the Neutral class, the model reached both precision and recall values close to perfect, at 1.00 and 0.98 respectively, leading to an F1-score of 0.99. The Positive class demonstrated precision of 0.96 and recall of 1.00, with an F1-score of 0.98. The overall accuracy of 98% is complemented by consistent macro and weighted averages of precision, recall, and F1-score at 0.99, 0.98, and 0.98, respectively. These findings highlight the Naïve Bayes model's high stability, generalization capability, and minimal bias across classes, even in the presence of class imbalance. The minor misclassifications observed, particularly between Neutral and Positive classes, suggest an area for potential improvement through further tuning of preprocessing techniques or exploring ensemble models. In conclusion, the Naïve Bayes algorithm proves to be an effective and efficient approach for sentiment classification in multilingual and diverse comment datasets such as those found on YouTube, especially when analyzing topics with varying public opinions like 5G technology.

5. Conclusion

Based on the results of this research, the discussion that has been carried out, it can be concluded as follows:

1. The application of Naïve Bayes algorithm in sentiment classification of YouTube users' comments on 5G technology videos showed excellent performance, with an accuracy of 95.56% on training data and 98.32% on test data. The model was able to classify comments into positive, neutral and negative categories with a very low error rate, and showed a high ability to predict sentiment on new data accurately and consistently.
2. Sentiment distribution analysis shows that the majority of comments are neutral (47.93%), followed by positive (39.36%), and negative (12.71%) comments, indicating that public perception of 5G technology tends to be neutral with a fairly strong positive tendency. This finding confirms that the Naïve Bayes algorithm is not only reliable in sentiment classification, but also able to provide an informative picture of public opinion towards technological developments.

References

[1] H. U. Mustakim, "Tantangan Implementasi 5G di Indonesia," *INTEGER J. Inf. Technol.*, vol. 4, no. 2, pp. 1–10, 2019, doi: 10.31284/j.integer.2019.v4i2.561.

[2] E. F. Saputra and M. R. Pribadi, "Analisis Sentimen Komentar Pada Kanal Youtube The Lazy Monday Menggunakan Algoritma Naive Bayes," *MDP Student Conf.*, vol. 2, no. 1, pp. 17–23, 2023, doi: 10.35957/mdp-sc.v2i1.4283.

[3] S. Mujahidin, B. Prasetio, and M. C. C. Utomo, "Implementasi Analisis Sentimen Masyarakat Mengenai Kenaikan Harga BBM Pada Komentar Youtube Dengan Metode Gaussian naïve bayes," *Voteteknika (Vocational Tek. Elektron. dan Inform.*, vol. 10, no. 3, p. 17, 2022, doi: 10.24036/voteteknika.v10i3.118299.

[4] F. A. Ramadhan, S. H. Sitorus, and T. Rismawan, "Penerapan Metode Multinomial Naïve Bayes untuk Klasifikasi Judul Berita Clickbait dengan Term Frequency - Inverse Document Frequency," *J. Sist. dan Teknol. Inf.*, vol. 11, no. 1, p. 70, 2023, doi: 10.26418/justin.v11i1.57452.

[5] D. Panca Wilie, "Analisis Sentimen Opini Publik Terhadap Chatgpt di Twitter Menggunakan Metode Naive Bayes," *J. Nas. Ilmu Komput.*, vol. 4, no. 4, pp. 2746–1343, 2023, [Online]. Available: https://www.kaggle.com/datasets/charunisa/ChatGPT-sentiment-analysis.

[6] R. Rachman and R. N. Handayani, "Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM," *J. Inform.*, vol. 8, no. 2, pp. 111–122, 2021, doi: 10.31294/ji.v8i2.10494.

[7] Fujiyama, "Pengertian Data Mining dan Implementasinya di Berbagai Sektor," surabaya telkom university. Accessed: Apr. 16, 2025. [Online]. Available: https://surabaya.telkomuniversity.ac.id/pengertian-data-mining-dan-implementasinya-di-berbagai-sektor/?utm_source=chatgpt.com

[8] S. M. Siroj, I. Arwani, and D. E. Ratnawati, "Analisis Sentimen Opini Publik pada Twitter terhadap Efek Pembelajaran Daring di Universitas Brawijaya menggunakan Metode K-Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 7, pp. 3131–3140, 2021, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/9494

[9] Rahel Lina Simanjuntak, Theresia Romauli Siagian, Vina Anggriani, and Arnita Arnita, "Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Shopee Dengan Menggunakan Algoritma Naïve Bayes," *J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 3, no. 3, pp. 23–39, 2023, doi: 10.55606/teknik.v3i3.2411.

[10] F. Syofiani, S. Alam, and M. I. Sulistyo S, "Analisis Sentimen Penilaian Masyarakat Terhadap Childfree Berdasarkan Komentar di Youtube Menggunakan Algoritma Naïve Bayes," *J. Teknol. Inform. dan Komput.*, vol. 9, no. 2, pp. 688–703, 2023, doi: 10.37012/jtik.v9i2.1661.

[11] M. Yusuf Rismanda Gaja, I. Maulana, and O. Komarudin, "Analisis Sentimen Opini Pengguna Aplikasi Vidio Pada Ulasan Playstore Menggunakan Algoritma Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 4, pp. 2767–2774, 2024, doi: 10.36040/jati.v7i4.7197.