Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4606

Submitted : Sep 09, 2024
Accepted : Oct 03, 2024
Published : Oct 11, 2024

Sentiment Analysis of Dune: Part Two Movie Reviews Using the Naive Bayes Method

Diyan Arum Maheswari[1)*], Ahmad Turmudi Zy[2)], Irfan Afriantoro[3)]
[1,2,3)]Universitas Pelita Bangsa, Bekasi, Indonesia
[1)]diyanarum28@gmail.com, [2)]turmudi@pelitabangsa.ac.id, [3)]irfanafriantoro@pelitabangsa.ac.id

ABSTRACT

Research on films is fascinating because of the profound changes that the development of information and communication technology has brought about in our interactions with and consumption of media content. This study performs sentiment analysis on "Dune: Part Two" movie reviews using the Naïve Bayes method. Review data was collected from IMDb and then processed through several stages such as preprocessing, feature selection with TF-IDF, data splitting, and data mining and evaluation. Naïve Bayes was chosen for its simplicity and ability to handle large datasets effectively. The test results showed a high accuracy rate of 95%, indicating that this model can identify positive, negative, and neutral sentiments well. The use of TF-IDF in feature selection allowed the model to focus on important words, enhancing its sentiment classification ability. This research can provide insights into audience perceptions of the film "Dune: Part Two," which is beneficial for the film industry.

Keywords: Sentimen Analysis; Naive Bayes; Movie Review; Dune: Part Two; IMDB, TF-IDF

INTRODUCTION

The film industry has seen significant changes as a result of the information and communication technology's rapid advancement in many facets of society. Films are no longer just a source of entertainment but have also become a fascinating subject of study for many.

In today's digital world, movie reviews written by users on social media platforms and film review websites provide a rich source of data for analysis. These reviews reflect the opinions, emotions, and experiences of viewers after watching a movie, offering valuable insights for filmmakers, researchers, and film enthusiasts.

One such film that has garnered considerable attention is “Dune: Part Two”, a sequel to the movie “Dune” which is adapted from the famous novel by Frank Herbert published in 1965 and first released as a film in 2021. "Dune: Part Two" has received widespread acclaim for its spectacular visual effects. Featuring top-tier actors and actresses such as Timothée Chalamet, Zendaya, and Florence Pugh, the film has become even more recognizable among Box Office movie enthusiasts. According to an article published on CNN Indonesia on March 2, 2024, the film achieved a score of 95% from film critics (*Kritikus Nilai Dune: Part Two 95 Persen, Rekor Baru Denis Villeneuve*, 2024).

In addition to receiving numerous positive comments and reviews praising the story, visuals, and acting, "Dune: Part Two" has also faced criticism and negative reviews. Some viewers may find the plot of the film too complex or slow-paced, making it difficult to follow for those unfamiliar with the original novel. Others have criticized the film's lengthy runtime and certain plot elements that may not align with the expectations of novel fans or general audiences. These criticisms are important to consider as they provide a more comprehensive perspective on how the film is received by different segments of the audience.

The debate surrounding "Dune: Part Two" demonstrates that reviews and opinions of this film have flooded social media, sparking strong and diverse responses from audiences. The film has received over 1,000 reviews on the IMDb website, creating valuable textual data for analysis.

These reviews offer valuable insight into how the film is perceived by a diverse audience, revealing their perceptions, appreciation, and satisfaction. Analyzing these reviews can help us understand which elements are most appreciated by viewers and which aspects may need improvement. Additionally, this study can provide an overview of current audience trends and preferences, which is crucial for the film industry in developing more effective marketing and production strategies in the future.

The aim of this research is to categorize and analyze the sentiment of online reviews of the film "Dune: Part Two" using the Naïve Bayes method. By conducting this analysis, it is expected that the underlying sentiment patterns and trends related to audience perceptions and satisfaction with the film can be identified.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4606

Submitted : Sep 09, 2024
Accepted : Oct 03, 2024
Published : Oct 11, 2024

LITERATURE REVIEW

Previous research titled "Sentiment Analysis of Movie Reviews Using Naïve Bayes Classifier with TF-IDF Features" used a dataset of 50,000 movie reviews, demonstrating that the inclusion of TF-IDF feature selection could enhance the accuracy of the Naïve Bayes classifier. The accuracy achieved by using Naïve Bayes with TF-IDF for IMDb data processing reached 86.48% (Muhammad Thaariq et al., 2023).

Additionally, in a study titled "Sentiment Analysis of Oppenheimer Movie Reviews on IMDb Using the Naïve Bayes Method," the sentiment of the film was successfully analyzed using 10,000 data points from the Kaggle website. The study found that it effectively differentiated between positive and negative sentiments, achieving an accuracy rate of 96%, with an average precision of 98%, an average recall of 90%, and an average F1-score of 93% (Anuar et al., 2023).

In the journal titled Sentiment Analysis of Social Media Twitter Towards the 2024 Indonesian Presidential Candidates Using Naïve Bayes Classification Algorithm, the study demonstrated that the research achieved a fairly high accuracy rate, with 67.23% for presidential candidate Anies, 83.42% for candidate Prabowo, and 88.15% for candidate Ganjar (Makmun et al., 2024).

Additionally, the journal Sentiment Analysis of Data Breaches on Twitter Using the Naive Bayes Algorithm indicated that Naive Bayes produced strong accuracy and results, with an accuracy score of 98.33%, precision of 100%, and recall of 97.13% (Turmudi Zy et al., 2022).

Lastly, in the journal Sentiment Analysis of Film Opinions Using Naïve Bayes and Lexicon Based Features, it was noted that sentiment analysis of film opinions yielded an accuracy of 0.8. However, using the Naïve Bayes method without Lexicon Based Features for film opinion sentiment analysis resulted in a higher accuracy of 0.95. Therefore, it can be concluded that the use of Lexicon Based Features is not superior to using the Naïve Bayes method alone (Kurniawan & Adinugroho, 2019).

In comparison to earlier studies, this research highlights several differences and innovations in the Sentiment Analysis of Dune: Part Two Movie Reviews using the Naïve Bayes method. In contrast to earlier research that examined Oppenheimer and other general movie reviews, this study focuses exclusively on the sentiment analysis of the film Dune: Part Two. Moreover, this study also applies the TF-IDF feature selection method but with a more targeted approach to recent and relevant review data for Dune: Part Two. This study collected review data from IMDb and processed it through several stages such as preprocessing, data splitting, and model evaluation. Although similar methods were used, this study achieved an accuracy of 95%, which is higher than general movie reviews but 1% lower than the research on Oppenheimer reviews. However, this still demonstrates good performance in identifying positive, negative, and neutral sentiments. This indicates that the model used in this study remains effective in the context of sentiment analysis for the film Dune: Part Two.

METHOD

In this research method, the steps or framework used are as follows:

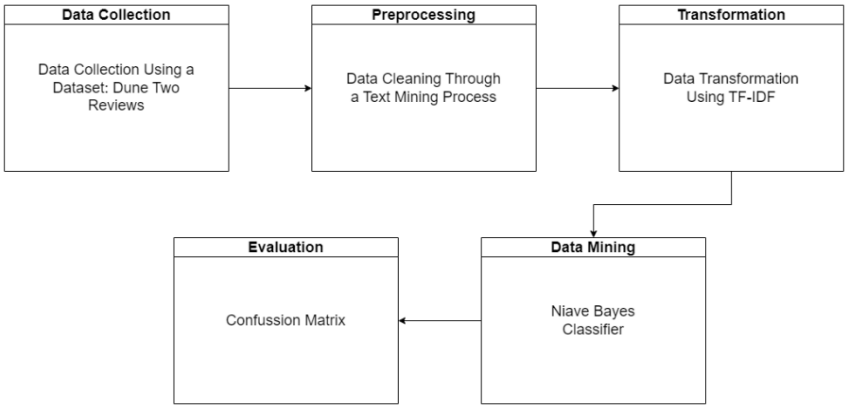

Fig. 1 Research Framework

From the above framework, the steps can be outlined. First, this research utilizes qualitative data from "Dune: Part Two" movie reviews, which were previously collected from a trusted Kaggle site. This data then undergoes preprocessing to clean the text, simplify it, and calculate the importance of each word in the document. The processed data is then used to train the Naïve Bayes algorithm, which creates a sentiment categorization model. The accuracy of

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4606

Submitted : Sep 09, 2024
Accepted : Oct 03, 2024
Published : Oct 11, 2024

this model in forecasting whether movie reviews would be favorable or unfavorable is then evaluated through testing. Metrics including accuracy, precision, recall, and F1-score are used to gauge how well the model performs.

The selection of secondary data was based on its ability to offer comprehensive and contextual information that is not possible with quantitative data. This study uses two attributes: the text containing the movie reviews from viewers and the rating, which is a score ranging from 1-10 to reflect the viewers' sentiment.

Naïve Bayes

Naïve Bayes is one of the popular algorithms used for data mining due to its ease of use, fast processing, and simple implementation structure, which offers high effectiveness (Khotimah & Utami, 2022).

Naïve Bayes is defined as a machine learning method that uses probability computations in another definition. This algorithm works under the premise that classes are independent of one another and makes use of fundamental statistics and probability (Afriansyah et al., 2023).

The Naïve Bayes algorithm calculates the probability of a particular sample characteristic belonging to class h (posterior) by multiplying the probability of class x with the occurrence of the sample characteristic in class c (likelihood). The following equation represents the Naïve Bayes method in its generic form:

$$P(c|x) = \frac{(P(x|c)xP(c))}{P(X)} \qquad (1)$$

Description:

x : is the data with an unknown class.
c : is the hypothesis that data x belongs to a certain class.
$P(c|x)$: is the posterior probability of class c (target) given the predictor x (attribute).
$P\ (c)$: is the prior probability of the class.
$P(x|c)$: is the likelihood of the predictor x occurring in a given class c.
$P\ (x)$: is the prior probability of the predictor x.

RESULT

The review data used by the author comes from a dataset available on the website https://www.kaggle.com/. Kaggle itself is a popular online platform among data scientists and machine learning researchers for sharing dataset, publishing code, and participating in data science competitions.

The number of data used in this study is 1,293 entries from the IMDb Movie Reviews website, which were then reduced to 1,065 entries. It is estimated that 1,065 entries are sufficient to represent the reviews of the film "Dune: Part Two." The review data for "Dune: Part Two" can be sourced from various online platforms, such as movie review websites, social media, and others.

Table 1
Dataset After Cutting

No	Text	Rating
1	Phenomenal stuff. I'll probably calm down tomorrow but right now my heart's beating like crazy and my…	9
2	This is the kind of movie that is impossible to do justice, just by talking about it! It is the kind of…	10
…	…	…
…	…	…
1064	Dune: Part Two is a cinematic triumph, a colossal and ambitious sequel that not only meets but…	7
1065	I know. Poor Stilgar. Without him, we're lost…	?

The next step, after obtaining the data, is to determine the accuracy by going through several stages, including:

Labeling Data

Labeling data is the process of classifying sentiment into specific categories (Muhammad Fernanda Naufal Fathoni et al., 2024). The researcher labels the data by modifying or converting the existing rating column into a column containing sentiment labels. The parameter used to determine the label is based on the rating score column already

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4606

Submitted : Sep 09, 2024
Accepted : Oct 03, 2024
Published : Oct 11, 2024

available in the previous dataset. Below is the process of data labeling using Python programming language.

```
def sentiment_label(score):
  if score > 7:
    return 'Positive'
  elif score < 4:
    return 'negative'
  else:
    return 'neutral'
```

Fig. 2 Data Labeling Process

Below is the result of data labeling.

Table 2. Labeling Data Result

No	Text	Sentiment
1	Phenomenal stuff. I'll probably calm down tomorrow but right now my heart's beating like crazy and my…	positive
2	This is the kind of movie that is impossible to do justice, just by talking about it! It is the kind of…	positive
…	…	…
…	…	…
1064	Dune: Part Two is a cinematic triumph, a colossal and ambitious sequel that not only meets but…	neutral
1065	I know. Poor Stilgar. Without him, we're lost…	neutral

Preprocessing Data

Preprocessing is the initial stage where documents are transformed into data that will be used in subsequent steps. The text is broken down into smaller parts, resulting in segments or tokens. Additionally, unnecessary symbols are removed. Along with symbols, digits are also eliminated (Nur Rozi & Harini Sulistyawati, 2019). This preprocessing stage involves several steps, including:

Cleaning: Cleaning is the process of removing characters that do not conform to the established rules, such as letters or characters outside the alphabet a-z (including punctuation marks), and eliminating links or URLs, hashtags, and usernames (Suryati et al., 2023).

Tokenizing: At this stage, a character-by-character check is performed. If the ith character is not a word separator such as a period (.), comma (,), space, or other delimiter, it is concatenated with the subsequent character (Salam et al., 2018).

Filtering: Using a stoplist method to exclude less relevant terms or a wordlist to keep significant words, filtering entails removing essential words from the tokenized results (Ferryawan et al., 2020).

Stemming: The process of stemming involves taking all affixes from derived words, such as prefixes, infixes, suffixes, and confixes (a combination of prefixes and suffixes), in order to determine the base form of a word (Guterres et al., 2019).

Transform Cases: Case transformation is a common preprocessing step, especially for parameters such as TF-IDF that are available in Rapidminer. In this step, all words within a document are converted to lowercase (Rizkina & Hasan, 2023).

Stopwords: Stopword removal is a process in text preprocessing that involves eliminating less significant words from the tokenized text using a stoplist. Stopwords are common words that frequently appear in large quantities and are considered to have little semantic value (Kelvin et al., 2022).

There is the process of preprocessing data using Python programming language.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4606

Submitted : Sep 09, 2024
Accepted : Oct 03, 2024
Published : Oct 11, 2024

```python
def preprocess_text(text):
    # Converting text to lowercase
    text = text.lower()
    # Removing punctuation
    text = text.translate(str.maketrans ('', '', string.punctuation))
    # Removing numbers
    text = re.sub (r'\d+', '', text)
    # Tokenizing
    word_tokens = word_tokenize (text)
    # Removing stop words
    stop_words = set (stopwords.words ('english'))
    filtered_text = [word for word in word_tokens if word not in stop_words]
    # Stemming stemmer
    stemmer = PorterStemmer()
    stemmed_text = [stemmer.stem (word) for word in filtered_text]
    # Reconstructing the cleaned words into a single string cleaned_text
    cleaned_text = ''.join(stemmed_text)

    return cleaned_text
```

Fig. 3 Preprocessing Data

There is the result of preprocessing data using Python programming language.

Table 3. Preprocessing Data Result

No	Text	Sentiment
1	phenomen stuff ill probabl calm tomorrow right heart beat like crazi brain scream bloodi…	positive
2	kind movi imposs justic talk kind experi never thought would get movi prove wrongthi movi…	positive
…	…	…
…	…	…
1064	dune part two cinemat triumph coloss ambiti sequel meet exce expect set predecessor deni…	neutral
1065	know poor stilgar without lost pretti much minu deplor rave scene thank villeneuvea hero…	neutral

Feature Selection

Feature Selection is a data mining technique frequently used during the pre-processing stage to reduce the complexity of attributes that will be managed in processing and analysis (Made et al., 2019).

TF-IDF is a word-weighting method used in natural language processing that evaluates a word's significance in a given text in relation to the corpus of documents as a whole. TF-IDF weighting is commonly used in information retrieval. This method assigns greater weight to a word that appears frequently in a specific document, while reducing its contribution if the word occurs across multiple documents (Yutika et al., 2021). The feature selection process is implemented using the syntax code below.

```python
# processing Review Data
df['processed_review'] = df['Review Content'].apply(preprocess_text)
# Using TfidfVectorizer
vectorizer = TfidfVectorizer (max_features=1000)
X = vectorizer.fit_transform(df ['processed_review'])
```

Fig. 4 Feature Selection Process

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4606

Submitted : Sep 09, 2024
Accepted : Oct 03, 2024
Published : Oct 11, 2024

Splitting Data

Splitting data is a method of dividing a dataset into two or more parts. Typically, this technique is used to separate the data into two primary segments: training data and testing data. The goal of splitting data is to evaluate how effectively a model performs on unseen data during the training phase (Putra et al., 2024).

A machine learning model is trained using a subset of the dataset. In order to generate predictions on fresh, unseen data, the model uses the labeled data to help it identify patterns and correlations within the data. More training data in higher quality and quantity enhances the model's performance. On the other hand, testing data is an alternative subset that's utilized to assess how well the trained model works. The training step does not make use of testing data. For the testing data, the model predicts labels; these predictions are compared to the actual labels to get the final assessment metrics, which include F1-score, accuracy, precision, and recall.

The process of splitting data can be seen in the following syntax code:

```
# Splitting the dataset into an 80:20 training and testing split
train, test = train_test_split(df, test_size=0.2, random_state = 42)
```

Fig. 5 Splitting Data Process

Data Latih:

	processed_review	sentiment
905	cinematographi movi best cinematographi ever c...	positive
44	feel like im take crazi pill miss dose compreh...	neutral
309	watch dune expect continu interest first part ...	positive
790	epitom scifi think genr timeramaz film act acr...	positive
331	anoth director shouldnt allow near scienc fict...	negative
...	...	...
330	cant believ eye experienc theatr movi one best...	positive
466	didnt quit fulli understand mean behind everyo...	positive
121	someon typic dislik fantasi scifi surprisingli...	positive
1044	expans stori dune told part complet refresh to...	positive
860	dune captur essenc univers dilemma address fra...	positive

Fig. 6 Splitting Data Train Result

Data Uji:

	processed_review	sentiment
31	said deni villeneuv master put grand spectacl ...	positive
807	plot imageri charact emot sceneri everyth ther...	positive
413	sorri rip dream apart task get everybodi back ...	neutral
1025	ive use imdb sinc havent taken time review mov...	positive
798	movi particularli gorgeou excit give key momen...	positive
...	...	...
55	movi wasnt hour long say hour id probabl give ...	neutral
120	act cinematographi art direct costum sound mus...	positive
1008	first review imdb think movi worth itjust one ...	positive
25	couldnt see christoph walken embodi charact sa...	positive
72	feel like film back back first dune caus look ...	positive

Fig. 7 Splitting Data Testing Result

GridSreachCV

GridSearchCV is a function that allows for finding the best combination of parameters for each algorithm to

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4606

Submitted : Sep 09, 2024
Accepted : Oct 03, 2024
Published : Oct 11, 2024

minimize the error produced by the algorithm's predictions (Yoshua et al., 2021). In this study, GridSearchCV uses cross-validation to evaluate each combination of parameters. Here, the researcher divides the dataset into five subsets or folds. Cross-validation is used to evaluate the performance of the sentiment classification model.

```
# Find best parameters using GridsearchCV
param_grid = {'alpha': [0.01, 0.1, 0.5, 1.0]}
nb_model - MultinomialNB()
grid_search = GridSearch CV (nb_model, param_grid, cv-5, scoring- accuracy")
grid_search.fit(x_train, y_train)
```

Fig. 8 GridSearchCV Process

There is the result from GridSearchCV using Python programming language.

```
GridSearchCV Results:
Mean: 0.943 Std: 0.012 with: {'alpha': 0.01}
Mean: 0.937 Std: 0.011 with: {'alpha': 0.1}
Mean: 0.918 Std: 0.014 with: {'alpha': 0.5}
Mean: 0.902 Std: 0.020 with: {'alpha': 1.0}

Accuracy: 0.9460966542750929
```

Fig. 9 GridSearchCV Result

The evaluation's findings show that this model's accuracy rate was 94.60%.

Classification

Classification is the process of identifying a set of patterns or functions that describe and distinguish one data class from another. It is used to predict the class of data that has not yet been categorized (Setio et al., 2020). During the data classification phase, the Naïve Bayes model is trained on the training data, learning the relationships between the data features and their corresponding labels. The prediction outcomes are then assessed using metrics such as accuracy, classification report, and confusion matrix.

Following that, measures like accuracy, classification report, and confusion matrix are used to evaluate the prediction results. These assessments aid in comprehending the efficacy of the sentiment classification process as well as the model's performance. The process of classification can be seen in the following syntax code:

```
# Predict and evaluate the model
y_pred=nb_model.predict(X_test)

print("\nAccuracy:", accuracy_score(y_test, y_pred))
print("\nClassification Report: \n", classification_report (y_test, y_pred))
print("\nConfusion Matrix: \n", confusion matrix (y_test, y_pred))
```

Fig. 10 Clasiffication Process

There is the result of classification can be seen in below:

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4606

Submitted : Sep 09, 2024
Accepted : Oct 03, 2024
Published : Oct 11, 2024

```
Accuracy: 0.9460966542750929

Classification Report:
              precision    recall  f1-score   support

    negative       0.99      0.94      0.97       174
     neutral       0.87      0.99      0.93       177
    positive       0.99      0.90      0.94       187

    accuracy                           0.95       538
   macro avg       0.95      0.95      0.95       538
weighted avg       0.95      0.95      0.95       538


Confusion Matrix:
 [[164   9   1]
 [  0 176   1]
 [  1  17 169]]
```

Fig. 11 Clasiffication Result

Based on the classification testing results, it shows that the trained model is able to classify test data with an accuracy, precision, and recall of 95%. With such a high accuracy rate, it demonstrates that the Naïve Bayes model is capable of identifying positive, negative, and neutral sentiments in the "Dune: Part Two" movie reviews. A visualization of the above results can be seen in the image below.

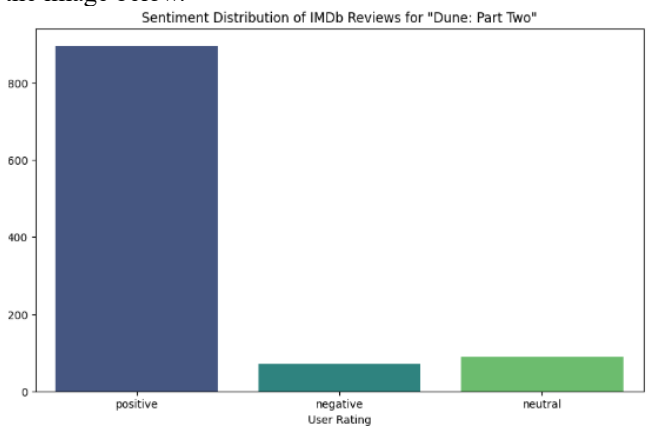

Fig. 12 Visualization of classification results

DISCUSSIONS

This study uses a dataset of reviews for the film "Dune: Part Two" from the IMDb Movie Reviews website. The data from this dataset is processed through several stages such as preprocessing, feature selection, splitting data, data mining, and evaluation. In this study, the researcher used the Naïve Bayes algorithm to classify the sentiment of "Dune: Part Two" movie reviews. The results of the above tests show that the Naïve Bayes method has successfully achieved a very high accuracy rate of 95%. This indicates that Naïve Bayes is capable of identifying positive, negative, and neutral sentiments in "Dune: Part Two" movie reviews with good precision.

Besides the success of the Naïve Bayes method in this study, the use of the TF-IDF method in feature selection also allows for focusing on the most important words in a review, thereby enhancing its ability to classify sentiment.

CONCLUSION

Based on the sentiment analysis of "Dune: Part Two" movie reviews using the Naïve Bayes method, it is clear that this approach is effective in classifying sentiment, achieving accuracy, precision, and recall rates of 95%. The results suggest that Naïve Bayes is a reliable method for analyzing text data and identifying sentiment patterns in movie reviews. The insights from this analysis are valuable to the film industry, helping to understand audience preferences and improve film production quality in the future. For future research, it is recommended to explore other methods, such as Support Vector Machines (SVM) or neural networks, to potentially achieve higher accuracy levels.

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4606

Submitted : Sep 09, 2024
Accepted : Oct 03, 2024
Published : Oct 11, 2024

Additionally, using larger and more diverse datasets, combined with advancements in natural language processing techniques, could further enhance sentiment classification results and provide deeper insights.

REFERENCES

Afriansyah, M., Saputra, J., Sa'adati, Y., & Valian Yoga Pudya Ardhana. (2023). Optimasi Algoritma Nai?ve Bayes Untuk Klasifikasi Buah Apel Berdasarkan Fitur Warna RGB. *Bulletin of Computer Science Research*, *3*(3), 242–249. https://doi.org/10.47065/bulletincsr.v3i3.251

Anuar, F., Putra, R., Firman Fadilah, F., Enri, U., & Karawang, U. S. (2023). *ANALISIS SENTIMEN ULASAN FILM OPPENHEIMER PADA SITUS IMDB MENGGUNAKAN METODE NAIVE BAYES*.

Ferryawan, R., Kusrini, K., & Wibowo, F. W. (2020). Analisis Sentimen Wisata Jawa Tengah Menggunakan Naïve Bayes. *Jurnal Informa : Jurnal Penelitian Dan Pengabdian Masyarakat*, *5*(3), 55–60. https://doi.org/10.46808/informa.v5i3.146

Guterres, A., Gunawan, & Santoso, J. (2019). Stemming Bahasa Tetun Menggunakan Pendekatan Rule Based. *Teknika*, *8*(2), 142–147. https://doi.org/10.34148/teknika.v8i2.224

Kelvin, K., Banjarnahor, J., -, E. I., & NK Nababan, M. (2022). Analisis perbandingan sentimen Corona Virus Disease-2019 (Covid19) pada Twitter Menggunakan Metode Logistic Regression Dan Support Vector Machine (SVM). *Jurnal Sistem Informasi Dan Ilmu Komputer Prima(JUSIKOM PRIMA)*, *5*(2), 47–52. https://doi.org/10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2365

Khotimah, A. C., & Utami, E. (2022). COMPARISON NAÏVE BAYES CLASSIFIER, K-NEAREST NEIGHBOR AND SUPPORT VECTOR MACHINE IN THE CLASSIFICATION OF INDIVIDUAL ON TWITTER ACCOUNT. *Jurnal Teknik Informatika (JUTIF)*, *3*(3). https://doi.org/10.20884/1.jutif.2022.3.3.254

Kritikus Nilai Dune: Part Two 95 Persen, Rekor Baru Denis Villeneuve. (2024). https://www.cnnindonesia.com/hiburan/20240301191201-220-1069435/kritikus-nilai-dune-part-two-95-persen-rekor-baru-denis-villeneuve

Kurniawan, A., & Adinugroho, S. (2019). *Analisis Sentimen Opini Film Menggunakan Metode Naïve Bayes dan Lexicon Based Features* (Vol. 3, Issue 9). http://j-ptiik.ub.ac.id

Made, I., Adnyana, B., Jln, S. B., & Puputan, R. (2019). *Penerapan Feature Selection untuk Prediksi Lama Studi Mahasiswa*.

Makmun, M., Zy, A. T., & Arwan, A. (2024). *Analisis Sentimen Media Sosial Twitter Terhadap Calon Presiden RI Tahun 2024 Menggunakan Klasifikasi Algoritma Naïve Bayes*. https://doi.org/10.47065/josyc.v5i3.5210

Muhammad Fernanda Naufal Fathoni, Eva Yulia Puspaningrum, & Andreas Nugroho Sihananto. (2024). Perbandingan Performa Labeling Lexicon InSet dan VADER pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM. *Modem : Jurnal Informatika Dan Sains Teknologi.*, *2*(3), 62–76. https://doi.org/10.62951/modem.v2i3.112

Muhammad Thaariq, Dade Nurjanah, & Hani Nurrahmi. (2023). *Analisis Sentimen Review Film Menggunakan Naive Bayes Classifier Dengan Fitur TF-IDF*.

Nur Rozi, F., & Harini Sulistyawati, D. (2019). *KLASIFIKASI BERITA HOAX PILPRES MENGGUNAKAN METODE MODIFIED K-NEAREST NEIGHBOR DAN PEMBOBOTAN MENGGUNAKAN TF-IDF* (Vol. 15, Issue 1).

Putra, F., Tahiyat, H. F., Ihsan, R. M., Rahmaddeni, R., & Efrizoni, L. (2024). Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(1), 273–281. https://doi.org/10.57152/malcom.v4i1.1085

Rizkina, N. Q., & Hasan, F. N. (2023). Analisis Sentimen Komentar Netizen Terhadap Pembubaran Konser NCT 127 Menggunakan Metode Naive Bayes. *Journal of Information System Research (JOSH)*, *4*(4), 1136–1144. https://doi.org/10.47065/josh.v4i4.3803

Salam, A., Zeniarja, J., Septiyan, R., & Khasanah, U. (2018). *ANALISIS SENTIMEN DATA KOMENTAR*

* Corresponding author

Journal of Computer Networks, Architecture and High Performance Computing
Volume 6, Number 4, October 2024
https://doi.org/10.47709/cnahpc.v6i4.4606

Submitted : Sep 09, 2024
Accepted : Oct 03, 2024
Published : Oct 11, 2024

SOSIAL MEDIA FACEBOOK DENGAN K-NEAREST NEIGHBOR (STUDI KASUS PADA AKUN JASA EKSPEDISI BARANG J&T EKSPRESS INDONESIA).

Setio, P. B. N., Saputro, D. R. S., & Winarno, B. (2020). *PRISMA, Prosiding Seminar Nasional Matematika Klasifikasi dengan Pohon Keputusan Berbasis Algoritme C4.5*. *3*, 64–71. https://journal.unnes.ac.id/sju/index.php/prisma/

Suryati, E., Ari Aldino, A., Penulis Korespondensi, N., & Suryati Submitted, E. (2023). *Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM)*. *4*(1), 96–106. https://doi.org/10.33365/jtsi.v4i1.2445

Turmudi Zy, A., Nugroho, A., Rivaldi, A., & Afriantoro, I. (2022). Analisis Sentimen Terhadap Pembobolan Data pada Twitter dengan Algoritma Naive Bayes. *Jurnal Teknologi Informatika Dan Komputer*, *8*(2), 202–213. https://doi.org/10.37012/jtik.v8i2.1240

Yoshua, I., Bunyamin, H., & Si, S. (2021). *Pengimplementasian Sistem Rekomendasi Musik Dengan Metode Collaborative Filtering* (Vol. 3).

Yutika, C. H., Adiwijaya, A., & Faraby, S. Al. (2021). Analisis Sentimen Berbasis Aspek pada Review Female Daily Menggunakan TF-IDF dan Naïve Bayes. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *5*(2), 422. https://doi.org/10.30865/mib.v5i2.2845

* Corresponding author