Journal of Student Research Exploration

https://shmpublisher.com/index.php/josre

p-ISSN 2964-1691 | e-ISSN 2964-8246

The Comparison of Statement Analysis on Disney Case Using Naive Bayes, SNM, and Logistic Algorithm Methods

Rizkiyanti Choirunnisa[1], Budi Prasetyo[2]

[12] Information System, Faculty of Mathematics and Science, Universitas Negeri Semarang

Article Info

Article history:

Received July 8, 2024
Revised April 2, 2026
Accepted April 13, 2026

Keywords:

Sentiment
Support vector machine
Naïve bayes
Disney
Logistic algorithm

ABSTRACT

The Walt Disney Company or also known as Disney, is one of the most famous companies in the world that focuses on the production of animation and film. Disney has been served in this entertainment industry for over 90 years. Since Disney's first film was released, Disney was become very famous until this day, especially when Disney has collaborated with many companies, it's not only focusing on animation production but also making films and many live-action versions of the animations. Recently, Disney has been a hot topic among Disney's movie fans due to the selection of actors for live action movie characters. Therefore, on this time, the author will conduct a sentiment analysis of Disney using analysis methods called Naïve Bayes, Support Vector Machine, and Logistic Algorithm. After passing through all the testing stages, the highest accuracy result is analysis using the Support Vector Machine (SVM) method. The test results show a high accuracy of 0.60 or 60% with the highest F1-Score in the Positive sentiment class by 67%.

1. Introduction

The Walt Disney Company or also known as Disney, is one of the most famous companies in the world that focused on the production of animation and film. Disney was founded on October 16, 1923, by Walt Disney and Roy Oliver Disney. Disney first named is Disney Brothers, the company eventually changed name to The Walt Disney Company in 1986, this was known through Gabler's book on Walt Disney states, "when the company had been formed back in 1923, it was a partnership between Walt and Roy" [1]. Disney has been served in this entertainment industry for over than 90 years, starting with the first film animation

[1] ***Corresponding Author:***
Rizkiyanti Choirunnisa,
Information System, Faculty of Mathematics and Science, Universitas Negeri Semarang
Kampus Pusat: Kampus Sekaran Gunungpati Semarang, Indonesia. 50229
Email: rizkiyantichoirunnisa@students.unnes.ac.id
DOI: https://doi.org/10.52465/josre.v4i1.388

titled "Snow White and The Seven Dwarfs" which is released on December 27, 1937. This first Disney animation was very success at that time when Disney was still named Disney Brothers [2]. After the first animation of its film, Disney became famous until right now. Additionally, Disney has collaborated with many companies, thus not only focusing on animation production but also making films and live-action adaptations of its animations.

Recently, Disney has been a hot topic among Disney's movie fans due to the selection of actors to portray its characters. Since 2019, Disney has focused on enhancing VAR (Virtual Augmented Reality) technology, using CGI to create a more realistic atmosphere, starting with the hugely successful animation titled "The Lion King" released On September 10, 2023, then Disney announced the release of a live-action film of the animation "The Little Mermaid" and announced that Halle Bailey as the main character, Ariel. This announcement provokes various reactions from fans, many fans welcoming the announce, while the others disagreed with the casting of Halle Bailey as Ariel.

Therefore, on this occasion, the author will conduct a sentiment analysis of Disney using methods called Naïve Bayes, Support Vector Machine, and Logistic Algorithm. "Analisis sentiment adalah salah satu cara yang tepat untuk mengetahui dan mengukur suatu reputasi merek dan consumer behavior" [3]. In business, sentiment is important to understand consumer emotions or sentiments towards a brand. Knowing this can help in planning and strategizing for the brand's sustainability. "Sentiment yang digunakan untuk analisis berasal dari *platform* twitter yang merupakan salah satu sumber data *real time* dunia nyata tentang perilaku yang mungkin mewakili konsumen dan lebih neturalistik dalam volume besar" [4].

The main method that used in this sentiment analysis is text mining with an approach to public sentiment, or as we know as netizen sentiment. "Pendekatan analisis sentiment bertujuan untuk memahami bagaimana orang menggambarkan dan mengungkapkan perasaan mereka terhadap Disney baik tanggapan secara positif, negative dan positif [5]". The main methods in sentiment analysis, as explained by Felix et al., are supported by Akbar's explanation, stating that " Text mining dapat didefinisikan secara luas sebagai suatu proses menggali informasi Dimana seorang user berinteraksi dengan sekumpulan dokumen menggunakan tools analisis yang merupakan komponen dalam data mining yang salah satunya adalah kategorisasi" [6].

Besides using the primary method of text mining for data collection, several other classification calculation methods are used, such as the Naïve Bayes algorithm, Support Vector Machine (SVM), and Logistic Algorithm. The Naïve Bayes method is usually used for binary and multiclass classification, applying supervised object classification techniques for future use. This method assigns class labels to instances using conditional probabilities. "Pada dasarnya metode naïve bayes merupakan suatu metode klasifikasi yang berakar pada teorima bayes dengan melakukan pengklasifikasian menggunakan metode probabilitas dan statistic yang memprediksi peluang masa depan berdasarkan pengalaman di masa sebelumnya" [7].

"Metode klasifikasi Support Vectore Machine atau biasa disebut SVM ini merupakan sebuah metode dalam supervised learning yang biasanya digunakan untuk klasifikasi dan regresi" [8]. "Dalam pemodelan klasifikasi, SVM memiliki konsep yang lebih matang dan lebih jelas secara

matematid dibandingkan dengan Teknik klasifikasi lainnya" [8]. "Sedangkan metode Logistic Algorithm adalah suatu jenis regresi analisis yang digunakan untuk menjelaskan hubungan antara variable depeden dan independent untuk menghubungkan satu atau lebih variable bebas demgam variable jenis terkait" [9]. This analysis aims to provide a clear picture of netizen comments or feedback, allowing Disney to evaluate and make informed decisions in the future.

2. Method

The method used for sentiment analysis in this study includes Naïve Bayes, SVM, and logistic algorithm, with an approach using text mining methods. As explained in the background, the Naïve Bayes method is a technique that can be used to predict the future based on current observations. The SVM method is highly popular for solving classification problems, while the logistic algorithm is one of the frequently used methods for sentiment analysis. These three algorithms are classification algorithms that fundamentally utilize a pattern to determine and classify data. "Machine learning merupakan ssuatu teknologi yang memungkinkan untuk membuat Keputusan secara otomatis menggunakan cara berpikir manusia" [10].

In sentiment analysis, which falls under data mining, the presence of machine learning is crucial because machine learning is trained to make decisions in the data processing process. The data used in this analysis is obtained from Twitter using the data crawling method. "Crawling data merupakan sebuah proses yang dilakukan oleh sebuah computer untuk mengumpulkan data dari berbagai sumber" [11]. This data is collected using the keyword 'Disney' with a limit of 500 data. The total data obtained amounts to 644 records without undergoing a cleaning process.

3. Results and Discussion

In conducting this sentiment analysis, several stages must be passed, starting from data crawling to obtain samples to achieving the highest accuracy using the data. Data crawling is essential to understand public responses to Disney from September 2023 to May 2024. The collected data amounted to 644 records, which had not undergone cleaning or preprocessing. After preprocessing, the data was reduced to 398 records. Below is an overview of the testing stages of sentiment analysis using the Naïve Bayes and Logistic Algorithm methods.

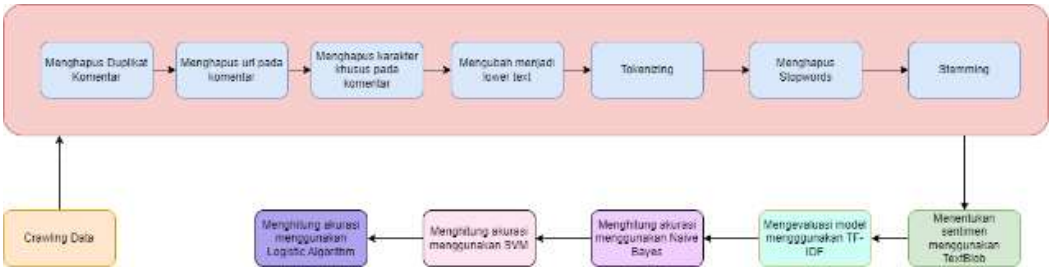

Figure 1. Testing stage

The initial step in this sentiment analysis testing is a crucial stage called preprocessing. “Tahap preprocessing ini merupakan suatu tahap untuk membuat data agar siap untuk masuk ke tahap selanjutnya. Data preprocessing atau praproses data biasanya dilakukan melalui cara eliminasi data yang tidak sesuai” [12]. This stage includes cleaning, converting text to lowercase, tokenizing, removing stopwords, and stemming. These steps are used to ensure that the data produced after preprocessing is more organized than before. The difference can be seen in the tables below.

Table 1. Sentiment data before pre-processing

No	Full text	Username
1.	So this is Disney plus https://t.co/RdlQDnLMss	EverythingOOC
2.	She thought she was in a Disney movie https://t.co/WSkPrUxcE3	cctvidiots
3.	EPSTEIN CLIENT LI…..	ShaykhSulaiman
4.	The kid was living Disney life. Woke up to reality htttps//….	cctvidiots
5.	Disney just announced Holotile floor…	heyBarsee
6.	disney on january 1st 2024 https://t.co/IHbmabl7Rb	InternetH0F
7.	So this is Disney plus https://t.co/RdlQDnLMss	EverythingOOC

Table 2. Sentiment data after pre-processing

No	Full text	Username	Clean text
1.	So this is Disney plus https://t.co/RdlQDnLMss	EverythingOOC	disney plus
2.	She thought she was in a Disney movie https://t.co/WSkPrUxcE3	cctvidiots	thought disney movie
3.	EPSTEIN CLIENT LI…..	ShaykhSulaiman	epstein client…
4.	The kid was living Disney life. Woke up to reality htttps//….	cctvidiots	kid live disney life woke realiti
5.	Disney just announced Holotile floor…	heyBarsee	disney announc holotil
6.	disney on january 1st 2024 https://t.co/IHbmabl7Rb	InternetH0F	disney januari 1st 2024
7.	So this is Disney plus https://t.co/RdlQDnLMss	EverythingOOC	disney movi come true

After the data passed through the preprocessing stage, the next step is to determine sentiment. Testing for sentiment analysis in this study uses Python tools, utilizing TextBlob to determine whether text falls into positive, negative, or neutral sentiment categories. Textblob is an open-source package provided by Python, useful for processing text data. The sentiment categories used in this analysis categorize sentiment as negative if the polarity is less than 0, positive if the polarity is greater than 0, and neutral if the polarity equals 0. Polarity refers to a reference value returned to the data object to indicate a level of positivity or negativity in a text. Below is the data that has undergone sentiment analysis using TextBlob.

Table 3. Sentiment at each text

No	full_text	username	sentiment
1.	So this is Disney plus https://t.co/RdlQDnLMss	EverythingOOC	neutral
2.	She thought she was in a Disney movie https://t.co/WSkPrUxcE3	cctvidiots	neutral
3.	EPSTEIN CLIENT LI…..	ShaykhSulaiman	neutral
4.	The kid was living Disney life. Woke up to reality htttps//….	cctvidiots	neutral
5.	Disney just announced Holotile floor…	heyBarsee	neutral
6.	disney on january 1st 2024 https://t.co/IHbmabI7Rb	InternetH0F	neutral
7.	So this is Disney plus https://t.co/RdlQDnLMss	EverythingOOC	positive

Sentiment analysis testing does not stop when data is assigned sentiment; rather, the data must undergo model evaluation using TF-IDF Vectorizer and tested with Naïve Bayes, SVM, and Logistic Algorithm methods. Term Frequency Inverse Document Frequency (TF-IDF) is a technique used to determine the relative frequency of a word compared to the proportion of that word in all documents [13]. This method has been proven as a robust technique in text data processing or data mining. After calculating accuracy in the model evaluation, the next step is to calculate accuracy using the Naïve Bayes, SVM, and Logistic Algorithm models. The data used in testing the model accuracy uses a 90:10 ratio comparison, with the highest accuracy seen in the SVM model with a total accuracy of 0.6. Meanwhile, the accuracy using other models, namely Naïve Bayes and Logistic Algorithm, only reached 0.55. The results of the calculations can be seen in the following tables.

Table 4. Evaluation result of naïve bayes model

NB Accuracy : 0.55				
Classification Report	Precision	Recall	F1-Score	support
negative	0.00	0.00	0.00	9
neutral	0.83	0.36	0.50	14
positive	0.50	1.00	0.67	17
accuracy			0.55	40
Macro avg	0.44	0.45	0.39	40
Weighted avg	0.59	0.55	0.46	40

In the table of Naïve Bayes model evaluation results, the accuracy value indicates that 55% of the classified data is correct. The precision evaluation results are as follows:

1. Precision of the negative class classification has a value of 0.00, indicating that none of the data classified as the negative class is correct. Meanwhile, the recall of the negative class indicates a accuracy value of 0, meaning 0% of the data correctly belongs to the negative class. The F1-score accuracy of the negative class indicates a value of 0.00, showing that the performance of Naïve Bayes in classifying the negative class is poor.
2. Precision of the neutral class classification has a value of 0.83, indicating that all data classified as the neutral class is correct. Meanwhile, the recall of the neutral class

indicates an accuracy value of 0.50, meaning only 50% of the data correctly belongs to the neutral class. The F1-score accuracy of the neutral class indicates a value of 0.65, showing that the performance of SVM in classifying the neutral class is moderate.

3. Precision of the positive class classification has a value of 0.50, indicating that all data classified as the positive class is correct. Meanwhile, the recall of the positive class indicates an accuracy value of 1.00, meaning 100% of the data correctly belongs to the positive class. The F1-score accuracy of the positive class indicates a value of 0.67, showing that the performance of SVM in classifying the positive class is moderate.

Table 5. Evaluation result of SVM model

SVM Accuracy : 0.55				
Classification Report	Precision	Recall	F1-Score	support
negative	1.00	0.11	0.20	9
neutral	0.50	0.54	0.62	14
positive	0.58	0.82	0.68	17
accuracy			0.60	40
Macro avg	0.73	0.53	0.50	40
Weighted avg	0.68	0.60	0.55	40

In the table of SVM model evaluation results, the accuracy value indicates that 60% of the classified data is correct. The precision evaluation results are as follows:

1. Precision of the negative class classification has a value of 1, indicating that all data classified as the negative class is correct. Meanwhile, the recall of the negative class indicates an accuracy value of 0.11, meaning only 11% of the data correctly belongs to the negative class. The F1-score accuracy of the negative class indicates a value of 0.20, showing that the performance of SVM in classifying the negative class is low.
2. Precision of the neutral class classification has a value of 0.60, indicating that all data classified as the neutral class is correct. Meanwhile, the recall of the neutral class indicates an accuracy value of 0.64, meaning 64% of the data correctly belongs to the neutral class. The F1-score accuracy of the neutral class indicates a value of 0.62, showing that the performance of SVM in classifying the neutral class is moderate.
3. Precision of the positive class classification has a value of 0.58, indicating that all data classified as the positive class is correct. Meanwhile, the recall of the positive class indicates an accuracy value of 0.82, meaning 82% of the data correctly belongs to the positive class. The F1-score accuracy of the positive class indicates a value of 0.68, showing that the performance of SVM in classifying the positive class is good.

Table 6. Evaluation result of logistic regression model

LR Accuracy : 0.55				
Classification Report	Precision	Recall	F1-Score	support
negative	0.00	0.00	0.00	9
neutral	0.62	0.36	0.45	14
positive	0.53	1.00	0.69	17
accuracy			0.55	40
Macro avg	0.39	0.45	0.38	40
Weighted avg	0.44	0.55	0.45	40

In the table of Logistic Regression model evaluation results, the accuracy value indicates that 55% of the classified data is correct. The precision evaluation results are as follows:

1. Precision of the negative class classification has a value of 0.00, indicating that none of the data classified as the negative class is correct. Meanwhile, the recall of the negative class indicates an accuracy value of 0, meaning 0% of the data correctly belongs to the negative class. The F1-score accuracy of the negative class indicates a value of 0.00, showing that the performance of Naïve Bayes in classifying the negative class is poor.
2. Precision of the neutral class classification has a value of 0.62, indicating that all data classified as the neutral class is correct. Meanwhile, the recall of the neutral class indicates an accuracy value of 0.36, meaning only 36% of the data correctly belongs to the neutral class. The F1-score accuracy of the neutral class indicates a value of 0.45, showing that the performance of SVM in classifying the neutral class is moderate.
3. Precision of the positive class classification has a value of 0.53, indicating that all data classified as the positive class is correct. Meanwhile, the recall of the positive class indicates an accuracy value of 1.00, meaning 100% of the data correctly belongs to the positive class. The F1-score accuracy of the positive class indicates a value of 0.

4. Conclusion

After conducting sentiment analysis testing using three methods: Naïve Bayes, Support Vector Machine (SVM), and Logistic Algorithm, the results show that the highest accuracy among these three methods is achieved by the SVM method, with an accuracy value of 0.60 using a 90:10 data split ratio. The accuracy value of 0.60 indicates that the SVM classification model is capable of predicting sentiment analysis on Disney quite well, with the best accuracy in the positive comment category with an F1-score value of 0.67 or 67%, indicating that the SVM model in classifying positive comments is moderate but the highest compared to negative and neutral categories. The SVM method is known to be very powerful but sensitive to noise in data, which explains why the SVM model has difficulty in classifying neutral and positive categories. Additionally, the support for negative comments in this data is the least compared to other categories, making it difficult for the SVM model to learn the text patterns of negative comments. “Noise data sendiri merupakan jenis data yang mengandung suatu eror yaitu nilai atribut yang tidka benar dan menyimpang dibandingkan nilai lainnya” [14]. For further testing, it may be better if the support for negative, neutral, and positive classes is equal, which would improve the accuracy of the chosen classification model.

REFERENCES

[1] N. Gabler, *Walt Disney: The Triumph of the American Imagination*. New York: Alfred A. Knopf, 2006.

[2] A. Taharudin, “Wajib Tahu, Ini Dia 5 Film Pertama yang Diproduksi Walt Disney Studios!” Accessed: Apr. 13, 2026. [Online]. Available: https://www.ggwp.id/entertainment/movie/film-animasi-pertama-walt-disney

[3] F. Himma, “Analisis Sentimen adalah: Pengertian, Contoh, Tipe.” Accessed: Apr. 13, 2026. [Online]. Available: https://majoo.id/solusi/detail/analisis-sentimen-adalah

[4] S. Yoon, N. Elhadad, and S. Bakken, “A Practical Approach for Content Mining of Tweets,” *Am. J. Prev. Med.*, vol. 45, no. 1, pp. 122–129, Jul. 2013, doi: 10.1016/j.amepre.2013.02.025.

[5] F. F. Mailo, L. Lazuardi, and D. Manajemen, “Analisis sentimen data twitter menggunakan metode text mining tentang masalah obesitas di indonesia,” *Journal of Information Systems for Public Health*, vol. 6, no. 1, p. 44, 2021.

[6] Z. I. Akbar, “Apa itu Text Mining?” Accessed: Apr. 13, 2026. [Online]. Available: https://sis.binus.ac.id/2021/04/23/apa-itu-text-mining/

[7] M. H. Widianto, “Algoritma Naive Bayes.” Accessed: Apr. 13, 2026. [Online]. Available: https://binus.ac.id/bandung/2019/12/algoritma-naive-bayes/

[8] Samsudey, “Penjelasan Sederhana tentang Apa Itu SVM?” Accessed: Apr. 13, 2026. [Online]. Available: https://medium.com/@samsudiney/penjelasan-sederhana-tentang-apa-itu-svm-149fec72bd02

[9] F. D. Pramakrisna, F. D. Adhinata, and N. A. F. Tanjung, “Aplikasi Klasifikasi SMS Berbasis Web Menggunakan Algoritma Logistic Regression,” *Teknika*, vol. 11, no. 2, pp. 90–97, Jun. 2022, doi: 10.34148/teknika.v11i2.466.

[10] D. Feby, “Pengertian Machine Learning Model & Implementasinya.” Accessed: Apr. 13, 2026. [Online]. Available: : https://dqlab.id/pengertian-machine-learning-model-and-implementasinya

[11] Team, “Apa Itu Crawling Data? Pengertian, Proses dan Caranya.” Accessed: Apr. 13, 2026. [Online]. Available: https://codingstudio.id/blog/crawling-data-adalah/

[12] R. N. R. Suripto and A. S. Kirana, “Teknik pre-processing dan classification dalam data science,” *Student corner, Aug*, vol. 26, 2022.

[13] Harishamzah, “Perbandingan Perhitungan Bobot TF-IDF secara Manual dan Menggunakan Python.” Accessed: Apr. 13, 2026. [Online]. Available: https://medium.com/bisa-ai/perbandingan-perhitungan-bobot-tf-idf-secara-manual-dan-menggunakan-python-377392a165c6

[14] “DATA BIAS, NOISE, & OUTLIER,” Data and Analytics. Accessed: Apr. 13, 2026. [Online]. Available: https://medium.com/@AI_and_DataAnalytics/data-bias-noise-outlier-903ebb8f6b65