JCBD
JOURNAL OF COMPUTERS AND DIGITAL BUSINESS
Volume 5, Number 1, January 2026, pp. 23-30
Homepage: https://jurnal.delitekno.co.id/index.php/jcbd

Classification of Korean Drama Popularity Based on Ratings Using Naïve Bayes

Afthar Kautsar

Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

jjtaryoung21@gmail.com

https://doi.org/10.56427/jcbd.v5i1.814

ARTICLE INFO

Article History
Received: November 30, 2025
Revised: January 22, 2026
Accepted: January 25, 2026

Keywords
Korean Drama
Classification
Naïve Bayes
Rating
Popularity

ABSTRACT

This study aims to classify the popularity of Korean dramas based on ratings obtained from the MyDramaList website. With the rapid growth of digital entertainment platforms, evaluating drama popularity has become increasingly important for understanding audience preferences and supporting decision-making in the content industry. The Naive Bayes algorithm is employed as the classification method due to its computational efficiency and suitability for handling categorical and numerical features. The dataset comprises 351 Korean dramas with attributes including title, year of release, genre, tags, number of episodes, cast information, synopsis, and user ratings. Ratings serve as the primary label for categorizing dramas into three classes: Top Dramas (rating ≥ 8.5), Popular (7.5–8.4), and Less Popular (< 7.5). The classification pipeline involves data preprocessing, feature encoding, and model training using Naive Bayes. Evaluation results yield an overall accuracy of 79%, with per-class performance assessed through precision, recall, and F1-score metrics. Supplementary visualizations, including pie charts, bar charts, and word clouds, are employed to analyze the distribution of dominant genres and tags across popularity categories. The findings indicate that the proposed approach provides a viable baseline for drama popularity classification while revealing content patterns, such as the prevalence of specific genres and thematic tags among top-rated dramas, that may inform content curation strategies on digital platforms.

1. Introduction

Korean dramas have become a global phenomenon, attracting millions of viewers worldwide. The abundance of titles on digital platforms makes audiences often rely on rating scores as indicators of drama quality. However, ratings alone sometimes fail to fully capture viewer satisfaction, especially when multiple dramas receive similar scores. Therefore, rating-based classification becomes a relevant approach to categorize dramas into popular, average, or less popular groups.

The Naive Bayes algorithm has been widely employed due to its simplicity, efficiency, and reliable performance with both textual and numerical data. For instance, Junianto et al. successfully used Naive Bayes to analyze public opinions on documentary films, demonstrating the method's speed in processing data [1]. Sooai and Laniwati also highlighted the ability of Naive Bayes to extract sentiment from film reviews on Twitter, particularly within fan communities [2]. Additionally, Rahmawaty et al. combined Naive Bayes with Cosine Similarity to detect public opinions on illegal film platforms on social media, showcasing the algorithm's flexibility [3].

Several studies compared Naive Bayes with other algorithms to evaluate its performance. Zulkarnain et al. examined Naive Bayes alongside K-Nearest Neighbor and Decision Tree for Netflix rating classification, finding that Naive Bayes remained competitive [4]. Putri et al. applied the algorithm to classify IMDb reviews, achieving satisfactory accuracy [5]. Ramadhan et al. compared Naive Bayes with Decision Tree on IMDb movie ratings, confirming the method's efficiency in sentiment prediction [6].

Other research emphasized practical applications on digital platforms. Rieuwpassa et al. implemented Naive Bayes to classify Netflix user reviews on Google Play, producing stable results [7]. Nurian and Sari analyzed Google Play application reviews, demonstrating Naive Bayes' ability to handle real user opinions [8]. Toyibah et al. compared Multinomial Naive Bayes with

Logistic Regression for IMDb ratings, emphasizing Naive Bayes' computational efficiency [9]. Aprilia and Lestari employed the method to analyze Korean drama sentiment on social media, proving its suitability for Indonesian-language text data [10].

In other online entertainment contexts, Meisty et al. compared Naive Bayes with Random Forest for Korean drama rating prediction, finding that while Random Forest achieved slightly higher accuracy, Naive Bayes excelled in efficiency [11]. Nurtikasari et al. emphasized Naive Bayes' effectiveness in analyzing public opinions on Twitter about films, confirming its relevance for social media data [12]. Alyandi et al. applied Naive Bayes to reviews of the Honor of Kings game on Playstore, generating a balanced classification between positive and negative sentiment [13]. Nurwanda and Rizkiani compared Naive Bayes with SVM on Twitter lifestyle reviews, showing that Naive Bayes remained competitive [14]. Finally, Amrullah et al. incorporated Chi-Square feature selection to enhance the accuracy of Naive Bayes for movie review sentiment analysis [15].

Based on these findings, the Naive Bayes algorithm proves to be a suitable method for classifying Korean drama popularity based on viewer ratings and for efficiently analyzing sentiment from digital reviews.

2. Research Methodology

This study adopts a quantitative approach using supervised learning classification with the Naive Bayes algorithm. The research procedure consists of several main stages, including data collection, labeling, data preprocessing, model implementation, and evaluation.

Data Collection

The dataset was collected from the MyDramaList website and consists of 351 Korean dramas. Each data entry includes several attributes, namely title, release year, number of episodes, genre, tags, description, actors, and rating.

Popularity Category Labeling

Drama popularity categories were determined based on rating values and divided into three classes:

1. Top Dramas: rating ≥ 8.5
2. Popular: 7.5 ≤ rating < 8.4
3. Less Popular: rating < 7.5

These categories were used as class labels in the classification process.

Data Preprocessing

The preprocessing stage was carried out to improve data quality and prepare the dataset for modeling. This stage includes:

1. Removal of missing and duplicate data
2. Encoding genre and tag attributes into numerical values
3. Mapping target classes into numerical labels (0, 1, and 2)

Naive Bayes Implementation

The Naive Bayes algorithm was applied to learn the distribution patterns of features toward the target classes. The dataset was divided into 80% training data and 20% testing data. The model was trained to classify Korean dramas based on numerical features, while rating values were used solely as the basis for forming class labels.

Evaluation and Visualization

Model performance was evaluated using classification metrics, including precision, recall, and f1-score. Prediction results were visualized using a confusion matrix to compare predicted and actual classes. In addition, word cloud visualization was employed to illustrate the frequency of dominant genres and tags. Pie charts and bar charts were also used to analyze the distribution of popularity categories and the most common genres.

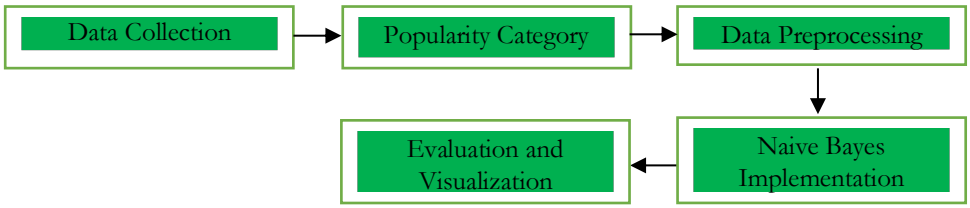

Figure 1. Methodology

3. Results and Discussion

This subsection discusses the results of the classification model applied to Korean drama data. The evaluation focuses on assessing the model's ability to distinguish popularity categories using *precision*, *recall*, *f1-score*, and *accuracy*. In addition, a *confusion matrix* is used to provide a clearer comparison between the model's predictions and the actual data.

Distribution of Popularity Categories

The initial stage of this study involved labeling the data based on drama ratings. The labeling criteria were defined as follows: dramas with a rating of ≥ 8.5 were classified as Top-Rated Dramas, dramas with ratings were categorized as Popular

Dramas, and dramas with ratings of < 7.5 were classified as Less Popular Dramas. Based on these criteria, the distribution of popularity categories from a total of 351 Korean drama records is as follows:

1. Top-Rated Dramas: 159 dramas
2. Popular Dramas: 169 dramas
3. Less Popular Dramas: 22 dramas

Table 1. Distribution of Korean Drama Popularity Categories

Popularity Category	Number of Dramas
Top-Rated Dramas	159
Popular Dramas	169
Less Popular Dramas	22
Total	351

To facilitate a clearer understanding of the proportion of each category, the distribution is further visualized using a pie chart. This visualization illustrates that the majority of Korean dramas fall into the Popular Dramas and Top-Rated Dramas categories, while the Less Popular Dramas category represents the smallest proportion.

The category distribution is visualized in the form of a pie chart to observe the proportion between categories:

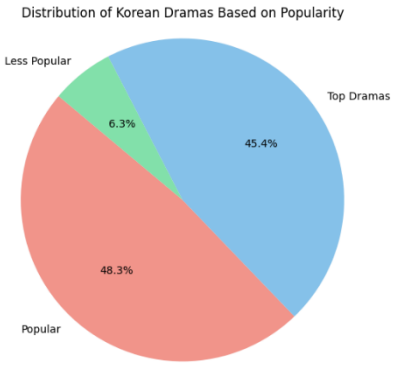

Figure 2. Distribution of Korean Dramas Based on Popularity

3.2 Classification of Dramas Based on Category

Below are the top 10 dramas with the highest ratings from each category:

a. Top Drama (Rating ≥ 8.5)

Table 2. Top Drama Category

Rank	Title	Year	Episodes	Rating	Genre	Actors
1	Move to Heaven	2021	10	9.2	Life, Drama, Family	Lee Je Hoon, Tang Jun Sang
2	Twinkling Watermelon	2023	16	9.2	Romance, Youth, Fantasy	Ryeoun, Choi Hyun Wook
3	Moving	2023	20	9.1	Action, Thriller, Mystery	Ryu Seung Ryong, Han Hyo Joo, Zo In Sung
4	The Trauma Code	2025	8	9.1	Action, Medical, Drama	Ju Ji Hoon, Choo Young Woo
5	Flower of Evil	2020	16	9.1	Thriller, Romance, Crime	Lee Joon Gi, Moon Chae Won
6	Hospital Playlist	2020	12	9.1	Romance, Life, Medical	Jo Jung Suk, Yoo Yeon Seok
7	Hospital Playlist 2	2021	12	9.1	Romance, Life, Medical	Jo Jung Suk, Jung Kyung Ho
8	My Mister	2018	16	9.1	Psychological, Drama, Family	Lee Sun Kyun, IU
9	Reply 1988	2015	20	9.1	Comedy, Youth, Family	Hyeri, Ryu Jun Yeol, Go Kyung Pyo
10	Weak Hero Class 1	2022	8	9.1	Action, Youth, Drama	Park Ji Hoon, Choi Hyun Wook

b. Popular (Rating 7.5–8.4)

Table 3. Popular Drama Category

Rank	Title	Year	Episodes	Rating	Genre	Actors
1	Lost	2021	16	8.4	Life, Drama, Melodrama	Jeon Do Yeon, Ryu Jun Yeol
2	The Atypical Family	2024	12	8.4	Comedy, Romance, Fantasy	Jang Ki Yong, Chun Woo Hee
3	Queen In Hyun's Man	2012	16	8.4	Historical, Romance	Ji Hyun Woo, Yoo In Na
4	One Dollar Lawyer	2022	12	8.4	Comedy, Law, Drama	Namkoong Min, Kim Ji Eun
5	If You Wish Upon Me	2022	16	8.4	Romance, Medical, Drama	Ji Chang Wook, Choi Soo Young
6	I'm Not a Robot	2017	32	8.4	Romance, Comedy, Sci-Fi	Yoo Seung Ho, Chae Soo Bin
7	Memorist	2020	16	8.4	Thriller, Mystery	Yoo Seung Ho, Lee Se Young
8	Voice 2	2018	12	8.4	Thriller, Mystery	Lee Ha Na, Lee Jin Wook
9	Doom at Your Service	2021	16	8.4	Romance, Fantasy, Drama	Park Bo Young, Seo In Guk
10	Duel	2017	16	8.4	Thriller, Sci-Fi	Jung Jae Young, Yang Se Jong

c. Less Popular (Rating < 7.5)

Table 4. Less Popular Category

Rank	Title	Year	Episodes	Rating	Genre	Actors
1	Queen of Divorce	2024	12	7.4	Action, Comedy, Romance	Lee Ji Ah, Kang Ki Young
2	Newtopia	2025	8	7.4	Action, Horror, Romance	Ji Soo, Park Jeong Min
3	Frankly Speaking	2024	12	7.4	Comedy, Romance, Drama	Go Kyung Pyo, Kang Han Na
4	The Bequeathed	2024	6	7.4	Thriller, Mystery	Kim Hyun Joo, Park Hee Soon
5	Pandora: Beneath Paradise	2023	16	7.4	Action, Thriller, Drama	Lee Ji Ah, Lee Sang Yoon
6	Queen of Masks	2023	16	7.3	Thriller, Drama	Kim Sun Ah, Oh Yoon Ah
7	Motel California	2025	12	7.3	Romance, Drama	Lee Se Young, Na In Woo
8	LTNS	2024	6	7.3	Comedy, Drama	Esom, Ahn Jae Hong
9	The Real Has Come!	2023	50	7.3	Romance, Drama	Baek Jin Hee, Ahn Jae Hyun
10	Kokdu: Season of Deity	2023	16	7.2	Comedy, Fantasy, Medical	Kim Jung Hyun, Im Soo Hyang

Classification

The evaluation results are presented through the following classification report

Table 5. Classification

	precision	recall	f1-score	support
0	0.75	0.86	00.80	35
1	0.83	0.71	0.77	35
accuracy	0.79	0.79	0.79	70
macro avg	0.79	0.79	0.78	70
weighted avg	0.79	0.79	0.78	

The classification model evaluation results are presented in the form of a classification report with two classes represented by numerical labels 0 and 1. These labels are the result of a label encoding process, where each corresponds to the Popular and Top Drama categories, respectively. Although the original data were grouped into three popularity categories, this study focuses the classification process on the two main classes due to the limited number of data in the Less Popular category, which could potentially lead to class imbalance.

The results indicate that the model achieves an accuracy of 79% with an average F1-score of 0.78. This demonstrates that the model performs fairly well in distinguishing the two main classes. Next, the prediction results are visualized using a confusion matrix to compare the predicted labels with the actual data.

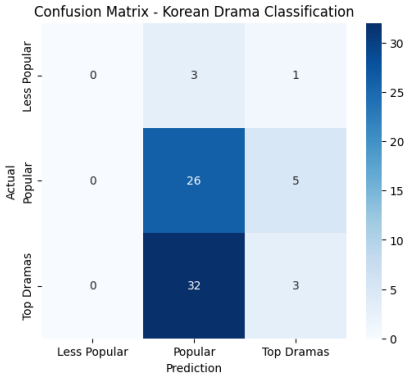

Figure 3. Confusion Matrix

Analysis of Korean Drama Genres and Tags

To understand the main characteristics of Korean dramas in the dataset, an analysis was conducted on the Genre and Tags columns. Data from these columns were collected and visualized using word clouds, which display the most frequently occurring words. The larger the word, the more often it appears.

The following figure shows the word cloud for the Genre column. Words such as Romance, Drama, Comedy, and Action are the most dominant genres in Korean dramas.

Figure 4. Word Cloud of Genres

The word Romance appears most frequently, indicating that romantic elements are a major attraction in the Korean drama industry. Meanwhile, the Tags column highlights typical elements in Korean dramas such as Male Lead, Female Lead, Strong Female, Supporting Character, and Smart Female**.**

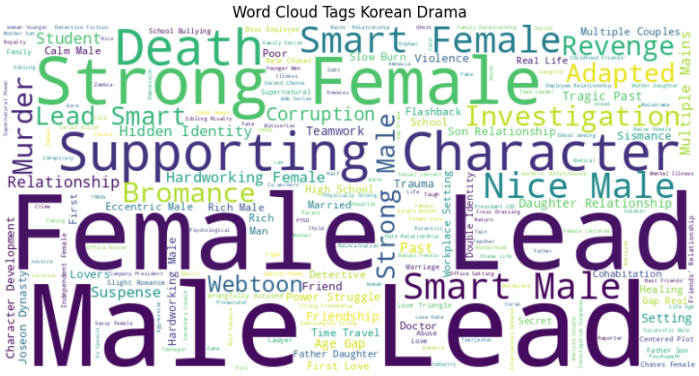

Figure 5. Word Cloud of Tags

Tags like Strong Female, Smart Female**,** and Nice Male reflect the strong and memorable personality traits of the characters. Meanwhile, the presence of tags like Bromance, Revenge, Death, Investigation, and Murder indicates that Korean dramas often explore emotional, mystery, and personal conflict themes. Furthermore, tags such as Adapted from Webtoon, Hidden Identity, and Time Travel indicate modern narrative trends that combine fantasy elements with realistic storytelling.

This pattern shows that Korean dramas do not only focus on romance but also expand into action, psychological, and supernatural genres.

Most Popular Genres

A further analysis was conducted by calculating the frequency of each genre. Genres were split based on commas (,) and counted using Counter. The top ten genres are displayed in a bar chart**.**

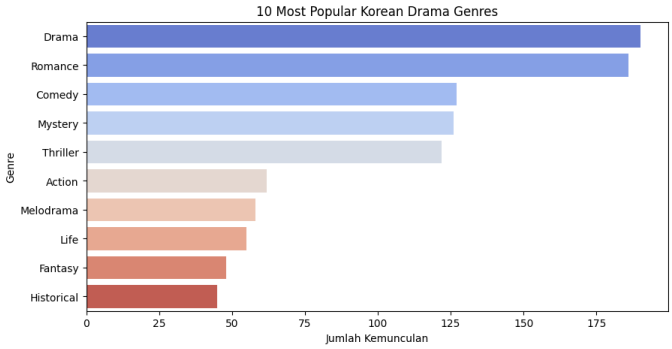

Figure 6. Bar Chart of Top 10 Korean Drama Genres

Genres such as Drama, Romance**,** and Comedy dominate the top ten, reinforcing the findings from the word cloud that emotional and lighthearted elements are very popular among Korean drama viewers.

Naïve Bayes Classification

To understand how the Naive Bayes algorithm works, a manual classification was performed on three example Korean dramas from each class: Best Drama, High Interest, and Low Interest. Numerical features used are Rating, Episode, and Year of Release. Posterior probabilities are computed using the Gaussian method, assuming each numerical feature follows a normal distribution.

a. Naïve Bayes Formula

$$P(Ci \mid X) = \frac{P(X|Ci)\cdot P(Ci)}{P(X)} \quad (1)$$

Since P(X) is the same for all classes:

$$P(Ci \mid X) \propto P(Ci)\cdot \prod_{j=1}^{n} P(xj \mid Ci) \quad (2)$$

$$P(xj \mid Ci) = \frac{1}{\sqrt{2\pi\sigma^2 Ci}} . e^{-\frac{(xj-\mu Ci)^2}{2\sigma^2 Ci}} \quad (3)$$

Notes:

$P(C_i)$ = class prior

μC_i, σC_i = mean and standard deviation of feature *j* di class C_i

x_j = value of feature j from test data

$P(C_i \mid X)$ = posterior probability that X belongs to class C_i

b. Training Data

Table 6. Training Data

Class	Title	Rating	Episode	Year
Top Drama	Move to Heaven	9.2	10	2021
	Twinkling Watermelon	9.2	16	2023
Popular	Lost	8.4	16	2021
	The Atypical Family	8.4	12	2024
Less Popular	Queen of Divorce	7.4	12	2024
	Newtopia	7.4	8	2025

c. Class Statistics

Table 7. Class Statistics

Class	μ Rating	σ Rating	μ Episode	σ Episode	μ Year	σ Year
Top Drama	9.2	0.00	13.0	4.24	2022.0	1.41
Popular	8.4	0.00	14.0	2.83	2022.5	2.12
Less Popular	7.4	0.00	10.0	2.83	2024.5	0.71

d. Prior Probabilities

Total training data: 6 dramas (2 per class)

$$P(\text{Top Drama})=P(\text{Popular})=P(\text{Less Popular})=\frac{6}{2}=0.333$$

e. Posterior Calculation Examples

Move to Heaven (Rating: 9.2, Episode: 10, Year: 2021)
P(Rating=9.2)= 1
$P(10 \mid 13,4.24)\approx 0.193$
$P(2021 \mid 2022,1.41)\approx 0.219$
Posterior:
$P = 0.333\times1\times0.193\times0.219=0.0141$
Rating ≠ μ dan σ = 0 ⇒ P=0
Prediction: Top Drama

Lost (Rating: 8.4, Episode: 16, Year: 2021)
P(Rating=8.4)= 1
$P(16 \mid 14,2.83)\approx 0.112$
$P(2021 \mid 2022.5,2.12)\approx 0.178$
Posterior:
$P = 0.333\times1\times 0.112 \times 0.178 =0.0066$
Rating ≠ μ dan σ = 0 ⇒ P=0
Prediction:: Popular

Queen of Divorce (Rating: 7.4, Episode: 12, Year: 2024)
P(Rating=7.4)= 1
$P(12 \mid 10,2.83)\approx 0.112$
$P(2024 \mid 2024.5,0.71)\approx 0.534$
Posterior:
$P = 0.333\times1\times 0.112 \times 0.534 =0.0199$
Rating ≠ μ dan σ = 0 ⇒ P=0
Prediction: Less Popular

Table 8. Classification Results

Title	P (Best Drama)	P (High Interest)	P (Low Interest)	Prediction
Move to Heaven	0.0141	0	0	Top Drama
Lost	0	0.0066	0	Popular
Queen of Divorce	0	0	0.0199	Less Popular

The calculations above are an illustrative example using the Gaussian Naive Bayes method, with three drama samples from each category, to demonstrate how posterior probabilities are computed. The Rating feature primarily determines the prediction since σ = 0, while Episode and Year only slightly adjust the posterior probabilities.

4. Conclusion

The Naive Bayes algorithm was successfully applied to classify the popularity of Korean dramas into three categories based on rating values. The classification results indicate that the "Popular" category contains the majority of dramas, followed by "Top Drama" and "Less Popular." The model achieved an overall accuracy of 79% with an average F1-score of 0.78, demonstrating that it performs reasonably well in distinguishing between different popularity levels. Analysis of the dataset revealed that the most popular genres are Drama, Romance, and Comedy, highlighting the importance of emotional and light-hearted elements in Korean dramas. Additionally, data visualizations, such as word clouds and bar charts, provided deeper insights into the trends of genres and character tags, helping to better understand the distinctive features of Korean dramas.

References

[1] D. I. S. S. H. Junianto, R. E. Saputro, B. A. Kusuma, "Eksplorasi Sentimen Publik terhadap Film 'Dirty Vote' melalui Metode Naïve Bayes dan Logistic Regression," *JEPIN J. Edukasi dan Penelit. Inform.*, vol. 10, no. 3, 2024, doi: 10.26418/jp.v10i3.78520.

[2] A. G. S. dan M. Laniwati, "Ekstraksi Ulasan Sentimen Film dari Twitter dengan Naïve Bayes pada Situs Web Media Sosial Penggemar Film," *Insyst*, vol. 3, no. 1, 2021, doi: 10.52985/insyst.v3i1.186.

[3] A. P. R. N. Rahmawaty, D. Indrayana, "Penerapan Metode Naïve Bayes dan Cosine Similarity dalam Analisis Sentimen

terhadap Platform Film Ilegal di Media Sosial X (Twitter)," *J. Jamastika*, vol. 3, no. 1, 2024, doi: 10.35473/jamastika.v3i1.3059.

[4] H. A. Z. Zulkarnain, R. Mutia, J. A. Ariani, Z. A. Barik, "Performance Comparison K-Nearest Neighbor, Naive Bayes, and Decision Tree Algorithms for Netflix Rating Classification," *IJATIS*, vol. 1, no. 1, 2024, doi: 10.57152/ijatis.v1i1.1104.

[5] H. M. A. Putri, K. Umam, "Opinion Classification on IMDb Reviews Using Naïve Bayes Algorithm," *J. Appl. Informatics Comput.*, vol. 9, no. 6, 2025, doi: 10.30871/jaic.v9i6.9831.

[6] D. A. R. N. J. Ramadhan, V. A. P. Putri, "Eksplorasi Analisis Sentimen pada Rating Film IMDb: Pendekatan Perbandingan menggunakan Decision Tree dan Naive Bayes," *Innov. J. Soc. Sci. Res.*, vol. 4, no. 3, 2024, doi: 10.31004/innovative.v4i3.10438.

[7] T. W. J. A. Rieuwpassa, S. Sugito, "IMPLEMENTASI METODE NAIVE BAYES CLASSIFIER UNTUK KLASIFIKASI SENTIMEN ULASAN PENGGUNA APLIKASI NETFLIX PADA GOOGLE PLAY," *J. Gaussian*, vol. 12, no. 3, pp. 362–371, 2024, doi: 10.14710/j.gauss.12.3.362-371.

[8] A. N. dan B. N. Sari, "Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naïve Bayes," *ITET (Jurnal Inform. dan Tek. Elektro Ter.*, vol. 11, no. 3, 2023, doi: 10.23960/jitet.v11i3%20s1.3348829.

[9] A. T. N. N. Z. B. Toyibah, Y. N. Putri, P. Puandini, Z. M. Widodo, "Perbandingan Kinerja Algoritma Multinomial Naïve Bayes dan Logistic Regression pada Analisis Sentimen Movie Ratings IMDB," *J. Ilm. Edutic Pendidik. dan Inform.*, vol. 10, no. 2, pp. 181–189, 2024, doi: 10.21107/edutic.v10i2.28150.

[10] P. D. A. dan S. Lestari, "Analisa Sentimen Drama Korea Melalui Media Sosial X dengan Menggunakan Algoritma Naïve Bayes," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 3, pp. 3248–3261, 2024, doi: 10.35870/jimik.v5i3.997.

[11] H. Meisty, L., Utami, R., & Saputra, "Perbandingan Algoritma Naive Bayes dan Random Forest dalam Klasifikasi Ulasan Drama Korea," *J. Ilm. Komput. dan Inform.*, vol. 7, no. 1, pp. 45–53, 2024, doi: 10.20956/ejsa.v5i1.26942.

[12] T. I. H. Y. Nurtikasari, S. Alam, "Analisis Sentimen Opini Masyarakat Terhadap Film Pada Platform Twitter Menggunakan Algoritma Naive Bayes," *Insologi*, vol. 1, no. 4, 2022, doi: 10.55123/insologi.v1i4.770.

[13] I. I. L. Alyandi, L. Hadiani, "Implementasi Naive Bayes dalam Analisis Sentimen Ulasan Game Honor of Kings di Playstore," *Busiti*, vol. 6, no. 1, 2025, doi: 10.33096/busiti.v6.

[14] F. N. dan J. R. Rizkiani, "Perbandingan Metode Naive Bayes Classifier dan Support Vector Machine pada Analisis Sentimen Twitter Topik Lifestyle," *J. Ilm. Wahana Pendidik.*, vol. 9, no. 21, pp. 314–323, 2023, doi: 10.5281/zenodo.10077023.

[15] M. A. J. H. A. Z. Amrullah, A. S. Anas, "Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square," *BITE*, vol. 2, no. 1, 2023, doi: 10.30812/bite.v2i1.804.