Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2025. Vol. 4. No. 2; e-ISSN: 2808-4519

The Improvement of Indonesian Film Genre Clustering Model Using the K-Means Algorithm in Film Production Decision-Making

Wiratriyana[1*], Martanto[2], Arif Rinaldi Dikananda [3], Mulyawan [4]

[1,2,3,4]*STMIK IKMI Cirebon*
wiratriyana20@gmail.com [1*]

Abstract

The Indonesian film industry is expanding rapidly, but understanding audience preferences remains a significant challenge for producers. This study aims to cluster Indonesian films by genre and synopsis using the K-Means algorithm to aid in marketing strategies and content development. The dataset comprises 1,271 Indonesian film entries, including attributes like release year, genre, synopsis, and user ratings. The research follows the Knowledge Discovery in Databases (KDD) framework, which involves data selection, preprocessing, transformation, clustering with K-Means, and evaluation using the Elbow method to identify the optimal number of clusters. The results show that the K-Means algorithm successfully grouped the films into three clusters: drama, horror, and others. The analysis indicates that drama films dominate the high-rating cluster, while horror films are more commonly found in the low-rating category. The use of Principal Component Analysis (PCA) in the visualization aids in interpreting the clustering results, providing a clearer view of the data distribution. These findings highlight the potential for improving film production strategies by aligning content with audience preferences. By understanding genre patterns and ratings, producers can make more informed decisions in marketing and content development.

Keywords: *K-Means Algorithm, Film Genre, Film Clustering, Marketing Strategy, Data Analysis.*

1. Introduction

Significant advancements in informatics have impacted various aspects of life, including technology, business, education, and entertainment. Digital technologies enable automation, large-scale data analysis, and faster, more accurate decision-making. In recent decades, integrating information technology into content production, distribution, and audience preference analysis has transformed the entertainment industry, especially filmmaking. Algorithms like machine learning and data clustering are now essential for understanding audience behavior, optimizing content, and improving global competitiveness. In Indonesia, the film sector has grown rapidly with the rise of online content consumption.

However, the variety of film genres and synopses poses challenges in understanding fragmented audience preferences. Producers often struggle to develop effective production and marketing strategies, particularly when audience data is not processed optimally. The abundance of available content complicates efforts to identify trends and patterns that could guide production and promotion.
Previous studies have used the K-Means algorithm for content clustering and recommendations[1] applied K-Means to classify 1,272 Indonesian films into recommended and less recommended categories based on attributes like ratings and genres, though their study did not explore advanced evaluation metrics[2] clustered cinema screening data but overlooked the shift to digital platforms. [3] developed a web-based recommendation system using K-Means, allowing users to filter by genre, year, and other criteria. While effective, this approach lacked deeper cluster quality evaluation and algorithm optimization for diverse datasets.
Advanced algorithms like K-Means clustering offer promising solutions to these challenges. By grouping films based on genres and synopses, K-Means helps identify distinct audience segments and preferences. This enables targeted content creation and marketing strategies while providing a structured understanding of market demands to improve production decisions and optimize resources.
In the Indonesian film industry, adopting these algorithms could be transformative. As the industry grows, understanding audience behavior is vital for success in a competitive market. Data clustering techniques like K-Means allow producers to streamline decision-making and deliver content that better aligns with diverse audience preferences. This approach enhances both the quality and profitability of the industry in the digital era.

Providing relevant, personalized recommendations is increasingly important in the Indonesian film industry, especially with the growing availability of films on digital platforms. While K-Means is effective for clustering, its application in categorizing Indonesian films based on synopses and genres requires further exploration.
Therefore, this study aims to answer several key questions related to the application of this algorithm and the preferences of Indonesian film audiences, as outlined in the following problem statement:

1. What are the main patterns or clusters of Indonesian films based on genre, user ratings, and release year?
2. Do certain genres dominate in high-rated or low-rated film categories?

This research is conducted to answer the questions arising from the problem statement regarding the application of the K-Means algorithm in clustering Indonesian films. By explaining the implementation of this method and understanding audience preferences, the study is expected to provide valuable contributions to optimizing content recommendations, enhancing film production strategies, and supporting more effective marketing approaches in the Indonesian film industry.

2. Literature Review

2.1. Related Research Results

This literature highlights the widespread application of the K-Means clustering algorithm across various fields, demonstrating its effectiveness in solving clustering problems and uncovering valuable insights. [4] analyzed customer shopping patterns using K-Means with WEKA, identifying five distinct clusters to better understand consumer behavior based on their shopping intensity. As a result, this study provides a clearer picture of consumer preferences and helps in formulating more effective marketing strategies. [5] applied K-Means to cluster frozen food sales into two groups, enabling more accurate mapping of customer preferences in the e-commerce sector. This approach helps companies improve their market segmentation and tailor their product offerings to meet consumer needs.
In a broader context, [6] used a hybrid approach combining Support Vector Machine (SVM) and SMOTE for sentiment analysis on Vtuber-related content, achieving an impressive accuracy rate of 88.18%. This method successfully identified dominant sentiments that could be used to design more relevant content and enhance user experience.

In higher education, [7] utilized K-Means to strategize university promotions by clustering new student data and uncovering valuable insights for designing targeted and efficient campaigns to attract prospective students. [8] compared K-Means with DBSCAN and Hierarchical clustering for market segmentation in the automotive industry, using KNIME software to determine the most effective clustering model. The findings from this study provide useful insights for developing more specific marketing strategies, making it easier for car manufacturers to target the right consumers. Meanwhile, [9] explored employee discipline patterns through e-attendance data using K-Means and Orange software, aiming to identify trends in punctuality and attendance. This research is significant in human resource management as it can help organizations improve employee discipline and design more targeted policies.

[10] applied K-Means to Netflix data to cluster films based on genre and rating, providing recommendations more suited to viewer preferences. With more accurate clustering, this algorithm can enhance the user experience and make it easier for viewers to select content according to their tastes. [11] researched song attributes on Spotify and found that attributes such as instrumentalness and valence significantly influenced song popularity. This study used the Davies-Bouldin Index to validate clustering results and identified nine optimal clusters, allowing Spotify to better understand user preferences and improve their music recommendation service. K-Means has also been widely applied in the retail sector. [12] used the algorithm to analyze sales trends in a footwear store, aiming to optimize stock and improve marketing decisions. [13] leveraged K-Means for product segmentation, ultimately improving inventory management and ensuring better product availability in the market.

Research by [14] also showed that K-Means could be used to identify popular café menus, allowing café owners to adjust their menu offerings and improve customer satisfaction. In the healthcare field, [15] utilized K-Means to classify patient diseases, which could help in designing more effective preventive measures. Lastly, [16]explored the role of social media in Indonesian film marketing, emphasizing its influence on audience satisfaction and purchase intent. The study revealed that interactive and informative film marketing campaigns on social media could boost customer engagement, increase ticket sales, and enhance viewer satisfaction. This highlights that social media has become a highly effective tool in attracting audience attention and influencing their decisions to watch films.

Overall, these studies illustrate the versatility of K-Means clustering across various fields, including retail, entertainment, education, healthcare, and marketing. This is further supported by advancements in tools like RapidMiner, KNIME, and Orange, which simplify its implementation and increase its practicality in various real-world applications. In summary, the literature review highlights the effectiveness of the K-Means algorithm in clustering films based on various attributes such as genre, user ratings, and synopses. Several studies have demonstrated the algorithm's ability to group films into meaningful clusters, which can improve the personalization of film recommendations. These findings support the potential of K-Means clustering in the Indonesian film industry, where a better understanding of audience preferences could lead to more targeted and effective content creation and marketing strategies.

3. Research Methods

The research method used in this study is Knowledge Discovery in Databases (KDD). KDD is a systematic process for discovering useful information or patterns from large and complex datasets. This process consists of several main stages, including data selection, data preprocessing, data transformation, application of data mining techniques, and interpretation and evaluation of results. In this study, the K-Means algorithm is applied during the data mining phase to cluster Indonesian films based on their synopsis and genre.

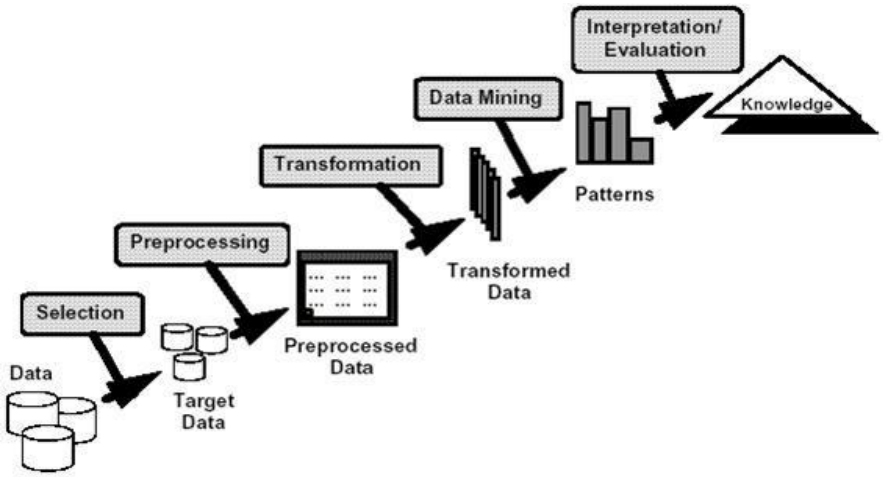

Figure 1: Research Stages of KDD Method

The data processing in this study is performed in stages to ensure accuracy in the results. The process begins with data selection, where relevant information such as genre, release year, and user ratings are chosen for analysis. This is followed by data preprocessing, which involves cleaning the dataset by removing incomplete or erroneous data to maintain its quality. These initial steps are crucial to preparing the data for the subsequent analysis using the K-Means algorithm.

4. Result And Discussion

4.1. Data Selection

Data Selection aims to choose and prepare data that is relevant to the analysis objectives. This process includes loading the dataset, performing an initial inspection to understand the data structure, and identifying the columns to be used in the subsequent steps. This step is essential to ensure that the data used meets the requirements for the analysis, allowing for more accurate and meaningful results.

Table 1: Dataset Reading Results

No	Judul Film	Tahun	…	Genre	Rating	…	…	…	…	Runtime
1	#FriendButMarried 2	2020	...	Biography	13+	...	...	...	...	100 min
2	4 Mantan	2020	...	Thriller	17+	...	...	...	...	80 min
3	Aku Tahu Kapan Kamu Mati	2020	...	Horror	13+	...	...	...	...	92 min
4	Anak Garuda	2020	...	Adventure	13+	...	...	...	...	129 min
5	Dignitate	2020	...	Drama	17+	...	...	...	...	109 min
...	...	...	...	...	...	...	...	...	...	...
1267	The Tiger from Tjampa	1953	...	Drama	NaN	...	...	...	...	97 min
1268	Enam Djam di Djogja	1951	...	Drama	NaN	...	...	...	...	116 min
1269	Darah dan Doa	1950	...	Drama	NaN	...	...	...	...	150 min
1270	Resia Boroboedoer	1928	...	Adventure	NaN	...	...	...	...	NaN
1271	Loetoeng Kasaroeng	1926	…	Fantasy	NaN	..	…	…	…	60 min

4.2. Transformation Data

Data transformation refers to the process of converting data from its original format or structure into a format that is more suitable for visualization, analysis, or use in machine learning models. The goal of data transformation is to enhance data quality, align data on the same scale, or simplify its analysis. It is a critical step in data preprocessing, as it ensures the data is in the proper format for further analysis and modeling.

Table 2: Results of Data After Standardization

	year	users_rating	genre_encoded
0	0.982768	0.258968	-0.807159
1	0.982768	0.187072	2.406.914
2	0.982768	-0.531898	1.442.692
3	0.982768	2.128.288	-1.449.973
4	0.982768	1.049.834	0.157063
...	...	...	...
1231	-4.270.249	0.187072	0.157063
1232	-4.427.055	0.115175	0.157063
1233	-4.505.458	0.330865	0.157063

1234	-6.230.329	0.618453	-1.449.973
1235	-6.387.136	0.762247	0.799878

Table 3: Results After Performing PCA

	pca_1	pca_2
0	-0.340173	0.987975
1	1.952.760	0.886794
2	1.709.397	0.528113
3	-1.912.208	1.982.689
4	-0.139966	1.384.427
...	...	...
1267	-1.654.664	-3.537.363
1268	-1.672.181	-3.708.523
1269	-1.831.989	-3.661.907
1270	-3.796.919	-4.946.403
1271	-2.369.064	-5.048.639

The table above explains the results of Principal Component Analysis (PCA), which was used to reduce the dataset's dimensions into two principal components, "pca_1" and "pca_2," each contributing the most to the data's variability. The "pca_1" column represents the first principal component, which explains the greatest variance in the data. The two value columns show the data's position in the new space, based on the linear combination of the original features. For example, the values "pca_1 = -0.340173" and "pca_2 = 0.987975" found in the first row represent the data coordinates in the reduced-dimensional space. These values allow the data to be visualized in two dimensions, reducing the complexity of the analysis while retaining important information. This process is often used to identify patterns or relationships within the data and improve the efficiency of machine learning models.

4.3. Data Mining

Data mining is the process of discovering hidden patterns or information in data using analytical techniques such as classification, clustering, or association. The goal is to uncover insights or trends that can support decision-making. Data mining is a core stage in the KDD process, occurring after data cleaning and before result interpretation. Determining the optimal number of clusters using the Elbow method is the process of identifying the point at which the decrease in clustering cost starts to slow down. In this method, the cost (e.g., within-cluster sum of squares) is calculated for different numbers of clusters and plotted. The point where the cost reduction is no longer significant, or the "elbow," is considered the optimal number of clusters. This method helps find an efficient and easily interpretable division of the data.

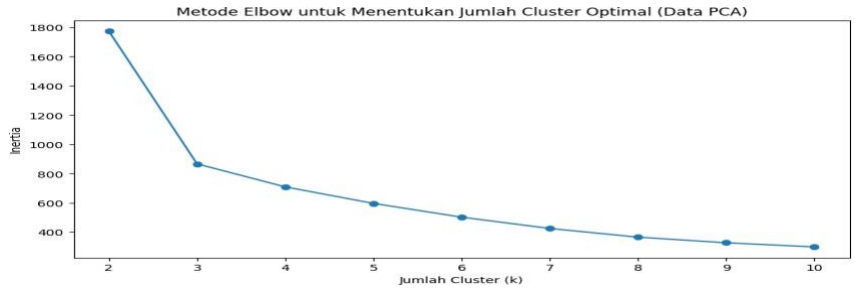

Figure 2: Elbow Method Visualization

Table 4: Clustering Results

	title	*cluster*
0	#FriendButMarried 2	0
1	4 Mantan	2
2	Aku Tahu Kapan Kamu Mati	2
3	Anak Garuda	0
4	Dignitate	0
1267	The Tiger from Tjampa	1
1268	Enam Djam di Djogja	1
1269	Darah dan Doa	1
1270	Resia Boroboedoer	1
1271	Loetoeng Kasaroeng	1

4.4. Interpretation/Evaluation

The Interpretation and Evaluation stage aims to understand the clustering results and assess the quality of the K-Means model. After clustering, patterns within each cluster are analyzed, and the quality is evaluated using metrics such as Silhouette Score or WCSS (Within-Cluster Sum of Squares). Visualization of clustering results helps to graphically depict the data distribution within clusters, making it easier

to understand the patterns formed. By using Principal Component Analysis (PCA), high-dimensional data can be reduced to two or three dimensions, allowing the clusters to be visualized in scatter plots or other diagrams.

```
# Visualisasi hasil clustering pada data hasil PCA
plt.figure(figsize=(10, 6))
sns.scatterplot(x=data['pca_1'], y=data['pca_2'], hue=data['cluster'],
                palette='viridis')
plt.title(f'Clustering Film Indonesia dengan PCA (k = {optimal_k})')
plt.xlabel('Komponen Utama 1')
plt.ylabel('Komponen Utama 2')
plt.show()
```

Figure 3: Code for Visualizing Clustering Results

The figure above shows the code that generates the visualization of the clustering results for Indonesian film data using PCA (Principal Component Analysis) to reduce the data dimensions into two principal components. The resulting scatter plot displays the data distribution along the Principal Component 1 and Principal Component 2 axes, with different colors representing each cluster based on the results of K-Means Clustering.

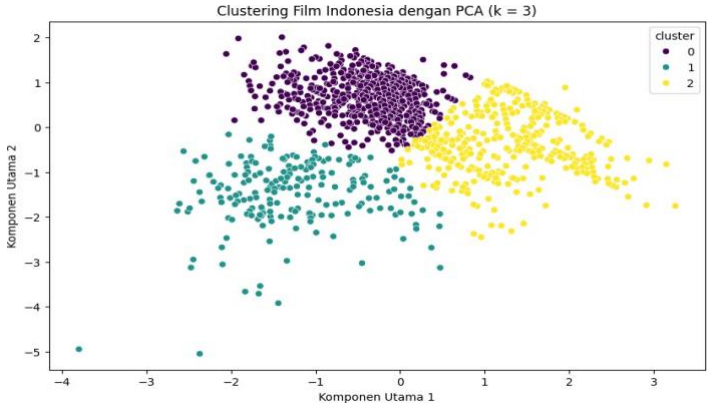

Figure 4: Clustering Visualization Results

The figure above shows the clustering visualization of Indonesian film data, with dimensionality reduced using PCA (Principal Component Analysis) into two principal components, "Principal Component 1" and "Principal Component 2," to facilitate visualization. The data is grouped into three clusters (k = 3), each represented by a different color (purple, bluish-green, and yellow), with each point representing a single film. The distribution of points indicates that the data is relatively well-separated into three distinct groups based on certain patterns, although there are areas showing some overlap between clusters. This visualization aids in understanding the grouping of films based on specific features such as genre, duration, or rating, which form the basis of the analysis.

Table 5: Characteristics of Each Cluster

cluster	year_Min	year_Max	year_Rata-rata	users_rating_Min	users_rating_Max	users_rating_Rata- rata	genre_name_<lambda>
0	1995	2020	2.013.388.889	3.6	9.4	6.869.290	Drama
1	1926	2001	1.983.049.505	2.9	8.7	6.473.267	Drama
2	1981	2020	2.010.297.927	1.2	7.5	4.740.674	Horror

The table above provides a summary of the characteristics of each cluster resulting from the clustering of film data based on several key features. Cluster 0 includes films released between 1995 and 2020, with an average release year of 2013, user ratings ranging from 3.6 to 9.4 and an average of 6.87, with Drama being the dominant genre. Cluster 1 consists of films released between 1926 and 2001, with an average release year of 1983, user ratings ranging from 2.9 to 8.7 and an average of 6.47, also dominated by the Drama genre. Meanwhile, Cluster 2 contains films released between 1981 and 2020, with an average release year of 2010, user ratings varying from 1.2 to 7.5 and an average of 4.74, with Horror as the dominant genre. This table illustrates the significant differences in the characteristics of each cluster, based on release periods, rating quality, and genre preferences.

Table 6: Results of Separating Films Based on Ratings

High Rated Movies :	title	users_rating	genre_name
0	#FriendButMarried 2	6,5	Biography
1	4 Mantan	6,4	Thriller

3	Anak Garuda	9,1	Adventure
4	Dignitate	7,6	Drama
6	Hunter in the Blue Side of Manchester	6.2	Biography
Low Rated Movies :	title	users_rating	genre_name
2	Aku tahu Kapan Kamu Mati	5.4	Horror
5	Guru-Guru Gokil	6.1	Comedy
7	Janin	6.1	Horror
8	Mangkujiwo	6.1	Horror
9	Milea	6.1	Drama

Table 7: Results After Calculating Genre Dominance in Each Rating Category

Genre	Count (High Rated)	Count (Low Rated)
Drama	354	102
Comedy	147	140
Action	72	60
Horror	51	180
Biography	28	-
Adventure	28	21
Thriller	8	4
Romance	8	3
Crime	6	3
Fantasy	4	6
Animation	3	1
Family	2	-
War	2	-
History	1	-
Sci-Fi	1	1

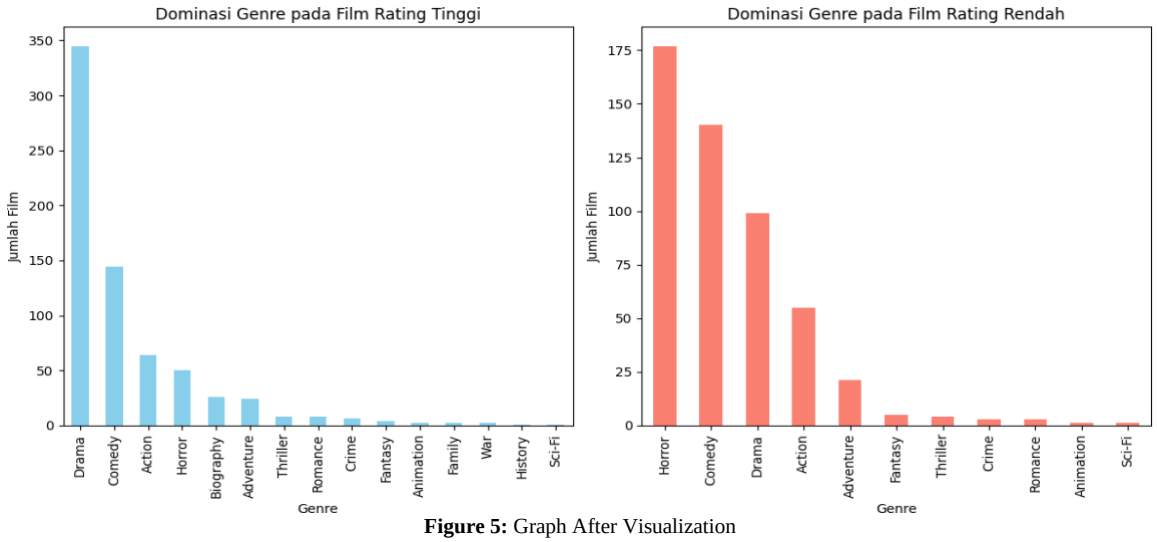

Figure 5: Graph After Visualization

Based on the figure above, the drama genre dominates films with high ratings, followed by comedy, action, and horror, indicating that drama tends to be more appreciated by viewers. On the other hand, for films with low ratings, the horror genre ranks highest, followed by comedy, drama, and action, suggesting that horror films are more susceptible to poor reception if not executed well. This data underscores the importance of production quality within specific genres to achieve successful ratings.

The analysis results indicate that the K-Means algorithm is effective in clustering film data based on genre and related attributes. The preprocessing phase involved cleaning the data by removing missing values, converting categorical columns into numeric format using label encoding, and standardizing features like release year and user ratings to align the data scale. These steps ensured better data quality and minimized biases during the clustering analysis. Additionally, the dimensionality of the data, which was initially complex, was successfully reduced using Principal Component Analysis (PCA), resulting in two principal components that facilitated visualization and sped up the clustering process without losing crucial information.
The elbow graph used to determine the optimal number of clusters shows a significant decrease in inertia until the third cluster, after which the curve flattens. This suggests that dividing the data into three clusters is the optimal choice for achieving a representative grouping. The clustering results visualized in a scatter plot reveal a clear distribution of data into three clusters based on specific patterns, though there are some areas that show overlap between clusters. This visual representation helps confirm the effectiveness of the clustering process and highlights the distinction between different groups of films.

Further analysis of genre distribution based on ratings shows that films in the high-rating category tend to be dominated by the drama genre, followed by comedy, action, and horror. This indicates that drama films are generally more appreciated by viewers. Conversely, horror films occupy the top position in the low-rating category, followed by comedy, drama, and action. This suggests that horror films are more susceptible to poor reception if not executed well, and highlights the importance of quality production within each genre to achieve higher ratings.

The findings emphasize the critical role of data preprocessing, dimensionality reduction, and proper cluster analysis in uncovering meaningful patterns in film data. Understanding the relationship between genre, ratings, and user preferences provides valuable insights for filmmakers and marketers to tailor their content and improve their strategies for film production and distribution.

5. Conclusion

Based on the study on the K-Means Algorithm to Enhance the Indonesian Film Genre Clustering Model, it can be concluded that this algorithm is effective in clustering Indonesian films into three main clusters with distinct characteristics based on genre, synopsis, release year, and user ratings. The first cluster (Cluster 0) is dominated by drama films, with release years ranging from 1995 to 2020, and an average user rating of 6.87, the highest among the clusters. The second cluster (Cluster 1) consists of classic drama films released between 1926 and 2001, with an average user rating of 6.47. Meanwhile, the third cluster (Cluster 2) is dominated by horror films, released between 1981 and 2020, and has the lowest average user rating of 4.74. Analysis of the relationship between genre and user ratings shows that drama films tend to fall into the high-rating category, while other genres, such as horror, are more likely to have lower ratings, indicating that execution quality significantly impacts the reception of specific genres.

Acknowledgement

With this, I would like to express my gratitude for the opportunity to complete this task. The analysis and implementation of the K-Means algorithm to cluster Indonesian films based on genre, synopsis, release year, and user ratings have been both insightful and rewarding. It is my hope that the findings contribute to a deeper understanding of film data and help in enhancing the strategies for film production, marketing, and audience engagement. I sincerely appreciate the support and guidance throughout the process, and I am glad to have successfully completed this research. Thank you for your attention and the chance to contribute to this valuable project.

References

[1] Elsa Vania, Salma Nuraini, and D. S. Y. Kartika, "Penggunaan Algoritma K-Means Clustering Untuk Menentukan Rekomendasi Film Indonesia," *Pros. Semin. Nas. Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 207–214, 2022, doi: 10.33005/sitasi.v2i1.299.

[2] M. N. Dayat, N. Suarna, and Y. A. Wijaya, "Analisa Clustering untuk Mengelompokan Data Penayangan Film Bioskop Menggunakan Algoritma K-Means," *Intern. (Information Syst. Journal)*, vol. 6, no. 1, pp. 68–78, 2023, doi: 10.32627/internal.v6i1.686.

[3] P. Parlaungan, F. Alva Mustika, and H. Dhika, "Sistem Rekomendasi Film Menggunakan Metode K-Means Clustering Berbasis Web," *J. SIMETRIS*, vol. 13, no. 2, pp. 1–17, 2022, [Online]. Available: jurnal.umk.ac.id

[4] A. Falakhi, "Pengolahan Data Pelanggan Dengan Tenik Clustering K-Means Di Aplikasi Weka," *J. Comput. Sci. Inf. Syst. J-Cosys*, vol. 3, no. 2, pp. 54–60, 2023, doi: 10.53514/jco.v3i2.394.

[5] T. Amalina, D. Bima, A. Pramana, and B. N. Sari, "Metode K-Means Clustering Dalam Pengelompokan Penjualan Produk Frozen Food," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 15, pp. 574–583, 2022, [Online]. Available: https://doi.org/10.5281/zenodo.7052276

[6] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.

[7] M. R. Alhapizi, M. Nasir, and I. Effendy, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang," *J. Softw. Eng. Ampera*, vol. 1, no. 1, pp. 1–14, 2020, doi: 10.51519/journalsea.v1i1.10.

[8] S. D. K. Wardani, A. S. Ariyanto, M. Umroh, and D. Rolliawati, "Perbandingan Hasil Metode Clustering K-Means, Db Scanner & Hierarchical Untuk Analisa Segmentasi Pasar," *JIKO (Jurnal Inform. dan Komputer)*, vol. 7, no. 2, p. 191, 2023, doi: 10.26798/jiko.v7i2.796.

[9] R. S. T. Atmojo, "Analisis Data E-Absensi untuk Menganalisis Perbandingan Pola Disiplin Kerja menggunakan Algoritma Clustering K-Means," *Electrician*, vol. 13, no. 1, p. 19, 2019, doi: 10.23960/elc.v13n1.2088.

[10] C. Budihartanti, C. I. Ifaru, A. Zahra, and M. H. Aenuddin, "Pengelompokkan Film Pada Platform Netflix Menggunakan Metode K-Means Clustering Sebagai Rekomendasi Film," vol. 5, no. 4, pp. 1392–1402, 2024, doi: 10.47065/josh.v5i4.5482.

[11] M. Ikhsan Firmansyah, R. Saepul Rohman, and E. Marsusanti, "Penerapan Algoritma Klastering K-Means Untuk Fitur Atribut Pada Layanan Streaming Musik Spotify," *Indones. J. Comput. Sci.*, vol. 2, no. 2, pp. 58–66, 2023, doi: 10.31294/ijcs.v2i2.2465.

[12] W. W. Kristianto and C. Rudianto, "Penerapan Data Mining Pada Penjualan Produk Menggunakan Metode K-Means Clustering (Studi Kasus Toko Sepatu Kakikaki)," *J. Pendidik. Teknol. Inf.*, no. 5, pp. 90–98, 2020.

[13] A. Triningsih and H. Supriyono, "Aplikasi Data Mining Berbasis Web Menggunakan Metode K-Means Clustering Untuk Pengelompokan Penjualan," *J. Insypro (Information Syst. Process.*, vol. 4, no. 1, pp. 1–7, 2019, [Online]. Available: https://journal.uin-alauddin.ac.id/index.php/insypro/article/view/6889

[14] S. D. Prasetiani and N. Rochmawati, "Penerapan Data Mining Untuk Clustering Menu Favorit Menggunakan Algoritma K-Means (Studi Kasus Kedai Expo)," *J. Informatics Comput. Sci.*, vol. 3, no. 03, pp. 278–286, 2022, doi: 10.26740/jinacs.v3n03.p278-286.

[15] R. Bayu Prasetyo, Y. Agus Pranoto, and R. Primaswara Prasetya, "Implementasi Data Mining Menggunakan Algoritma K-Means Clustering Penyakit Pasien Rawat Jalan Pada Klinik Dr. Atirah Desa Sioyong, Sulteng," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 4, pp. 2144–2151, 2023, doi: 10.36040/jati.v7i4.7419.

[16] M. F. Djamaly, D. Djumarno, R. Astini, and D. Asih, "Literature Review: Peran Media Sosial Dalam Pemasaran Film Indonesia: Analisis Kepuasan Dan Niat Beli Penonton," *Sci. J. Reflect. Econ. Accounting, Manag. Bus.*, vol. 6, no. 3, pp. 647–659, 2023, doi: 10.37481/sjr.v6i3.706.