IMPLEMENTATION OF A GAME RECOMMENDATION SYSTEM USING THE K-MEANS CLUSTERING AND CONTENT-BASED FILTERING METHODS

Rianggi Silvi Anti-[1], Nanang Ruhyana-[2*)], Syarah Seimahura-[3], Andri Agung Riyadi-[4]

Prodi Sains Data
Faculty of Information Technology
Universitas Nusa Mandiri
15210018@nusamandiri.ac.id-[1], nanang.ngy@nusamandiri.ac.id-[2], syarah.yrs@nusamandiri.ac.id-[3]
andri.agv@nusamandiri.ac.id-[4]

(*) Corresponding Author

Abstract

The rapid growth of the digital game industry has made it increasingly difficult for users to find games that match their preferences due to the limited availability of personalized information. This study aims to develop a web-based game recommendation system using a hybrid approach that combines K-Means Clustering and Content-Based Filtering to improve the accuracy and relevance of recommendations. The dataset was obtained from the RAWG API and consisted of 1,000 game records with key attributes including game name, genre, platform, rating, and age classification (ESRB). The research methodology involved data collection, data preparation, exploratory data analysis, data transformation and preprocessing, application of the K-Means algorithm for game segmentation, and similarity calculation using the cosine similarity method. The hybrid recommendation approach was implemented by restricting the recommendation process to games within the same cluster before applying content similarity analysis. The results indicate that the integration of clustering and content-based methods produces more relevant and contextual recommendations compared to single-method approaches. Visualization using UMAP and t-SNE demonstrates clear and well-defined cluster separation. The developed recommendation system was implemented using the Django framework and deployed on the Google Cloud Platform, enabling real-time access and providing an efficient, adaptive, and scalable solution for game recommendation..

Keywords: K-Means Clustering, Hybrid Recommendation, Content-based Filtering, Game, System Recommendation

Abstrak

Pertumbuhan industri game digital yang sangat pesat menyebabkan pengguna menghadapi kesulitan dalam menemukan game yang sesuai dengan preferensi mereka akibat keterbatasan informasi yang terpersonalisasi. Penelitian ini bertujuan untuk mengembangkan sistem rekomendasi game berbasis web dengan pendekatan hybrid yang mengombinasikan metode K-Means Clustering dan Content-Based Filtering guna meningkatkan akurasi serta relevansi rekomendasi. Dataset yang digunakan berasal dari RAWG API dengan total 1.000 data game yang memiliki atribut utama berupa nama game, genre, platform, rating, dan kategori usia (ESRB). Tahapan penelitian meliputi proses pengumpulan data, data preparation, analisis eksploratif, transformasi dan preprocessing data, penerapan algoritma K-Means untuk melakukan segmentasi game, serta perhitungan tingkat kemiripan menggunakan metode cosine similarity. Pendekatan hybrid diterapkan dengan membatasi proses rekomendasi hanya pada game yang berada dalam cluster yang sama sebelum dilakukan perhitungan kesamaan konten. Hasil penelitian menunjukkan bahwa kombinasi kedua metode mampu menghasilkan rekomendasi yang lebih relevan dan kontekstual dibandingkan penggunaan satu metode tunggal. Visualisasi menggunakan UMAP dan t-SNE memperlihatkan pemisahan cluster yang jelas dan representatif. Sistem rekomendasi yang dikembangkan diimplementasikan menggunakan framework Django dan dideploy pada Google Cloud Platform sehingga dapat diakses secara real-time, adaptif, dan efisien bagi pengguna..

Kata kunci: K-Means Clustering, Hybrid Rekomendasi, Content-based Filtering, Game, Sistem Rekomendasi

INTRODUCTION

The mobile gaming industry has experienced rapid growth in recent years due to improvements in internet network quality, device sophistication, and digital technology innovation. This phenomenon has transformed games from mere entertainment into interactive learning media and sources of income (Agustina et al., 2022). Indonesia itself is one of the largest markets globally, with 94.5% of internet users aged 16–64 recorded as playing games, placing it third in the world (Dihni, 2022).

However, the increasing number of games has created a new problem, namely the difficulty users have in finding games that suit their preferences due to the limited amount of personalized information. This situation highlights the urgency of developing a recommendation system that can help users choose relevant games (Widaraeni & Vivianti, 2021) (Safitri et al., 2024). Research by (Pragusma et al., 2023) in "Game Recommendation Using Content-based Algorithm" shows that a content-based approach is effective in providing game recommendations that match player characteristics.

However, the content-based approach is considered to be the most widely used and has been applied in various fields, such as music, movies, and video games. -Based Filtering, which emphasizes individual preferences (Rochmad Wahono et al., 2024), and the study "A Novel Video Game Recommender System Using Content-Based" (Chythanya N et al., 2019), which proves the effectiveness of content-based algorithms in the gaming domain through the use of attributes such as genre, rating, and game mechanics. The other approach is Collaborative Filtering, which utilizes similarities between users (Widya et al., 2025). In addition, the research "Content-Based Player and Game Interaction Model for Game Recommendation" (Viljanen et al., 2020) emphasizes the importance of modeling the interaction between players and game content in improving the accuracy of recommendation systems.

Another widely used method is clustering, particularly the K-Means Clustering algorithm, which has the potential to improve recommendation quality by grouping games based on similar characteristics (Fitri et al., 2023) (Yudhistira & Andika, 2023)). Several recent studies, such as "Game Recommendation Algorithms Based on Hybrid" (Y. Wang et al., 2024), confirm that combining content-based and collaborative filtering approaches produces more varied and relevant recommendations for users. Additionally, the study "DRGame: Diversified Recommendation for Multi-category Video Games" (Liu et al., 2023) also highlights the importance of diversification in game recommendations to avoid monotonous content similarity for users.

Several other studies have also focused on sentiment analysis to understand user perceptions of games. The study "Sentiment Analysis and Rating Video Game Dimensions via NLP" (Yuan et al., 2025)shows that analyzing user reviews can more accurately describe satisfaction levels and gaming experiences. Research by Renaldi Fauzi Adnan and Ikrimach (2024) developed an Android-based PC game encyclopedia application using the SAW method with an accuracy rate of 85%. Meanwhile, Rohim Nur Rahman, Abdul Rahim, and Wawan Joko Pranoto (2025) applied the Naïve Bayes algorithm for sentiment analysis on eFootball 2024 game reviews and achieved an accuracy of 85% (Nurdy et al., 2024). Furthermore, the study "Using Game Reviews to Recommend Games" (Meidl et al., n.d.) proves that user reviews can be used as direct input in the game recommendation process. Additionally, the study "Extraction of User Opinions by Adjective-Context Co-clustering for Game Review Texts" (Meidl et al., n.d.) shows that the co-clustering technique is effective in understanding patterns of user opinions relevant to the K-Means clustering concept. The study "What Causes Wrong Sentiment Classifications of Game Reviews?" (Viggiato et al., n.d.) also identifies key challenges in sentiment classification, such as language ambiguity and sarcasm in user reviews.

Nevertheless, the integration of Content-Based Filtering and K-Means Clustering in game recommendation systems is still rarely done, resulting in a research gap in this field. Therefore, this study aims to develop a game recommendation system by combining Content-Based Filtering and K-Means Clustering to improve the accuracy and relevance of recommendation results. This approach is in line with the latest research trends such as "Solving the Content Gap in Roblox Game Recommendations: LLM-Based Profile Generation and Reranking" (C. Wang et al., 2025), which utilizes generative artificial intelligence to improve recommendation personalization. The system proposed in this study was developed web-based using the Django framework and implemented on the Google Cloud Platform (GCP) service so that it can be accessed in real-time by users..

RESEARCH METHODS

This research uses a quantitative approach to develop a web-based game recommendation system using K-Means Clustering and Content-Based Filtering methods. The following are the steps taken in this study.

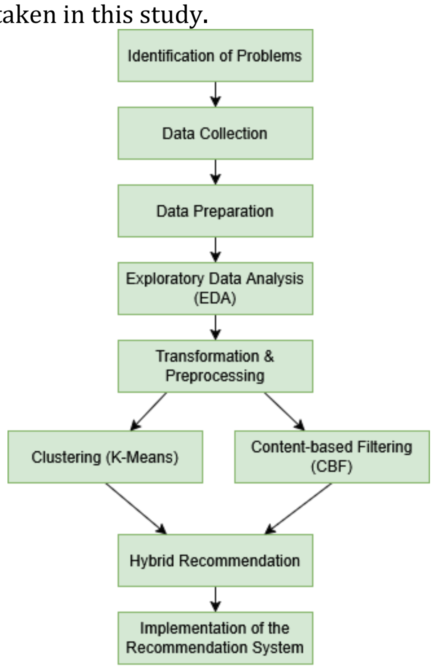


Figure 1. Research Concept

The Identification of Problems

The researcher identified the problem that necessitated a game recommendation system. With the rapid growth of the gaming industry, users often find it difficult to find games that suit their preferences. The large number of games available on various platforms makes the game selection process inefficient and time-consuming. In addition, previous studies generally only utilize one recommendation method, such as collaborative filtering or content-based filtering, without combining the two with clustering techniques. Therefore, this study focuses on developing a hybrid-based recommendation system, which combines the K-Means algorithm and content-based filtering to improve the quality of the recommendations produced.

Data Collection

After identification of the problem, the next step is data collection. The data used in this study comes from the RAWG Video Games Database API, an international game database that provides information on more than 500,000 games. From this source, researchers collected 1,000 game data with attributes considered important in determining the relevance of recommendations, namely:

a. Game name
b. Rating
c. Genre
d. Platform
e. Age category (ESRB)

The data collection process was carried out using the Python programming language in Google Colaboratory. The extracted data was then saved in Comma Separated Values (CSV) format for easy processing in the next stage.

Data Preparation

The Data Preparation stage is carried out to ensure that the dataset is clean and ready for use in modeling. This process involves:

1. Checking data completeness, which is ensuring that all necessary attributes are available for each entry.
2. Identifying and removing duplicate data, so that there is no bias in the analysis.
3. Selecting relevant attributes, by only retaining attributes that really influence the recommendation process.

Thus, the dataset used in this study is truly focused on important features that support the recommendation process.

Exploratory Data Analysis (EDA)

This stage aims to understand the characteristics of the dataset in greater depth by exploring data distribution patterns. The analysis conducted includes:

1. Game rating distribution, which shows that most games have ratings in the range of 3.0–4.5.
2. Age category distribution (ESRB), where the majority of games fall into the Mature category.
3. Game genre distribution, which shows the dominance of the Action, Adventure, and Indie genres.
4. Game platform distribution, with PC as the dominant platform, followed by Android and PlayStation 4.

The results of the data exploration are visualized in the form of histograms, bar charts, and distribution graphs. This EDA is important because it provides an overview of the dataset and supports decision making in the preprocessing and modeling stages.

Transformasi & Preprocessing

The Transformation and preprocessing are performed to clean the data and change its format to suit the algorithm's requirements. The steps taken are as follows:

1. Handling missing values: numerical attributes such as ratings are filled with median values, while categorical attributes such as Genre, platform, and ESRB are filled with modes.
2. Removal of duplicate data to prevent redundancy.
3. Standardization of numerical data (ratings): performed using StandardScaler so that the values are within the same range.
4. Encoding of categorical attributes: ESRB is converted to numerical form using Label Encoding, while Genre and platform are converted using multi-hot encoding because one game can have more than one category.

This stage is important for producing clean, uniform data that can be better processed by Clustering and content-based filtering algorithms.

Clustering (K-Means)

Clustering is used to group games based on similar attributes. The algorithm used in this study is K-Means, which divides the dataset into several clusters according to the specified K value. The optimal number of clusters is determined using the Elbow and Silhouette Score methods, which then produce a value of K = 3 as the best number of clusters.

The K-Means process begins by calculating the distance between each data point and all initial centroids using the Euclidean Distance formula:

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^{n} (x_{ik} - c_{jk})^2}$$

Where x_I represents the i-th data point, c_j is the centroid of cluster *j* and *n* is the number of features. Each data point is assigned to the cluster with the smallest distance. After the assignment step, the centroid is updated by calculating the mean of all data points within the cluster using:

$$c_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i$$

This process is repeated iteratively until the centroids converge or no significant changes occur.

The clustering results show a relatively balanced division of the dataset, with three clusters each representing a group of games with specific characteristics. To further clarify the results, visualization is performed using the UMAP and t-SNE algorithms, allowing the distribution of clusters to be observed visually in two-dimensional space.

Content-based Filtering (CBF)

Content-based Content-based filtering is used to generate game recommendations based on content similarity. This process is carried out by calculating the cosine similarity values between games based on the processed attributes such as **Genre, Platform, Rating**, and **ESRB**. The similarity between two games A and B is computed using the cosine similarity formula:

$$\text{CosineSimilarity}(A, B) = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\ \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Where:

1. A_I and B_i represent the feature values of game A and game B, and
2. *n* is the total number of features used.

A higher cosine similarity value indicates that two games share more similar content characteristics.

For example, when a user selects Grand Theft Auto V (GTA V), the system calculates its similarity with all other games. Those with the highest similarity scores—such as Skyrim, Assassin's Creed, and Elden Ring—are then recommended. This ensures that users receive recommendations aligned with their preferences based on the content of the selected game.For example, when a user selects Grand Theft Auto V (GTA V), the system will recommend games with a high level of similarity, such as Skyrim, Assassin's Creed, and Elden Ring. This way, users can obtain recommendations that match their preferences based on the content of the selected game.

Hybrid Recommendation

The hybrid recommendation approach is implemented by combining the results of K-Means Clustering and content-based filtering. The mechanism used is to limit the recommendation search only to games that are in the same cluster, then calculate the similarity value using cosine similarity.

This way, the system does not only rely on content similarities between games, but also considers similarities within clusters. This makes recommendations more relevant, varied, and tailored to user preferences.

Implementation of the Recommendation System

The latest stage of the research is the implementation of the recommendation system into a web-based application. The system was built using the Django framework that supports Python-based development. The main features of the system include:

1. User authentication (login and registration).
2. Navigation based on genre, platform, rating, and ESRB.
3. Game detail page that displays complete information.
4. Game recommendation features using clustering, content-based filtering, and hybrid approaches.

The completed system was then deployed to Google Cloud Platform (GCP) so that it could be accessed in real time via the internet. Thus, the results of the research are not only theoretical but can also be used practically by users. j

RESULTS AND DISCUSSION

Data Collection

Data collection in this study was conducted using secondary data obtained from the RAWG Video Games Database API, an international digital game information provider platform.

Table 1. Result Data Collection

ID	Nama	Release	Rating	*Genre*	...
13536	Portal	2007-10-09	4.49	Action, Puzzle	...
13537	Half-Life 2	2004-11-16	4.48	Action, Shooter	...
802	Borderlands 2	2012-09-18	4.01	Action, Shooter, RPG	...

Based on Table 1 above, examples of data collection results obtained through the use of the Application Programming Interface (API) from the RAWG website are shown. This data collection process was carried out in a structured manner, resulting in a dataset of 1,000 games with various relevant attributes, including game ID, name, release year, rating, genre, platform, store, and supporting tags. The data obtained has a fairly high level of diversity because it covers various types of genres, age categories, and different platforms. Thus, this dataset can be used as the main foundation in the process of analysis, exploration, and development of a more representative content-based recommendation system.

Exploratory Data Analysis (EDA)

Data exploration was conducted to understand the characteristics of the digital game dataset in order to identify data distribution patterns and determine the important features to be used in the content-based recommendation system. The following are the steps in the EDA process:

1. Game Rating Distribution

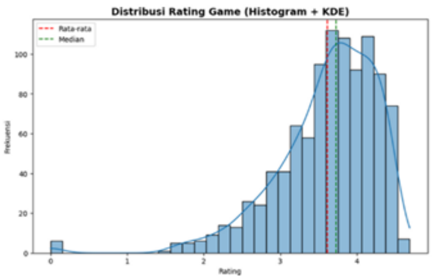


Figure 2. Rating Distribution Visualization

Figure 2 shows the distribution of ratings in the form of a histogram supplemented with a KDE curve to clarify the data distribution pattern. Most games have ratings in the range of 3.0–4.5, indicating a tendency toward high scores. The dotted red line indicates the mean, while the green line indicates the median. The proximity of the two indicates a relatively symmetrical distribution with a slight skew to the left, but still depicts a good and normal distribution of ratings.

2. Game Distribution Based on ESRB

After determining the rating distribution, the next step is to analyze the ESRB (Entertainment Software Rating Board) distribution. ESRB attributes provide important information about the recommended age limit for playing a game.

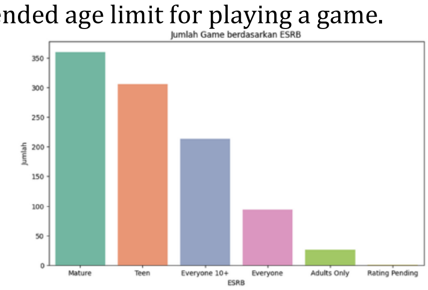


Figure 3. ESRB Distribution Visualization

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i1.444
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

Figure 3 shows the distribution of ESRB categories, with the majority of games classified as "Mature," followed by 'Teen' and "Everyone 10+." Very few games are classified as "Adults Only," while "Rating Pending" is almost non-existent. This information is important for filtering recommendations according to the age group of users.

3. Distribution of the Most Popular Game Genres

Genre is the attribute that best reflects users' personal preferences.

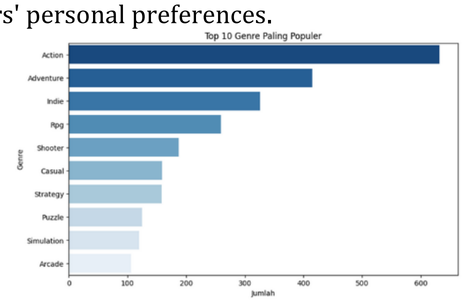


Figure 4. Visualization of the Most Popular Game Genres

Figure 4 shows the ten most popular game genres in the dataset. The Action genre dominates, followed by Adventure and Indie, which emphasize narrative exploration and the creativity of independent developers. RPG and Shooter also occupy significant positions, reflecting interest in deep characterization and shooting action. Meanwhile, Casual, Strategy, and Puzzle are in the middle range, while Simulation and Arcade close out the top ten list, though they still have a relevant following.

4. Distribution of the Most Popular Gaming Platforms

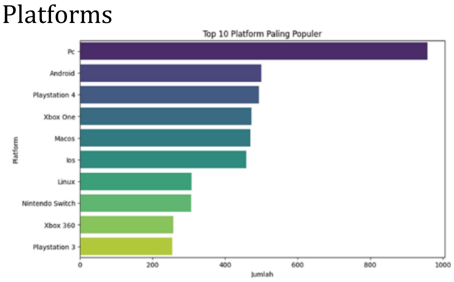


Figure 5. Visualization of the Most Popular Gaming Platforms

Figure 5 shows the ten most popular platforms in the dataset. PCs dominate with the largest number of games, confirming their flexibility in supporting various genres. The next positions are occupied by Android, PlayStation 4, and Xbox One, reflecting the dominance of mobile and modern consoles. Platforms such as MacOS, iOS, and Linux are also quite significant, while Nintendo Switch, Xbox 360, and PlayStation 3 are at the bottom of the list, yet still play an important role. This distribution emphasizes the importance of platform attributes in recommendations, as device compatibility greatly influences user preferences.

Preprocessing Data

The preprocessing stage is crucial in the development of a recommendation system, as it aims to clean, organize, and prepare raw data into a format that can be used by machine learning algorithms. In this study, preprocessing was carried out in three main stages, namely:

Selection of Attributes

The first step in this stage involves selecting relevant attributes to support the development of the recommendation system. Not all of the many attributes available in the RAWG dataset are used, as some attributes do not contribute significantly to the analysis process.

Table 2. Attribute Selection Results

Name	Rating	Genres	Platforms	ESRB
Portal	4.49	Action, Puzzle	macOS, PC, Android, PlayStation 3	Teen
Half-Life 2	4.48	Action, Shooter	PC, macOS, Xbox 360, Linux, Xbox, Android	Mature
Borderlands 2	4.01	Action, Shooter, RPG	PlayStation 3, macOS, PC, Android, Linux	Mature

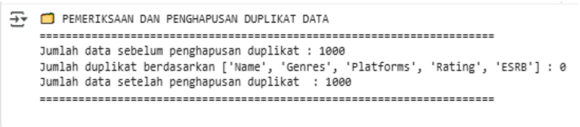
Table 4.2 shows the results of the research attribute selection, namely Name, Rating, Genres, Platforms, and ESRB. These attributes were retained because they describe the important characteristics of each game. For example, Portal, with a rating of 4.49, is in the Action and Puzzle genres, is available on various platforms, and is classified as Teen. Half-Life 2 has a rating of 4.48, belongs to the Action and Shooter genres on several platforms, and is classified as Mature. Borderlands

P-ISSN: 2656-1743 |E-ISSN: 2656-1735 JURNAL RISET INFORMATIKA
DOI: https://doi.org/10.34288/jri.v8i1.444 Vol. 8, No. 1. December 2025
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

2 has a rating of 4.01, belongs to the Action, Shooter, and RPG genres, and is also classified as Mature. With this selection, the data becomes more focused on features relevant to analysis and recommendation system development..

Handling Missing Values and Duplication

This step aims to ensure data quality prior to analysis. The dataset from the RAWG API still contains deficiencies, such as missing values and duplicates. To address this, a thorough check is performed: incomplete data is deleted, empty values in numerical attributes are imputed with the median, while in categorical attributes they are replaced with the mode. Duplicate data is also deleted to avoid redundancy. The result is a cleaner, more consistent dataset that is ready for use in the next stage of analysis.



```
PEMERIKSAAN DAN PENGHAPUSAN DUPLIKAT DATA
======================================================================
Jumlah data sebelum penghapusan duplikat : 1000
Jumlah duplikat berdasarkan ['Name', 'Genres', 'Platforms', 'Rating', 'ESRB'] : 0
Jumlah data setelah penghapusan duplikat  : 1000
======================================================================
```

Figure 6. Results of Missing Values and Duplicates

Data Transformation

The data undergoes a transformation stage to meet the requirements of the machine learning algorithm. Numeric Rating features are standardized using StandardScaler, while categorical ESRB features are converted to numeric form with one-hot encoding. For Genres and Platforms attributes that have many values in a single entry, multi-hot encoding is used through MultiLabelBinarizer. The result of this transformation is a binary representation that makes it easier for the model to recognize preference patterns based on Genre and platform.

Scale Transformation in Rating

In this research, normalization was performed using the StandardScaler method from the scikit-learn library, which changes the rating distribution to have a mean close to zero and a standard deviation of one.

Table 3. StandardScaler Rating

Name	*Rating*	*Rating_scaled*
Portal	4.49	1.318571
Half-Life 2	4.48	1.303528
Borderlands	4.01	0.596520
Life Is Strange	4.12	0.761990
Limbo	4.14	0.792075

Table 3 shows the normalization results for the top five games. Games with high ratings, such as Portal and Half-Life 2, have Rating_scaled values above 1, while games with ratings close to the average produce lower values. This indicates that the scaling process works well, maintaining the proportions between ratings but on a more controlled scale that is ready for use in the model.

ESRB Rating Conversion to Numerical Ratings

Following normalization of numerical attributes, categorical features such as ESRB are also transformed so they can be processed numerically. The transformation is performed using label encoding techniques, which convert each category into a unique number so it can be recognized by machine learning algorithms.

Table 4. ESRB's Transformation to Numerical Ratings

Name	*ESRB*	*ESRB*_Cat
Portal	Teen	5
Half-Life 2	Mature	3
Borderlands	Mature	3
Life Is Strange	Mature	3
Limbo	Teen	5

Table 4 shows the results of converting ESRB labels, for example Teen to 5 and Mature to 3. These numbers only serve as unique identifiers without representing any particular order or weight. This encoding process is important so that algorithms such as K-Means and cosine similarity can process ESRB attributes along with other numerical features.

Multi-hot encoding for Genre

In additions to ESRB, the Genres attribute includes a multi-category feature because one game can have more than one Genre, such as Action and Adventure. To process this, a multi-hot encoding technique is used that converts each Genre into a binary column. Each column represents one Genre, with a value of 1 indicating that the Genre is owned by a particular game, while a value of 0 indicates the opposite..

Table 5. Multi-hot encoding for Genres

Action	*Adventure*	*Arcade*	*Board Games*	*Card*	**....**
1	0	0	0	0	**....**
1	0	0	0	0	**....**
1	1	0	0	0	**....**

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i1.444
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

0	1	0	0	0	....
1	1	0	0	0	....

Table 5 shows the results of multi-hot encoding, where the first row represents a single genre game (Adventure), while the following rows reflect genre combinations such as Adventure and Action. This technique enables flexible multi-label numerical representation for use in recommendation systems..

Multi-hot encoding for Platforms

Platform attributes are also transformed using multi-hot encoding to represent multiplatform games. Each platform, such as Android, Apple II, Atari 2600, and Classic Macintosh, is converted into a binary column. A value of 1 indicates that the game is available on that platform, while 0 means it is not available..

Table 6. Multi-hot encoding for Platform

3DO	Android	Apple II	Atari 2600	Atari 5200	....
0	1	0	0	0	....
0	1	0	0	0	....
0	1	0	0	0	....
0	1	0	0	0	....
0	1	0	0	0	....

Table 6 shows the encoding results of the first five games with a total of 47 identified platforms. This representation enriches the features and allows the system to provide recommendations that are not only similar in terms of content, but also compatible with the user's device.

Clustering with K-Means

The clustering process aims to find similarities between games by grouping them into several clusters as the basis for the recommendation system. The main stages in this process include:

1. Featuring Clustering Preparation
   The initial stage of clustering is to prepare relevant numerical features. This study uses the transformation of Rating, ESRB, Genres, and Platforms attributes in numerical form, while descriptive attributes such as Name, Genres (string), Platforms (string), and ESRB (string) are excluded. Rating is standardized with StandardScaler, ESRB is converted to a numerical category, while Genres and Platforms are transformed with MultiLabelBinarizer into multi-hot encoding that represents the presence of Genre or platform in each game.
2. Cluster Number Determination Process
   Determining the number of clusters (K) is very important in K-Means because it affects the quality of data segmentation. In this study, the optimal K value was determined by combining two evaluation methods, namely the Elbow Method and Silhouette Score.

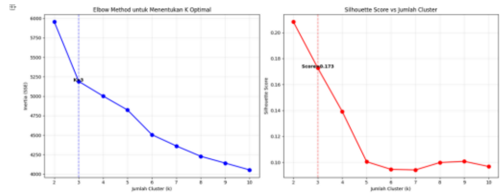


Figure 7. Elbow and Silhouette Score for Determining K

3. Implementation of K-Means Clustering
   The K-Means model iteratively clusters game data into three clusters based on feature similarity. The result of this process is cluster labels that are added to the dataset as a new column named Cluster, which indicates each game's membership to a particular cluster.
4. Cluster Balance Analysis
   Cluster balance analysis was performed to assess the distribution of data in each cluster.

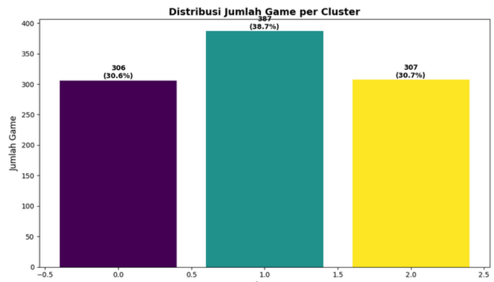


Figure 8. Distribution of Number of Games per Cluster

5. Visualization of Clustering Results
   To understand the distribution of clustering results visually, visualization was performed using two dimension reduction techniques, namely UMAP and t-SNE.

a. Additional visualization was performed using UMAP (Uniform Manifold Approximation and Projection), a non-linear dimension reduction technique that aims to map high-dimensional data into a

Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

two-dimensional space while preserving the global and local structure of the data.

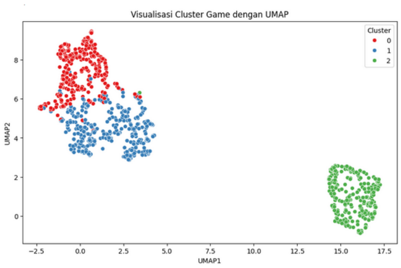


Figure 9. Game Cluster Visualization with UMAP

b. Visualization of clustering results using the t-SNE (t-Distributed Stochastic Neighbor Embedding) approach in Figure 11 shows a two-dimensional mapping of game data after the K-Means process. The t-SNE technique was used because of its ability to capture non-linear relationships between high-dimensional data while maintaining the local structure between data points.

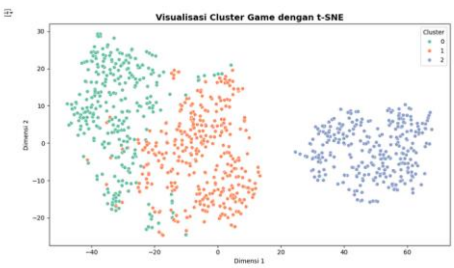


Figure 10. Visualization of Game Clusters with t-SNE

6. Game Recommendations Based on Clusters

The cluster-based recommendation system works on the principle that games in the same cluster have similar content characteristics, such as genre, platform, rating, and ESRB classification. Therefore, when a user likes a game, the system will recommend other games from the same cluster in the hope that they will have similar preferences.

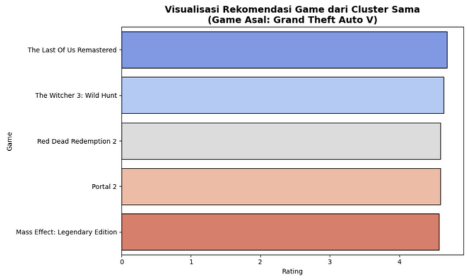


Figure 11. Visualization of Game Recommendations from the Same Cluster

The visualization in Figure 11 displays game recommendations generated based on clusters similar to Grand Theft Auto V. This horizontal bar chart presents the top five games from Cluster 1 that share similarities with the original game in terms of genre, platform, and rating.

Recommendations Based on Content-Based Filtering

The Content-Based Filtering method recommends games based on feature similarity. The system compares attributes such as genre, platform, rating, and ESRB to find games that are most similar to the reference game. The more similarities there are in these attributes, the higher the similarity score obtained. The following shows the recommendation results generated by this approach:

1. Recommendation Results with Content-based Filtering

Table 7. Recommendations for Content-based Filtering for Grand Theft Auto V

No	*Name*	*Rating*	*Genres*	*Platforms*	*ESRB*	*Similarity*
1	The Elder Scrolls V: Skyrim	4.42	Action, Rpg	*PlayStation*	Mature	0.916
2	Assassin'S Creed Revelations	3.99	*Action*	*PlayStation*	Mature	0.875
3	Assassin'S Creed Ii	4.42	*Action*	*PlayStation*	Mature	0.852
4	Elden Ring	4.40	Action, Rpg	*PlayStation*	Mature	0.852

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i1.444
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

5	Assassin'S Creed Brotherhood	4.27	*Action*	*PlayStation*	Mature	0.846

The recommendation results show five games with the highest similarity scores to Grand Theft Auto V. Some of them are The Elder Scrolls V: Skyrim, Assassin's Creed Revelations, and Elden Ring, which are recommended because they have similar genres (Action, RPG), multiplatform support, and high ratings. The similarity score indicates the level of closeness to the reference game, where a score close to 1 indicates a very strong similarity.

2. Recommendation Visualization with Content-based Filtering

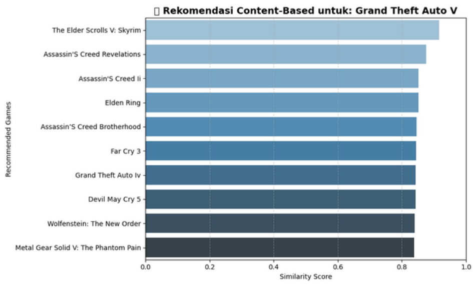


Figure 12. Visualization of Content-based Filtering Recommendations.

Figure 12 shows a visualization of game recommendation results using Grand Theft Auto V as a reference. The horizontal bar chart displays ten games with the highest similarity scores based on cosine similarity in terms of genre, platform, rating, and ESRB. The Elder Scrolls V: Skyrim, Assassin's Creed Revelations, and Elden Ring are at the top of the list with scores close to 1.0, indicating a high similarity to GTA V. The darker the bar color, the lower the similarity score, while also representing the ranking order of feature proximity.

Hybrid Recommendation

The hybrid recommendation method in this study combines clustering-based filtering and content-based filtering to improve the quality and relevance of recommendations. The process begins with the application of the K-Means algorithm to divide the game dataset into several clusters based on feature similarity. Once the target game is known to be in a particular cluster, recommendations are then focused only on games in the same cluster.

Next, the content-based filtering method is applied locally to the cluster using cosine similarity to identify the most similar games in terms of content. This combined approach not only considers the similarity of content between games, but also reduces global bias by limiting the recommendation space to relevant segments. Thus, the system is able to provide more specific, personalized, and contextually appropriate recommendations for groups of similar games.

Deployment

The system implementation was carried out through a local deployment phase using a Django development server. The system was run locally at http://127.0.0.1:8000/, enabling direct testing of all designed user interface and backend functions. The following are the steps in implementing the recommendations.

1. User Authentication
a. Login Page

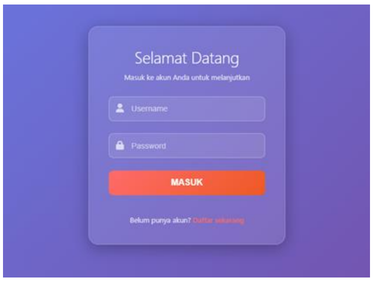


Figure 13. Login Page

The login page is designed to be simple and minimalist, with a focus on the authentication form. In the image above, you can see two input fields for entering your username and password, as well as a button to log in to the system. This page also displays a link to the registration page for new users.

b. Registration page

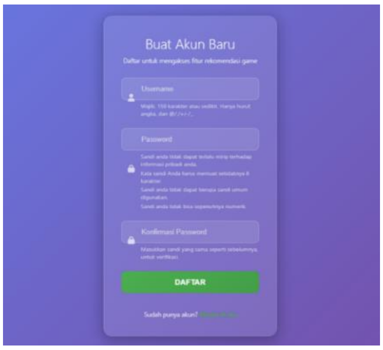


Figure 14. Registration Page

Figure 14 above shows the registration form consisting of several fields such as username, password, and password confirmation. The design is neat and easy to understand so that users can easily register a new account.

2. Main System Navigation
a. Dashboard Page

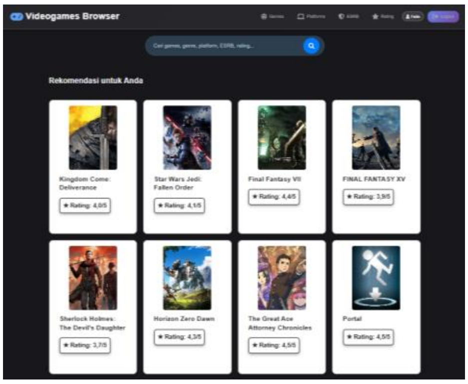


Figure 15. User Dashboard

The main page display consists of a horizontal navigation menu at the top, a search bar in the middle, and a responsive list of games. Users can use the Genre, Platform, ESRB, and Rating menus to filter the game display according to their needs.

b. Genre Page

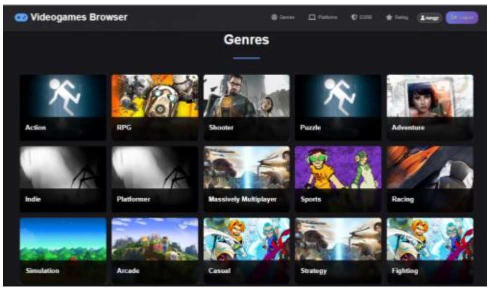


Figure 16. Genre Page

Figure 16 above shows the Genre interface in the form of a grid with background images and Genre names superimposed on top of them. This design makes it easy for users to recognize and select game categories according to their interests.

c. Platform Page

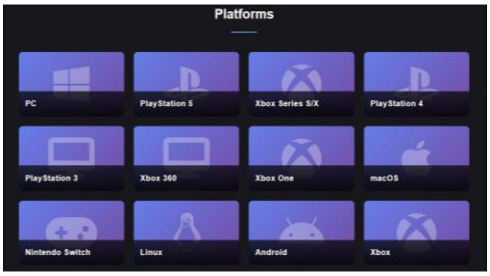


Figure 17. Platform Page

The platform page displays icons and names of various available platforms. After selecting a platform, users will be directed to a list of games that support that platform.

d. ESRB Page

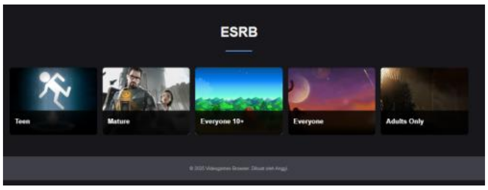


Figure 18. ESRB Page

Figure 18 above shows the ESRB page with an attractive visual display. Each card contains the ESRB logo, category name, and number of games included in that rating. This design helps users more easily filter games according to age restrictions.

P-ISSN: 2656-1743 |E-ISSN: 2656-1735
DOI: https://doi.org/10.34288/jri.v8i1.444
Accredited rank 4 (SINTA 4), excerpts from the decision of the DITJEN DIKTIRISTEK No. 230/E/KPT/2023

e. Rating Page

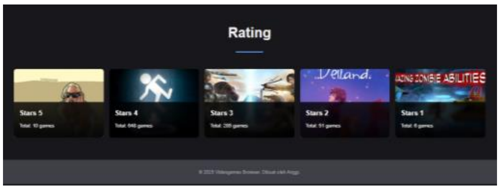


Figure 19. Rating Page

The rating page in Figure 19 above is designed using a neat and responsive grid layout. Users can select a specific rating category, and the system will display a list of games that match that category, from highest to lowest based on their rating scores.

3. Recommendation Features and Game Details

a. Game Details Page

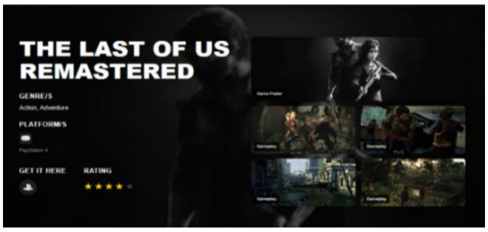


Figure 20. Game Detail Page

The game details page is designed to be informative and easy to read, with large images of the game followed by descriptive details and other supplementary information. Users can also see which platforms support the game.

b. Hybrid Recommendation Page

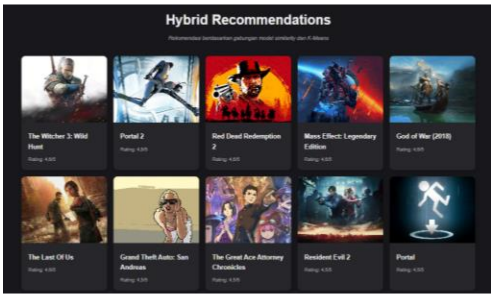


Figure 21. Hybrid Recommendation Page

Figure 21. above shows a list of recommended games arranged visually and interactively. Users can directly explore recommended games without having to perform a manual search.

CONCLUSIONS AND SUGGESTIONS

Conclusion

Based on the results of the research and the implementation of the game recommendation system using the K-Means Clustering and Content-Based Filtering methods, it can be concluded that the system was successfully developed by combining clustering and content-based approaches, enabling the generation of more contextual and relevant recommendations. The dataset consisting of 1,000 game entries obtained from the RAWG API—containing attributes such as genre, platform, rating, and age category (ESRB)—was effectively utilized in the modeling process. The K-Means algorithm was able to group games into homogeneous clusters, which were further visualized using t-SNE and UMAP to provide a clearer understanding of the data distribution. In addition, the Content-Based Filtering method using cosine similarity proved effective in recommending games with similar content characteristics. This hybrid approach improved the efficiency and accuracy of the recommendation process by narrowing the search space to clusters relevant to the selected game. Finally, the system was successfully deployed as a web application using Django and Google Cloud Platform, making it easily accessible for end users.

Suggestion

The recommendations for further research are as follows:

a. Add quantitative evaluation methods such as precision, recall, and F1-score to measure the quality of recommendations more objectively.

b. Use larger datasets and utilize real-time user feedback to improve the accuracy and relevance of recommendations.

c. Integrate historical user data (if available) so that the system can be developed towards more personalized collaborative filtering or hybrid systems.

d. Apply a similar approach to other digital products, such as movies, music, or books, with feature adjustments according to the characteristics of the domain..

REFERENCES

Agustina, N., Citra, D. H., Purnama, W., Nisa, C., & Kurnia, A. R. (2022). Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan

Shopee pada Google Play Store. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *2*(1), 47–54.

Chythanya N*, K., Bhargavi Y, K., & Rohan, B. (2019). A Novel Video Game Recommender System using Content Based Filtering -Vidya. *International Journal of Recent Technology and Engineering (IJRTE)*, *8*(4), 536–541.

Dihni, V. A. (2022). *Jumlah Gamers Indonesia Terbanyak Ketiga di Dunia*. Databoks. Diakses Tanggal 18 Agustus 2025 from https://databoks.katadata.co.id/teknologi-telekomunikasi/statistik/950b8ba78451f97/jumlah-gamers-indonesia-terbanyak-ketiga-di-dunia

Fitri, E. M., Suryono, R. R., & Wantoro, A. (2023). Klasterisasi Data Penjualan Berdasarkan Wilayah Menggunakan Metode K-Means Pada Pt Xyz. *Jurnal Komputasi*, *11*(2), 157–168.

Liu, K., Ma, J., Feng, S., Zhang, H., & Zhang, Z. (2023). *DRGame: Diversified Recommendation for Multi-category Video Games with Balanced Implicit Preferences*.

Meidl, M., Lytinen, S., & Raison, K. (n.d.). *Using Game Reviews to Recommend Games*. www.aaai.org

Nurdy, A. H., Rahim, A., & Arbansyah. (2024). Analisis Sentimen Ulasan Game Stumble Guys Pada Playstore Menggunakan Algoritma Naïve Bayes. *Teknika*, *13*(3), 388–395.

Pragusma, Z. B., Kwandinata, V., Natthannael, J., Meiliana, & Achmad, S. (2023). Game Recommendation Using Content-based Algorithm Title. *2023 International Conference on Informatics, Multimedia, Cyber and Information Systems, ICIMCIS 2023*, 313–317.

Rochmad Wahono, A., Aji Saputra, B., & Fadlu Rahman, F. (2024). Sistem Rekomendasi Film Menggunakan Metode Content-Based Filtering dan Algoritma K-Nearest Neighbors (KNN). *Prosiding Seminar Nasional Teknologi Informasi Dan Bisnis*, 1–6.

Safitri, J., Atina, V., & Sudibyo, N. A. (2024). *Rancang bangun sistem rekomendasi pemilihan drama korea dengan metode content-based filtering Design of a korean drama selection recommendation system using the content-based filtering method*. *5*, 175–189.

Viggiato, M., Lin, D., Hindle, A., & Bezemer, P. (n.d.). *What Causes Wrong Sentiment Classifications of Game Reviews?* https://github.com/asgaardlab/sentiment-analysis-Steam

Viljanen, M., Vahlo, J., Koponen, A., & Pahikkala, T. (2020). *Content Based Player and Game Interaction Model for Game Recommendation in the Cold Start setting*.

Wang, C., Wei, X., Jiang, Y., Ong, F., Gao, K., Yu, X., Hui, Z., Yoon, S., Yu, P., & Gong, M. (2025). *Solving the Content Gap in Roblox Game Recommendations: LLM-Based Profile Generation and Reranking*.

Wang, Y., Chu, C., & Li, K. (2024). Research on Game Recommendation Algorithms Based on Hybrid Models. *ACM International Conference Proceeding Series*, 267–274.

Widaraeni, F. S., & Vivianti. (2021). Tematik : Jurnal Teknologi Informasi Komunikasi (e-Journal) Vol. 8 No. 2 Desember 2021. *Tematik : Jurnal Teknologi Informasi Komunikasi (e-Journal)*, *8*(2), 160–175.

Widya, A., Gaffar, M., & Samsul, S. W. (2025). *Sistem Rekomendasi Pemilihan Hardware Komputer Menggunakan Metode Item-Based Collaborative Filtering*. *14*(April), 284–292.

Yuan, L., Ding, M., Meng, F., & Tian, Y. (2025). *Sentiment Analysis and Rating Video Game Dimensions via NLP*.

Yudhistira, A., & Andika, R. (2023). Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering. *Journal of Artificial Intelligence and Technology Information (JAITI)*, *1*(1), 20–28.