Enhancing Gaming Experience: A Comparative Study of Three K-Nearest Neighbors Variants in Game Recommendation

Yoon-Teck Bau 1, a), Nurul Ezzatie Mohd Norazizan 1, b) and Kian-Chin Lee1, c)

1Faculty of Computing and Informatics, Multimedia University, Persiaran Multimedia, 63100 Cyberjaya, Selangor, Malaysia.

b) Corresponding author: nezzatie254@gmail.com  
a) ytbau@mmu.edu.my   
c) kclee@mmu.edu.my

**Abstract.** In crowded online gaming markets, gamers are often overwhelmed by the sheer number of titles available to choose from and play. To address this, game publishers are increasingly using recommendation systems to develop games that align with user preferences, boost player engagement, and drive higher sales. The objective of this research is to implement a game recommendation that leverages machine learning techniques to provide game recommendation results to a gamer. Within the methodological construct lies machine learning techniques that study and analyze the preferences of the user, characteristics of games, and personalize the recommendation results to gamers. The data source for this research is acquired from data.world that provides an entire dataset entailing 200,000 rows along features of user identifications, game names, user behavior actions, and hours played. The performance of the recommendation is evaluated using three metrics, which mean absolute error, root mean squared error and mean squared error. More specifically, the machine learning techniques for improving the recommendation accuracy comprise of three k-nearest neighbor variants, which are k-nearest neighbors basic, k-nearest neighbors with mean, and k-nearest neighbors with z-score. K-nearest neighbors with means technique produces the lowest mean absolute error. K-nearest neighbors basic technique produces the lowest mean squared error and root mean squared error. The best technique is k-nearest neighbors basic for two out of three performance metrics. The study further extends by choosing this best technique for personalized recommendation results within the gaming domain.

# INTRODUCTION

In the ever-evolving landscape of the electronic gaming industry, both gamers and game developers face a common challenge, overwhelming abundance of game options. With many electronic games released annually across various platforms, from mobile to console to personal computer, selecting the right game that aligns with a player's unique preferences has become an increasingly complex task. This saturation has made the discovery of relevant and enjoyable games more difficult for gamers and intensified the competition for visibility among game developers.

To address this challenge, game recommendation systems have emerged as essential tools. These systems leverage machine learning techniques to analyze gamer data, gamer preferences, gamer behavior, gaming history, gameplay patterns, and game metadata, enabling them to streamline the game discovery process tailored to individual players. By doing so, recommendation systems not only enhance the gaming experience for gamers through personalized suggestions but also provide benefits to game developers and game publishers by improving customer targeting, understanding gamer behavior and market segmentation.

In particular, machine learning based recommendation algorithms utilize a variety of techniques to process large datasets related to gamer activity, preferences, and histories. Among these, the k-nearest neighbors algorithm is a widely used method for recommendation systems. K-nearest neighbors can deliver effective personalized recommendations by identifying similar users or games based on specific metrics. Such recommendation algorithm contributes to enhanced gamer engagement, improved user satisfaction, and ultimately, drove higher game sales.

The primary objective of this research is to develop a personalized game recommendation using three variants of the k-nearest neighbors algorithm. This study aims to evaluate and compare the performance of these three techniques to identify the most effective technique for generating accurate and relevant game recommendation results.

# LITERATURE REVIEW

The game industry has developed exponentially within the last ten years to give users genres, platforms, and types of games. It is apparent that increasing the number of available games makes it more difficult for users to identify which games they will enjoy the most. Game recommendation systems are designed to help users discover games that align with their preferences [1], [2], [3]. Recommendation systems, thus, use intelligent systems and techniques to suggest games that would interest the users [4], [5]. Three categories of recommendation techniques, which are content-based, collaborative filtering, and hybrid.

Collaborative filtering is one of the strategies that can be used to recommend games. This is a popular technique for recommending games based on the preferences of users with similar tastes [6], [7], [8], [9], [10], [11], [12]. It works by analyzing user-item interaction data to find patterns of similarity [13], [14]. For example, if user A likes games that user B also enjoys, the system might recommend games that user B has played but user A has not [9].

Moreover, the content based technique recommends games that are similar in content to those a user has previously enjoyed before [4]. It analyzes gaming attributes such as game name, rating, category, as well as time spent playing. For instance, if the person has been playing, or has enjoyed playing, in the category of action-adventures, then that system will make recommendations concerning other games as only being relevant via the genre they fall into or the mechanics involved in playing the game [9].

The hybrid recommendation method combines such typical features of collaborative filtering and content based filtering, thereby exploiting the strengths of both approaches to recommendations [4], [13]. Their systems have the aim of increasing the accuracy and diversity of their recommendations. One is to conduct collaborative filtering and the content based filtering separately to compile different lists and then merge the lists. Another example is when collaborative filtering predicts, while in parallel retaining a content based profile for each user [10] .

The field of game recommendation systems faces several significant research gaps that hinder progress and innovation. The lack of a common performance metric for game recommendation systems makes it difficult to compare the effectiveness of different techniques [5], [10] . The absence of standardized metrics makes it difficult to assess whether a recommendation system truly enhances player satisfaction, as different metrics may prioritize different aspects of the recommendation process. Another study that compared different techniques for collaborative filtering, including alternating least squares, singular value decomposition and non-negative matrix factorization, was limited in scope due to the smaller dataset used in the study [8] . The study acknowledged that this limited the ability to fully assess the performance of the techniques.

Additionally, the gaming field is under researched compared to other fields like movies or books. This is partly due to the unique characteristics of the game field such the difficulty in measuring player satisfaction. As a result, there is a need for more research on effective game recommendation techniques, particularly for those that can address the cold start problem.

# K-NEAREST NEIGHBOR VARIANT TECHNIQUES

K-nearest neighbors technique is under the supervised machine learning technique that can be used for recommendation. It is a non-parametric method, meaning it does not make assumptions about the underlying data distribution. K-nearest neighbors technique is lazy learning technique because it does not build a model explicitly from the training data. Rather, it retains the training data as they are and makes use of the information during the recommendation of the model.

K-nearest neighbors offer several advantages that make it a popular choice in machine learning. Firstly, the technique is simplicity and ease of implementation make it straightforward to understand and make it an excellent starting choice for beginners. Next, k-nearest neighbors can handle multi-class classification problems and work well with various data distributions without making strong assumptions about the underlying data. Last but not least, it is also widely used in numerous applications such as recommendation systems, image recognition, and anomaly detection. There are three variants of the k-nearest neighbors technique applied in this research. The variants are k-nearest neighbors basic, k-nearest neighbors with means and k-nearest neighbors with z-score.

## Dataset

The data acquired from the data.world, a well-known website that has a data catalog and platform to help users find data. The dataset collected in the year of 2017. The title of the dataset is Steam video game hours played. The dataset involves 200,000 samples and four features. It includes information about user identifications, game names, user behavior actions such as purchase or play, and the hours played. The dataset has four features for Steam video games. The features available on the dataset are described in Table 1.

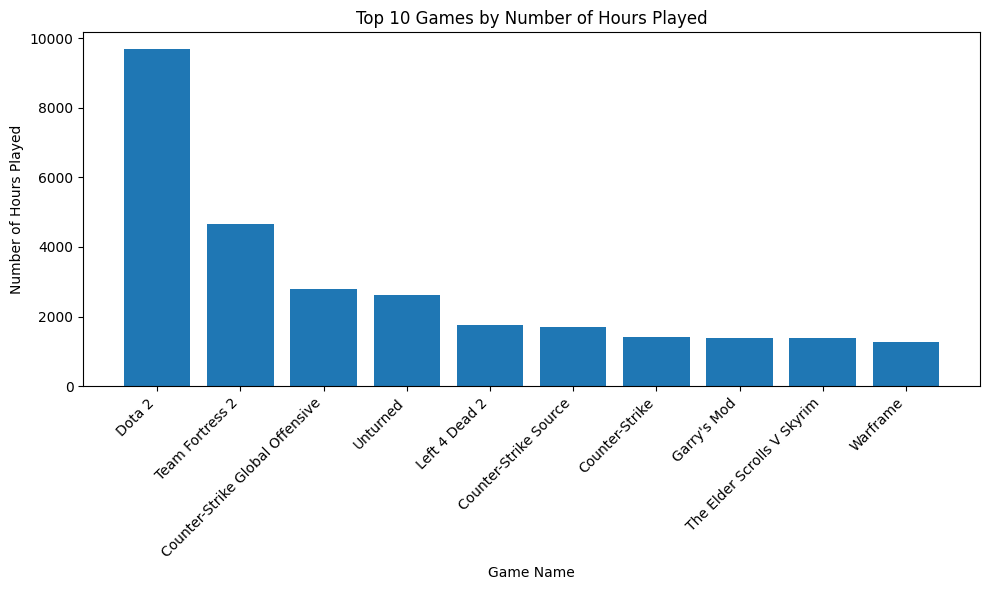
**TABLE 1.** Description of the dataset features

|  |  |
| --- | --- |
| **Feature** | **Description** |
| User identification | A unique numeric identification for each user. |
| Game name | The title of the game on Steam. |
| User behavior  Hours played | Describe user behavior action: purchase or play  If the behavior is "play," this indicates the number of hours played. If "purchase," the value is always 1.0. |

## Exploratory Data Analysis

Exploratory data analysis examines user gaming behavior through various visualizations, providing insights into gameplay trends and user engagement. Key analyses include identifying the most frequently played games, examining typical playtime distribution, and comparing purchasing patterns among users. These visualizations help uncover patterns in user interaction with games, such as preferences, activity levels, and purchasing behavior.

Figure 1 shows the top 10 games by number of hours played. The bar chart is sorted in descending orders by the number of hours played. The highest number of hours played is Dota2 game which is almost 10000 hours played, and the lowest number of hours played is below 2000 hours played is Warframe game out of top 10 games.



**Figure 1.** Bar chart of top 10 games by number of hours played

Figure 2 shows boxplot of hours for play behavior after outliers removed. The box represents the interquartile range, with the median marked as a vertical line inside the box. This boxplot highlights the central tendency and spread of hours played, excluding extreme values to provide a clearer view of typical play behavior which is approximately between 1 to 20 hours with median 4.5 hours.

## Data Preprocessing

The data preprocessing steps are taken to refine the dataset for meaningful analysis. Key actions include filtering users who played for at least two hours and retaining games played by a minimum of 10 unique users. Each game is then given rating from 1 to 5 based on average hours played. Final preprocessed dataframe contains three important features for game recommendation results that are user identification, game name and rating.

A blue rectangular object with white text

AI-generated content may be incorrect.

**Figure 2.** Boxplot of hours for play behavior without outliers

## Variant Technique 1: K-nearest Neighbor Basic

The k-nearest neighbors basic technique uses basic equation (see Equation (1)) without considering means and variances of data points. The equation below is the prediction of game item *i* rating by user *u*, in recommendation using k-nearest neighbors basic technique.

|  |  |
| --- | --- |
|  | (1) |

The recommendation process in user-based involves several keys. Firstly, the similarity sim(*u* ,*v*) between the target user *u* and potential neighbor user *v* is calculated. This similarity reflects how closely the two users align in their preferences. Next, the ratings given by the neighbors *N* from user *u* to user *v* for a specific game item *i* that are considered. Finally, the prediction rating is determined. This value represents how much the target user *u* is likely to enjoy the game item *i*, based on the similarities and ratings of their nearest neighbors. Pseudocode of recommendation technique using k-nearest neighbors basic is shown as follows.

-----------------------------------------------------------------------------------------------------------------------------

Recommendation Technique: K-Nearest Neighbors Basic

-----------------------------------------------------------------------------------------------------------------------------

1. Determine the number of neighbors (*k*).

2. Calculate the similarity between users (*u* and *v*).

3. Sort their similarity in descending order and select the top *k* users.

4. Calculate the prediction rating *r* by user *u* for game item *i* using k-nearest neighbor basic equation.

5. Return the list of recommended items for the target user (*u*).

-----------------------------------------------------------------------------------------------------------------------------

## Variant Technique 2: K-nearest Neighbor with Mean

The k-nearest neighbors with means technique uses Equation (2) that considers the mean ratings of each user. The equation below is the prediction of game item *i* rating by user *u*, in recommendation using k-nearest neighbors with means technique.

|  |  |
| --- | --- |
|  | (2) |

Two means have been added to the recommendation process in k-nearest neighbors with means compared to the first variant, k-nearest neighbors basic equation. Mean of ratings by user *u* is represented by and mean of ratings by user *v* is represented by user . Pseudocode for k-nearest neighbors with means variant is similar to the pseudocode for k-nearest neighbors basic whereby the step 4 is replaced by its own equation,

## Variant Technique 3: K-nearest Neighbor with Z-score

K-nearest neighbors with z-score technique is the third variation of the k-nearest neighbors. It uses the Equation (3) that considers z-score normalization with standard deviation of each user. The equation below is the prediction of game item *i* rating by user *u*, in recommendation using k-nearest neighbors with z-score technique.

|  |  |
| --- | --- |
|  | (3) |

Two standard deviations have been added to the recommendation process in k-nearest neighbors with z-score compared to the second variant, k-nearest neighbors with means equation. Standard deviation of ratings by user *u* is represented by and standard deviation of ratings by user *v* is represented by user . Pseudocode for k-nearest neighbors with z-score variant is similar to the pseudocode for k-nearest neighbors basic whereby the step 4 is replaced by its own equation,

# RESULTS AND DISCUSSION

Table 2 is a comparison table of three k-nearest neighbors variants and their performance results in terms of mean absolute error, mean squared error and root mean squared error. The evaluation of k-nearest neighbors techniques including k-nearest neighbors basic, k-nearest neighbors with means and k-nearest neighbors with z-score based on mean absolute error, mean squared error and root mean squared error. A recommendation model performs from different operational angles through these error evaluation metrics.

**TABLE 2.** Comparison table of three k-nearest neighbor variants and their performance results in terms of mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). The best results are in bold

|  |  |  |  |
| --- | --- | --- | --- |
| **Technique** | **MAE** | **MSE** | **RMSE** |
| K-nearest neighbors basic | 1.4544 | **2.6756** | **1.6357** |
| K-nearest neighbors with means | **1.4350** | 2.8121 | 1.6769 |
| K-nearest neighbors with z-score | 1.4359 | 2.8230 | 1.6801 |

Mean absolute error provides an average error measurement because it overlooks error directions, so each error type receives equal treatment for accurate overall performance assessment. The mean squared error calculates errors by squaring them before doing the averaging process which results in emphatic punishment of large error sizes for identifying models that generate occasional severe wrong predictions. Root mean squared error represents the square root calculation of mean squared error giving it the same measurement units found in the actual data for simpler interpretation.

The k-nearest neighbors basic model demonstrates the best performance in handling errors through both mean squared error (2.6756) and root mean squared error (1.6357) measurements. Second best performance and lowest mean absolute error of 1.4350 occurs using k-nearest neighbors with means based on the evaluated results. The errors generated by k-nearest neighbors with z-score show the least effective compared to both k-nearest neighbors basic and k-nearest neighbors with means when its values were the highest errors for both mean absolute error and root mean squared error when evaluating recommendation outcomes in this case.

K-nearest neighbors basic delivers better performance compared to k-nearest neighbors with means and k-nearest neighbors with z-score since it achieves better results on both mean squared error and root mean squared error by achieving average error reduction. Therefore, even though the mean absolute error is marginally higher for k-nearest neighbors basic, its ability to limit the severity of larger errors makes it the overall best performing technique in this comparison. Among the three k-nearest neighbors variants, the best performing technique which is k-nearest neighbors basic is applied for personalization of game recommendation results.

Figure 3 shows the personalized game recommendations for two different users. The game recommendations generated using the k-nearest neighbor basic technique. Top 10 list for the first user includes Brothers - A Tale of Two Sons, Car Mechanic Simulator 2014, Valiant Hearts The Great War, Call of Juarez The Cartel, Botanicula, Serious Sam HD The First, Encounter, Deadlight, The Wolf Among Us, Long Live The Queen and SOMA. These first user suggestions reflect the user's likely preference for immersive single player experiences. Top 10 list for the second user includes Construction-Simulator 2015, Dungeons & Dragons Online, F.E.A.R. Online, Football Manager 2010, Football Manager 2012, Football Manager 2013, Grey Goo, Poly Bridge, S.K.I.L.L. – Special Force 2, and Stronghold HD. These second user suggestions indicate the user's preference for simulation, strategy, and multiplayer experiences.

A screenshot of a computer

AI-generated content may be incorrect. A screenshot of a computer

AI-generated content may be incorrect.

**Figure 3.** Personalized game recommendation results for two different users

# CONCLUSION

In conclusion, throughout the phases, this research implemented a game recommendation results using best k-nearest neighbor among three variant techniques. The research began with a literature review of various recommendation techniques. Following that, a dataset comprising approximately 200,000 samples was acquired from data.world. This dataset included user identifications, game titles, user play behavior, user purchase behavior, and hours played. Exploratory data analysis played a crucial role in understanding the structure and dynamics of user-game interactions. A boxplot of hours played showed that most users played games for 1 to 20 hours with median 4.5 hours. Additionally, Dota 2 accounted for 33.8% of total playtime among the top 10 games. Furthermore, in this research, three the k-nearest neighbors technique variants were implemented and tested to identify the most effective method for game recommendation results. These models were trained and evaluated using three standard evaluation metrics that are mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). Among the three, k-nearest neighbors basic demonstrated the best overall performance, achieving the lowest mean squared error and root mean squared error scores, out of three metrics. The best k-nearest neighbors variant is chosen for personalized game recommendation results. Future research can be expanded through deep learning model development. The deep learning technique can offer improved personalization, and greater immunity to challenges like big data.

# REFERENCES

1. M. A. F. Ali and Z. K. A. Baizal, “Video game recommender system using deep reinforcement learning,” in *2023 International Conference on Advancement in Data Science, E-Learning and Information System (ICADEIS)*, (IEEE, Bali, Indonesia, 2023), pp. 1–6.
2. G. Andreev, D. Saxena, and J. K. Verma, “Impact of review sentiment and magnitude on customers’ recommendations for video games,” in *2021 International Conference on Computational Performance Evaluation (ComPE)*, (IEEE, Shillong, India, 2021), pp. 992–995.
3. N. Imanian, S. Havas, and P. Moradi, “Adaptive educational games based on recommendation strategy and learning style identification,” in *2022 9th International and the 15th National Conference on E-Learning and E-Teaching (ICeLeT)*, (IEEE, Sanandaj, Iran, 2022), pp. 1–6.
4. B. M. Adamson, M. J. Bathie, C. K. Leung, J. D. Portz, E. S. Valete, C. T. Voelpel, and A. G. M. Pazdor, “A recommender system for an online game platform in the metaverse,” in *2023 IEEE Smart World Congress (SWC)*, (IEEE, Portsmouth, United Kingdom, 2023), pp. 1–8.
5. H. Bian and Q. Lu, “HLRS: A deep reinforcement learning-based hero recommendation system for MOBA games,” in *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, (IEEE, Honolulu, Oahu, HI, USA, 2023), pp. 2040–2047.
6. S. Benkessirat, N. Boustia, and R. Nachida, “A new collaborative filtering approach based on game theory for recommendation systems,” *Journal of Web Engineering* (2021).
7. C. BharathiPriya, A. Sreenivasu, and S. Kumar, “Online video game recommendation system using content and collaborative filtering techniques,” in *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, (IEEE, Coimbatore, India, 2021), pp. 1–7.
8. I. R. Budianto, A. Y. Zakiyyah, and A. C. Sari, “Web-based video games recommendation system using collaborative filtering method,” in *2023 International Conference on Informatics, Multimedia, Cyber and Informations System (ICIMCIS)*, (IEEE, Jakarta Selatan, Indonesia, 2023), pp. 549–554.
9. A. H. Imran Abd Rouf, N. Musa, and I. Ismail, “Improving the accuracy of online game recommendations: A content-based filtering approach utilizing cosine similarity matrix,” in *2023 IEEE 8th International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, (IEEE, Kuala Lumpur, Malaysia, 2023), pp. 1–6.
10. W. E. Lomanto, V. Andrian, S. Achmad, and R. Sutoyo, “Collaborative filtering for Steam games recommendation,” in *2023 5th International Conference on Cybernetics and Intelligent System (ICORIS)*, (IEEE, Pangkalpinang, Indonesia, 2023), pp. 1–6.
11. B.-Y. Pei, D.-R. Liu, and J.-H. Chen, “Game recommendation based on reviews and neighbor interaction with generative adversarial networks,” in *2024 16th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI)*, (IEEE, Takamatsu, Japan, 2024), pp. 357–362.
12. Z. Tao, M. Cheung, J. She, and R. Lam, “Item recommendation using collaborative filtering in mobile social games: A case study,” in *2014 IEEE Fourth International Conference on Big Data and Cloud Computing*, (IEEE, Sydney, Australia, 2014), pp. 293–297.
13. J. Alanya-Beltran, “Personalized learning recommendation system in e-learning platforms using collaborative filtering and machine learning,” in *2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, (IEEE, Chennai, India, 2024), pp. 1–5.
14. Q. Pu and B. Hu, “Intelligent movie recommendation system based on hybrid recommendation algorithms,” in *2023 International Conference on Ambient Intelligence, Knowledge Informatics and Industrial Electronics (AIKIIE)*, (IEEE, Ballari, India, 2023), pp. 1–5.