Analyzing Video Game Ratings: A Machine Learning Approach Using Random Forest and Decision Tree Models with the RAWG Video Games Database API

Hasnain Imtiaz[1] and Sazzad Hussain Farhaan[2],
[1] Department of Television, Film and Media Studies, University of Dhaka
[2] Department of Computer Science and Engineering, United International University
Email: sakibimtiaz669@gmail.com[1], and sfarhaan2230039@mscse.uiu.ac.bd[2]

***Abstract*—The paper aims to gain insights into the factors influencing game ratings and create predictive models using machine-learning approaches. It combines the RAWG Video Games Database API with a Python application to collect data about video games and processes it for analysis. Two models, Decision Trees, and Random Forests, are trained on this data to uncover the relationships between specific features and game ratings. We assess the models' performance using Mean Squared Error (MSE) and R-squared (R2) scores, which provide quantitative measures of accuracy. Additionally, we analyze the importance of different features in influencing game ratings. The study also presents the results of regression analysis, including R-squared and Adj. R-squared values, which quantify the extent to which independent variables explain variability in the ratings. The findings shed light on critical factors such as Metacritic scores, playtime, and review text counts play statistically significant roles in predicting the impact of video game ratings. Finally, we identify potential issues with the model's assumptions, such as non-normal residuals and autocorrelation.**

***Index Terms*—Video games, Machine Learning, Decision Tree, Random Forests, Rating analysis, RAWG Video Games Database, Predictive Modelling.**

I. Introduction

In 2022, the U.S. video game industry generated a staggering $47.5 billion in content sales in revenue, where 88% of the gamers believe games expand their social circles and more than 82% of them believe games introduce new circles and networks, underlining its status as a multi-billion-dollar powerhouse (Entertainment Software Association, 2023). Furthermore, the Newzoo's Newest Free Global Games Market Report 2023 [1] reported that the number of players worldwide will reach 3.38 billion in 2026, where console gaming will be a major growth driver and Mobile will account for the highest share of the revenues. Tencent, Apple, Sony, and all other global supergiants have flourished in their ways to interact with video game users worldwide by utilizing the trends and advanced programmatic and technical mechanisms such as Generative AI, Procedural Content Generation (PCG), NPC Behavior, Personalized Gaming Experiences.

In a landscape where thousands of video game titles are available to players, the importance of creating a "good" game cannot be overstated. Game User researchers often employ surveys to collect feedback about games, using various measures to evaluate different aspects of the gaming experience, including presence, enjoyment, usability, flow, and social networks. However, existing gaming assessment scales have limitations [2], [3], including not adhering to best practices in scale development and validation, focusing on a single element of gaming, being tailored to specific game types, and containing unclear items.

The result of the Churn Prediction [17], which considers in-game activity, predicted that players' likelihood of abandoning the game is proportional to their time investment, both in-game and gamified system. This research result possesses the necessity of the intrinsic way to predict Churn and at the same time hybrid/neural recommendation system. For example, user behavior analysis including features like playtime, review text sentiment, and behavior will be beneficial to identify patterns in behavior In addition, user-specific metrics (e.g., the average rating given by a user) and game-specific metrics(e.g., the average rating for a game), can be used in churn prediction models. On the other hand, using Metacritic scores, playtime, and review text can be used for content-based recommendations for users who have enjoyed games with similar Metacritic scores or playtime might receive recommendations. Moreover, by analyzing user interactions and preferences, we can implement strategies to balance exploration(introducing users to new games) and exploitation (recommending games similar to those the user has enjoyed).

To achieve the best decisions based on the available data from the web platforms for the gaming industry, it is essential to learn from the user's perceptions and interactivity that are continuously produced by active users such as suggestions, recommendations, comments, rankings, preferences, and playtime. This offers a unique opportunity not only to obtain an accurate prediction on the player's preference and choices or behaviors but at the same time can be operated as personalized recommendations of in-game related items [6] that are likely relevant to them.

Therefore, this paper aims to unravel the factors influencing game ratings using predictive modeling techniques while considering player feedback from various scales. By understanding these elements, we can offer insights to game developers and enthusiasts alike, aiding in the creation of more enjoyable and successful games. Our research takes a multifaceted approach, encompassing machine learning, data analysis, and gaming expertise, to unveil the intricate dynamics driving game ratings and player experience.

II. Literature Review

The video games industry is a multi-billion-dollar industry with various platforms, genres, and classifications of games from AAA titles to indie games. A data-driven survey [2] involves analyzing data from a search engine and extracting patterns from query logs to gain a user perspective on video games. The frequency, Temporal distribution, and relevance of game-related queries are considered. This study has gone through game scores, several titles, and distribution by genre in the PV gaming industry based on data from the Metacritic database. The survey reveals increasing trends in the number of gaming titles during specific periods (2009-14 and 2014-2019), where top-ranking games including FPS/First Person Shooting Games, RPG(Role Playing Games) (e.g. Grand Theft Auto, Call of Duty, Counter-Strike, Bioshock, etc.) remain relatively constant in the score for each genre. The study confirms the accuracy of the evaluation in terms of conclusions about the success rate of games based on user scores. Thus, the emergence of serious games and the evaluation of successful games and genres for the gamification of real-world scenarios.

Another study by Santos et al.in 2019 [3] analyzes the existence and nature of the discrepancies in video game appraisal, including ratings, review texts, and timing of reviews. The research shows the predictive power of early experts and amateur reviews in forecasting video games' reputation in the short and long term. The research investigates around 1 million reviews on the Metacritic Platform and employs Latent Dirichlet Allocation (LDA) topic analysis to uncover differences in vocabulary and topics discussed by experts and amateurs in their reviews. Amateur reviews are remarkably predictive of game reputation in the short term, while both expert and amateur reviews are equally well suited for long-term predictions.

Recent advancements in hybrid recommendation systems [4] can be compatible with the video game ecosystem with advanced algorithms capable of learning heterogeneous sources of data and generating personalized recommendations for users. Traditional recommender system simply relies on user ratings as feedback and doesn't imply other relatable factors such as rating countings, review text, or suggestions in an analytical way that incorporates with sentimental understanding of the user's perception. Using multiple features according to the ratings of the video games can help to implement video game recommender systems such as collaborative filtering and content-based filtering [6], which are instrumental in helping users uncover hidden gems and niche titles within the vast gaming library. Using Random Forest and Decision Tree in recommending system throughout the diversified web platform can maintain transparency, improve predictive accuracy for newcomers or users, help to reduce overfitting issues, and finally quick prototyping will be serviceable [5] that can tailor recommendations to distinct user segments those who enjoy strategy games or even have interest in loots/crates may receive different recommendations than users who prefer offline games or role-play game.

III. Methodologies

In this section, the paper outlined the methodology and working process used in the analysis of the video games dataset and the development of machine learning models. The goal of this project is to gain insights into the factors influencing game ratings and to create predictive models. This Python application in this project integrated the machine learning approaches with the RAWG Video Games Database API to provide real-time predictions for game ratings using a Decision Tree and Random Forest Tree under the task of regression problem where the paper attempted to predict video game ratings, which typically involve continuous values. However, the paper also uses classification-related tasks such as identifying statistically significant predictors and analyzing the importance of different features. These tasks can be part of the overall analysis process in a regression problem to understand which independent variables are most influential in predicting the continuous target variable (ratings).

The method develops a Python application that integrates with the RAWG API to fetch data about video games. After that, it ensures a robust data acquisition pipeline that is designed to extract information from the JSON (JavaScript Object Notation) files, contained as a list of dictionaries. Python serves as the primary tool for data retrieval, granting access to an extensive repository of video game metadata. We then implied a data storage and management system for the refined dataset to ensure a secure repository within an SQLite3 database. This meticulous approach guarantees data integrity, enforces governance, and provides controlled access for future reference.

The Decision Tree and Random Forest models were trained on the training dataset to learn the relationship between selected features and game ratings. The models learned relationships between the selected features and game ratings during this phase. The refined dataset in .xlsx format undergoes partitioning into training and testing subsets, considering missing value imputation that provides readiness for subsequent model training.

Model performance and evaluation were assessed using evaluation metrics such as Mean Squared Error (MSE) and R-squared (R2) scores. These metrics helped gauge the accuracy and goodness of fit of the models. It also presents the coefficients for different features, indicating the estimated effect of independent variables on the dependent variable 'rating'. Feature importance analysis was conducted to understand

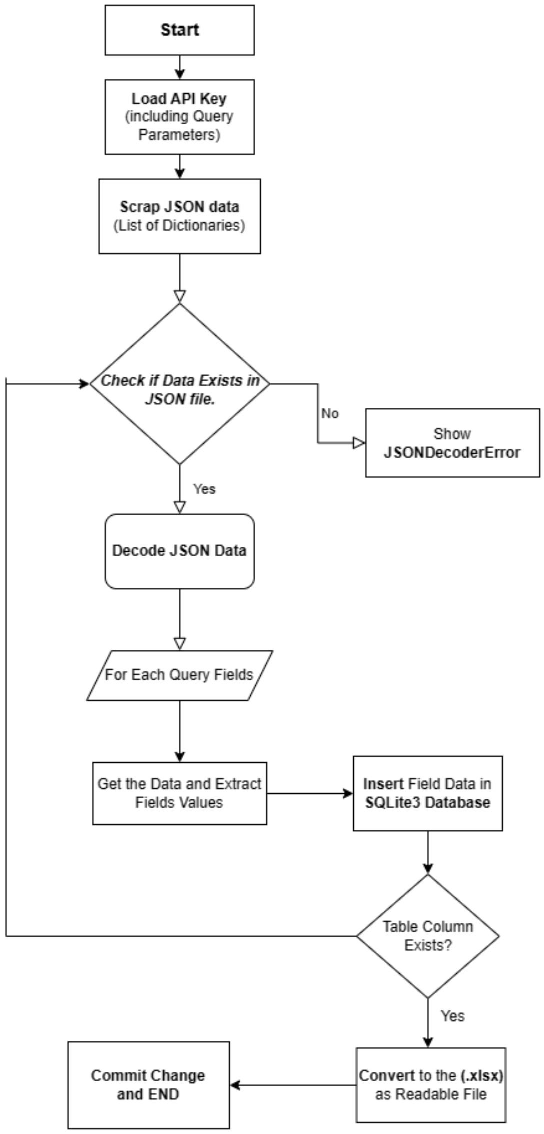

Fig. 1. Flowchart of Extracting and Analysis of video games data from RAWG API

which features had the most significant impact on game ratings. This analysis closely aligned with the principle of variable importance in machine learning, providing actionable insights into the factors driving game ratings.

IV. Results and Discussions

A. Dataset Summary

The dataset at the heart of this paper is a treasure trove of gaming data, meticulously sourced from the RAWG Video Games Data API. This rich and diversified collection of information provides invaluable insights into the dynamic world of video games, where the dataset offers details such as titles, release date, user ratings and review/ review counts, top ratings, Metacritic scores, playtime for each identified player, suggestions count according to the issues with gaming experiences and also review counts to understand the magnitude or interactivity with such games.

User Engagement is a focal point of this dataset. Key Indicators like user ratings, review text count, and Metacritic scores illuminate the player's perspective on the specific video games with overall satisfaction or sentiments. The 'rating' column quantifies user sentiment, while the 'review text count' hints at the depth of player feedback, and thus inclusion of Metacritic ratings provides an external validation of gaming quality. Multiple techniques such as user engagement metrics, Metric validation, and player feedback have been gone through the feature importance process by which training datasets are being filtered to deploy cutting-edge techniques like Decision trees or Random Forest Trees. In every scenario, we have maintained our testing data size as close as 20% of the data, and 80% of the data are being trained by the two mentioned models.

B. Experimental Setup

In this study, we conducted a comprehensive analysis of the video game rating system, where the dataset was collected from the RAWG Video Games Database API in the first week of September 2023. The computation-intensive task of training machine learning models for rating prediction was performed in both Pycharm(IDE) and Jupyter Notebook. It is inherently designed to capture the game rating system based on individual user ratings and added filters to help find suitable favorite games. We utilized Python as our primary programming language, employing popular libraries such as sci-kit-learn, pandas, matplotlib, Seaborn, stats models, Numpy, and Scipy for exploratory data analysis and model development.

For our training and testing datasets, we have applied cutting-edge techniques called random Forest, unde the ensembling learning method that builds multiple decision trees and combines their predictions to improve accuracy and reduce overfitting. In our dataset, 'rating' is the target variable that has been set up while other columns have been used as potential features. The algorithms that we have applied, both provide feature important scores which indicate features have the most significant impact on the 'rating'. Training parameters were fine-tuned to optimize model performance.

On one hand, in the context of video games popularity and user engagement, we have gone through the tree-like structures which form a series of questions or conditions based on input features(e.g., Metacritic score, playtime) to partition the dataset into subsets. Each decision node represents a feature, and each leaf node represents a rating prediction. Decision Trees capture non-linear relationships and are interpretable, allowing us to understand the decision-making process. On the other hand, for the training dataset in the Random Forest tree, we built bootstrapped subsets for the data that utilize random feature selections. The predictions from individual trees are then averaged or aggregated to produce a final rating

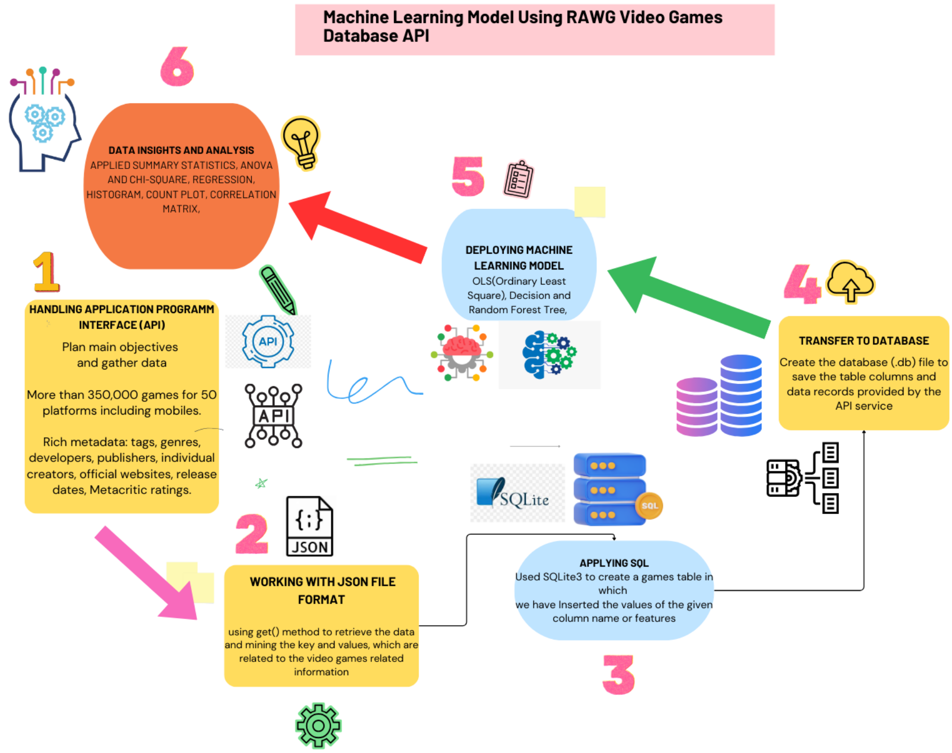

Fig. 2. Hardware and Software Setup For Data Extractions and Analysis Process

TABLE I
DESCRIPTIVE STATISTICS

Statistic	Unnamed: 0	id	tba	rating	rating_top	reviews_text_count	metacritic	playtime
Count	4000.000	4000.000	4000.000	4000.000	4000.000	4000.000	4000.000	4000.000
Mean	2000.500	2000.500	0.002250	3.352935	3.51250	2.213750	76.585504	4.567750
Std	1154.844867	1154.844867	0.047387	0.736663	1.15874	4.782602	7.748950	16.738024
Min	1.000	1.000	0.000	0.000	0.00000	0.000	23.000	0.000
25%	1000.750	1000.750	0.000	2.910000	3.00000	0.000	75.000	1.000
50%	2000.500	2000.500	0.000	3.460000	4.00000	1.000	76.585504	3.000
75%	3000.250	3000.250	0.000	3.920000	4.00000	2.000	80.000	4.000
Max	4000.000	4000.000	1.000	4.800000	5.00000	72.000	99.000	900.000

prediction. To enhance predictive accuracy, we measure feature importance scores, revealing which features have the most influence

C. Experimental Outcomes

From the technical perspective of the rating distribution analysis, among the 4000 video games, the average rating is approximately 3.35, with a minimum rating of 0 and a maximum ratio of 4.80. This suggests that most games in the dataset have ratings between 2.91 and 3.92, with some outliers(can be considered as either topmost or bottom ranked) on both ends. At the same time, the '$review_text_count'$ column shows the number of reviews for each game. On average, games have approximately 2.21 reviews. There is some variability, with the maximum number of reviews reaching 72. However, it's important to note that '$review_text_count'$ doesn't necessarily correlate directly with high ratings. Moreover, Metacritic rank, which is the largest and most popular website for aggregated reviews of television shows, video games, and albums, plays a dominant role in the gaming industry. Interestingly, the mean value for '$metacritic'$ (for 2649 entries) is approximately 76.59 with a minimum of 23 and a maximum of 99, suggesting that the games in the dataset generally receive favorable Metacritic ratings from the community of game players in an online or offline space.

In addition, the '$playtime'$ column represents the playtime or spending time for each gamer for a particular game. For example, on average, gamers tend to stick with a playtime of approximately 4.57 units, with a maximum playtime of 900 units. This suggests that most games are designed for relatively shorter play sessions, but there are some exceptions for longer video games such as Single-player mode/Storyline-based gaming experiences. Even, the '$suggestions_count'$ column also indicates the number of suggestions for each game. On average, games have around 410.59 suggestions, with a maximum of 1668. Higher suggestion counts may indicate

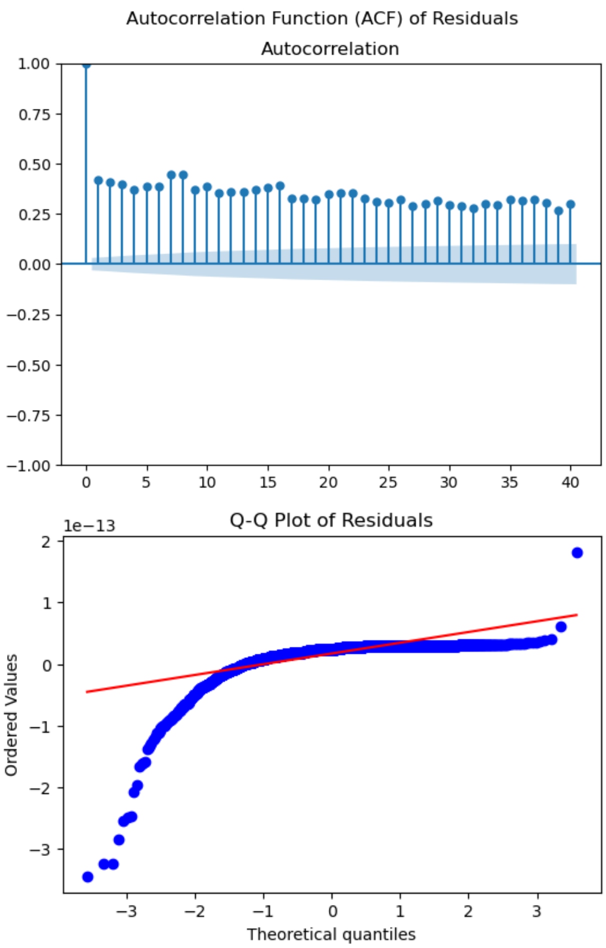

Fig. 3. Autocorrelation and Q-Q Plot for Residuals.

popular or highly discussed games

It is worth noting that some columns in the dataset, such as '$user_{g}ame'$' and '$metacritic'$', contain a significant number of missing values, possibly indicating that not all games have data available in these categories.

The Quantile-Quantile Plot (Q-Q Plot) analyzed (Figure 3) the observed data to see if the data are normally distributed in a way where most of the points would be on the line and understand how well the dots fit a straight line. In the figure, the quantile distribution for residuals suggests that the residuals in the regression model are not perfectly normally distributed. However, the concentration of residuals with small quantile values, from less than -3 to slightly greater than -1, depicts a skewness in the distribution, which means some observations have negative residuals, indicating potential model errors or outliers. At the same time, the Q-Q plot's regression line shows that residuals are approximately normally distributed for a substantial portion of the data. However deviations from the line, especially in the tails, suggest that there may be some cases where the residuals do not follow a normal distribution, possibly due to outliers and model limitations.

In Figure 3, Autocorrelation shows the values for different

TABLE II
REGRESSION MODEL SUMMARY STATISTICS

Statistic	Value
R-squared	0.434
Adj. R-squared	0.433
F-statistic	506.2

TABLE III
COEFFICIENT SUMMARY

Variable	Coef	Std Err	t-statistic	P-value
const	0.9202	0.075	12.288	0.000
reviews$_{t}ext_{c}ount$	0.0194	0.002	11.717	0.000
metacritic	0.0336	0.001	35.849	0.000
playtime	0.0029	0.001	2.479	0.013
suggestions$_{c}ount$	5.334e-05	4.81e-05	1.109	0.267

lags. The subsequent data points are between 0 and 0.50 on the Y-axis, suggesting that, at these lags, there is a positive but weaker autocorrelation. This indicates that values in the time series are somewhat correlated with values at these lagged positions, but the strength of the correlation diminishes as the lag increases. The fact that there are no data points below the X-axis (negative autocorrelation) suggests that there is no strong negative correlation between values at any of the examined lags. This means that values at a given lag do not tend to be inversely related to values at previous lags in our time series.

The following OLS(Ordinary Least Square) regression results from Table II are critically based on the characteristics of features or independent variables included in the OLS regression model. In the summary statistics, the numerical outcome for the R-squared (R2) and Adj. R-squared tests are 0.434 and 0.433. These values indicate that approximately 43.4% of the dependent variable 'rating' variability can be explained by the independent variables in the regression model.

TABLE IV
ADDITIONAL STATISTICS

Statistic	Value
Mean Squared Error (MSE)	0.18499388124793303
Chi-Square Test (reviews_text_count vs. playtime)	
Chi-Square Value	21323.421499697994
p-value	0.0
ANOVA Test (Metacritic vs. suggestions_count)	
F-statistic	1.5565091064256782
p-value	0.003768003756472688

TABLE V
ADDITIONAL OLS STATISTICS

Statistic	Value
Omnibus	183.127
Prob(Omnibus)	0.000
Jarque-Bera (JB)	248.670
Prob(JB)	1.00×10^{-54}
Skew	-0.600
Kurtosis	3.902
Durbin-Watson	1.964
Cond. No.	4.34×10^{3}

The Adj. R-squared is slightly lower, suggesting that some independent variables may not significantly contribute to the model's explanatory power.

In Table IV, MSE (Mean Squared Error) and R-squared values provide insights from the Decision Tree and Random Forest models in predicting 'rating' based on 'Review Text Count'. The Decision Tree model's MSE and R-squared values are consecutively 0.3768 and 0.3054, while the Random Forest model's values are 0.3769 and 0.3054. These analytical measurements suggest that MSE with having a lower MSE indicates model's predictions are closer to the actual values in terms of how close their predictions are to the actual 'rating' values. From TABLE III, the F-statistics tests the overall significance of the regression model. A very low p-value(0.00) and F-Statistics's value of 506.2 suggest that the model is statistically significant, meaning at least one of the independent variables has a significant impact on the dependent variable. The model appears to have a good fit, as indicated by the F-statistic and R-squared values. In addition, Coefficients values for different features such as 'const' coefficient (Intercept): 0.92 $'reviews_text_count' coefficient : 0.019$, $'Metacritic' coefficient : 0.0336$, $'playtime' coefficient : 0.0029$ and $'suggestions_count' coefficient : 5.334e-05$ suggest that the estimated effect of the independent variable on the dependent variable 'rating' while holding other variables constant.

Moreover, The Omnibus test ($p < 0.001$), the Durbin-Watson statistic (1.964), and the Jarque-Bera test ($p < 0.001$) in Table V suggest that residuals are not normally distributed and have potential issues with the model's assumptions. At the same time, there are some auto-correlations in the residuals and those residuals don't follow the normal distribution. In effect, $'reviews_count'$ and $'Metacritic'$ have positive coefficients, indicating that an increase in these variables is associated with higher ratings. $'playtime'$ (coefficient of 0.0029 and p-value of 0.013) has a positive coefficient but a smaller magnitude, suggesting a weaker relationship with ratings. $'suggestions_count'$ (coefficient of $5.334e-05$ and a p-value of 0.267 as) has a coefficient close to zero, indicating that it does not significantly influence ratings in this model. The skewness measures the asymmetry of the residuals with a skew value of -0.600 that suggests a slight left-skew with kurtosis of 3.902 indicating the distribution has heavier tails than a normal distribution. Though there are some non-linearity issues and multicollinearity problem arising, we still can see review text impact, metacritic influence and playtime associations have statistically significant impact on overall gaming experiences and preferences.

In Figure 4 and Figure 5, we can see that the feature '$suggestion_count'$ has the highest importance, with a value between 0.2 and 0.3, explaining the number of suggestions or recommendations associated with a review is a crucial factor in predicting the '$rating'$. When there are more suggestions for a particular review, it tends to have a significant impact on the overall rating. While not as dominant as Metacritic, the suggestions count still contributes substantially to the model's predictions. At the same time, '$review_text_count'$ and '$playtime'$ both have got moderate importance, with values of slightly greater than 0.1 and exactly 0.1. These numeric values indicate that the length or word count of the review text also plays a role in predicting '$rating'$, although it is not as influential as the '$suggestion_count'$. A longer detailed review may contribute positively to the overall rating. Besides, the amount of '$playtime'$ has a relatively lower importance score of around 0.1 which suggests that the amount of time players spend on a game, on its own, has a limited impact on the predicted ratings according to both models. Which is likely related to the product being reviewed, and has a relatively lower impact on predicting the '$rating'$ compared to the other features. In summary, when predicting the 'rating' on Metacritic, the number of suggestions or recommendations for a review is the most crucial factor, followers by the length of the review text. Games with higher Metacritic scores tend to receive higher ratings, as suggested by the high importance score.

- **Model Performance:** The regression model exhibits an R-squared value of 0.434, indicating that approximately 43.4% of the variance in the 'Rating' can be explained by the included independent variables. This suggests that the model provides a moderate level of fit to the data. The Adj. R-squared, a variant adjusted for the number of predictors, is 0.433, reinforcing the model's explanatory power.
- **Significant Predictors:** Among the independent variables, 'Reviews Text Count' and 'Metacritic' stand out as statistically significant predictors of 'Rating.' Both have p-values close to zero, indicating strong evidence that they are associated with changes in 'Rating.' 'Reviews Text Count' has a positive coefficient of 0.0194, suggesting that an increase in the number of reviews positively impacts the 'Rating.' Similarly, 'Metacritic' has a positive coefficient of 0.0336, implying that higher Metacritic scores are associated with higher 'Rating.'
- **Moderate Influences:** 'Playtime' also shows a statistically significant impact on 'Rating' with a positive coefficient of 0.0029. However, the effect size is smaller compared to 'Reviews Text Count' and 'Metacritic.' This indicates that while longer playtime can contribute to higher ratings, its influence is less pronounced.
- **Insignificant Predictor:** 'Suggestions Count' does not exhibit a statistically significant relationship with 'Rating' as its p-value is relatively high (0.267). Therefore, it may not play a substantial role in explaining variations in 'Rating' in this model.
- **Model Fit:** The F-statistic of 506.2 with a very low associated p-value (close to zero) suggests that the overall model is statistically significant, indicating that at least one of the predictors in the model has a non-zero effect on 'Rating.'

In Figure 4 and Figure 5, we can see that the feature '$suggestion_count'$ has the highest importance, with a value

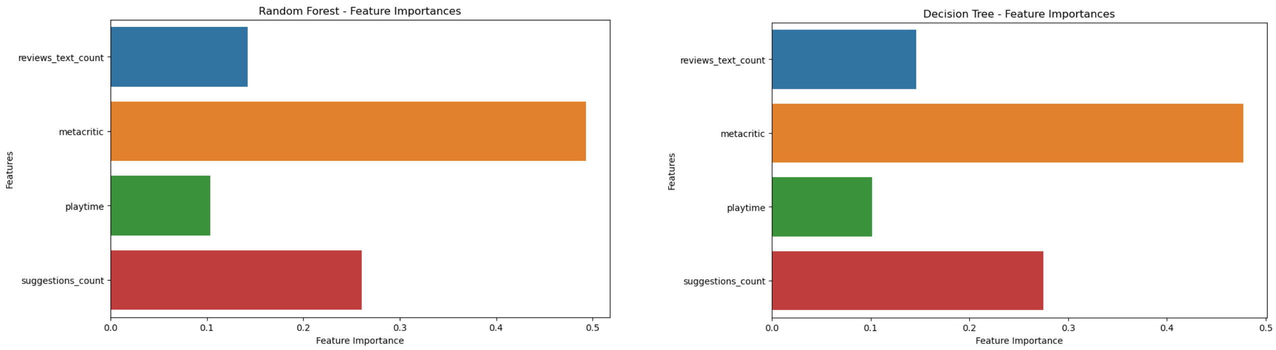

Fig. 4. Feature Importance, Bar Plot

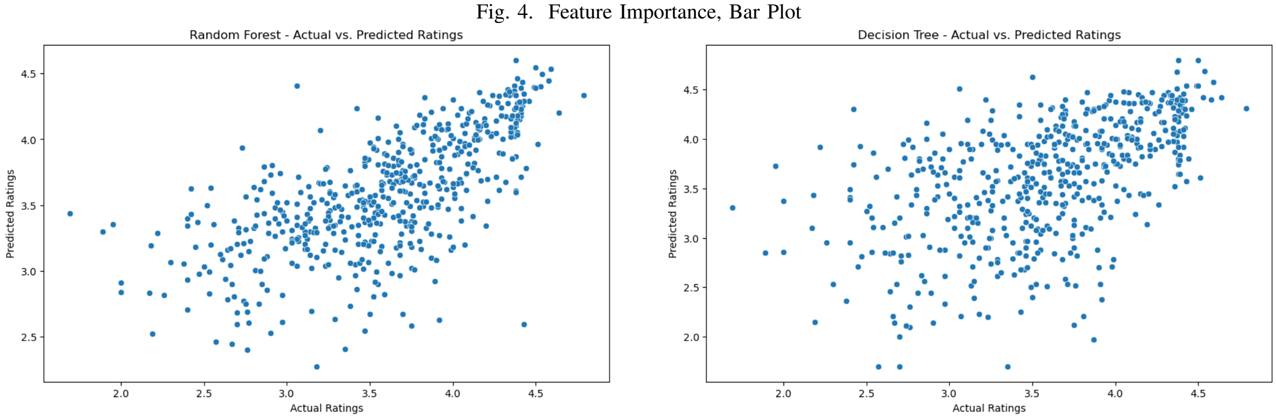

Fig. 5. Feature Importance, Scatter Plot

slightly less than 0.5 and that suggests the number of suggestions or recommendations associated with a review is a crucial factor in predicting the ‘$rating'$. When there are more suggestions for a particular review, it tends to have a significant impact on the overall rating. At the same time, ‘$review_text_count'$ and ‘$playtime'$ both have got moderate importance, with values of slightly greater than 0.1 and exactly 0.1. These numeric values indicate that the length or word count of the review text also plays a role in predicting ‘$rating'$, although it is not as influential as the ‘$suggestion_count'$. A longer detailed review may contribute positively to the overall rating. Besides, the amount of ‘playtime’. Which is likely related to the product being reviewed, has a relatively lower impact on predicting the ‘rating’ compared to the other features. In summary, when predicting the ‘$rating'$ on Metacritic, the number of suggestions or recommendations for a review is the most crucial factor, followers by the length of the review text.

In Decision Tree (Figure 6) metrics for each numeric feature inside the video games data significantly shows us that for the ‘$review_textcount'$ feature, the Decision Tree model has an MAE of 0.48, which means on average, its predictions deviate from the actual ratings. The MSE of 0.38 indicates the squared average deviation and the RMSE. In addition, ‘$metacritic'$ feature results in a slightly better performance with an MAE of 0.45, which indicates that predictions deviate by 0.45 on average. The MSE of 0.34 implies that squared errors are lower, leading to a lower RMSE of 0.59. As a consequence, ‘$playtime'$ indicates slightly larger prediction errors with having score MAE of 0.52 and MSE of 0.45 suggesting that squared errors are moderate, resulting in an RMSE of 0.67. Moreover, ‘$rating_top'$ demonstrates the lowest prediction errors. The MSE of 0.12 and RMSE of 0.34 also suggest good predictive accuracy.

In summary, the Decision Tree model’s performance varies for different numeric features. $'rating_top'$ shows the best predictive accuracy with the lowest MAE, MSE and RMSE values. $'metacritic'$ also performs reasonably well, while $'review_text_count'$ and $'playtime'$ have larger prediction errors. It’s important to consider these metrics when selecting features for modelling and interpreting the model’s performance. Further feature engineering or trying different mechine learning algorithms could potentially improve predictions for features with higher errors.

The analysis of the relationship between ’rating’ and the independent variables $'metacritic'$ $'playtime'$, ‘$review_text_count$ and $'ratingTop'$ using Random Forest regression (Figure 7) reveals some interesting patterns. In the graph depicting the relationship between $'metacritic'$ and ‘rating’, both the actual(in blue) and random forest predictions(in red) are scattered across the range of 30-90. However, the majority of data points are densely concentrated between 70 and 80 which suggests that games with Metacritic scores in this range tend to have consistent and high ratings,

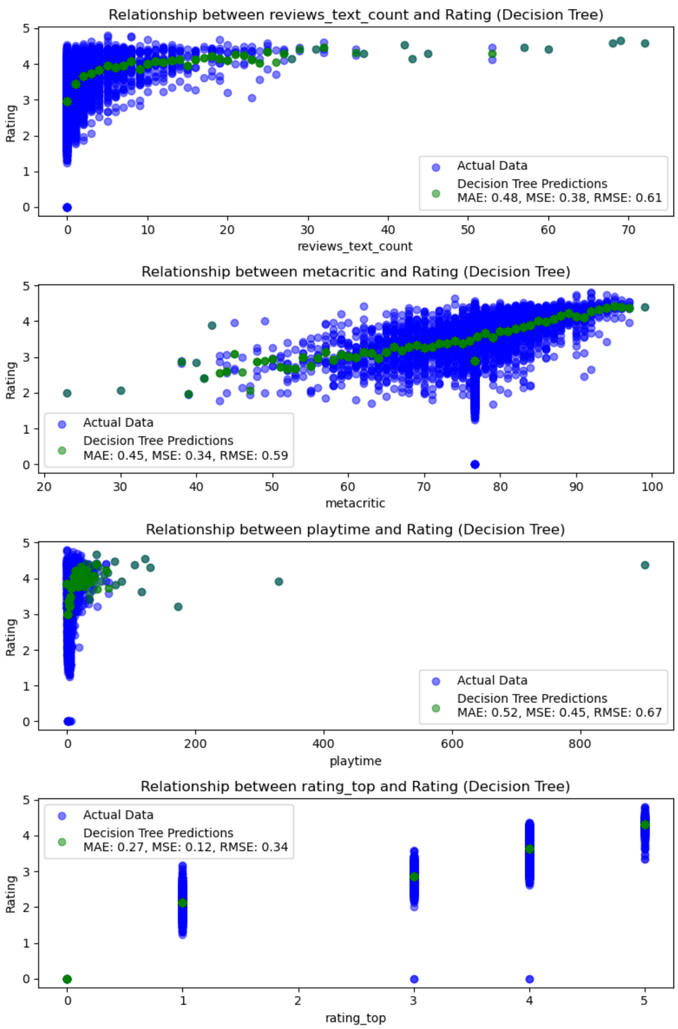

Fig. 6. Relationship Between the Features - Decision Tree Model

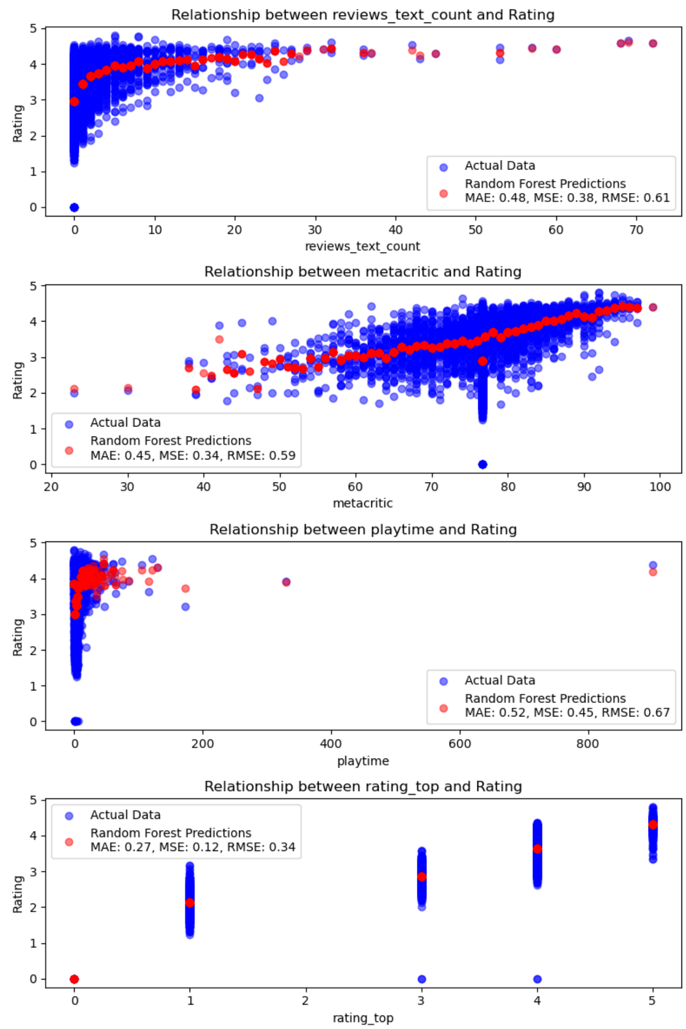

Fig. 7. Relationship Between Features - Random Forests

as indicated by the dense cluster of data points. In the graph (Relationship between ‘$review_text_count$ and ‘$rating'$) shows that there is a significant concentration of data points in the interval of 0 to 20 on the x-axis, indicating that most games receive a relatively low number of text-based reviews. This relation also portrayed that games with higher text-based review counts tend to have higher ratings.

- ‘$rating_top'(0.8701)$: This feature has the highest importance score among all the selected features. It suggests that the $'rating_top'$ feature is the most influential in predicting the rating score of a game. In other words, the top rating a game can achieve appears to be a critical factor in determining the overall rating.
- ‘$metacritic'(0.0570)$: The $'metacritic'$ feature also has a significant importance score, though lower than $'rating_top'$. This implies that the Metacritic score of a game contributes positively to predicting its rating. Games with higher Metacritic scores tend to have higher overall ratings.
- ‘$reviews_text_count'(0.0385)$: $'Reviews_text_count'$ has a lower importance score compared to the previous two features. It suggests that the number of reviews or the amount of text content related to a game may have a lesser impact on predicting its rating. However, it still contributes to some extent.
- ‘$playtime'(0.0344)$: $'Playtime'$ has the lowest importance score among the selected features. This indicates that the amount of playtime in a game may not be a strong predictor of its rating, or other features such as $'rating_top'$ and $'metacritic'$ are more influential.

In summary, according to the Random Forest model's feature importances, $'rating_top'$ and $'metacritic'$ are the most critical factors in predicting the rating of a game. Developers and analysts can use this information to focus on improving these aspects to potentially enhance a game's rating. However, it's important to note that this analysis is based on the specific model and dataset used, and real-world scenarios may involve more complex relationships between features and ratings.

REFERENCES

[1] Get Newzoo's newest free Global Games Market Report 2023. (2023, June 20). Newzoo. https://newzoo.com/resources/trend-reports/newzoo-global-games-market-report-2023-free-version

[2] Predescu, A., & Mocanu, M. (2020). A data driven survey of video games. 2020 12th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), 1–6. https://doi.org/10.1109/ECAI50035.2020.9223203

[3] Santos, T., Lemmerich, F., Strohmaier, M., Helic, D. (2019). What's in a Review: Discrepancies Between Expert and Amateur Reviews of Video Games on Metacritic. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1–22. https://doi.org/10.1145/3359242

[4] Elahi, M., Khosh Kholgh, D., Kiarostami, M. S., Oussalah, M., & Saghari, S. (2023). Hybrid recommendation by incorporating the sentiment of product reviews. Information Sciences, 625, 738–756. https://doi.org/10.1016/j.ins.2023.01.051

[5] Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[6] Bertens, P., Guitart, A., Chen, P. P., & Periáñez, Á. (2018). A Machine-Learning Item Recommendation System for Video Games. 2018 IEEE Conference on Computational Intelligence and Games (CIG), 1–4. https://doi.org/10.1109/CIG.2018.8490456

[7] Liu, H., Wang, Y., Peng, Q., Wu, F., Gan, L., Pan, L., & Jiao, P. (2020). Hybrid neural recommendation with joint deep representation learning of ratings and reviews. Neurocomputing, 374, 77–85. https://doi.org/10.1016/j.neucom.2019.09.052

[8] Hazrati, N., Elahi, M., & Ricci, F. (2020). Simulating the Impact of Recommender Systems on the Evolution of Collective Users' Choices. Proceedings of the 31st ACM Conference on Hypertext and Social Media, 207–212. https://doi.org/10.1145/3372923.3404812

[9] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[10] Marin, N., Makhneva, E., Lysyuk, M., Chernyy, V., Oseledets, I., & Frolov, E. (2022). Tensor-based Collaborative Filtering With Smooth Ratings Scale (arXiv:2205.05070). arXiv. http://arxiv.org/abs/2205.05070

[11] Porter, A. M., & Goolkasian, P. (2019). Video Games and Stress: How Stress Appraisals and Game Content Affect Cardiovascular and Emotion Outcomes. Frontiers in Psychology, 10, 967. https://doi.org/10.3389/fpsyg.2019.00967

[12] Banskota, A., & Ng, Y.-K. (2020). Recommending Video Games to Adults with Autism Spectrum Disorder for Social-Skill Enhancement. Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, 14–22. https://doi.org/10.1145/3340631.3394867

[13] Predescu, A., & Mocanu, M. (2020). A data driven survey of video games. 2020 12th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), 1–6. https://doi.org/10.1109/ECAI50035.2020.9223203

[14] Banskota, A., & Ng, Y.-K. (2020). Recommending Video Games to Adults with Autism Spectrum Disorder for Social-Skill Enhancement. Proceedings of the 28th ACM Conference on User Modeling, Adaptation and Personalization, 14–22. https://doi.org/10.1145/3340631.3394867

[15] Shelstad, W. J., Hosein, A. A., Keebler, J. R., & Chaparro, B. S. (2020). Using User Experience Scales to Predict Video Game Play and Purchasing Intent. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 64(1), 2065–2069. https://doi.org/10.1177/1071181320641500

[16] Dharma, H. R. C., & Setiadi, C. J. (2023). Adaptation in Game: Consumer Rating Result for Remade and Remastered Games (1997-2022). 2023 8th International Conference on Business and Industrial Research (ICBIR), 218–222. https://doi.org/10.1109/ICBIR57571.2023.10147693

[17] Darzi, A., Wondra, T., McCrea, S., Novak, D. (2019). Classification of Multiple Psychological Dimensions in Computer Game Players Using Physiology, Performance, and Personality Characteristics. Frontiers in Neuroscience, 13, 1278. https://doi.org/10.3389/fnins.2019.01278

[18] Aziz, A., Ismail, S., Othman, M. F., Mustapha, A. (2018). Empirical Analysis on Sales of Video Games: A Data Mining Approach. Journal Name, 1049(1), 012086.

[19] Nor, M. H. M., Liong, C. Y., Gunasekaran, H. (2023). Visualization and Prediction of User Ratings on Video Games Using Web Analytics. Nucleation and Atmospheric Aerosols.

[20] W. Yang, T. Huang, J. Zeng, L. Chen, S. Mishra and Y. E. Liu, "Utilizing Players' Playtime Records for Churn Prediction: Mining Playtime Regularity," in IEEE Transactions on Games, vol. 14, no. 2, pp. 153-160, June 2022, doi: 10.1109/TG.2020.3024829.

[21] H. Verma and S. Gochhait, "Review and Analysis of usage of Spatio-Temporal Game Analytics," 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2023, pp. 1816-1821, doi: 10.1109/ICICCS56967.2023.10142334.