

Feature Importance of a Multi-Layer Perceptron Applied to Elden Ring Reviews and User Gaming History

Eduardo N. S. Ramos¹

¹Departamento de Engenharia Elétrica
Pontifícia Universidade Católica do Rio de Janeiro (PUC-RJ)

eduardonsantiago@aluno.puc-rio.br

Abstract. User reviews in video games have been integral to digital platforms since their inception, guiding users toward products aligned with their preferences and steering away gamers who might not find certain products entertaining. However, few studies have focused on the specific gaming profiles that lead to positive or negative reviews, especially for games targeting niche markets but effectively serving those demographics. This paper aims to explore a method for predicting negative reviews based on users' gaming history using a Multi-Layer Perceptron (MLP), and to understand which variables positively and negatively influence user feedback for the game *Elden Ring*. Our results indicate that while the MLP did not achieve satisfactory metrics in this task, its output can still serve as a baseline for identifying gaming profiles suitable for the game.

1. Introduction and Motivation

The video game industry has generated over 400 billion dollars in revenue as of 2024, surpassing the movie, book, and music industries [Statista 2024]. This form of entertainment differs significantly from others because it relies on user input rather than passive consumption. This means that each user's experience with a game is uniquely shaped by their interactions with the content. This interactivity allows for a personalized and dynamic engagement that is not typically found in other forms of media, where the experience is more uniform and passive.

This individual experience fosters video game communities, such as those on [Reddit 2024], where players can share their achievements, ask questions, and discover new ways to interact with games they are already familiar with.

In the case of *Elden Ring*, the game is known for its difficulty and overwhelming challenges compared to other games. Additionally, games released by the same company (*From Software*) possess atmospheric and dark fantasy elements that further reinforce a hostile game world. As pointed out by [Lachowski 2024], the world of *Elden Ring* is filled with almost insurmountable adversities and dark, decaying ruins, while still maintaining an element of hope, with non-player characters commenting on the player's chance to defeat the antagonists and achieve victory.

[Lachowski 2024] references this contrast with the element of 'Grace' in *Elden Ring*.

The misfortune of being a Tarnished places us in the position of observers of the conflicts and their developments across the Lands Between, while

the contact with Grace conditions new transformations brought about by the battles with demigods and other secrets encountered along the way.

Regarding a user’s main form of product evaluation in the video game industry, the Steam review system [FromSoftware 2024] allows players to create binary reviews (either positive or negative) of their played games, along with a written text detailing their impressions. Despite winning the Game Awards for Best Game of 2022 [IGN 2022], *Elden Ring* currently has around 92% positive reviews across all languages and reviews on Steam.

A system capable of identifying, based on a user’s gaming history, whether they are more likely to enjoy *Elden Ring*, and understanding how their gaming habits can shape future installments of the franchise, can help mitigate negative reviews and provide marketing strategies better aligned with the target audience.

2. Related Works

While there is a lack of research on predicting a user’s gaming interests based on their history, some papers address the use of sentiment analysis and its pitfalls in gaming reviews, as well as other methods for predicting a user’s game review. Some papers related to gaming reviews will be detailed in this section, alongside a review of feature importance methods applied to a shallow MLP.

[Viggiato et al. 2022] and [Wang and Goh 2020] attempted to classify gaming reviews using sentiment analysis. The former concludes that analyzing text alone is problematic due to the prevalence of sarcastic reviews and bullet-point formats lacking contextual information necessary for classification models. In contrast, the latter achieves better results by incorporating topic modeling and extracting features such as word count and keywords directly from the text. However, neither study incorporates external information such as gaming hours to predict user reviews.

[He et al. 2020] discusses the application of SHAP [Lundberg and Lee 2017] and other feature importance methods to shallow MLPs, highlighting their utility in explaining model outputs. While the authors do not delve into deep neural networks (DNNs), the use of SHAP can still yield significant insights for interpreting model predictions and understanding how input data contribute to the model’s output.

3. Methodology

This section will delve into the methodology used for dataset preparation, dimensionality reduction, neural network training, and evaluation, followed by a SHAP analysis regarding the importance of the features in the model.

3.1. Dataset

The datasets were obtained via Steam’s Web API [Valve 2024]. The data comprised two datasets: one regarding user reviews of the game *Elden Ring*, and another containing the total playtime of users across all games in their account history. Both datasets include information related to the time of data extraction, which occurred after the reviews were written. For simplification purposes, users’ behavior will be considered consistent, meaning that their gaming behavior before playing *Elden Ring* should remain the same after playing the game.

The reviews dataset contains users' Steam IDs and review information, as seen in Table 1. This table contains the model's target variable, 'voted_up', alongside total number of reviews and number of games owned.

steamid	playtime_forever	num_reviews	num_games_owned	voted_up
76561199098034787	303	13	0	True
76561199013405597	82	12	0	False
76561199173536454	437	1	0	True
76561199118169519	88	2	0	False
76561198809081703	4957	2	18	True

Table 1. Reviews Dataset

The user information dataset in Table 2 contains the total of minutes each user played the games in the dataset. This table has 4462 samples, with 17656 columns, one for each game across all users. Games that have been played by less than 1% of users in the dataset have been excluded for computational reasons.

id	80_playtime_forever	20_playtime_forever	105600_playtime_forever
76561198356825656	63.0	12.0	25468.0
76561198880066501	NaN	NaN	226.0
76561198124297820	NaN	NaN	832.0
76561199521464666	NaN	NaN	NaN
76561199079408421	NaN	NaN	NaN

Table 2. Playtime Dataset

In order to fit this data into any machine learning model, the columns related to playtime were aggregated into predefined tags from the Steam Store page of each game. These tags are voted on by users and represent the overall experience users have while playing the game. Each game has a total of 10 tags, and the gameplay time has been aggregated entirely for each tag it belongs to. This process resulted in a total of 366 columns, as shown in Table 3.

game_count	Action	FPS	Military	Strategy	Adventure	target
143	87439.0	11477.0	568.0	19822.0	59639.0	0
24	12942.0	10925.0	7447.0	9067.0	1463.0	0
113	356913.0	187850.0	451.0	12415.0	240115.0	0
73	24741.0	11571.0	10649.0	11928.0	5022.0	0
215	191710.0	39479.0	19153.0	40466.0	180129.0	0

Table 3. Model Dataset

The dataset contains 5.5% of positive cases (when someone disliked the game) and maximum correlation with the target variable of 0.10, from the 'Heist' game tag feature. As seen in Table 4. This problem presents an unbalanced dataset with poor linear correlation to the target variable.

Feature	Absolute Correlation
Heist	0.105449
Gun Customization	0.096270
Economy	0.079738
num_reviews	0.078400
Epic	0.066750
Transhumanism	0.063779
Stealth	0.063666
Political	0.060928
Third-Person Shooter	0.060907
Based On A Novel	0.060147

Table 4. Correlation Between Features and Target

3.2. Dimensionality Reduction

The initial step in dimensionality reduction employed a correlation filter to analyze the model features. Features exhibiting an absolute correlation coefficient above 0.9 were identified and the feature with lesser variance was removed to prevent redundancy in the dataset.

Afterwards, LightGBM was employed for the initial classification prediction. LightGBM, introduced by [Ke et al. 2017], is a Gradient Boosting Decision Tree algorithm known for its efficiency and faster training times. The model was applied to a dataset using SMOTE [Chawla et al. 2002], a technique that oversamples the minority class and undersamples the majority class. This process was used to obtain the feature importances and remove the less important variables before fitting into a MLP, resulting in 166 columns.

3.3. Hyperparameter Tuning

To determine the optimal parameter combination for a Multi-Layer Perceptron (MLP) [Rumelhart et al. 1986], the dataset underwent three forms of data preprocessing: oversampling with SMOTE [Chawla et al. 2002], random undersampling [Lemaître et al. 2017], and no preprocessing. In the last case, weights of the Neural Network’s BCEWithLogitsLoss criterion in PyTorch [Paszke et al. 2017] were adjusted to address class imbalance, emphasizing variables crucial for discerning the minority class without using sampling techniques.

The hidden layers utilized ReLU activation functions [Agarap 2019] to mitigate the Vanishing Gradient problem, while the output layer employed a Sigmoid activation [Zhai et al. 2023] to produce a continuous value, which was rounded to a discrete number (1 or 0) for calculating performance metrics.

Afterwards, the model underwent the following hyperparameter iterations detailed in Table 5, across 5 folds of stratified cross-validation. The process optimized the F1 Score after rounding of the Neural Network output .The dataset was split into a 60%/20%/20% training/validation/test split. The data was scaled using the StandardScaler method provided by the library *Sci-Kit Learn* [Pedregosa et al. 2011]. Both the validation and testing

datasets were scaled according to the training data, and no sampling technique was applied to the samples preserve the original data structure.

Adjustments in momentum or loss criterion yielded no improvements in accuracy, precision, and recall. In cases where there were two hidden layers, the size of the second layer was half that of the first hidden layer.

Hyperparameter	Values
Hidden Layers	1 and 2
First Hidden Layer Size	256, 128, 64, 32 and 16
Learning Rates	0.01, 0.001, 0.0001 and 0.00001

Table 5. Hyperparameters Used in Training

3.4. Explainable AI

In order to comprehend the model output and assess the feature importance, the explainable framework SHAP was utilized. Introduced by [Lundberg and Lee 2017], SHAP quantifies the impact of each feature in a classification problem, indicating whether they positively or negatively influence the final probability. Additionally, SHAP reveals interactions between features, enabling a nonlinear analysis of their impact.

4. Results

The results in Table 6 represent the top 3 network configurations and their metrics in the three types of data preprocessing employed during hyperparameter tuning. Subsequently, after achieving the best results the best model per sampling strategy was run with the full dataset (except test and validation) to provide a final and accurate classification.

Sampling	Layers	HL Max Size	L Rate	Acc	Prec	Rec	F1
SMOTE	2	32	10e-3	0.9058	0.2903	0.2349	0.2551
SMOTE	2	64	10e-3	0.8850	0.2437	0.1980	0.2178
SMOTE	1	32	10e-4	0.8842	0.2511	0.1884	0.2121
Weights	1	32	10e-4	0.8778	0.1433	0.2101	0.1562
UnderSampling	2	16	10e-3	0.6296	0.0917	0.4755	0.1430
UnderSampling	1	16	10e-2	0.6052	0.0846	0.5063	0.1373
UnderSampling	1	256	10e-1	0.6545	0.0849	0.4618	0.1373
Weights	1	32	10e-3	0.8840	0.1234	0.1573	0.1322
Weights	1	16	10e-2	0.8722	0.1133	0.1697	0.1309

Table 6. Hyperparameters Used in Training

The final run, as seen in Table 7, demonstrated a disparity between the results of the SMOTE sampling strategy compared to the others. This indicates that while SMOTE could capture and classify individual folds, it struggled on a larger dataset. The undersampling and weight adjustment strategies provided better overall results, with the latter maintaining a higher F1 score, balancing precision and recall. Since no sampling strategy was conducted with the weights method, the threshold with greater F1 Score remained at 0.5.

Sampling	Test Acc	Test Prec	Test Rec	Test F1
Undersampling	0.8365	0.1031	0.2826	0.1511
Weights	0.8824	0.1445	0.26086	0.1860
SMOTE	0.7984	0.0864	0.3043	0.1346

Table 7. Final Run with Best Configuration per Sampling Strategy

None of the models were able to successfully classify whether a user would review *Elden Ring* positively or negatively based on the user’s gaming history. Nevertheless, the results can still provide information regarding feature importance and how it managed to classify the more obvious cases.

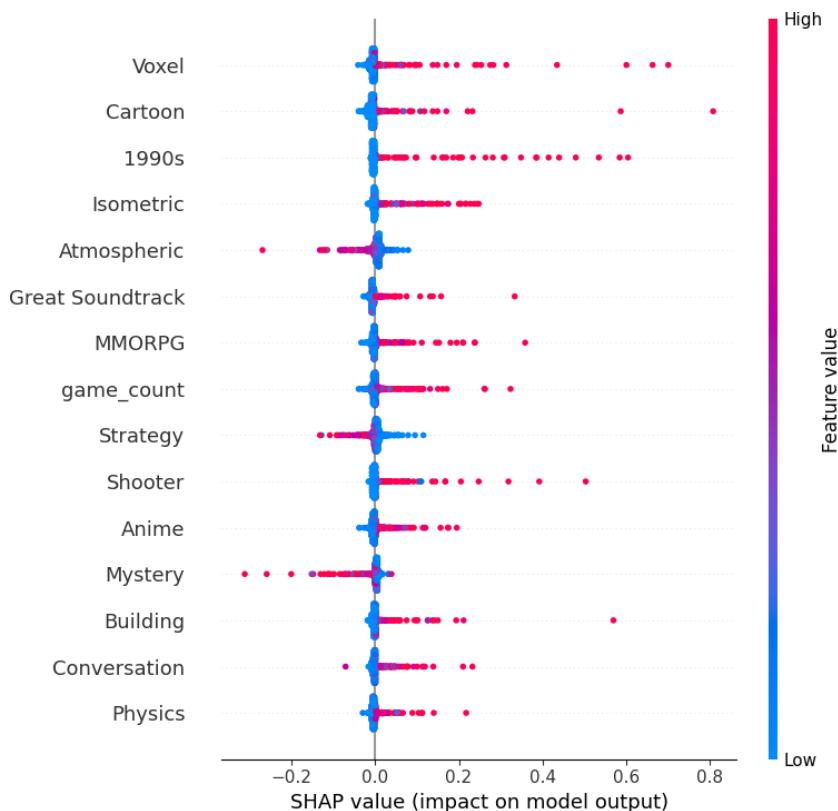


Figure 1. Impact of Features on Model Output

From Figure 1, the model successfully identified the features with the greatest negative impact on users’ final reviews of *Elden Ring*. The tags “Voxel,” “Cartoon,” “1990s,” and “Anime” suggest that gamers accustomed to more family-friendly products with cartoonish and simple graphical fidelity are not the primary audience for *Elden Ring*. Additionally, tags related to entirely different gaming experiences, such as “Physics” and “Conversation,” indicate a demographic that typically engages with games significantly different from the classic Action Adventure genre.

“Shooter” and “Building” represent groups of players accustomed to more challenging gaming experiences but who engage with products in different manners. The “Shooter” tag includes arena shooters characterized by fast-paced online gameplay, while

the "Building" tag encompasses large-scale survival experiences. These two forms of engagement differ significantly from the boss-based difficulty found in *Elden Ring*.

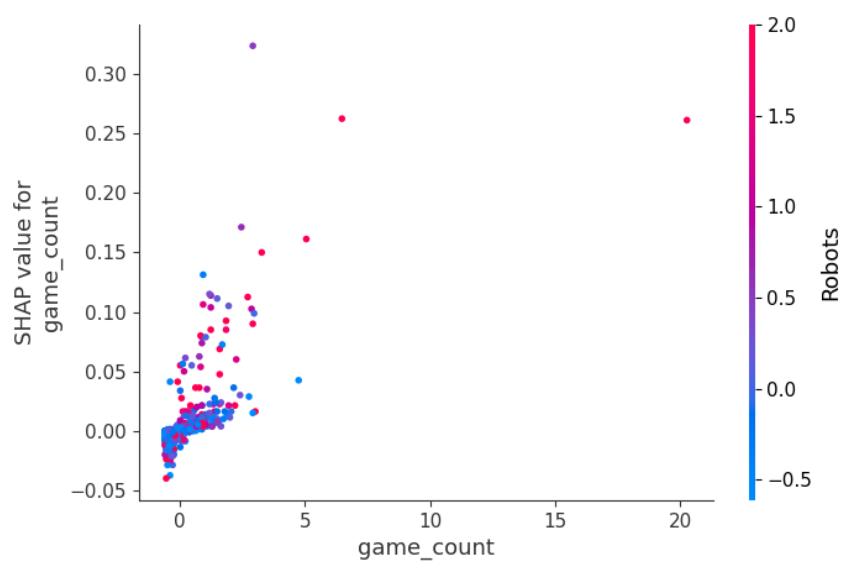


Figure 2. Impact of Features on Model Output

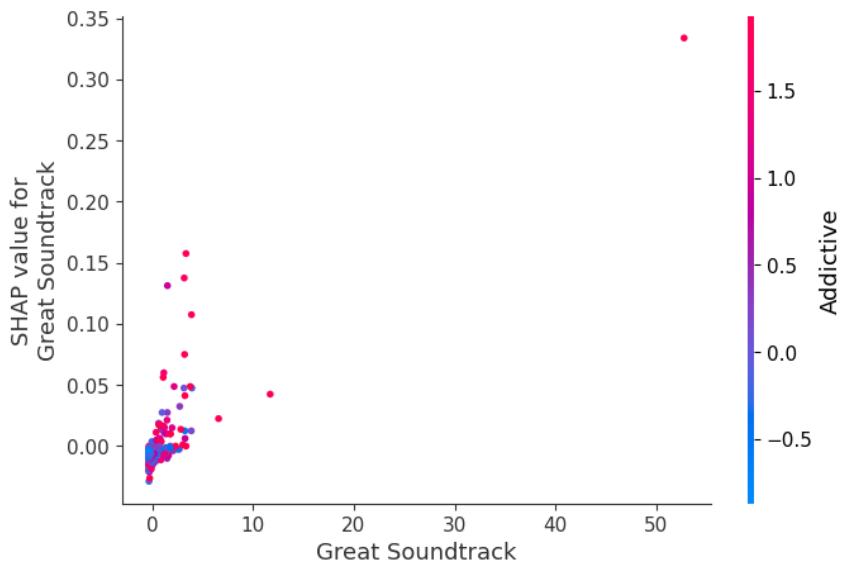


Figure 3. Impact of Features on Model Output

Figure 2 shows how the number of games in a user's account can negatively impact their reviews of *Elden Ring*. This variable interacts with the "Robot" tags. The figure suggests that while having fewer games in a user's account may lead to higher reviews of *Elden Ring*, possibly because it serves as the entry point for many newcomers to the soul-like genre, having many games has a smaller impact on disliking *Elden Ring*. Instead, having many games, especially with many hours played in typically sci-fi experiences (indicated by the "Robots" tag), may reflect a user's preference for that setting, making the dark fantasy experience of *Elden Ring* less appealing to them.

Meanwhile, the Figure 3 shows that the intersection of "Great Soundtrack" and high playtime of "Addictive" games leads to a disliking of *Elden Ring*. Games with both tags are typically shorter experiences with extensive replayability, often associated with the roguelike genre. Although these games are renowned for their difficulty, they frequently reset player progress and allow for shorter gameplay sessions. This contrasts with the more prolonged and continuous gameplay experience of *Elden Ring*.

5. Conclusion and Future Works

While a Multi-Layer Perceptron did not achieve good results in predicting whether a user, based on their gaming history, would enjoy *Elden Ring*, it still provides valuable insights into the target audience. Difficulty alone is not the primary selling point, as players who favor shooters and survival games are inversely correlated with enjoying *Elden Ring*. Additionally, themes of fantasy and longer gameplay sessions are more closely aligned with the preferences of *Elden Ring* enthusiasts.

Overall, the main audience for *Elden Ring* is already familiar with the fantasy genre and tends to avoid other difficult gaming experiences. These players prefer longer sessions of single-player or co-op gameplay, characterized by overcoming bosses and upgrading a single character throughout the game. This indicates a preference for immersive, sustained gameplay over shorter, more repetitive experiences.

Future research could explore the use of convolutional neural networks and incorporate more data on user experiences, such as gaming achievements, computer architecture, and gameplay behaviors prior to purchasing *Elden Ring*. Such studies could further refine the understanding of the factors influencing player enjoyment and enhance predictive models.

6. Code

All Python code used in this project is available at https://github.com/edu-santiago/NN_Elden_Ring

References

- Agarap, A. F. (2019). Deep learning using rectified linear units (relu).
- Chawla, N. V. et al. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357.
- FromSoftware (2024). Elden ring steam page. https://store.steampowered.com/app/1245620/ELDEN_RING/ [Accessed: 2024-05].
- He, C., Ma, M., and Wang, P. (2020). Extract interpretability-accuracy balanced rules from artificial neural networks: A review. *Neurocomputing*, 387:346–358.
- IGN (2022). The game awards 2022 winners: The full list. <https://www.ign.com/articles/the-game-awards-2022-winners> [Accessed: 2024-05].
- Ke, G. et al. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA*.

- Lachowski, V. F. (2024). Desgraça em elden ring: Expressões alegóricas de ruínas e cadáveres.
- Lemaître, G., Nogueira, F., and Aridas, C. K. (2017). Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *Journal of Machine Learning Research*, 18(17):1–5.
- Lundberg, S. and Lee, S.-I. (2017). A unified approach to interpreting model predictions.
- Paszke, A. et al. (2017). Automatic differentiation in pytorch. In *NIPS-W*.
- Pedregosa, F. et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Reddit (2024). r/ elden ring. <https://www.reddit.com/r/Eldenring/> [Accessed: 2024-05].
- Rumelhart, D. E. et al. (1986). Learning representations by back-propagating errors. *nature*, 323(6088):533–536.
- Statista (2024). Video game industry - statistics facts. <https://www.statista.com/topics/868/video-games/#topicOverview> [Accessed: 2024-05].
- Valve (2024). Steam web api documentation. <https://steamcommunity.com/dev> [Accessed: 2024-05].
- Viggiato, M. et al. (2022). What causes wrong sentiment classifications of game reviews? *IEEE Transactions on Games*.
- Wang, X. and Goh, D. H.-L. (2020). Components of game experience: An automatic text analysis of online reviews. *Entertainment Computing*, 33:100338.
- Zhai, X., Mustafa, B., Kolesnikov, A., and Beyer, L. (2023). Sigmoid loss for language image pre-training.