

Prediction of Sales in Video Games using Random Forest Algorithm in Comparison with Gradient Boosting Algorithm to Improve Accuracy

J. Sai Chandu¹, K.V. Kanimozhi²

Department of Computer Science and Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University,

Chennai, Tamil Nadu, India.

saichanduj1216.sse@saveetha.com, kanimozhikv.sse@saveetha.com

Abstract- Aim: With the goal of increasing the precision of purchases in gaming. In order to identify which algorithm most reliably predicts sales in games, this study will assess the efficiency of many methods. **materials and methods:** I used a dataset of video game sales records that included genre, platform, reviewer scores, player scores, and developer data. The number of iterations for each algorithm is ten, and the sample size was found to be ten per group. As a result, $tp=0.121$. Both were trained using identical parameters on a single set of data. **Result:** These investigations demonstrate that the main method and gradient boosting methods have significant precision differences. To be more specific, the Random Forest approach achieved an 86.40% level of precision, substantially greater compared to the gradient boosting method 78.75%. **Conclusion:** The article finds a brief example on how sales is going to happen, the Random Forest method performs better than the gradient boosting algorithm. The gradient boosting had a 78.75% precision, compared to 86.40% for the Random Forest technique. It implies that the intervention method does a great job of assessing sales of games. Therefore, random forest is the suitable option for the prediction of accurate sales in video games.

Keywords: Gradient boosting, random forest, comparison, accuracy, video game, sales, prediction, improvement, machine learning.

Introduction

When compared to the Gradient Boosting technique, the random forest algorithm is used to estimate gaming revenue with more reliability [12]. Perfect revenue forecasting is critical for designers in making informed options in today's world, when industry is rapidly developing & turning into a rival. This looks into the use of purchases information in order to forecast gaming purchases[3].

IEEE Xplore has 196 articles on Forecasting revenue for games, whereas 1750 articles in google scholar. Those documents demonstrate how Random Forest outperforms alternative methods forecasts [7]. In contrast, alternative

methodologies, experts has frequently demonstrated, because the algorithm of Random Forest predicts superior precision & robustness [3]. This result highlights random Forest's ability to capture the deep relationships found in video games sales [9].

The key issue in contemporary documents is the lack of emphasis given to analyzing Gradient Boosting approaches with random forest algorithms [1]. I began my research to discover an answer to this unanswered question because understanding the method that performs best is critical to improving the precision value [5]. My study intends to compare the ability of the algorithms to anticipate sales in video games [4]. My goal is to identify whether a method delivers accurate and trustworthy sales estimates by analyzing historical sales data [2].

Materials and Methods

This recommended task was conducted within SIMATS. This study has found two categories in total. 1st categorie used the random forest method, whereas 2nd one used the Gradient boosting approach. The random forest and Gradient boosting methods run at various intervals. The computation employed an $\alpha = 0.05$ and $\beta = 0.2$.

The set of data used for this study includes the game's name as well as sales statistics from different countries. Pre-processing the dataset regularly increases its quality, resulting in more accurate analysis. Following next steps are various cleaning procedures. The method of feature extraction entails determining and picking the most pertinent characteristics that have a significant effort. The aim is to determine the most crucial elements influencing purchases by focusing on vital components.

Random forest algorithm

This method is trained using a random subset. During tree construction, Random Forest introduces randomness in

two key ways. This randomness helps prevent overfitting. Random Forest offers several advantages, including robustness against overfitting, versatility across different types of tasks, the ability to determine feature importance, and scalability to handle large datasets[2]. Overall, Random Forest is valued for its simplicity, flexibility, and consistently high predictive performance.

Pseudocode for RF algorithm

1. Load necessary libraries for implementation of random forest algorithm
2. Load the dataset onto the software you are working.
3. Split the dataset into development set and the evaluation set
4. Create a random forest object.
5. Train the classifier on the development set of data.
6. Predict labels of testing data.
7. Now compare the development set with the evaluation set.
8. Find precision percentage.

Gradient boosting algorithm

This method constantly refines forecasts and lowering overall forecast error. Gradient Boosting excels in both regression and classification problems thanks to its iterative technique, especially on datasets with complicated, nonlinear interactions. Its versatility to different types of datasets makes it an effective tool for detecting subtle patterns and connections that other algorithms may miss. Gradient Boosting also provides versatility in terms of loss functions, allowing users to customize the learning process to specific objectives. Despite its advantages, Gradient Boosting is susceptible to problems such as overfitting, particularly when applied to small datasets or when hyperparameters are not properly calibrated. To overcome this, procedures such as regularization and cross-validation are commonly used to ensure that the model generalizes effectively to new data. Nonetheless, Gradient Boosting ability is to produce accurate forecasts and successfully handle complex datasets. Its methodical approach to increasing model performance makes it an important addition to the toolbox.

Pseudocode for gradient boosting algorithm

1. Load the necessary libraries.
2. Load a synthetic dataset related to the video game sales.
3. Initialize a gradient boosting classifier.
4. Split the classifier into the development set and evaluation set.
5. Train the classifier on the development set of data.
6. Compare this with evaluation set
7. Find accuracy.
8. Now assess this with rf algorithm

Statistical Analysis

The quantitative Analysis is done with IBM SPSS. I plan to use approaches such as Hypothesis Testing and Cross-Validation to determine the relevance of differences in precision value. My aim is to do quantitative analysis and determine the best technique for increasing the accuracy of video game sales forecast models.

Result

The Random Forest method predicted purchases of games with a precision percentage of 86.50%, which was greater than the Gradient Boosting approach's with precision percentage of 78.75%. rf performs better than Gradient Boosting throughout all iterations in Table 3. Table 4 displays the Random Forest method's results, which have a mean value of 81.50, a sd of 3.027. In contrast, the Gradient Boosting algorithm produces a mean of 73.50, a sd of 3.027. Modify these results in the best way.

Table1: Accuracy values of 10 iterations

Rf accuracy	Gb accuracy
85.2	78.3
84.21	74.5
84.3	75.6
86.4	76.4
85.6	78.9
82.4	87.04
85.9	79.8
86.9	75.2
83.2	73.5
80.2	78.2

Table 2:

	Group	N	Mean
Accuracy	CNN	10	97.81
	KNN	10	78.92

Table 3: Independent Sample Test

		Levene test for variances		t-test		
		f	Sig.	t	D	Sig. 2-tailed
Accuracy	Equal variances assumed	0.54	0.121	5.90	18	.000
	Equal Variances not assumed			5.90	18.0	.000

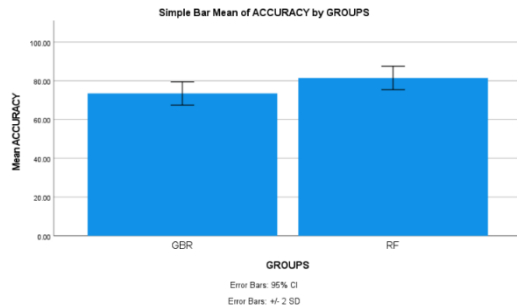


Figure 1: Gradient boosting algorithm vs Random Forest algorithm

Discussion

In the field of gaming purchases forecasting, this study found that the algorithm known as rf exceeds the Gradient Boosting technique, with a precision rate of 86.40% against 78.75%. A paired t-test, confirming the significant difference in accuracy[11]. The rf techniques strength and efficacy in discovering deeper trends in the gaming purchases data is demonstrated by its better accuracy. These findings, particularly in the context of game sales prediction, provide significant evidence for Random Forest's usefulness and dependability in predictive modeling tasks[9].

For an identical line, sandy M. reports that Random Forest outperforms various other methods in comparable prediction tasks[10]. chris E, on the other hand, found contradicting results, arguing that the Gradient Boosting Method outperforms alternative methods in some instances. Despite this gap, a consistent trend across

widely referenced articles reveals that Random Forest is better in each regard.

This study lends confirmation to the prevailing consensus in the scientific community that the algorithm known as Random Forest had the highest accuracy for forecasting gaming purchases [8]. The consistency of these findings with previous investigations demonstrates Random Forest's reliability and effectiveness in this field. Additional study is needed to completely comprehend the intricacies of method choice as well as optimization of parameters. Understanding such details is critical to developing better and practical forecasting models which can accurately depict the complexities of the multimedia game business [7]. Continued research in this field will enhance our understanding of forecasting and lead to more accurate and beneficial conclusions to businesses.

In terms of sales of video games forecasting, the present research demonstrates that the Random Forest methodology always beats the Gradient Boosting method [6]. The outcomes demonstrate that Random Forest has greater analytical potential in this domain, as evidenced by the greater level of accuracy it obtained.

Since sales prediction has a direct impact on revenue and success, it is a crucial component for both publishers and developers of video games [5]. With great accuracy, machine learning algorithms, in particular Random Forest and Gradient Boosting, provide effective tools for sales prediction. Gradient Boosting, on the other hand, creates trees one after the other, fixing the mistakes of the ones before it to create a powerful prediction model. Both systems have advantages and disadvantages when it comes to predicting video game sales[2].

Conclusion

This study found that the algorithm known as Random Forest excels over the Gradient Boosting method in forecasting sales of video games. rf predicted purchases with 86.40% precision, whereas the Gradient Boost algorithm obtained just 78.75% accuracy. Random Forest is more effective at predicting gaming sales. So, for those in the video game sector, Random Forest is an excellent option.

References

- [1]. T. D. Masterson et al., "Prevalence and comparisons of alcohol, candy, energy drink, snack, soda, and restaurant brand and product marketing on Twitch, Facebook Gaming and YouTube Gaming," Public Health Nutrition, vol. 25, no. 1, pp. 1–12, 2022.
- [2]. P. Etchells et al., "Loot box spending is associated with problem gambling but not mental wellbeing," Royal Society Open Science,

- vol. 9, no. 8, p. 220111, 2022.
- [3]. R. Hansch, "Video Games and the Law". Routledge, 2024.
 - [4]. M. Islam, "Effect of internet use and electronic game-play on academic performance of Australian children," *Scientific Reports*, vol. 10, no. 1, p. 21727, [2022].
 - [5]. J. James, "Random forests, sound symbolism and Pokémon evolution," *PloS One*, vol. 18, no. 1, p. e0279350, 2023.
 - [6]. R. Leung, "Designing an eSports intervention for middle-aged and older adults in Hong Kong: Social marketing approach," **PloS One**, vol. 18, no. 4, p. e0284504, 2022.
 - [7]. E. Mao, "How live stream content types impact viewers' support behaviors? Mediation analysis on psychological and social gratifications," vol. 13, no. 10, p. 951055, 2022.
 - [8]. W. K. Moon, "Nintendo: Makers of Mario and Zelda". ABDO, 2023.
 - [9]. K. Thiel, "How Are Video Games Made and Sold?" Cavendish Square Publishing, LLC, 2019.
 - [10]. H. Chen, J.-T. Chiang, N.-F. Miao, H.-Y. Chuang, and C.-C. Tseng, "Excessive gaming and online energy-drink marketing exposure associated with energy-drink consumption among adolescents," **Int. J. Environ. Res. Public Health**, vol. 19, no. 17, p. 10661, 2022. [Online]
 - [11]. Y. Yi, M. Yang, C. Fu, and Y. Li, "Gaming strategies within a green supply chain considering consumers' concern about the greenness and conformance quality of products," **Environ. Sci. Pollut. Res. Int.**, vol. 29, no. 45, pp. 69082–690100, 2022.
 - [12]. D. Zendle, "The changing face of desktop video game monetisation: An exploration of exposure to loot boxes, pay to win, and cosmetic microtransactions in the most-played Steam games of 2010-2019," **PloS One**, vol. 15, no. 5, p. e0232780, 2020.