Title 4:

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

J Sai Chandu[1], K. V. Kanimozhi[2]

J Sai Chandu
research scholar
department of computer science and engineering
SIMATS engineering
SIMATS institute of medical and technical sciences
SIMATS university, chennai, tamil nadu, india, pincode: 602105
saichanduj1216.sse@saveetha.com

K. V. Kanimozhi
Research guide, corresponding author
department of computer science and engineering
SIMATS engineering
SIMATS institute of medical and technical sciences
SIMATS university, chennai, tamil nadu, india, pincode: 602105
kanimozhikv.sse@saveetha.com

Keywords: Prediction, Random Forest Algorithm, Gradient Boosting Algorithm, Accuracy, Video Games, Machine Learning, Sales.

Abstract:
Aim: With the goal of increasing the precision of purchases in gaming, this research contrasts the random forest and Gradient Boosting. 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 dataset, which was divided into training and testing sets. Performance criteria such as precision, recall, and F1 score were used to assess predictive effectiveness. **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(Thiel 2019). Perfect revenue forecasting is critical for designers in making informed options in today's world, when industry is rapidly developing & turning into a rival(Zendle, Meyer, and Ballou 2020). This looks into the use of purchases information in order to forecast gaming purchases(Leung and Chu 2023).

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(Kilpatrick, Ćwiek, and Kawahara 2023). In contrast, alternative methodologies, experts has frequently demonstrated, because the algorithm of Random Forest predicts superior precision & robustness(Kilpatrick, Ćwiek, and Kawahara 2023; Mao 2022). This result highlights random Forest's ability to capture the deep relationships found in video games sales.(Yang et al. 2022).

The key issue in contemporary documents is the lack of emphasis given to analyzing Gradient Boosting approaches with random forest algorithms(Etchells, Morgan, and Quintana 2022). 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(Yi et al. 2022). My study intends to compare the ability of the random forest and Gradient Boosting algorithms to anticipate sales in video games(Islam, Biswas, and Khanam 2020). My goal is to identify whether a method delivers accurate and trustworthy sales estimates by analyzing historical sales data(Moon 2023).

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(Pavlov 2019). 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(Hansch 2024). 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. Overall, Random Forest is valued for its simplicity, flexibility, and consistently high predictive performance.

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 machine learning toolbox for a variety of applications.

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.

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%,(Wade and Glynn 2020). A paired t-test, confirming the significant difference in accuracy. 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(Wade 2020).

For an identical line, sandy M. reports that Random Forest outperforms various other methods in comparable prediction tasks(Saupin 2022). chris E, on the other hand, found contradicting results, arguing that the Gradient Boosting Method outperforms alternative methods in some instances(Sharma 2018). 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(Mao 2022). 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. 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(Edwards et al. 2022). The outcomes demonstrate that Random Forest has greater analytical potential in this domain, as evidenced by the greater level of accuracy it obtained(Gard and Ronald Gard 2017).

Since sales prediction has a direct impact on revenue and success, it is a crucial component for both publishers and developers of video games. 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.

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.

Declaration:

This report has no self serving influences.

Authors Contribution:

JSC was responsible for writing the manuscript, as well as data analysis and collection. KVK was in charge of conceptualizing, validating data, and critically assessing the content.

Acknowledgements:

The authors appreciate the SIMATS for providing the support needed to complete this work effectively.

Funding:

We express our gratitude to the subsequent institutions who provided monetary contributions that have made it possible for us to finish this study.

1. INautix Technologies, India.
2. SIMATS Engineering.
3. SIMATS university.
4. SIMATS institute of medical and technical sciences.

References:

Edwards, Caitlyn G., Catherine C. Pollack, Sara J. Pritschet, Keally Haushalter, John W. Long, and Travis D. Masterson. 2022. "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* 25 (1): 1–12.

Etchells, Peter J., Alexandra L. Morgan, and Daniel S. Quintana. 2022. "Loot Box Spending Is Associated with Problem Gambling but Not Mental Wellbeing." *Royal Society Open Science* 9 (8): 220111.

Gard, Elizabeth Townsend, and W. Ronald Gard. 2017. *Video Games and the Law*. Routledge.

Hansch, Ronny. 2024. *Handbook of Random Forests: Theory and Applications for Remote Sensing*. World Scientific Publishing Company.

Islam, Md Irteja, Raaj Kishore Biswas, and Rasheda Khanam. 2020. "Effect of Internet Use and Electronic Game-Play on Academic Performance of Australian Children." *Scientific Reports* 10 (1): 21727.

Kilpatrick, Alexander James, Aleksandra Ćwiek, and Shigeto Kawahara. 2023. "Random Forests, Sound Symbolism and Pokémon Evolution." *PloS One* 18 (1): e0279350.

Leung, Ka-Man, and William Chu. 2023. "Designing an eSports Intervention for Middle-Aged and Older Adults in Hong Kong: Social Marketing Approach." *PloS One* 18 (4): e0284504.

Mao, Eric. 2022. "How Live Stream Content Types Impact Viewers' Support Behaviors? Mediational Analysis on Psychological and Social Gratifications." *Frontiers in Psychology* 13 (October): 951055.

Moon, Walt K. 2023. *Nintendo: Makers of Mario and Zelda: Makers of Mario and Zelda*. ABDO.

Pavlov, Yu L. 2019. *Random Forests*. Walter de Gruyter GmbH & Co KG.

Saupin, Guillaume. 2022. *Practical Gradient Boosting: A Deep Dive into Gradient Boosting in Python*. guillaume saupin.

Sharma, Nonita. 2018. *XGBoost. The Extreme Gradient Boosting for Mining Applications*. GRIN Verlag.

Thiel, Kristin. 2019. *How Are Video Games Made and Sold?* Cavendish Square Publishing, LLC.

Wade, Corey. 2020. *Hands-On Gradient Boosting with XGBoost and Scikit-Learn: Perform Accessible Machine Learning and Extreme Gradient Boosting with Python*.

Wade, Corey, and Kevin Glynn. 2020. *Hands-On Gradient Boosting with XGBoost and Scikit-Learn: Perform Accessible Machine Learning and Extreme Gradient Boosting with Python*. Packt Publishing Ltd.

Yang, Chung-Ying, Fong-Ching Chang, Ru Rutherford, Wen-Yu Chen, Chiung-Hui Chiu, Ping-Hung Chen, Jeng-Tung Chiang, Nae-Fang Miao, Hung-Yi Chuang, and Chie-Chien Tseng. 2022. "Excessive Gaming and Online Energy-Drink Marketing Exposure Associated with Energy-Drink Consumption among Adolescents." *International Journal of Environmental Research and Public Health* 19 (17). https://doi.org/10.3390/ijerph191710661.

Yi, Yongxi, Min Yang, Chunyan Fu, and Yuqiong Li. 2022. "Gaming Strategies within a Green Supply Chain Considering Consumers' Concern about the Greenness and Conformance Quality of Products." *Environmental Science and Pollution Research International* 29 (45): 69082–100.

Zendle, David, Rachel Meyer, and Nick Ballou. 2020. “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* 15 (5): e0232780.

Table 1: pseudocode for RF algorithm.

Input: video games sales prediction dataset
Output: improved accuracy for video game sales prediction
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.

Table 2: pseudocode for gradient boosting algorithm

Input: video games sales prediction dataset
Output: improved accuracy for video game sales prediction
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

Table 3: Accuracy values of 10 iterations:

Iteration number	Rf accuracy	Gb accuracy
1	85.2%	78.3%
2	84.21%	74.5%
3	84.3%	75.6%
4	86.4%	76.4%
5	85.6%	78.9%
6	82.4%	78.7%
7	85.9%	79.8%
8	86.9%	75.2%
9	83.2%	73.5%
10	80.2%	78.2%

Table 4: group statistics:

group	n	mean	standard deviation	standard error mean
RF	10	81.50	3.027	0.954
GB	10	73.50	3.027	0.957

Table 5: Independent Sample Test

independent samples test										
		levene's test for equality of variances		t-test for equality of means						
		f	sig.	t	df	sig. (2-tailed)	mean difference	std. error difference	95% confidence interval of the difference: lower	upper
accuracy	equal variances assumed	0.52	0.121	5.908	18	0.00	8.00	1.35401	5.15534	10.84464
	equal variances not assumed			5.908	18.00	0.00	8.00	1.35401	5.15524	10.84423

Figure 1: Gradient boosting algorithm **vs** Random Forest algorithm