Title 2:

Efficient Prediction of Sales in Video Games Using Random Forest Algorithm in Comparison with Decision Tree Algorithm to Improve Accuracy.

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

J Sai Chandu
research scholar
BE computer science
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: Video Games, Sales, Prediction, random forest, Decision Tree, accuracy, Comparison.

abstract:
Aim: To increase the precision percentage of sales in video games prediction, this study compares the Random Forest and Decision Tree Techniques. This paper compares working of several algorithms and identify the most accurate one for video game sales prediction. **Materials and Methods:** This study examined two distinct methods for properly predicting the number of video game copies that would be sold: the RF and Decision Tree algorithms. We examined the performance of these two approaches by looking at video games sales data. Since there are ten iterations for each algorithm and ten sample sizes per group, the p=0.124. These results provide information about whether the approach is better at correctly assuming video game sales. **Result:** This investigation reveals a significant difference in precision level of the RFand DT algorithms. More specifically, the RF achieved an accuracy score of 86.40%, which was notably higher than the 79.32% accuracy rate of the Decision tree algorithm. **Conclusion:** In this research's conclusion, the superiority of the RF method over the Decision Tree method for predicting video game sales is clearly demonstrated. The RF technique obtained 86.40% accuracy compared to the Decision tree algorithm's 79.32%. This suggests that RF is more accurate in analyzing sales of video games. For professionals in the video game sector who need to make exact sales projections, Random Forest is therefore a superior choice.

Keywords: Comparison, Decision Tree, Video Games, random forest, Sales Prediction, Accuracy Improvement, Machine Learning.

Introduction:

Compared to the Decision Tree method, the Random Forest method is used to estimate sales in video games with the aim of improving accuracy(Kowert and Quandt 2020). In today's world, where the gaming sector is evolving swiftly and becoming more competitive, game designers, advertisers, and marketers need to be able to make informed decisions. This requires accurate sales projection(Mohan et al. 2023). This work looks into using machine learning techniques to examine historical sales data in order to forecast future trends in video game sales.

In total, there are 220 papers in Google Scholar and 1520 in IEEE Xplore on the topic of sales of video games forecasting using machine learning techniques(Thiel 2019). These articles show that Random Forest routinely outperforms other methods in sales of video games forecasting(Parnell et al. 2023). Researchers have consistently shown that the Random Forest method has greater anticipated accuracy and durability when compared to other methods. This pattern shows how successfully the complex connections and patterns found in video game sales data are captured by Random Forest(Leung and Chu 2023). In addition, Random Forest's flexibility and scalability make it a great option for handling large datasets and adapting to the game industry's constant change(Acton et al. 2023).As a result, many researchers who want to develop accurate and consistent revenue projection models for the video game business have found that Random Forest is the best algorithm.

A significant issue with recent publications is how little thought is given to comparing the Random Forest and Decision Tree approaches, especially when it comes to projecting purchases of video games(Kilpatrick, Ćwiek, and Kawahara 2023). I began this study in an attempt to address this open-ended problem because determining which algorithm performs best in this scenario is critical to improving the accuracy of sales prediction in the gaming industry(Mao 2022). My study attempts to assess and compare the video game sales forecasting capabilities of Random Forest and Decision Tree algorithms. My goal is to ascertain whether the approach generates accurate and trustworthy sales estimates by looking over historical sales data and evaluating both algorithms.

Materials and Methods:

I finished this suggested task in the SSE Lab. For this title, a total of two parts has acknowledged. While part 2 employed a decision tree, part 1 employed a technique called random forest (Yang et al. 2022). random forest & Decision tree methods are applied at varied stages to a dataset of items and a sample size of 20. The computation involved the use of α (0.05) and β (0.2).

A set of information utilized for this study are the game's name and sales information of several nations, including Japan, Europe. These features are essential for increasing the Precision of purchases. This dataset is continuously improved via preprocessing, that makes the precision more reliable.(Etchells, Morgan, and Quintana 2022). The method of FE involves identifying and selecting the most relevant attributes that significantly affect the accuracy in sales forecasts. With an emphasis on vital components, this goal is to pinpoint the important factors influencing sales performance.

Random Forest Algorithm:

In this study, I employed this to forecast purchases of games, and it produces a lot of decision trees during training. Every tree is built independently utilizing some samples and attributes from the dataset in order to produce assumptions. The algorithm integrates the forecasts from each tree to produce a last forecast. By using this ensemble method, prediction accuracy is increased and overfitting is decreased. Because each tree has more variation due to the algorithm brought forth in sample selection, robust assumptions are generated. Given its ease of handling large datasets and reduced susceptibility to bias, Random Forest is a strong option for prediction.

Decision Tree Algorithm:

A supervised learning technique used in machine learning for both regression and classification problems is the Decision Tree algorithm. In order to make judgments, it works by recursively dividing the feature space into subsets according to the values of input features. A class label or regression value is correlated with each leaf node in the tree, which represents a feature and a decision rule. By increasing information gain or decreasing impurity metrics like entropy or Gini impurity, the algorithm finds the best decision rules at each node. Until a halting condition is satisfied, like reaching a maximum tree depth or seeing no further progress in impurity reduction, this procedure keeps going. Because they are simple to use and intuitive, decision trees are a good tool for situations where it's critical to comprehend the decision-making process. When trained on noisy or complicated datasets, they are prone to overfitting, however this can be minimized with the help of strategies like pruning or ensemble methods. Notwithstanding their drawbacks, Decision Trees are nevertheless a popular and adaptable technique in many machine learning applications.

statistical analysis:

The tool utilized for evaluation is IBM SPSS. Utilizing techniques like validation for hypotheses. Moreover, this significance of feature_evaluation will be carried out to determine the variables impacting sales projections for each method. By doing a comprehensive statistical analysis, I hope to identify the optimal methods for improving the accuracy of sales of video games forecast models.

Results:

The random forest Algorithm outperformed the Decision Tree algorithm in sales prediction of video games. It is clear that random forest continuously performs better than Decision Tree throughout all iterations shown in Table 3. An average of 81.50, an SD of 3.02, are displayed by the Random Forest algorithm in Table 4. The Decision Tree, on the other hand, shows an average of 74.50 along with SD of 3.027. These findings implies when contrasted to the Decision Tree technique, The random forest approach generally obtains more precision, exhibits small differences. Figure 1 illustrates how random forest Algorithm sells in Video games more accurately than the Decision Tree approach.

discussion:

This study demonstrates, when it comes to Video Game Sales forecast, RF works more efficiently than DT method(Yi et al. 2022). Through statistical Analysis, This considerable accuracy difference was confirmed; a paired t-test produced a p-value < 0.05. The algorithm's increased accuracy serves as a testament to its excellence and effectiveness in revealing the deeper trends in the set of data(Tabares-Tabares et al. 2022). These findings provide significant evidence for Random Forest's utility and reliability in predictive tasks, specifically in the field of video game sales prediction.

The findings of this study validate previous work by Joseph et al. by demonstrating that random forest method has a higher accuracy(Lupinek et al. 2021). Joseph et al. have found that Random

Forest outperforms decision trees in analogous challenges. However, research like Erik K. 's shows conflicting results, supporting Decision Tree as a superior option in particular circumstances(Islam, Biswas, and Khanam 2020). Despite the disparity, extremely referenced articles shares the same pattern, random forest is better in all aspect.

This article held belief in that the algorithm known as random forest haves the highest predictive value. The consistency of the findings with those of previous studies demonstrates the dependability and effectiveness of Random Forest in this domain(Labrador et al. 2021). Further investigation is still required to completely comprehend the subtleties of Parameter Optimization and algorithm selection. To develop more dependable and widely applicable prediction algorithms that can accurately capture the complexities of the video game industry, it is imperative to explore these subtleties. Continuing working in this field will enhance our understanding and provide more accurate and practical insights(Moon 2023).

According to this article, random forest methodology outperforms the Decision Tree method in forecasting video game sales(Schöning et al. 2023). These results demonstrate that Random Forest has superior prediction abilities in this field(Fisher 2021). The statistical significance discovered in this investigation attests to Random Forest's supremacy and verifies its accuracy.(Castillo 2023). Overall, my results demonstrate random Forest's potential as a top choice for modeling and prediction positions in the video game business, highlighting its versatility and shown effectiveness in properly predicting sales.

Conclusion:

The results show that when it comes to video game sales prediction, the Random Forest algorithm outperforms the Decision Tree method. Random Forest did substantially better than Decision Tree, which had an accuracy of 79.32%, with an accuracy of 86.40%. This significant gap highlights how much better Random Forest is at identifying patterns in video game sales data. As such, Random Forest is the best option for video game industry professionals looking for accurate sales forecasts. Its capacity to deliver precise projections is essential for well-informed decision-making and tactical planning in the fiercely competitive and rapidly evolving video game industry. Industry participants can develop successful strategies and stay ahead of the curve by utilizing Random Forest's predictive capabilities to obtain insightful information on consumer preferences, market trends, and sales performance. Furthermore, Random Forest is a vital tool for managing the complexity and uncertainties present in the ever-changing video game business. This is due to its robustness and dependability. Therefore, using Random Forest to anticipate sales has the potential to be very beneficial, giving business managers the ability to make data-driven decisions and succeed in a constantly changing market.

Declaration:

There are no dual loyalties with this document.

Authors Contribution:

JSC had been in control of writing the manuscript, as well as data analysis and collection. Author KVK was in charge of conceptualizing, validating data, and critically evaluating the content.

Acknowledgements:

The authors thank the SIMATS engineering, SIMATS institute of medical and technical science for supplying the essential support to complete this study effectively.

Funding:

The ensuing institution's contributions allowed me to finish this research, and we are grateful for their assistance.

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

References:

Acton, Rachel B., Mariangela Bagnato, Lauren Remedios, Monique Potvin Kent, Lana Vanderlee, Christine M. White, and David Hammond. 2023. "Examining Differences in Children and Adolescents' Exposure to Food and Beverage Marketing in Canada by Sociodemographic Characteristics: Findings from the International Food Policy Study Youth Survey, 2020." *Pediatric Obesity* 18 (6): e13028.

Castillo, Angel. 2023. *Free Fire: Advanced Strategies: Tips and Tricks to Dominate the Battlefield in the Popular Battle Royale Game*. farfangram.

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.

Fisher, Magnus. 2021. *CHESS TACTICS AND MOVE PREDICTION: Beginners Guide to Strategies and Basics Opening and Closing Tactics! Learn How to Visualize the Game and Predict Your Opponent's Intentions!*

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.

Kowert, Rachel, and Thorsten Quandt. 2020. *The Video Game Debate 2: Revisiting the Physical, Social, and Psychological Effects of Video Games*. Routledge.

Labrador, F. J., M. Bernaldo-de-Quirós, I. Sánchez-Iglesias, M. Labrador, M. Vallejo-Achón, I. Fernández-Arias, and F. J. Estupiñá. 2021. "Advertising Games of Chance in Adolescents and Young Adults in Spain." *Journal of Gambling Studies / Co-Sponsored by the National Council on Problem Gambling and Institute for the Study of Gambling and Commercial Gaming* 37 (3): 765–78.

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.

Lupinek, Joshua M., Jinhee Yoo, Eugene A. Ohu, and Eric Bownlee. 2021. "Congruity of Virtual Reality In-Game Advertising." *Frontiers in Sports and Active Living* 3 (October): 728749.

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.

Mohan, Deepika, A. James O'Malley, Julia Chelen, Meredith MacMartin, Megan Murphy, Mark Rudolph, Jaclyn A. Engel, and Amber E. Barnato. 2023. "Using a Video Game Intervention to Increase Hospitalists' Advance Care Planning Conversations with Older Adults: A Stepped Wedge Randomized Clinical Trial." *Journal of General Internal Medicine* 38 (14): 3224–34.

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

Parnell, Stephanie A., Joelie Mandzufas, Justine Howard, Anna T. Gannett, and Gina S. A. Trapp. 2023. "A Massive Hit That Targets Kids Quite a Bit: Where and How Australian School Children See Energy Drinks." *Health Promotion Journal of Australia: Official Journal of Australian Association of Health Promotion Professionals* 34 (4): 736–41.

Schöning, Julius, Jan Kettler, Milena I. Jäger, and Artur Gunia. 2023. "Grand Theft Auto-Based Cycling Simulator for Cognitive Enhancement Technologies in Dangerous Traffic Situations." *Sensors* 23 (7). https://doi.org/10.3390/s23073672.

Tabares-Tabares, Marcela, Luis A. Moreno Aznar, Virginia Gabriela Aguilera-Cervantes, Edgar León-Landa, and Antonio López-Espinoza. 2022. "Screen Use during Food Consumption: Does It Cause Increased Food Intake? A Systematic Review." *Appetite* 171 (April): 105928.

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

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.

Table 1: pseudocode for random forest algorithm

Input: video games sales prediction dataset
Output: improved accuracy for video game sales prediction
step 1: To use the Random Forest, load packages such as random_forest_classifier. step 2: To anticipate purchases, work on software and import the CSV containing the data. step 3: Divide the set of data, validation and coaching subsets. step 4: Initialize a random forest Classifier with the parameters for finding total assumers, the total no.of trees. step 5: The method creates several decision trees of training data after training the Random Forest Model with the coaching data. step 6: Make assumptions based on the coaching and validation data. step 7: Compare the actual and anticipated sales amounts with validation data. step 8: Determine precision, evaluate it against alternative methods.

Table 2: pseudocode for decision tree algorithm

Input: video games sales prediction dataset
Output: improved accuracy for video game sales prediction
Step 1: To represent a Decision Tree, create a class called Tree. Step 2: Create a method that uses the dataset to initialize the Decision Tree. Step 3: Put into practice a technique to determine the dataset's entropy. Step 4: Create a method to identify the feature for data splitting that optimizes information gain. Step 5: Make a function that splits the dataset according to the chosen attribute. Step 6: Specify a procedure to create child nodes for every conceivable attribute value. Step 7: Conduct steps 3-6 iteratively for every child branch until a halting condition is satisfied. Step 8: Create a function that uses the Decision Tree that has been built to categorize new instances. Step 9: stop.

Table 3: Increased accuracy in the prediction of video game sales (86.40% accuracy in the Random Forest algorithm and 79.32% accuracy in the decision tree technique).

Iteration number	Rf accuracy	DT accuracy
1	85.2%	79.1%
2	84.21%	78.9%
3	84.3%	75.4%
4	86.4%	72.6%
5	85.6%	72.5%
6	82.4%	73.5%
7	85.9%	72.1%
8	86.9%	75.8%
9	83.2%	74.5%
10	80.2%	79.3%

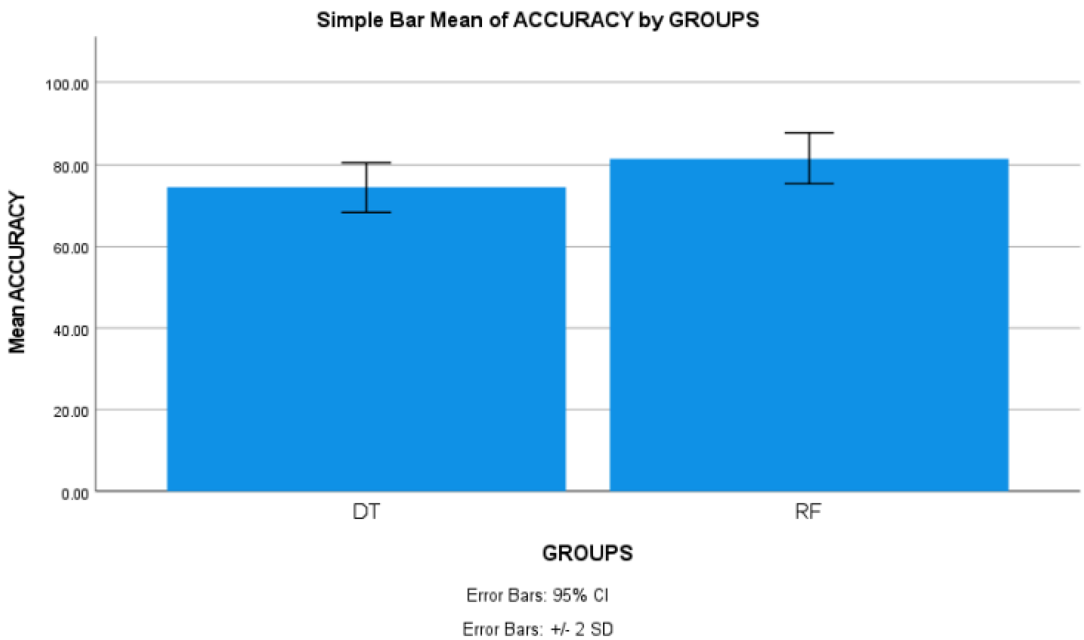
Table 4: group statistics

Group	n	mean	standard deviation	standard error mean
RF	10	81.50	3.02765	0.95743
DT	10	74.50	3.02765	0.954723

Table 5: independent sample test

independent samples test										
		levene's test for equality of variances		t-test for equality of means						
									95% confidence interval of the difference	
		f	sig.	t	df	sig. (2-tailed)	mean difference	std. error difference	lower	upper
accuracy	equal variances assumed	2.4	0.124	5.170	18	0.00	7.00	1.35401	4.15534	9.84466
	equal variances not assumed			5.170	18	0.00	7.00	1.35401	4.15534	9.84466

Figure 1:



Simple Bar Mean of ACCURACY by GROUPS
100.00
80.00
60.00
40.00
20.00
0.00
Mean ACCURACY
DT
RF
GROUPS
Error Bars: 95% CI
Error Bars: +/- 2 SD