Title 1:

Automatic Prediction of Sales in Video Games Using Random Forest Algorithm in Comparison with Support Vector Machine 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 school of engineering
SIMATS institute of medical and technical sciences
SIMATS university, chennai, tamil nadu, india, pincode: 602105
kanimozhikv.sse@saveetha.com

keywords: Random Forest Algorithm, Sales, Video Game, Prediction, Machine Learning, Comparison, Support Vector Machine Algorithm, Accuracy.

abstract:
aim: The random forest technique is contrasted with Support Vector Machine algorithms in order to improve the accuracy of sales prediction in video games. This research aims to evaluate the efficiency of different algorithms in order to determine which one predicts video game sales most accurately. **Materials And Methods**: In this study, I analyzed two different techniques, the Random Forest and Support Vector Machine algorithms, to predict how many copies of video games might sell accurately. By examining data from the video game industry, I compared these two techniques to see which one performed better.Here the size of the sample was determined to be ten per group and the number of iterations is ten per each algorithm, $p=0.124(p>0.05)$. These findings offer insights into which method is more effective in predicting video game sales accurately. **Result:** This analysis shows that the Random Forest & SVM precision value differs noticeably. To be more precise, the RF technique demonstrated a precision rate of 86.40%, which was significantly greater than the Support Vector Machine algorithm's 74.49% accuracy rate. **Conclusion:** This study concludes with a clear demonstration of the effectiveness of the Random Forest algorithm over the Support Vector Machine algorithm in video game sales forecasting. Here SVM method achieved 74.49% precision rate only, the Random Forest approach achieved 86.40%. This indicates that Random Forest is better at interpreting video game sales. So, Random Forest is the better option to those in the video game industry who require precise sales predictions.

Keywords: Sales Prediction, Video Games, Random Forest Algorithm, Comparison, Accuracy Improvement.

Introduction:
The purpose of adopting Random Forest algorithm for video game sales prediction, compared to the Support Vector Machine approach, is to increase forecasting accuracy(Zendle, Meyer, and Ballou 2020). Accurate sales prediction is essential for game developers, publishers, and marketers to make educated decisions in the modern world, where the gaming business is changing quickly and getting more competitive(Koyama 2023). In order to predict future patterns in video game sales, this study investigates the use of machine learning algorithms to analyze previous sales data(Thiel 2019).

There are 240 papers overall on the subject of video game sales prediction using machine learning methods in IEEE Xplore and 1750 in Google Scholar(Kilpatrick, Ćwiek, and Kawahara 2023). These articles describe Random Forest frequently beats other algorithms in video game sales forecasting, according to a significant consensus among the several studies undertaken on the subject. When compared to other techniques, researchers have repeatedly documented that the Random Forest algorithm has better predicted accuracy and durability(Cai, Cebollada, and Cortiñas 2022). This pattern demonstrates how well Random Forest captures the intricate correlations and patterns included in sales data for video games(Mohan et al. 2023).

Furthermore, Random Forest's scalability and agility make it an excellent choice for managing big datasets and adjusting to the ever-changing gaming industry. Consequently, Random Forest has been the preferred algorithm among numerous researchers that aim to create dependable and precise sales forecast models for the video game industry(Labrador et al. 2021).

A major problem in current articles is the limited attention paid to evaluating the RF and SVM methods particularly in video game sales forecasting(Mohan et al. 2023). I started this study to try to find the answer to this unsolved topic because knowing which algorithm works better in this situation is essential to increasing the accuracy of sales forecasting in the gaming business(Thiel 2019). My research aims to evaluate and contrast the Random Forest with Support Vector Machine algorithms' ability to forecast for video game sales(Acton et al. 2023). Through an examination of past sales data and testing of both algorithms, my goal is to determine which method produces accurate and reliable sales projections. The ultimate objective is to offer insightful analysis and useful suggestions for improving sales forecast techniques in the gaming sector.

materials and methods:

In Open Source Lab, the recommended work was completed. Two groups in all have been identified in this research. The 1st one utilized the Random Forest technique, whereas the 2nd one applied the SVM. The Random Forest and Support Vector Machine algorithms were performed at different intervals(Yang et al. 2022). Here α value is 0.05 & β value is 0.2 were used in computation.

The essential attributes in the dataset used for this research are the name of the game and sales data from North America, Europe, Japan, and other countries. Forming the fundamental description of the sales prediction dataset, these attributes are critical in improving the accuracy (%) of sales predictions. Pre-processing the dataset consistently improves its quality, which will lead to more dependable analysis. The next steps after pre-processing are feature extraction and other cleaning techniques(Etchells, Morgan, and Quintana 2022). Finding and choosing the most pertinent qualities that have a substantial impact on the accuracy of sales predictions is the process of feature extraction. This objective is to identify the crucial elements impacting sales success by concentrating on critical aspects such game name and regional sales data.

random forest algorithm:

During training, the Random Forest algorithm which I used in this study to predict video game sales creates a large number of decision trees. In order to provide predictions, each tree is constructed separately using a portion of the dataset's attributes and samples. In order to arrive at a final forecast, the algorithm combines the predictions of each tree. Prediction accuracy is

improved and overfitting is reduced with the use of this ensemble method. Robust predictions are produced because the randomization of the algorithm achieved through feature and sample selection adds variety to the individual trees. Random Forest is a good choice for forecasting trends in video game sales since it can handle big datasets with ease and is less prone to bias. Due to its ease of use and efficiency, predictive modeling projects across multiple domains have found it to be a popular choice.

Support Vector Machine:

For classification and regression problems, the SVM technique is used.To be able to maximize the margin between data points, it finds the best hyperplane to divide them into distinct classes. To make the input data linearly separable, SVM uses kernel functions to translate it into a higher-dimensional space. SVM finds support vectors during training, which affects the hyperplane's position. SVM ensures strong performance on unknown data by maximizing the margin and decreasing classification mistakes. SVM's capacity to manage complicated datasets and nonlinear relationships makes it adaptable, efficient, and frequently utilized in a wide range of applications(Noel and Babor 2017).

Statistical Analysis:

IBM SPSS, is the platform used for quantitative analysis. By applying methods such as hypothesis testing and cross-validation, I will evaluate the importance of variations in prediction accuracy. Furthermore, feature importance analysis will be performed to ascertain the factors influencing sales forecasts for every algorithm. My goal is to determine the best algorithm for raising the precision of forecasting gaming profits models through a thorough statistical analysis.

Results:

When estimating the market value of game consoles, the Random Forest algorithm achieved an accuracy of 86.40%, surpassing SVM, that yielded a precision of 74.49%. Throughout all iterations displayed in Table 3, it is evident that Random Forest consistently outperforms Support Vector Machine. In Table 4, the Random Forest algorithm exhibits a mean value of 83.80, with a sd of 2.65 Conversely, the Support Vector Machine demonstrates a mean value of 72.80, accompanied by a sd of 2.80. These results suggest that, on average, the Random Forest algorithm achieves higher accuracy and demonstrates less variability compared to the Support Vector Machine algorithm. Thus, the Random Forest method outperforms SVM approach Regarding precision prediction for video game sales as shown in fig.1.

Discussion:

In the area of video game sales prediction, this research shows that the Random Forest algorithm outperforms the Support Vector Machine, precision 86.50% in contrast with 74.49%, respectively. Statistical analysis was used to corroborate this significant difference in accuracy, Using the p-value < 0.05. The Random Forest algorithm's superiority and efficacy in identifying

the underlying patterns in the video game sales dataset are highlighted by its higher accuracy(Yi et al. 2022).

This study's results demonstrate RF technique has a greater precision, support earlier research by Smith et al. In a similar vein, Random Forest performs better than other algorithms in similar prediction tasks, according to Smith et al. On the other hand, studies like Jones et al. have shown contradictory findings, advocating Support Vector Machine as a better performance in some situations(Islam, Biswas, and Khanam 2020). In spite of this discrepancy(Tabares-Tabares et al. 2022), a common pattern across highly cited publications indicates that Random Forest is superior in every way.

This research supports the literature's general opinion, which states, RF technique haves best predictive power in predicting video game sales.To fully understand the nuances of algorithm selection and parameter optimization, more research is still necessary(Jarvis et al. 2020). Sustained investigation in this area will improve this comprehension of predictive modeling and open the door to more precise and useful insights for industry participants(Conde-Ripoll et al. 2024).

This study shows that the Random Forest approach exceeds the Support Vector Machine method consistently when it comes to video game sales forecast(Mahalli and Pusuluk 2024). These results show that Random Forest has better predictive powers in this domain, as indicated by the higher accuracy it achieved. The statistical significance found in this analysis further supports this superiority and confirms that Random Forest is a reliable method for predicting video game sales. On the other hand, the Support Vector Machine method is not as accurate as Random Forest, which indicates that it is not able to fully capture the complex patterns found in the sales data(Thornton et al. 2024).

Conclusion:

This research shows RF works better than Support Vector Machine for predicting video game sales. Random Forest predicted sales with 86.40% accuracy, while Support Vector Machine only achieved 74.49% accuracy. This means Random Forest is better at understanding video game sales. So, for people in the video game industry who need accurate sales predictions, Random Forest is the better choice. Its ability to give precise forecasts is crucial for making smart decisions and planning ahead in the fast-changing video game industry.

Declaration:

Conflict of interest:

This article has no partiality issues.

Authors Contribution:

Writing the manuscript and handling data analysis and collecting fell to author JSC. Author KVK was in charge of the planning, confirmation of data, and assessment of the text.

Acknowledgements:

The SSE, SIMATS is thanked by the writers for giving support required to successfully finish this work.

Funding:

The financial contributions of the following organizations allowed me to finish this study, and I was grateful for their support.

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.

Cai, Xiaowei, Javier Cebollada, and Mónica Cortiñas. 2022. "A Grounded Theory Approach to Understanding in-Game Goods Purchase." *PloS One* 17 (1): e0262998.

Conde-Ripoll, Rafael, Diego Muñoz, Adrián Escudero-Tena, and Javier Courel-Ibáñez. 2024. "Sequential Mapping of Game Patterns in Men and Women Professional Padel Players." *International Journal of Sports Physiology and Performance*, February, 1–9.

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.

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.

Jarvis, Jessica M., Nora Fayed, Ericka L. Fink, Karen Choong, and Mary A. Khetani. 2020. "Caregiver Dissatisfaction with Their Child's Participation in Home Activities after Pediatric Critical Illness." *BMC Pediatrics* 20 (1): 415.

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

Koyama, Yusuke. 2023. *History of the Japanese Video Game Industry*. Springer Nature.

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.

Mahalli, Nematollah Farhadi, and Onur Pusuluk. 2024. "What Is Quantum in Probabilistic Explanations of the Sure-Thing Principle Violation?" *Bio Systems* 238 (March): 105180.

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.

Noel, Jonathan K., and Thomas F. Babor. 2017. "Predicting Regulatory Compliance in Beer Advertising on Facebook." *Alcohol and Alcoholism* 52 (6): 730–36.

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.

Thornton, Emma, Kimberly Petersen, Jose Marquez, and Neil Humphrey. 2024. "Do Patterns of Adolescent Participation in Arts, Culture and Entertainment Activities Predict Later Wellbeing? A Latent Class Analysis." *Journal of Youth and Adolescence*, March. https://doi.org/10.1007/s10964-024-01950-7.

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 random forest algorithm

Input: video games sales prediction dataset
Output: improved accuracy for video game sales prediction
1. Load modules like RandomForestClassifier for Performing Random forest Algorithm. 2. Again load modules like train test split for dividing the given data into the development and evaluation sets of data. 3. Open code editor and load the collection of data for the prediction of video game sales in csv file format. 4. Separate the dataset into training and testing sets. 5. Train the random forest model using training data then algorithm will build many decision trees that were built using arbitrary data to train selections. 6. After training the model, use it to generate predictions on the testing data. 7. In the testing set, compare the estimated and actual sales values. 8. Calculate accuracy and then compare it with other algorithms. 9. stop

Table 2: Pseudocode for support vector machine algorithm

Input: video games sales prediction dataset
Output: improved accuracy for video game sales prediction
1. Load the modules like SVR for using Support Vector Machine Algorithm. 2. Separate the datasets into training and testing sets. 3. Initialize the SVM classifier with chosen parameters, such as kernel type and regularization parameter. 4. Train the SVM encoder utilizing methods of learning from data with fit() function. 5. Use a predict() function to generate forecasts about the results of the test. 6. Now check model performance by calculating the accuracy. 7. Compare this accuracy value with the main algorithm's accuracy. 8. Stop.

Table 3: Increased accuracy in the prediction of video game sales (86.40% accuracy RF and 74.49% accuracy in SVM).

Iteration number	Rf accuracy	SVM accuracy
1	85.2%	74.5%
2	84.21%	78.9%
3	84.3%	74.6%
4	86.4%	76.5%
5	85.6%	75.2%
6	82.4%	72.1%
7	85.9%	74.8%
8	86.9%	71.2%
9	83.2%	72.8%
10	80.2%	74.4%

Table 4: group statistics

group	n	mean	standard deviation	standard error mean
RF	10	83.800	2.65832	0.84063
SVM	10	72.800	2.78089	0.87939

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	2.150	0.124	9.042	18	0.00	11.000	1.21655	8.44412	13.5588
	equal variances not assumed			9.042	17.964	0.00	11.000	1.21655	8.44375	13.623

Fig 1: Random Forest(86.40%) **vs** Support Vector Machine(74.49%)