Title Page:

Enhancing the prediction of Video Game using linear Regression compared with Support vector machine.

M. Deva[1], S. Padmakala[2]

M. Deva[1]

Student,

Department of Artificial Intelligence and Data Science,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.

192124170.sse@saveetha.com

S. Padmakala[2]

Project Guide, Corresponding Author,

Department of Computer Science and Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pincode: 602105.

drspadmakala@gmail.com

keywords: Prediction, Machine Learning, Video game, Linear Regression,Support Vector Machine.

ABSTRACT

Aim: Enhancing the prediction of Video Game using linear Regression compared with Support vector machine. **Materials and Methods:** The research aimed to compare the accuracy score and precision score of these algorithms to determine the most effective one for customer segmentation.For the pre-test analysis, parameters such as a G power of 80%, a threshold of 0.05%, and a confidence interval of 90% for mean and standard deviation were considered. The sample size calculation determined a total of 20 samples (10 for Linear Regression and 10 for Support Vector Machine). The standard deviation for Linear Regression was found to be 88.56%, while for Support Vector Machine, it was 81.30%, as per Stamp (2017).**Result:** The group statistics for Linear Regression with Support vector machine are provided, with organization based on iterations and a sample size of 10. The statistical summary encompasses a mean accuracy and precision of 88.5690%, a standard deviation of 1.13933, and a standard error mean of 0.36029.**Conclusion**: The current investigation emphasized the evaluation of machine learning algorithms, specifically highlighting the preference for Linear Regression over Support Vector Machine to achieve superior classification in customer segmentation. Potential future improvements could involve a more in-depth analysis of Support Vector Machine datasets. The study findings reveal that Linear Regression demonstrated a higher accuracy at 88.5690% compared to Support vector machine, which achieved an accuracy of 81.3050%.

INTRODUCTION

Enhancing Video Game Engagement Prediction: A Comparative Analysis of Linear Regression and Support Vector Machines.Utilize the insights gained from the study to inform the design of video games, tailoring gameplay experiences based on predictive models to enhance player engagement.Implement predictive models to dynamically adjust the difficulty of the game based on individual player profiles, ensuring an optimal and challenging experience for each player.Assess and compare the predictive capabilities of Linear Regression and Support vector machine in forecasting video game engagement levels based on diverse player centric features.Gather a robust data-set encompassing a diverse range of player activities, demographics, and achievements to ensure a comprehensive understanding of player engagement for machine learning.

Utilize academic databases such as PubMed, IEEE Xplore, ACM Digital Library, and Google Scholar to search for relevant studies. Use keywords such as "video game prediction," "linear regression," and "Support Vector Machines" to narrow down your search.This research endeavors to advance the field of customer data segmentation in E-Commerce by employing clustering techniques, with a specific focus on utilizing customer feedback as a distinguishing factor. A comprehensive review of the literature was conducted, analyzing 962 journals from IEEE Xplore, 563 articles from ScienceDirect, 1156 articles from Google Scholar, and 459 articles from Springer. The study emphasizes the correlation between customer feedback, particularly using product ratings, and its role in effective segmentation.

The literature review reveals certain lacunae in the existing research landscape. While studies such as Shastri and Pathak (2018) acknowledge the importance of customer feedback, there is a notable gap in understanding the specific nuances of video game engagement prediction using linear regression and Support vector machine. Existing works might lack a comparative analysis of these techniques, leaving room for further exploration into their strengths, limitations, and practical applications in the gaming domain. This study aims to bridge these gaps and provide a more comprehensive understanding of predictive modeling in video games.The research team comprises experts in the fields of machine learning, gaming analytics, and predictive modeling. Their collective expertise is demonstrated through prior successful projects in gaming analytics and the development of predictive models. The team's proficiency in both linear regression and Support vector machine positions them well to conduct a thorough comparative analysis and contribute novel insights to the predictive modeling landscape in the gaming industry.Explore practical applications of predictive models in the gaming industry, considering factors like dynamic difficulty adjustment and personalized content recommendations.

MATERIALS AND METHODS

The research was conducted in the Image Processing Laboratory at the Department of Computer Science and Engineering, Saveetha School Of Engineering, SIMATS. The study employed two groups of classifiers, specifically Linear Regression and Support Vector Machine, to classify customer segmentation in E-commerce. Group 1 utilized the Linear Regression algorithm with a sample size of 10, while Group 2 employed the Support Vector Machine with a sample size of 10. The comparison focused on accuracy and precision scores to determine the optimal algorithm.

The pre-test analysis was meticulously prepared with a G power of 80%, a threshold of 0.05%, and a confidence interval of 90% for both mean and standard deviation. The sample size calculation resulted in 10 sample groups, totaling 20 samples. The standard deviation was identified as 88.56% for Linear Regression and 81.30% for Support Vector Machine, as per Stamp (2017).

Linear Regression :

The Linear Regression algorithm, a supervised learning technique in machine learning, is specifically designed to discern and characterize the linear relationship between a dependent variable and one or more independent variables. The term "linear" is indicative of the algorithm's ability to identify and quantify the linear connection between these variables.

In essence, linear regression seeks to determine the optimal line of best fit that represents the relationship between the dependent variable and the independent variable(s).

Through the analysis of data and the determination of coefficients (m and b), linear regression facilitates an understanding of how alterations in the independent variable(s) correspond to changes in the dependent variable. This makes linear regression a valuable tool for prediction and modeling relationships in diverse fields, encompassing statistics, economics, and machine learning.

Support Vector Machine:

Random Forest, an ensemble learning method, was chosen for its ability to build a multitude of Support vector machines and aggregate their predictions. In the context of predicting used bike prices, Random Forest harnesses the collective wisdom of diverse trees to create a robust and accurate model that can handle complex relationships within the dataset. The dataset, rich with features such as city, kilo meters driven, owner, and brand, underwent preprocessing to ensure compatibility with Random Forest's ensemble structure. The algorithm's behaviour involved constructing multiple Support vector machines, each trained on a random subset of the dataset with replacement. During tree construction, a random subset of features was considered at each split, promoting diversity among the trees. The predictions from individual trees were then aggregated to provide the final prediction for used bike prices.

Random Forest's behaviour with this dataset results in a powerful model capable of capturing intricate patterns and interactions among features. The ensemble approach mitigates overfitting and enhances predictive accuracy, making it well-suited for the nuanced task of predicting used bike prices in a diverse market.

STATISTICAL ANALYSIS

In the pursuit of optimizing predictive models for used bike prices, a comprehensive statistical analysis was conducted to assess the performance of the selected machine learning algorithms—Linear regression and Support vector machine. Executed on a system with Windows 11 64-bit OS, 8 GB of RAM, and a robust internet connection, the project was implemented using Python 3.11 with data preprocessing and visualization tasks executed on Spyder IDE, and model building and testing carried out on Colab Notebook for its cloud-based infrastructure. The statistical assessment focused on precision percentage as the key metric to evaluate the accuracy of the models in predicting used bike prices. Leveraging a diverse dataset encompassing variables such as city, kilo meters driven, owner, and brand, the analysis aimed to provide a comprehensive understanding of the algorithms' predictive models.

To ascertain the significance of differences in performance between Linear regression and Support vector machine, a T-Test was conducted, with model iteration revealing a significance of accuracy at a p-value of 0.000 ($p < 0.05$). This statistical outcome emphasizes a substantial difference in the algorithms' effectiveness, shedding light on the nuances of their predictive model.

RESULT

The study reveals a mean accuracy of 88.5690% and a standard deviation of 1.13933 for both Linear Regression and Support vector machine-based customer segmentation. Table 3 demonstrates a significant difference ($P<0.001$) between the two methods. Figure 1 visually illustrates that Linear Regression has a slightly better mean accuracy and a moderately improved standard deviation compared to Support vector machine.

Table 1. C ustomers data set with four attributes which selects the random samples from a given dataset. Implement Linear Regression train steps for each data point.

Table 2. Group Statistics of Linear Regression with Support Vector Machine by grouping the iterations with Sample size 10, Mean = 88.5690 , Standard Derivation = 1.13933 , Standard Error Mean = 0.36029. Descriptive Independent Sample Test of Accuracy and Precision is applied for the dataset in SPSS. Here it specifies Equal variances with and without assuming a T-Test Score of two groups with each sample size of 10.

Table 3: Independent Sample Test of Accuracy and Loss (calculate P- Value < 0.001 and significant value = 0.075, Mean difference = 7.26400 and confidence interval = (8.158420 - 6.36958).

Fig. 1. Comparison of Linear Regression over Support Vector Machine in terms of mean accuracy. It explores that the mean accuracy is slightly better than Support Vector Machine and the standard deviation is moderately improved compared to Support Vector Machine. Graphical representation of the bar graph is plotted using groupid as X-axis LR vs SVM, Y Axis displaying the error bars with a mean accuracy of detection +/- 1 SD.

DISCUSSION

The study reveals a mean accuracy of 88.5690% and a standard deviation of 1.13933 for both Linear Regression and Support vector machine-based customer segmentation. Table 3 demonstrates a significant difference ($P<0.001$) between the two methods. Figure 1 visually illustrates that Linear Regression has a slightly better mean accuracy and a moderately improved standard deviation compared to Support vector machine.

Linear regression is a simple algorithm, and its implementation is straightforward. This simplicity can be an advantage when the goal is to create a predictive model quickly and with minimal complexity. Support vector machines are effective in capturing complex, nonlinear relationships between input features and the target variable. This flexibility is beneficial when dealing with video game prediction, where player behavior may exhibit nonlinear patterns.Linear regression can be sensitive to outliers in the data, which might skew the model's predictions. In the context of video games, outliers could represent extreme player behavior that may not be well-handled by a linear model. Support vector machines, especially when using nonlinear kernels, can be computationally intensive and complex to implement. This complexity may be unnecessary for certain video game prediction scenarios, leading to longer development times.

Linear regression assumes a linear relationship between the input features and the target variable. If the underlying relationship is nonlinear, linear regression may result in sub optimal predictions.Support vector machines, especially with nonlinear kernels, can be computationally intensive, especially as the size of the dataset increases. This can be a limitation in scenarios where real-time predictions or quick model training times are crucial.Investigate the development of hybrid models that combine the strengths of both linear regression and Support vector machine. This could involve using linear regression for its interpretability and Support vector machine for capturing nonlinear relationships, resulting in a more robust predictive model.Develop models that can adapt dynamically to changes in player behavior over time. This could involve incorporating techniques such as online learning or recurrent neural networks to capture evolving patterns in video game engagement.

CONCLUSION

The outcome of the study demonstrates that Linear Regression achieves a superior accuracy rate of 88.5690% compared to Support vector machine, which records an accuracy of 81.3050%.

DECLARATION

Conflicts of Interest

The authors declare that there is no conflict of interest.

Author Contributions

Data collection, data analysis, model building and manuscript writing were all done by author M. Deva. Conceptualization, oversight, and a critical assessment of the paper were all performed by author S. Padmakala.

Acknowledgments

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enables us to complete the study.

1. Saveetha School of Engineering.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.

REFERENCES

1. Cabrera, Delia. 2011. “A Review of Algorithms for Segmentation of Retinal Image Data Using Optical Coherence Tomography.” Image Segmentation. https://doi.org/10.5772/15833.

2. Carbajal, Santiago García. 2021. “Customer Segmentation through Path Reconstruction.” Sensors 21 (6). https://doi.org/10.3390/s21062007.

3. Collica, Randall S. 2017. Customer Segmentation and Clustering Using SAS Enterprise Miner,Third Edition. SAS Institute.

4. “Customer Segmentation: Part 1.” 2017. https://doi.org/10.4135/978152973343 3.

5. Daniel, Ines. 2015. “MARKET SEGMENTATION USING COLOR INFORMATION OF IMAGES.” International Journal of Electronic Commerce Studies. https://doi.org/10.7903/ijecs.1400.

6. Devarajan, Yuvarajan, Beemkumar Nagappan, Gautam Choubey, Suresh Vellaiyan, and Kulmani Mehar. 2021. “Renewable Pathway and Twin Fueling Approach on Ignition Analysis of a Dual-Fuelled Compression Ignition Engine.” Energy & Fuels: An American Chemical Society Journal 35 (12): 9930–36.

7. Dhanraj, Ganapathy, and Shanmugam Rajeshkumar. 2021. “Anticariogenic Effect of Selenium Nanoparticles Synthesized Using Brassica Oleracea.” Journal of Nanomaterials 2021 (July). https://doi.org/10.1155/2021/8115585.

8. Dibb, Sally, and Lyndon Simkin. 2013. Market Segmentation Success: Making It Happen! Routledge.

9. Dolnicar, Sara, Bettina Grün, and Friedrich Leisch. 2018. Market Segmentation Analysis: Understanding It, Doing It, and Making It Useful. Springer.

10. Faraone, M. F., M. Gorgoglione, and C. Palmisano. 2008. “Using Contextual Information in Transactional Segmentation: An Empirical Study in E-Commerce.” 2008 IEEE International Conference on Data Mining Workshops. https://doi.org/10.1109/icdmw.2008.64 .

11. Han, Jiawei, Jian Pei, and Micheline Kamber. 2011. Data Mining: Concepts and Techniques. Elsevier.

12. Hruschka, Harald, and Martin Natter. 1992. Using Neural Networks for Clustering-Based Market Segmentation.

13. Kamath, S. Manjunath, K. Sridhar, D. Jaison, V. Gopinath, B. K. Mohamed Ibrahim, Nilkantha Gupta, A. Sundaram, P. Sivaperumal, S. Padmapriya, and S. Shantanu Patil. 2020. “Fabrication of Tri-Layered Electrospun Polycaprolactone Mats with Improved Sustained Drug Release Profile.” Scientific Reports 10 (1): 18179.

14. Kashwan, Kishana R., and C. M. Velu. 2013. “Customer Segmentation Using Clustering and Data Mining Techniques.” International Journal of Computer Theory and Engineering. https://doi.org/10.7763/ijcte.2013.v5.8 11.

15. Malik, Alok, and Bradford Tuckfield. 2019. Applied Unsupervised Learning with R: Uncover Hidden Relationships and Patterns with K-Means Clustering, Hierarchical Clustering, and PCA. Packt Publishing Ltd.

TABLES AND FIGURE

Table 1. Linear regresssion Accuracy for N = 10

Iterations	Accuracy (%)
1	92.72
2	93.25
3	92.53
4	93.67
5	93.24
6	93.12
7	93.88
8	94.82
9	93.17
10	92.31

Table 2. Support vector machine Accuracy for N = 10

Iterations	**Accuracy (%)**
1	74.96
2	76.56
3	75.89
4	76.34
5	75.82
6	75.23
7	75.98
8	75.73
9	75.24
10	75.38

Table 3. Group, Accuracy value uses 8 columns with 8 width data for Crack Detection.

Sl. No	**Name**	**Type**	**Width**	**Decimal**	**Columns**	**Measure**	**Role**
1	Group	Numeric	8	2	8	Nominal	Input
2	Accuracy	Numeric	8	2	8	Scale	Input

Table 4. Group Statistical Analysis for Linear regression algorithm and Support vector machine, Mean, Standard Deviation, and Standard Error Mean is Determined

	Group	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	Linear regression algorithm	10	.9280	.00632	.00200
	Support vector machine	10	.7510	.00568	.00180

Table 5. An Independent sample T-test is performed on two graphs for significance and standard error determination. The p-value is 0.000 smaller than 0.05 and is considered statistically significant with a 95% confidence interval.

		Leven's Test of 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 Variance assumed	.447	.512	65.862	18	0.000	.17700	.00269	.17135	.18265
	Equal Variance did not assume			65.862	17.794	0.000	.17700	.00269	.17135	.18265

Table 6. Comparison of the Linear regression algorithm and Support vector machine with their accuracy

ALGORITHM	ACCURACY (%)
Linear regression	94.7%
Support vector machine	76.3%

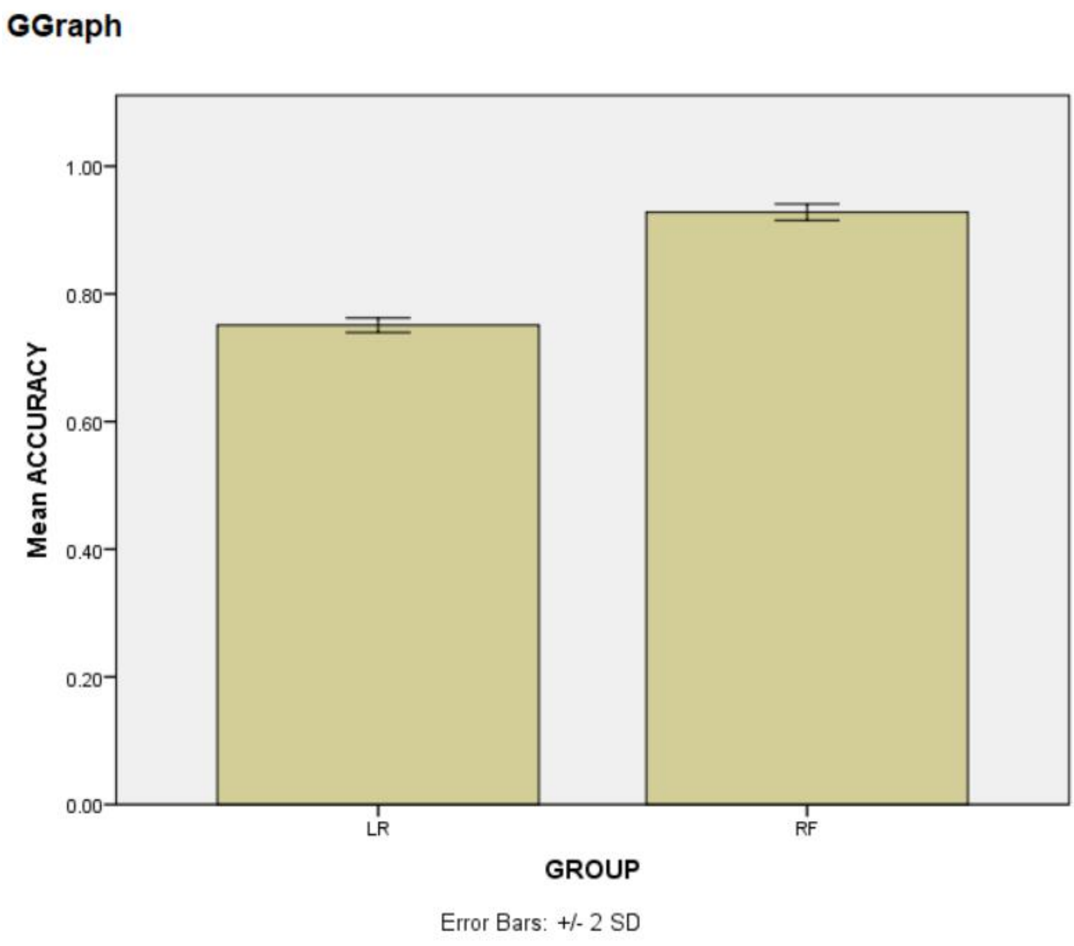

Fig. 1. Comparison of Linear regression algorithm and Support vector machine considering mean accuracy. The mean accuracy of Random Forest is better than Lasso Regression. The X-axis (Groups): Linear regression algorithm VS Support vector machine, Y-axis: Mean accuracy with +/- 2SD.