RESEARCH ARTICLE

Check for updates

RSP Science Hub

Analysing the Impact of video game on Consumer Engagement and Brand Loyalty: A Comparative Study of Traditional Marketing and Machine Learning -based Strategies

A Manimuthu[1], U Udhayakumar[2], A Cathrine Loura[3], Peterjose P [4], Antony S Alexander[5]

[1]Assistant Professor, Department of Data Science, Loyola College, Chennai, India

[2]Associate Professor, Department of Computer Science, Shanmuga Industries Arts and Science College, Tiruvannamalai, India

[3]Assistant Professor Department of Compute Applications, Loyola College, Vettavalam, Tiruvannamalai, India

[4]Assistant Professor, Department of Computer Science, Mount Carmel College, Bengaluru, India

[5]Assistant Professor, Department of Computer Science, Loyola College, Chennai, India

Email: msgtomanimuthu@gmail.com

Abstract

In contemporary times, the video game industry has experienced remarkable growth, captivating a wide audience with its immersive offerings. Undoubtedly, it stands as a significant global contributor to revenue generation. This sector wields a considerable influence, drawing in individuals with sharp and innovative skills to foster the expansion of video games worldwide. Exploring the substantial profit generated this sector, machine learning technologies have become instrumental in creating highly effective models that can analyse and forecast computer game sales well in advance. The realm of machine learning offers a diverse array of models for predicting future sales, employing techniques such as Linear and Multiple Regression, Random Forest, Decision Trees, Support Vector Machines, among others. Each of these approaches processes data using various mathematical concepts and formulas to estimate sales. The selection of an appropriate model depends on a thorough comparison of their accuracy and performance, considering the nature of the data. Model accuracy is commonly assessed by measuring the total number of correct predictions relative to all predictions made. As a key performance metric for evaluating the efficacy of the models, the R-square statistic is widely employed. Four algorithms have been tested on a selected dataset, and their performance has been compared to identify the most effective model for the given data

1. Introduction

The Modern world the online games are play major role in human life in that video games is imported Though the Video games was introduced way back in 50s, it has gained more limelight in the last few years. It provides instant entertainment and enjoyment to millions of people out there. Reports states that the gaming industry has now surpassed movie and music industry combined to become the largest engaging industry. Apparently, this industry is a very complex field involving tedious tasks to be accomplished. With great efforts, each and every step is completed and proceeded further. But the best part is about the profit generation which ben-

Analysing the Impact of video game on Consumer Engagement and Brand Loyalty: A Comparative Study of Traditional Marketing and Machine Learning -based Strategies

2023, Vol. 05, Issue 11 November

efits tons of people who has developed these games putting in enormous efforts. It's no surprise that this business produces profits in billion dollars. Video sport industry wishes correct income in an exponential market increase. Over the past decade, the sales generated from computer and video games in the United States have seen a significant and impressive increase. So we must are expecting of numerous video game fans via using historical sales information. This take a look at entails extracting the online game income facts and analyzing which recreation has extra sales globally whilst in comparison to other nations. Using machine learning techniques, we predicted the market sales of video games. This approach has proven beneficial for numerous industries seeking to forecast sales data (Ansari, Talreja, and Desai).While this being the case, it is a great innovative knowledge to combine the concepts of Machine learning and Video game industry to end up with wonderful and magnificent results.

1.1. Machine Learning:

As the name suggests, it is the capability of a machine to learn by itself without the assistance of an external program coded by a technical person. This technology has benefited a heavy number of fields with its countless added advantages and its potential to deal with big volume of data. Eventually, Machine learning's ability is to shape the dataset and lastly mould the model to its best so that the ultimate performance and accuracy is extracted out of it. On that account, a merge of these concept will lead to beneficial insights of the data. The ultimate goal is to predict the future results by analyzing the patterns and trends of previous data by applying statistical concept, mathematical intuition and the chosen appropriate model for the selected dataset.Machine Learning algorithms used for talking about supervised Learning, it contains labelled input features that are mapped to labelled target features(outputs). The machine learns from the labelled data by itself. Supervised data is further sorted into Regression and Classification. Regression is extensively used for this chosen dataset to solve and predict the global sales (Outcome or target variable). With the help of statistics and by plotting graph, it is also utilized to understand and analyze the trend line. These graphs help to acknowledge more about the dataset with diagram observation. Graphs such

as pie chart, bar graphs, line charts and more gives out information which is an easy and quick task to learn about the figures. It is a great solution when the data is large in size and is much complicated. In Unsupervised Learning, the machine learns from unlabeled data and the target feature is not available. The model is trained with the help of data that is collected in hand and produces its own output by grouping the Obtainable facts. From a business operation perspective, the sales prediction is very much important as it notifies if the company is incurring loss or making a fast buck. More the sales would definitely indicate higher the rate of gaining people's attention. Sales prediction also helps in understanding in which genre people are more interested so that more effort could be put in that particular criteria

2. Literature Study

(Michaltrneny)MichalTrněný ,the author proposed, "There is a strong evidence of the Gaming industry growing worldwide but no detailed studies or proof on the topic of predicting the success on this market". The data has no information about the age of the people whose attention has been caught extensively. The author has applied six algorithms such as Baseline, Linear model, RPART,Random Forest,Gaussian Process and SVM to calculate the output separately. He has predicted that this particular dataset has covered 33% of the data. (Yufa et al.)Alice Yufa, Jonathan L. Yu, Henry Chan, Paul D. Berger, the authors considered critic and user reviews would be very significant in generating higher sales. The authors had also mentioned some of the limitations which stated that some of the data were missing and some of the features such as Gender, not being a part of the dataset as it would give us an idea as to how many male and female are interested in different types of genre games. The authors have examined the dataset by performing anova and stepwise regression analysis. (Aziz et al.)Amar Aziz1 ,Shuhaida Ismail2 , Muhammad Fakri Othman1 , Aida Mustapha1, the authors proposed, " The primary objective of this study is to identify the key factors that contribute to the success of video games as blockbusters. They have implemented four models namely Using RapidMiner Studio as the tool for importing the dataset, we compared the results across accuracy for Naïve Bayes, K-Nearest Neighbor, Random For-

est, and Decision Tree algorithms, recall and precision scores. Lastly, it is seen that Naïve bayes has been concluded that is more suitable for that particular dataset. (Chavarria)Ignacio Chavarria, the author has carried out heatmaps to mainly focus on the co-relation between each and every features. The author has identified the games with best sales globally along with the developers who had gained a lot of profit through their work. He has compared each and every feature to other feature to comment on the strong co-relation amongst and has concluded that "year of release" feature has a strong relation towards the"platform" feature. He has performed Linear Regression, Random forest models and has determined their accuracies. Vishal Shrivastava, the author proposed, clustering sales methods ("A study of various clustering algorithms on retail sales data"), this paper discusses the 4 principal clustering algorithm ok-approach, density based totally, filtered, farthest first clustering set of rules and evaluating the performances of those principles clustering set of rules at the element of Correctly class smart cluster constructing capacity of set of rules. "Prediction of Sales Value in Online purchasing the usage of Linear Regression", (Gopalakrishnan, Riteshchoudhary, and Prasad) this paper proposed is to analyse the sales of a huge store, projecting future revenues to enhance profitability and strengthen the brand's competitive edge. Achieving customer satisfaction aligns with market trends. The authors, Paul Bertens and Anna Guitart, utilized the Linear Regression Algorithm, a popular technique in Machine Learning, for sales prediction. In a separate study, they proposed "Games and Big Data: A Scalable Multi-Dimensional Churn Prediction Model" (Bertens, Guitart, and Perianez) The article introduces a strategy for forecasting game churn using survival ensembles, allowing precise predictions on both the time when players exit the game and their total accumulated playtime. This model is ideal for conducting real-time churn analyses, even for games with millions of daily active users.

3. Methodology:

This research applies various feasible prediction modelling algorithms to see which affords the fine effects. The analysis incorporated the application of Linear Regression, Decision Tree, Random Forest, and Support Vector Regression algorithms to examine video game sales data (Alfons).While building a model, there are certain methods and procedures to be followed so that the model is built in an accurate way. The fact must be clear that Machine learning idea is not only about model creation, there are plenty of steps that must be followed beforehand. Certain steps are to be done with utmost attention, which also requires quite a lot of time when the data dealt with is in large number.

3.1. Dataset Description:

The dataset is collected from the website names The selected dataset from Kaggle comprises 11 features and 16,600 rows. The aim is to predict the global sales of each and every game with the help of the Regression concept. The chosen dataset contains the following features. The digital groups have detected the use of the pinnacle spreaders identified with the assistance of rating algorithms. Around 5 percent of the whole links are expected inside the experiments with the decision log dataset.

- Rank - Ranking of overall sales
- Name - The games name
- Platform - Platform of the games release (i.e. PC,PS4, etc.)
- Year - Year in which the game released
- Genre - Genre of the game
- Publisher - Publisher of the game
- NA_Sales - Sales in North America (in millions)
- EU_Sales - Sales in Europe (in millions)
- JP_Sales - Sales in Japan (in millions)
- Other_Sales - Sales in the rest of the world (in millions)
- Global_Sales - Total worldwide sales (Target Variable)

3.2. Data Preparation and analysis:

The data collected initially is likely unsuitable and requires multiple preparatory steps before being input into the model. The subsequent steps form part of the model-building process

3.3. Statistical overview of the data:

Check the number of rows and columns of the data .Look over how the features are distributed across through the medium of graphs and charts.Study about the outliers and select a specific way to handle it. Find out the statistical measures for the numerical features available. Observe the co-relation between

Analysing the Impact of video game on Consumer Engagement and Brand Loyalty: A Comparative Study of Traditional Marketing and Machine Learning -based Strategies

2023, Vol. 05, Issue 11 November

the features and note the features that has higher co-relation towards the target variable.

3.4. Data Pre-processing:

Handling null values and replacing them with any of the appropriate chosen method.

4. Feature Selection:

This step is performed to avoid "Curse of dimensionality". It states that less the number of features, more the accuracy and performance of the model. The ultimate aim always is to try out and perform various techniques with the perspective of converting the complex data into a simpler version.

4.1. Encoding Categorical Values :

Encoding is a necessary step that is carried out in any machine learning model as it is converts the categorical features in to numerical features which in turn helps the machine to understand further about the data. There are two methods namely Label Encoding and One Hot Encoding frequently used.

4.2. Chi-Square Test:

It is a statistical technique involving hypothesis testing that is applied to extensively find out the co-relation between categorical features. An important point to be noted is that, not all features is to be included for model training rather highly co-related features are to be only considered.

It is shown that people are more interest in Action and Sports genre. And it's no surprise as sports has always never failed to gain its popularity and attention from the audience. The third position is gained by shooter games which has also had quite an impact among the gamers. And the last place is gathered by the puzzle genre as it is seen only very few people comparatively is interest in that genre.

5. Multiple Regression:

This is one of the types of multivariate statistical analysis which is applied in a scenario where in the output variable is determined by multiple independent variables. The aim in multiple linear regression is to compute the intercept and slopes of each and every independent features to determine how much each feature influences the dependent variable.

$$y = b_1 x_1 + b_2 x_2 + \ldots + b_n x_n + c \text{————————-1}$$

Where, y is the dependent variable whose value is to be found out, b_1 is the intercept, b_2 is the slope and X is the independent variable

6. Support Vector Regression

SVR model is one such models that could be used in both regression and classification use cases. It builds a hyperplane in such a way that it passes to both the nearest positive and negative points. The entire distance from the nearest positive and the negative point across the hyperplane is called as margin and our main aim is to maximize the margin. SVM kernels are used to solve non linearly separable points that convert a 2D or a low dimension to a higher dimension. The Math equation used in Support Vector Machine is as follows, Support Vector Regression (SVR) employs the SVM classification algorithm to predict continuous variables. Various regression models are utilized to minimize the disparity between the predicted value and the actual value (Mahdevari et al.). SVR endeavors to fit a quality line among predefined error rates, encompassing key terms such as kernel, hyperplane, boundary line, and support vector. It is particularly effective for linear data points.

$$f(x) = \sum n = 1 \, N \, (\alpha n - \alpha n^*)(x n' x) + b$$
–2

Support Vector Regression for Non-Linear points :

$$\beta = \sum n = 1 \, N \, (\alpha n - \alpha n^*) \, x \, n \text{————————}$$
—3

7. Decision Tree Regressor:

This technique tries to divide the data into many diversions until it reaches the leaf nodes. It follows the basic rule of If-else and then takes a decision accordingly. After each and every split, the entropy is measured (level of impurity) and it slowly gets reduced at each and every step. Information gain is also measured at each level which tends to gets increased at every split. The Split also happens based on a concept names gini impurity which is found out to be better than entropy-based splitting.
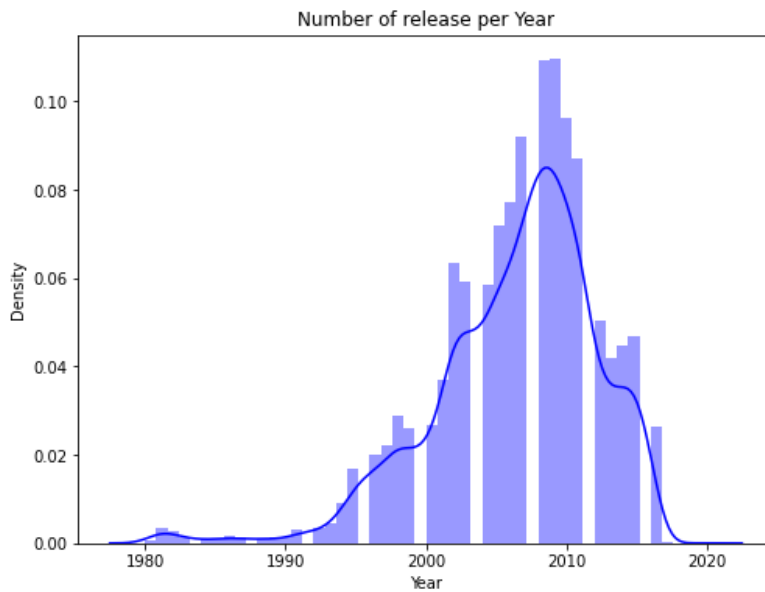
7.1. Decision Tree

A decision tree is a type of supervised machine learning model that defines "the relationship between the input and the corresponding output based on our data (Ho)." Its primary objective is to predict the value of a target variable.
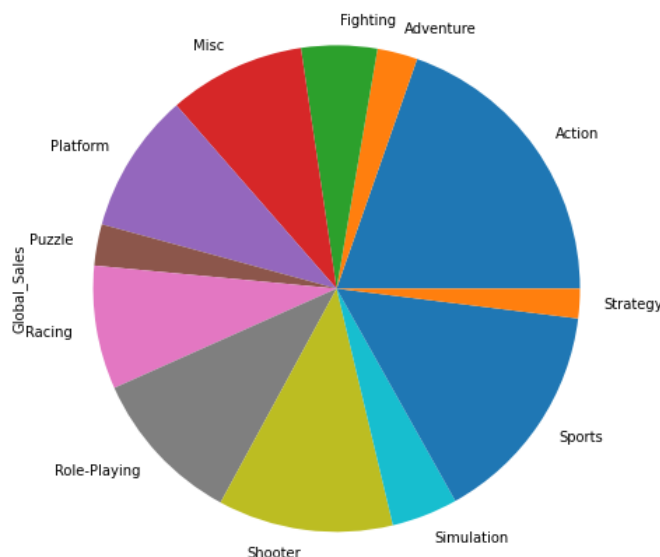
8. Random Forest:

Random forest Regressor is an ensemble model that uses the bagging technique. Several decision trees

**FIGURE 1.** **Graph depicts the entire sale records**

FIGURE 2. **The percentage of growth of a particular genre in Gaming industry.**

together form a random forest. Row sampling and feature sampling are done for each and every decision tree. In regression, every decision tree is trained and tested on a sample of randomly selected data which in turn produces different continuous values and the mean of those values is calculated to get the final outcome. Through this process, we would get a model that has low variance and low bias leading to high performance, and good accuracy by the model (Boinee, Angelis, and Foresti).

9. Prediction and Results:

Considering the performances of above four mentioned algorithms on the chosen dataset, it is seen that Multiple Regression shows maximum accuracy amongst followed by Random forest. Multiple Regression indicates the association between the independent variables towards the dependent feature. Any change in the value of any independent feature impacts the value of the dependent variable. Regression is a statistical and analytical method that is commonly used in a lot of fields. Meanwhile, Random Forest also done an average esti-
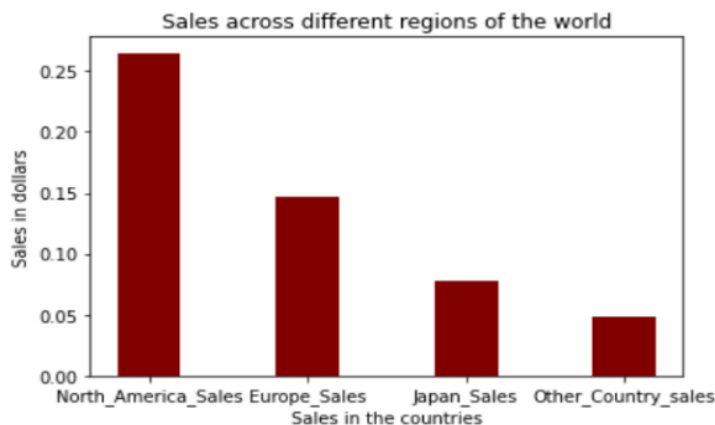
FIGURE 3. Sales collected throughout the regions such as North America, Europe, Japan and other regions.
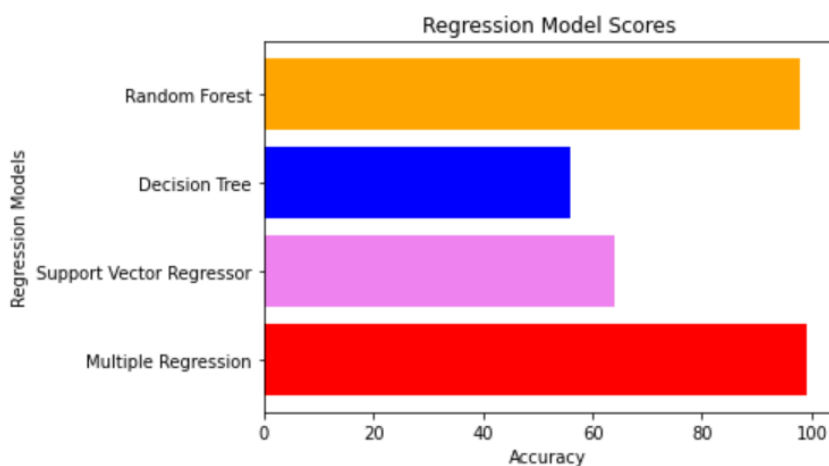
FIGURE 4. Regression Model Scores

mation and is most likely to perform even better in real world scenarios where the dataset is very large in volume. An added advantage of Random forest involves bootstrap sampling which involves sampling of the independent features with replacement. It is also extensively used for dimensionality reduction method.

10. Conclusion and Future work

Businesses that are involved in sales across the entire global has always paved with a lot of investment settled by the publishers. In that event, it is really a major task to calculate the sales generated. This is a crucial task that requires utmost seriousness. Fortunately, the gaming industry has been benefited vastly with latest technology that could pour a lot of predictions of the future events with the help of the history data that is provided. Machine Learning along with Artificial Intelligence has added benefits

to deal with the most complex data and convert them into simpler version of data. In future we can the data from real time source like Amazon and other online selling platform and predict the modes using various algorithms

References

"A study of various clustering algorithms on retail sales data". *Vishal shrivastava* 1 (2012).

Ansari, Nazneen, Maahi Talreja, and Vaishali Desai. "Data Mining in Online Social Games". *Advances in Intelligent Systems and Computing*. Springer India, 2013. 801–805.

Aziz, Amar, et al. "Faculty of Computer Science and Information Technology". *Empirical Analysis on Sales of Video Games: A Data Mining Approach*.

Bertens, Paul, Anna Guitart, and Africa Perianez. "Games and big data: A scalable multidimensional churn prediction model". *2017 IEEE Conference on Computational Intelligence and Games (CIG)* (2017): 978–979.

Boinee, P, A D Angelis, and G Foresti. "Meta random forests". *International Journal of Computational Intelligence* 2.3 (2005): 138–147.

Ho, T. "The random subspace method for constructing decision forests". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20.8 (1998): 832–844.

Mahdevari, Satar, et al. "A support vector regression model for predicting tunnel boring machine penetration rates". *International Journal of Rock Mechanics and Mining Sciences* 72 (2014): 214–229.

Michaltrneny. *Master's thesis, "Machine learning for Predicting success of Video games",FacultasArtisinfomatice.* Universitasmasarykiana, Brno, Spring, 2017.

Yufa, Alice, et al. *Predicting Global Video game Sales*. Bentley University Waltham, MA 02452 USA.