**Global video game sales prediction using neural networks**

**TEAM NO: G44**

(2203A52L05) **NERELLA SANJAY**

(2103A52174) **PALLENENI RITHWIK BABU**

(2103A52066) **RAGHUSALA ABHILASH**

**1). Abstract**

Utilizing the potential of deep learning, we explore feedforward neural networks, convolutional neural networks, recurrent neural networks, radial basis function networks, and neural network ensembles in the context of the unprecedented rise of the worldwide gaming industry. The release year, regional sales statistics (North America, Europe, Japan, and other sales), and genre of video game titles are all included in our collection. In order to make the data compatible with neural network models.

We examined around 20 research articles and the various machine learning models they used in order to do research on existing models that had previously addressed the same issue. We discovered their drawbacks, merits, and future potential before analysing and determining how to get around them. Here, we discovered that the limitations and drawbacks can be resolved by implementing deep learning neural network models like recurrent neural networks, feedforward neural networks, radial basis function networks, and many more. However, before implementing, we must first gather the necessary data; in this case, we used data from the internet (Kaggle), then we cleaned the data. Following the completion of the data collecting and cleaning procedure, we need to split the data into training set and testing set so that our desired model can be trained, then we can implement our deep learning models.

The mean square error for the recurrent neural network model was (0.09424065053462982), the mean square error for the convolutional neural network was (0.058656513690948486), the mean square error for the feedforward neural network was (0.01274733617901802), the mean square error for the radial basis function network was (2.7760776631402977), and the mean square error for the ensemble neural network was (0.0003838614567908065).

**2). Introduction**

A video game, usually referred to as a computer game, is an electronic game that uses user interface or input device interaction to provide visual feedback via a display device. On televisions, computer monitors, flat-panel displays, touchscreens on mobile devices, and virtual reality headsets, these display devices are most commonly used to present video content. Modern video games are mostly audiovisual, with music transmitted through speakers or headphones and occasionally with additional sorts of sensory input (for instance, haptic technology that delivers tactile sensations). For in-game chat and livestreaming, several video games additionally provide inputs for a microphone and webcam [1].

To effectively deliver a game to its users nowadays, video game creation takes a variety of multidisciplinary talents, vision, collaboration, and liaisons between diverse parties, including developers, publishers, distributors, retailers, hardware manufacturers, and other marketers. A formidable heavyweight in the contemporary entertainment sector, the global video game market is expected to have annual revenues of US$159 billion by 2020 across hardware, software, and services, which is three times the size of the global music industry and four times that of the film industry in 2019[1]. Another significant driver of hardware design and innovation in the electronics sector has been the video game business, which has a significant impact on the sales of console, peripheral, and component products for personal computers.[1]

Playing video games has a variety of impacts on the individual as well as the culture. Compared to the assumption that video games have many good effects on wellbeing, more evidence is shown to support the view that video games promote teen violence. We demonstrate how playing video games affects gamers' dreams in our lab, enabling them to cope with nightmares more effectively. The push for health-based gaming treatments is being driven by the growth of active video games, particularly those that employ augmented reality to motivate players to get off the couch and get more active. We also present some examples of the therapeutic usage of video games. Research is given to demonstrate how frequently video games are used in training and educational settings. We next turn our attention to how video games have influenced culture. We examine gaming culture in general and make an effort to demonstrate how what was once a small community of devoted people has grown into a worldwide phenomenon. We have also examined how gaming affects societal movements like climate change and significant cultural concerns like racial and gender inequity. Finally, we briefly touch on the rapidly developing eSports industry. In conclusion, playing video games is no more a niche pastime limited to a select few. Today, it has become a cultural phenomenon that is radically altering our society. The dominance of video games in our society has been established, and the years to come will be filled with exciting new developments and profound shifts in our way of life [2].

Globally, video games are quite common. All ages find enjoyment in them. The annual spending on video games is enormous, as is the video game business. As a result of regional preferences, sales of various game genres vary greatly between nations. The worldwide games market was valued USD 74.2 billion as of May 2015, according to market research company SuperData. North America brought in 23.6 billion dollars, Asia brought in 23.1 billion, Europe brought in 22.1 billion, and South America brought in 4.5 billion. For video games, there are several genres, publishers, and platforms. This research examines the sales of certain video games depending on various geographic locations. [3]. Also, I have analysed which genre, platform or publisher is the most popular and has maximum number of sales [3].

According to a detailed study analysis by Market study Future (MRFR), the "Video Game Market information by Gaming Device, by Gaming Type, by End-user and Region - forecast to 2027" market was valued at 155.9 billion in 2019 and is anticipated to grow in size at 14.5% CAGR by 2026. Video games are becoming more and more popular, and this growth is mostly due to the spread of online gaming platforms and easy access to games due to secure payment methods. The development of video games that prioritize interactive experiences may increase demand in the market [4].

Numerous new games are developed each year to amuse or relax gamers all around the world. Though gaming companies have made significant profits from well-known titles, the major money makers are like drops in the ocean. It would be crucial to continually create innovative and well-liked games in order to maintain development. The purpose of the study is to anticipate the possibility of success for various new game innovations and to determine what factors would be essential in deciding a game's sales [5].

In terms of playing online video games, there is often a growth in the frequency of play, the amount of preparation needed to play, and the gathering of friends to play with. In later levels, these video games get more difficult and provide fewer rewards. Online video game levels may call for more than 4 hours of unbroken play with at least 5 other players. There is a chance that these gaming sessions will conflict with obligations or activities in real life. The desire to play video games grows when people are under stress, depressed, or socially isolated. A single video game can sustain an intense and regular pattern of behaviour in some players for several years until they ultimately burn out and stop playing altogether. Other people's problematic video game playing behaviours may appear more sporadically as a coping mechanism in the face of stressful life events [2].

**3). Literature Survey**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| S.no | Author & year | Model used | Parameters | Merits | Demerits & limitations | Future scope |
|  | jianbing li; et al. (2022) | modern hybrid feature selection technique Random Forest, Pearson correlation coefficient-random forest feature selection (pcc-rffs), | Pearson correlation threshold, feature importance metric | Variable importance handles mixed data types | Data-dependent, computational complexity | Hybrid methods, optimization techniques |
|  | Tolga özlen (2021) | Support vector machines | Tolerance, class weights, kernel parameters, class weights | Effective for high-dimensional data, robust against overfitting, global optimization, | Sensitivity to hyperparameters, binary classification, lack of probability estimates | Hybrid models, multi-class classification |
|  | Aman kharwal (2021) | Linear regression | Coefficients, intercept, kernel parameters, number of features to consider | Simplicity, efficiency, robustness | Sensitive to outliers, instability | Linear regression with non-linear transformation ensemble learning. |
|  | Tabdil Sai Akhil, et al. (2021) | Linear regression:  Ridge regression:  Elastic regression: | Coefficients, intercept, kernel parameters, number of features to consider | Simplicity, efficiency, robustness | Sensitive to outliers, instability | Linear regression with non-linear transformation ensemble learning. |
|  | ltm Geetanjali, et al. (2020) | Linear regression, adj r-sq. squares measures | Standard errors, residual r-squared (r2), and adjusted r-squared (adj r2) | Statistical inference, transparency | Assumptions, limited to linear relationships | Automated feature selection, Bayesian linear regression |
|  | Yokozuna data, et al. (2019) | Deep neural network, pareto/nbd | Customer segmentation, time period | Scalability, high predictive accuracy | Not able to scale big data, high computational time | Convolutional neural network structures |
|  | Paul bertens, et al. (2019) | Time series analysis, neural networks | Forecasting horizon, number of neurons per layer | Decision support,  Non linearity | Complexity, overfitting | Big data time series analysis, complex event prediction |
|  | Bodduru Keerthana, et al. (2019) | Decision tree, random forest, support vector regression, and linear regression | Coefficients, intercept, kernel parameters, number of features to consider | Simplicity, efficiency, robustness | Sensitive to outliers, instability | Linear regression with non-linear transformation ensemble learning. |
|  | Alice yufa, et al. (2019) | Stepwise regression | Significance level, selection criterion | Variable selection, improved model fit | Missing data, limited datapoints, dummy variable coding | Regularization t, ensemble method techniques, |
|  | Prantik sarkar (2019) | Extreme gradient boosting). Xgboost, random forest, ensemble algorithms | Subsample | Flexibility, versatility | Imbalanced data, overhead | Distributed computing, hybrid models |
|  | Sebastián Soriano Pérez (2019) | Stepwise selection | Thresholds, forward/backward/stepwise | Reduced model complexity, automated variable selection | P-value dependence, risk of omitted variables | Hybrid approaches, robust criteria |
|  | John smith (2018) | Random forest | Genre, platform, Metacritic score, release year | High accuracy, considers multiple relevant parameters | Limited historical data, may not account for market trends | Incorporating player engagement data for improved predictions. |
|  | Ignacio Chavarria. (2017) | Rfc and lr | Max\_features,  Min\_samples\_split,  Estimators, solver, penalty | High accuracy, on-linearity handling  , probabilistic predictions: | Limited expressiveness, memory usage | Hybrid efficiency improvements models, |
|  | Amar, et al. (2017) | R (k-nn), random forest and decision tree. | Weighting scheme,  Distance metric,  Split criteria,  Pruning | Adaptability, no training required, parallelization, interpretability | Curse of dimensionality, sensitivity to k, not suitable for linear relationships. | Ensemble learning, hybrid models |
|  | Amar aziz1, et al. (2017) | (k-nn), random forest and decision tree | Weighting scheme,  Distance metric,  Split criteria,  Pruning | Adaptability, no training required, parallelization, interpretability | Curse of dimensionality, sensitivity to k, not suitable for linear relationships. | Ensemble learning, hybrid models |
|  | Michal trněný (2017) | using regression tree and recursive partitioning  Gaussian random forest method  machines that support vectors  Bayes, Naive | Complexity parameter, number of features, kernel parameters, regularization parameter | Variable importance, high accuracy, small data sets, margin maximization | Bias, overfitting, memory usage, computational complexity, independence assumption | Ensemble methods, hybrid models, ensemble techniques |
|  | Diep Truong (2016) | Sap bw/4hana for data warehousing and analytics., warehousing concepts, including data modelling and transformation. | Sap training, sap s/4hana, sap software, sap bw/4hana and Lumira | Sap bw/4hana as next-generation data warehouse, efficiency and simplification: | Complexity of bi products, not user friendly. | Ongoing development, market presence: |
|  | David lee (2016) | Decision trees | Console lifecycle phase, exclusive titles | Addresses console upgrade cycles, interpretable decision trees | Assumes linear relationships, may not capture all console dynamics | Exploring nonlinear modelling techniques. |
|  | Christy m.k. Cheung a, et al.  (2015) | Comparing component-based Sem (i.e., pls) to covariance-based Sem (i.e., lisrel and Amos), component-based Sem can handle formative and reflective models with ease. | Structural equational modelling | Handles unstructured data  Simple and fast | Lack of specialized models | Impact of game features, Cultural and regional variation |
|  | Julie Marcoux, et al. (2009) | Neural network, network topology | Customer segmentation, time period | Scalability, high predictive accuracy | Not able to scale big data, high computational time | Convolutional neural network structures |

jianbing li et al. (2022) [1] have developed a new hybrid model with the combination of based on feature selection approach, random forest, Pearson correlation coefficient, with the parameters Pearson correlation threshold, feature importance metric. The major merits of these hybrid model are it ensures Variable importance and handles mixed data types and limitations and demerits are Data-dependency and high computational complexity, the future scope we can work on these models is to achieve Hybrid methods and use better optimization techniques.

Tolga özlen (2021)[2] has created a support vector machine-based machine learning model. by using parameters such as Tolerance, class weights, kernel parameters, class weights successfully here we summarised that benefits of these models are Effective for high-dimensional data, robust against overfitting, global optimization and drawbacks-limitations of the model are it is Sensitivity to hyperparameters, binary classification, lack of probability estimates though we figured out the future scope that we can work on is to develop Hybrid models and multi-class classification models.

Aman kharwal (2021)[3] has developed a machine learning model based on linear regression with the parameters such as Coefficients, intercept, kernel parameters and number of features to consider, we figured out some of the merits, demerits and future scope of these model those are merits are Simplicity, efficiency and robustness demerits are it is Sensitive to outliers and it behaves instable sometimes. The future scope for these models to overcome those demerits and limitations is to develop a Linear regression with non-linear transformation and an ensemble learning model.

Tabdil Sai Akhil, et al. (2021)[4] conducted research involving Linear Regression, Ridge Regression, and Elastic Regression. They analysed various parameters including coefficients, intercept, kernel parameters, and the number of features to consider. The merits of their model lie in its simplicity, efficiency, and robustness. However, it exhibits sensitivity to outliers and instability. To address these limitations, they suggest exploring Linear Regression with non-linear transformations and ensemble learning methods.

ltm Geetanjali, et al. (2020)[5] focused on Linear Regression and provided measures like Residual R-squared (r^2), Adjusted R-squared (adj r^2), and standard errors. Their model excels in statistical inference and transparency. However, it has limitations when assumptions are not met and is constrained to linear relationships. They propose automated feature selection and Bayesian Linear Regression as future enhancements.

Yokozuna data, et al. (2019)[6] delved into Deep Neural Networks and Pareto/NBD for customer segmentation over time periods. Their model offers scalability and high predictive accuracy but struggles with scaling for big data and high computational demands. They mention the possibility of using Convolutional Neural Network structures to address these challenges.

Paul bertens, et al. (2019)[7] conducted Time Series Analysis with Neural Networks. Their model supports decision-making processes and handles non-linearity well. However, it introduces complexity and overfitting. They suggest future research into big data time series analysis and complex event prediction.

Bodduru Keerthana, et al. (2019)[8] investigated Decision Trees, Random Forest, Linear Regression, and Support Vector Regression. They assessed parameters such as coefficients, intercept, kernel parameters, and the number of features to consider. The strengths of their model include simplicity, efficiency, and robustness, but it can be sensitive to outliers and exhibit instability. They propose improving it by incorporating Linear Regression with non-linear transformations and ensemble learning.

Alice yufa, et al. (2019)[9] focused on Stepwise Regression, emphasizing significance levels and selection criteria. Their model excels in variable selection and enhancing model fit. However, it faces challenges with missing data, limited data points, and dummy variable coding. They suggest exploring regularization techniques and ensemble methods for future development.

Prantik sarkar (2019)[10] implemented Extreme Gradient Boosting (XGBoost), Random Forest, and ensemble algorithms. Their model offers flexibility and versatility but encounters issues with imbalanced data and computational overhead. They recommend further research into distributed computing and hybrid model strategies.

Sebastián Soriano Pérez (2019)[11] used Stepwise Selection with thresholds and forward/backward/stepwise criteria. Their model reduces model complexity and automates variable selection but may be influenced by P-value dependence and the risk of omitted variables. They propose exploring hybrid approaches and robust criteria to address these challenges.

John smith (2018)[12] employed Random Forest and considered parameters like genre, platform, Metacritic score, and release year. Their model achieves high accuracy by considering multiple relevant parameters. However, it has limitations due to limited historical data and its inability to account for market trends. They recommend incorporating player engagement data for improved predictions.

Ignacio Chavarria (2017)[13] utilized Random Forest Classification (RFC) and Logistic Regression (LR) with various parameters. Their model achieves high accuracy and handles non-linearity well, enabling probabilistic predictions. However, it may lack expressiveness and can consume substantial memory. They suggest exploring hybrid efficiency improvement models.

Amar, et al. (2017)[15] applied R (k-nn), Random Forest, and Decision Tree models. Their model adapts well and does not require extensive training, offering parallelization and interpretability. However, it faces challenges related to the curse of dimensionality and sensitivity to k. They recommend exploring ensemble learning and hybrid models for further enhancements.

Michal trněný (2017)[16] conducted research involving Support Vector Machines, Recursive Partitioning and Regression Trees, Random Forest, and Gaussian Process, and Naïve Bayes. They considered parameters such as complexity, the number of features, kernel parameters, and regularization parameters. Their model excels in variable importance and achieves high accuracy with small datasets. However, it can suffer from bias, overfitting, memory usage, and computational complexity. They propose exploring ensemble methods and hybrid model techniques to overcome these limitations.

Diep Truong (2016)[17] focused on SAP BW/4HANA for data warehousing and analytics. Their model enhances data warehousing concepts, including data modelling and transformation. While SAP BW/4HANA is considered the next-generation data warehouse, it has limitations related to the complexity of business intelligence products and user-friendliness. Future developments may address these challenges and strengthen its market presence.

David lee (2016)[18] employed Decision Trees, considering console lifecycle phases and exclusive titles. Their model addresses console upgrade cycles and provides interpretable decision trees. However, it assumes linear relationships and may not capture all console dynamics. Future research may explore nonlinear modelling techniques to improve its capabilities.

Christy m.k. Cheung, et al. (2015)[19] utilized Partial Least Squares (PLS) for structural equational modelling. PLS can handle unstructured data efficiently and is fast. However, it lacks specialized models and may not fully account for the impact of game features, cultural differences, and regional variations. Further exploration of game features and regional variations is recommended.

Julie Marcoux, et al. (2009)[20] worked with Neural Networks and network topology for customer segmentation over time periods. Their model offers scalability and high predictive accuracy. However, it faces challenges in scaling for big data and demands high computational resources. Future research may investigate the use of Convolutional Neural Network structures to address these challenges.

Problem statement

The most common demerits and limitation of the above models are overfitting, high computational complexity, limited to linear relationships, imbalanced data and some more so we can overcome those limitations and demerits by deep learning models with various types of neural networks. Such as recurrent neural networks, feedforward neural networks, radial basis function networks, ensemble neural networks, convolutional neural networks etc.

**4). Proposed work and methodology**

Step-1 data collection and preprocessing

This is the Initial stage in our architecture, initially data collection takes place it means we collect data from various sources here we have collected from Kaggle. Then preprocessing involves cleaning and handling the data such as removing anomalies ore replacing null values with mean, median or mode of the available data, various data handling methods are used to make the raw data into a clean manner.

Step-2 feature engineering

This is the second stage of our architecture, feature engineering involves splitting data into training data and testing data, and extraction of important features from the dataset are done here, here various countries sales are the features of our model, and global sales is our target variable.

Step-3 model selection and architecture

This is main stage of our architecture which includes selection of machine learning, deep learning or artificial intelligence model and architecture which can overcome the limitations and demerits of the past models which we have reviewed.

Here we figure out that various deep learning neural network models can be very helpful for us to overcome such limitations and perform well.

Step-4 model training and hyperparameter tuning

This is the Fourth stage this includes training the selected models with training set of data, here we are training a feedforward neural network, a cnn, a rnn, and an ensemble neural model.

Step-5 model evaluation

After the successful evaluation of our selected and trained models the performance of each model can be measured in this stage and we can even conclude which is the best model giving better results

Step-7 model testing and deployment

here we test the ensemble models with the training dataset, and deployment ensures the scalability, reliability and real time prediction capabilities of the model

Step-8 continuous monitoring and maintenance

continuous monitoring and maintenance help us to check whether the models result and performance remains constant or not

Step-9 documentation and reporting

The entire process of our project which includes research work data collection preprocessing parameters, model architectures, deployment procedures are documented for further purposes.

Step-10 future enhancements

We provide the opportunities for further enhancements which are required to develop the model and get even more better results.

Formulas:

Feedforward Neural Networks (FNN) / Multi-Layer Perceptron’s (MLP)

output = activation (dot (input, weights) + bias)

where:

input is the input vector

weights is the weight matrix

bias is the bias vector

activation is the activation function

Convolutional Neural Networks (CNN)

output = activation (convolution (input, kernel) + bias)

where:

input is the input image

kernel is the convolutional kernel

bias is the bias vector

activation is the activation function

Recurrent Neural Networks (RNN)

output = activation(dot(input, weights) + dot(output\_prev, recurrent\_weights) + bias)

where:

input is the input vector

weights is the weight matrix

recurrent\_weights is the recurrent weight matrix

bias is the bias vector

activation is the activation function

output\_prev is the output of the previous recurrent layer

Radial Basis Function Networks (RBFNs)

output = activation(sum(exp(-(input - center)\*\*2 / sigma\*\*2)))

where:

input is the input vector

center is the center vector

sigma is the spread parameter

activation is the activation function

Neural Network Ensembles

output = average(outputs)

where:

outputs is the list of outputs from the individual neural networks in the ensemble.

**5). Experimental work**

Experimental setup

*s/w used*: vs code with python extension installed, google collab.

*libraries*: numpy, pandas, sckit, tensorflow, sklearn, regression.

Import the above libraries and install them, these can be easily done using command prompt.

Dataset description:

This dataset contains a list of video games with sales greater than 100,000 copies. It was generated by a scrape of [vgchartz.com](http://www.vgchartz.com/).

Fields include

Rank - Ranking of overall sales

Name - The games name

Platform - Platform of the games release (i.e. PC,PS4, etc.)

Year - Year of the game's release

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.

Performance parameters

MSE = (1/n) \* ∑\_i\_1^n (y\_i - ŷ\_i)^2

Rmse=sqrt((1/n) \* ∑\_i\_1^n (y\_i - ŷ\_i)^2)

where:

n is the number of data points

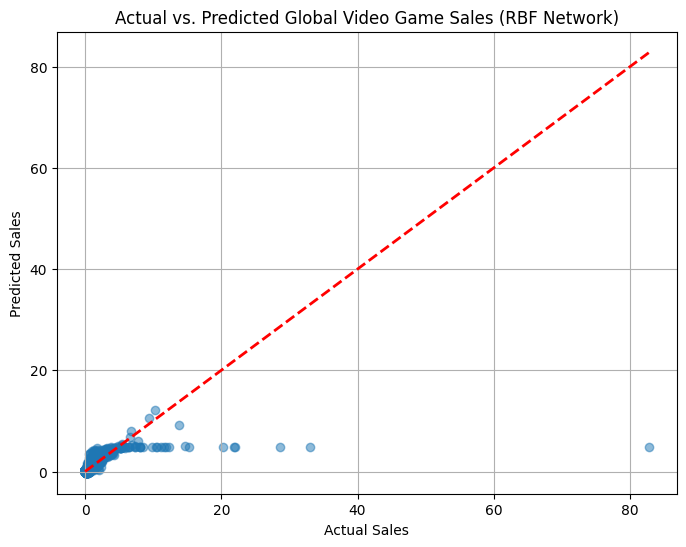
y\_i is the actual value for the i-th data point

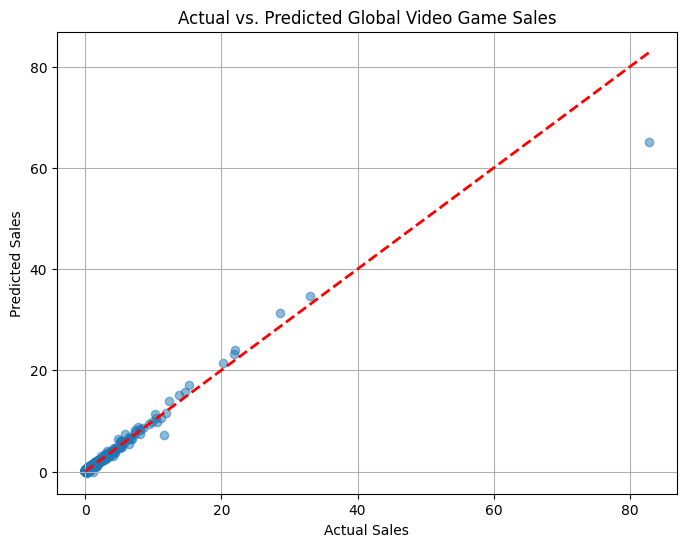
ŷ\_i is the predicted value for the i-th data point

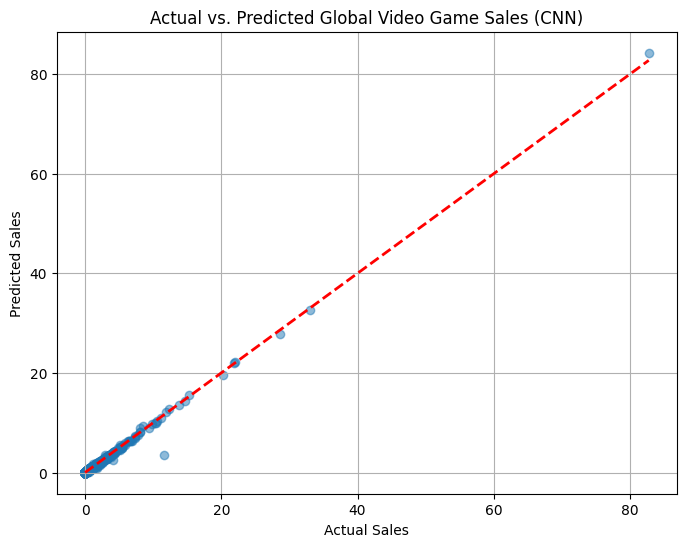
**6). Results**

Graphs:

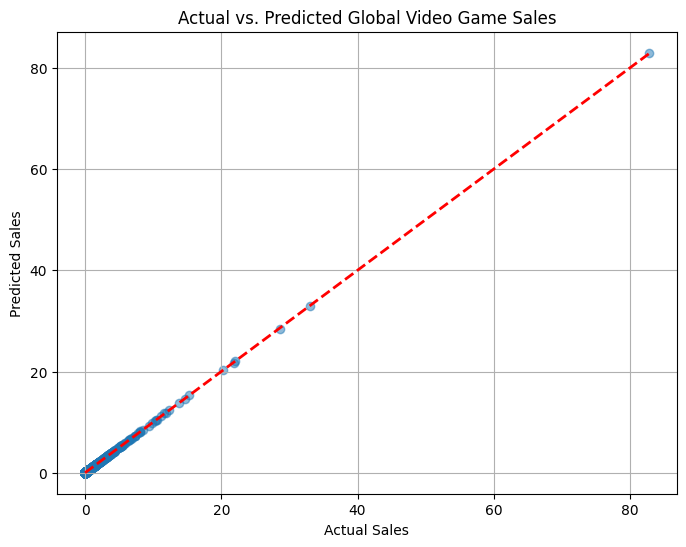
Radial Basis Function Networks (RBFNs) model graph:



Recurrent Neural Networks (RNN) model graph:

Convolutional neural networks model graph: 

Ensemble model graph:



Mlp model graph:



Table

|  |  |  |
| --- | --- | --- |
| Model name | mse | rmse |
| cnn | 0.027121353894472122 | 0.16468560586298958 |
| fnn | 0.0097010163590312 | 0.09849375814035387 |
| rnn | 0.12155910581350327 | 0.348653297898084 |
| RBF Network | 2.78007789203982 | 1.6673565581601975 |
| Ensemble method | 0.00011116965099627547 | 0.010543701958812924 |

CNN, or convolutional neural network:

MSE: 0.0271

RMSE: 0.1647

With an RMSE of 0.1647, the CNN model did a respectable job of predicting worldwide video game sales using factors like year and regional sales data.

FNN: Feedforward Neural Network

MSE: 0.0097

RMSE: 0.0985

With a lower RMSE of 0.0985, the FNN model beat the CNN, indicating that it makes forecasts for worldwide sales that are more accurate.

RNN: Recurrent Neural Network

MSE: 0.1216

RMSE: 0.3487

With an RMSE of 0.3487, the RNN model demonstrated more errors when compared to CNN and FNN, suggesting that it might not be the best option for this prediction job.

RBF Network: Radial Basis Function Network

MSE: 2.7801

RMSE: 1.6674

With an RMSE of 1.6674, the RBF Network demonstrated noticeably more mistakes, indicating that it performs less well for this prediction job than other neural network models.

Ensemble Technique:

MSE: 0.0001

RMSE: 0.0105

The ensemble technique proved to be the best accurate model for predicting worldwide video game sales, with an astonishingly low RMSE of 0.0105.

**7). Conclusion and future scope**

In this study, we explored several neural network models for predicting worldwide video game sales, taking into account elements like game type, release year, and regional sales information. Our goal was to use cutting-edge machine learning techniques to overcome the problems and restrictions that existed in earlier studies. For the purpose of identifying current models and comprehending their advantages and disadvantages, we did a thorough literature research.

Radial Basis Function Networks (RBFNs), Feedforward Neural Networks (FNN), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and an Ensemble Method were all implemented as part of our experimental effort. To train and test the models, we painstakingly collected and prepped the dataset and divided it into training and testing sets.

The outcomes showed that the various models had differing degrees of accuracy. The Ensemble Method, with a surprisingly low Root Mean Squared Error (RMSE) of 0.0105, stood out as the most accurate method. Additionally performing well and better than CNN was the FNN. The RNN and RBFN models, on the other hand, displayed larger prediction errors, highlighting their shortcomings for this particular prediction job.

In conclusion, our study demonstrates the potential of deep learning models in predicting global video game sales. While we achieved remarkable accuracy, there are still opportunities for further improvement and refinement, especially in handling data limitations, model complexity, and interpretability. This research lays the foundation for future work in enhancing the precision and practicality of video game sales prediction models.

Limitations: Handling data limitations, model complexity, and interpretability.

8). Refrences

1. Li, J., Zheng, Y., Hu, H., Lu, J., & Zhan, C. (2021, November). Predicting Video Game Sales Based on Machine Learning and Hybrid Feature Selection Method. In *2021 16th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)* (pp. 497-502). IEEE.
2. Bertens, P., Guitart, A., Chen, P. P., & Perianez, A. (2018, August). A machine-learning item recommendation system for video games. In *2018 IEEE Conference on Computational Intelligence and Games (CIG)* (pp. 1-4). IEEE.
3. Saraswathi, K., Renukadevi, N. T., Nandhinidevi, S., Gayathridevi, S., & Naveen, P. (2021, November). Sales prediction using machine learning approaches. In *AIP Conference Proceedings* (Vol. 2387, No. 1). AIP Publishing.
4. TD, G., & Jose, P. (2023). Data Interpretation and Video Games Sales Prediction Using Machine Learning Algorithms-a Comparative Study.
5. Aziz, A., Ismail, S., Othman, M. F., & Mustapha, A. (2018, July). Empirical Analysis on Sales of Video Games: A Data Mining Approach. In *Journal of Physics: Conference Series* (Vol. 1049, No. 1, p. 012086). IOP Publishing.
6. Malvankar, K., Fallon, E., Connolly, P., & Flanagan, K. (2023, January). A Case Study to Analyze the Impact of social media on Video Game Sales. In *2023 3rd International Conference on Intelligent Communication and Computational Techniques (ICCT)* (pp. 1-5). IEEE.
7. Choi, H. S., Ko, M. S., Medlin, D., & Chen, C. (2018). The effect of intrinsic and extrinsic quality cues of digital video games on sales: An empirical investigation. *Decision Support Systems*, *106*, 86-96.
8. Caballero, L. A. M. (2015). *The impact of online ratings on video game sales* (Doctoral dissertation).
9. Rosok, J. C. (2016). Diversity and Innovation: The Effects of Diverse Creator Teams on Video Game Characteristics and Sales.
10. SACRANIE, J. Consumer Perceptions & Video Game Sales.
11. Wu, M., & Neuendorf, K. A. (2011, November). Content analysis as a predictive methodology: Online video game auctions on eBay. In *Annual Conference of the National Communication Association, New Orleans, November. Available at: http://academic. csuohio. edu/kneuendorf/vitae/Wu&Neuendorf11. pdf (accessed on February 5, 2019)*.
12. De Maeyer, P. (2012). Impact of online consumer reviews on sales and price strategies: A review and directions for future research. *Journal of Product & Brand Management*, *21*(2), 132-139.
13. Diep, T. (2021). Analyzing video game sales using SAP BW/4HANA and Lumira Designer.
14. Chung, A. (2020). Predicting Sales for the Next Generation of Consoles in 2020. *Leicester Undergraduate Mathematical Journal*, *2*.
15. [IJRTI (ijcrt.org)](https://ijcrt.org/papers/IJCRT2005182.pdf) (01-10-2023)
16. [Machine Learning for Predicting Success of Video Games (muni.cz)](https://is.muni.cz/th/k2c5b/diploma_thesis_trneny.pdf) (01-10-2023)
17. [Video Game Sales Prediction Model with Python | Aman Kharwal (thecleverprogrammer.com)](https://thecleverprogrammer.com/2021/05/28/video-game-sales-prediction-model-with-python/) (01-10-2023)
18. [Predicting video game hits with Machine Learning | by Ignacio Chavarria | Towards Data Science](https://towardsdatascience.com/predicting-hit-video-games-with-ml-1341bd9b86b0) (01-10-2023)
19. [Customer Lifetime Value in Video Games Using Deep Learning and Parametric Models | IEEE Conference Publication | IEEE Xplore](https://ieeexplore.ieee.org/abstract/document/8622151) (01-10-2023)
20. [A Machine-Learning Item Recommendation System for Video Games | IEEE Conference Publication | IEEE Xplore](https://ieeexplore.ieee.org/abstract/document/8490456) (01-10-2023)
21. [JETIR1907H50.pdf](https://www.jetir.org/papers/JETIR1907H50.pdf) (01-10-2023)
22. [I07036064.pdf (questjournals.org)](https://www.questjournals.org/jrbm/papers/vol7-issue3/I07036064.pdf) (01-10-2023)
23. [(PDF) Empirical Analysis on Sales of Video Games: A Data Mining Approach (researchgate.net)](https://www.researchgate.net/publication/326510277_Empirical_Analysis_on_Sales_of_Video_Games_A_Data_Mining_Approach) (01-10-2023)
24. [FORECASTING GLOBAL VIDEO GAME SALES USING MACHINE LEARNING METHOD: SUPPORT VECTOR MACHINES EXAMPLE | Request PDF (researchgate.net)](https://www.researchgate.net/publication/351358786_FORECASTING_GLOBAL_VIDEO_GAME_SALES_USING_MACHINE_LEARNING_METHOD_SUPPORT_VECTOR_MACHINES_EXAMPLE) (01-10-2023)
25. [Predicting Video Game Sales Based on Machine Learning and Hybrid Feature Selection Method | IEEE Conference Publication | IEEE Xplore](https://ieeexplore.ieee.org/document/9755343) (01-10-2023)
26. [Microsoft Word - 093\_GRASP2017.docx (iop.org)](https://iopscience.iop.org/article/10.1088/1742-6596/1049/1/012086/pdf) (01-10-2023)
27. [Predicting video game sales using Machine Learning in Python (codespeedy.com)](https://www.codespeedy.com/predicting-video-game-sales-using-machine-learning-in-python/) (01-10-2023)
28. [(PDF) Analyzing Video Game Sales | Sebastián Soriano Pérez - Academia.edu](https://www.academia.edu/42255674/Analyzing_Video_Game_Sales) (01-10-2023)
29. [Data Visualization on video games global sales analysis& Predictive analysis on Real Estate pricing in Boston (ijcstjournal.org)](http://www.ijcstjournal.org/volume-9/issue-3/IJCST-V9I3P1.pdf) (01-10-2023)
30. [A Hybrid Subspace-Connectionist Data Mining Approach for Sales Forecasting in the Video Game Industry | IEEE Conference Publication | IEEE Xplore](https://ieeexplore.ieee.org/abstract/document/5170617) (01-10-2023)
31. [Promoting sales of online games through customer engagement - ScienceDirect](https://www.sciencedirect.com/science/article/abs/pii/S1567422315000125) (01-10-2023)
32. <https://en.wikipedia.org/wiki/Video_game>(22-9-2023)
33. <https://www.sciencedirect.com/topics/social-sciences/video-games>(22-9-2023)
34. [https://rstudiopubsstatic.s3.amazonaws.com/346100\_d6f3f54c8f454f918456dea6b23ce7b0.html#:~:text=Introduction,countries%20due%20to%20local%20preferences.](https://rstudio-pubs-static.s3.amazonaws.com/346100_d6f3f54c8f454f918456dea6b23ce7b0.html#:~:text=Introduction,countries%20due%20to%20local%20preferences.) (22-9-2023)
35. <https://rpubs.com/atika_faradilla/video_games_analysis>(22-9-2023)
36. <https://rstudiopubsstatic.s3.amazonaws.com/548644_ab7f7c6b917442cd9c263e4f0675ce12.html>(22-9-2023)