

Name: T. Sai
Chandu

Dept: CSE

course

Reg. No: 192211216

Guide: Dr. K. U. Kanimozhi

Code: SPIC438

Title → 3

3 Enhanced prediction of sales in video games using Random Forest Algorithm in comparison with Generalized Linear Model Algorithm to improve accuracy.

Introduction:

Para 1:

Definition: The video game sales prediction with the help of Random Forest Algorithm in comparison with Generalised linear Algorithm to improve accuracy (Paul Bertens, 2020).

Importance: Sales predictions on video games can guide marketing teams in devising effective promotional campaigns, pricing strategies (Patrick Kapoor, 2019) (A. Alfonso, 2019)

Applications:

→ Marketing field (Alice Yefka, 2019)

→ Entertainment field (Hyacinato Andrade, 2020)

→ Inventory field (David Buckley, 2022)

Para 2:

1. Total no. of articles published in this topic over the past 5 years

Google Scholar - 950

IEEE explore - 30

2. Most cited:

* Future Sales prediction on video game sales (Jacob sweeton, Daniel Simpson, 2021).

* The effective reviews effect on video game sales (Michal Grensoulen, 2019).

* online Rating - who decides what video games you will buy? (Erik Antansson, 2022)

* Emerging technologies in video game industry: Anticipation of future (Martina Semanova, 2020).

3. Best among them.

"changing the game: How video games are transforming the future of business" (D edery, 2018).

Para 3:

1. Handling of numeric data and sensitivity to irrelevant features.
 - accuracy is less.
 2. Existing experience in research
 - I learnt about Random Forest algorithm and generalised linear Model algorithm from coursera and Great learning.
 - After that I had a discussion with my guide and came to conclusion for providing better accuracy in videogame sales prediction.
 3. Aim of study:
 - prediction of video game sales
 - Improving the accuracy.
- Materials and Methodology:
- Improved accuracy in prediction of video game sales using Random Forest algorithm compared with Generalized Linear Model Algorithm.

Materials and Methods:

Para 1:

study setting: SIMATS, SSE, programming Lab.

No. of groups: 2

i) Group 1: Random Forest Algorithm

ii) Group 2: Generalized Linear Model algorithm

sample size: 340

Dataset: videogame sales (kaggle.com)

Para 2: Sample preparation of Group 1.

Random Forest Algorithm

- pre-processing Dataset

- applying Random Forest algorithm

- calculate accuracy.

Para 3: Sample preparation of Group 2.

Generalized linear Model Algorithm

→ pre-processing Dataset

→ Applying Generalized linear Model algorithm.

→ calculate accuracy.

Para 4:

* Testing Set-up

- Google colab / Jupyter lab

- i5 intel 10th gen.

- 8 GB RAM

- Windows OS

* Testing procedure:

- pre-processing the dataset

- Train 70% of dataset

- Train 30% of dataset

- classification and comparative analysis

Para 5: Data Collection

video game sales from kaggle.com

Para 6: statistical software used: SPSS

Independent variable: Game names, data values

Dependent variable: Accuracy.

Analysis Done: Yes.

Results and Discussion:

Para 1: In this study we observed that Random Forest has better prediction than Generalized Linear Model Algorithm.

Point 2: the graph represents the comparison of Prediction of video game sales using Random Forest algorithm and Generalized linear Model → Random Forest has high accuracy. → prediction of sales in video games.

Point 3: Data collection

→ Dataset is trained

site: www.kaggle.com.

Point 4: statistical software: SPSS

Independent variable: Game name, data values

Dependent variable: Accuracy.

Analysis Done: Yes.

→ comparison of Random Forest algorithm with Generalised linear Model Algorithm.

Limitations: The only limitation is small size

Future scope: Accuracy is improved using Random Forest.

Conclusion:

→ the overall model Prediction of video game sales on past information of data.

→ the accuracy of Random Forest is higher than Generalised linear Model.

→ the proposed algorithm Random Forest has 86.40% of accuracy over Generalized linear Model has 74.89%.

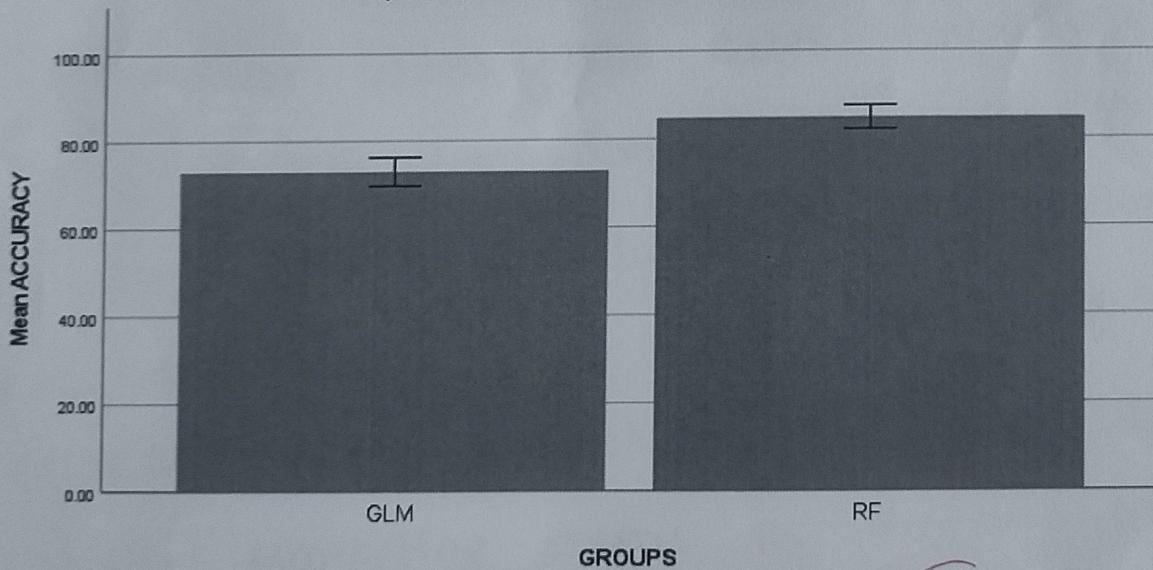
Group Statistics

	GROUPS	N	Mean	Std. Deviation	Std. Error Mean
ACCURACY	RF	10	84.7000	1.33749	.42295
	GLM	10	73.0000	1.63299	.51640

Independent Samples Test

ACCURACY	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	.066	.800	17.528	18	<.001	11.70000	.66750	10.29764	13.10236
	Equal variances not assumed			17.528	17.327	<.001	11.70000	.66750	10.29372	13.10628

Simple Bar Mean of ACCURACY by GROUPS



Error Bars: 95% CI

Error Bars: +/- 2 SD

Verified
by [Signature]
(Handwritten)