Code: SSE_26_11_216_4

Name: T. Sai chandu	Dept: CSE	Course Code: SPIC2438
Reg.No: 192211216	Guide: Dr. K.V. Kanimozhi	

Title → 4

4. Prediction of sales in video Games using Random Forest Algorithm in comparison with Gradient Boosting Algorithm to improve accuracy.

*Introduction:

Para 1:

Definition: The prediction of video game sales with the help of Random Forest Algorithm and comparing with Gradient boosting algorithm (Angelis, 2019).

Importance: The correct sales predictions contribute to building and Maintaining inves -tor confidence. (Yagis Z, 2022)

Applications:

→ Entertainment field (G. Foresti, 2020)

→ Marketing field (J. Zhange, 2019)

→ Inventary Management (M. Zalrej, 2023)

↳

Para 2:

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

Google scholar - 1300

IEEE explore - 40

2. Most cited.

*video Game sales data in North america (Aditya Rezky, Bima maulana, 2023).

*video Game sales between real rules and fictional worlds using ML (Jesper, 2019).

*customer sales perceptions on video game sales: A Machine learning approach (Amar Aziz, 2018).

*Emphical analysis on video game sales: (John sacraine, 2021).

3. Best among them.

"Violent video games sales; the newest media violence hazard" (Douglas A Gentile, 2023).

Para 3:

1. Handling of Numeric data and sensitivity to in-relavent Features.

2. Existing experience in research:

- I learnt about Random Forest algorithm and Gradient boosting Regressor algorithm from coursera and the Great learning.
- After that I had a discussion with my guide and came to a conclusion for providing better accuracy in video game sales prediction.

3. Aim of Study:

→ Prediction of video Game Sales

→ Improving the accuracy

Materials and Mathodology:

Improved accuracy in prediction of video game sales using Random Forest algorithm compared with Gradient boosting Regressor.

Materials and Methods:

Para 1: study setting: SIMATS, SSE, programming lab

No of Groups: 2

i) Group 1: Random Forest algorithm

ii) Group 2: Gradient Boosting Algorithm

sample size : 340

Dataset : video game sales (Kaggle.com).

Para 2: Sample preparation of Group 1.

Random Forest Algorithm

- preprocessing Datasets
- applying Random Forest Algorithm.
- calculate accuracy.

Para 3: Sample preparation of Group 2.

Gradient Boosting Regressor Algorithm

- preprocessing Dataset.
- applying Gradient Boosting Regressor Algorithm
- calculate accuracy.

Para 4:

* Testing set up:
  - → Google colab / Jupiter lab
  - → i5 intel 12th gen
  - → 8 GB RAM
  - → Windows OS
* Testing procedure:
  - → pre-processing the dataset
  - → Train 70% of the dataset
  - → Train 30% of the dataset
  - → classification and comparitive analysis.

Para 5: Data collection

Video game sales, from kaggle.com

Para 6: statistical software used: SPSS

Independent Variables: Game names, data value

Dependent Variables: Accuracies

Analysis Done: Yes

Results and Discussion:

Para 1: In this study we observed that Random Forest has better prediction than Gradient boosting Regressor Algorithm.

Para 2: The graph represents the comparison of Prediction of video game sales using Random Forest and Gradient boosting Regressor algorithm.

→ Random Forest have high accuracy.

→ prediction of sales in video games.

Para 3: Data collection

site: www.kaggle.com.

Para 4: Stastical software - SPSS

Independent variable - Game names, data values

Dependent variable - accuracy.

Analysis Done : Yes

→ comparison of Random Forest and Gradient boosting Algorithm.

Limitations: the only limitation is small size.

Future scope: accuracy improved using Random Forest algorithm.

Conclusion:

→ The overall model prediction of video game sales based on past information of data.

→ The accuracy of Random Forest is higher as we compared to Gradient boosting regressor Algorithm.

→ The proposed algorithm Random Forest has 86.40% of accuracy over Gradient boosting Regressor algorithm" has 78.75%.

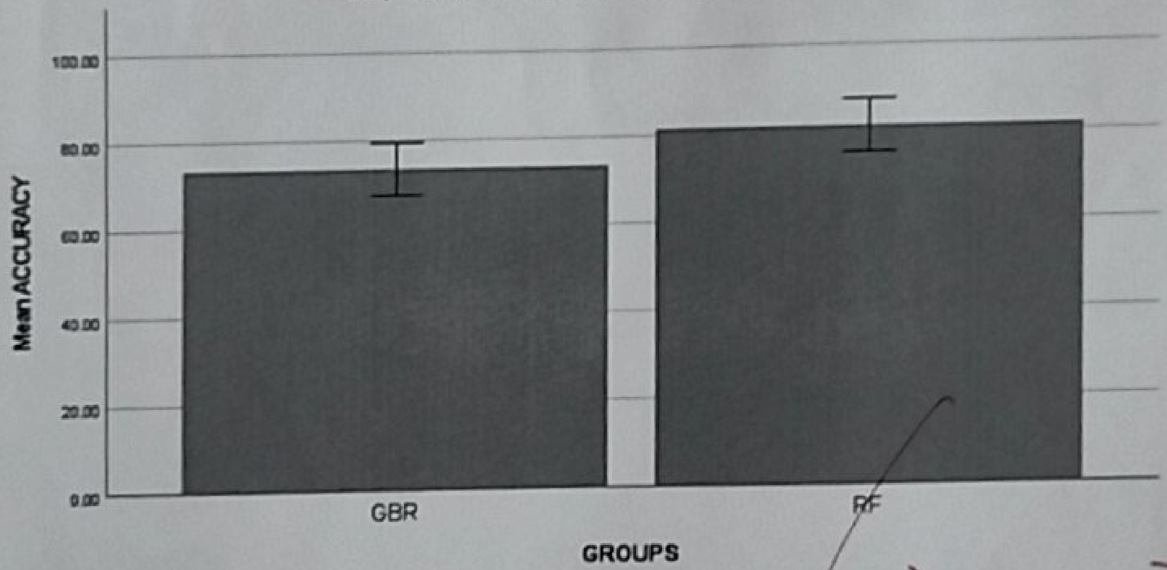
Group Statistics

	GROUPS	N	Mean	Std. Deviation	Std. Error Mean
ACCURACY	RF	10	81.5000	3.02765	.95743
	GBR	10	73.5000	3.02765	.95743

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	.000	1.000	5.908	18	<.001	8.00000	1.35401	5.15534	10.84466
	Equal variances not assumed			5.908	18.000	<.001	8.00000	1.35401	5.15534	10.84466

Simple Bar Mean of ACCURACY by GROUPS



Error Bars: 95% CI

Error Bars: +/- 2 SD

Verified (Guide)