

Name: T. Sai chandy Dept: CSE

course code:

Reg.NO: 192211216

SPICM438

Guide: Dr. K.V. Kanimozhi

Title → 2

2. Efficient prediction of sales in video games using Random Forest Algorithm in comparison with Decision Tree Algorithm to improve accuracy.

Introduction:-

Para 1:

Definition: prediction of video game sales with the help of Random Forest Algorithm and comparing with decision tree algorithm to improve accuracy.

Importance: Accurate sales predictions are crucial for game developers and publishers to plan their budgets and to allocate resources effectively.

(Patrick Kapen, 2019)

Applications:

→ Global Market analysis (Bentens, 2018)

→ Entertainment field (Guitart, 2018)

→ Economic factors (Zackariasson, 2023)

Para 2:

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

in Google scholar - 15200

in IEEE explore - 35

2. Most cited:

* Recommendation of video games with logics (H calderon - vilca, NM chavez, 2020).

* A case study to analyze the impact of social media on video game sales (M Malvankar, F Fallon, 2023).

* Machine learning for prediction success of video games (M Tinney, 2019).

* Promotion sales of video games through custom engagement (CMIC cheung, 2020)

3. Best among them is:

"How to sell a video game? Analyze the drivers of successful video game sales" (Mark Eulerich, 2023).

Para 3:

1. Handling of numeric data and sensitivity to irrelevant features.
 - accuracy is less.
2. Existing experience in tree-search
 - I learnt about Random Forest algorithm and Decision Tree algorithm from coursera and Great learning with the help of my Guide.
 - After that I had a discussion with my guide and came to conclusion for providing better accuracy in video game sales prediction

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 Decision Tree 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: Decision Tree Algorithm

Sample size: 340

Dataset: video game 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

Decision Tree Algorithm

- pre-processing dataset

- applying Decision Tree Algorithm

- calculate accuracy.

Para 4:

* Testing setup:

→ Google colab / Jupiter lab

→ i5 intel 12th gen

→ 8 GiB RAM

→ windows OS

* Testing Procedure:

→ pre-processing the dataset

→ Train 70% of data set

→ 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 variables: Game names, data values

Dependent Variables: accuracies

Analysis Done: Yes

Results and discussion:

Para 1: In this study we observed that Random Forest Algorithm has better prediction than Decision

Tree Algorithm.

Para 2: The graph represents the comparison of prediction of video game sales using Random Forest algorithm and Decision Tree Algorithm.

→ Random Forest Algorithm have high accuracy

→ Prediction of sales in video games

Para 3: Data Collection

→ Dataset is trained

site: www.kaggle.com.

Para 4: statistical software - SPSS

Independent variable - Game names, data value

dependent variable - accuracy.

Analysis Done - Yes.

→ comparison of Random Forest algorithm and Decision Tree Algorithms.

Limitations: The only limitation is small size.

Future scope: Accuracy improved using Random Forest.

Conclusion:

→ The overall model prediction of Video game sales based on Past information of data

→ The accuracy of Random Forest has higher compared to Decision Tree algorithm.

→ The proposed algorithm Random Forest has 86.40% of accuracy over Decision Tree has 79.32%.

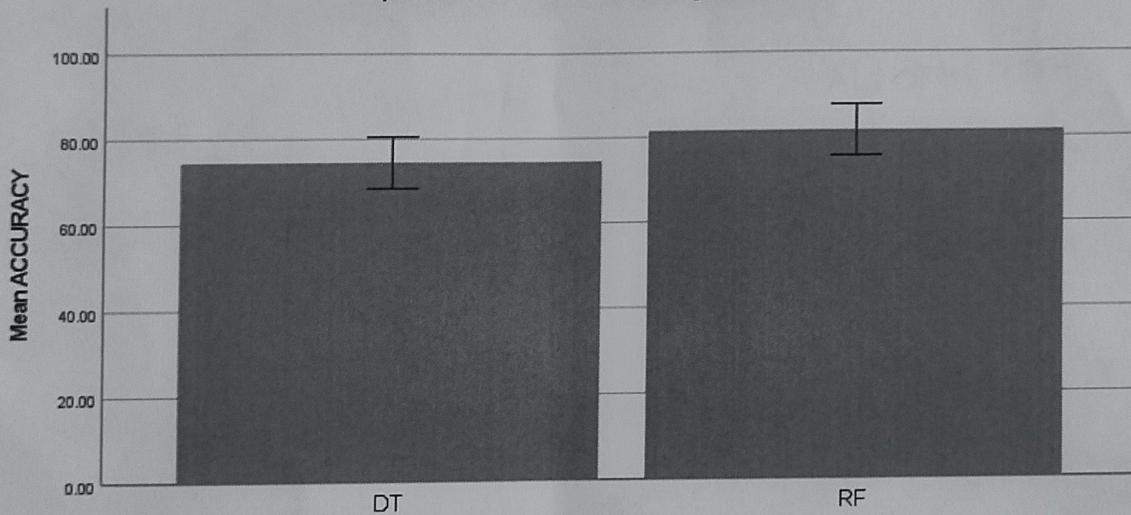
Group Statistics

	GROUPS	N	Mean	Std. Deviation	Std. Error Mean
ACCURACY	RF	10	81.5000	3.02765	.95743
	DT	10	74.5000	3.02765	.95743

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	Lower
Equal variances assumed	.000	1.000	5.170	18	<.001	7.00000	1.35401	4.15534	9.84466
Equal variances not assumed			5.170	18.000	<.001	7.00000	1.35401	4.15534	9.84466

Simple Bar Mean of ACCURACY by GROUPS



Error Bars: 95% CI

Error Bars: +/- 2 SD

Verified graphically (Anide).