Name : T. Sai Chandu Dept: CSE

Reg.no: 19221216 Guide: Dr. K.V. Kanimozhi

course Code : SP1C438

Title →1

1. Automatic prediction of sales in video game using Random forest algorithm in comparis -on with support vector Machine algorithm to improve accuracy.

Introduction:

Paragraph 1 :

Definition: To predict the video game sales with the help of Random Forest Algorithm and comparing with support vector Machine algorithm for improving accuracy (Y Zheng, 2021)

Importance:

Predicting sales of the video game can pre-adjust sales strategies and development plans in advance (Haoran Hu), 2021)

Applications:

→ video games are entertainment experiences (Mary Beth oliver, 2019).

→ Inventary Management (Gowri T.D, 2023)

Para 2:

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

Google scholar - 17500

IEEE explore - 43

- Most cited:

2. * A hybrid subspace-connectionlist data mining approach for sales prediction in video game industry (J Marcoux, 2019).

* Research on the prediction of sales in video games (W Huang, 2023).

* Automatic Machine learning based Data analysis for video game industry (Z Zhou, 2022)

* Video game sales prediction using Machine learning algorithms - a comparitive study (Peter Jose, 2023).

3. Best among them according to me "A time series approach to analyze the video game sales" (Malvankar, 2023)

Para 3:

1. Handling of numeric data, and sensitivity to in-relavent features
   - accuracy is less.
2. Existing experience in research.
   - I learnt about Random forest and support vector Machine algorithms from courser and Great learning with the help of my Guide.
   - Next 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 Methodology:

Improved accuracy in prediction of video game sales using Random Forest algorithm compared with support vector machine 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: support vector Machine

sample size : 340

Dataset : video game_sales (kaggle.com).

Para 2: Sample preparation of Group 1.

- Random forest Algorithm
- preprocessing dataset of video games sales
- applying Random Forest algorithm
- calculate accuracy.
- calculate prediction.

Para 3: sample preparation of Group 2.

support vector Machine algorithm.

- pre-processing dataset
- applying support vector Machine
- calculate accuracy.

Para 4:

* Testing setup:
  - → Google colab/ Jupitar lab
  - → is intel 12th gen
  - → 8 GiB RAM
  - → Windows OS
* Testing procedure:
  - → Pre-processing the dataset
  - → Train 70% of dataset
  - → Train 30% of dataset
  - → classification and comparitive analysis

Para 5: Data collection

Video game sales from kaggle.com.

Para 6:

stastical software used : SPSS

Independent variables : Machine learning (Game names, data values).

Dependent variables : Accuracies

Analysis Done : Yes

Results and Discussion:

Para 1: In this study we observed that Random Forest has better prediction than support vector Machine algorithm.

Para 2: The graph represents the comparison of prediction of video game sales using Random Forest and support vector Machine.

→ Random Forest haves high accuracy.

→ prediction of sales in so video games

Para 3: Data collection

→ Dataset is trained

sites: www.kaggle.com

Para 4:

statistical software - SPSS

Independent variable - Game names, data values

Dependent variable - accuracy.

Analysis Done : yes.

→ comparison of Random Forest and support vector Machine 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 is higher compared to support vector Machine algorithm.

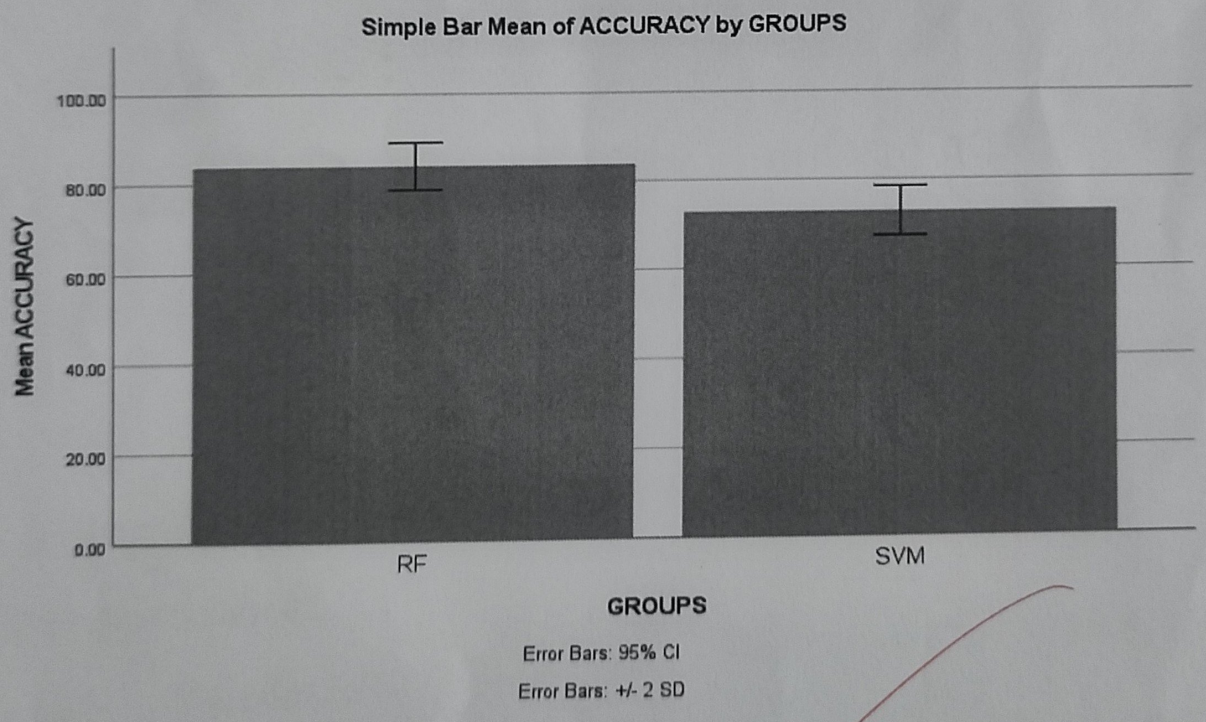
→ The proposed algorithm Random Forest has 86.40% of accuracy over SVM has 74.49%

Group Statistics

	GROUPS	N	Mean	Std. Deviation	Std. Error Mean
ACCURACY	RF	10	83.8000	2.65832	.84063
	SVM	10	72.8000	2.78089	.87939

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	9.042	18	<.001	11.00000	1.21655	8.44412	13.55588
	Equal variances not assumed			9.042	17.964	<.001	11.00000	1.21655	8.44375	13.55625



Verified
15/2/24
(Guide).