Video Games Sales Prediction Using Wide and Dense Deep Networks

Megha Sivasankar
Abishek Thamizharasan
Sakshi Kumari (62768)

Dataset & EDA

Dataset: https://github.com/jacobholm10/Global-Sales-Prediction-for-Video-Games

	Platform	Year	Genre	Publisher	NA_Sales	JP_Sales	Global_Sales	Critic_Score	Critic_Count	User_Score	User_Count
0	Wii	2006.0	Sports	Nintendo	41.36	3.77	1	76.0	51.0	8	322.0
2	Wii	2008.0	Racing	Nintendo	15.68	3.79	1	82.0	73.0	8.3	709.0
3	Wii	2009.0	Sports	Nintendo	15.61	3.28	1	80.0	73.0	8	192.0
6	DS	2006.0	Platform	Nintendo	11.28	6.50	1	89.0	65.0	8.5	431.0
7	Wii	2006.0	Misc	Nintendo	13.96	2.93	1	58.0	41.0	6.6	129.0

Key Features:

-
- Categorical and numerical features

Model Architecture & Training

Model.png: https://drive.google.com/file/d/1_qFhH7M3CaqhYgEbEHbWhRdHMqtGOx2D/view?usp=sharing

- Input Layers: Cross, Categorical, Numerical
- Dense Layer: Units-50, 25,10; ReLU Activation
- Output Layer: 1 Unit, Sigmoid Activation

Training Process:

- Optimizer: adagrad, Adam
- Loss: mean_squared_error, Binary Crossentropy
- Epochs: 50, Batch Size: 32

Model Evaluation & Best Model

optimizer='adagrad'

loss='mean_squared_error'

```
[[626   0]
 [ 39  17]]
              precision    recall  f1-score   support

           0       0.94      1.00      0.97       626
           1       1.00      0.30      0.47        56

    accuracy                           0.94       682
   macro avg       0.97      0.65      0.72       682
weighted avg       0.95      0.94      0.93       682
```

0.97

0.9428152492668622

optimizer='adam'

loss='Binary Crossentropy'

```
[[626   0]
 [ 27  29]]
              precision    recall  f1-score   support

           0       0.96      1.00      0.98       626
           1       1.00      0.52      0.68        56

    accuracy                           0.96       682
   macro avg       0.98      0.76      0.83       682
weighted avg       0.96      0.96      0.95       682
```

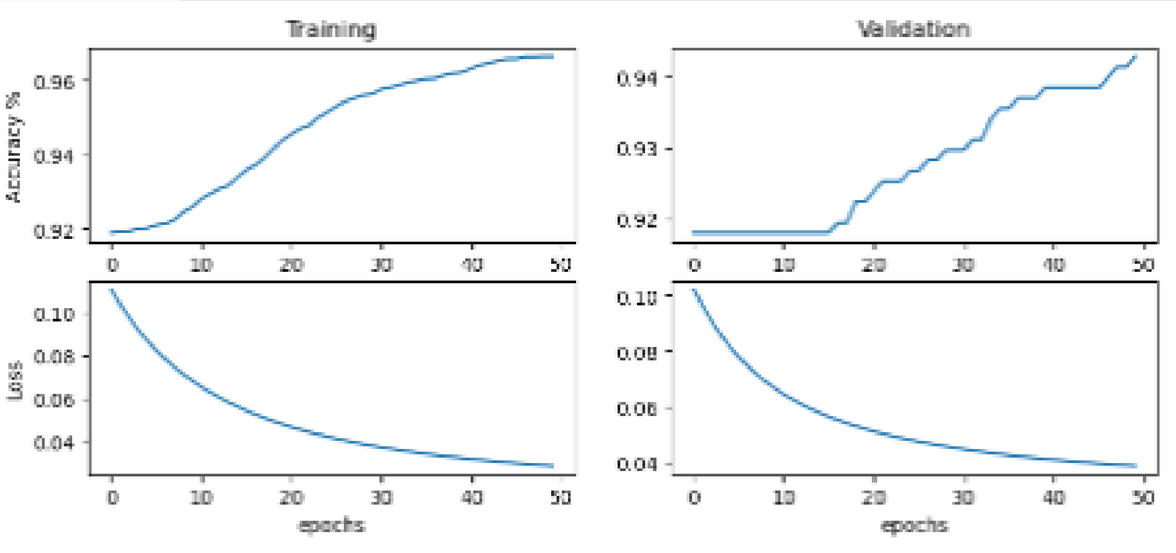
0.98

0.9604105571847508

Model Evaluation & Best Model

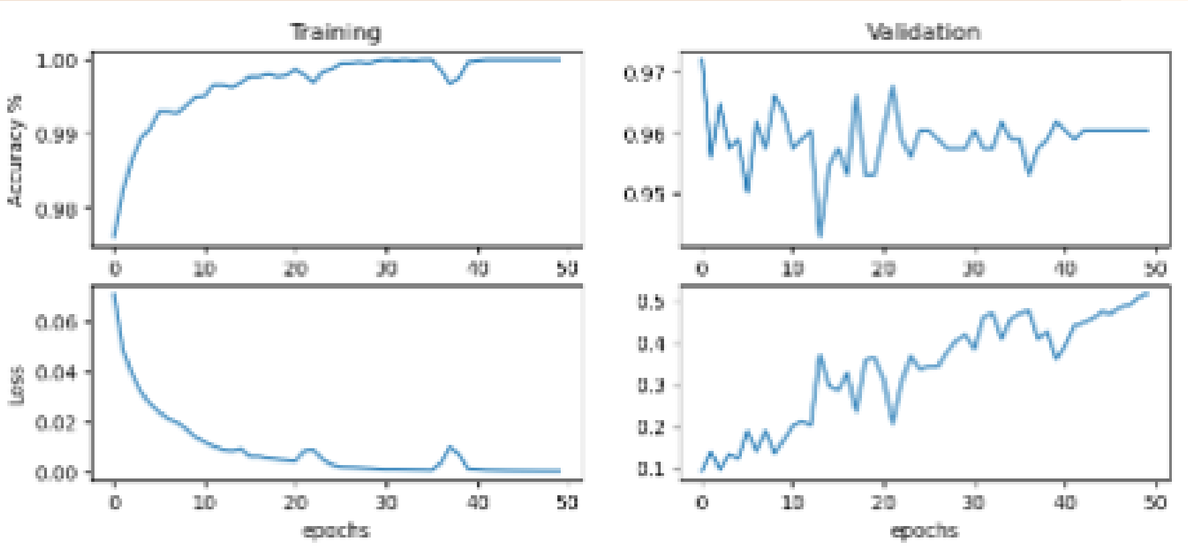
optimizer='adagrad'

loss='mean_squared_error'



optimizer='adam'

loss='Binary Crossentropy'



Hyperparameter Tuning (Keras Tuner)

- RandomSearch
- Max Trials: 5
- Best val_accuracy: 0.982
- Total elapsed time: 00h 04m 08s

```
[[625   1]
 [  7  49]]
```

	precision	recall	f1-score	support
0	0.99	1.00	0.99	626
1	0.98	0.88	0.92	56
accuracy			0.99	682
macro avg	0.98	0.94	0.96	682
weighted avg	0.99	0.99	0.99	682

Testing

Sales Estimation and Data Merging:

- Estimating Sales from Predictions:
  - We've successfully generated predictions using our pre-trained model. However, these predictions are binary, representing whether a game's sales are estimated to be above or below 2 million units.
  - To translate these binary predictions into estimated sales numbers, we use a straightforward approach. We calculate the average sales for each category in the original dataset:
    - For games predicted to have sales above 2 million, we take the average of the rating in that category.
    - For games predicted to have sales below or equal to 2 million, we compute the average rating within that category.
  - These averages serve as our estimated sales values, allowing us to gauge the expected performance of each game.

Output

- Accuracy: 0.4938
- Precision: 0.2288
- Recall: 0.1226
- F1 Score: 0.3846

- Confusion Matrix:

True Positives (TP): 69

True Negatives (TN): 3301

False Positives (FP): 2961

False Negatives (FN): 494

updated_predicted_video_game_sales (1)

Name	index	Platform	Year	Genre	Publisher	Global_Sales	Critic_Score	Critic_Count	User_Score	User_Count	Developer	Rating	Predicted_Sales_Category
Wii Sports	0	Wii	2006.0	Sports	Nintendo	1	76.0	51.0	8.0	322.0	Nintendo	E	Less than 2M
Super Mario Bros.	2	Wii	2008.0	Racing	Nintendo	1	82.0	73.0	8.3	709.0	Nintendo	E	Less than 2M
Mario Kart Wii	3	Wii	2009.0	Sports	Nintendo	1	80.0	73.0	8.0	192.0	Nintendo	E	Less than 2M
Wii Sports Resort	6	DS	2006.0	Platform	Nintendo	1	89.0	65.0	8.5	431.0	Nintendo	E	More than 2M
Pokemon Red/Pokemon Blue	7	Wii	2006.0	Misc	Nintendo	1	58.0	41.0	6.6	129.0	Nintendo	E	Less than 2M
Tetris	8	Wii	2009.0	Platform	Nintendo	1	87.0	80.0	8.4	594.0	Nintendo	E	Less than 2M
New Super Mario Bros.	11	DS	2005.0	Racing	Nintendo	1	91.0	64.0	8.6	464.0	Nintendo	E	Less than 2M
Wii Play	13	Wii	2007.0	Sports	Nintendo	1	80.0	63.0	7.7	146.0	Nintendo	E	Less than 2M
New Super Mario Bros. Wii	14	X360	2010.0	Misc	Microsoft Game Studios	1	61.0	45.0	6.3	106.0	Good Science Studio	E	Less than 2M
Duck Hunt	15	Wii	2009.0	Sports	Nintendo	1	80.0	33.0	7.4	52.0	Nintendo	E	Less than 2M
Nintendogs	16	PS3	2013.0	Action	Take-Two Interactive	1	97.0	50.0	8.2	3994.0	Rockstar North	M	Less than 2M
Mario Kart DS	17	PS2	2004.0	Action	Take-Two Interactive	1	95.0	80.0	9.0	1588.0	Rockstar North	M	Less than 2M
Pokemon Gold/Pokemon Silver	19	DS	2005.0	Misc	Nintendo	1	77.0	58.0	7.9	50.0	Nintendo	E	More than 2M
Wii Fit	23	X360	2013.0	Action	Take-Two Interactive	1	97.0	58.0	8.1	3711.0	Rockstar North	M	Less than 2M
Kinect Adventures!	24	PS2	2002.0	Action	Take-Two Interactive	1	95.0	62.0	8.7	730.0	Rockstar North	M	Less than 2M
Wii Fit Plus	26	DS	2005.0	Puzzle	Nintendo	1	77.0	37.0	7.1	19.0	Nintendo	E	More than 2M
Grand Theft Auto V	28	PS2	2001.0	Racing	Sony Computer Entertainment	1	95.0	54.0	8.4	314.0	Polyphony Digital	E	Less than 2M
Grand Theft Auto: San Andreas	29	X360	2011.0	Shooter	Activision	1	88.0	81.0	3.4	8713.0	Infinity Ward, Sledgehammer Games	M	Less than 2M
Super Mario World	32	X360	2010.0	Shooter	Activision	1	87.0	89.0	6.3	1454.0	Treyarch	M	Less than 2M
Brain Age: Train Your Brain in Minutes a Day	34	PS3	2012.0	Shooter	Activision	1	83.0	21.0	5.3	922.0	Treyarch	M	Less than 2M
Pokemon Diamond/Pokemon Pearl	35	X360	2012.0	Shooter	Activision	1	83.0	73.0	4.8	2256.0	Treyarch	M	Less than 2M

Displaying Predictions and Saving Data:

- Displaying Predictions:
  - Our combined DataFrame provides a detailed overview of our predictions. It includes the names of the video games and their respective predicted sales categories.
  - By visualizing this data, we gain insights into how our model categorizes each game's potential sales performance.
  - This presentation of predictions alongside game names facilitates further analysis, aiding stakeholders in making informed decisions about their video game titles.
- Saving Data for Further Analysis:
  - Recognizing the importance of data preservation and future investigations, we offer the option to save the combined dataset to a new CSV file.
  - This CSV file, named 'predicted_video_game_sales.csv,' contains all the relevant information about video game titles and their predicted sales categories.

THANK YOU SO MUCH!