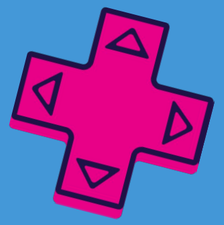


WHAT MAKES A VIDEO GAME A “HIT”

– A MACHINE LEARNING APPROACH

TEAM 4B:

Meghna Prakash, Edward Dai, Alex Huang,

Srinidhi Chevvuri, Preme Chinpattanakul

START

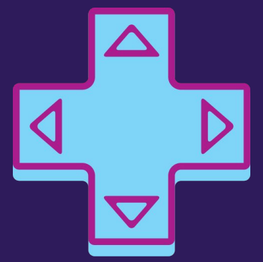
Problem Statement

Can video games be classified as a “hit” or “non-hit” using their attributes such as platform, genre, publisher, and global sales?

Business Implications: why this matters!

If we can predict “hits” early, companies can:

- ➔ Highly competitive gaming market
- ➔ Allocate marketing budget efficiently
- ➔ Prioritize promising game concept
- ➔ Reduce financial risk in new game development
- ➔ Forecast demand more accurately



Dataset Overview

1907

Video Game Data Points

13

Variables (Categorical and Numerical)

10.0

Usability Rate Recently published/updated

13 Attributes

INTEGER

- Rank
- Year
- North America
- Europe
- Japan
- Rest of World
- Global

STRING

- Game Title
- Platform
- Genre
- Publisher

FLOAT

- Review
- is_hit (created)

Data Preparation

- Dropped 31 missing values → now only have **1878** data points
- Created binary target variables **is_hit** using review score ≥ 75
- Consistent **80:20** train:test
  - (use weights to accommodate uneven hit/non-hit ratio)
  - → all models
- Drop columns → Index , Rank
  - Why Rank?
    - Allows the model to “cheat”

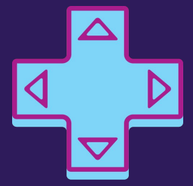
```
gaming.isnull().sum()
```

```
index              0
Rank               0
Game Title         0
Platform           0
Year              29
Genre              0
Publisher          2
North America      0
Europe             0
Japan              0
Rest of World      0
Global             0
Review             0
dtype: int64
```

Define a Hit vs. Non-Hit Game

Hit Game

- **Review score ≥ 75**
- “Generally Favorable” on Metacritic
- Strong critical perception and **higher audience approval**



Non-Hit Game

- **Review score < 75**
- “Mixed or Lower” reviews
- Average or below-average perception

Why Metacritic?

- Widely used industry benchmark
- Aggregates professional critic reviews
- Less influenced by review bombing

Model Analysis

Logistic Regression

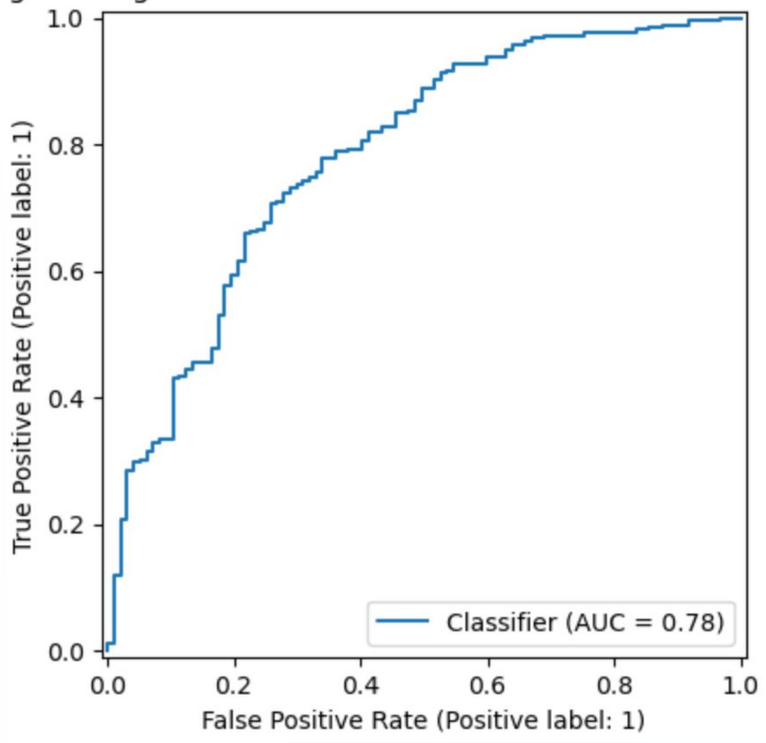
Decision Tree

Random Forest

Naive Bayes

Logistic Regression

- **Accuracy** : 73%
- **Non-Hits (class 0):**
  - predicts “non-hit” correct 48% of the time
  - identifies only 68% of actual non-hits
  - F1: 56%
- **Hits (class 1)** :
  - predicts “hit” correct 87% of the time
  - identifies 74% of actual hits
  - F1: 80%
- **AUC**:
  - 0.780

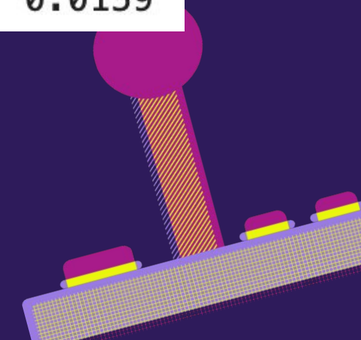



```
Accuracy: 0.726063829787234

Confusion Matrix:
 [[ 66  31]
 [ 72 207]]

Classification Report:
              precision    recall  f1-score   support

           0       0.48      0.68      0.56        97
           1       0.87      0.74      0.80       279

    accuracy                           0.73       376
   macro avg       0.67      0.71      0.68       376
weighted avg       0.77      0.73      0.74       376
```

Logistic Regression (cont.)

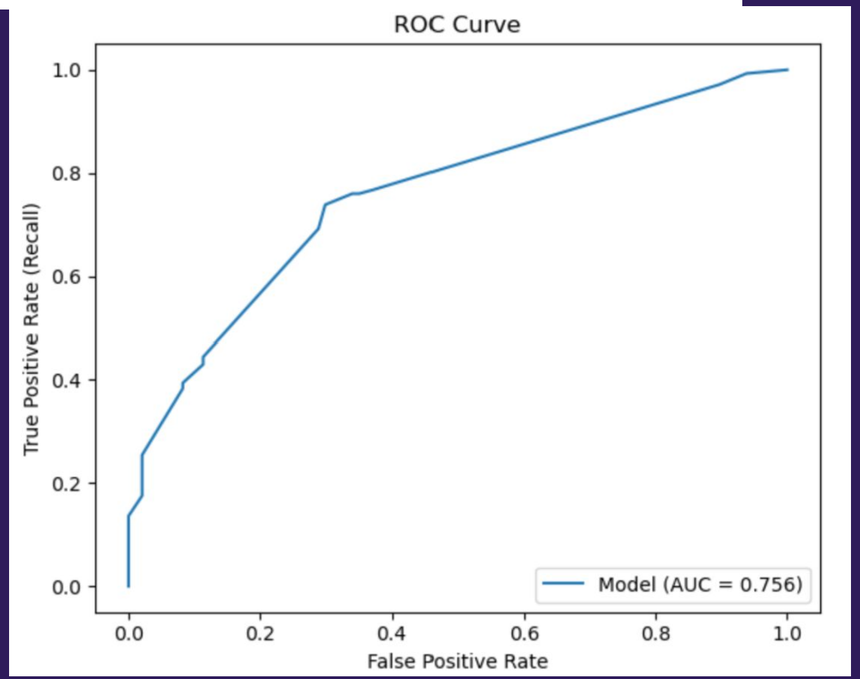
- **Publishers** : High importance in predicting hits → Nintendo
- **Genre** : Fighting and Roleplaying
- **Regional** : High sales in North America
- **Platform** : Available on Nintendo and PC platforms

```
Top 15 most important features:
Publisher_Nintendo                          0.0342
Genre_Fighting                              0.0243
North America                               0.0232
Platform_Wii                                0.0201
Platform_DS                                 0.0193
Publisher_MTV Games                         0.0191
Genre_Role-Playing                          0.0185
Publisher_505 Games                         0.0183
Platform_GEN                                0.0176
Publisher_Take-Two Interactive              0.0173
Platform_PC                                 0.0173
Publisher_Enix Corporation                  0.0170
Publisher_Sony Computer Entertainment       0.0165
Publisher_Universal Interactive             0.0162
Publisher_Empire Interactive                0.0159
```



Decision Tree

```
Accuracy on test set: 0.6968085106382979

Classification Report:
              precision    recall  f1-score   support

           0       0.45      0.71      0.55        97
           1       0.87      0.69      0.77       279

    accuracy                           0.70       376
   macro avg       0.66      0.70      0.66       376
weighted avg       0.76      0.70      0.71       376

Confusion Matrix:
 [[ 69  28]
 [ 86 193]]
```



- **Accuracy** : ~70%
- **Non-Hits (class 0)** :
  - predicts “non-hit” correct 45% of the time
  - identifies only 71% of actual non-hits
  - F1: 55%
- **Hits (class 1)** :
  - predicts “hit,” correct 87% of the time
  - identifies 69% of actual hits
  - F1: 77%
- **AUC**:
  - 0.756

Decision Tree (cont.)

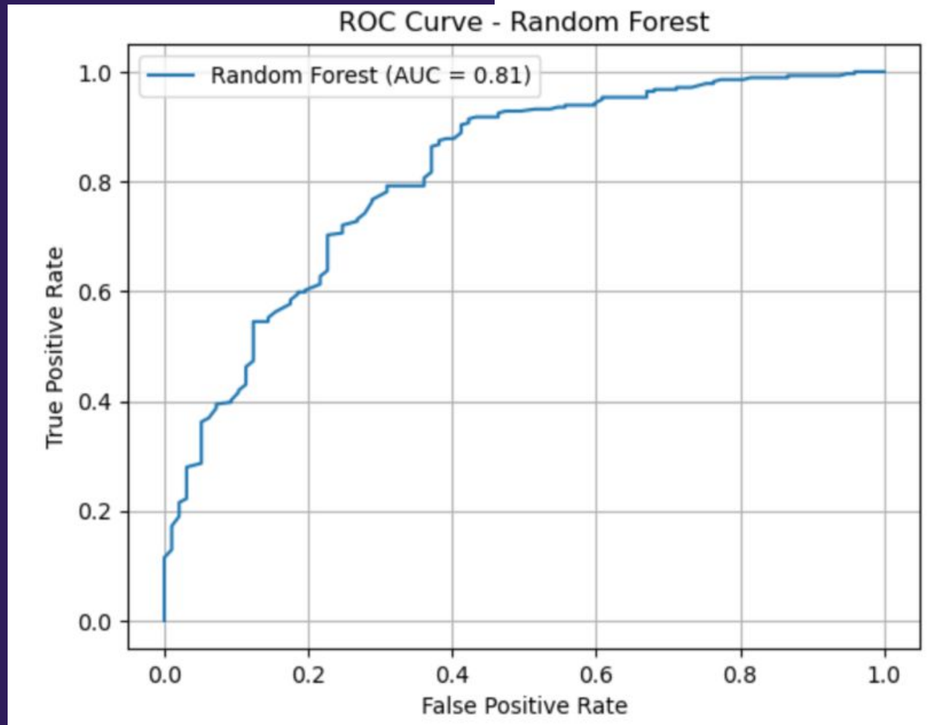
```
Top 15 most important features:

Global                                  0.262654
Europe                                  0.164984
Japan                                   0.107223
Platform_PC                             0.088289
Genre_Misc                              0.082919
Genre_Sports                            0.074923
Platform_Wii                            0.056466
Platform_XB                             0.040789
North America                           0.026530
Publisher_Take-Two Interactive          0.025476
Genre_Role-Playing                      0.025276
Year                                    0.022301
Rest of World                           0.022168
Platform_NES                            0.000000
Platform_PSV                            0.000000
dtype: float64
```

- **Regional Sales** : Global sales → high influence on hit
- **Platform** : PC, Nintendo, Xbox → positive influence
- **Genre** : Sports, role playing, & miscellaneous → likelier to be hits
- **Publisher** : Take-Two Interactive →hit
- **Year** : Release time → positively influence likeliness to become a hit

Random Forest

- **Accuracy** : 80%
- **Non-hits (class 0):**
  - predicts “non-hit” correctly 71% of the time
  - identifies 40% of the actual non-hits
  - F1: 51%
- **Hits (class 1):**
  - predicts “hit” correct 82% of the time
  - identifies 94% of actual hits
  - F1: 88%
- **AUC**: 0.81

```
Accuracy: 0.8031914893617021

Confusion Matrix:
 [[ 39  58]
 [ 16 263]]

Classification Report:

              precision    recall  f1-score   support

           0       0.71      0.40      0.51        97
           1       0.82      0.94      0.88       279

    accuracy                           0.80       376
   macro avg       0.76      0.67      0.69       376
weighted avg       0.79      0.80      0.78       376


ROC-AUC: 0.8128256290876844
```



Random Forest (cont.)

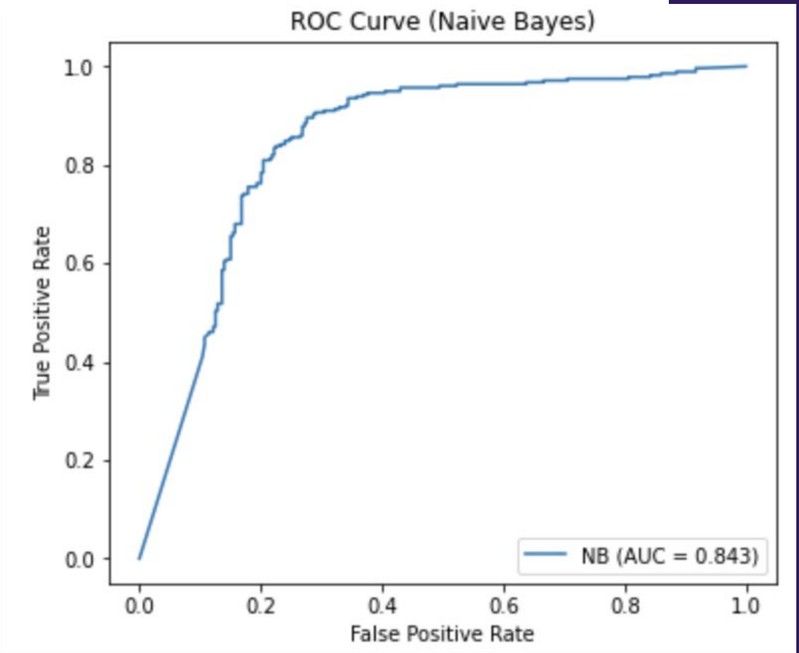
- **Regional Sales** : higher sales globally → hits
  - North America, Europe, Japan
- **Year** : Release timing→ influence on being a hit
- **Genre:** Sports, Role playing, Misc, Platform, Action→ likelier to be hits
- Platform: Nintendo Console → hits
  - Wii, DS

Top 15 Feature Importances:

	Feature	Importance
120	Global	0.119876
116	North America	0.117920
117	Europe	0.097769
119	Rest of World	0.091399
115	Year	0.079227
118	Japan	0.074159
18	Platform_Wii	0.020486
25	Genre_Misc	0.019801
32	Genre_Sports	0.019613
55	Publisher_Electronic Arts	0.015372
29	Genre_Role-Playing	0.013871
26	Genre_Platform	0.012439
102	Publisher_THQ	0.012115
22	Genre_Action	0.011926
2	Platform_DS	0.011907

Naive Bayes

```
Model Accuracy: 0.6696588868940754

Classification Report (Balanced Data):
              precision    recall  f1-score   support

           0       0.91      0.38      0.53       279
           1       0.61      0.96      0.74       278

    accuracy                           0.67       557
   macro avg       0.76      0.67      0.64       557
weighted avg       0.76      0.67      0.64       557


Confusion Matrix (Balanced Data):
[[105 174]
 [ 10 268]]
```



- **Accuracy** : ~67%
- **Non-Hits (class 0)** :
  - predicts “non-hit” correct 91% of the time
  - identifies 38% of actual non-hits
  - F1: 53%
- **Hits (class 1)** :
  - predicts “hit” correct 61% of the time
  - identifies 96% of actual hits
  - F1: 74%
- **AUC**:
  - 0.843

Naive Bayes (cont.)

Top 15 most important features:

Feature	Importance
Platform_Wii	0.340933
Genre_Misc	0.340737
Genre_Role-Playing	0.329201
Japan	0.308380
Global	0.298403
Platform_DS	0.291094
North America	0.266401
Platform_PC	0.252413
Genre_Fighting	0.248308
Publisher_Disney Interactive Studios	0.237907
Rest of World	0.227684
Publisher_Nintendo	0.218879
Europe	0.217961
Publisher_Take-Two Interactive	0.210606
Platform_PS3	0.207172

- **Platform:** Wii, DS, PC → high influence on hit
- **Genre** : Misc., role playing, & fighting → likelier to be hits
- **Regional** : Japan, Global, North America→ hit
- **Publisher** : Nintendo, Take-two Interactive → positively influence likeliness to become a hit

Best ML Approach is...

Logistic Regression

73% Accuracy

Decision Tree

70% Accuracy

Random Forest

80% Accuracy

Naive Bayes

67% Accuracy

Best ML Approach is...

Random Forest

80% Accuracy

Insights

- Global sales performance and positive reviews go hand-in-hand
- Publisher reputation matters
  - Nintendo, Take-Two Interactive
- Platform ecosystems influence success
  - PC, Nintendo, Xbox, PlayStation
- Genre patterns emerge
  - Sports, Role Playing, Miscellaneous, and Fighting
- Release timing holds influence on success
- Models mostly agree on core drivers
  - Tend to measure hits more accurately than non-hits even with weights

What this analysis represents... or not

What This Analysis Represents:

- ➔ A data-driven blueprint
- ➔ A comparative summary
- ➔ A foundation for decision-making

What This Analysis Does Not Represent:

- ➔ It does not capture creativity or cultural impact
- ➔ It does not predict long-term franchise success
- ➔ It is limited to structured features

Citations

- Movie Reviews, TV Reviews, Game Reviews, and Music Reviews. https://www.metacritic.com/. Accessed 1 Dec. 2025.

- https://www.kaggle.com/datasets/thedevastator/discovering-hidden-trends-in-global-video-games. Accessed 1 Dec. 2025.

THANK YOU!