

The background is a light blue gradient with various gaming-related icons scattered around. In the top left, there's a grey and blue game controller. In the top right, there's a black and white striped ball and a pair of blue and white headphones. In the bottom left, there's a blue and white VR headset. In the bottom right, there's a blue handheld console and a grey handheld console. There are also several colorful spheres (purple, orange, pink) and a large pink 'X' shape.

Predicting Video Games Success with Genres

With Decision Trees and Random Forest

Tommy Phung

Overview



Stakeholders: Games Developers and Game Investors



Goal: Suggest genres to new video games to hit profitable benchmarks

Best Model Accuracy: **77.08%**

Best Genres Overall: **Adventure, RPG, Indie**

Suggested Genres: **Puzzle and Strategy**

Business Understanding

High Demand: **26%** increase in demand **2019 - 2021**

➤ Projected to **increase**



Relevance: **145 Billions** Gaming vs **42 Billions** Box Office

➤ Most Lucrative Industry 

Low Success Rate: **20%** of Video Games **Success Rate**

➤ **34** Games produced daily (Steam)

Data Understanding

Rank	Top Average Rating	Most Common
1	Turn Based Strategy	Adventure
2	Visual Novel	RPG
3	Tactical	Indie
4	RPG	Shooter
5	Adventure	Platform

First Impressions:
Safest Options -
RPG
Adventure

Not Enough Data for other Top
Rated Genres

Data Preparation



Source 1: Kaggle

16,598 titles

Feature: Global_Sales



Combined

7,593 Overlapped Titles

Dups, Multiple Platforms

1st Dataset

Second Dataset

Third Dataset

Final Dataset



Source 2: Backloggd

60,000 titles

Features: Genres



Cleaned

4,488 Unique Titles

24 Genre Features

Metrics

 **Model Performance Accuracy:** Number of correct predictions

Amount of Sale Needed for Profit

Larger Developers \Rightarrow 1,000,000



Indie Developers \Rightarrow 100,000



Recommendations - Indie



Accuracy: 78.51%

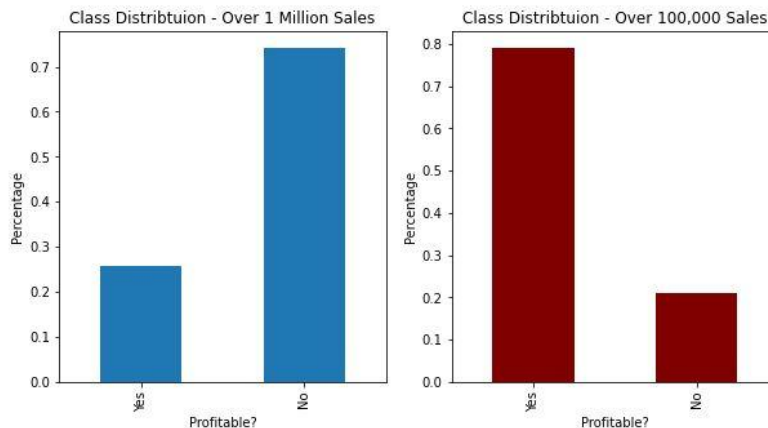
Suggestion: **Strategy** and **Arcade**

Rank	Genres
1	RPG
2	Platformers
3	Adventure
4	Strategy
5	Arcade

Class Imbalance

Definition: An **uneven** distribution in the targeted feature.

Solution: **Synthetically** create more data for minority



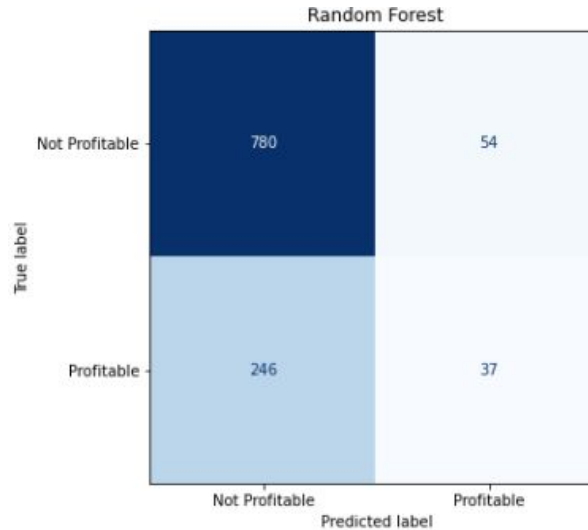
Confusion Matrix

True Negative:

- Pred → Negative
- Actual → Negative

False Positive:

- Pred → Positive
- Actual → Negative



False Negative:

- Pred → Negative
- Actual → Positive

True Positive:

- Pred → Positive
- Actual → Positive

Class Imbalance Effect

False Positive: Assume a video game is **successful**
→ Recommends **wrong genres**

False Negative: Assume a video game is **unsuccessful**
→ **Ignores** potentially good genres

Unbalanced unique genre:

➤ **Strategies** 

Balanced Unique Genres:

➤ **Brawler** 

Recommendations - Developed



Accuracy: 59.18% (Balanced)

Suggestion: **Puzzle** and **Simulator**

Rank	Genres
1	RPG
2	Brawler
3	Adventure
4	Puzzle
5	Simulator

Recommendations



Experienced Developer

Safe	Experiment
Adventure	Simulator
RPG	Puzzle



Indie Developers

Safe	Experiment
Adventure	Strategy
RPG	Arcade

Next Step

More Games

- **Limited** by matches

More features

- **Developer**, budget, platform

Identify Trends:

- Audience genres trends - **MOBA, Arcade**, etc.



Future Steps

New Modeling Technique: **Light GBM**

- **Potential Reduce Error during modeling**

Possible Benefit:

- **Predict number of sales** rather than profitability

Example: Average Error (RMSE) = **3.03 Million**

Predicted Number of Sales	Actual Number of Sales	Difference in Sales
3.83 Million	2.66 Million	1.17 Million

Questions?

Email: phungtommy109@gmail.com

Github: <https://github.com/Tommyphung1>

Notebook: <https://github.com/Tommyphung1/Project>

Thanks !

