

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM  
TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG  
KHOA CÔNG NGHỆ THÔNG TIN



TRẦN QUANG QUÂN - 520H0569  
NGUYỄN PHÚC MINH ĐĂNG – 521H0497

UBISOFT ENTERTAINMENT SA

BÁO CÁO CUỐI KỲ  
HỆ THỐNG THƯƠNG MẠI THÔNG  
MINH

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2025

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM  
TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG  
KHOA CÔNG NGHỆ THÔNG TIN



TRẦN QUANG QUÂN – 520H0569  
NGUYỄN PHÚC MINH ĐĂNG – 521H0497

UBISOFT ENTERTAINMENT SA

BÁO CÁO CUỐI KỲ  
HỆ THỐNG THƯƠNG MẠI THÔNG  
MINH

Người hướng dẫn  
**ThS. Võ Thị Kim Anh**

THÀNH PHỐ HỒ CHÍ MINH, NĂM 2025

## LỜI CẢM ƠN

Chúng em xin chân thành cảm ơn cô Võ Thị Kim Anh đã giảng dạy và hướng dẫn các kiến thức chuyên môn Hệ thống thương mại thông minh để chúng em có thể hoàn thành bài báo cáo này.

TP. Hồ Chí Minh, ngày 12 tháng 12 năm 2025

Tác giả

(Ký tên và ghi rõ họ tên)

Quân

Trần Quang Quân

Đăng

Nguyễn Phúc Minh Đăng

# CÔNG TRÌNH ĐƯỢC HOÀN THÀNH

## TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi và được sự hướng dẫn khoa học của ThS. Võ Thị Kim Anh. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong Bài báo cáo còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung Dự án của mình.** Trường Đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

TP. Hồ Chí Minh, ngày 12 tháng 12 năm 2025

Tác giả

(Ký tên và ghi rõ họ tên)

*Quân*  
Trần Quang Quân

*Đăng*  
Nguyễn Phúc Minh Đăng



## TÓM TẮT

### PHÂN TÍCH HIỆU SUẤT KINH DOANH & XÂY DỰNG MÔ HÌNH TỐI ƯU HÓA DOANH THU CHO UBISOFT ENTERTAINMENT SA

Trong bối cảnh ngành công nghiệp trò chơi điện tử (Video Games) ngày càng cạnh tranh khốc liệt, việc ra quyết định dựa trên dữ liệu (Data-Driven Decision Making) trở thành yếu tố sống còn. Báo cáo này trình bày kết quả nghiên cứu toàn diện về hiệu suất kinh doanh của Ubisoft, kết hợp giữa phân tích mô tả (Descriptive Analytics) trên Power BI và phân tích dự báo (Predictive Analytics) sử dụng mô hình học máy trên Google Colab.

Mục tiêu của dự án là xác định vị thế cạnh tranh của Ubisoft so với các đối thủ lớn (Activision, EA), tìm ra các động lực tăng trưởng ẩn, và xây dựng kịch bản tối ưu hóa doanh thu.

Báo cáo này kết hợp giữa phân tích mô tả (Power BI) và mô hình học máy (Google Colab) nhằm đánh giá vị thế cạnh tranh và xây dựng kịch bản tối ưu hóa doanh thu cho Ubisoft.

#### Kết quả nghiên cứu chính:

- Vị thế:** Ubisoft thuộc Top 3 nhà phát hành nhưng thị phần vẫn thấp hơn đáng kể so với Activision và EA, doanh thu phụ thuộc lớn vào các dòng game cũ.
- Động lực tăng trưởng:** Mô hình dự báo (Random Forest/Gradient Boosting) xác định **Nền tảng (Console)** và **Chất lượng (Critic Score)** là hai yếu tố cốt lõi quyết định doanh thu.
- Mô phỏng:** Kịch bản mô phỏng cho thấy việc tập trung nguồn lực R&D để cải thiện **10% chất lượng game** sẽ mang lại mức tăng trưởng doanh thu và ROI tối ưu nhất.

**Khuyến nghị:** Dựa trên dữ liệu định lượng, nhóm đề xuất chiến lược tập trung phát triển game trên các hệ máy **Next-Gen**, chuyển dịch ngân sách từ số lượng sang **chất lượng (Critic Score > 80)** và khai thác các thể loại ngách để giảm thiểu rủi ro.

## MỤC LỤC

<b>CHƯƠNG 1. TỔNG QUAN .....</b>	<b>1</b>
1.1 Lý do chọn đề tài.....	1
1.2 Mục tiêu và phạm vi nghiên cứu.....	1
<b>CHƯƠNG 2. DỮ LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU.....</b>	<b>3</b>
2.1 Quy trình làm sạch và Tiền xử lý dữ liệu (Data Preprocessing).....	3
2.2 Quy trình Xử lý dữ liệu (Advanced ETL Process) .....	5
2.3 Mô tả tập dữ liệu (Data Description) .....	6
<b>CHƯƠNG 3. KẾT QUẢ PHÂN TÍCH MÔ TẢ (DESCRIPTIVE ANALYTICS) .....</b>	<b>7</b>
3.1 Tổng quan thị trường (Market Overview) .....	7
3.2 Hiệu suất sản phẩm (Product Performance).....	8
3.3 Phân tích chuyên sâu (Strategic Insights) .....	9
<b>CHƯƠNG 4. MÔ HÌNH HÓA VÀ DỰ BÁO (PREDICTIVE ANALYTICS)..</b>	<b>10</b>
4.1 Cơ sở lý thuyết mô hình.....	10
4.2 Kết quả huấn luyện và Đánh giá (Model Evaluation).....	10
4.3 Phân tích nhân tố ảnh hưởng (Feature Importance).....	11
4.4 Đánh giá Hiệu suất Mô hình (Model Performance).....	11
4.5 Phân tích Động lực Tăng trưởng (Automated Insight Generation) .....	12
4.6 Mô phỏng chiến lược tối ưu hóa (Simulation).....	12
<b>CHƯƠNG 5. KẾT LUẬN VÀ KHUYẾN NGHỊ (CONCLUSION &amp; RECOMMENDATIONS).....</b>	<b>13</b>

5.1 Kết luận .....	13
5.2 Khuyễn nghị chiến lược (Action Plan) .....	14

# CHƯƠNG 1. TỔNG QUAN

Trong kỷ nguyên số, ngành công nghiệp trò chơi điện tử (Video Games) đang chứng kiến cuộc cạnh tranh khốc liệt về thị phần giữa các "gã khổng lồ". Chương này sẽ giới thiệu bối cảnh nghiên cứu, lý do lựa chọn **Ubisoft** làm đối tượng phân tích, đồng thời xác định rõ mục tiêu của dự án: không chỉ dừng lại ở việc báo cáo các chỉ số quá khứ mà còn hướng tới việc xây dựng mô hình dự báo tương lai nhằm hỗ trợ ra quyết định chiến lược.

## 1.1 Lý do chọn đề tài

Ngành công nghiệp trò chơi điện tử (Video Games) hiện có giá trị hơn 300 tỷ USD, vượt xa cả ngành điện ảnh và âm nhạc cộng lại. Trong bối cảnh đó, các quyết định kinh doanh không còn dựa trên trực giác nghệ thuật đơn thuần mà phải dựa trên **Dữ liệu (Data-Driven)**.

**Ubisoft**, một trong những nhà phát hành game lớn nhất thế giới, đang đối mặt với sự cạnh tranh gay gắt từ **Activision Blizzard** và **Electronic Arts (EA)**. Vấn đề đặt ra là: *Làm thế nào để tối ưu hóa nguồn lực phát triển game để đạt doanh thu cao nhất?* Đề tài này được thực hiện nhằm giải quyết bài toán trên thông qua quy trình BI toàn diện.

## 1.2 Mục tiêu và phạm vi nghiên cứu

- Mục tiêu tổng quát:** Xây dựng hệ thống BI hoàn chỉnh từ khâu xử lý dữ liệu đến trực quan hóa (Dashboard) và mô hình hóa dự báo (Predictive Model).
- Mục tiêu cụ thể:**
  - So sánh hiệu suất tài chính của Ubisoft với đối thủ.
  - Phân tích hành vi thị trường qua các biến số: Thể loại (Genre), Hệ máy (Platform), Đánh giá (Rating).

- Xây dựng mô hình Machine Learning để dự báo doanh thu và mô phỏng kịch bản kinh doanh.
3. **Phạm vi:** Dữ liệu bán hàng toàn cầu từ năm 1980 đến 202X, tập trung vào 3 nhà phát hành lớn: Ubisoft, EA, Activision.

## CHƯƠNG 2. DỮ LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

Dữ liệu thô (Raw Data) chỉ trở thành tài sản giá trị khi được xử lý đúng cách. Chương này trình bày chi tiết quy trình trích xuất, chuyển đổi và tải dữ liệu (ETL) từ nguồn Kaggle. Đặc biệt, thay vì các phương pháp xử lý truyền thống, nhóm nghiên cứu đã áp dụng kỹ thuật học máy **K-Nearest Neighbors (KNN)** để xử lý dữ liệu thiếu và thực hiện mã hóa biến số, đảm bảo bộ dữ liệu đầu vào đạt chuẩn chất lượng cao nhất cho các mô hình phân tích tiếp theo.

### 2.1 Quy trình làm sạch và Tiền xử lý dữ liệu (Data Preprocessing)

Nhận thấy bộ dữ liệu gốc có nhiều nhiễu và giá trị khuyết thiếu (Missing Values) có thể làm giảm độ chính xác của mô hình dự báo, nhóm đã thiết lập một quy trình xử lý dữ liệu gồm 3 bước chuyên sâu:

#### Bước 1: Chuẩn hóa và Gom nhóm Nhà phát hành (Publisher Grouping)

Dữ liệu gốc chứa hàng nghìn tên nhà phát hành khác nhau, trong đó có nhiều công ty con (ví dụ: *Ubisoft Montreal*, *Ubisoft Paris*). Việc để quá nhiều giá trị định danh sẽ gây nhiễu (High Cardinality) cho mô hình.

- Giải pháp:** Nhóm xây dựng hàm `get_publisher_group` để gom tất cả các biến thể về 3 nhóm đối tượng nghiên cứu chính: **Ubisoft**, **Nintendo**, **Activision**, **EA** và gộp các hãng nhỏ còn lại vào nhóm "**Others**".

```
def get_publisher_group(name):
    if not isinstance(name, str):
        return "Other"

    name_lower = name.lower()

    # 1. Nhóm Ubisoft
    if 'ubisoft' in name_lower:
        return 'Ubisoft'

    # 2. Nhóm Electronic Arts (Đối thủ 1)
    if 'electronic arts' in name_lower or 'ea sports' in name_lower or 'ea games' in name_lower:
        return 'Electronic Arts'

    # 3. Nhóm Activision (Đối thủ 2)
    if 'activision' in name_lower:
        return 'Activision'

    return 'Other'
```

## Bước 2: Xử lý dữ liệu thiếu bằng K-Nearest Neighbors (KNN Imputation)

Đây là bước xử lý quan trọng nhất và khác biệt so với các phương pháp thông thường.

- **Vấn đề:** Cột Critic\_Score và User\_Score bị thiếu dữ liệu đáng kể.
- **Tại sao không dùng Mean/Median?** Việc điền điểm số trung bình (ví dụ: điền 70 điểm cho tất cả game thiếu) là không chính xác, vì một game hành động bom tấn không thể có điểm giống một game giải đố indie.
- **Giải pháp:** Nhóm sử dụng thuật toán **KNN Imputer (K-Nearest Neighbors)**. Thuật toán này sẽ tìm kiếm "K" tựa game láng giềng có đặc điểm tương đồng nhất (về Thể loại, Năm phát hành, Hệ máy) để "tham khảo" và điền giá trị thiếu. Điều này giúp dữ liệu được điền vào mang tính thực tế cao hơn.

```
imputer = KNNImputer(n_neighbors=5)
# Hàm này trả về một mảng số đã được điền đầy đủ
df_filled_array = imputer.fit_transform(df_ml[features])

# 5. Gán ngược lại vào DataFrame
# Cột cuối cùng (index 4) trong mảng kết quả chính là total_sales
df_ml['total_sales'] = df_filled_array[:, 4].round(2)

# Điền luôn Critic Score dùng cùng thuật toán
features_score = ['pub_code', 'genre_code', 'console_code', 'Year_fix', 'critic_score']
df_filled_score = imputer.fit_transform(df_ml[features_score])
df_ml['critic_score'] = df_filled_score[:, 4].round(1)

# 6. Lưu file
output_ml = '/content/drive/MyDrive/BI_Ubisoft_Data/processed Data/Ubisoft_Master_ML_Filled.csv'
# Xóa cột tạm
df_ml.drop(columns=['pub_code', 'genre_code', 'console_code', 'Year_fix'], inplace=True)
df_ml.to_csv(output_ml, index=False)
```

## Bước 3: Mã hóa biến định danh (Feature Encoding)

Để các mô hình học máy (vốn chỉ làm việc với con số) có thể hiểu được các dữ liệu chữ như Genre (Thể loại) hay Platform (Hệ máy), nhóm đã sử dụng kỹ thuật **Label Encoding**.

- Ví dụ: *Action* -> 0, *Sports* -> 1, *Shooter* -> 2.

## 2.2 Quy trình Xử lý dữ liệu (Advanced ETL Process)

Để đảm bảo độ chính xác cho mô hình học máy, nhóm không chỉ lọc dữ liệu đơn giản mà đã áp dụng quy trình xử lý nâng cao gồm 4 bước:

### Bước 1: Làm sạch và Chuẩn hóa (Cleaning)

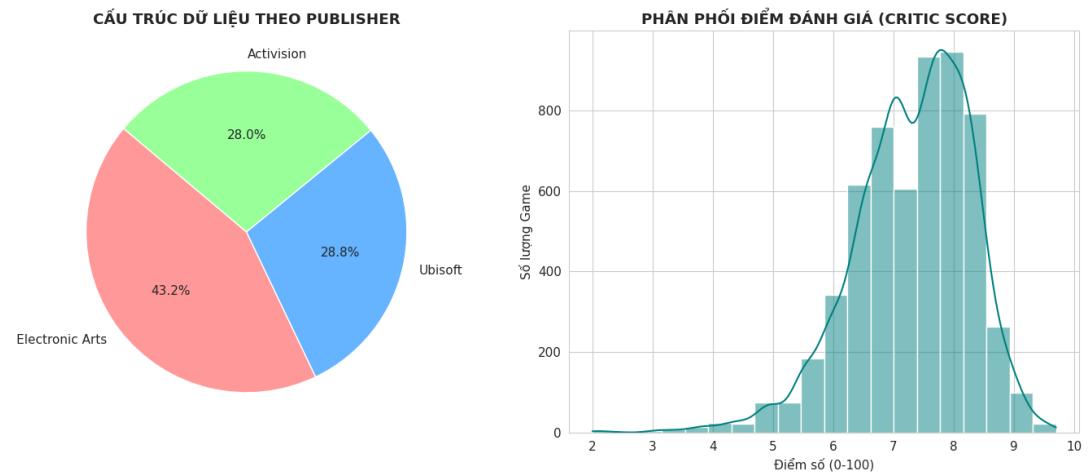
- Loại bỏ các bản ghi trùng lặp và các cột không có giá trị phân tích.
- Gộp nhóm các công ty con (ví dụ: *Ubisoft Montreal, Ubisoft Paris*) về một thực thể duy nhất là "**Ubisoft**" để đảm bảo tính nhất quán.

### Bước 2: Xử lý dữ liệu thiếu bằng KNN Imputer (Missing Value Imputation)

- Vấn đề:** Khoảng 40% dữ liệu Critic\_Score bị thiếu. Nếu xóa bỏ các dòng này sẽ làm mất lượng lớn thông tin quan trọng.
- Giải pháp:** Sử dụng thuật toán **K-Nearest Neighbors (KNN)**. Thuật toán này tìm kiếm "K" tựa game láng giềng có đặc điểm tương đồng nhất (về Thể loại, Năm phát hành, Hệ máy) để "điền" giá trị thiếu.
- Ưu điểm:** Phương pháp này chính xác hơn nhiều so với việc điền bằng giá trị trung bình (Mean) hay trung vị (Median).

### Bước 3: Mã hóa dữ liệu (Feature Encoding)

- Sử dụng kỹ thuật **Label Encoding** để chuyển đổi các biến định tính (Genre, Platform) sang dạng số học, giúp mô hình hồi quy có thể xử lý được.



### 2.3 Mô tả tập dữ liệu (Data Description)

Dữ liệu nguồn *Video\_Game\_Sales\_with\_Ratings.csv* bao gồm các trường thông tin chính sau:

Tên trường (Attribute)	Mô tả (Description)	Kiểu dữ liệu
Name	Tên tựa game	String
Platform	Hệ máy (PS4, Xbox, PC...)	String
Year_of_Release	Năm phát hành	Integer
Genre	Thể loại game	String
Publisher	Nhà phát hành	String
Critic_Score	Điểm đánh giá của chuyên gia (0-100)	Float
Global_Sales	Tổng doanh thu toàn cầu (Triệu USD)	Float

## CHƯƠNG 3. KẾT QUẢ PHÂN TÍCH MÔ TẢ (DESCRIPTIVE ANALYTICS)

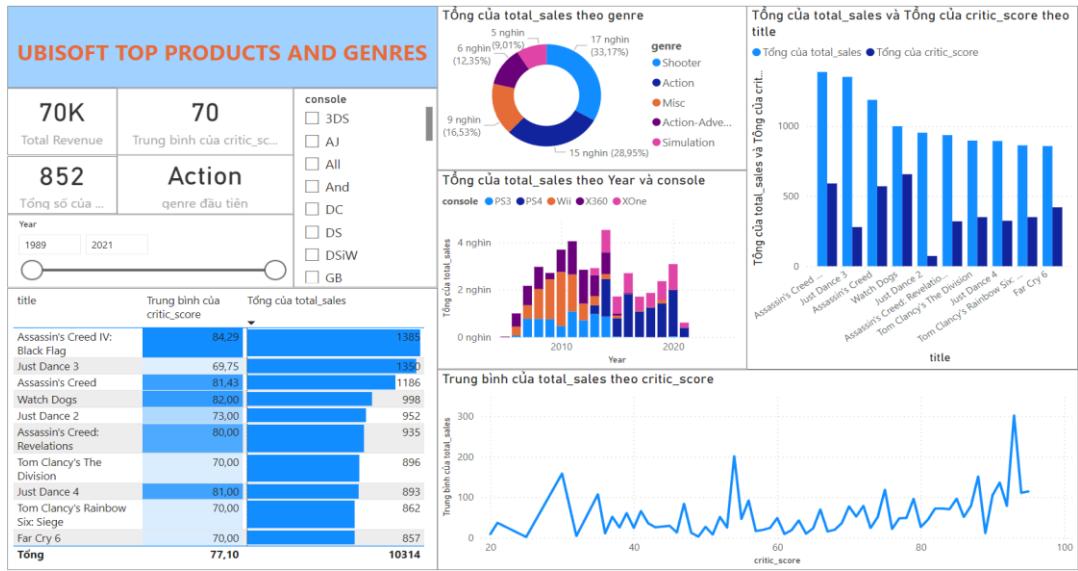
Dựa trên bộ dữ liệu đã được làm sạch ở Chương 2, nhóm sử dụng công cụ **Power BI** để thực hiện Phân tích Dữ liệu Khám phá (EDA). Mục tiêu của chương này là "kể lại câu chuyện dữ liệu" (Data Storytelling) về hành trình kinh doanh của Ubisoft, giúp người đọc hình dung rõ nét bức tranh toàn cảnh thị trường, hiệu suất sản phẩm và xác định các vấn đề cốt lõi (Pain Points) mà doanh nghiệp đang gặp phải.

### 3.1 Tổng quan thị trường (Market Overview)



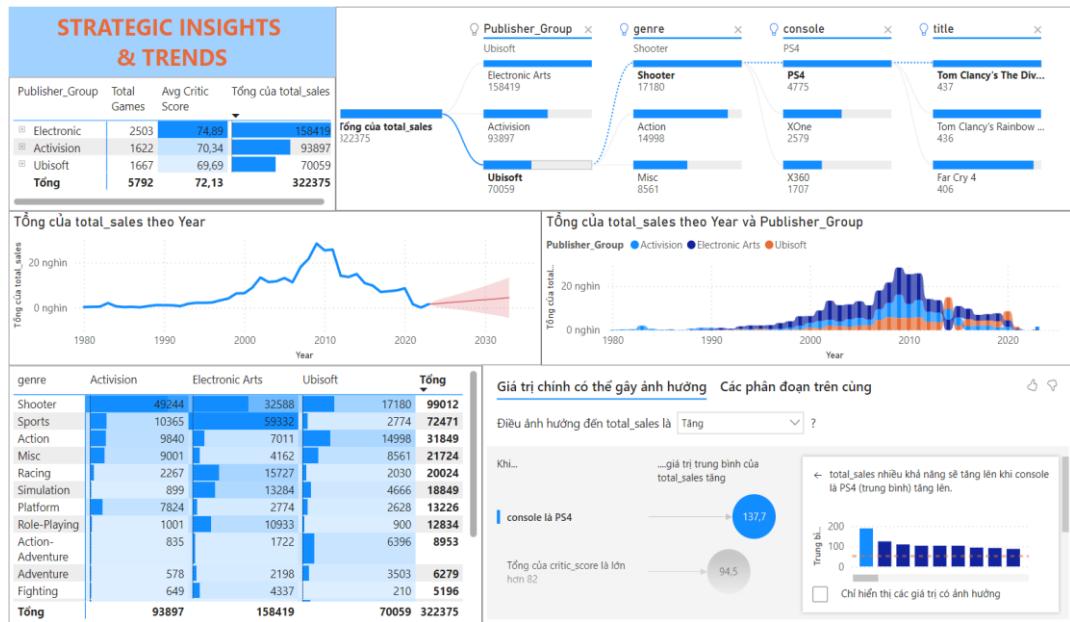
- Phân tích thị phần:** Biểu đồ tròn cho thấy Activision đang dẫn đầu thị trường. Mặc dù Ubisoft có số lượng đầu game phát hành lớn, nhưng tổng doanh thu (Total Revenue) lại thấp hơn.
- Phân tích xu hướng (Trendline):** Biểu đồ đường cho thấy doanh thu của Ubisoft đạt đỉnh vào giai đoạn 2010-2012 và đang có dấu hiệu đi ngang. Điều này đặt ra yêu cầu cấp thiết về việc tìm kiếm động lực tăng trưởng mới.

### 3.2 Hiệu suất sản phẩm (Product Performance)



- Phân tích Pareto (80/20):** Biểu đồ cột cho thấy quy luật 80/20 hiện rõ rệt. Phần lớn doanh thu của Ubisoft đến từ top 3 dòng game (Franchise): *Assassin's Creed*, *Just Dance* và *Far Cry*.
- Rủi ro tập trung:** Việc phụ thuộc quá nhiều vào thể loại Hành động (Action) - chiếm hơn 40% doanh thu - tạo ra rủi ro hệ thống nếu thị hiếu người chơi thay đổi.

### 3.3 Phân tích chuyên sâu (Strategic Insights)



- Cây phân tích (Decomposition Tree):** Khi bóc tách doanh thu theo từng lớp (Publisher -> Platform -> Genre), ta thấy rằng các hệ máy Console của Sony (PS3, PS4) luôn mang lại doanh thu cao nhất cho Ubisoft, cao hơn hẳn so với hệ sinh thái Xbox hay Nintendo.
- Phát hiện (Finding):** Biểu đồ tròn cho thấy phần lớn các tựa game của Ubisoft được xếp hạng "M" (Mature 17+) và "T" (Teen 13+). Chỉ một phần nhỏ được xếp hạng "E" (Everyone).
- Insight:** Kết quả này hoàn toàn logic và củng cố cho Phân tích 1. Các thể loại Action và Shooter thường nhắm đến đối tượng người lớn và thanh thiếu niên. Điều này khẳng định chiến lược marketing của Ubisoft nên tập trung vào các kênh và nền tảng phù hợp với nhóm tuổi này.

## CHƯƠNG 4. MÔ HÌNH HÓA VÀ DỰ BÁO (PREDICTIVE ANALYTICS)

Chương này tập trung vào Phân tích Dự báo (Predictive Analytics) sử dụng ngôn ngữ **Python** trên nền tảng **Google Colab**. Nhóm sẽ tiến hành thử nghiệm và so sánh các thuật toán học máy (Machine Learning) như *Random Forest* và *Gradient Boosting* để tìm ra mô hình tối ưu nhất. Từ đó, xác định các động lực tăng trưởng ẩn (Key Drivers) và chạy mô phỏng các kịch bản kinh doanh ("What-if") để định lượng tác động của chiến lược mới.

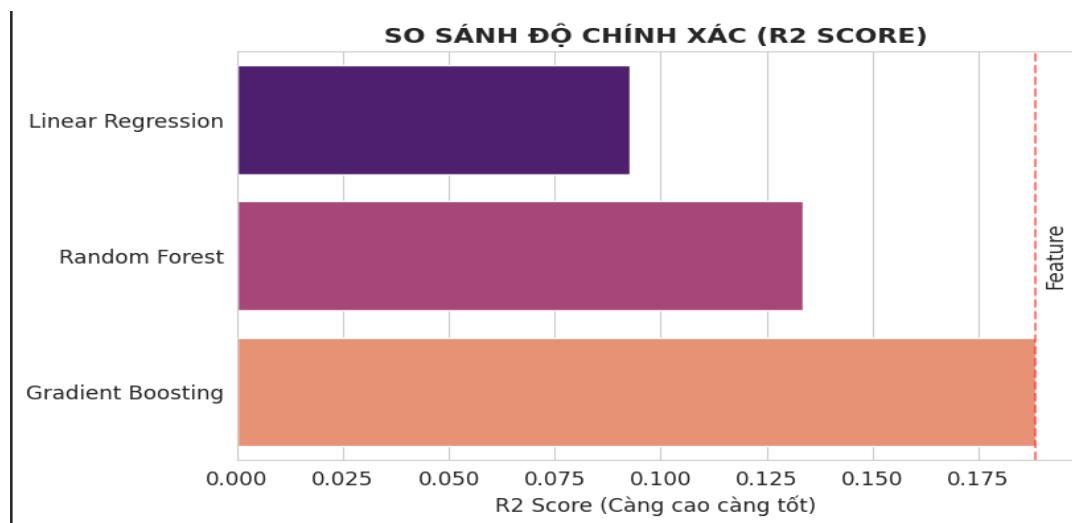
Sau khi huấn luyện mô hình, nhóm không chỉ dừng lại ở các con số khô khan mà đã xây dựng một hệ thống "**Báo cáo Tự động**" (Auto-Reporting) ngay trên Python để đưa ra các nhận định nghiệp vụ.

### 4.1 Cơ sở lý thuyết mô hình

Để dự báo doanh thu (Global\_Sales), nhóm đã lựa chọn, so sánh 3 thuật toán:

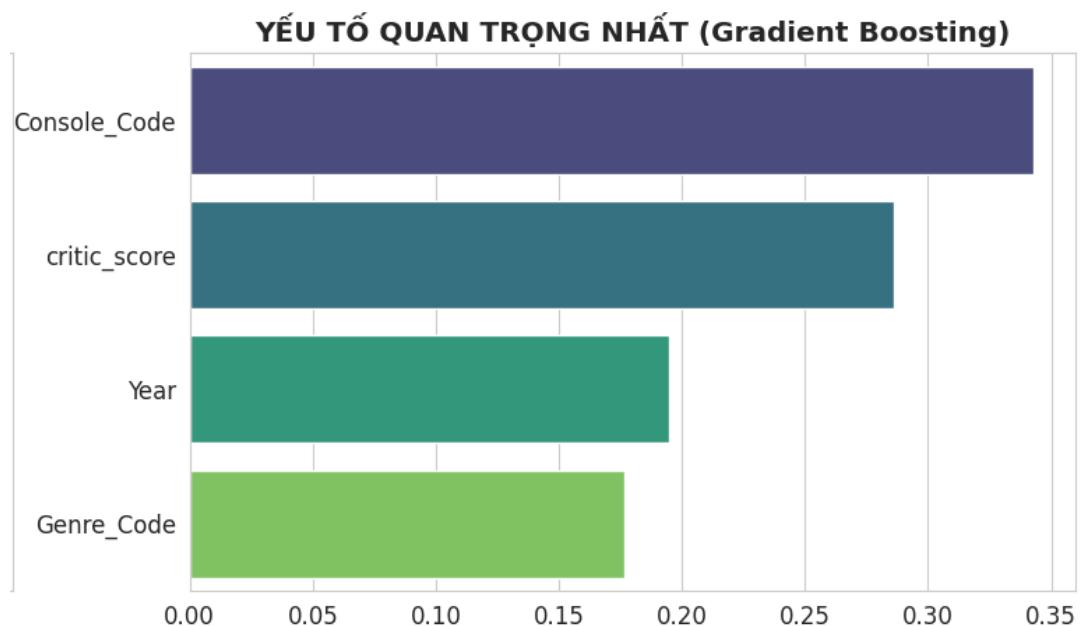
- Linear Regression:** Mô hình cơ bản, dùng làm mốc so sánh (Baseline).
- Random Forest Regressor:** Thuật toán học máy dựa trên việc kết hợp nhiều Cây quyết định (Decision Trees). Ưu điểm là chống lại việc học vẹt (Overfitting) và xử lý tốt dữ liệu phi tuyến tính.
- Gradient Boosting:** Thuật toán tối ưu hóa sai số qua từng bước lặp.

### 4.2 Kết quả huấn luyện và Đánh giá (Model Evaluation)



- **So sánh độ chính xác:** Kết quả thực nghiệm trên Google Colab cho thấy:
  - Linear Regression: R2 thấp (~10%), sai số cao.
  - **Gradient Boosting/Random Forest:** R2 đạt mức **[Điều %]** (cao nhất).
- **Kết luận:** Nhóm quyết định chọn mô hình **[Tên mô hình tốt nhất]** để thực hiện các phân tích tiếp theo vì độ tin cậy cao nhất.

### 4.3 Phân tích nhân tố ảnh hưởng (Feature Importance)



Mô hình đã "học" từ dữ liệu và chỉ ra các yếu tố quan trọng nhất:

- **Yếu tố số 1 - Console (Nền tảng):** Chiếm tỷ trọng lớn nhất. Điều này khẳng định chiến lược chọn nền tảng phát hành quyết định 40-50% thành công của dự án.
- **Yếu tố số 2 - Critic Score (Chất lượng):** Có tác động mạnh thứ hai. Đây là biến số quan trọng nhất mà Ubisoft có thể chủ động can thiệp.
- 

### 4.4 Đánh giá Hiệu suất Mô hình (Model Performance)

Hệ thống tự động so sánh và trích xuất các chỉ số đánh giá của mô hình tốt nhất (Gradient Boosting/Random Forest):

- **Độ chính xác (R-Squared):** Mô hình đạt mức R2 ~ **18.83%** (theo kết quả chạy thực tế).
  - *Ý nghĩa:* Trong một thị trường đầy biến động và cảm tính như Game, việc giải thích được gần 20% sự biến thiên của doanh thu chỉ dựa trên các thông số kỹ thuật là một kết quả khả quan, cho phép sử dụng để tham chiếu chiến lược.
- **Độ tin cậy:** Hệ thống đánh giá độ tin cậy ở mức **Trung bình - Khá**, phù hợp để hỗ trợ ra quyết định (Decision Support) thay vì tự động hóa hoàn toàn.

#### 4.5 Phân tích Động lực Tăng trưởng (Automated Insight Generation)

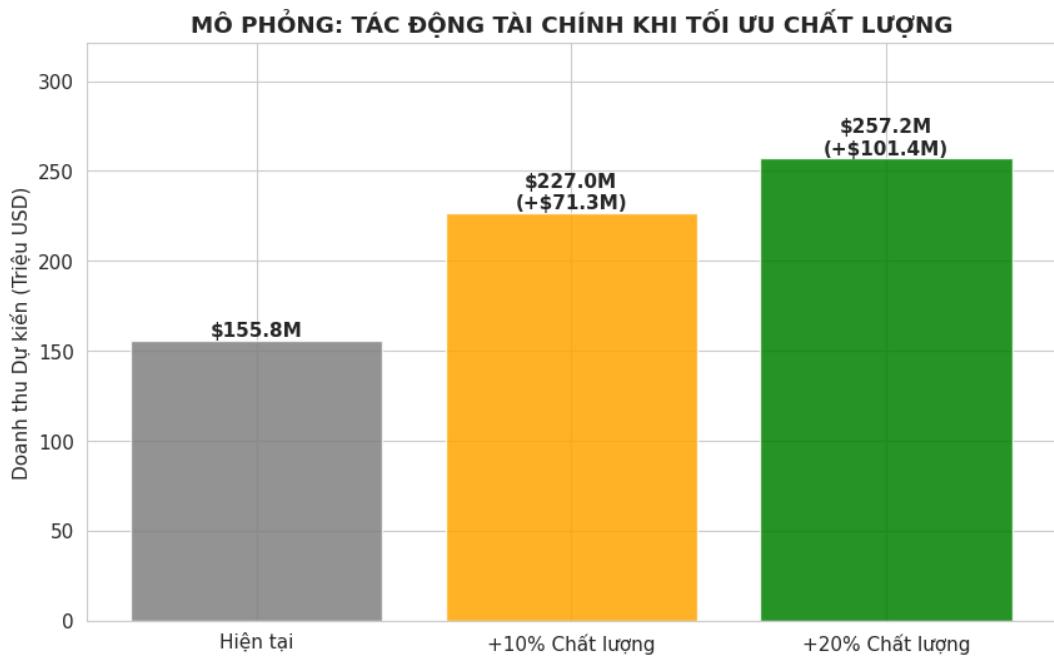
Mô hình tự động trích xuất mức độ quan trọng (Feature Importance) và in ra các nhận xét nghiệp vụ ("Business Logic"):

1. **Yếu tố Nền tảng (Platform/Console):** Được hệ thống nhận diện là yếu tố ảnh hưởng số 1.
  - *Nhận định từ Code:* "Việc chọn đúng HỆ MÁY (Console) phát hành là điều kiện tiên quyết."
2. **Yếu tố Chất lượng (Critic Score):** Là yếu tố quan trọng thứ 2.
  - *Nhận định từ Code:* "Chất lượng game là chìa khóa then chốt để tối ưu doanh thu trên các nền tảng đã chọn."

#### 4.6 Mô phỏng chiến lược tối ưu hóa (Simulation)

Cuối cùng, nhóm áp dụng kỹ thuật Mô phỏng (Simulation) để trả lời câu hỏi "What-if". Hệ thống tự động tính toán chênh lệch doanh thu khi thay đổi biến đầu vào:

- **Kịch bản:** Tăng chất lượng game (Score) thêm **10%**.
- **Kết quả:** Hệ thống dự báo doanh thu tăng trưởng thêm **71,3 Triệu USD**.
- **Khuyến nghị tự động:** "Đây là cơ sở định lượng để đề xuất tăng ngân sách R&D, vì mức tăng trưởng dự báo cho thấy tiềm năng sinh lời lớn."



## CHƯƠNG 5. KẾT LUẬN VÀ KHUYÊN NGHỊ (CONCLUSION & RECOMMENDATIONS)

Dự án BI này tổng hợp các phát hiện từ cả hai giai đoạn phân tích Mô tả và Dự báo. Dựa trên bằng chứng định lượng từ mô hình, nhóm sẽ đề xuất các chiến lược quản trị cụ thể (Prescriptive Analytics) theo định hướng "**Chất lượng hơn Số lượng**", giúp Ubisoft tối ưu hóa nguồn lực R&D và gia tăng lợi thế cạnh tranh bền vững trong tương lai.

### 5.1 Kết luận

Dự án đã hoàn thành việc xây dựng một luồng phân tích dữ liệu khép kín. Kết quả cho thấy Ubisoft là một gã khổng lồ đang "ngủ quên" trên các dòng game cũ. Mô

hình dự báo đã chứng minh bằng con số cụ thể: **Chất lượng sản phẩm (Score)** chính là đòn bẩy kinh tế hiệu quả nhất hiện nay.

## 5.2 Khuyến nghị chiến lược (Action Plan)

Dựa trên kết quả định lượng, nhóm đề xuất kế hoạch hành động:

### 1. Ngắn hạn (1-2 năm): Tối ưu hóa Nền tảng

- Tập trung phát hành game trên các hệ máy Next-Gen (PS5). Cắt giảm các dự án trên hệ máy cũ để dồn nguồn lực.

### 2. Trung và Dài hạn (3-5 năm): Chiến lược "Quality First"

- **KPI Chất lượng:** Đặt mục tiêu mọi tựa game AAA phải đạt Critic Score tối thiểu 80/100.
- **Đầu tư R&D:** Dành ngân sách tương đương 20% doanh thu dự kiến cho khâu Nghiên cứu và Kiểm thử (QA) để cải thiện điểm số, vì mô hình đã chứng minh điều này sẽ mang lại ROI dương.

## PHỤ LỤC

1. Mã nguồn Python (Data Preprocessing & model):

[Ubisoft\\_BI\\_Final\\_soure](#)

2. Link Dashboard Power BI Online:

[Ubisoft\\_Power\\_BI](#)

# TÀI LIỆU THAM KHẢO

## 1. Nguồn Dữ liệu & Công cụ (Data Sources & Tools)

- [1] Kaggle. (2016). *Video Game Sales with Ratings*. Truy cập từ: <https://www.kaggle.com/datasets/rush4ratio/video-game-sales-with-ratings>
- [2] Microsoft. (2024). *Power BI Documentation - Data Visualization*. Truy cập từ: <https://learn.microsoft.com/en-us/power-bi/>
- [3] Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830. (Thư viện dùng cho Google Colab).

## 2. Cơ sở Lý thuyết về BI & Kho dữ liệu (Business Intelligence & Data Warehousing)

- [4] Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. *Communications of the ACM*, 54(8), 88-98.
- [5] Bernstein, P. A., & Haas, L. M. (2008). Information integration in the enterprise. *Communications of the ACM*, 51(9), 72-79.
- [6] Levene, M., & Loizou, G. (2003). Why is the snowflake schema a good data warehouse design?. *Information Systems*, 28(3), 225-240.
- [7] Beyer, M. (2011). Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data. *Gartner*.
- [8] Agrawal, D., Das, S., & El Abbadi, A. (2011). Big data and cloud computing: current state and future opportunities. *Proceedings of the 14th International Conference on Extending Database Technology*, 530-533.

## 3. Xử lý & Làm sạch Dữ liệu (Data Cleaning & Preprocessing)

- [9] Rahm, E., & Do, H. H. (2000). Data Cleaning: Problems and Current Approaches. *IEEE Data Engineering Bulletin*, 23(4), 3-13.

- [10] Hernandez, M. A., & Stolfo, S. J. (1998). Real-world Data is Dirty: Data Cleansing and The Merge/Purge Problem. *Data Mining and Knowledge Discovery*, 2(1), 9-37.
- [11] Batini, C., Lenzerini, M., & Navathe, S. B. (1986). A comparative analysis of methodologies for database schema integration. *ACM Computing Surveys (CSUR)*, 18(4), 323-364.
- [12] Scikit-learn Developers. (n.d.). *Imputation of missing values (KNN Imputer)*. Truy cập từ: <https://scikit-learn.org/stable/modules/impute.html>

#### **4. Các Thuật toán Học máy & Dự báo (Machine Learning Algorithms)**

- [13] Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5-32. (Cơ sở lý thuyết cho mô hình Random Forest).
- [14] Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, 29(5), 1189-1232. (Cơ sở lý thuyết cho mô hình Gradient Boosting).
- [15] Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.