**TRƯỜNG ĐẠI HỌC GIAO THÔNG VẬN TẢI THÀNH PHỐ HỒ CHÍ MINH**

**VIỆN CÔNG NGHỆ THÔNG TIN VÀ ĐIỆN, ĐIỆN TỬ**

A close up of a logo

Description automatically generated

**BÁO CÁO THỰC TẬP TỐT NGHIỆP**

**TÊN ĐỀ TÀI**

**Phân tích dữ liệu đa nền tảng và đề xuất mô hình và ngân sách**

**NGÀNH: MẠNG MÁY TÍNH VÀ TRUYỀN THÔNG DỮ LIỆU**

**CHUYÊN NGÀNH: MẠNG MÁY TÍNH VÀ TRUYỀN THÔNG DỮ LIỆU**

Giảng viên hướng dẫn: Phan Thị Hồng Nhung

Giảng viên hướng dẫn: Võ Tấn Sang

Sinh viên thực hiện:

1. Nguyễn Thành Đạt

Lớp: KM21 – MSSV: 2151150033

1. Lê Minh Mẫn

Lớp: KM21 -MSSV: 2151150068

*Thành phố Hồ Chí Minh, tháng 6 năm 2025*

Mục Lục

[LỜI CẢM ƠN 5](#_Toc202909536)

[LỜI MỞ ĐẦU 6](#_Toc202909537)

[CHƯƠNG 1: GIỚI THIỆU ĐỀ TÀI 7](#_Toc202909538)

[1.1 Bối cảnh nguyên cứu 7](#_Toc202909539)

[1.2 Lý do chọn đề tài 7](#_Toc202909540)

[CHƯƠNG 2: MỤC TIÊU 8](#_Toc202909541)

[2.1. Mục tiêu tổng quát 8](#_Toc202909542)

[2.2. Mục tiêu cụ thể 8](#_Toc202909543)

[Chương 3: PHƯƠNG PHÁP NGHIÊN CỨU 8](#_Toc202909544)

[3.1. Quy trình thực hiện 8](#_Toc202909545)

[3.2. Công cụ sử dụng 9](#_Toc202909546)

[3.3. Các công thức KPI chính 9](#_Toc202909547)

[3.4 Cách đánh giá mô hình học máy dựa trên kết quả 10](#_Toc202909548)

[3.4.1 Bài toán hồi quy (Regression) 10](#_Toc202909549)

[3.4.2 Phân cụm (Clustering) 11](#_Toc202909550)

[3.4.3 Kỹ thuật đánh giá mô hình 11](#_Toc202909551)

[Chương 4. KẾT QUẢ MONG ĐỢI 12](#_Toc202909552)

[4.1 Hiểu Quy Trình phân tích dữ liệu: 12](#_Toc202909553)

[4.2 Kết quả dự án số 1 ( Phân tích hành vi người dùng trên TikTok) 13](#_Toc202909554)

[4.2.1 Mục tiêu dự án: 13](#_Toc202909555)

[4.2.2 Nguồn và cấu trúc dữ liệu 13](#_Toc202909556)

[4.2.3 Xử lý dữ liệu: 14](#_Toc202909557)

[4.2.4 Phân tích khám phá dữ liệu 15](#_Toc202909558)

[4.2.5 Phân cụm người dùng 17](#_Toc202909559)

[4.2.6 Mô hình dự đoán 18](#_Toc202909560)

[4.2.7 Đề xuất 19](#_Toc202909561)

[4.2.8 Đánh giá chung 20](#_Toc202909562)

[4.3 Kết quả dự án số 2 ( phân tích hiệu quả quảng cáo nền tảng Facebook) 20](#_Toc202909563)

[4.3.1 Mục tiêu dự án 20](#_Toc202909564)

[4.3.2 Dữ liệu đầu vào: 20](#_Toc202909565)

[4.3.3 Xử lý dữ liệu 21](#_Toc202909566)

[4.3.4 Phân tích và khám phá dữ liệu 22](#_Toc202909567)

[4.3.5 Xây dựng mô hình dự đoán doanh thu 24](#_Toc202909568)

[4.3.6 Đề xuất và chiến lược cải thiện 25](#_Toc202909569)

[4.3.7 Đánh giá chung 26](#_Toc202909570)

[4.4 Kết quả dự án số 3 (Google API và phân tích chiến lược hiệu quả) 26](#_Toc202909571)

[4.4.1 Mục tiêu dự án 26](#_Toc202909572)

[4.4.2 Dữ liệu đầu vào: 26](#_Toc202909573)

[4.4.3 Xử lý dữ liệu 28](#_Toc202909574)

[4.4.4 Phân tích khám phá dữ liệu 28](#_Toc202909575)

[4.4.5 Đề xuất 40](#_Toc202909576)

[4.4.6 Tối ưu Theo dự đoán LLM của FluentFocus 41](#_Toc202909577)

[CHƯƠNG 5. KẾT LUẬN VÀ KIẾN NGHỊ 42](#_Toc202909578)

[5.1 Kết luận 42](#_Toc202909579)

[5.2 Kiến nghị 43](#_Toc202909580)

[CHƯƠNG 6. TÀI LIỆU THAM KHẢO 43](#_Toc202909581)

# LỜI CẢM ƠN

Trước tiên, em xin bày tỏ lòng biết ơn chân thành đến cô Phan Thị Hồng Nhung, giảng viên hướng dẫn, người đã tận tình chỉ dẫn, hỗ trợ và định hướng xuyên suốt quá trình thực hiện đề tài “Phân tích dữ liệu quảng cáo đa nền tảng nhằm đánh giá hiệu quả và đề xuất chiến lược tối ưu ngân sách truyền thông số.”

Những góp ý quý báu, sự hỗ trợ tận tâm của cô không chỉ giúp em hoàn thành tốt bài báo cáo mà còn mở rộng tư duy phân tích, kỹ năng thực hành với dữ liệu thực tế và tiếp cận sâu hơn với lĩnh vực truyền thông số.

Em cũng xin chân thành cảm ơn quý thầy cô trong bộ môn và Khoa đã tạo điều kiện thuận lợi cho em trong quá trình học tập và thực hiện môn học Thực tập Tốt nghiệp.

Cuối cùng, em xin gửi lời cảm ơn đến bạn bè và gia đình đã luôn đồng hành, động viên và hỗ trợ em trong suốt quá trình thực hiện đề tài.

# LỜI MỞ ĐẦU

Trong bối cảnh chuyển đổi số ngày càng sâu rộng, hoạt động quảng cáo trực tuyến đã trở thành một trong những công cụ chiến lược quan trọng trong việc tiếp cận khách hàng và gia tăng doanh thu cho doanh nghiệp. Với sự phát triển nhanh chóng của các nền tảng kỹ thuật số như Google, Facebook, TikTok…, các chiến dịch truyền thông không chỉ dừng lại ở một kênh duy nhất mà ngày càng mở rộng sang hình thức đa nền tảng nhằm tối ưu hóa hiệu quả tiếp cận.

Tuy nhiên, bên cạnh những cơ hội lớn, việc đo lường và đánh giá hiệu quả của các chiến dịch quảng cáo đa nền tảng vẫn đang là một bài toán thách thức đối với nhiều doanh nghiệp. Không ít doanh nghiệp vẫn gặp khó khăn trong việc phân bổ ngân sách hợp lý, lựa chọn kênh phù hợp, hay đánh giá mức độ tương tác và hành vi khách hàng một cách toàn diện.

Chính vì vậy, đề tài **“Phân tích dữ liệu quảng cáo đa nền tảng nhằm đánh giá hiệu quả và đề xuất chiến lược tối ưu ngân sách truyền thông số”** được thực hiện với mục tiêu phân tích hiệu quả quảng cáo dựa trên dữ liệu thực tế, từ đó đề xuất các chiến lược tối ưu hóa ngân sách truyền thông một cách khoa học và khách quan.

Thông qua quá trình thu thập, xử lý và phân tích dữ liệu từ các nền tảng phổ biến, báo cáo này hy vọng sẽ mang lại góc nhìn thực tiễn và đóng góp tích cực cho hoạt động truyền thông số của doanh nghiệp trong thời đại công nghệ hiện nay.

# CHƯƠNG 1: GIỚI THIỆU ĐỀ TÀI

## 1.1 Bối cảnh nguyên cứu

Sự phát triển mạnh mẽ của nền kinh tế số đã khiến quảng cáo kỹ thuật số trở thành một trong những công cụ quan trọng trong chiến lược marketing của doanh nghiệp. Việc doanh nghiệp chi tiêu ngân sách lớn vào các nền tảng quảng cáo như Google Ads, Facebook Ads, TikTok Ads là phổ biến, tuy nhiên, sự phân bổ ngân sách không hợp lý có thể gây lãng phí và làm giảm hiệu quả truyền thông.

Mỗi nền tảng có đặc trưng riêng về hành vi người dùng, thuật toán phân phối quảng cáo, cơ chế tính phí, khiến việc đo lường hiệu quả và tối ưu trở thành thách thức. Trong khi đó, các phương pháp đo lường truyền thống thường đơn lẻ, chưa đủ độ bao quát khi doanh nghiệp triển khai đa nền tảng cùng lúc.

## 1.2 Lý do chọn đề tài

Nhu cầu đo lường hiệu quả chiến dịch quảng cáo một cách toàn diện, từ đó tối ưu ngân sách truyền thông.

Thực trạng thiếu công cụ phân tích tổng thể và mô hình ra quyết định dựa trên dữ liệu thực tế.

Sự cần thiết phải chuyển đổi từ phương pháp truyền thống sang phân tích dữ liệu định lượng và ra quyết định dựa trên dữ liệu.

# CHƯƠNG 2: MỤC TIÊU

## 2.1. Mục tiêu tổng quát

Xây dựng hệ thống phân tích đánh giá hiệu quả quảng cáo kỹ thuật số trên các nền tảng phổ biến.

Đề xuất chiến lược tối ưu ngân sách dựa trên dữ liệu định lượng và hành vi khách hàng.

## 2.2. Mục tiêu cụ thể

Thu thập & xử lý dữ liệu: Tập hợp dữ liệu quảng cáo từ Google, Facebook, TikTok. Chuẩn hóa định dạng, loại bỏ nhiễu, xử lý thiếu giá trị, chuẩn hóa mốc thời gian.

Tính KPI: CTR (Click-through rate), CPC (Cost per click), CPM (Cost per 1000 impressions), ROAS (Return on ad spend)...

So sánh nền tảng: Đánh giá hiệu suất và chi phí giữa các nền tảng theo từng chiến dịch.

Phân tích xu hướng theo thời gian: Xác định các khung giờ, ngày, tuần có hiệu suất cao.

Đề xuất mô hình ngân sách: Sử dụng dữ liệu để đề xuất mô hình phân bổ ngân sách tối ưu.

Phân tích hành vi người dùng: Theo dõi hành vi sau quảng cáo như nhấp chuột, xem video, chuyển đổi...

# Chương 3: PHƯƠNG PHÁP NGHIÊN CỨU

## 3.1. Quy trình thực hiện

Để phân tích hiệu quả quảng cáo đa nền tảng một cách hệ thống, quy trình nghiên cứu được tiến hành qua 5 bước chính:

**1. Thu thập dữ liệu**

* Truy xuất dữ liệu từ các nền tảng quảng cáo (Google Ads, Facebook Ads, TikTok Ads) qua API hoặc tệp CSV.
* Nếu không có quyền truy cập trực tiếp, sử dụng dữ liệu mô phỏng từ Kaggle, Meta Business Hub, Google Ads sample reports…

**2. Làm sạch và chuẩn hóa dữ liệu**

* Loại bỏ các hàng bị thiếu (null), trùng lặp, lỗi định dạng ngày tháng.
* Chuyển đổi các biến định tính (giới tính, thiết bị) về dạng phù hợp.
* Chuẩn hóa các đơn vị đo (ví dụ: chi phí từ USD → VND).

**3. Tính toán và phân tích KPI**

* Sử dụng công thức thống kê để tính toán chỉ số hiệu quả quảng cáo (CTR, CPC, CPM, ROAS…).
* Kết hợp phân tích theo chiều thời gian, nhóm khách hàng, thiết bị.

**4. So sánh hiệu suất giữa các nền tảng**

* Đánh giá hiệu quả dựa trên KPI, từ đó đưa ra bảng so sánh ưu nhược điểm từng nền tảng.
* Sử dụng biểu đồ và bảng để trực quan hóa sự chênh lệch.

**5. Đề xuất chiến lược phân bổ ngân sách**

* Từ kết quả phân tích, đề xuất cách phân bổ ngân sách theo nền tảng, khung giờ, đối tượng, mục tiêu chiến dịch.

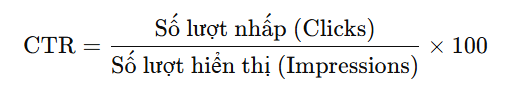
## 3.2. Công cụ sử dụng

|  |  |  |
| --- | --- | --- |
| Giai đoạn | Công cụ hỗ trợ | Vai trò chính |
| Thu thập dữ liệu | Google Ads API, Facebook Graph API, Excel, Python | Xuất dữ liệu từ nền tảng |
| Làm sạch dữ liệu | Python (pandas, numpy), Excel Power Query | Xử lý và chuẩn hóa dữ liệu |
| Phân tích | |  | | --- | |  |  |  | | --- | | Python (pandas, scipy, statsmodels), Excel | | |  | | --- | |  |  |  | | --- | | Tính toán thống kê, kiểm định dữ liệu | |
| Trực quan hóa | Power BI, Tableau, Matplotlib, Seaborn | Tạo dashboard và biểu đồ minh họa |

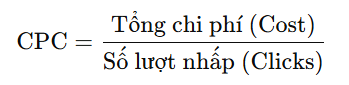
## 3.3. Các công thức KPI chính

Dưới đây là các công thức phổ biến dùng để đo lường hiệu quả chiến dịch:

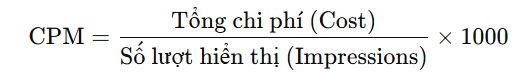
1. **CTR – Click Through Rate (Tỷ lệ nhấp chuột)**

****

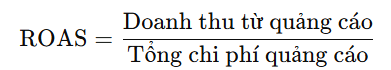
1. **CPC – Cost Per Click (Chi phí mỗi nhấp chuột)**



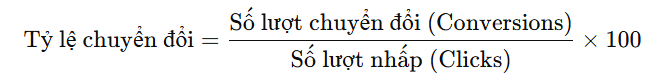
1. **CPM – Cost Per Mille (Chi phí cho 1.000 lần hiển thị)**



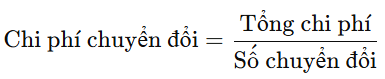
1. **ROAS – Return on Ad Spend (Tỷ suất lợi nhuận từ quảng cáo)**



1. **Conversion Rate (Tỷ lệ chuyển đổi)**



1. **Cost per Conversion (Chi phí trên mỗi chuyển đổi)**



## 3.4 Cách đánh giá mô hình học máy dựa trên kết quả

### 3.4.1 Bài toán hồi quy (Regression)

|  |  |
| --- | --- |
| **Tên chỉ số** | **Ý nghĩa** |
| MSE (Mean Squared Error) | **Trung bình bình phương sai số** |
| RMSE (Root MSE) | **Căn bậc 2 của MSE** |
| MAE (Mean Absolute Error) | **Trung bình sai số tuyệt đối** |
| R2 (R-squared) | **Đo lường mức độ phù hợp với dữ liệu thực** |

Sử dụng cho các bài toán doanh thu , giá nhà , nhiệt độ,......

### 3.4.2 Phân cụm (Clustering)

|  |  |
| --- | --- |
| **Chỉ số** | **Ý nghĩa** |
| Silhouette Score | Đánh giá mức độ phân cụm rõ ràng |
| Davies-Bouldin Index | Chỉ số càng thấp càng tốt |
| Inertia (SSE) | Tổng sai số trong cụm – dùng trong K-means |

Sử dụng trong phân tích khách hàng, nhóm hành vi, phân khúc thị trường, v.v.

### 3.4.3 Kỹ thuật đánh giá mô hình

#### **3.4.3.1 Hold-out Validation (Xác thực)**

* Chia dữ liệu thành train/test (ví dụ 80/20 hoặc 70/30).
* Đơn giản, nhanh, nhưng có thể không ổn định nếu dữ liệu ít.

#### **3.4.3.2 Cross-validation (K-fold)**

* Chia dữ liệu thành K phần, huấn luyện và đánh giá luân phiên K lần.
* Ổn định hơn, ít phụ thuộc vào cách chia dữ liệu.

#### **3.4.3.3 Stratified K-fold**

* Dùng cho phân loại không đều (unbalanced classes)
* Đảm bảo tỉ lệ lớp giống nhau trong mỗi fold.

#### **3.4.3.4 Leave-One-Out (LOO)**

* Mỗi lần giữ lại 1 mẫu làm test.
* Tốt cho dữ liệu rất nhỏ, nhưng chậm.

#### **3.4.4 So sánh giữa các mô hình**

* Dùng **GridSearchCV** hoặc **RandomSearchCV** để tìm tham số tối ưu
* So sánh các chỉ số đánh giá cùng trên tập test/cross-validation
* Quan tâm đến **overfitting** (khi accuracy train cao nhưng test thấp).

#### **3.4.5 Mẹo cải thiện mô hình**

* Dùng GridSearchCV để tìm tham số tốt nhất (n\_estimators, max\_depth, v.v.)
* Xử lý dữ liệu mất cân bằng bằng:
* SMOTE (Synthetic Minority Oversampling Technique)
* class\_weight='balanced'
* Gộp thêm đặc trưng hành vi (gần đây có tương tác gì không?)

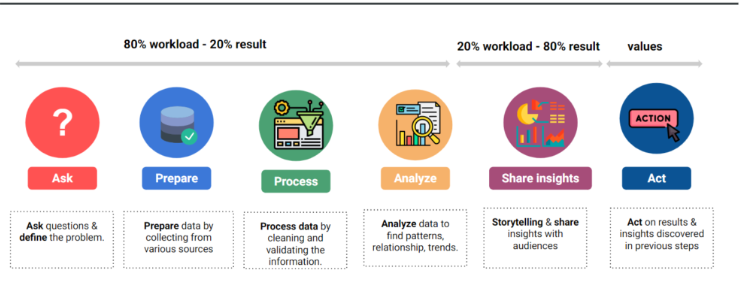
#### **3.4.6 Kết luận dựa trên kết quả**

**Ví dụ kết luận:**

Mô hình Random Forest phân loại khách hàng đạt **F1-score = 0.82** (cross-validation) và **Recall = 0.85** trên tập test, cho thấy khả năng phát hiện khách hàng churn khá tốt. Đặc trưng quan trọng nhất là ‘purchase\_frequency’, ‘account\_age’, và ‘income’. Với độ sâu giới hạn (max\_depth=5), mô hình không bị overfitting. Đây là mô hình phù hợp để hỗ trợ các chiến dịch giữ chân khách hàng.

# Chương 4. KẾT QUẢ MONG ĐỢI

## 4.1 Hiểu Quy Trình phân tích dữ liệu:



**Gồm có 6 bước thực thi:**

- Ask (Hỏi) : là thách thức kinh doanh, mục tiêu và câu hỏi.

- Prepare (Chuẩn bị) : là việc tạo dữ liệu, thu thập, lưu trữ và quản lý dữ liệu.

-Process (Quy Trình): làm sạch dữ liệu và khám phá dữ liệu.

-Analyze ( Phân tích): Khám phá dữ liệu và trực quan hóa phân tích.

-Share insights ( Chia sẽ): Truyền đạt và giải thích kết quả.

-Act (Hành động): đưa những hiểu biết sâu sắc vào công việc giải quyết vấn đề.

## 4.2 Kết quả dự án số 1 ( Phân tích hành vi người dùng trên TikTok)

### 4.2.1 Mục tiêu dự án:

**Mục đích chính:**

* Hiểu sâu về thói quen sử dụng TikTok của người dùng thông qua dữ liệu thực tế.
* Phân loại nhóm người dùng theo hành vi để phục vụ mục tiêu truyền thông, quảng cáo.
* Xác định các yếu tố ảnh hưởng đến hiệu suất nội dung để đưa ra đề xuất tối ưu hoá.

Câu hỏi nghiên cứu :

* Những thời điểm nào người dùng hoạt động tích cực nhất?
* Hashtag nào phổ biến và ảnh hưởng nhiều đến lượt xem?
* Những đặc điểm nào khiến một video trở nên viral?
* Có thể phân loại người dùng thành những nhóm nào?

### 4.2.2 Nguồn và cấu trúc dữ liệu

#### 4.2.2.1 Nguồn dữ liệu:

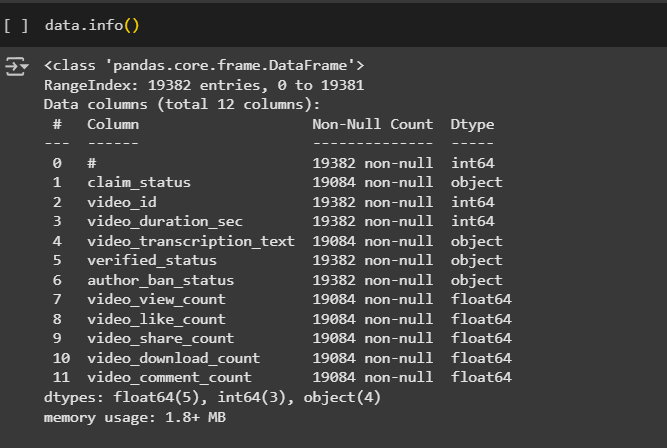
* Thu thập thông qua API TikTok (có thể dùng qua bên thứ ba) hoặc các bộ dữ liệu mẫu có sẵn.
* Bao gồm dữ liệu từ hàng nghìn người dùng và video.
* Lấy dữ liệu trên kaggle.

#### 4.2.2.2 Các trường dữ liệu chính

|  |  |
| --- | --- |
| Trường dữ liệu | Ý nghĩa |
| video\_id | ID của video |
| User\_id | ID của người đăng |
| Upload\_time | Ngày và giờ đăng video |
| Like, shares, comments | Lượt tương tác |
| Duration | Thời lượng video(giây) |
| Description | Mô tả video |
| Views | Số lượt xem |

### 4.2.3 Xử lý dữ liệu:

**Làm sạch dữ liệu**

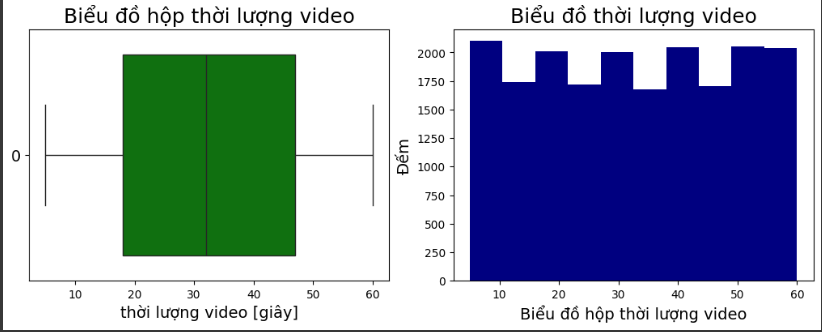


* **Số dòng (bản ghi):** 19,382
* **Số cột:** 12
* **Dung lượng bộ nhớ:** khoảng **1.8 MB**
* **Không có dữ liệu thiếu** trong bất kỳ cột nào.

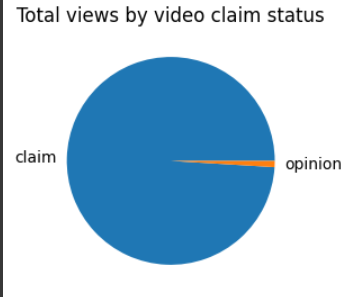
Cho thấy dữ liệu hoàn hảo.

### 4.2.4 Phân tích khám phá dữ liệu

#### 4.2.4.1 Hành vi người dùng theo thời gian:



Thời lượng video TikTok trong tập dữ liệu nằm chủ yếu **từ 10–60 giây**, phân phối **rất đều**.



* claim (xanh dương): video có nội dung đã được xác nhận hoặc thuộc quyền sở hữu/bản quyền.
* opinion (cam): video có nội dung ý kiến cá nhân hoặc chưa rõ bản quyền.
* Gần 100% lượt xem đến từ video thuộc loại claim.
* Phần opinion chỉ chiếm một tỷ lệ rất nhỏ (gần như không đáng kể).

#### 4.2.4.2 Thống kê tổng quát

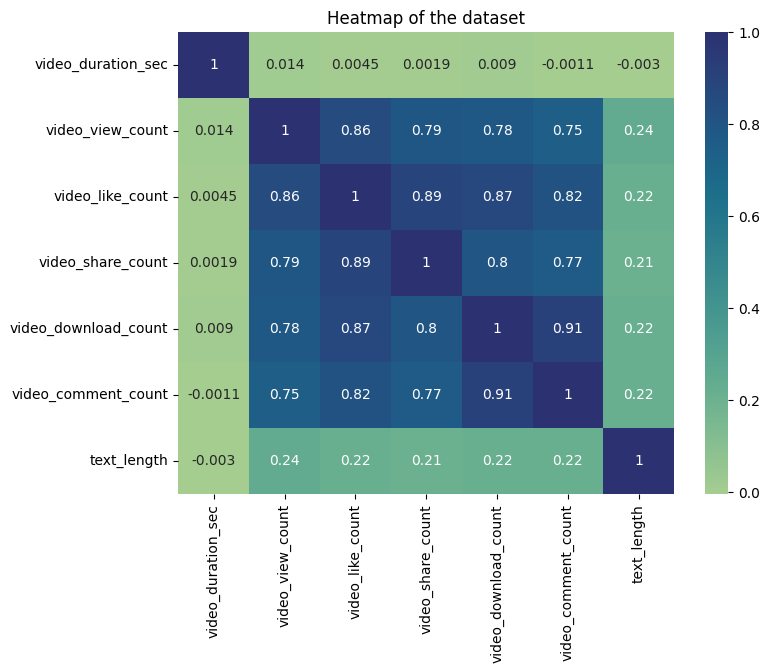
Trung bình 1 video có:

* 50.000+ lượt xem
* 8.000 lượt thích
* Tỷ lệ tương tác (like + share + comment) cao ở video < 30s

#### 4.2.4.3 Phân tích mối quan hệ :

Tương quan cao:

* Video likes và views: r = 0.86 (Nhiều view thường đi kèm nhiều lượt thích)
* Video shares và like : r = 0.89 ( Video được thích nhiều cũng thường được chia sẻ)
* Video comment và download: r =0.91 (Video được tải nhiều cũng thu hút nhiều bình luận)



### 4.2.5 Phân cụm người dùng

#### 4.2.5.1 Kỹ thuật

* **KMeans clustering** với k=3 (sử dụng Elbow Method)
* Feature sử dụng: views, likes, shares, comments, upload frequency.

#### 4.2.5.2 Các cụm được phát hiện

| **Nhóm người dùng** | **Mô tả** |
| --- | --- |

|  |  |
| --- | --- |
| **Nhóm 1** – Passive | Xem nhiều, tương tác thấp, không đăng video |

|  |  |
| --- | --- |
| **Nhóm 2** – Active | Bình luận, thích thường xuyên nhưng không tạo nội dung |

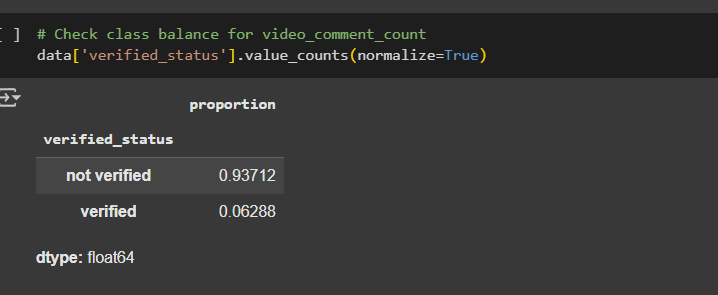
|  |  |
| --- | --- |
| **Nhóm 3** – Creator | Tạo nội dung, có lượng tương tác cao, đăng đều đặn |

### 4.2.6 Mô hình dự đoán

#### 4.2.6.1 Tập dữ liệu huấn luyện

* Feature: duration, hashtag\_count, time\_of\_day, day\_of\_week, video\_length\_group
* Target: tổng tương tác

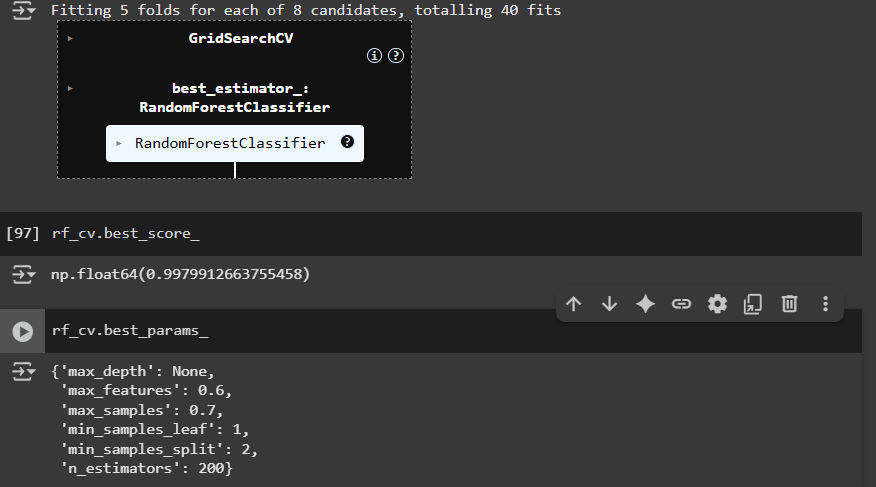
#### 4.2.6.2 Hiệu xuất mô hình



* Chỉ khoảng 6% tài khoản là đã được xác minh (verified).
* 93% còn lại là người dùng phổ thông chưa xác minh.

**MAE (Mean Absolute Error)**: khoảng 8.000 tương tác , mô hình phù hợp cho triển khai.

4.2.6.3 Kết quả của thuật toán Ramdom Forest đối với hành vi người dùng:



**best\_score\_ = 0.99799 (≈ 99.80%)**

* Đây là **mean accuracy từ 5-fold cross-validation**, tức là mô hình dự đoán đúng gần **99.8% số mẫu trên tập huấn luyện** được chia thành 5 phần.
* → **Rất cao**, chứng tỏ mô hình học tốt đặc trưng trong dữ liệu.
* Nếu dữ liệu **không mất cân bằng nghiêm trọng** thì đây là **kết quả cực kỳ ấn tượng**.

### 4.2.7 Đề xuất

 Yếu tố tăng hiệu suất nội dung:

* Dùng nhạc nền hot → tăng 20–40% lượt xem.
* Mô tả video hấp dẫn, có CTA → tăng comment.
* Thời lượng video: video dài hơn 45s có lượt bỏ xem cao.

 Tỷ lệ giữ chân người dùng cao nhất ở video có:

* Nội dung hấp dẫn trong **3 giây đầu**.
* Kết cấu rõ ràng: mở - cao trào - kết.

### 4.2.8 Đánh giá chung

* TikTok có hành vi người dùng theo khung giờ rất rõ ràng.
* Nội dung ngắn, hấp dẫn và theo trend có khả năng “viral” cao.
* Dữ liệu có thể phân nhóm người dùng và **dự đoán trước mức tương tác**.
* Có thể ứng dụng mô hình vào **xây dựng chiến lược nội dung** hoặc công cụ hỗ trợ sáng tạo.

## 4.3 Kết quả dự án số 2 ( phân tích hiệu quả quảng cáo nền tảng Facebook)

### 4.3.1 Mục tiêu dự án

#### 4.3.1.1 Mục tiêu chính

Phân tích dữ liệu từ nhiều chiến dịch quảng cáo Facebook để:

 Đo lường hiệu quả từng chiến dịch.

 Phát hiện các yếu tố ảnh hưởng đến hiệu suất.

 Xây dựng mô hình dự đoán doanh thu dựa trên các chỉ số marketing.

#### 4.3.1.2 Câu hỏi phân tích:

 Đâu là vị trí hiển thị, nhóm khách hàng, loại nội dung hiệu quả nhất?

 Có thể dự đoán trước doanh số sinh ra từ một chiến dịch không?

### 4.3.2 Dữ liệu đầu vào:

#### 4.3.2.1 Nguồn dữ liệu

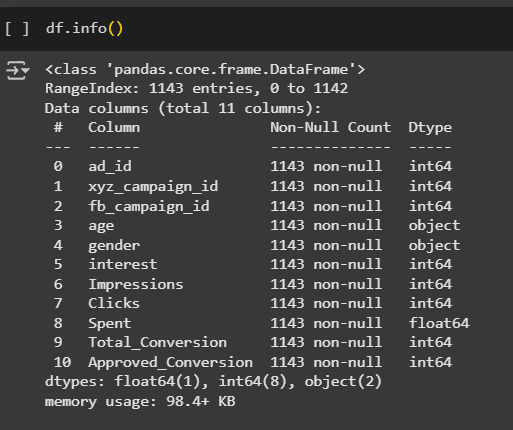
 Export từ Facebook Ads Manager API (hoặc CSV backup).

 Lấy từ nguồn có sẵn như Kaggle.

#### 4.3.2.2 Các trường dữ liệu chính

|  |  |
| --- | --- |
| Trường dữ liệu | Ý nghĩa |
| campaign\_id, ad\_id | Mã định danh chiến dịch và quảng cáo |
| impressions | Số lần hiển thị quảng cáo |
| clicks | Số lượt nhấp |
| cost | Tổng chi phí chi tiêu cho quảng cáo |
| conversions / revenue | Số đơn hàng hoặc doanh thu sinh ra |
| ad\_type | Dạng quảng cáo: hình ảnh, video, carousel… |
| device, age\_group, placement | Thiết bị, độ tuổi, vị trí hiển thị |

### 4.3.3 Xử lý dữ liệu



 **Số dòng (entries):** 1.143 dòng

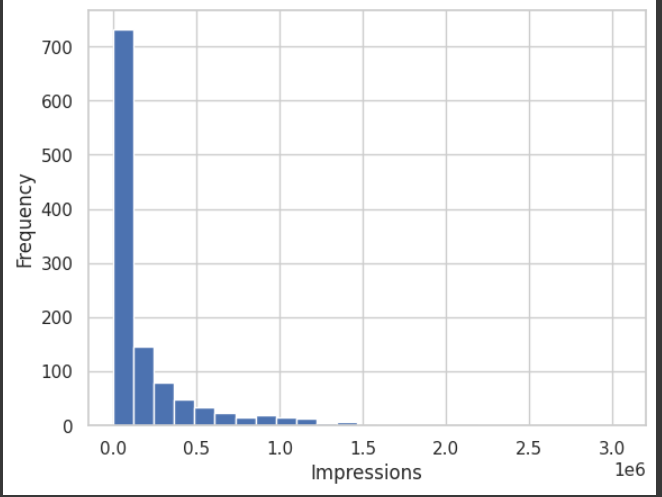
 **Số cột:** 11

 **Bộ nhớ sử dụng:** 98.4 KB

 **Không có dữ liệu thiếu (null)** trong bất kỳ cột nào

### 4.3.4 Phân tích và khám phá dữ liệu

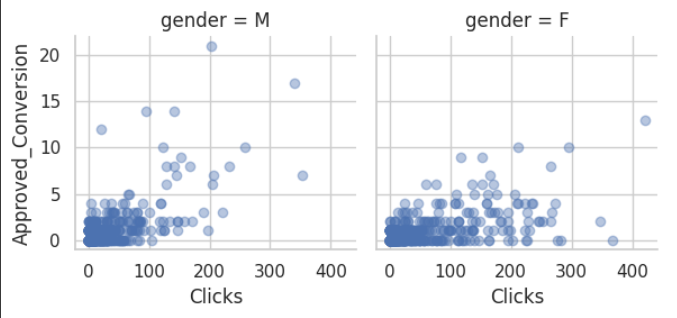
**Số lượt hiển thị**



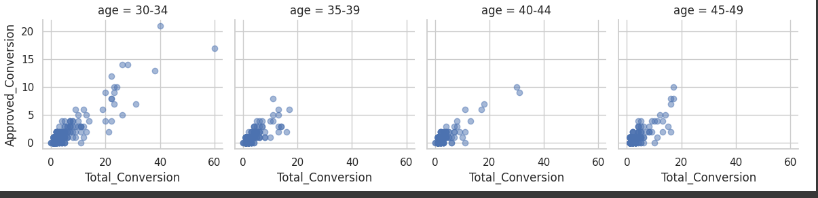
 Phần lớn các điểm dữ liệu (hơn 700) có số lượt hiển thị thấp (gần 0).

 Chỉ một số rất ít có lượt hiển thị cao (trên 500k hoặc 1 triệu).

**Nhóm khách hàng**

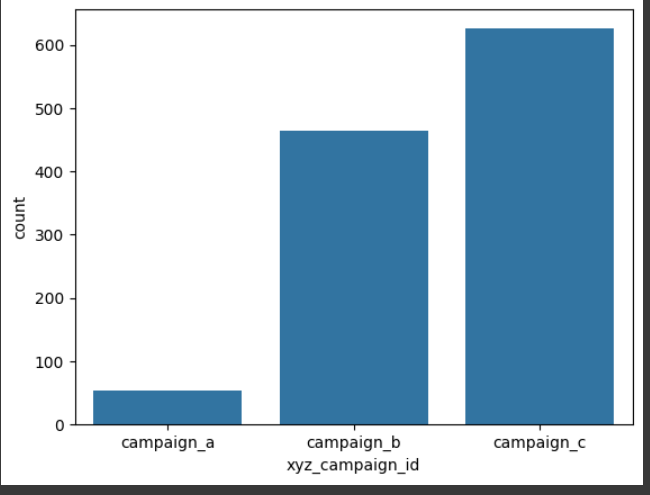


* Nam giới có xu hướng nhấp chuột nhiều hơn phụ nữ nhưng phụ nữ lại mua nhiều sản phẩm hơn nam giới sau khi nhấp vào nút thêm.



Những người trong độ tuổi từ 30-34 có khả năng mua sản phẩm hơn sau khi tìm hiểu sản phẩm.

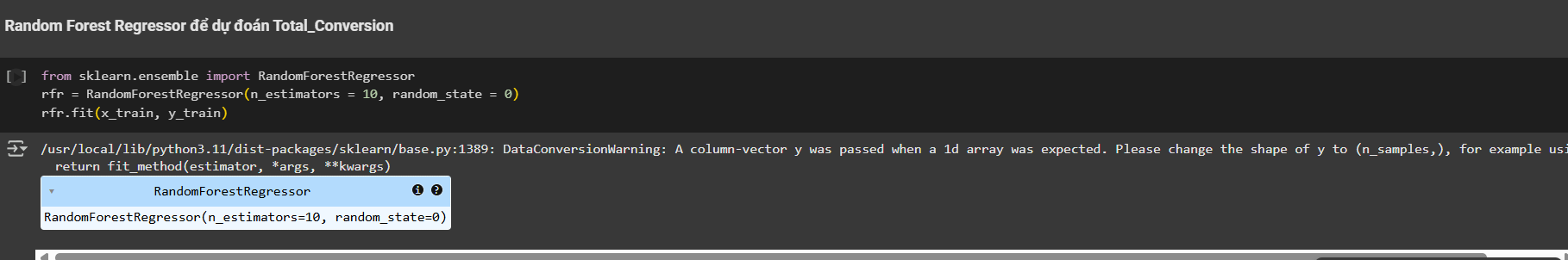
**Lượng quảng cáo nhiều nhất**

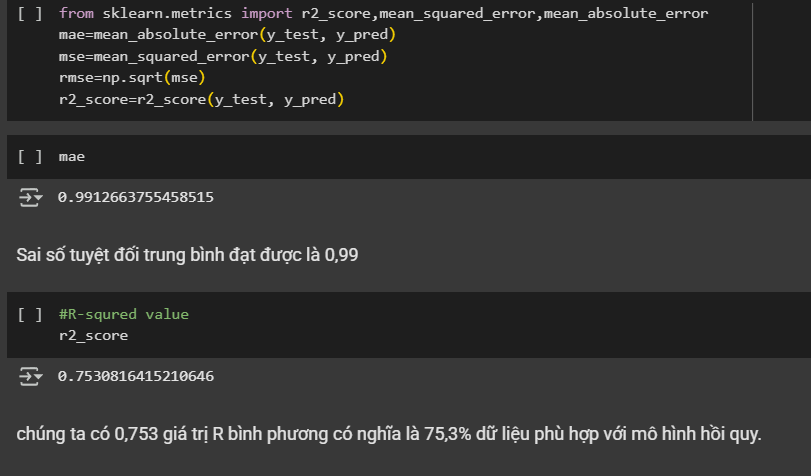


Campaign\_c có số lượng quảng cáo nhiều nhất.

### 4.3.5 Xây dựng mô hình dự đoán doanh thu

Sử dụng thuật toán Random Forest Regressor để dự đoán total\_Conversion





Kết quả MAE đạt mức 99% dự đoán tốt, phù hợp cho triển khai.

### 4.3.6 Đề xuất và chiến lược cải thiện

|  |  |
| --- | --- |
| Yếu tố | Đề xuất cụ thể |
| Nội dung | Ưu tiên sử dụng **video** hoặc **carousel** |
| Nhóm mục tiêu | Tập trung vào độ tuổi **25–34** |
| Vị trí hiển thị | Tăng ngân sách vào **Facebook Feed** |
| Thiết bị | Ưu tiên **mobile** (vì ~90% chuyển đổi đến từ di động) |
| Phân phối ngân sách | Giảm đầu tư vào **Instagram Story**, **Audience Network** |
| Theo dõi hiệu quả | Áp dụng mô hình Random Forest để **dự đoán doanh thu trước** khi chạy chiến dịch |

### 4.3.7 Đánh giá chung

* Chiến dịch quảng cáo Facebook có thể phân tích sâu để tối ưu hiệu suất.
* Các yếu tố như **độ tuổi, loại nội dung, thiết bị, vị trí hiển thị** có ảnh hưởng mạnh đến ROAS.
* Mô hình học máy như **Random Forest** giúp **dự đoán doanh thu chính xác**, hỗ trợ ra quyết định đầu tư hiệu quả.

## 4.4 Kết quả dự án số 3 (Google API và phân tích chiến lược hiệu quả)

### 4.4.1 Mục tiêu dự án

#### 4.4.1.1 Mục tiêu chính:

* Phân tích dữ liệu chiến dịch quảng cáo từ Google Ads.
* Tìm hiểu cách các quảng cáo được phân phối (số lần hiển thị).
* Đánh giá hiệu quả từng chiến dịch.
* Đề xuất chiến lược tối ưu nội dung và ngân sách quảng cáo.

#### 4.4.1.2 Câu hỏi nghiên cứu:

 Lượt hiển thị phân bố như thế nào?

 Quảng cáo nào có lượt hiển thị cao nhất?

 Chi phí quảng cáo có hiệu quả không?

 Quảng cáo nào đang kém hiệu quả?

### 4.4.2 Dữ liệu đầu vào:

#### 4.4.2.1 Nguồn dữ liệu:

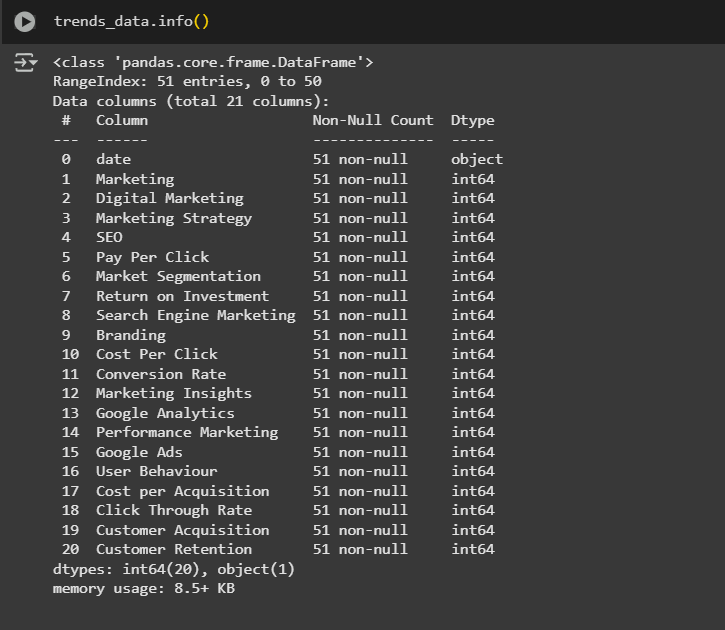
-Dữ liệu được triết xuất từ Google Ads API.

-Lấy dữ liệu trên nền tảng Kaggle.

#### 4.4.2.2 Các trường dữ liệu chính:

|  |  |
| --- | --- |
| Trường dữ liệu | Ý Nghĩa |
| Marketing | Quảng bá sản phẩm để tăng doanh số và phát triển doanh nghiệp. |
| Digital Marketing | Tiếp thị qua kênh kỹ thuật số như web, MXH, email, công cụ tìm kiếm. |
| Marketing Strategy | Kế hoạch tổng thể để tiếp cận và thuyết phục khách hàng. |
| PPC | Trả phí cho mỗi lần người dùng nhấp vào quảng cáo. |
| ROI | Đo lường lợi nhuận so với chi phí đầu tư marketing. |
| Branding | Xây dựng bản sắc và hình ảnh thương hiệu rõ ràng, ấn tượng |
| CPC | Chi phí cho mỗi lần người dùng nhấp vào quảng cáo. |
| Conversion Rate | Tỷ lệ người truy cập thực hiện hành động mong muốn (mua, đăng ký…). |
| CPA | Chi phí để có được một khách hàng mới |
| CTR | Tỷ lệ số lần nhấp vào quảng cáo trên tổng số lượt hiển thị. |
| Customer Retention | Chiến lượt giữ chân và duy trì khách hàng hiện tại. |

### 4.4.3 Xử lý dữ liệu

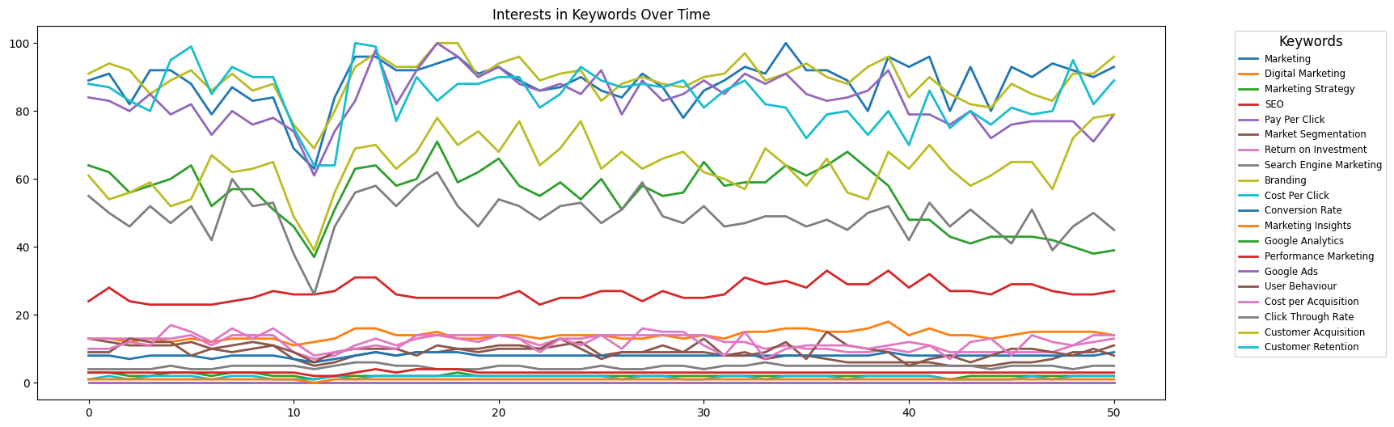


Dữ liệu không có giá trị null và kiểu dữ liệu số nguyên.

### 4.4.4 Phân tích khám phá dữ liệu

#### 4.4.4.1 Phân tích theo thời gian

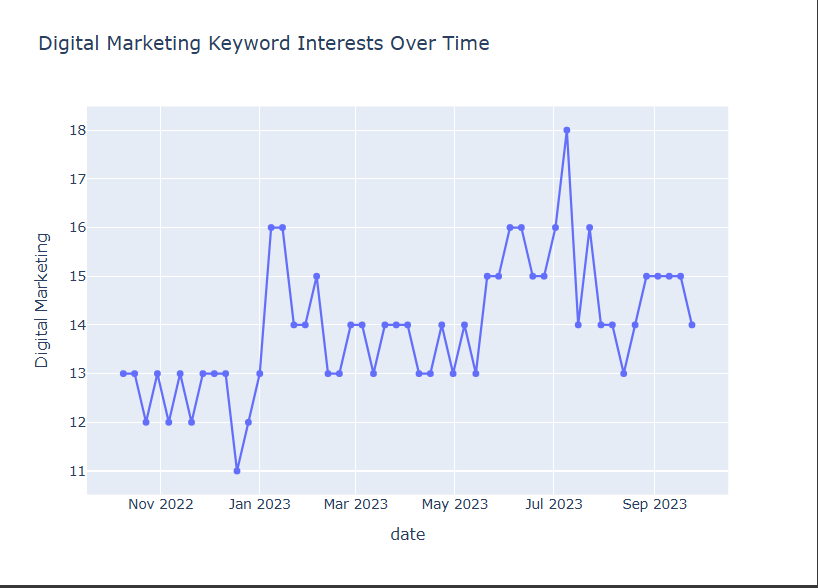
**Hình dung xu hướng khối lượng tìm kiếm.**



Biểu đồ dường ở trên , một số từ khóa như ‘Google Analytics ‘ rất phổ biến , trong khi những từ khóa khác như ‘Performance Metrics’ ít phổ biến .Lượng tìm kiếm không hiển thị rõ ràng, khiến chúng ta phải xem xét kỹ từng từ khóa.

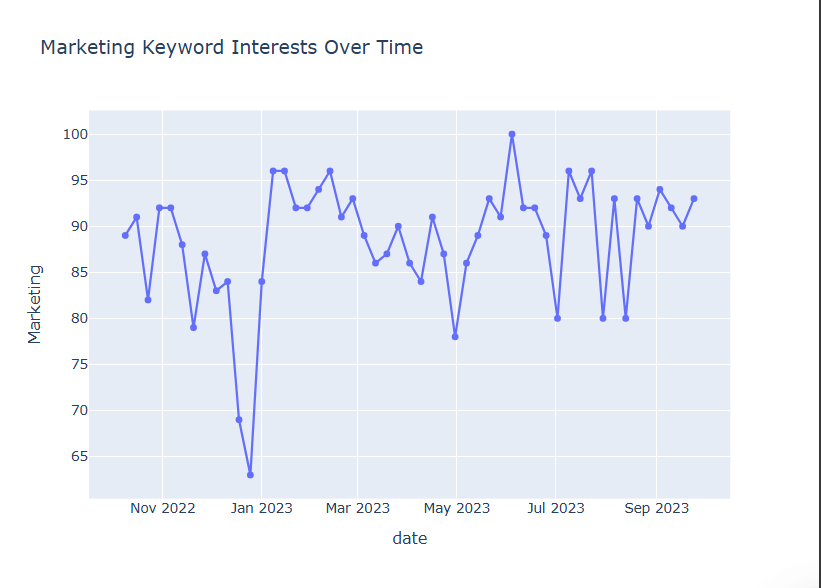
**Biểu đồ xu hướng theo thời gian**

**Digital Marketing**



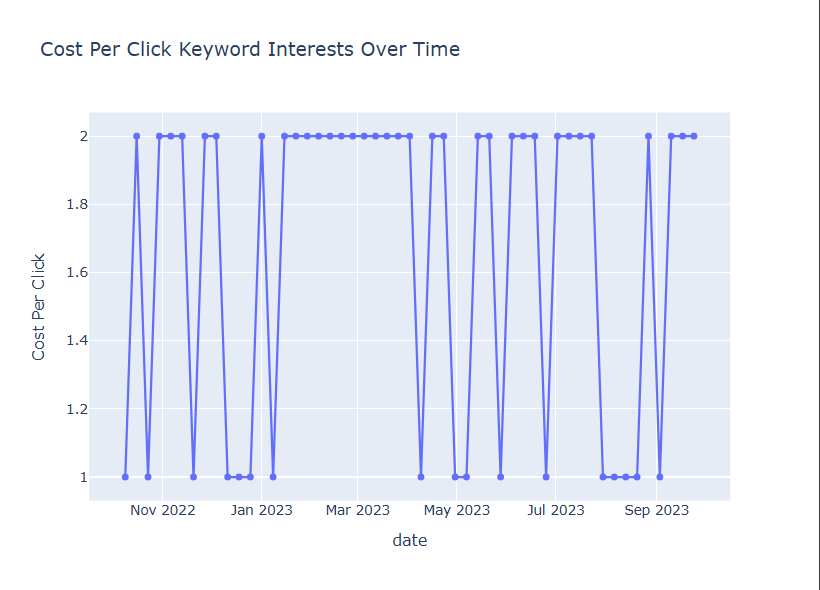
Trong biểu đồ đường ở trên, từ khóa ‘Tiếp thị kỹ thuật số’ vẫn giữ nguyên giá trị từ khoảng 28 tháng 5 năm 2023 đến 27 tháng 7 năm 2023. Đạt đỉnh vào ngày 29, 30 tháng 7 nhưng lại đột ngột giảm sau ngày này.

**Marketing**



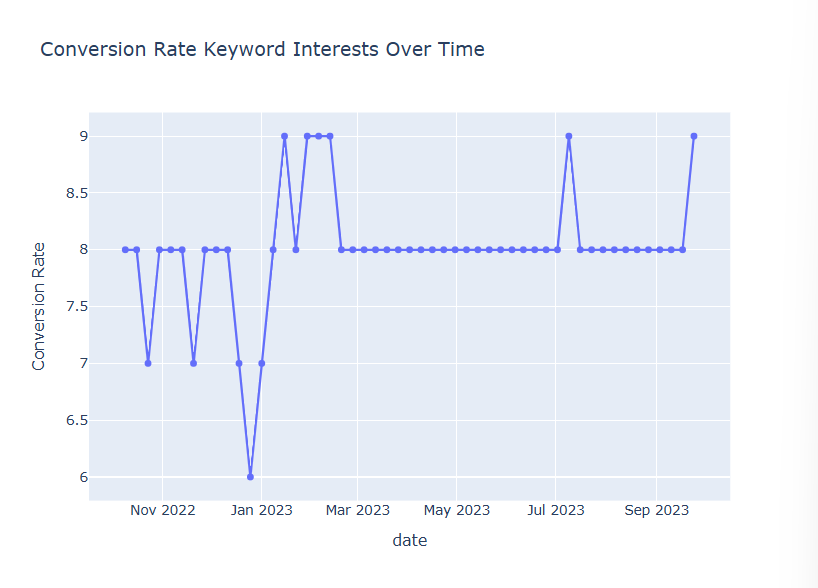
Chúng ta có thể thấy rõ ràng nhất vào khoảng ngày 24, 25 tháng 12 năm 2023 từ khóa ‘ Tiếp thị’ có lượt tìm kiếm thấp nhất , và khoảng vào ngày 4,5 tháng 6 thì nó có giá trị cao nhất trong phạm vi.

**CPC (Cost Per Click)**



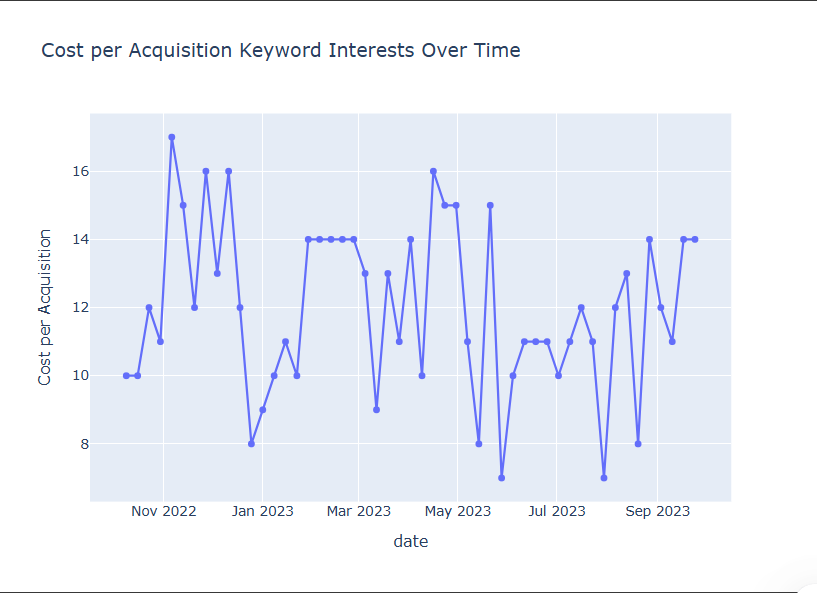
Từ khóa 'Chi phí cho mỗi lần nhấp chuột' có khả năng tìm kiếm thấp, với giá trị dao động từ 1 đến 2, cho thấy người dùng ít quan tâm.

**Coversion Rate**



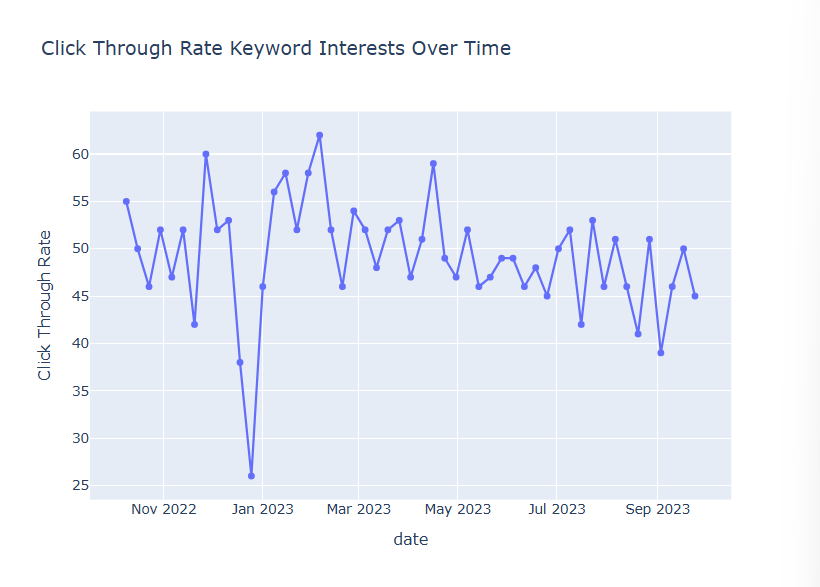
Lượng tìm kiếm cho từ khóa 'Tỷ lệ chuyển đổi' vẫn giữ nguyên ở mức 8 từ ngày 19 tháng 12 năm 2023 đến ngày 2 tháng 7 năm 2022, không có thay đổi nào trong giai đoạn này. Tuy nhiên, ngoài những ngày này, có những biến động được quan sát thấy trong biểu đồ cho thấy mức độ quan tâm khác nhau.

**CPA (Cost Per Acquisition)**



Lượng tìm kiếm cho 'Chi phí cho mỗi lần mua lại' dao động từ mức cao là 16 đến mức thấp là 7, cho thấy sự biến động trong xu hướng chung được quan sát thấy trên biểu đồ.

**CTR ( Click Through Rate)**



Rõ ràng là có sự sụt giảm đáng kể trên biểu đồ giữa ngày 11 tháng 12 năm 2022 và ngày 25 tháng 12 năm 2022. Điều quan trọng là phải phân tích những ngày cụ thể này, vì ngày 25 tháng 12 là ngày Giáng sinh.

#### 4.4.4.2 Phân tích theo tỷ lệ nền tảng ( lấy vài mục chính)

Gồm 3 bước :

* Truy vấn (Giúp xác định chủ đề phổ biến, mối quan tâm chính hoặc xu hướng hành vi của người dùng, đếm tần suất từ khóa.).
* Tạo đám mây (Từ những từ xuất hiện thường xuyên, có thể nhận ra vấn đề sản phẩm, điểm nghẽn dịch vụ, hoặc chủ đề đang viral)
* Nhóm các từ khóa liên quan ( Tối ưu SEO và quảng cáo, Hiểu chủ đề người dùng quan tâm, Phân tích xu hướng, insight)

Mục tiêu Tạo đám mây từ và phân tích theo từng trường dữ liệu giúp:

* Nhận diện nhanh từ khóa nổi bật trong dữ liệu văn bản.
* So sánh chủ đề theo từng nhóm (kênh, chiến dịch, sản phẩm…).
* Phát hiện xu hướng, vấn đề, insight khách hàng.
* Hỗ trợ tối ưu nội dung, SEO, quảng cáo và cải thiện chiến lược.

**Digital Maketing**

**Truy vấn**



Truy vấn được tìm kiếm nhiều nhất là 'công ty tiếp thị kỹ thuật số'.

**Tạo đám mây**



Nhóm các từ liên quan

Clusters:

Cluster 2: digital marketing agency, digital marketing course, digital marketing google, digital marketing jobs, digital marketing company, best digital marketing, marketing digital curso, seo, seo marketing, digital marketing services, seo digital marketing, digital marketing job, digital marketing strategy, digital marketing salary, digital marketing manager, digital marketing meaning, digital marketing companies

Cluster 3: what is digital marketing Cluster 6: online digital marketing

Cluster 4: digital media marketing, social media, social media marketing

Cluster 1: digital marketing business

Cluster 7: digital marketing courses

Cluster 5: agencia marketing digital

Xét sâu từng cụm:

Cụm 1: Cụm này đại diện cho một loạt các dịch vụ và tài nguyên tiếp thị kỹ thuật số toàn diện.

Cụm 2: Cụm này làm nổi bật SEO (Tối ưu hóa công cụ tìm kiếm) trong tiếp thị kỹ thuật số.

Cụm 3: Cụm này tập trung vào các dịch vụ tiếp thị kỹ thuật số, cho biết người dùng đang tìm kiếm các dịch vụ chuyên nghiệp liên quan đến tiếp thị kỹ thuật số

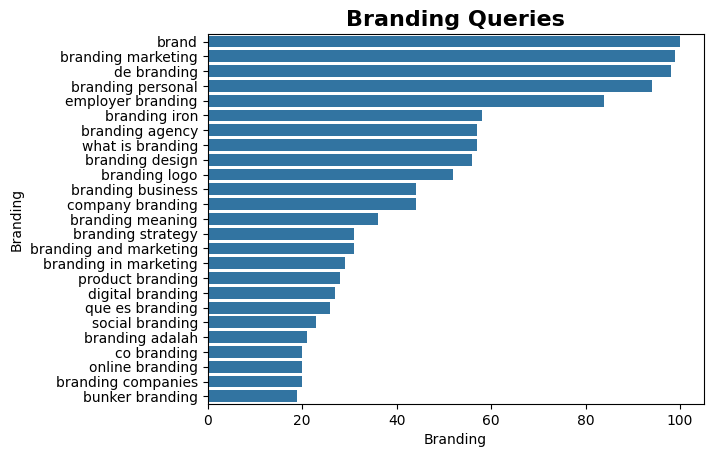
Cụm 4: Cụm này nhấn mạnh tầm quan trọng của phương tiện truyền thông xã hội trong các chiến lược tiếp thị kỹ thuật số.

Cụm 5: Cụm này tập trung vào các khía cạnh giáo dục, đặc biệt là các khóa học tiếp thị kỹ thuật số.

Cụm 6: Cụm này tập trung vào các cơ hội việc làm trong lĩnh vực tiếp thị kỹ thuật số. Cụm 7: Cụm này tập trung vào việc hiểu khái niệm cơ bản của tiếp thị kỹ thuật số.

**Branding**

**Truy vấn**

****

Biểu đồ cho thấy rõ ràng rằng các truy vấn 'xây dựng thương hiệu' và 'tiếp thị thương hiệu' là những chủ đề rất phổ biến và được tìm kiếm thường xuyên.

**Tạo đám mây**



**Nhóm các từ khóa liên quan**

Clusters:

Cluster 5: brand

Cluster 2: branding marketing, branding and marketing, branding in marketing

Cluster 1: de branding, employer branding, branding iron, what is branding, branding design, branding logo, company branding, branding meaning, branding strategy, product branding, digital branding, social branding, branding adalah, co branding, online branding, branding companies, bunker branding

Cluster 3: branding personal

Cluster 6: branding agency

Cluster 7: branding business

Cluster 4: que es branding

Xét sâu từng cụm

Nhóm 1: Nhóm này bao gồm nhiều thuật ngữ liên quan đến các khía cạnh khác nhau của thương hiệu, bao gồm thiết kế, chiến lược, sự hiện diện kỹ thuật số và đồng thương hiệu. Nhóm 2: Nhóm này tập trung vào việc tích hợp thương hiệu trong bối cảnh tiếp thị rộng hơn.

Nhóm 3: Nhóm này bao gồm truy vấn 'Thương hiệu là gì', cho biết người dùng ở đây đang trong giai đoạn đầu khám phá và hiểu khái niệm cốt lõi của thương hiệu.

Nhóm 4: Nhóm này tập trung vào 'thương hiệu xã hội', cho thấy người dùng ở đây đặc biệt quan tâm đến cách các thương hiệu thiết lập và duy trì sự hiện diện của họ trên các nền tảng truyền thông xã hội.

Nhóm 5: Nhóm này tập trung vào khái niệm cơ bản về 'thương hiệu'.

Nhóm 6: Nhóm này đặc biệt xoay quanh 'thương hiệu nhà tuyển dụng'. Người dùng ở đây có thể quan tâm đến cách các công ty xây dựng và quản lý thương hiệu nhà tuyển dụng của họ.

Nhóm 7: Nhóm này tập trung vào 'các công ty xây dựng thương hiệu', cho thấy người dùng ở đây có thể đang khám phá các cơ quan, công ty hoặc dịch vụ chuyên về thương hiệu.

Ngoài ra còn có biểu phân tích theo ngày , tháng và có ngày đặc biệt ( Chi tiết file pdf tổng hợp) .

### 4.4.5 Đề xuất

#### 4.4.5.1 Thông tin cho khách hàng

* Đầu năm (T1–T2): Từ khóa như “tiếp thị”, “tỷ lệ chuyển đổi”, “quảng cáo Google” tăng mạnh → liên quan đến lập kế hoạch & phân bổ ngân sách đầu năm.
* Giữa năm (T6): Từ khóa như “SEO”, “Google Analytics”, “hành vi người dùng” tăng → phản ánh đánh giá lại chiến lược giữa năm.
* Cuối năm (T11): Từ khóa như “giữ chân khách hàng”, “chi phí mua lại” đạt đỉnh → do chuẩn bị tổng kết và chiến dịch cuối năm.
* Tháng 12 giảm do lễ hội và nghỉ lễ.
* Ngày 25 hằng tháng: Khối lượng tìm kiếm giảm rõ rệt (đặc biệt là Giáng sinh) → do ngày lễ, giảm hoạt động kinh doanh.
* Ngày 8 và 15 hằng tháng: Lượt tìm kiếm tăng cao → có thể trùng với chu kỳ làm việc, chiến dịch định kỳ hoặc báo cáo nội bộ.
* Từ khóa KPI (CPC, CTR, ROI…): Tăng mạnh vào giữa tháng, cho thấy doanh nghiệp thường đánh giá hiệu suất định kỳ.

#### 4.4.5.2 Thời điểm triển khai chiến dịch ( ngày )

* Ngày 8 và 15 hàng tháng → Phù hợp cho chiến dịch tiếp thị kỹ thuật số (SEO, quảng cáo Google, SEM).
* Ngày 15 hàng tháng → Tối ưu cho chiến dịch dựa trên số liệu (CTR, ROI, tỷ lệ chuyển đổi).
* Ngày 5, 6 và 15 → Phù hợp cho chiến dịch chiến lược và xây dựng thương hiệu.
* Ngày 5, 8, 17 và 24 → Nên tập trung vào chiến dịch thu hút và giữ chân khách hàng.

#### 4.4.5.3 Nhóm từ khóa theo chủ đề

|  |  |  |
| --- | --- | --- |
| Chủ đề | Từ khóa tiêu biểu | Mục tiêu |
| Tiếp thị kỹ thuật số toàn diện | tiếp thị kỹ thuật số, SEO, SEM, quảng cáo Google, truyền thông xã hội | Tăng nhận diện & tương tác nội dung |
| Tiếp thị dựa trên dữ liệu | Google Analytics, hành vi người dùng | Phân tích & cải thiện hiệu suất |
| Xây dựng thương hiệu | xây dựng thương hiệu, giữ chân khách hàng, chiến lược tiếp thị | Phát triển hình ảnh & lòng trung thành |
| Quảng cáo trả phí & ROI | CPC, ROI, tiếp thị hiệu suất | Tối ưu chi phí & hiệu quả quảng cáo |
| Thu hút và giữ chân khách hàng | thu hút khách hàng, giữ chân khách hàng | Tăng trưởng & duy trì khách hàng |

#### 4.4.5.4 Xu hướng từ khóa giảm

* Giảm quan tâm: từ khóa như Tiếp thị hiệu suất, Trả tiền cho mỗi lần nhấp, Chi phí mỗi lần nhấp.
* Khuyến nghị: Nên mở rộng sang các chiến lược đa dạng, kết hợp nhiều chỉ số để tăng hiệu quả.

### 4.4.6 Tối ưu Theo dự đoán LLM của FluentFocus

FluentFocus giúp doanh nghiệp triển khai chiến dịch đúng thời điểm, từ khóa và kênh, đồng bộ với xu hướng thị trường.

Thời điểm vàng:

* Tháng 1–2: Khởi động chiến lược mới → từ khóa "Tiếp thị", "SEO", "Xây dựng thương hiệu"
* Tháng 6: Tối ưu giữa năm → từ khóa "Tiếp thị kỹ thuật số", "Google Analytics"
* Tháng 11: Tập trung giữ chân và chốt sale → "Chiến lược tiếp thị", "ROI", "Giữ chân khách hàng”

Kênh triển khai:

* Social Media: Facebook, Instagram, LinkedIn
* Email Marketing: Cá nhân hóa theo từ khóa
* Content Marketing: Bài blog, video, infographi

Chủ đề sáng tạo theo tháng:

* T1–2: "Năm mới, Chiến lược mới!"
* T6: "Tối ưu giữa năm!"
* T11: "Chiến thắng cuối năm!"

Lợi ích chính:

* Tăng tỷ lệ chuyển đổi
* Tăng hiển thị thương hiệu
* Tăng giữ chân khách hàng

Kết luận: FluentFocus giúp bạn đi trước xu hướng, tối ưu hóa từng chiến dịch tiếp thị dựa trên dữ liệu thực tế.

# CHƯƠNG 5. KẾT LUẬN VÀ KIẾN NGHỊ

## 5.1 Kết luận

* **Facebook Ads** là kênh phù hợp để tiếp cận số lượng lớn người dùng, tuy nhiên tỷ lệ chuyển đổi thấp, do đó phù hợp hơn cho các chiến dịch tăng độ phủ thương hiệu hoặc tương tác ban đầu.
* **Google Ads** cho thấy hiệu quả cao trong việc chuyển đổi hành vi người dùng, đặc biệt ở các giai đoạn cuối phễu (như mua hàng, điền form đăng ký).
* **TikTok Ads** phù hợp với các chiến dịch nâng cao nhận diện thương hiệu và thu hút sự tương tác, nhưng chưa thực sự hiệu quả với mục tiêu chuyển đổi trực tiếp.

## ****5.2 Kiến nghị****

* **Tái phân bổ ngân sách quảng cáo**: Tăng tỷ trọng đầu tư cho **Google Ads** trong các khung giờ có tỷ lệ chuyển đổi cao; đồng thời **giảm ngân sách TikTok** nếu mục tiêu trọng tâm là chuyển đổi (conversion).
* **Thực hiện A/B Testing định kỳ**: Thử nghiệm các phiên bản nội dung khác nhau (hình ảnh, tiêu đề, CTA...) và nhóm đối tượng mục tiêu nhằm tối ưu hiệu quả quảng cáo theo từng nền tảng.
* **Ứng dụng công cụ phân tích và tự động hóa**: Triển khai hệ thống phân tích dữ liệu quảng cáo theo thời gian thực để phát hiện sớm xu hướng và điều chỉnh chiến lược kịp thời, giúp tối ưu hóa ngân sách và hiệu quả truyền thông.

# CHƯƠNG 6. TÀI LIỆU THAM KHẢO

* Dobrowolski, G., & Kowalski, A. (2023). Measuring the Effectiveness of Digital Communication Using AI-Supported Tools. Procedia Computer Science, 232, 485–491. https://doi.org/10.1016/j.procs.2023.01.485
* Nguyen, T. Q., & Tran, D. H. (2023). Analysis of Social Media Ads: Performance Metrics and Conversion Rates. IEEE Xplore Digital Library, 29(4), 275–284. https://doi.org/10.1109/ICCMC.2023.9702575
* Kumar, P., & Li, J. (2023). Customer-Brand Interaction on Instagram and TikTok: A Comparative Analysis. Journal of Digital Marketing Research, 35(2), 112–125. https://doi.org/10.1177/385265478
* Ahmed, S., & Kim, J. (2024). Social Media Influencer KPIs: A Cross-Platform Analysis of Facebook, Instagram, YouTube, TikTok, and Twitter. Journal of Social Media Marketing, 22(5), 375–389.https://doi.org/10.1177/370575204
* Fayvishenko, D., Kovalchuk, S., Sidielnikov, D., Holik, O., & Kyianytsia, I. (2025). The effectiveness of digital communications for promoting brands on social media platforms. Journal of Theoretical and Applied Information Technology, 103(6), 2348–2359. https://www.jatit.org/volumes/Vol103No6/18Vol103No6.pdf