

Graduation Project Proposal

Phân Tích & Dự Đoán Xu Hướng Mạng Xã Hội

Người thực hiện:

Nguyễn Mạnh Chí – DA_240930

Tên dữ liệu sử dụng:

Viral Social Media Trends & Engagement Analysis

Nguồn dữ liệu:

<https://www.kaggle.com/datasets/atharvasoundankar/viral-social-media-trends-and-engagement-analysis>

1. Giới Thiệu Dự Án

Mục Tiêu

Dự án này nhằm phân tích các xu hướng lan truyền trên nhiều nền tảng mạng xã hội (TikTok, Instagram, Twitter, YouTube) thông qua các chỉ số tương tác và dự đoán xu hướng hashtag trong tương lai. Mục tiêu là cung cấp các phân tích chuyên sâu giúp tối ưu hóa chiến lược nội dung và marketing trên mạng xã hội.

Mục Tiêu Chính

- Xác định các yếu tố quan trọng ảnh hưởng đến mức độ lan truyền và tương tác của nội dung.
- So sánh mức độ tương tác giữa các nền tảng và loại nội dung khác nhau.
- Xây dựng mô hình dự đoán xu hướng hashtag trong tương lai.
- Cung cấp thông tin hữu ích để tối ưu hóa chiến lược marketing trên mạng xã hội.

2. Mô Tả Dữ Liệu

Tập Dữ Liệu: Viral_Social_Media_Trends.csv

Tập dữ liệu này gồm 5.000 bài đăng viral trên mạng xã hội với các đặc điểm sau:

- **Post_ID:** Mã định danh duy nhất của bài đăng.
- **Platform:** Nền tảng mạng xã hội (TikTok, Instagram, Twitter, YouTube).
- **Hashtag:** Hashtag đang thịnh hành gắn với bài đăng.
- **Content_Type:** Loại nội dung (Reel, Video, Post, Shorts, Tweet, v.v.).
- **Region:** Quốc gia nơi bài đăng đạt được sức hút lớn.
- **Views:** Tổng số lượt xem.
- **Likes:** Số lượt thích.
- **Shares:** Số lần chia sẻ.
- **Comments:** Số bình luận.
- **Engagement_Level:** Mức độ tương tác (Thấp, Trung bình, Cao).

3. Câu Hỏi Nghiên Cứu & Hướng Phân Tích

A. Phân Tích Tổng Quan Xu Hướng & Engagement

- Nền tảng nào tạo ra mức độ tương tác cao nhất?
- Loại nội dung nào phổ biến nhất trong các bài đăng viral?
- Có sự khác biệt nào về tương tác giữa các khu vực không?
- Mối quan hệ giữa lượt xem, lượt thích, chia sẻ và bình luận là gì?

Phương pháp:

- Phân tích thống kê mô tả (mean, median, phân phối).
- Trực quan hóa dữ liệu bằng biểu đồ (bar chart, heatmap).
- Phân cụm các bài đăng dựa trên engagement để tìm pattern phổ biến.

B. Phân Tích Hashtag Trends

- Những hashtag nào xuất hiện thường xuyên nhất trong các bài viết viral?
- Có mối liên hệ giữa hashtag cụ thể và mức độ tương tác không?

- Một hashtag có hiệu suất tương tác khác nhau trên từng nền tảng hay không?
- Các hashtag phổ biến nhất theo khu vực là gì?

Phương pháp:

- Phân tích tần suất hashtag.
- Tính trung bình các chỉ số tương tác theo hashtag.
- So sánh hiệu quả của hashtag giữa các nền tảng và vùng địa lý.
- Phân tích mối tương quan giữa hashtag và Engagement_Level.

4. Công Cụ Sử Dụng

- Ngôn ngữ: Python
- Thư viện: Pandas, NumPy, Matplotlib, Seaborn
- Trực quan hóa nâng cao: Power BI

5. Kết Quả Mong Đợi

1. Báo cáo tổng quan về hành vi tương tác trên các nền tảng mạng xã hội.
2. Danh sách các hashtag có hiệu suất cao và phân tích đặc trưng của chúng.
3. Dashboard tương tác trình bày các phát hiện chính.
4. Gợi ý chiến lược sử dụng hashtag hiệu quả theo từng nền tảng/khu vực.

7. Lợi Ích & Ứng Dụng

- Tối ưu hóa chiến lược nội dung: Chọn đúng loại nội dung và nền tảng phù hợp với nhóm người dùng mục tiêu.
- Sử dụng hashtag thông minh: Tăng cơ hội tiếp cận và tương tác bằng cách chọn đúng hashtag.
- Nắm bắt xu hướng kịp thời: Ra quyết định dựa trên dữ liệu thay vì cảm tính.

8. Kết Luận

Dự án mang lại cái nhìn sâu sắc về xu hướng nội dung lan truyền và chiến lược sử dụng hashtag hiệu quả trên mạng xã hội. Các kết quả thu được sẽ hỗ trợ mạnh mẽ cho doanh nghiệp, influencer và marketer trong việc phát triển nội dung tối ưu và gia tăng tương tác.