

**ĐỀ CƯƠNG CHI TIẾT**  
**KHÓA LUẬN TỐT NGHIỆP NGÀNH CÔNG NGHỆ THÔNG TIN**

Họ tên sinh viên: Trần Văn Sang

MSSV: 110121094

Lớp: DA21TTB

Khóa: 2021

Tên đề tài: Xây dựng hệ thống phân tích dữ liệu và dự đoán xu hướng video thịnh hành trên nền tảng YouTube.

1. Mục tiêu của đề tài:

Xây dựng hệ thống thu thập thông tin về video trên nền tảng Youtube. Từ đó phân tích dữ liệu và dự đoán xu hướng video nào sẽ trở thành thịnh hành trên nền tảng này.

2. Nội dung thực hiện:

- Nghiên cứu tài liệu liên quan đến YouTube Data API v3, thuật toán Extreme Gradient Boosting (XGBoost), ngôn ngữ Python và thư viện liên quan, Django và Tailwind CSS framework.

- Thực hiện gọi API để thu thập thông tin video trên Youtube, lưu trữ dữ liệu thu thập được dưới dạng CSV, tiền xử lý dữ liệu thu thập được, phân tích dữ liệu sau khi tiền xử lý.

- Xây dựng hệ thống website gồm các chức năng thu thập, xử lý, phân tích dữ liệu, áp dụng thuật toán XGBoost vào mô hình dự đoán với backend sử dụng Django và frontend dùng Tailwind CSS.

3. Phương pháp thực hiện:

- Phương pháp nghiên cứu lý thuyết: Nghiên cứu YouTube Data API v3, các đặc điểm của các video đang thịnh hành trên nền tảng Youtube, thuật toán học máy XGBoost, ngôn ngữ Python và các thư viện trực quan hóa dữ liệu, Django và Tailwind CSS framework.

- Phương pháp thực nghiệm: Thu thập dữ liệu từ YouTube thông qua API; dùng Django, Tailwind CSS framework, các thư viện của Python và thuật toán XGBoost để xây dựng hệ thống website gồm các chức năng tiền xử lý, phân tích dữ liệu và dự đoán khả năng một video trở thành xu hướng.

#### 4. Bố cục đề tài:

- Chương 1: Tổng quan: Tổng quan về các bài báo khoa học và công trình nghiên cứu đã thực hiện về hệ thống dự báo và dự đoán xu hướng video, đề cập đến các thuật toán phổ biến được sử dụng để dự đoán các video có khả năng thịnh hành, thuật toán YouTube sử dụng để xác định video thịnh hành, giới thiệu về Django framework trong việc phát triển hệ thống website phân tích dữ liệu.

- Chương 2: Nghiên cứu lý thuyết: Chương này trình bày về các khái niệm lý thuyết liên quan đến phân tích dữ liệu, các thuật toán dự đoán xu hướng, cũng như cách sử dụng API và thư viện để thu thập dữ liệu từ YouTube. Đồng thời, nghiên cứu cũng đề cập đến việc ứng dụng Tailwind CSS để xây dựng giao diện người dùng và Django như một framework backend mạnh mẽ.

- Chương 3: Hiện thực hóa nghiên cứu: Trong chương này, trình bày chi tiết về yêu cầu bài toán dự đoán video thịnh hành trên nền tảng Youtube. Để giải quyết vấn đề, nghiên cứu sẽ trình bày các bước từ thu thập, làm sạch, đến phân tích dữ liệu. Tiếp theo, nghiên cứu sẽ đưa ra mô hình cho bài toán dự đoán sử dụng thuật toán XGBoost và áp dụng Tailwind CSS, Django framework nhằm xây dựng hệ thống website theo yêu cầu đã đề ra.

- Chương 4: Kết quả nghiên cứu: Trong chương này trình bày kết quả đạt được sau quá trình thực hiện đề tài, bao gồm sản phẩm đã hoàn thiện, hình ảnh minh họa giao diện hệ thống, cũng như các chức năng chính đã được xây dựng và triển khai thành công.

- Chương 5: Kết luận – Kiến nghị: Chương này sẽ đưa ra nhận xét và đánh giá về hệ thống đã xây dựng. Bên cạnh đó, báo cáo đề xuất hướng phát triển trong tương lai cũng như khả năng mở rộng cho đề tài.

#### 5. Tài liệu tham khảo:

[1] C. M. Bishop (2006), *Pattern Recognition and Machine Learning*, Springer.

[2] W. McKinney (2017), *Python for Data Analysis*, 2nd ed., O'Reilly Media.

[3] Django Documentation, *Django Software Foundation*, 2025. [Online]. Available: <https://docs.djangoproject.com/en/stable/>. [Accessed: Apr. 8, 2025].

[4] Tailwind CSS Documentation, *Tailwind CSS*, 2025. [Online]. Available: <https://tailwindcss.com/docs>. [Accessed: Apr. 10, 2025].

[5] YouTube API, *YouTube*, 2025. [Online]. Available: <https://www.youtube.com/yt/dev/>. [Accessed: Apr. 17, 2025].

[6] YouTube Data API v3 Documentation, *Google Developers*, 2025. [Online]. Available: <https://developers.google.com/youtube/v3>. [Accessed: Apr. 18, 2025].

#### 6. Kế hoạch thực hiện đề tài

Tuần	Từ ngày - đến ngày	Công việc thực hiện	Ghi chú
1	Từ ngày 08/04/2025 đến ngày 14/04/2025	<ul style="list-style-type: none"><li>- Tìm hiểu YouTube API: các loại dữ liệu cung cấp, hạn chế.</li><li>- Nghiên cứu các mô hình học máy: XGBoost</li><li>- Tìm hiểu các kỹ thuật trực quan hóa dữ liệu: biểu đồ phân phối, heatmap,...</li><li>- Viết báo cáo.</li><li>- Báo cáo tiến độ với GVHD.</li></ul>	
2	Từ ngày 15/04/2025 đến ngày 21/04/2025	<ul style="list-style-type: none"><li>- Tìm hiểu các yếu tố nổi bật ảnh hưởng đến xu hướng thịnh hành.</li><li>- Nghiên cứu Django và Tailwind CSS để chuẩn bị xây dựng hệ thống website.</li><li>- Xây dựng mô hình dự đoán.</li><li>- Viết báo cáo.</li><li>- Báo cáo tiến độ với GVHD.</li></ul>	
3	Từ ngày 22/04/2025 đến ngày 28/04/2025	<ul style="list-style-type: none"><li>- Xây dựng công cụ thu thập dữ liệu từ YouTube bằng API.</li><li>- Thiết kế giao diện người dùng.</li><li>- Viết báo cáo.</li><li>- Chạy demo.</li><li>- Báo cáo tiến độ với GVHD.</li></ul>	
4	Từ ngày 29/04/2025 đến ngày 05/05/2025	<ul style="list-style-type: none"><li>- Xây dựng chức năng tiền xử lý.</li><li>- Viết báo cáo.</li><li>- Báo cáo tiến độ với GVHD.</li></ul>	

5	Từ ngày 06/05/2025 đến ngày 12/05/2025	- Xây dựng chức năng thống kê và phân tích dữ liệu. - Chạy demo. - Viết báo cáo. - Báo cáo tiến độ với GVHD.	
6	Từ ngày 13/05/2025 đến ngày 19/05/2025	- Huấn luyện các mô hình học máy: XGBoost. - Chạy demo. - Viết báo cáo. - Báo cáo tiến độ với GVHD.	
7	Từ ngày 20/05/2025 đến ngày 26/05/2025	- Xây dựng chức năng dự đoán. - Đánh giá mô hình. - Viết báo cáo. - Báo cáo tiến độ với GVHD.	
8	Từ ngày 27/05/2025 đến ngày 02/06/2025	- Phát triển giao diện người dùng bằng Tailwind CSS. - Thiết kế UI đơn giản, dễ sử dụng, phản ánh đúng quy trình. - Chạy demo. - Viết báo cáo. - Báo cáo tiến độ với GVHD.	
9	Từ ngày 03/06/2025 đến ngày 09/06/2025	- Kiểm thử toàn bộ hệ thống. - Tối ưu tốc độ xử lý, sửa lỗi nếu phát sinh. - Viết báo cáo. - Báo cáo tiến độ với GVHD.	
10	Từ ngày 10/06/2025 đến ngày 16/06/2025	- Hoàn chỉnh toàn bộ hệ thống, hoàn chỉnh báo cáo.	

**GIẢNG VIÊN HƯỚNG DẪN**

*Trà Vinh, ngày tháng 4 năm 2025*  
**SINH VIÊN THỰC HIỆN**