

TRƯỜNG KỸ THUẬT VÀ CÔNG NGHỆ
KHOA CÔNG NGHỆ THÔNG TIN

**ĐỀ CƯƠNG CHI TIẾT
THỰC TẬP ĐỒ ÁN CHUYÊN NGÀNH**

Tên đề tài: Tìm hiểu về Phân tích dữ liệu (Data Analytics)

Giảng viên hướng dẫn: TS. Nguyễn Bảo Ân

Thời gian thực hiện: 03/11/2025-28/12/2025

Sinh viên thực hiện: Cô Nhân Quý

Mã số sinh viên: 110122150 - **Mã lớp:** DA22TTB

Nội dung đề tài:

- Mô tả:**

Các chuyên gia và nhà quản lý đã nhấn mạnh, dữ liệu là nguồn tài nguyên chiến lược của nền kinh tế số và chuyển đổi số quốc gia, cần được bảo đảm an toàn thông tin, an ninh mạng và bảo vệ dữ liệu cá nhân. Việc nâng cao năng lực phân tích, xử lý và quản trị dữ liệu là yêu cầu cấp thiết với mọi tổ chức, doanh nghiệp và cá nhân.

Đó là một trong những nội dung của Tọa đàm “Xử lý và phân tích dữ liệu: Động lực cho chuyển đổi số quốc gia” diễn ra vào ngày 28/05/2025 đã cho thấy tầm quan trọng của việc phân tích dữ liệu. Phân tích dữ liệu vốn là một môn học cũng như một ngành khoa học cung cấp phương pháp xử lý các dữ liệu thô thông qua các công cụ nhằm tìm hiểu các quy luật của dữ liệu được cho là bất thường hoặc không rõ ràng trong các tập dữ liệu.

Và mục tiêu của đề tài này là tìm hiểu quy trình phân tích dữ liệu (Data Analytics) - bao gồm các bước: thu thập, làm sạch, xử lý, và trực quan hóa dữ liệu. Thông qua việc sử dụng các thư viện Python phổ biến như Pandas, Matplotlib, Seaborn, sinh viên sẽ xây dựng một ứng dụng nhỏ phân tích dữ liệu CSV (ví dụ: điểm

sinh viên hoặc doanh thu bán hàng), kèm dashboard trực quan đơn giản để thể hiện kết quả phân tích.

- **Phương pháp thực hiện:**

- 1. Nghiên cứu lý thuyết:**

Nghiên cứu tổng quan về quy trình phân tích dữ liệu, các khái niệm về dữ liệu thô, dữ liệu sạch, và các phương pháp trực quan hóa dữ liệu.

- 2. Lựa chọn công cụ, thư viện sử dụng và chuẩn bị dữ liệu:**

Sử dụng ngôn ngữ Python và các thư viện Pandas, Matplotlib, Seaborn, Plotly hoặc Streamlit để xây dựng dashboard.

Xác định và thu thập nguồn dữ liệu đầu vào từ file CSV (tập tin điểm sinh viên hoặc doanh thu bán hàng).

- 3. Tiến hành thực nghiệm:**

Bước 1: Làm sạch và tiền xử lý dữ liệu:

Xử lý các giá trị thiếu, loại bỏ dữ liệu trùng lặp hoặc ngoại lai, chuyển đổi định dạng dữ liệu.

Bước 2: Phân tích và thống kê:

Thực hiện các phép tính thống kê mô tả (trung bình, phương sai, tương quan), phân tích xu hướng và tỷ lệ.

Bước 3: Trực quan hóa dữ liệu:

Xây dựng các biểu đồ (cột, đường, tròn, boxplot,...) để minh họa kết quả phân tích một cách trực quan.

Bước 4: Xây dựng Dashboard:

Tích hợp các biểu đồ và kết quả phân tích vào một giao diện dashboard tổng quan, cho phép lọc và tương tác cơ bản.

4. Kiểm thử, điều chỉnh và đánh giá:

Kiểm thử toàn bộ quy trình phân tích dữ liệu, từ khâu làm sạch đến trực quan hóa nhằm đảm bảo chương trình chạy ổn định, không có lỗi và cho ra kết quả chính xác.

Đánh giá tính chính xác của kết quả phân tích cùng khả năng truyền tải thông tin của các biểu đồ, dashboard.

5. Tổng kết và viết báo cáo:

Tổng hợp toàn bộ quá trình thực hiện, kết quả đạt được và bài học kinh nghiệm vào một báo cáo chi tiết. Chuẩn bị nội dung và slide để báo cáo, thuyết trình đồ án.

- **Kết quả đạt được:**

- Hiểu rõ quy trình phân tích dữ liệu và các bước làm sạch, xử lý, trực quan hóa dữ liệu.
- Biết cách sử dụng các thư viện Pandas, Matplotlib, Seaborn trong thực tế.
- Xây dựng được ứng dụng phân tích dữ liệu nhỏ và dashboard trực quan.
- Biết cách đánh giá và trình bày kết quả phân tích dưới dạng báo cáo.
- Hoàn thiện báo cáo và nội dung trình bày kết quả thực nghiệm và nộp đồ án .

- **Kế hoạch thực hiện:**

Tuần	Thời gian	Nội dung công việc	Kết quả dự kiến
Chuẩn bị	Từ ngày 03/11 đến 9/11/2025	<ul style="list-style-type: none">- Tìm hiểu tổng quan về Data Analytics và các bước trong quy trình phân tích dữ liệu.- Thực hành cơ bản với Pandas: đọc, xử lý, làm sạch dữ liệu CSV.	<ul style="list-style-type: none">- Nắm được khái niệm, vai trò và các giai đoạn của Data Analytics.- Hiểu và thực hành được các thao tác xử lý dữ liệu bằng Pandas.

Tuần 1	Từ ngày 10/11 đến 16/11/2025	<ul style="list-style-type: none"> - Thực hành phân tích dữ liệu: tính toán, tổng hợp, thống kê mô tả. - Tìm hiểu trực quan hóa dữ liệu với Matplotlib và Seaborn. 	<ul style="list-style-type: none"> - Hoàn thành phần phân tích dữ liệu cơ bản. - Hiểu và nhận biết dữ liệu, xu hướng và so sánh.
Tuần 2	Từ ngày 24/11 đến 30/11/2025	<ul style="list-style-type: none"> - Thực hành trực quan hóa dữ liệu với Matplotlib và Seaborn. - Thiết kế dashboard hiển thị kết quả phân tích (dùng Streamlit/Plotly) 	<ul style="list-style-type: none"> - Tạo được các biểu đồ trực quan thể hiện xu hướng dữ liệu. - Hoàn thành dashboard trực quan đơn giản.
Tuần 3	Từ ngày 08/12 đến 14/12/2025	<ul style="list-style-type: none"> - Kiểm thử và sửa lỗi tổng thể. - Viết báo cáo về quá trình thực hiện và kết quả thu được. 	<ul style="list-style-type: none"> - Chương trình chạy ổn định, ít lỗi. - Hoàn thành phần lý thuyết và hình ảnh minh họa trong báo cáo
Tuần 4	Từ ngày 22/12 đến 28/12/2025	<ul style="list-style-type: none"> - Kiểm thử và sửa lỗi tổng thể lần cuối. - Hoàn thiện toàn bộ báo cáo, kiểm tra và chỉnh sửa nội dung 	Báo cáo hoàn chỉnh
Kết thúc	29/12/2025	Chuẩn bị nộp đề tài và hoàn tất thủ tục.	Hoàn thành đồ án chuyên ngành và nộp cho giảng viên hướng dẫn.

Xác nhận của GVHD

Ngày 02 tháng 11 năm 2025

Sinh viên thực hiện

Nguyễn Bảo Ân

Cô Nhân Quý