

LAB THỰC HÀNH SỐ 2. THỰC HÀNH VỚI THƯ VIỆN NUMPY

LAB 2.1. Phân tích dữ liệu khí hậu sử dụng numpy

Thời gian hoàn thành: (30 phút)

Mục Tiêu:

- Hiểu và thực hành cách tạo mảng và thao tác cơ bản trên mảng Numpy.
- Áp dụng fancy indexing và chỉ mục kết hợp để phân tích dữ liệu.

Các Bước Thực Hiện:

1. Tạo Dữ Liệu Mô Phỏng Nhiệt Độ: (10 phút)

- Hãy tạo một mảng Numpy mô phỏng dữ liệu nhiệt độ hàng ngày trong một tháng là số thực. Làm tròn dữ liệu đến 2 chữ số sau dấu phẩy
- Cho biết nhiệt độ trung bình trong tháng là bao nhiêu?

2. Phân Tích Xu Hướng Nhiệt Độ: (5-10 phút)

- Xác định ngày có nhiệt độ cao nhất, thấp nhất cho biết đó là ngày thứ mấy trong tháng?
- Thống kê sự chênh lệch nhiệt độ giữa các ngày, tìm ngày có sự biến đổi nhiệt độ cao nhất?

3. Áp dụng Fancy Indexing: (10 phút)

Đưa ra tất cả các ngày có nhiệt độ cao hơn 20 độ C.

Lấy nhiệt độ của ngày 5, 10, 15, 20, và 25.

Tìm nhiệt độ của các ngày có nhiệt độ trên trung bình.

Lấy nhiệt độ của các ngày chẵn/lẻ trong tháng.

LAB 2.2. Phân tích dữ liệu điểm số sinh viên sử dụng numpy.

Thời gian hoàn thành: (45-60 phút)

Mục tiêu:

- Tăng cường kỹ năng sử dụng NumPy
- Nắm vững kiến thức lý thuyết về xử lý và phân tích dữ liệu sang các tình huống thực tế thông qua việc xử lý dữ liệu từ file CSV.
- Cách thực hiện các phân tích cơ bản như tính toán tổng, trung bình, độ lệch chuẩn và sử dụng các kỹ thuật như broadcasting và fancy indexing trong NumPy.
- Kỹ năng giải quyết vấn đề và tư duy lập trình qua việc áp dụng các phương pháp phân tích trên dữ liệu cụ thể.

Các Bước Thực Hiện:

Giả sử có file dữ liệu 'diem_hoc_phan.csv' chứa dữ liệu bao gồm: 'id', 'tên sinh viên', 'HP 1', 'HP 2', 'HP3' các điểm HP1, HP2, HP3 theo thang điểm 10.

1. Đọc Dữ Liệu Từ File CSV: (10 phút)

- Đọc dữ liệu điểm số từ file '**diem_hp.csv**' vào một list.

- Chuyển đổi list này thành một mảng NumPy để thực hiện các phân tích. Phân Tích Điểm Số:
- Qui đổi từ thang điểm 10 sang điểm tín chỉ: (15 phút)
 Biết rằng:
 - 8.5 <= điểm <= 10 thì tương ứng điểm 'A'
 - 8.0 <= điểm <= 8.4 ứng với điểm 'B+'
 - 7.0 <= điểm < 8 ứng điểm B
 - 6.5 <= điểm < 7 điểm C+
 - 5.5 <= điểm < 6 điểm C
 - 5.0 <= điểm < 5.5 điểm D+
 - 4.0 <= điểm < 5 điểm D
 - điểm < 4 ứng điểm F
 - Chia tách Dữ Liệu theo học phần để phân tích : (5-10 phút)
 Chia tách dữ liệu để phân tích điểm số của mỗi học phần riêng biệt.
 - Phân tích dữ liệu theo từng học phần : (10-15 phút)
 Phân tích tổng, trung bình, và độ lệch chuẩn cho mỗi học phần.
 - Phân tích điểm tổng hợp. (5-10 phút)
 Tính toán số lượng sinh viên đạt từng loại điểm chữ (A, B+, B,...) trên tổng số học phần.

LAB 2.3. Phân tích hiệu suất sản xuất dựa trên ca làm việc

Thời gian hoàn thành: (45-70 phút)

Mục tiêu: Giúp sinh viên củng cố kiến thức và kỹ năng lập trình của mình, đặc biệt là trong lĩnh vực xử lý và phân tích dữ liệu

Giả sử có hai tập tin: **efficiency.txt** và **shifts.txt**. Tập tin **efficiency.txt** chứa dữ liệu về hiệu suất sản xuất (đo bằng số lượng sản phẩm hoàn thành) của các nhân viên trong một ngày làm việc, mỗi dòng trong tập tin ứng với số lượng sản phẩm hoàn thành bởi một nhân viên. Tập tin **shifts.txt** chứa thông tin về ca làm việc của họ ('Morning', 'Afternoon', 'Night'). Mỗi dòng tương ứng với ca làm việc của một nhân viên.

Yêu cầu của bài lab như sau:

- Đọc dữ liệu từ 2 tập tin **efficiency.txt** và **shifts.txt** vào 2 list là **efficiency** và **shifts** : (10-phút)
- Tạo numpy array **np_shifts** từ list **shifts** và kiểm tra kiểu dữ liệu của **np_shifts**. (10 phút)
- Tạo numpy array **np_efficiency** từ list **efficiency** và kiểm tra kiểu dữ liệu của **np_efficiency**. (10 phút)
- Tính hiệu suất sản xuất trung bình của những nhân viên làm việc vào ca 'Morning'. (10 phút)
- Tính hiệu suất sản xuất trung bình của những nhân viên làm việc trong các ca khác (Không phải là 'Morning'). (10 phút)

6. Tạo mảng dữ liệu có cấu trúc (Structure Array tự định nghĩa workers gồm 'shift' (kiểu văn bản U10) và 'efficiency' (kiểu 'float')). (10 phút)
7. Sắp xếp mảng workers theo efficiency, và xác định ca làm việc nào có hiệu suất cao nhất và thấp nhất. (10 phút)