

Học viện Công nghệ Bưu chính Viễn thông
Khoa: Công nghệ thông tin

Báo cáo bài tập lớn Python

Họ tên: Nguyễn Hữu Đức
Mã sinh viên: B22DCKH033
Nhóm: 11
Giảng viên: Kim Ngọc Bách

Bài 1:

Bài gồm 3 phần:

* Players:

Mục Đích: Thu thập thông tin của từng cầu thủ

Các bước làm bài:

Bước 1: Phân tích yêu cầu của bài

Xác định các chỉ số cần thu thập của cầu thủ.

Bước 2: Thu thập dữ liệu từ website fbref.com

Thư viện cần thiết: requests để gửi yêu cầu http, BeautifulSoup để phân tích html và pandas để xử lý dữ liệu và ghi vào file csv.

Gửi yêu cầu đến trang web fbref.com để truy cập trang thống kê các cầu thủ Ngoại hạng Anh mùa giải 2023 – 2024

Phân tích đoạn mã html để tìm bảng dữ liệu

Bước 3: Thu thập dữ liệu:

Lọc các cầu thủ có thời gian thi đấu hơn 90 phút

Các thông tin cơ Bản: Tên, quốc tịch, vị trí chơi bóng, tuổi,..

Chỉ số chơi bóng

Chỉ số mong đợi

Điền giá trị thiếu bằng “N/a”

Bước 4: ghi dữ liệu vào file “results.csv”

* Sorted_players:

Mục đích: sắp xếp lại danh sách cầu thủ theo First name, nếu trùng tên thì xếp theo độ tuổi từ lớn đến nhỏ

* Title:

Mục đích: đặt lại cấu trúc cột dữ liệu, để xác định các chỉ số cần lấy từ trang web, đảm bảo dữ liệu được sắp xếp đúng thứ tự trong file csv

Bài 2:

Bước 1: đọc dữ liệu từ file results.csv

Bước 2: tính toán trung vị, trung bình và độ lệch chuẩn của mỗi chỉ số cho các cầu thủ trong toàn giải

Dùng thư viện pandas và numpy để tính toán

Bước 3: tìm 3 cầu thủ cao nhất và thấp nhất cho mỗi chỉ số

Bước 4: ghi kết quả vào file results2.csv

Bước 5: vẽ biểu đồ histogram cho từng chỉ số

Dùng thư viện matplotlib để vẽ biểu đồ histogram

Bài 3:

Phân loại cầu thủ bằng Kmeans

Xác định số lượng cụm, ví dụ chọn cụm bằng 3

Áp dụng Kmeans

Nhận xét nhóm

Dùng thuật toán PCA để giảm số chiều dữ liệu xuống còn 2 chiều

Dùng matplotlib để vẽ các điểm dữ liệu trong không gian 2D