

# Mô tả Đồ án

## Bài toán Khí hậu

Ngày 8 tháng 3 năm 2022

## 1 Giới thiệu

Trong đồ án này, các bạn sẽ được làm quen với việc thu thập, xử lý và phân tích dữ liệu thông qua một bài toán cụ thể mà ở đây là các bài toán về đề tài khí hậu. Để đơn giản, ta sẽ sử dụng nguồn dữ liệu từ Cơ quan Quản lý Khí quyển và Đại dương Quốc gia (NOAA) của Mỹ: <https://www.ncdc.noaa.gov/cdo-web/datasets>. Đồ án sẽ được chia thành các phần nhỏ tương ứng với từng bước giải bài toán như sau: Đặt vấn đề; Thu thập và xử lý dữ liệu; Phân tích, đánh giá và kết luận.

## 2 Yêu cầu

### 2.1 Đặt vấn đề (4đ)

Trong phần này, các bạn cần xác định rõ những khía cạnh của bài toán gồm:

- Xác định và hình thức hóa mục tiêu bài toán: Đặt ra câu hỏi cần giải quyết? Từ câu hỏi đó, mô tả nó bằng ngôn ngữ toán học như thế nào?
- Loại (dạng) bài toán: Mô tả (nêu ra các đặc trưng của dữ liệu), Khám phá (tìm các quan hệ chưa được biết trong dữ liệu), Suy luận (đặt ra các giả thuyết và kiểm định các giả thuyết đó), Dự đoán (từ dữ liệu hiện có trong hiện tại và quá khứ để đưa ra dự đoán dữ liệu trong tương lai chưa biết)?
- Đối tượng được chọn cho bài toán: Nhiệt độ, độ ẩm, lượng mưa, tốc độ gió, hướng gió?
- Phạm vi, mức độ, quy mô của bài toán: Theo không gian, dữ liệu trong một thành phố, tỉnh, vùng, quốc gia, toàn thế giới? Theo thời gian, dữ liệu trong một tháng, năm, thập kỷ, thế kỷ, toàn bộ lịch sử?

Ví dụ một bài toán: Chứng minh khí hậu nóng lên → Kiểm định giả thuyết biến nhiệt độ trung bình năm toàn cầu và biến thời gian theo năm là tương quan → Loại bài toán suy luận → Chọn các biến liên quan đến nhiệt độ → Lấy dữ liệu ở phạm vi toàn cầu theo năm trong suốt toàn bộ lịch sử.

**Yêu cầu:** Trong báo cáo, các bạn nêu và thực hiện giải quyết **hai** bài toán. Mỗi bài thuộc một dạng bài toán khác nhau.

**Lưu ý:** Các bạn tự chọn **hai** bài toán trong khuôn khổ nguồn dữ liệu khí hậu của NOAA.

## 2.2 Thu thập và xử lý dữ liệu (4đ)

Từ đối tượng và phạm vi đã nêu ở trên, các bạn trích xuất và tổng hợp lại dữ liệu theo yêu cầu của bài toán mà mình đặt ra. Các bạn cũng có thể thực hiện các bước tiền xử lý dữ liệu nếu thấy cần thiết.

**Yêu cầu:** Trong báo cáo, các bạn nêu vắn tắt các bước các bạn đã thực hiện.

## 2.3 Phân tích, đánh giá và kết luận (4đ)

Từ dạng bài toán đã nêu ở trên, các bạn chọn ra các mô hình phù hợp. Sau khi chạy các mô hình, các bạn phân tích kết quả thu được từ các mô hình đã chọn và đưa ra kết luận cuối cùng.

**Yêu cầu:** Trong báo cáo, các bạn nêu ra các mô hình sẽ dùng, cơ sở lý thuyết hoặc lý do tại sao chọn các mô hình đó, lập bảng kết quả, cho biết ý nghĩa của các kết quả này là gì và đưa ra kết luận.

## 3 Các quy định khác

- Do đề án yêu cầu làm 2 bài toán nên cần tách ra làm 2 thư mục, mỗi thư mục làm 1 bài toán. Khi đó, mỗi thư mục sẽ gồm có 3 phần: Report (phần báo cáo), Data (phần dữ liệu), Source (phần mã nguồn).
- Trong phần báo cáo, các bạn cần nêu rõ họ tên, MSSV, bài toán mình đang làm là gì và trình bày theo như mô tả trong phần Yêu cầu ở trên.
- Trong phần dữ liệu, các bạn chỉ nộp phần dữ liệu có liên quan đến bài toán. Nếu dữ liệu quá lớn không thể nộp trên moodle thì các bạn có thể nộp trên Google Drive, OneDrive hoặc những nơi lưu trữ tương tự, rồi chèn đường dẫn vào.
- Trong phần mã nguồn, các bạn nộp toàn bộ mã nguồn mà các bạn sử dụng cho bài toán.
- Đây là bài làm cá nhân, không làm nhóm dưới mọi hình thức.
- Trình bày bằng tiếng Việt. Đối với các thuật ngữ, các bạn có thể sử dụng tiếng Anh.
- Thang điểm cho đề án này là 12/10 điểm. Phần điểm dư ra (nếu có) sẽ không được tính.
- Nghiêm cấm các hành vi sao chép bài làm của nhau, của sinh viên các khóa trước, các lớp khác, nhờ người làm hộ. Được phép trích dẫn giống như các ấn phẩm, bài báo khoa học, nhưng nghiêm cấm các hành vi đạo văn từ các nguồn tài liệu. Hình thức xử phạt tối thiểu là bị 0 điểm đề án.