

I. ĐỀ BÀI TIỂU LUẬN

Trong thời gian qua ở nước ta đã có rất nhiều sự thay đổi trong xã hội để thích ứng với tình hình dịch COVID phức tạp. Có thể nhận thấy rõ ràng là ảnh hưởng của COVID-19 lên nhiều mặt của xã hội, gây ra hậu quả không nhỏ. Do đó đề tài này có mục tiêu chính là phân tích diễn biến bệnh và dự đoán tình hình dịch bệnh COVID-19 trong thời gian tới. Một số nhiệm vụ được đặt ra trong đề tài như sau:

1. Thu thập dữ liệu tình hình bệnh dịch của các tỉnh thành trong nước, tất cả các thông số đều được quan tâm.
2. Đưa ra phân tích và diễn biến trong thời gian từ tháng 1/2021 đến tháng 3/2022. Các con số thống kê có thỏa phân phối nào không? (e.g. normal, poisson, gamma, exponential, etc)
3. Xây dựng 1 mô hình dự đoán về tình hình dịch bệnh trong khoảng 1 tuần - 3 tháng cho
 - o Toàn lãnh thổ Việt Nam
 - o 3 tỉnh có nhiều ca nhiễm nhất, tính tới thời điểm làm thống kê, mà không phải thành phố thuộc trung ương.
 - o 5 thành phố trực thuộc trung ương

Lưu ý:

- Có thể sử dụng nhiều mô hình toán (e.g. linear regress) khác nhau hoặc sử dụng các công cụ học máy, học sâu....
 - Có thể chia khoảng thời gian tùy chọn để kiểm tra mô hình, ví dụ như train trong khoảng tháng 1/2021-11/2021 và dự đoán cho khoảng thời gian 12/2021-2/2022, hoặc train toàn bộ data và dự đoán cho khoảng thời gian 1 tuần tới 3 tháng tới.
 - Mục tiêu chính là kiểm tra xem có khả năng xây dựng mô hình dự đoán vào bất kì khoảng thời gian nào hay không. Phân tích các kết quả đạt được một cách chi tiết.
4. Kiểm định, phân tích mô hình, các yếu tố ảnh hưởng, và khả năng sử dụng mô hình chéo giữa các tỉnh thành phố. Ví dụ như dùng data của HCM dự đoán cho HN.
 5. Đánh giá dữ liệu và khả năng dự đoán của mô hình vào thời điểm trước và sau khi tiêm vaccin, vaccin liệu có ảnh hưởng gì tới các thông số như ca lây nhiễm, người chết....
 6. Theo anh chị, những ngành kinh tế nào bị ảnh hưởng nặng nhất do dịch bệnh (tối đa 3 ngành), hãy dùng các dữ liệu thu thập được cùng với phân tích để chứng minh điều này, đánh giá khả năng phục hồi dựa vào thông tin từ dữ liệu (có thể sử dụng dữ liệu chứng khoán).

II. MỘT SỐ LINK THAM KHẢO

[1]

https://covid19.who.int/?gclid=CjwKCAjwjqT5BRAPEiwAJIBuBa9_ei3XT8h73c_TCRV3_HHfAMdTm1OGByxNQr1ik2huOwExJP8jbBoCdSYQAvD_BwE

[2] <https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge>

[3] <https://covid19.gov.vn/>

III. YÊU CẦU

1. Lập nhóm gồm tối đa 5 người và ít nhất 4.
2. Chọn ngẫu nhiên nhóm để báo cáo và nội dung báo cáo có thể là toàn bộ hay 1 phần câu hỏi. Ngày báo cáo sẽ được thông báo trước. Thời gian báo cáo không quá 25 phút cho 1 nhóm.
3. Nghiêm cấm mọi hành vi sao chép, trao đổi báo cáo.
4. Nộp mọi tài liệu thu thập được + 1 slide báo cáo + 1 report + các demo liên quan. Chỉ duy nhất nhóm trưởng nộp và duy nhất 1 file zip chứa tất cả.
5. Trong slide báo cáo bắt buộc phải để % nỗ lực của từng thành viên trên tổng thể (tổng nỗ lực là 100%) ở **trang đầu tiên**.
6. Thời gian hiện thực 4 tuần. Deadline ngày 15/4/2022 23h55 phút
7. Bắt buộc mỗi nhóm chuẩn bị ít nhất 2 câu hỏi cho các nhóm khác về nội dung bài báo cáo.