

TRƯỜNG ĐẠI HỌC NGUYỄN TẤT THÀNH
KHOA CÔNG NGHỆ THÔNG TIN

CUỘC THI TRÍ TUỆ NHÂN TẠO
KHOA CÔNG NGHỆ THÔNG TIN NĂM 2022

ĐỀ THI VÒNG CHUNG KẾT

Thời gian: từ 8h30 đến 11h30 ngày 23/07/2022

I. Bài toán: Xác định mức độ lây lan và dự báo số người bị nhiễm, người tử vong và đã âm tính virus Covid-19.

Trong vòng loại này, chúng tôi cung cấp cho các nhóm tập tin **covid_19_data.csv** chứa nhiều thông tin liên quan đến virus Covid-19.

Các nhóm hãy thực hiện việc rút trích thông tin liên quan đến virus Covid-19 về một quốc gia nào đó trên dữ liệu này, đồng thời thực hiện các yêu cầu sau:

1. Xác định mức độ lây lan virus Covid-19 đến quốc gia khác

- Dựa trên bộ dữ liệu mà các nhóm đã rút trích các nhóm hãy xây dựng chương trình dự báo mức độ lây lan virus Covid-19 đến các nước khác.
- Giả sử rằng, cứ sau 14 ngày, những người bị nhiễm sẽ khỏi bệnh nếu không bị tử vong.

- Yêu cầu:

- Sử dụng Logic mờ để giải quyết bài toán xác định mức độ lây lan
- Các nhóm đề xuất mức độ lây lan và giải thích ý nghĩa trước khi áp dụng để mờ hóa
- Các nhóm đề xuất các cấp độ của tập mờ và giải thích ý nghĩa

- Hướng dẫn:

- Nên phân tích và xây dựng bài toán theo 05 giai đoạn
- Căn cứ vào dữ liệu rút trích để xác định bộ tiên đề (Antecedent) phù hợp

2. Dự báo số người bị nhiễm, người tử vong và đã âm tính virus Covid-19

- Dựa trên bộ dữ liệu mà các nhóm đã rút trích các nhóm hãy xây dựng chương trình dự báo số người bị nhiễm, người tử vong và đã âm tính virus Covid-19 cho 04 ngày tiếp theo.

- Yêu cầu:

- Ngoài việc đưa ra kết quả dự đoán, các nhóm cần đưa ra độ sai lệch của dự báo
- Giải thích việc lựa chọn phương pháp hồi quy
- Kết quả xuất được thể hiện dưới dạng: số lượng (+/- độ sai lệch), xem ví dụ dưới đây

Ngày thứ	Số người nhiễm	Số người tử vong	Số người đã âm tính
1	532 (+/- 5)	8 (+/- 1)	348 (+/- 4)
2	501 (+/- 5)	6 (+/- 1)	408 (+/- 4)
3	378 (+/- 5)	0 (+/- 1)	255 (+/- 4)
4	491 (+/- 5)	3 (+/- 1)	317 (+/- 4)

- Hướng dẫn:

- Các nhóm có thể sử dụng các phương pháp hồi quy cổ điển để dự báo kết quả
- Có thể sử dụng dữ liệu ngày trước là cơ sở để dự đoán cho ngày tiếp theo hoặc sử dụng các phương pháp nội suy (tuyến tính, sai phân tiến, ta tuyến, nội suy đa thức, kriging, số gần nhất, Lagrange, ...) để tính ra kết quả, sau đó dự báo độ sai lệch của phương pháp nội suy.

II. Bộ dữ liệu - Dataset

Bộ dữ liệu gồm 1 file: **covid_19_data.csv**, cụ thể như sau:

- Bộ dữ liệu chứa nhiều thông tin liên quan đến virus Covid-19 được mô tả chi tiết trong **covid_19_data** với các trường dữ liệu như sau:

Trường	Kiểu dữ liệu	Mô tả
Số thứ tự	Integer	Thứ tự cập nhật dữ liệu
Ngày ghi nhận	String	Ngày ghi nhận các ca bị nhiễm, tử vong và đã âm tính do virus Covid-19
Tiểu ban/tỉnh	String	Tên tiểu ban hoặc tỉnh
Quốc gia/vùng	String	Tên quốc gia hoặc vùng lãnh thổ
Thời gian ghi nhận cuối cùng	Date/Time	Thời gian ghi nhận các ca bị nhiễm, tử vong và đã âm tính do virus Covid-19 cuối cùng trong ngày
Số người nhiễm	Integer	Số người nhiễm virus Covid-19 tính đến hết

		Thời gian ghi nhận cuối cùng
Tử vong	Interger	Số người ủ vong do virus Covid-19 tính đến hết Thời gian ghi nhận cuối cùng
Âm tính	Interger	Số người âm tính với virus Covid-19 tính đến hết Thời gian ghi nhận cuối cùng

Bảng 1: Bộ dữ liệu liên quan đến virus Covid-19

III. Yêu cầu

Toàn bộ quy trình bao gồm nhiều bước từ tiền xử lý đến đánh giá. Các nhóm được yêu cầu khai phá dữ liệu và báo cáo công việc của mình từng bước sau đây:

1. Xử lý dữ liệu - Data Imputation (2 điểm):

Lựa chọn những dữ liệu đáp ứng yêu cầu.

2. Khám phá dữ liệu (Data Exploration) và Trích xuất đặc trưng (Feature Extraction) (4 điểm):

- Nhóm cần khám phá dữ liệu để hiển thị một số thông tin thống kê và phân tích của tập dữ liệu đã lựa chọn.
- Hãy đề xuất cách trích xuất đặc trưng từ bộ dữ liệu đã cho, cung cấp lý do và giải thích cách làm của các nhóm.

Hướng dẫn:

- Các nhóm có thể sử dụng các biểu đồ để biểu diễn các đặc trưng của dữ liệu.
- Phương pháp Ranking.
- Biểu diễn sự phân bố dữ liệu.
- ...

3. Xác định mức độ lây lan (5 điểm):

Hãy xây dựng chương trình dự báo mức độ lây lan virus Covid-19 đến các nước khác.

4. Dự đoán- Prediction (7 điểm):

Hãy đề xuất bộ Train/Test và các độ đo phù hợp, sau đó tiến hành dự báo số người bị nhiễm, người tử vong và đã âm tính virus Covid-19 cho 04 ngày tiếp theo.

5. Thảo luận (1 điểm):

Đây là một nhiệm vụ của ngành khoa học dữ liệu và Trí tuệ nhân tạo khi giải quyết những thử thách trong nhu cầu thực tế hiện nay, các nhóm có thể ra ý kiến, giải pháp của mình những vấn đề liên quan. Các nhóm vui lòng cung cấp các thảo luận, ý tưởng của nhóm về việc thu thập các đặc trưng mới ột cách hiệu quả để tăng hiệu suất của mô hình.

IV. Báo cáo

- Báo cáo của các nhóm phải được sử dụng ngôn ngữ tiếng Việt.

- Báo cáo phải có cấu trúc rõ ràng cho từng yêu cầu ở trên, với mỗi yêu cầu đều bao gồm: phần giới thiệu, cách tiếp cận, đánh giá, thảo luận, v.v. phù hợp để các nhóm hoàn thành các yêu cầu ở trên một cách hiệu quả (*theo mẫu đính kèm*).

- Báo cáo trình bày dưới dạng Power Point.

V. Nộp bài

Vui lòng gửi báo cáo của nhóm (ở định dạng pdf) và mã (bằng Python) tại địa chỉ <http://hackathon.daihocsonntt.edu.vn/> trước **11h30 ngày 23/07/2022**.

**BAN TỔ CHỨC - Cuộc thi Trí tuệ nhân tạo - Khoa Công nghệ thông tin
Năm 2022**

**TM. BAN TỔ CHỨC CUỘC THI
TRƯỞNG BAN TỔ CHỨC
TRƯỞNG KHOA CNTT**



TS. Nguyễn Kim Quốc