Khóa học Big Data

Capstone Project   
Kế hoạch hành động

ⓒ2023 SAMSUNG. All rights reserved.

Tài liệu này là tài sản văn học được bảo vệ bởi luật bản quyền nên việc in lại và sao chép mà không được phép đều bị nghiêm cấm.

Để sử dụng tài liệu này ngoài chương trình giảng dạy của Samsung Innovation Campus, bạn phải nhận được sự đồng ý bằng văn bản từ chủ sở hữu bản quyền.

| Khóa học | Khóa học Big Data |
| --- | --- |
| Tên nhóm | BD01 - Nhóm 4 |
| Nhóm trưởng/  Thành viên | Võ Thị Ngọc Châu /  Lê Quang An, Nguyễn Minh Gia Bảo, Phạm Thành Đạt, Đặng Vạn Như Ý, Hà Vân Yến |
| Tên đề tài | COVID-19 Analysis and Prediction |
| Mục tiêu |  |
| Xây dựng mô hình dự đoán khả năng mắc COVID-19 của bệnh nhân dựa vào các yếu tố đầu vào cụ thể. | |
| Tóm tắt |  |
| Trực quan hóa giữa các yếu tố đầu vào và khả năng mắc bệnh để tìm ra sự tương quan giữa chúng, từ đó lựa chọn các thuật toán để xây dựng mô hình. Tiến hành đánh giá và lựa chọn mô hình tốt nhất. | |
| Phương pháp |  |
| * Sử dụng thư viện pandas load tập dữ liệu vào data frame, các phương thức có sẵn để tiền xử lý dữ liệu. * Sử dụng thư viện seaborn, matplotlib để trực quan hóa dữ liệu. * Sử dụng thư viện … để xây dựng các mô hình | |

| Dữ liệu |  |
| --- | --- |
| * Tìm kiếm các tập dữ liệu phù hợp trên Kaggle và Github * Tiến hành tiền xử lý, trực quan và xây dựng mô hình với tập dữ liệu phù hợp. | |
| Kết quả mong muốn |  |
| * Xây dựng mô hình dự đoán khả năng mắc COVID-19 của bệnh nhân dựa vào các yếu tố đầu vào cụ thể với độ chính xác càng cao càng tốt. * Từ đó, ngoài biết được một người có mắc bệnh hay không, ta còn có thể biết được cách phòng tránh mắc bệnh dựa vào đánh giá tương quan giữa các yếu tố với biến phụ thuộc | |
| Vai trò của các thành viên |  |
| * Võ Thị Ngọc Châu: nhóm trưởng, lên kế hoạch, phân chia công việc, hỗ trợ các bạn thực hiện deadline (nếu có), đánh giá kết quả và góp ý chỉnh sửa, thực hiện các nhiệm vụ được giao. * Lê Quang An: tìm kiếm dữ liệu, góp ý xây dựng dự án và thực hiện các nhiệm vụ được giao. * Nguyễn Minh Gia Bảo: tìm kiếm dữ liệu, góp ý xây dựng dự án và thực hiện các nhiệm vụ được giao. * Phạm Thành Đạt: tìm kiếm dữ liệu, góp ý xây dựng dự án và thực hiện các nhiệm vụ được giao. * Đặng Vạn Như Ý: tìm kiếm dữ liệu, góp ý xây dựng dự án và thực hiện các nhiệm vụ được giao. * Hà Vân Yến: tìm kiếm dữ liệu, góp ý xây dựng dự án và thực hiện các nhiệm vụ được giao. | |

| Tóm tắt kế hoạch thực hiện |  |
| --- | --- |
| 1. Xác định đề tài và thu thập/tìm kiếm dữ liệu: 20/06/2024 - 04/07/2024 2. Tiền xử lý dữ liệu: 05/07/2024 - 12/07/2024 3. Trực quan hóa dữ liệu: 13/07/2024 - 20/07/2024 4. Huấn luyện mô hình bằng nhiều thuật toán: 21/07/2024 - 30/07/2024 5. Đánh giá và lựa chọn mô hình: 31/07/2024 - 07/08/2024 6. Viết báo cáo + Thiết kế slide + Thuyết trình thử: 08/08/2024 - 14/08/2024 | |
| Nhận xét và đánh giá của giảng viên |  |
|  | |