

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN



ĐỒ ÁN MÔN HỌC
LẬP TRÌNH PYTHON CHO MÁY HỌC

KHOA: KHOA HỌC MÁY TÍNH

HOMEWORK: ĐĂNG KÝ ĐỒ ÁN

GV hướng dẫn: TS. Nguyễn Vinh Tiệp

Nhóm thực hiện:

1. Trương Thành Thắng – 20521907
2. Ngô Ngọc Sương – 20521852
3. Trần Văn Lực – 20521587
4. Huỳnh Viết Tuấn Kiệt – 20521494
5. Nguyễn Đức Anh Phúc – 20520276
6. Ngô Văn Tấn Lưu – 20521591
7. Nguyễn Văn Toàn - 20522028

01.01 Bài toán:

01.01.01 Bài toán phân lớp: Dự đoán bệnh nhân nguy kịch bởi covid

Bệnh vi-rút corona (Covid-19) là một bệnh truyền nhiễm gây ra bởi một loại vi-rút corona mới được phát hiện. Hầu hết những người bị nhiễm vi-rút Covid-19 đều phát triển bệnh hô hấp từ nhẹ đến trung bình và hồi phục mà không cần điều trị đặc hiệu. Người lớn tuổi và những người mắc các bệnh nền như bệnh tim mạch, tiểu đường, bệnh hô hấp mãn tính và ung thư có nhiều khả năng phát triển bệnh nặng hơn.

Một trong những vấn đề chính mà các nhà cung cấp dịch vụ chăm sóc sức khỏe phải đối mặt trong suốt đại dịch là thiếu nguồn lực y tế và kế hoạch phù hợp để phân bổ các nguồn lực đó một cách hiệu quả. Trong những thời điểm khó khăn này, việc có thể dự đoán những nguồn lực mà một cá nhân có thể cần khi họ xét nghiệm dương tính hoặc thậm chí trước khi họ xét nghiệm dương tính sẽ giúp các cơ quan chức năng lên lịch các nguồn lực cần thiết để cứu sống bệnh nhân.

Vì vậy, trong đồ án này nhóm em sẽ xây dựng các mô hình máy học để dự đoán liệu bệnh nhân covid-19 có nguy kịch hay không dựa trên các triệu chứng, tình trạng và tiền sử bệnh hiện tại của họ.

Mô hình bài toán:



Hình 1: Mô hình bài toán. Output: 1: Bệnh nhân nguy kịch, 0: Bệnh nhân không nguy kịch

01.01.02 Dataset:

Covid-19 Dataset¹: Là bộ dữ liệu mang thông tin về triệu chứng, trạng thái và tiền sử bệnh của các bệnh nhân Covid-19 được cung cấp bởi chính phủ Mexico, bao gồm 1.048.576 điểm dữ liệu với 21 đặc trưng lần lượt là: sex, age, classification, patient type, pneumonia, pregnancy, diabetes, copd, asthma, inmsupr, hypertension, cardiovascular, renal chronic, other disease, obesity, tobacco, usmr, medical unit, intubed, icu, death.

01.02 Protocol đánh giá:

Bộ dữ liệu chia tỉ lệ *train: test* là: 7: 3, được chia 3 lần để thực nghiệm, *random_state* lần lượt là: 7, 14, 21.

Mỗi mô hình được thực nghiệm 3 lần trên 3 bộ dữ liệu đã chia trên, sau đó tính trung bình cộng hiệu suất trên mỗi mô hình.

Độ đo: *f1_score*, *accuracy*.

¹ <https://www.kaggle.com/datasets/meirizri/covid19-dataset>

01.03 Phân công:

STT	Họ và tên	MSSV	Vai trò	Mô hình
1	Trương Thành Thắng	20521907	Nhóm trưởng	Linear Discriminative Analysis
2	Ngô Ngọc Sương	20521852	Thành viên	Logistic Regression
3	Trần Văn Lực	20521587	Thành viên	K-nearest Neighbor
4	Huỳnh Viết Tuấn Kiệt	20521494	Thành viên	Support Vector Machines
5	Nguyễn Đức Anh Phúc	20520276	Thành viên	Neural Networks
6	Ngô Văn Tấn Lưu	20521591	Thành viên	Decision Trees
7	Nguyễn Văn Toàn	20522028	Thành viên	Naïve Bayes