Trả lời câu hỏi :

1. Công nghệ bài dùng :

2 phương pháp :

+ phương pháp k-fold cross validation lấy mẫu để đánh giá

+ phương pháp xáo trộn dữ liệu đầu vào .

1 thuật toán :

+ svm

1. Svm thuộc loại học máy nào :

+ svm là thuật toán thuộc loại học có giám sát.

1. Thống kê như nào lựa chọn thuộc tính dữ liệu như nào :

+ dự vào các thuộc tính ảnh hưởng trực tiếp tới kết quả.

+ dự trên cách tiếp cận thuộc tính của người bệnh.(dựa vào sổ khám sức khỏe của người bệnh)

(ở trong sổ này thì 1 cái khi đi khám sức khỏe đa phần sẽ đều kiểm tra đó là xét nghiệm máu, hay xét nghiệm sinh hóa, .....)

+ dự trên nghiên cứu của “hiệp hội tim mạch” để lựa chọn (dự án đưa AI vào dự đoán đột quỵ ở Anh bắt đầu từ năm 2017 cũng dựa trên các thuộc tính này và mới đc đưa ra cống chúng 27/12/2022 mới đây).

1. Đầu vào là các thuộc tính gì ý nghĩa ? (10 thuộc tính đầu vào được chọn )

+ Tuổi (người già có nguy cơ đột quỵ cao nhất là sau 55 tuổi)

+ Giới tính (cân nặng / bình phương chiều cao)

+ Chỉ số khối

+ nhịp tim

+ huyết áp

+ đường máu (glucose : xác định bệnh tiểu đường ảnh hưởng biến chứng đến nhiều cơ quan đặc biệt là thận mạch máu vv.)

+ Cholesterol ( là một chất béo có trong máu, Nó có thể gây hẹp hoặc tắc nghẽn mạch máu dẫn đến nhồi máu cơ tim hoặc đột quỵ.)

+ **Triglycerides** ( là một loại chất béo khác. Nếu Triglycerides cao sẽ làm tăng nguy cơ nhồi máu cơ tim hoặc đột quỵ.)

+ Tiền sử bị bệnh

+ rankin : (thang đo nhận thức)

1. Tại sao chọn SVM mà ko chọn cái khác ?

* svm là phương pháp phân loại nhị phân .
  + - SVM rất hiệu quả để giải quyết bài toán dữ liệu có số chiều lớn.
    - SVM là phương pháp phân lớp nhanh.
    - SVM có hiệu suất tổng hợp tốt và hiệu suất tính toán cao.
    - SVM tiết kiệm bộ nhớ khi sử dụng.

Và đặc biệt là nó được sử dụng nhiều trong mô hình chẩn đoán bệnh

* đây là phương pháp SVM luôn cố gắng cực đại hóa margin là khoảng cách từ mặt phẳng tới biên do đó SVM có thể giảm thiểu việc phân lớp sai (misclassification) đối với điểm dữ liệu mới đưa vào.
* Với bài toán dự đoán bệnh đột quỵ này thì chúng ta ko biết sác xuất ảnh hưởng của từng thuộc tính tới kết quả là bao .nên dùng svm là hợp lý vì svm cũng ko tính  được xác suất hiện của một thuộc tính trong một nhóm

1. Dựa vào đâu bảo cái này tốt hơn cái kia .

+ Sau khi huấn luyện xong mô hình , dùng mô hình SVM đã huấn luyện để dự đoán nhãn cho tập kiểm thử thực tế. sử dụng hàm accuracy\_score để tính toán điểm chính xác cho một tập hợp các nhãn được dự đoán so với các nhãn thực tế .

1. Sử dụng k\_fold cross có ý nghĩ j :

+ giúp chúng ta đánh giá một model đầy đủ và chính xác hơn khi chúng ta có một tập dữ liệu không lớn. Để sau đó chúng ta đưa ra quyết định model đó có phù hợp với dữ liệu, bài toán hiện tại hay không để mà đưa ra hành động tiếp theo.

1. Sử dụng xáo trộn dữ liệu ngẫu nhiên hay xáo trộn dữ liệu trong danh sách một cách ngẫu nhiên.
2. Tại sao đoạn code này có ý nghĩa j .
3. ....