

PHÂN TÍCH VÀ DỰ ĐOÁN NGUY CƠ BỆNH SUY TIM BẰNG MÔ HÌNH XGBOOST

SINH VIÊN: BÙI ĐỨC ANH - GVHD: TH.S NHỮ VĂN KIẾN

Giới thiệu

Đề tài tập trung vào ứng dụng mô hình XGBoost để dự đoán nguy cơ bệnh suy tim dựa trên các chỉ số y tế nhằm đưa ra kết quả với độ chính xác và độ tin cậy cao.

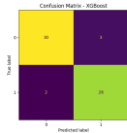
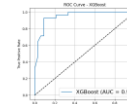
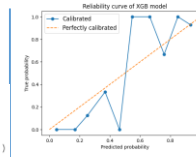
Quy trình triển khai

- Quy trình: Thu thập dữ liệu → Tiền xử lý dữ liệu → Huấn luyện mô hình → Đánh giá kết quả → Tinh chỉnh tham số → Ứng dụng dự đoán nguy cơ mắc bệnh suy tim.
- XGBoost với nhiều ưu điểm: độ chính xác cao, vượt trội hơn so với các mô hình máy học máy cũ, có khả năng mở rộng và linh hoạt.

Kết quả và Ý nghĩa

0	0.90	0.85	0.88	33
1	0.83	0.89	0.86	28
accuracy			0.87	61
macro avg	0.87	0.87	0.87	61
weighted avg	0.87	0.87	0.87	61

Medical Metrics of XGB
Sensitivity: 0.928754328754286 (95% CI: 0.891818181818182 - 1.0)
Specificity: 0.900000000000000 (95% CI: 0.7948022928852495 - 1.0)
PPV: 0.89655724137915 (95% CI: 0.774293548878968 - 1.0)
NPV: 0.9375 (95% CI: 0.843623991935484 - 1.0)
F1: 0.9121807817543859 (95% CI: 0.82352941217647858 - 0.9824718886285459)



- Kết quả cho thấy mô hình đạt độ chính xác cao và hiệu suất vượt trội (accuracy: 0.92, AUC: 0.95), khả năng phát hiện cả dương tính và loại bỏ cả âm tính giả tốt (Sensitivity: 0.928, NPV: 0.937)
- Một số hạn chế nhỏ vẫn tồn tại khi mô hình bỏ sót một vài ca bệnh, tuy nhiên có thể được cải thiện bằng việc tối ưu tham số và mở rộng dữ liệu huấn luyện.
- Kết quả này khẳng định XGBoost là một mô hình hiệu quả và đáng tin cậy trong hỗ trợ chẩn đoán sớm suy tim, góp phần giảm thiểu tỷ lệ bỏ sót ca bệnh và nâng cao độ chính xác trong dự đoán lâm sàng.

Giao diện ứng dụng

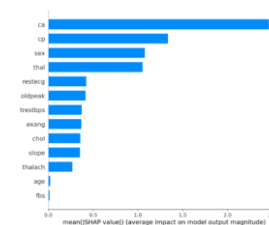
Dự đoán nguy cơ mắc bệnh tim

Form input fields for patient data:

- Tuổi: 50
- Giới tính (0=nam, 1=nữ): 0
- Loại huyết áp (0=0, 1=1): 0
- Huyết áp tâm thu (mmHg): 130
- Cholesterol (mg/dL): 200
- Hàm lượng đường huyết (100 mg/dL (11.0, 11.0 mmol/L)): 0
- Khi quả tim đập (0=0, 1=1): 0
- Mức tiêu thụ năng lượng (100): 100
- Số lượng mạch vành (0=0, 1=1): 0
- Thalassemia (0=normal, 1=first defect, 2=second defect): 0

Giải thích kết quả dự đoán

Biểu đồ dưới đây cho thấy mức độ ảnh hưởng của từng đặc trưng.



Output section of the application:

- Đã đoán: ☐
- Nguy cơ mắc bệnh tim thấp (số suất: 0.13)
- Reset

Các chức năng chính

- Dự đoán nguy cơ mắc bệnh suy tim dựa trên các chỉ số y tế của bệnh nhân.
- Giải thích kết quả dự đoán dựa trên mức độ ảnh hưởng của từng đặc trưng.
- Tải về file kết quả dự đoán

Kết luận và hướng phát triển

- Đề tài đã xây dựng và phát triển thành công mô hình XGBoost với độ chính xác và tính ổn định cao, vượt trội hơn so với các mô hình truyền thống, góp phần hỗ trợ chẩn đoán và ra quyết định sớm.
- Hướng phát triển: Mở rộng tập dữ liệu huấn luyện, phát triển giao diện trực quan, nghiên cứu các phương pháp để tối ưu hơn độ chính xác dự đoán.