

ĐỒ ÁN MÔN HỌC KHAI THÁC DỮ LIỆU

LỚP IS252.O22.HTCL

1. Tên nhóm: HOHO

2. Tên đề tài: DIABETES PREDICTION

3. Tóm tắt nội dung đề tài (không quá 10 dòng):

Đề tài này nhằm phát triển mô hình dự báo chính xác nguy cơ mắc bệnh tiểu đường. Các dữ liệu liên quan như tuổi, giới tính, cholesterol, BMI, hút thuốc, tiền sử bệnh tim mạch, hoạt động thể chất, huyết áp và chế độ ăn sẽ được thu thập và làm sạch. Quá trình này sẽ bao gồm xử lý các giá trị khuyết, ngoại lệ và chuẩn hóa dữ liệu, đồng thời thực hiện phân tích thăm dò để hiểu rõ hơn về các đặc trưng.

Các kỹ thuật học máy tiên tiến như Decision Tree, Random Forest, KNN, SVM, XGBoost sẽ được áp dụng. Kết quả từ mô hình này sẽ giúp cải thiện việc chẩn đoán sớm, phòng ngừa và điều trị bệnh tiểu đường một cách hiệu quả hơn. Các mô hình này sẽ được xây dựng, sau đó so sánh hiệu suất trên các tiêu chí như độ chính xác, độ nhạy và độ đặc hiệu. Mô hình tốt nhất sẽ được lựa chọn dựa trên kết quả so sánh.

4. Thông tin nhóm:

STT	MSSV	Họ tên	Mô tả công việc được giao	Tổng công việc được phân (% công việc được phân công)	Khả năng hoàn thành công việc được giao (% hoàn thành)	Vai trò	Ký tên
1	21520596	Trần Thị Kim Anh	Tiền xử lý dữ liệu Mô hình SVM Báo cáo	25%	100%	Thành viên	
2	21521049	Hồ Quang Lâm	Tiền xử lý dữ liệu Mô hình Random Forest Báo cáo	25%	100%	Thành viên	
3	21521586	Lê Thị Lệ Trúc	Tiền xử lý dữ liệu Mô hình KNN Báo cáo	25%	100%	Nhóm trưởng	
4	21521882	Lê Minh Chánh	Mô hình Decision Tree Mô hình XGBoost Báo cáo	25%	100%	Thành viên	