**ĐỒ ÁN MÔN HỌC KHAI THÁC DỮ LIỆU**

**LỚP IS252.N21.HTCL**

1. **Tên nhóm:** Nhóm 8
2. **Tên đề tài:** Chuẩn đoán bệnh tim
3. **Tóm tắt nội dung đề tài (không quá 10 dòng):**

Ở đề tài này tập trung vào phân tích dữ liệu Personal Key Indicators of Heart Disease. Sử dụng Python, Jupyter Notebook và Google Colaboratory, nhóm áp dụng kiến thức về tiền xử lí, phân tích và trực quan hóa dữ liệu. Thực nghiệm được thực hiện trên các mô hình như RandomForest, DecisionTree, K-Neighbors, Naïve Bayes và Logistic Regression, đánh giá dựa trên độ đo Recall để tìm ra mô hình tốt nhất. Mục tiêu là xây dựng một mô hình chẩn đoán bệnh tim nhanh nhất và chính xác nhất dựa trên triệu chứng và chỉ số y tế.

**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 thanh công việc được giao (% hoàn thanh) | Vai trò | Ký tên |
| 1 | 20521865 | Dương Bảo Tâm | - Chạy thuật toán Decision Tree  -Tiền xử lý dữ liệu  - Phân tích | 25% | 100% | Trưởng nhóm |  |
| 2 | 20521271 | Nguyễn Trần Lưỡng Hà | - Chạy thuật toán Naïve Bayes  - Chỉnh sửa báo cáo  - Mô tả dữ liệu bài toán | 25% | 100% |  |  |
| 3 | 20521255 | Phạm Thái Mỹ Duyên | - Chạy thuật toán Logistic Regression  - Tiền xử lý dữ liệu  - Phân tích | 25% | 100% |  |  |
| 4 | 20521260 | Lê Thanh Giang | - Chạy thuật toán Random Forest  - Chỉnh sửa báo cáo  - Mô tả dữ liệu bài toán | 25% | 100% |  |  |