**ĐỒ ÁN MÔN HỌC DATA MINING**

**LỚP IS252.J11.HTCL**

1. **Tên nhóm: HTP**
2. **Tên đề tài: Dữ liệu tổng điều tra dân số 1994**
3. **Tóm tắt nội dung đề tài (không quá 10 dòng):**

Chọn dataset, giới thiệu tổng quan dataset. Nguồn cung cấp dữ liệu và hướng dẫn tải dữ liệu.

Phát biểu bài toán cho dữ liệu đã chọn: Dự đoán một người có đặt tính như thế nào thì có mức lương hằng năm nhiều hơn 50.000$.

Tiến hành tiền xử lý dữ liệu: thực hiện làm sạch dữ liệu tích hợp dữ liệu, thu giảm rút gọn dữ liệu.

Mô tả chi tiết các thuộc tính của dữ liệu sau khi tiền xử lý dữ liệu.

Giới thiệu tổng quan về các công cụ đã sử dụng: weka

Giới thiệu thuật toán mà nhóm đã chọn. Tiền hành thực hiện dữ liệu với thuật toán đã chọn. Đọc kết quả từng thuật toán và so sánh kết quả từng tệp train - test.

Thuât toán đã thực hiện: Cây quyết định J48, NaiveBayes, K-Means

So sánh kết quả chung giữa các thuật toán.

Đưa ra kết luận: những kết quả đạt được, những hạn chế chưa giải quyết được, hướng phát trong tương lai và tài liệu tham khảo trong quá trình thực hiện, bảng phân công nhiệm vu.

1. **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 | 15520214 | Nguyễn Hoàng Hiệp | Chọn dataset,  Nghiên cứu tính cần thiết để khai thác dữ liệu,  Mô tả bài toán  Tiền xủ lý dữ liệu  Chọn thuật toán  Thực hiện dữ liệu với thuật toán đã chọn  Giải thích kết quả thu được Viết báo cáo | 50% | 100% | Nhóm trưởng |  |
| 2 | 15520844 | Nguyễn Trường Thịnh | Nghiên cứu tính cần thiết để khai thác dữ liệu  Chọn thuật toán,  Thực hiện dữ liệu với thuật toán đã chọn,  Giải thích kết quả thu được | 30% | 100% | Thành viên |  |
| 3 | 15520639 | Lê Hữu Phúc | Mô tả bài toán,  Nghiên cứu tính cần thiết để khai thác dữ liệu,  Giải thích kết quả thu được  Viết báo cáo | 20% | 100% | Thành viên |  |