1. **Thông tin về thành viên nhóm:**

- Trần Thị Mai Anh - 23001497

- Đỗ Thị Như Quỳnh - 23001556

- Hồ Huyền Trang - 23001565

1. **Phân công:**

|  |  |
| --- | --- |
| **Thành viên** | **Nhiệm vụ** |
| Trần Thị Mai Anh | Lý thuyết và demo mô hình KNN, phân cụm bằng Kmean, DBSCAN |
| Đỗ Thị Như Quỳnh | Lý thuyết và demo mô hình Sorftmax Regression, đưa bài toán phân loại Softmax, SVM về bài toán hồi quy |
| Hồ Huyền Trang | Tiền xử lý dữ liệu + giảm chiều, lý thuyết và demo mô hình SVM |

1. **Cách tổ chức chương trình:**

- Sử dụng tập dữ liệu alzheimers\_disease\_data.csv đã up lên drive;

- Chương trình được chạy trên Google Colab;

- Sử dụng các mô hình thích hợp từ thư viện sklearn, trực quan hóa dùng matplot, seaborn, ngoài ra còn cần numpy, pandas,...

1. **Các kịch bản thực nghiệm:**

- Tiền xử lý dữ liệu: hiểu bộ dữ liệu, kiểm tra và xử lý các giá trị thiếu, lặp, không hợp lệ hoặc outliers;

- Trực quan hóa dữ liệu;

- Giảm chiều: các phương pháp thực nghiệm gồm PCA, LDA và UMAP;

- Chạy mô hình phân loại:

+ Các phương pháp thực nghiệm gồm KNN và Sortmax Regression, SVM;

+ Các kịch bản:

\* Chạy trên dữ liệu ban đầu, dữ liệu đã giảm chiều;

\* Thử nghiệm với các tỉ lệ: 8:2, 7:3, 6:4;

\* So sánh dữ liệu ban đầu với dữ liệu đã giảm chiều của cả hai mô hình;

\* So sánh kết quả thực nghiệm của hai mô hình KNN và Sortmax Regression.