**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO**

**Lab 01: Trực quan hóa dữ liệu với Python**

**Bộ môn: Trực quan hóa dữ liệu**

**NHÓM 22**

|  |  |
| --- | --- |
| **Nhóm trưởng** | 20120128 – Nguyễn Thị Cẩm Lai |
| **Thành viên**  **Thành Viên** | 20120037 – Trần Thị Minh Anh  20120125 – Bùi Anh Kiệt |
| **Thành viên** | 20120232 – Võ Duy Trường |
| **Thành viên** | 20120547 – Võ Thành Phong |

**Giáo viên hướng dẫn:** Nguyễn Thị Thu Hằng

Nguyễn Bảo Long

Lê Nhựt Nam

**Thành phố Hồ Chí Minh – 4/2023**

1. **Thông tin chung**
2. **Thông tin thành viên**

|  |  |  |
| --- | --- | --- |
| **Họ và tên** | **MSSV** | **Email** |
| Nguyễn Thị Cẩm Lai (nhóm trưởng) | 20120128 | 20120128@student.hcmus.edu.vn |
| Trần Thị Minh Anh | 20120037 | [20120073@student.hcmus.edu.vn](mailto:20120073@student.hcmus.edu.vn) |
| Bùi Anh Kiệt | 20120125 | [20120125@student.hcmus.edu.vn](mailto:20120125@student.hcmus.edu.vn) |
| Võ Duy Trường | 20120232 | [20120232@student.hcmus.edu.vn](mailto:20120232@student.hcmus.edu.vn) |
| Võ Thành Phong | 20120547 | 20120547@student.hcmus.edu.vn |

* **Github nhóm:** https://github.com/DataVisualizationGroup22/DataVisualization\_Lab01

1. **Phân công công việc**

|  |  |  |  |
| --- | --- | --- | --- |
| **Phần** | **Yêu cầu** | **Người thực hiện** | **Mức độ hoàn thành** |
| A.Thu thập dữ liệu | Lựa chọn dataset | Cả nhóm | 100% |
| Trình bày ngữ cảnh và động cơ lựa chọn | Cẩm Lai | 100% |
| B. Khám phá dữ liệu (thường đan xen với pha tiền xử lý dữ liệu) | [1. Đọc dữ liệu và tính số dòng và cột](https://github.com/ntclai/LTKHDL_ProjectFinal-TH/blob/647b8b81232fc6e7d2ba8cabe106a4420c8488c2/#2.2.1) | Duy Trường | 100% |
| [2. Mỗi dòng có ý nghĩa gì? Có vấn đề các dòng có ý nghĩa khác nhau không?](https://github.com/ntclai/LTKHDL_ProjectFinal-TH/blob/647b8b81232fc6e7d2ba8cabe106a4420c8488c2/#2.2.2) |
| [3. Dữ liệu có các dòng bị lặp không?](https://github.com/ntclai/LTKHDL_ProjectFinal-TH/blob/647b8b81232fc6e7d2ba8cabe106a4420c8488c2/#2.2.3) |
| [4. Tỷ lệ giá trị thiếu và thống kê mô tả của từng cột](https://github.com/ntclai/LTKHDL_ProjectFinal-TH/blob/647b8b81232fc6e7d2ba8cabe106a4420c8488c2/#2.2.4) |
| [5. Kiểu dữ liệu của mỗi cột](https://github.com/ntclai/LTKHDL_ProjectFinal-TH/blob/647b8b81232fc6e7d2ba8cabe106a4420c8488c2/#2.2.5) |
| [6. Xem xét tập giá trị của các thuộc tính phân loại](https://github.com/ntclai/LTKHDL_ProjectFinal-TH/blob/647b8b81232fc6e7d2ba8cabe106a4420c8488c2/#2.2.6) | Anh Kiệt | 100% |
| [7. Xem xét sự phân bố giá trị của các cột dữ liệu dạng số](https://github.com/ntclai/LTKHDL_ProjectFinal-TH/blob/647b8b81232fc6e7d2ba8cabe106a4420c8488c2/#2.2.7) |
| [8. Xem xét sự phân bố giá trị của các cột dữ liệu không phải dạng số](https://github.com/ntclai/LTKHDL_ProjectFinal-TH/blob/647b8b81232fc6e7d2ba8cabe106a4420c8488c2/#2.2.8) |
| C. Khám phá mối quan hệ trong dữ liệu | Khám phá mối quan hệ đơn biến (độc lập) | Yêu cầu tối thiểu:   * Minh Anh (3 khám phá) * Cẩm Lai (3 khám phá) * Anh Kiệt (2 khám phá) * Duy Trường (2 khám phá) * Thành Phong (1 khám phá) | 100% |
| Khám phá mối quan hệ đa biến/nhân quả (tích hợp) |
| D. Mô hình học máy | Sử dụng thuật toán học máy đơn giản để hiểu thêm về dữ liệu | Thành Phong | 100% |
| E. Viết báo cáo |  | Cẩm Lai, Thành Phong | 100% |

1. **Đánh giá mức độ hoàn thành**

|  |  |  |
| --- | --- | --- |
| **Phần** | **Yêu cầu** | **Mức độ hoàn thành** |
| A. Thu thập dữ liệu | Lựa chọn dataset | 100% |
| Trình bày ngữ cảnh và động cơ lựa chọn | 100% |
| B. Khám phá dữ liệu | Khám phá dữ liệu | 100% |
| Tiền xử lý dữ liệu | 100% |
| C. Khám phá mối quan hệ trong dữ liệu | Khám phá mối quan hệ đơn biến | 100% |
| Khám phá mối quan hệ đa biến/nhân quả | 100% |
| Trực quan hóa trên nhiều loại biểu đồ | 100% |
| D. Mô hình học máy | Chạy một số thuật toán học máy đơn giản để hiểu thêm về dữ liệu | 100% |
| Trình bày chi tiết thuật toán | 100% |

1. **Chi tiết thuật toán**
2. **Tài liệu tham khảo**