**Danh sách các công việc cần thực hiện.**

|  |  |  |
| --- | --- | --- |
| **Task** | **Mô tả** | **Ghi chú** |
| Viết báo cáo hàng tuần | - Tổng hợp các công việc hàng tuần để báo cáo tiến độ công việc. |  |
| Lên kế hoạch và tìm hiểu đề tài | - Liệt kê các công việc cụ thể theo tuần và theo ngày.  - Ước lượng thời gian thực hiện công việc: thời gian bắt đầu, kết thúc.  - Khảo sát về để tài và xu hướng của đề tài: lập bảng câu hỏi, khảo sát, thống kê bằng công cụ. |  |
| Tìm hiểu API của Facebook | - Tìm hiểu điều khoản sử dụng.  - Nghiên cứu các về các thông số của API.  - Nghiên cứu cách hoạt động | - Lựa chọn một trong hai API. |
| Tìm hiểu API của Twiter |
| Thực hiện chạy thử API Facebook và Twiter | - Đăng ký tài khoản.  - Chạy thử. |  |
| Lấy dữ liệu trên web về và lưu trữ | - Xây dựng chương trình.  - Chạy, kiểm thử và sửa lỗi.  - Tiến hành chạy lấy dữ liệu và lưu trữ. | - Việc chạy lấy dữ liệu có thể phải treo máy vài ngày nếu dữ liệu nhiều và yêu cầu môi trường đủ điều kiện. |
| Tìm hiểu các thư viện: Pandas, matploid, numpy... | - Đọc tài liệu xem cách hoạt động.  - Xem ví dụ minh hoạ.  - Chạy thử. |  |

|  |  |  |
| --- | --- | --- |
| Xử lý dữ liệu thô | - Phác hoạ cấu trúc dữ liệu ban đầu.  - Liệt kê những kiểu dữ liệu cho trường cụ thể.  - Tìm các lỗi gây mất dữ liệu, sai kiểu dữ liệu,… - Tìm giải pháp.  - Xây dựng chương trình xử lý dữ liệu thô.  - Chạy, kiểm thử và sửa lỗi.  - Tiến hành xử lý. | - Quá trình xử lý có thể bỏ qua một số lỗi về định dạng. |
| Nghiên cứu thuật toán phân cụm | - Đọc tài liệu về thuật toán.  - Xem các ví dụ minh hoạ.  - Mô tả một số bài toán cần áp dụng. |  |
| Tìm hiểu thuật toán phân tích nội dung |
| Xử lý nhiễu dữ liệu | - Phác thảo cấu trúc dữ liệu theo bài toán.  - Liệt kê các trường hợp có thể bị mất hoặc không có dữ liệu đưa ra cách giải quyết.  - Xây dựng chương trình.  - Chạy, kiểm thử và sửa lỗi.  - Tiến hành xử lý. | - Chia ra các phiên bản để dễ quản lý và không làm mất mát dữ liệu. |
| Tìm hiểu Markov Chain | - Đọc tài liệu về thuật toán.  - Xem các ví dụ minh hoạ.  - Mô tả một số bài toán cần áp dụng. |  |
| Phân tích dữ liệu | - Xác định bài toán theo dữ lệu.  - Vẽ biểu đồ: histogram, plot,..  - Đưa ra kết quả, kiểm chứng. |  |
| Vẽ biểu đồ phân cụm | - Xây dựng chương trình.  - Chạy và lưu kết quả. |  |
| Chạy thử, kiểm tra và sửa lỗi |  |  |
| Lấy kết quả và thông số phân tích |  |  |
| Viết báo cáo khóa luận |  |  |
| Làm PowerPoint và video demo |  |  |
| Tối ưu giải pháp | - Tối ưu thời gian  - Tối ưu tốc độ.  - Tối ưu độ chính xác. | - Kế hoạch phát triển thêm nếu còn thời gian thực hiện |
| Tìm hiểu giải pháp khác | - Tìm kiếm giải pháp khác cho phân tích dữ liệu  - Nghiên cứu giải pháp có phù hợp để áp dụng |
| Tích hợp, phát triển các giải pháp khác(nếu có) |  |
| Lấy dữ liệu mới nhất và phân tích | - Chạy chương trình với thời gian gần nhất. |