# Các câu hỏi nên đưa ra với Nhà Khoa Học Dữ Liệu (Chiến lược: Top down + Sắp xếp câu hỏi: Pyramid + Chủ yếu câu hỏi đóng)

| **Yêu cầu** | **Câu hỏi** |
| --- | --- |
| Mục tiêu phân tích | 1. Về mức độ chi tiết của phân tích, bạn muốn dự đoán món ăn tiếp theo nên theo cấp độ món ăn cụ thể hay theo danh mục món ăn? 2. Về độ chính xác, bạn có yêu cầu về mức độ chính xác cụ thể hay tối thiểu của mô hình không? 3. Bạn muốn dự đoán món ăn tiếp theo được tính toán và phản hồi ngay lập tức theo thời gian thực hay có thể cập nhật qua từng phiên truy cập? 4. Bạn có muốn cá nhân hóa dự đoán món ăn cho từng người dùng riêng lẻ hay áp dụng theo nhóm người dùng có hành vi tương tự? 5. Bạn có kịch bản sử dụng cụ thể nào không, ví dụ như hiển thị món ăn gợi ý trên trang chủ hay đề xuất trong giỏ hàng? 6. Có muốn tìm hiểu nguyên nhân khiến người dùng từ bỏ giỏ hàng không? |
| Về dữ liệu cần thiết | 1. Khoảng thời gian dữ liệu lịch sử bạn muốn sử dụng là bao lâu? 2. Bạn cần dữ liệu tương tác người dùng được ghi nhận với mức độ chi tiết nào, ví dụ lượt click hay cần theo dõi cả thời gian trên từng trang? 3. Lịch sử đặt món cần chi tiết đến mức nào? Có cần thông tin về số lượng , thời gian đặt, hoặc địa điểm giao hàng không ? 4. Bạn có muốn phản hồi của người dùng được phân loại theo mức độ tích cực và tiêu cực không? 5. Về dữ liệu thời gian thực, bạn có muốn xử lý ngay lập tức hay bạn chấp nhận gom lại và xử lý theo batch? 6. Bạn có muốn sử dụng dữ liệu bên ngoài như thời tiết hoặc sự kiện đặc biệt để cải thiện mô hình không? |
| Yêu cầu hệ thống | 1. Bạn muốn lưu trữ dữ liệu trong khoảng thời gian bao lâu? 2. Bạn cần hệ thống của mình hỗ trợ bao nhiêu người dùng đồng thời, vì điều này sẽ ảnh hưởng đến hiệu năng xử lý dữ liệu? 3. Bạn muốn cập nhật dữ liệu lịch sử và phản hồi người dùng với tần suất nào: hàng ngày, hàng giờ, một khoảng thời gian hay thời điểm cụ thể? 4. Ban có cần đảm bảo độ tin cậy (reliability) cao không? Nếu một phần dữ liệu bị mất hoặc cập nhật chậm thì có ảnh hưởng nghiêm trọng đến phân tích không? |

# Chuyển các yêu cầu của Nhà Khoa Học Dữ Liệu sang Yêu Cầu Hệ Thống

| Yêu cầu | Chi tiết yêu cầu | | |
| --- | --- | --- | --- |
| Chức năng | Thu thập và lưu trữ dữ liệu | * Lưu trữ lịch sử đặt món của người dùng theo thời gian. * Ghi nhận dữ liệu tương tác của người dùng với món ăn (click, thời gian xem, thêm vào giỏ hàng nhưng không mua). * Lưu phản hồi của người dùng về nhà hàng, món ăn, và giao hàng. * Cập nhật dữ liệu thời gian thực từ ứng dụng (click, session, giỏ hàng). | |
| Xử lý và làm sạch dữ liệu | * Chuẩn hóa dữ liệu món ăn theo danh mục, giá cả, và thời gian giao hàng. * Loại bỏ dữ liệu không hợp lệ hoặc nhiễu trong phản hồi người dùng. | |
| Hỗ trợ mô hình dự đoán | * Cung cấp dữ liệu lịch sử cho mô hình machine learning. * Xây dựng API để truy vấn món ăn gợi ý theo thời gian thực. * Lưu trữ và phục vụ kết quả dự đoán theo yêu cầu của hệ thống đề xuất. | |
| Phân tích lý do bỏ giỏ hàng | * Ghi nhận dữ liệu khi người dùng thêm món vào giỏ hàng nhưng không thanh toán. * Phân tích các yếu tố ảnh hưởng như giá, thời gian giao hàng, phí ship, các review tiêu cực. | |
| Phi chức năng | Hiệu năng (Performance) | Tốc độ (Speed) | * Hệ thống phải có khả năng xử lý dữ liệu thời gian thực với độ trễ thấp (<500ms cho truy vấn gợi ý). * API dự đoán món ăn phải có thời gian phản hồi nhanh (<1s). |
| Tính mở rộng (Scalability) | * Hỗ trợ tăng số lượng người dùng mà không ảnh hưởng đến hiệu năng. * Dữ liệu phải được lưu trữ theo cách có thể mở rộng (ví dụ: sử dụng kiến trúc microservices và lưu trữ trên cloud). |
| Tính tin cậy (Reliability) | * Đảm bảo dữ liệu đầy đủ và không bị mất trong quá trình xử lý. * Cung cấp dữ liệu sạch và cập nhật để hỗ trợ mô hình dự đoán. |
| Bảo mật và quyền riêng tư | * Bảo vệ dữ liệu người dùng theo các tiêu chuẩn bảo mật. * Ẩn danh dữ liệu nếu cần để tuân thủ các quy định về quyền riêng tư. | |