### **Guideline Gán Nhãn Đánh Giá Sản Phẩm - Độ Tin Cậy**

#### **1. Mục đích của dự án**

Mục đích của dự án này là xây dựng mô hình học máy để phân loại đánh giá sản phẩm trên các sàn thương mại điện tử thành hai nhóm: **"Đáng Tin Cậy"** và **"Không Đáng Tin Cậy"**. Quá trình gán nhãn chính xác sẽ giúp huấn luyện mô hình để dự đoán độ tin cậy của các đánh giá mới trong tương lai.

#### **2. Các Tiêu Chí Phân Loại**

**A. Đánh Giá Đáng Tin Cậy** Đánh giá sẽ được coi là "Đáng Tin Cậy" khi đáp ứng các yếu tố sau:

1. **Số sao hợp lý**:  
   * Đánh giá có số sao phù hợp với nội dung và cảm xúc. Ví dụ: Một đánh giá tích cực về sản phẩm sẽ có từ 4 đến 5 sao.
2. **Nội dung đánh giá rõ ràng và chi tiết**:  
   * Đánh giá cung cấp thông tin chi tiết về sản phẩm, như chất liệu, kích thước, hiệu quả sử dụng, hoặc trải nghiệm cá nhân của người dùng.
3. **Cảm xúc và ngữ cảnh nhất quán**:  
   * Cảm xúc trong đánh giá (tích cực hoặc tiêu cực) phải phù hợp với số sao. Ví dụ, đánh giá tích cực sẽ có cảm xúc vui vẻ, còn đánh giá tiêu cực sẽ có cảm xúc không hài lòng.
4. **Không có dấu hiệu spam hoặc thao túng**:  
   * Đánh giá không chứa các từ ngữ lặp lại, quá ngắn gọn, hoặc không có sự hỗ trợ nào từ văn bản. Nội dung đánh giá phải có vẻ tự nhiên và hợp lý.

**B. Đánh Giá Không Đáng Tin Cậy** Đánh giá sẽ bị gán nhãn là "Không Đáng Tin Cậy" khi có các dấu hiệu sau:

1. **Số sao không hợp lý**:  
   * Đánh giá có số sao không hợp lý với nội dung. Ví dụ: Đánh giá tiêu cực nhưng lại cho 5 sao hoặc ngược lại.
2. **Nội dung mơ hồ hoặc không chi tiết**:  
   * Đánh giá quá ngắn, không cung cấp đủ thông tin về sản phẩm hoặc không rõ ràng. Ví dụ: "Sản phẩm tốt" mà không có mô tả thêm.
3. **Dấu hiệu spam hoặc thao túng đánh giá**:  
   * Đánh giá có vẻ như được sao chép, sử dụng từ ngữ quảng cáo, hoặc không liên quan đến sản phẩm. Các đánh giá có nhiều lỗi chính tả, lặp lại từ ngữ, hoặc chỉ là các câu như "Tuyệt vời" mà không có nội dung chi tiết.
4. **Cảm xúc và ngữ cảnh không nhất quán**:  
   * Đánh giá có cảm xúc trái ngược hoặc không phù hợp với số sao (ví dụ: viết tiêu cực nhưng lại cho 5 sao).

#### **3. Quy Trình Gán Nhãn**

1. **Đọc kỹ đánh giá**:  
   * Đảm bảo rằng bạn đọc toàn bộ nội dung đánh giá và xem xét các yếu tố: số sao, nội dung văn bản và hình ảnh (nếu có).
2. **Đánh giá độ tin cậy của văn bản**:  
   * Kiểm tra xem đánh giá có đủ chi tiết, có cảm xúc phù hợp với số sao không, và có dấu hiệu của spam hoặc thao túng không.
3. **Kiểm tra hình ảnh (nếu có)**:  
   * **Chất lượng hình ảnh**: Hình ảnh phải rõ ràng, không mờ, và có thể nhận diện được sản phẩm.
   * **Liên quan đến sản phẩm**: Hình ảnh phải thể hiện đúng sản phẩm và có liên quan trực tiếp đến nội dung đánh giá.
   * **Không có dấu hiệu chỉnh sửa**: Hình ảnh không có dấu hiệu bị chỉnh sửa hoặc lấy từ các nguồn không liên quan.
4. **Gán nhãn**:  
   * Sau khi đánh giá kỹ các yếu tố trên, hãy gán nhãn cho đánh giá:  
     + **Đánh Giá Đáng Tin Cậy**: Nếu nội dung, số sao, và hình ảnh (nếu có) phù hợp và không có dấu hiệu bất thường.
     + **Đánh Giá Không Đáng Tin Cậy**: Nếu có dấu hiệu của nội dung mơ hồ, số sao không hợp lý, hình ảnh giả mạo, hoặc các dấu hiệu thao túng.

#### **4. Các Câu Hỏi Hướng Dẫn**

Khi gặp các tình huống không rõ ràng, hãy tự hỏi các câu sau:

1. **Đánh giá này có cung cấp đủ thông tin chi tiết không?** Nếu không, đánh giá có thể không đáng tin cậy.
2. **Số sao có phù hợp với nội dung và cảm xúc không?** Nếu không, đánh giá có thể không đáng tin cậy.
3. **Hình ảnh có rõ ràng, có liên quan và không bị chỉnh sửa không?** Nếu không, đánh giá có thể không đáng tin cậy.
4. **Có dấu hiệu của spam hoặc thao túng không?** Nếu có, đánh giá không đáng tin cậy.

#### **5. Ví Dụ Minh Họa**

**Ví dụ 1**:

* **Đánh giá**: "Sản phẩm này rất tuyệt vời! Chất liệu tốt, màu sắc đẹp, và sử dụng rất dễ dàng. Tôi đã sử dụng hàng tuần và không gặp vấn đề gì."
* **Số sao**: 5 sao
* **Hình ảnh**: Hình ảnh rõ nét, hiển thị chi tiết sản phẩm.
* **Nhận xét**: Đây là một đánh giá **đáng tin cậy** vì cả nội dung và hình ảnh đều hợp lý và chất lượng.

**Ví dụ 2**:

* **Đánh giá**: "Tôi không hài lòng với sản phẩm này. Nó bị hỏng sau 1 tuần sử dụng."
* **Số sao**: 1 sao
* **Hình ảnh**: Hình ảnh mờ, không thể nhận diện rõ sản phẩm.
* **Nhận xét**: Đây là một đánh giá **đáng tin cậy** mặc dù hình ảnh không rõ, nhưng nội dung đánh giá phù hợp với số sao.

**Ví dụ 3**:

* **Đánh giá**: "Sản phẩm tốt, nhưng tôi nhận được một sản phẩm khác."
* **Số sao**: 2 sao
* **Hình ảnh**: Hình ảnh không phải của sản phẩm đã được đánh giá mà là hình ảnh mẫu từ trang web.
* **Nhận xét**: Đây là đánh giá **không đáng tin cậy** vì hình ảnh không phải của sản phẩm thực tế.

**Ví dụ 4**:

* **Đánh giá**: "Ok"
* **Số sao**: 1 sao
* **Hình ảnh**: Không có hình ảnh.
* **Nhận xét**: Đây là đánh giá **không đáng tin cậy** vì nội dung quá mơ hồ và không đủ thông tin.

#### **6. Lưu Ý Quan Trọng**

1. **Xử lý hình ảnh**: Hình ảnh là một phần quan trọng của đánh giá, vì vậy hãy đảm bảo rằng hình ảnh có chất lượng và liên quan đến sản phẩm.
2. **Đảm bảo tính nhất quán**: Các annotators cần phải tuân thủ các tiêu chí phân loại để đảm bảo tính nhất quán trong việc gán nhãn.
3. **Kiểm tra nhãn lại**: Sau khi gán nhãn, nếu có thể, hãy kiểm tra lại nhãn của các đánh giá để đảm bảo độ chính xác.