**I. Quy Trình Thực Hiện**

Quy trình này gồm **5 bước chính**, từ thu thập dữ liệu đến triển khai ứng dụng.

**1. Thu thập dữ liệu**

📌 **Nguồn dữ liệu**

* **Nền tảng thương mại điện tử**: Shopee, Tiki, Lazada
* **Website đánh giá dịch vụ**: Google Reviews

📌 **Cách thu thập dữ liệu**

* **Web scraping**: Dùng **BeautifulSoup, Scrapy, Selenium** để lấy bình luận
* **Thu thập tay**: với trường hợp web không cho sử dụng tool cào dữ liệu

📌 **Lưu ý**  
✔ Loại bỏ dữ liệu rác (spam, quảng cáo)  
✔ Cân bằng số lượng bình luận tích cực, tiêu cực, trung tính

**2. Tiền xử lý dữ liệu**

📌 **Bước 1: Làm sạch dữ liệu**  
✔ Loại bỏ stop words, dấu câu, số, URL  
✔ Chuyển đổi chữ viết hoa thành chữ thường  
✔ Xóa emoji hoặc chuyển emoji thành text

📌 **Bước 2: Biến đổi dữ liệu**  
✔ **Tokenization**: Chia câu thành các từ nhỏ  
✔ **Padding & Truncation**: Đảm bảo độ dài câu cố định  
✔ **Word Embedding**: Dùng BERT tokenizer để mã hóa văn bản thành vector số

📌 **Bước 3: Gán nhãn dữ liệu**  
✔ **Dữ liệu có nhãn**: Bình luận có sẵn nhãn cảm xúc (positive, neutral, negative)

**3. Huấn luyện mô hình AI**

📌 **Mô hình sử dụng**  
**- BERT-base**

**- BERT-LSTM**

**- BERT-RCNN**

📌 **So sánh mô hình**

* Đánh giá qua **Accuracy, Precision, Recall, F1-score**
* Chọn mô hình có hiệu suất tốt nhất

**4. Đánh giá & Điều chỉnh mô hình**

📌 **Các bước đánh giá**  
✔ Dùng tập test để kiểm tra độ chính xác của mô hình  
✔ Tinh chỉnh tham số **learning rate, batch size** để tối ưu kết quả

*Kết quả mô hình Vreview dataset (AIVIVN) (DATA)*

| **Model** | **Precision(%)** | **Recall(%)** | **F1(%)** |
| --- | --- | --- | --- |
| BERT-base | 86.08 | 88.44 | 87.2 |
| BERT-LSTM | 85.25 | 89.9 | 87.5 |
| BERT-RCNN | 87.08 | 89.38 | **88.22** |

*Kết quả mô hình* *NTC-SV Dataset (FOODY)*

| **Model** | **Precision(%)** | **Recall(%)** | **F1(%)** |
| --- | --- | --- | --- |
| BERT-base | 88.13 | **94.02** | 90.9 |
| BERT-LSTM | **89.78** | 92.08 | 90.91 |
| BERT-RCNN | 88.76 | 93.68 | **91.15** |