1. Đào dữ liệu:  
   + đào dữ liệu từ 1 số trang web rồi cho chọn 2-3 trang  
   + Đào dữ liệu sản phẩm, đánh giá, dùng cái đánh giá để train rồi đưa ra sản phẩm
2. Thu thập dữ liệu  
   + Dịch ngược

+ Thay thế từ đồng nghĩa

+ Thay thế thực thể

+ Gây nhiễu dữ liệu

+ Đổi chỗ ngẫu nhiên một số từ

+THêm ngẫu nhiên một số từ

+ Xóa ngẫu nhiên một số từ

+ dịch ngược

1. Làm sạch dữ liệu

+ Xóa dấu chấm câu

+ Xóa số

+ Xóa khoảng trắng

+ Xóa các biểu tượng

+ Sentence Tokenization

+ Word Tokenization

+ Xóa stop words

+ viết thường các ký tự

+ Sửa lại các từ viết tắt

+ Sửa lại các từ viết sai

+ POS tagging

+ Stemming

+ Lemmatization

+ Parsing

+ NER

1. Vecto hóa dữ liệu:

+ One – hot encoding

+ Bag of words

+ Bag of N-Grams

+ TF – IDF

+ Word Embedding

+ Glove

+ Fast test

+ ELMo

+ GPT

+BERT

1. Classification

+SVM

1. Recommendation system
2. Chatbot

Chức năng:

* Chọn cột từ dữ liệu nhập vào
* Tô đậm các câu đã chỉnh sửa

APP – PIPELINE

* Thanh 1: Chọn dữ liệu vào

+ khung chọn dữ liệu từ các dữ liệu có sẵn – có phần để chọn cột cmt và ratings (label)

+ Khung chọn data từ một số trang web có sẵn – amazon, Lazada, shoppe

+ khung nhập them dữ liệu đầu vào

* Thanh 2: tiền xử lý dữ liệu

+ Thanh chọn các biện pháp so sánh dữ liệu, chọn số lượng câu làm sạch, chọn stt câu để hiện thị sau đó có 2 khung 1 khung câu góc, 1 khung câu đã làm sạch, them phần hiển thị màu sắc cho các điểm khác biệt

* Thanh 3: tang cường dữ liệu  
  + thanh chọn các biện pháp tang cường dữ liệu, chọn số lượng câu tăng cường, chọn stt câu để hiển thị sau đó có 2 khung 1 khung câu góc, 1 khung câu đã làm sạch, them phần hiển thị màu sắc cho các điểm khác biệt
* Thanh chọn model training đánh giá

+ chọn model muốn train in ra confuse matrix,…

* Cho người dùng nhập vào 3-4 câu cmt và đánh giá sau đó đưa ra đề xuất cho người dùng
* Chatbot có prompt để gợi ý cho người dùng các sản phẩm từ các đánh giá, hỏi về giới thiệu sản phẩm