510 Group Project Proposal

Track of the project: Development track

Group member name & email:

1. Jiayi Gu: jiayigu4@illinois.edu
2. Xiaoqing Yao: xyao14@illinois.edu
3. Kanghong Zhao: kz8@illinois.edu
4. Wentao Zhou(coordinator): wentao6@illinois.edu

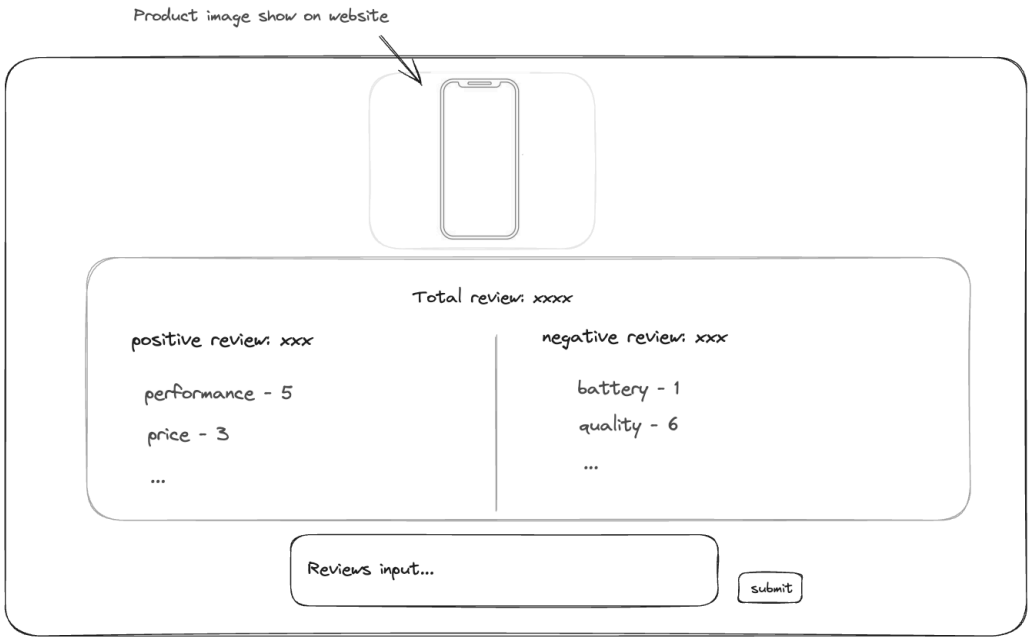
Functions and Users

We are developing an extension for existing web applications. This extension will integrate with popular online shopping platforms, like Amazon, to provide users with summarized insights from product reviews.

Major Functions:

1. Review Aggregation: Collect and aggregate reviews for a given product from the web page it is associated with.
2. Natural Language Processing (NLP): Analyze the text of the reviews using NLP techniques to understand sentiment, identify key themes, and extract useful insights.
3. User Interface (UI) Integration: Display a list of those keywords in positive and negative sections in a user-friendly manner on the top of the review section, making it easily accessible to users as they shop.

The primary users of this tool are Individuals browsing online shopping websites like Amazon who seek a quick and easy way to understand product reviews without reading through every single one. This tool helps them make informed purchasing decisions more efficiently.



Significance

Our proposed tool is vital because it streamlines the process of analyzing online reviews, which is currently a cumbersome task. Although Amazon has a similar feature as our project, their keywords are preset instead of dynamically generated. Also, most other shopping websites only provide rates and individual reviews. Addressing this pain point makes navigating consumer feedback more manageable. Moreover, in today's digital marketplace, where online reviews heavily influence purchase decisions, our tool serves a societal need for transparent and informed consumer choices. Its implementation fosters trust between consumers and sellers, contributing to a more efficient e-commerce ecosystem.

Approach

The main idea of our project is to extract sentiment and keywords from the reviews. We are planning to use two existing Machine Learning models to achieve this. One is to analyze sentiment, and the other one is to extract keywords. Since the core part is the models, we will build a simple webpage that simulates the review sections for demo purposes. In this case, we will use simple HTML for the frontend, Python for the backend server and models, and MongoDB to store the model output. One potential barrier we can anticipate is the performance of the existing models. If the models are not performing well in the evaluation stage, we will do some fine-tuning to achieve optimal performance.

Evaluation

We will crawl open sourced labeled data and also manually gather and label data. We will then validate the correctness and usefulness of the models by employing F1, precision and recall metrics.

Timeline

– First week: Research models, design web application and collect relevant data
– Second and third week: Implement models and web application and prepare evaluation
– Fourth week: Finish integrating the web application, gather feedback for improvement, and evaluate the models
– Fifth week: Prepare for final report and presentation

Task division

Jiayi Gu – Implement frontend of the web application
Xiaoqing Yao – work on backend server and database implementation
Kanghong Zhao – collect and preprocess data and implement model evaluation
Wentao Zhou – research, design and implement models