**CSCE 5290: Natural Language Processing**

**Project Proposal**

**Title: [Customer Review Summarization for Products]**

**Name: [Nick Wylie]**

**1. Motivation**

Customer Reviews play a major role in whether a product/service is purchased. Reviews can often be lengthy paragraphs filled with detail that is not often useful. This can easily turn a customer off from making a purchase. The goal of this application is to summarize customer reviews giving the more important detail and leaving out unneeded text. With a summarization tool in play, people will read triple the amount of review summaries and comprehend more of peoples’ experience using the product/service. In my experience looking through online reviews, I had trouble before looking at the descriptions of people’s experience, and making a judgement call if the product is worth the value of what others have praised or complained. This is why my motivation in this project is to change the way that reviews are displayed for an easier judgment on the customers perspective.

**2. Significance**

Not only for consumers, but company owners will benefit from this review-condensing application. Whenever a product being sold is a great high-quality item, there will be honest positive reviews. The amount of positivity the customer will be able to skim through increases, furthering their chances of purchasing the product/service. How this application can benefit the customers is to make the reviews easier to tailor to their specific needs. For example, let’s say someone is looking for a new mini fridge. Their old mini fridge had a problem where water was leaking out of the back and making a mess. When this customer is looking at a new mini fridge online and sees condensed reviews that mention a leaking, they have a better chance of avoiding this and can get a better mini fridge that does not have the same issues. The significance of this application will give company owners a bigger clientele, and help customers find a product/service that tailors to their needs.

**3. Objectives**

For this program to be successful it needs to have a quick run time. The theme of this program is quick condensed summaries of reviews, so a slow run time will defeat the purpose. This program will also need to be scalable. Once data shows that summarized reviews increase sales in product, this application should be able to run on big platforms like Yelp, Google Reviews, Facebook Reviews, etc.. These review platforms have a large dataset of reviews in numerous entities, so for this program to be scalable it should have the ability to run on large sets of data. Lastly, I aim to make the code clear and maintainable. This will encourage new features and improvements for the future.

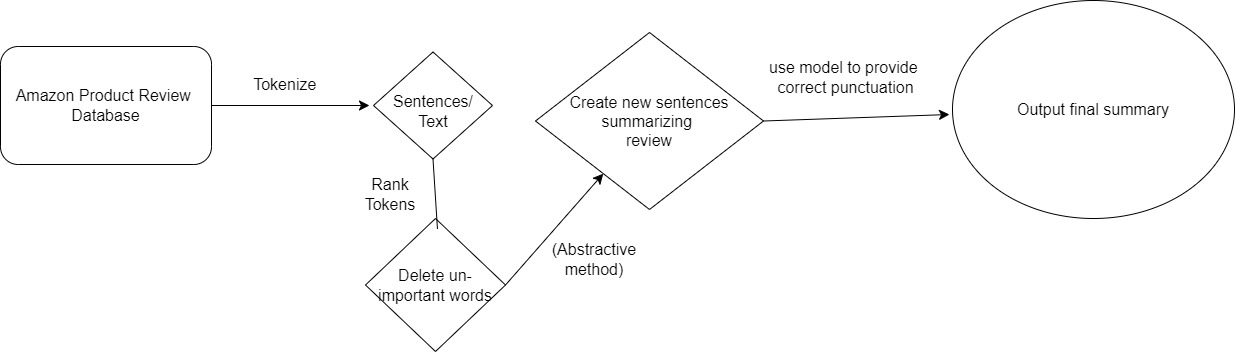
**4. Features**

The features that will be implemented in this project will have the user able to read condensed reviews over numerous products. They will be able to navigate through different products and skim through different opinions via review. If the user finds that they are interested in the review and want to review they will have an option to read the full review and that exact review will be expanded for them. This is a distinctive element in this project because there it can help business learn about the product and see reoccurring reviews and their similarities. This gives quick and effective feedback for everyone.

**5. Dataset.**

The dataset given to me is Amazon’s list of product reviews. This is a range of long/short paragraphs of reviews. This dataset contains 500,000 reviews ranging back from October 2012. This data will be processed for analyzation and condensing text. The data set also contains additional details such as user information, product information, ratings, and reviews from multiple amazon categories. This dataset is given in a .csv file.

**6. Visualization**

****

This diagram above shows the process behind what will happen in this application. The first step is to extract the long text reviews out of the database for analysis. The next step would be to tokenize the words or sentences in this text for organization. Next, I will use a trained language model to delete unimportant text out of these summaries. This will ensure that the summaries of these reviews will only provide important detail. After this, my application will use the abstractive method of generating new sentences into the summary that will add good context of the review. Lastly, punctuation will be corrected to ensure the finally output is readable.

**GitHub link:**

**https://github.com/NickHubGitHub/ProductReviews**