Project proposal

Team members: Jiaqi Cao (jiaqic7), Naifu Zheng (naifuz2), Yige Feng (yigef2), Yuxin Wang (yuxinw5)

Captain: Jiaqi Cao (jiaqic7)

Topic chosen: Intelligence Browsing - Chrome extension to summarize Amazon product review using topic analysis and sentiment analysis.

Why is it a problem?

There are summaries about the product reviews on Amazon, but only for the products with more than hundreds of reviews. However, when browsing the product review summaries, there are some inaccuracies in the summaries. For example, there was a summary with the tag "even though", and another two redundant summaries of "easy to set" and "easy to set up", which can be simplified and improved.

How does it relate to the theme and to the class?

The summary tags are going to be simplified by using the text retrieval techniques that we learned in this class such as stemming.

Dataset:

[Amazon Customer Reviews Dataset](#)

Algorithms:

Tokenization, stemming, POS tagging

Feasibility Description: Customers of Amazon provide comments, either positive or negative, on the products. Our extension performs topic analysis on the comments and captures keywords from them. For each key aspect, the extension performs sentiment analysis on the keywords and identifies whether customers like or dislike the product. Also, there might be a minority of

people who make highly biased comments on the products with high subjectivity, and it is part of our goal to differentiate between objective and subjective comments.

Programming languages: Javascript, Python

Tasks:

Module	Estimated time
Scrape and crawl product reviews from a target url	~10hrs
Data cleaning	~5hrs
Tokenization and stemming	~5hrs
Topic analysis and sentiment analysis	~10hrs
Make the module as an extension	~15hrs
Frontend & user interface	~20hrs
Integration of each parts	~15hrs