1) Which tasks have been completed?

I altered my topic to now be: Analyze comments from recipe blog posts, focusing on classifying reviews as positive or negative (or more sentiments, depending on the datasets I can find).

I have identified some datasets, but have not entirely solidified which one to use. The topmost choices are: https://www.kaggle.com/code/gemmin/sentiment-analysis, https://www.kaggle.com/datasets/farukalam/yelp-restaurant-reviews/data, and https://www.kaggle.com/datasets/joebeachcapital/restaurant-reviews/code . These are not datasets that are directly related to comments on recipe blogs, but they are about reviews of recipes and reviews of restaurants.

I have reached out to authors of a paper "Sentiment Analysis of Food Recipe Comments", since they used a dataset that is more relevant to the topic I am working on. I hope they can send me their dataset, in which case I will only use that one. If not, I will be utilizing the above datasets- specifically, the review text column and the rating. All the ratings are 1-5, so I will be classifying 1 and 2 and negative, 3 and neutral, and 4 and 5 as positive reviews.

I have written some code using pandas to clean and partition whichever dataset I will be using, into training and testing data.

I have written a Naive Bayes algorithm draft using Python.

I have written a draft of a Chrome extension using JavaScript, based on a tutorial from the official Chrome Developer website, which sets up the extension and selects the correct HTML elements to scrape comments from the blog (in this case, King Arthur Baking).

2) Which tasks are pending?

The task of finalizing a dataset and cleaning it remains to be done.

The task of completing and evaluating/fine-tuning the Naive Bayes algorithm remains to be done.

The Chrome extension must be completed still- it cannot be turned on currently. It also needs to have a way to scroll to collect the second/third/etc page of comments from the blog post. This may not be fixable given the time frame, but I will attempt to get it done. The extension should also be passing the comments to the model.

3) Are you facing any challenges?

Selecting the datasets has been challenging, since I cannot find specific comment analysis datasets.

The Chrome extension has not been easy to write so far, but I am looking into debugging practices here.