**Methodology**

In summary, our deep learning pipeline, akin to transfer learning, employs two distinct machine learning models for natural language processing and prediction. Following dataset cleaning and balancing, we enhance reviews with pseudo-labels using the 4-billion-parameter BART model as a zero-shot classifier, assuming accurate ratings in the source dataset. Subsequently, we calculate likelihoods for each review's association with pseudo-label classes, defined arbitrarily based on product sentiments and represented as numbers from 0 to 1. Finally, our neural network, utilizing both pseudo-labels and target ratings, is trained to provide robust predictions for a user's rating based solely on the detailed content.

When profiling the dataset in the data cleaning and normalization step, we noticed that there are more than 100,000 ratings with 5 stars, while others are below 50,000 or even 20,000, indicating that the balance of the rating distribution is problematic. Additionally, the distribution of comment length is also imbalanced: That is, while the length is around 500 characters with a standard deviation of 700 characters, which translates to around 50 to 100 words in total length, there are some that are more 30,000 characters long or, on the other end of the spectrum, have zero words.