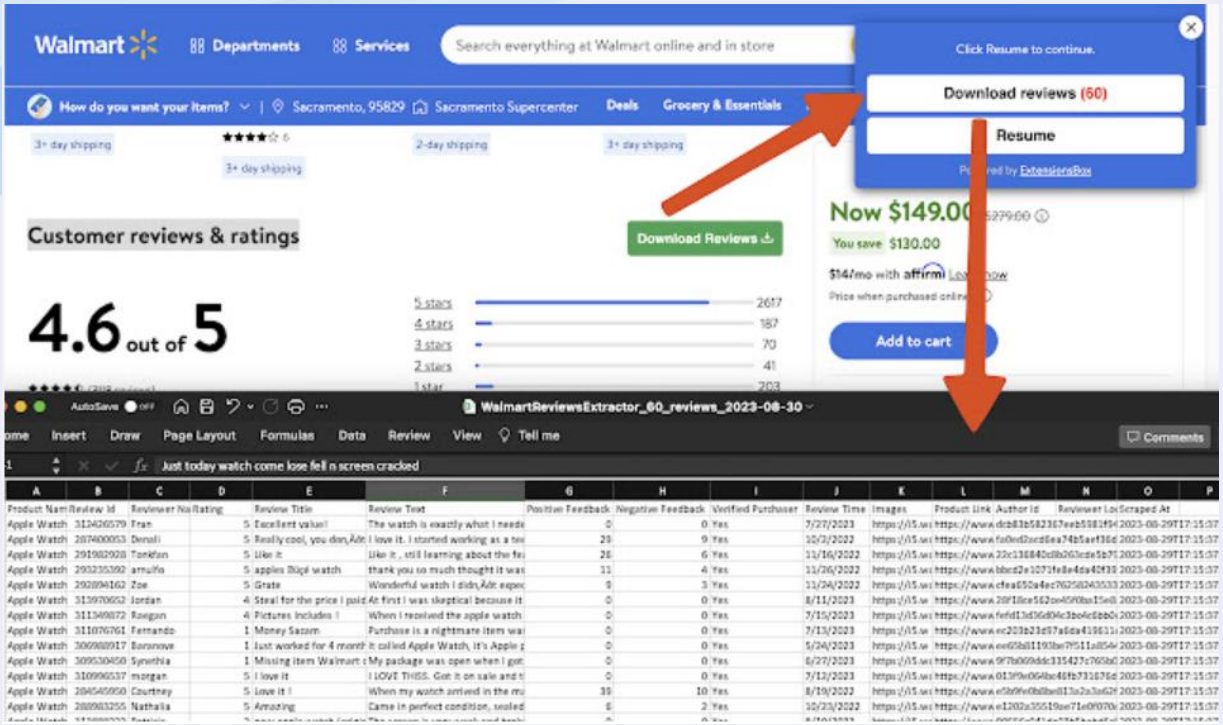
Natural language sentiment analysis and modeling on Walmart customer reviews

Wenhui Wang
Springboard
02/22/2024

The problem



- The Walmart Customer Reviews Dataset offers a wealth of insights into consumer sentiment and product feedback related to one of the world's largest retail giants. This dataset contains a vast collection of customer reviews, star ratings, and other relevant information that has been gathered through web scraping and data compilation.

Targets of the project

- Running sentiment analysis to find the key features that will promote or impair customers' experience.

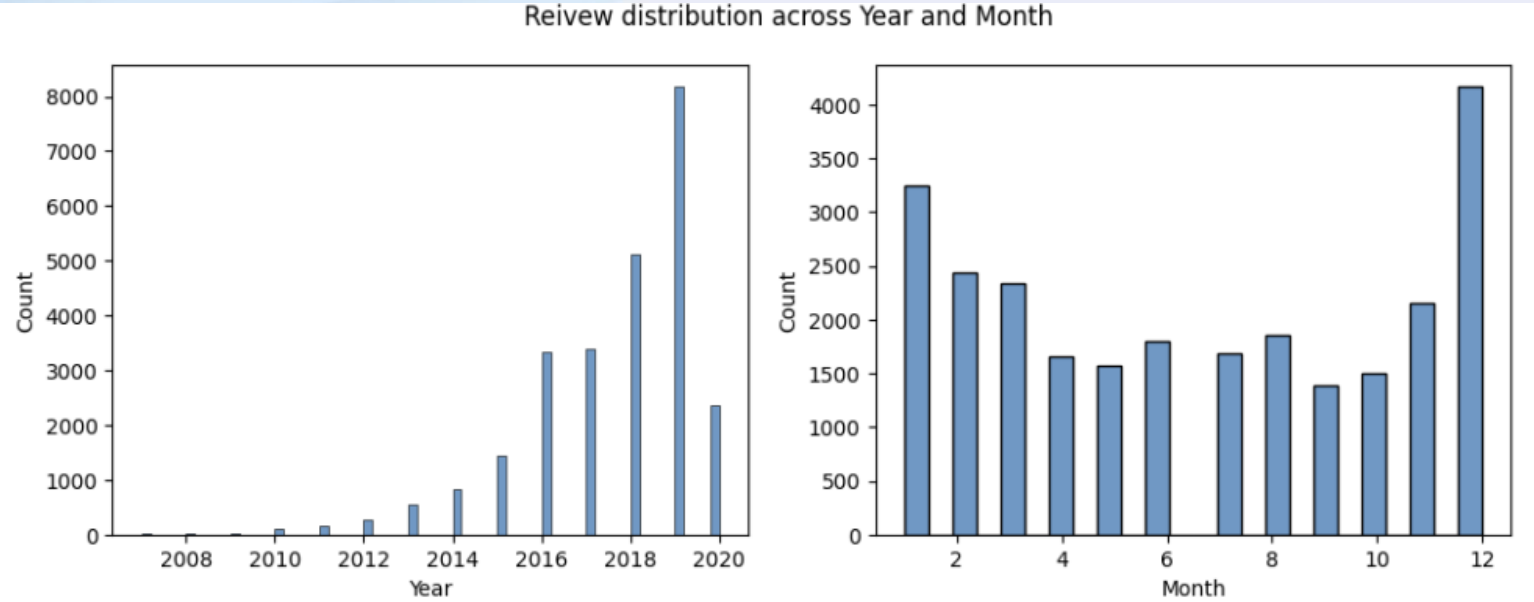
- Building model to predict customer rating based on review language by transfer learning and fine tuning on of pre-trained RoBERTa model.

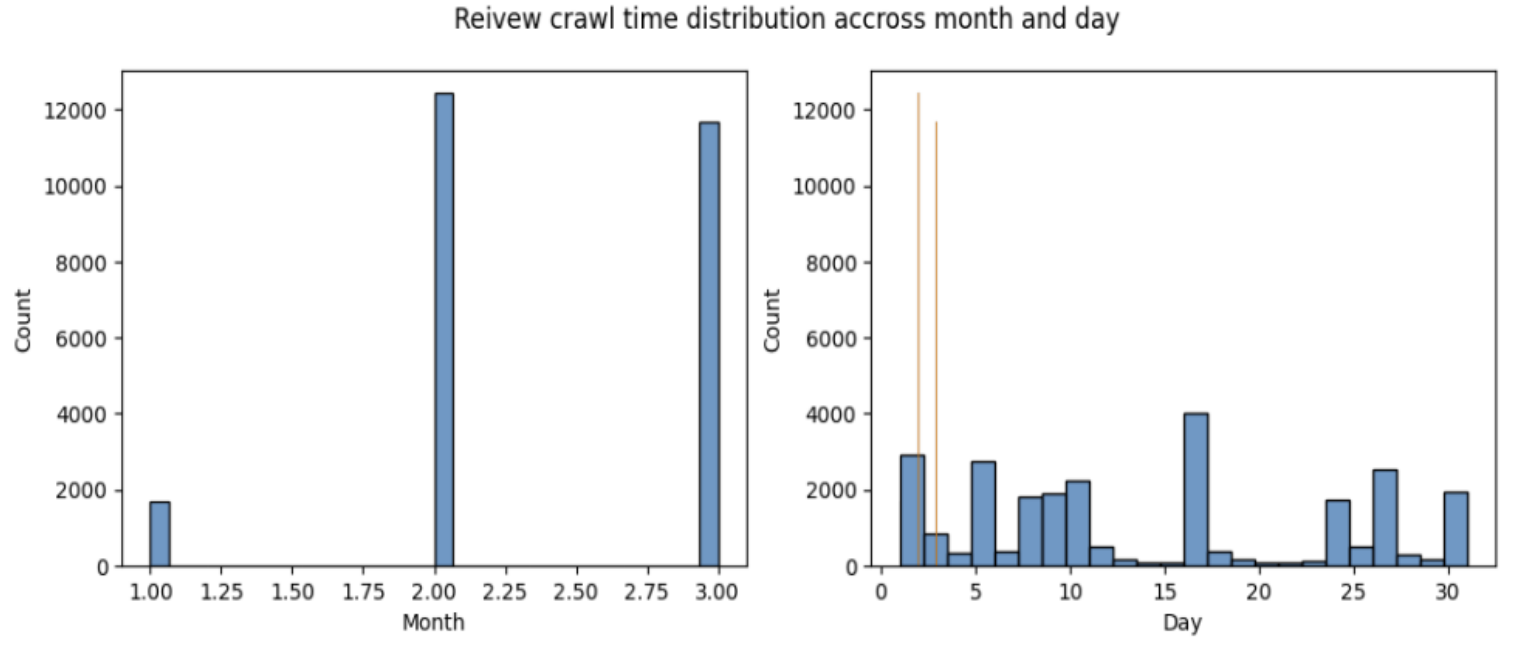
Data wrangling and clean

- Two independently collected Walmart review data. Only 1009 reviews overlapped. Both contain around 30k reviews (30006 vs 29997).

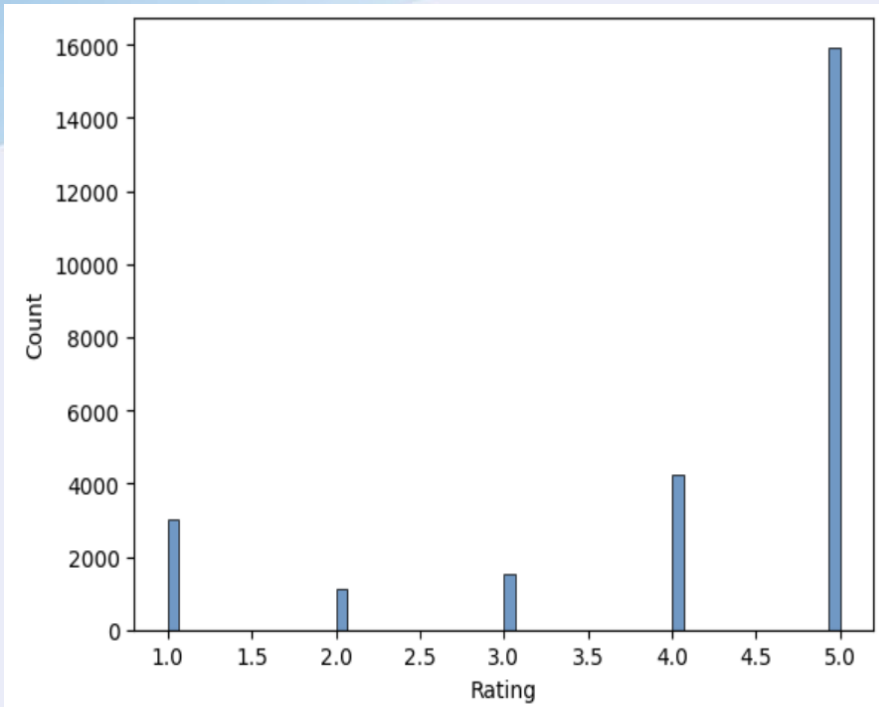
- After dropping 6 unrelated features, I got 13 features including Uniq id, Crawling time, pageurl, website, title, rating, review, reviewer name, review upvotes, review downvotes, verified purchaser, review date, etc. After remove records containing NA, I got 25822 records.

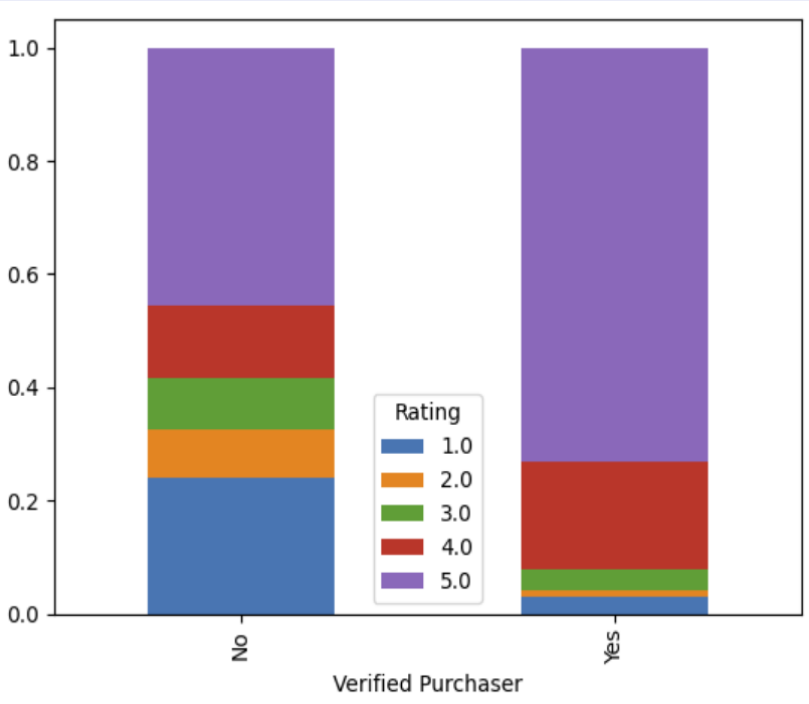
Data Exploration



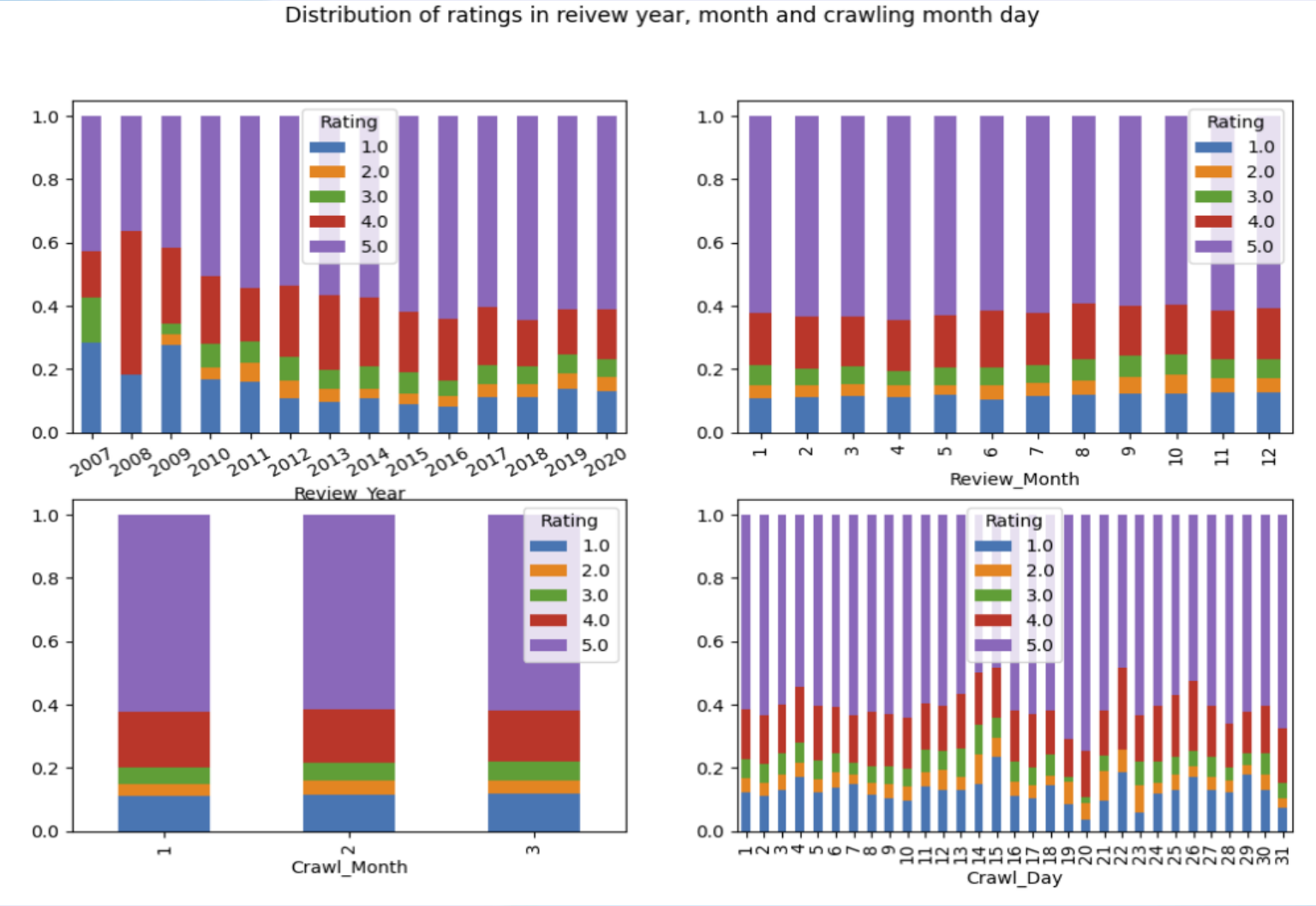


Rating Destribution



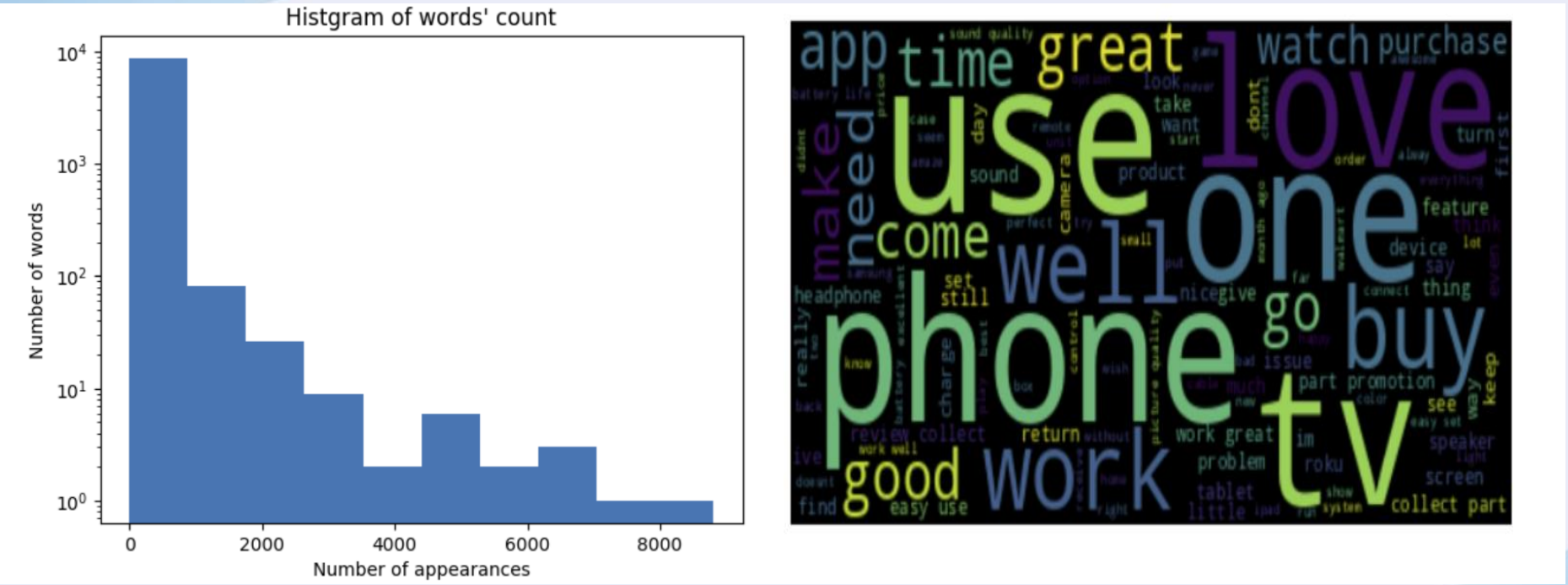


Rating distribution with review date and crawl date.

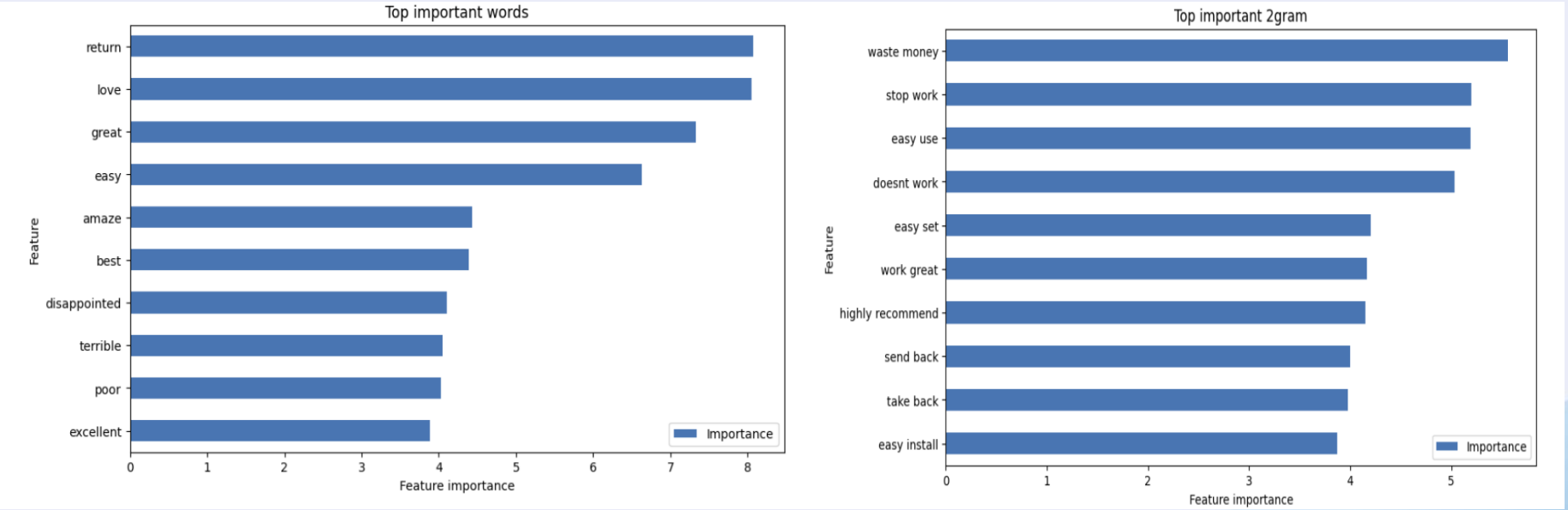


Sentiment analysis

- Words count distribution and top words



Top single word and 2-gram contribute to rating from logistic regression based on tf_idf encoding

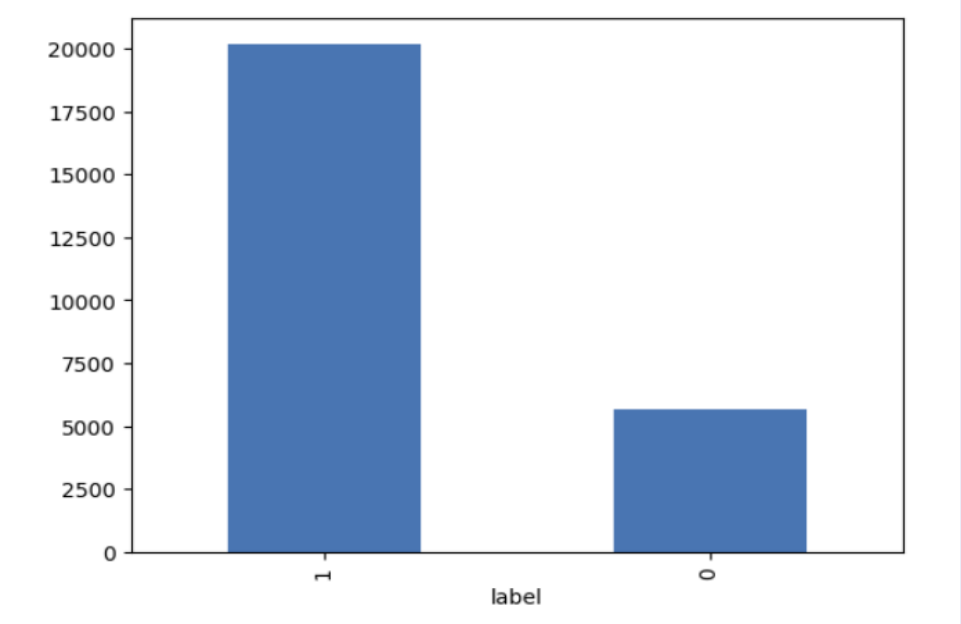


RoBERTa model

- The pre-trained RoBERTa model we choose is downloaded from hugging face cardiffnlp/twitter-roberta-base-sentiment-latest. This is a RoBERTa-base model trained on ~154M tweets from January 2018 to December 2022, and fine-tuned for sentiment analysis with the TweetEval benchmark. According to their paper, even for the original model, their performance on sentiment is better than base RoBERTa model.

- I used GPU to decrease the computational time significantly. Accuracy is selected as the metric to tuning the parameters. All the data are preprocessed to padding with max_length=128. The learning rate is chosen as 1e-5, weight_decay as 0.01. The model is trained for 10 epochs.

Rating distribution



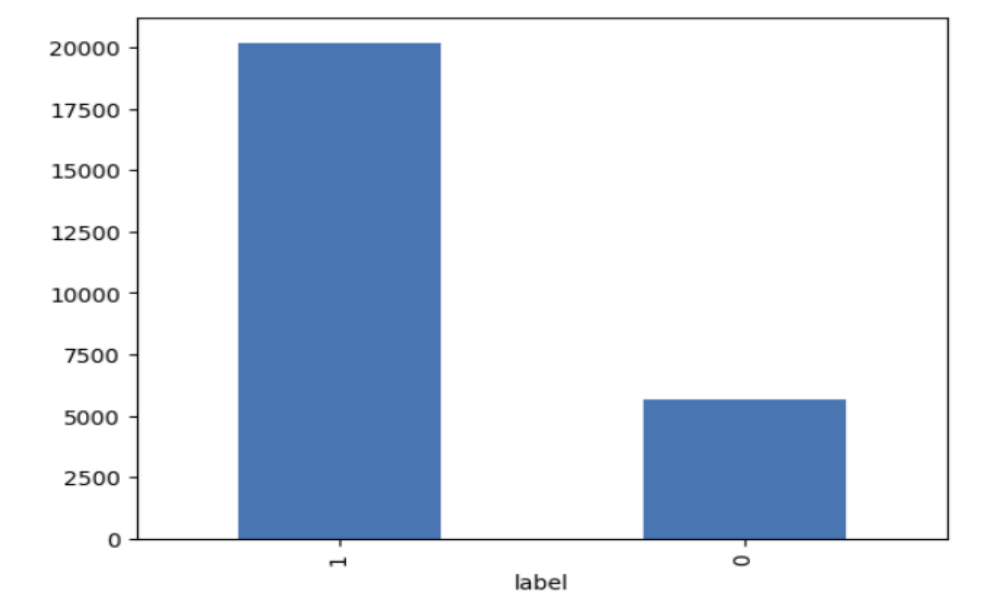
	Train	Valid	Test
Postive	14093	4015	2067
Negative	4033	1098	515

Performance on test data

label	Precision	Recall	F1-score
0	0.8	0.77	0.78
1	0.94	0.95	0.95

The overall accuracy is 92%.

Independent validation



label	Precision	Recall	F1-score
1	0.84	0.74	0.78
0	0.93	0.96	0.94

The overall accuracy is 91%

Take away

- Sentiment analysis can help the retailer to find out the key points to improve their service to promote customers experience.

- Tf-idf encoding performs great in the analysis.

- With pre-trained RoBERTa model from tweet and fine-tuning with Walmart review data, our model achieve accuracy 91%. The model performs even better on positive rating reviews which is as high as 95% in F1-score.

Future Plan

- We can build a real-time application that will automatically update the review data and detect the key points to improve the customer service.

- We can also run the sentiment analysis with different time range which will help us understanding customer's request changing and validating the outcome of improving customer service.

- We find that there is still a big room to improve the performance on negative rating prediction.