

Natural language sentiment analysis and modeling on Walmart customer reviews

Wenhui Wang

Springboard

02/22/2024

The problem

The screenshot shows the Walmart website interface for an Apple Watch. The top navigation bar includes the Walmart logo, search bar, and links to Departments and Services. A search bar contains the text "Search everything at Walmart online and in store". Below the navigation bar, there are filters for shipping options (3+ day shipping, 2-day shipping) and a "Download Reviews (60)" button. The main content area displays the product name "Apple Watch", a star rating of 4.6 out of 5, and a price of \$149.00. A green button labeled "Download Reviews" is visible. An orange arrow points from the "Download Reviews" button on the website to a similar button in a software application window titled "WalmartReviewsExtractor_60_reviews_2023-08-30". The software window shows a spreadsheet with columns for Product Name, Review ID, Reviewer Name, Review Title, Review Text, Positive Feedback, Negative Feedback, Verified Purchaser, Review Time, Images, Product Link, Author ID, Reviewer Location, and Scraped At. The spreadsheet contains 60 rows of data, each representing a customer review.

Product Name	Review ID	Reviewer Name	Review Title	Review Text	Positive Feedback	Negative Feedback	Verified Purchaser	Review Time	Images	Product Link	Author ID	Reviewer Location	Scraped At
Apple Watch	312405579	Tran	5 Excellent value!	The watch is exactly what I need	0	0	Yes	7/27/2023		https://www.dob83b52367eeb558194	2023-08-29T17:15:07		
Apple Watch	207400053	Daniel	5 Really cool, you don't	I love it. I started working as a	29	9	Yes	10/2/2022		https://www.folled2ced6ea74b5aef36d	2023-08-29T17:15:07		
Apple Watch	291902020	Tonkian	5 Like it	like it, still learning about the fe	26	6	Yes	11/16/2022		https://www.22c136840db263c5e5b75	2023-08-29T17:15:07		
Apple Watch	293235392	amulfo	5 apples 3dgl watch	thank you so much thought it was	11	4	Yes	11/26/2022		https://www.bbec2e1071fe8e4da42435	2023-08-29T17:15:07		
Apple Watch	292094162	Zoe	5 Grate	Wonderful watch I did, Ake espe	9	3	Yes	11/24/2022		https://www.cfeaf55c4ec7625d243533	2023-08-29T17:15:07		
Apple Watch	313970052	Jordan	4 Steel for the price I paid	As first I was skeptical because it	0	0	Yes	8/11/2023		https://www.20f18ca562ca490ba15a6	2023-08-29T17:15:07		
Apple Watch	311349872	Rougan	4 Pictures Includes I	When I received the apple watch	0	0	Yes	7/15/2023		https://www.fefdf13d36d94c3b04e5b6b1	2023-08-29T17:15:07		
Apple Watch	311076761	Fernando	1 Money Sacem	Purchase is a nightmare item wal	0	0	Yes	7/13/2023		https://www.nc20362d77a6da413611	2023-08-29T17:15:07		
Apple Watch	309909117	Borrorive	1 Just worked for 4 month	I called Apple Watch, it's Apple i	0	0	Yes	5/24/2023		https://www.nc050a1193ba79711a05a4	2023-08-29T17:15:07		
Apple Watch	309530450	Symethia	1 Missing item Walmart	cMy package was open when I got	0	0	Yes	6/27/2023		https://www.0f7b09b46-315427c766ac	2023-08-29T17:15:07		
Apple Watch	310996037	morgan	5 I love it	I LOVE THIS. Got it on sale and i	0	0	Yes	7/12/2023		https://www.0139fw068ba46fb73167d4	2023-08-29T17:15:07		
Apple Watch	294545950	Courtney	5 Love it!	When my watch arrived in the nu	39	10	Yes	8/19/2022		https://www.e10d9fw068ba46fb73167d4	2023-08-29T17:15:07		
Apple Watch	288983255	Nathalia	5 Amazing	Came in perfect condition, sealed	8	2	Yes	10/23/2022		https://www.e1202a30519a71e0f0707a	2023-08-29T17:15:07		

- 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

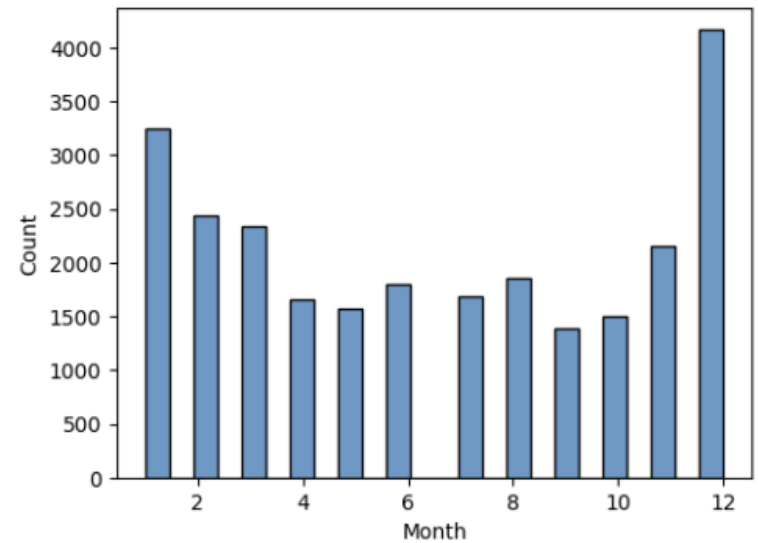
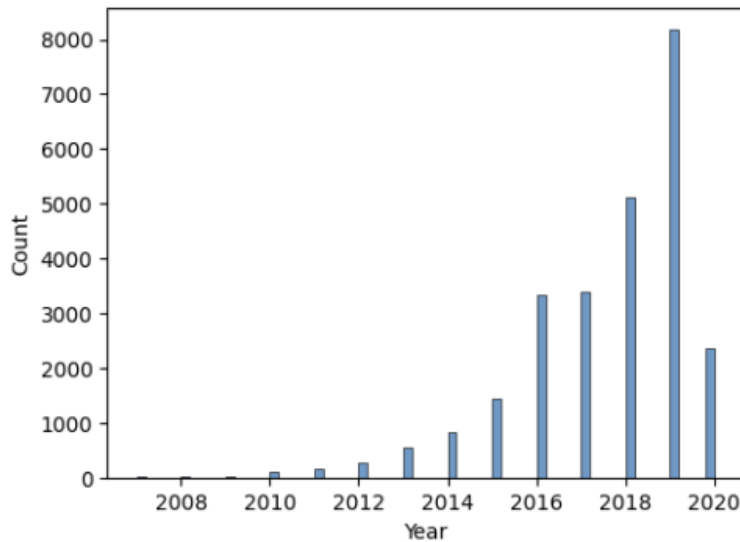
- Running sentiment analysis to find the key features that will promote or impair customers' experience.
- Building model to predict customer rating based on review language by transfer learning and fine tuning on of pre-trained RoBERTa model.

Data wrangling and clean

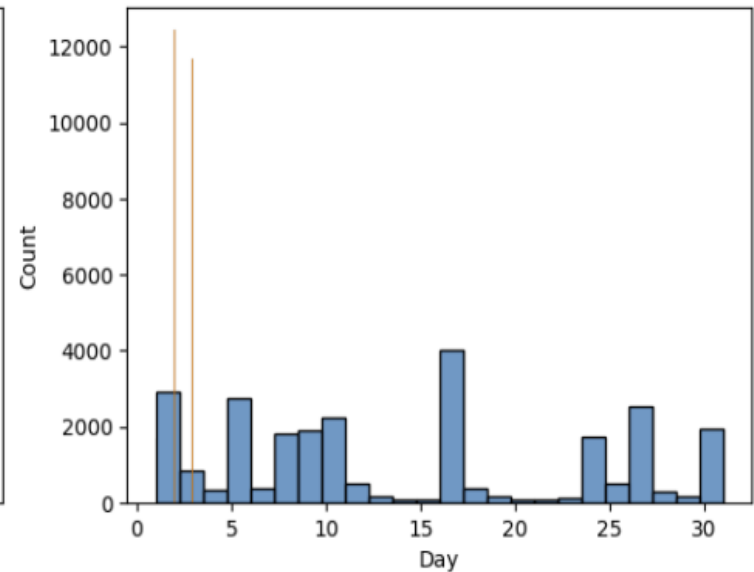
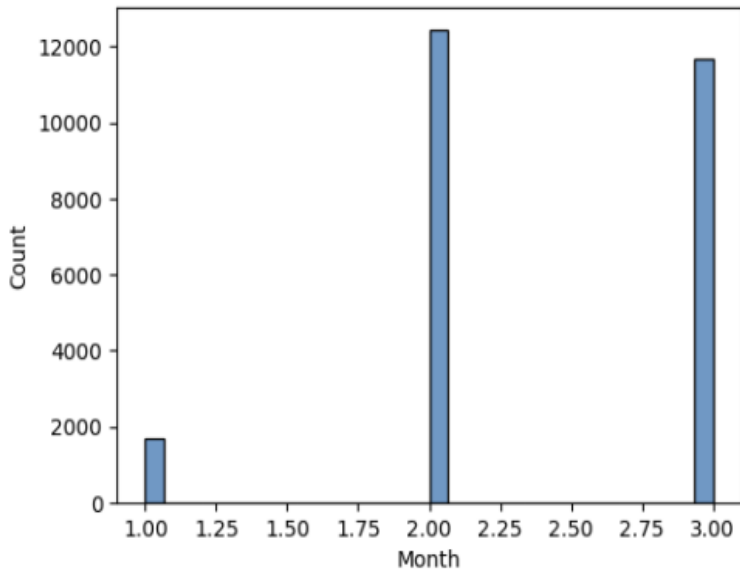
- Two independently collected Walmart review data. Only 1009 reviews overlapped. Both contain around 30k reviews (30006 vs 29997).
- After dropping 6 unrelated features, I got 13 features including Uniq id, Crawling time, pageurl, website, title, rating, review, reviewer name, review upvotes, review downvotes, verified purchaser, review date, etc. After remove records containing NA, I got 25822 records.

Data Exploration

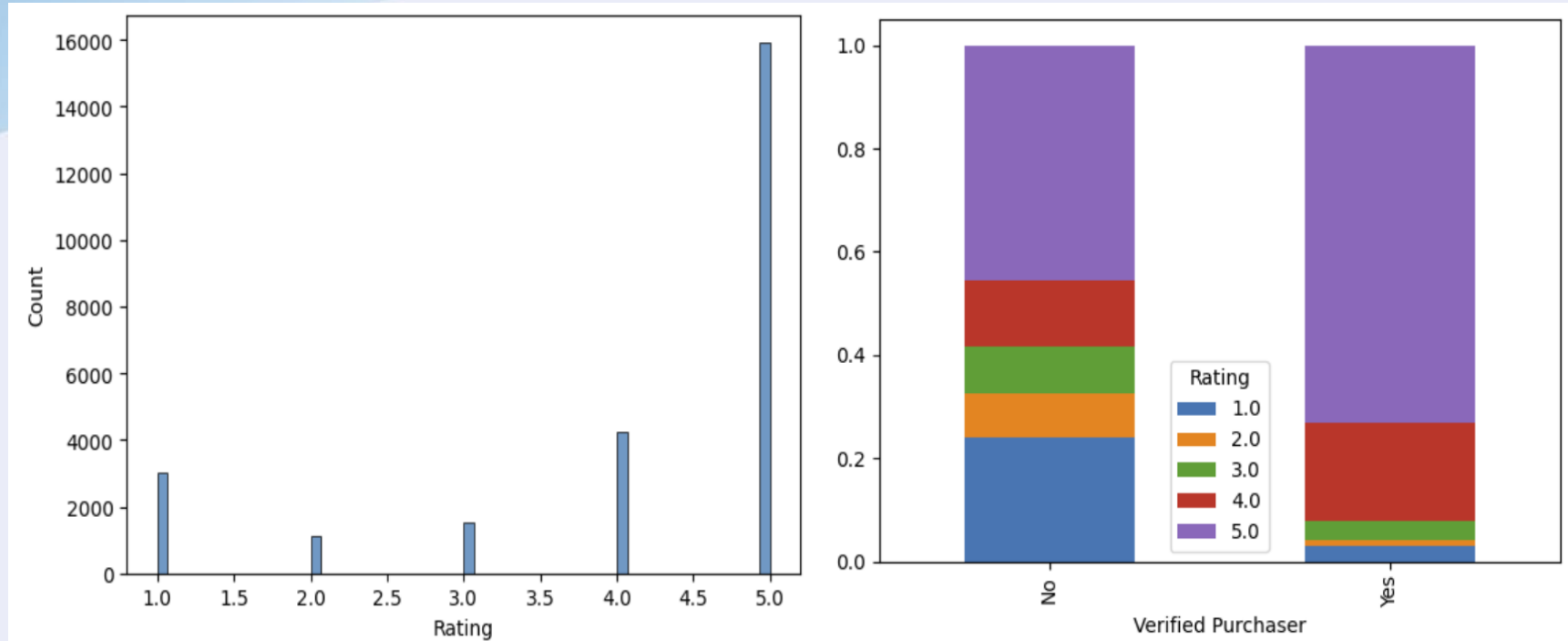
Reivew distribution across Year and Month



Reivew crawl time distribution accross month and day

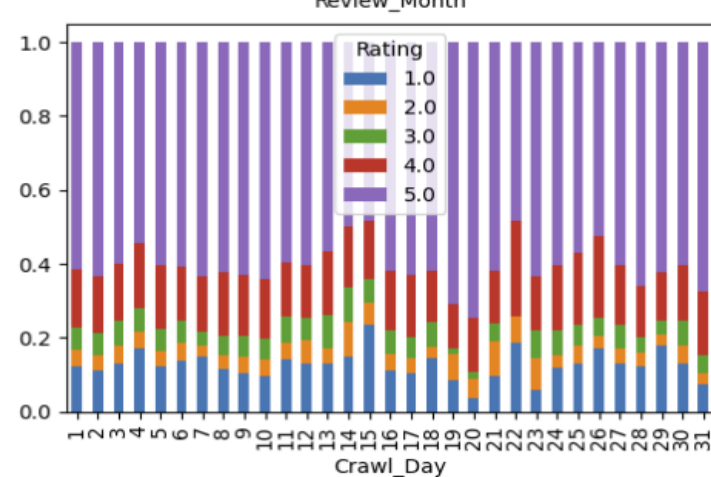
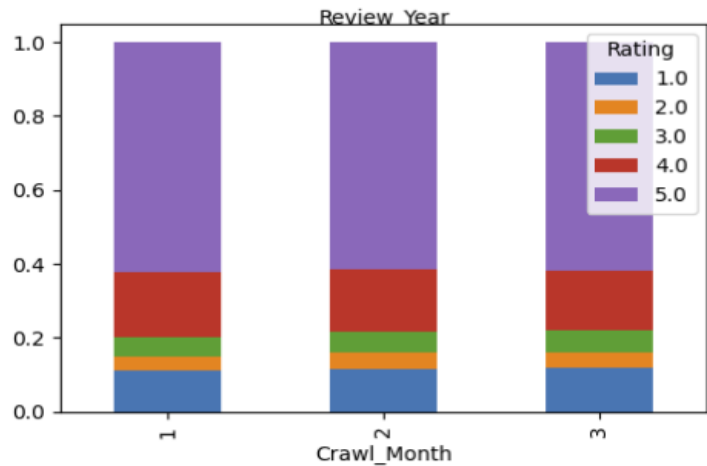
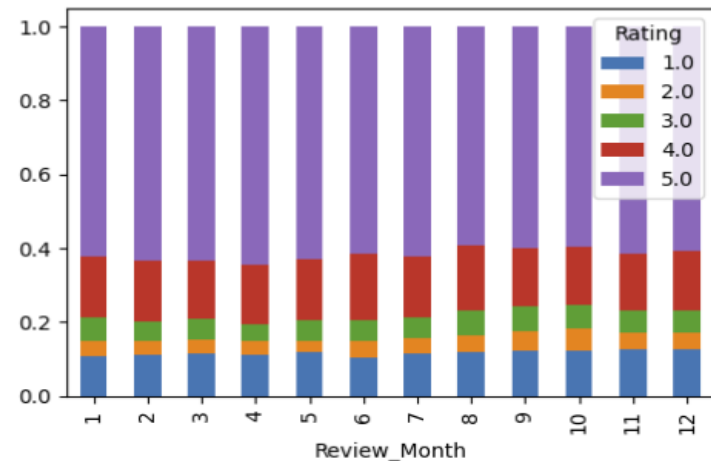


Rating Distribution



Rating distribution with review date and crawl date.

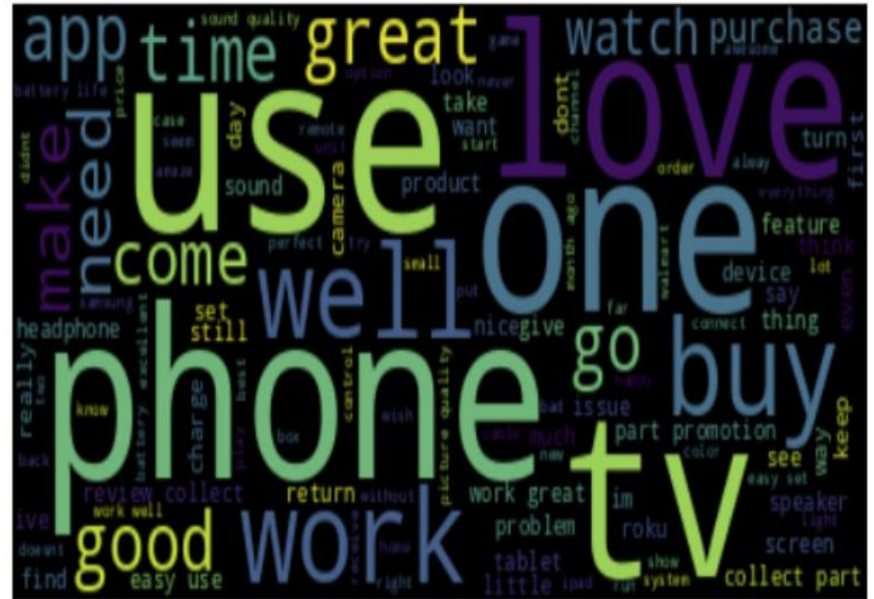
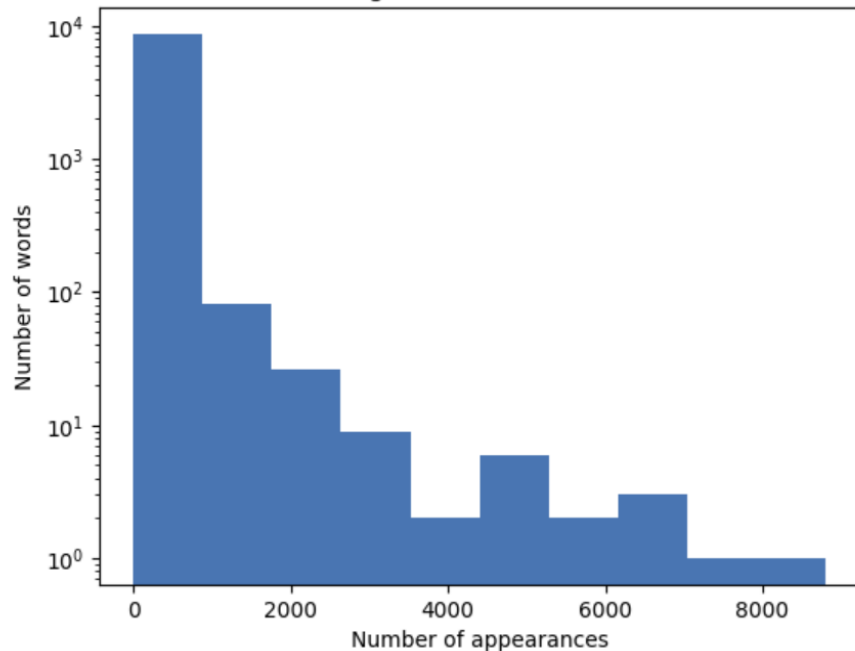
Distribution of ratings in review year, month and crawling month day



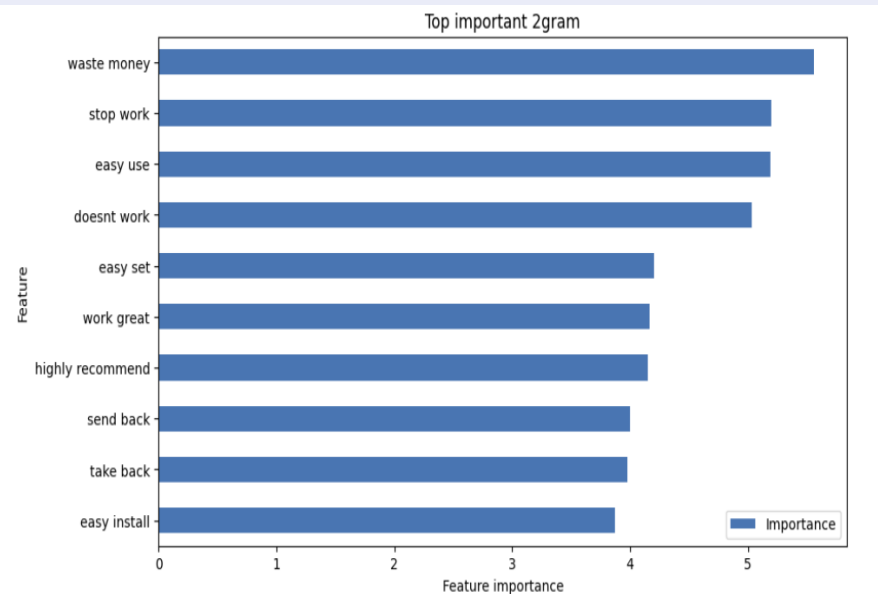
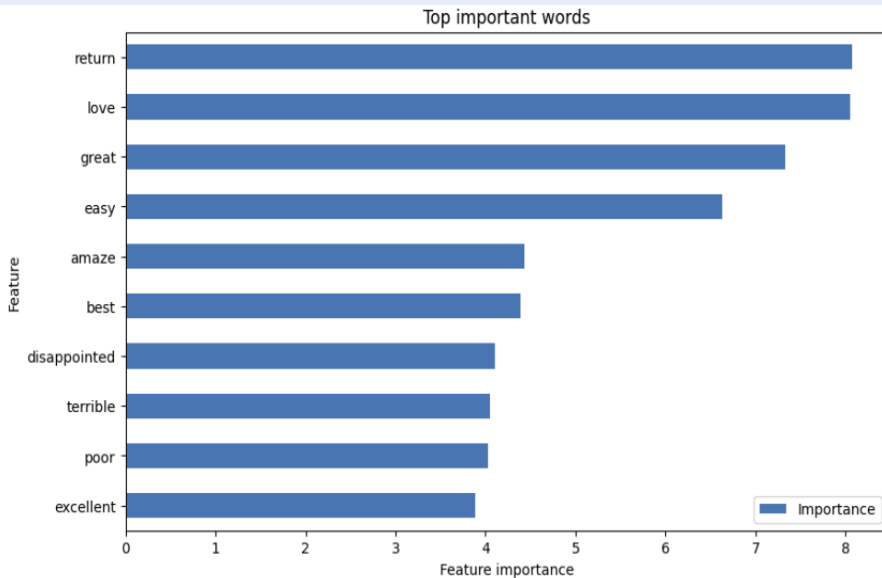
Sentiment analysis

- Words count distribution and top words

Histogram of words' count



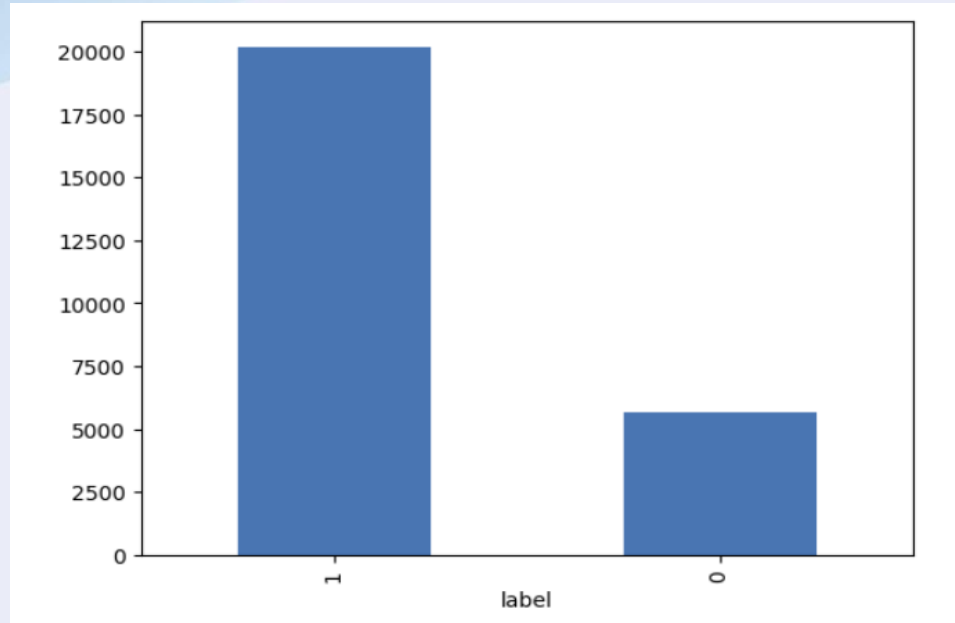
Top single word and 2-gram contribute to rating from logistic regression based on tf_idf encoding



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



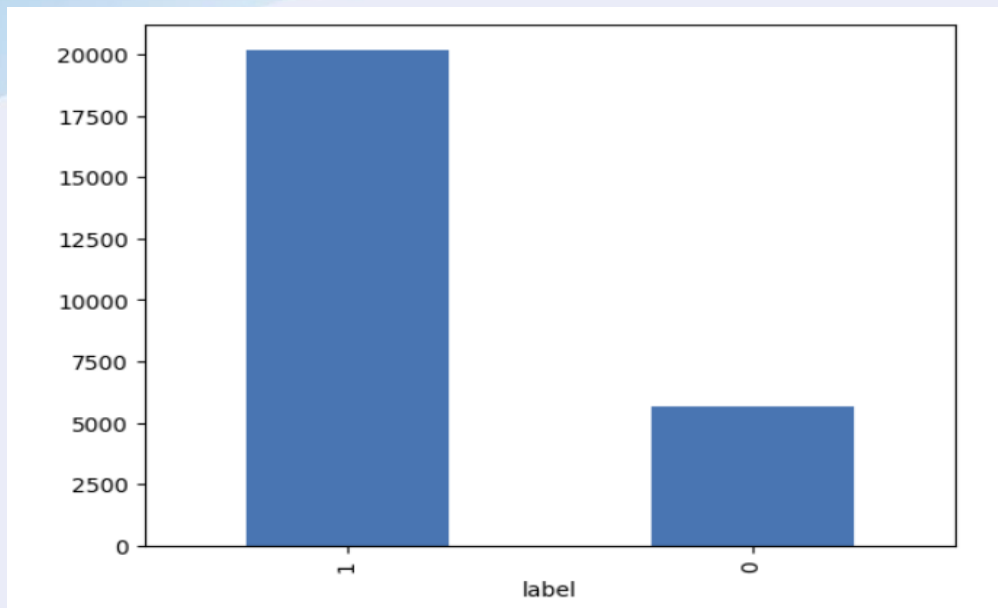
	Train	Valid	Test
<u>Postive</u>	14093	4015	2067
Negative	4033	1098	515

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.