Stephon Mensah

12/5/2023

Advanced Software Engineering

Final Project Report

Many social media apps are used and are given feedback, but most of the time changes and updates are made that the users don't expect or ask for. For my final project I wanted to evaluate the reviews for an app and see if the developers of that said app were actually taking into account the real time reviews of their users and seeing if they actually attempt to improve the app based on that feedback. The specific app that I wanted to use for this analysis and evaluation was Instagram. The social media application Instagram has been out for more than 10 years and the consumer base has only grown since then. It is one of the most downloaded and used application on the apple app store and google play store. I wanted to use Instagram because of the visual nature of it. Aside from captions and comments everything on Instagram is visual. So, when it comes to the comsumer/user it is very easy to identify things that they do or don't like in the application and write reviews, them being positive or negative for the application. Because of the way that this is, I thought that Instagram would be a great application to evaluate for reviews pre and post update and seeing if reviews are actually handled in the proper way and if they are overall handled after significant updates.

For my final project I had to acquire the dataset for the Instagram reviews. Initially I had some issues acquiring this dataset because, I was attempting to web scrape the google play store webpage and create a tool that would automatically get the reviews using selenium. This failed as the tool would only capture a certain amount of the reviews then stop prematurely leaving me with only 3 or 4 reviews. I then moved on the process of using a dataset of Instagram reviews that I got from the website Kaggle.com. From there I was able to get a dataset of Instagram reviews that was up to date and had reviews from 2018 to 2023. The dataset consisted of the following columns. The Instagram review description, the 1 to 5 star rating attached to that review and the date of that review. To add to this great dataset being very simple, easy to read, and easy to use. It also had an extensive amount of reviews for me to use and gave the ability to sample when I was using the different machine learning method. There were more than 210,000 reviews in this dataset.

(Visual of the dataset on the next page)

	review_description	rating	review_date
2	The app is good for connecting with friends, family and e	3	7/11/2023 23:57
3	Used to be my favorite social media app, but "improvem	2	7/22/2023 21:37
4	Instagram is the best of all the social media. IG is not just	5	7/25/2023 3:24
5	I love this app.. but as of late, I have been having probler	2	7/9/2023 4:49
6	Used to be a great app but there are so many more bugs	3	7/17/2023 16:47
7	The app has been glitching alot. Videos will not play and	1	7/11/2023 5:04
8	I like this app a lot, but the messaging part of the app is v	3	7/18/2023 12:48
9	Eventhough sometimes it works, most of the time it does	2	7/13/2023 14:01
10	I'm not sure what the point of leaving a review will accor	1	7/14/2023 2:26
11	This app has so many frustrating aspects. The themes wc	2	7/13/2023 7:46
12	The only complaint i have is about the video calling featu	4	7/20/2023 14:29
13	So a while ago, I noticed that when I talk to people my b	3	7/12/2023 5:02
14	Love the app but everytime I open it, after about ten to t	1	7/9/2023 6:07
15	I was fine with all the updates until I watch video posts o	2	7/26/2023 4:34
16	Says I've had a msg for a year & none of the troubleshoo	1	7/11/2023 2:53
17	Was better before.. Now the comments are one long lin	3	7/23/2023 20:06
18	App switches to dark mode and unuseable. For the past f	1	7/12/2023 13:18
19	Even with light mode on, sometimes the app will turn the	1	7/10/2023 17:21
20	Pretty good app, and I use it frequently, but there's been	3	7/10/2023 10:39
21	As a Photographer, Web Devloper, and Graphic Designer,	3	7/10/2023 0:45
22	Easy to use for the most part, good to talk to friends, but	4	7/16/2023 2:26
23	The app will NOT open anymore. Any attempts to do so	1	7/9/2023 14:50
24	First, I lose the ability to use reels with no real indication	1	7/9/2023 12:49
25	It's currently impossible to edit reels in the app. They sto	2	6/28/2023 11:50
26	The app is honestly my favorite social medias, and the or	3	6/28/2023 21:55

For the techniques and implementation I used a number of different techniques for the implementation of the project. Initially for the cleaning and preprocessing of the dataset I had to extract and clean the Instagram user reviews, then I had to apply text normalization and filtering to the enhanced data quality. I had to filter the reviews using certain keywords so that I could ensure I that I was only using reviews that were relevant to the evaluation that I was trying to do. These were the key words, update, fix, bug, issue, crash, glitch, improve, patch, version. Next I used Named Entity recognition or NER. This is a technique that utilized a python library called spaCy NLP which extract the entity. The next portion or technique was a very important technique for the overall evaluation of the project, that being sentiment analysis. Using a python library called NLTK's sentiment intensity analyzer, I had to classify the sentiment of the reviews in the data set after filtering. So this tool would basically categorize the review as positive or negative, later in the paper will be referred to as class 0 and class 1. 0 for negative and 1 for positive. Next, was the TF-IDF vectorizer, which is a tool that prepares the dataset for the support vector machine. This TF-IDF converts text data into numerical format so that machine learning algorithms can process it. Then lastly for final evaluation I used the SVM for classification of the sentiments

For the final evaluation the plan was to use specific evaluation metrics to measure the performance of first the quality of the dataset using SVM, then lastly the quality of the actual tool that I developed. This was done by creating an alignment score, which determined how many issues or features mentioned in the reviews had corresponding mention in subsequent change logs. And User satisfaction post update, which analyzed the message of the reviews after the update to gauge if user sentiment improved, confirming tat the updates addressed concerns. For the SVM, I wanted to capture the Accuracy, precision, recall and F1 score of the dataset and the two sentiments.

Metric	Score	Class 0	Class 1	Support (Class 0)	Support (Class 1)
Accuracy	0.86	-	-	-	-
Precision	-	0.86	0.78	2861	492
Recall	-	1.00	0.08	2861	492
F1-Score	-	0.92	0.14	2861	492

So, as you can see for the results from the SVM analysis, we measured accuracy, precision and recall. Class 0 is the negative sentiments and class 1 is the positive sentiments. So, from the results we glean that the overall accuracy of the sampled dataset is good at 86 percent. For class 0 and 1 the precision is also strong at 86 and 78 percent. Recall is strong for class 0 but not for class 1 and this is most like because of the imbalance of the amounts for the sentiments, the slow quantity of sentiments for class 1 lead to the low recall. Lastly the f1 scores reflect the recall for the two classes. The support is just the number of occurencess of each class in the specific dataset that was sampled.

Metric	Score
Alignment Score	-0.014967372699510095
Average Sentiment Pre-Update	0.17486522911051214
Average Sentiment Post-Update	0.13585149612116734
Change in User Satisfaction	-0.039013732989344796

For this evaluation we got the alignment score, the average sentiment pre update and post update and the change in user satisfaction. The alignment score tells us that there are less reported issues or different focus areas in the post update reviews. The average sentiment pre update number tells us that most of the reviews were negative. The Average sentiment post update tells us that there was an increase in negative reviews. This is contradictory of the alignment score because there could be and increase in negative reviews but they may not be addressing the update that occurred or just off topic in general. Lastly, the Change in user satisfaction, which tells us that the user satisfaction worsened following the update. Overall we glean that for the specific update date that I evaluated, the reviews did not get any better.

In conclusion, some improvement for the future would be to evaluate different update dates rather than one and see the performance of those and compare. Also, sampling a larger dataset.