DS 5110

Final Project Report
Sentiment Analysis of Spotify Comments and Topic Modeling and Analysis

Pengli Shao

Deadline: 07/31/2024, 11:59 PM

1 Introduction

App reviews contain users' feedback and their sentiment toward the app. Companies cannot find out the hidden messages and sentiments from their users only by app scores. Moreover, it is expensive to hire people to analyze the hidden messages from comments manually.
Spotify is one of the mainstream music apps that has tons of hidden information in it's reviews, some users may give high ratings on Spotify but express their dissatisfaction with the app in their reviews. Therefore, identifying user's sentiment by their comments and finding out the topics of negative reviews is the key of solving existing problems in the next version.

2 Literature Review

In *How do users like this feature? a fine grained sentiment analysis of app reviews.* [1] by Guzman and Maalej, 2014, they conducted a sentiment analysis and topic modeling research on 32,210 reviews from 7 apps on IOS and Android. Their average precision is about 51% and the maximum percision is about 91%, their maximum recall is 73% and the average recall is also 51%. However, they used a tool called SentiStrength that is design for sentiment analysis for short and low quality contents based on rules and dictionary.
In this project, I used BERT(Bidirectional Encoder Representations from Transformers), a model developed by Google in 2018 that read the content from both direction in pre-training. In *BERT: a sentiment analysis odyssey* [2] by Alaparthi and Mishra, they conducted sentiment analysis for 50,000 reviews and used 4 kinds of sentiment analysis tools including BERT. They found out that precision of Bert for sentiment analysis is 92.35%, and the recall score of 92.31% represents BERT is capable of correctly identify users' sentiments. The major different of BERT and all other rule & dictionary based sentiment classifiers is that BERT is a unsupervised model that is pre-trained by large datasets like Wikipedia, and it is capable of reading the content from both direction to understand the context.

3 Methodology

1. Data Collection

The Spotify reviews data is from:
https://www.kaggle.com/datasets/ashishkumarak/spotify-reviews-playstore-daily-update
This dataset originally contains 84,165 rows of data from year 2018 to 2024 on Google Play Store.

2. Data Structure

Datasets contains columns:

- review id
- user name
- content
- score
- thumbs up count
- review created version
- review date
- app version

About 40% of the rating score for Spotify is 1. Most reviews are from 2024.

3. Data Cleaning

For this project, review is the most important feature, I simply dropped all the rows that contains NA values. Filling NAs with random value or mean value does not apply in this project because all important variables are strings or date time type.
In addition, I transform the review date column's date type from string to datetime data type.

4. Text Mining

- Remove all the non-English characters
- Remove all symbols except question mark and exclamation mark
- Lower the text
- Removed stop words
  Ex: `[a, an, the, you, he, in, on, and,...]`
- Lemmatization
  Ex: `['running', 'became','ate', 'makes']` -¿ `['run', 'become', 'eat', 'make']`

5. Sentiment Analysis

I conducted sentiment classification by three different tools:

vaderSentiment.SentimentIntensityAnalyzer

A rule and dictionary based sentiment analysis tool designed for short and low quality review contents. It rate the sentiment intensity score for each review, the range of score is between -4 to +4. Vader also consider intensifiers, negations, and punctuation marks such as question mark and exclamation.

- Pros:
  - Fast
  - Understand slangs, typo, and abbreviations
  - Process intensifiers, negations, and punctuation marks.
- Cons:
  - Can't understand context
  - Limited understanding ability when the target word is not in it's dictionary

The output generated by Vader suggests that there are over 60% of the comments are positive. Some positive reviews were actually negative.

TextBlob

A sentiment classification tool based on dictionary, patterns, and machine learning algorithms. It tokenize the words and rate them from -1(negative) to 1(positive),

- Pros:
  - Fast
  - Easy
  - A certain level of ability of understanding the context
- Cons:
  - Does not process intensifiers, negations, and punctuation marks.

The output of TextBlob suggests that there are over 60% of the reviews are positive. It is not capable of understanding the context.

BERT

BERT is a deep learning model that read the content from both direction (right-to-left and left-to-right), and pre-trained by large datasets. It will tokenize the sentence and embedding the words into a continuous low dimension vector. Those features make BERT capable of understanding complex context and classify sentiment of each reviews with high precision and recall.

- Pros:
  - Understand context
  - Understand slangs, typo, and abbreviations
  - Process intensifiers, negations, and punctuation marks.
- Cons:
  - High computational resource requirements
  - It only classify sentiment to Positive and Negative, no neutral.

The output of BERT suggests that there are over 80% of the comments are negative. I compared the performance of three tools and decided to use BERT for further analysis based on the precision and recall score provided from research conducted by Alaparthi and Mishra [2].

6. Keyword Extraction

TF-IDF(Term Frequency-Inverse Document Frequency) is a method to extract keywords from a set of documents.

$$TermFrequncy(t, d) = TF(t, d) = \frac{f_{t,d}}{N_d} \tag{1}$$

$$\tag{2}$$

where $f_{t,d}$ is the time of text t that appears in review d
N_d is the number of text in review d

$$InverseDocumentFrequency(t, D) = IDF(t, D) = \frac{N_D}{n_{t,D}} \tag{3}$$

where N_D is the number of total reviews
$n_{t,D}$ is the number of reviews that contains text t

TF-IDF method extract those words that is frequent in some specific reviews(high TF) and relatively less common in the whole dataset(High IDF). I extracted 200 keywords by TF-IDF and examine them. However, it is impossible to identify the hidden message within those keywords without futher process.
Afterward, the reviews were separated into positive and negative groups and calculated the average TF-IDF score for each keywords within each group of reviews to determine which words are most representative in positive and negative reviews.

7. Topic Modeling

Topic modeling is a method to find the hidden topics/structure of a set of reviews. I used LDA(Latent Dirichlet Allocation) algorithm to find out those topics.
LDA is a probilistic distribution algorithm that assumes each review is a combination of multiple topics, each topic is a combination of multiple keywords. It infers the topic distribution θ for each reviews, the word distribution ϕ for each topic, and the topic assignment for z for each word by Gibbs sampling. Therefore, each review is related to

some topics, and each topic is related to some keywords.
After topic modeling, the specific problem of each topic can be found by showing the sample reviews of each topic.

4 Results

The result of BERT sentiment classifier illustrate that there are over 60,000 reviews are negative and only about 15,000 reviews are positive, as shown in figure 1.

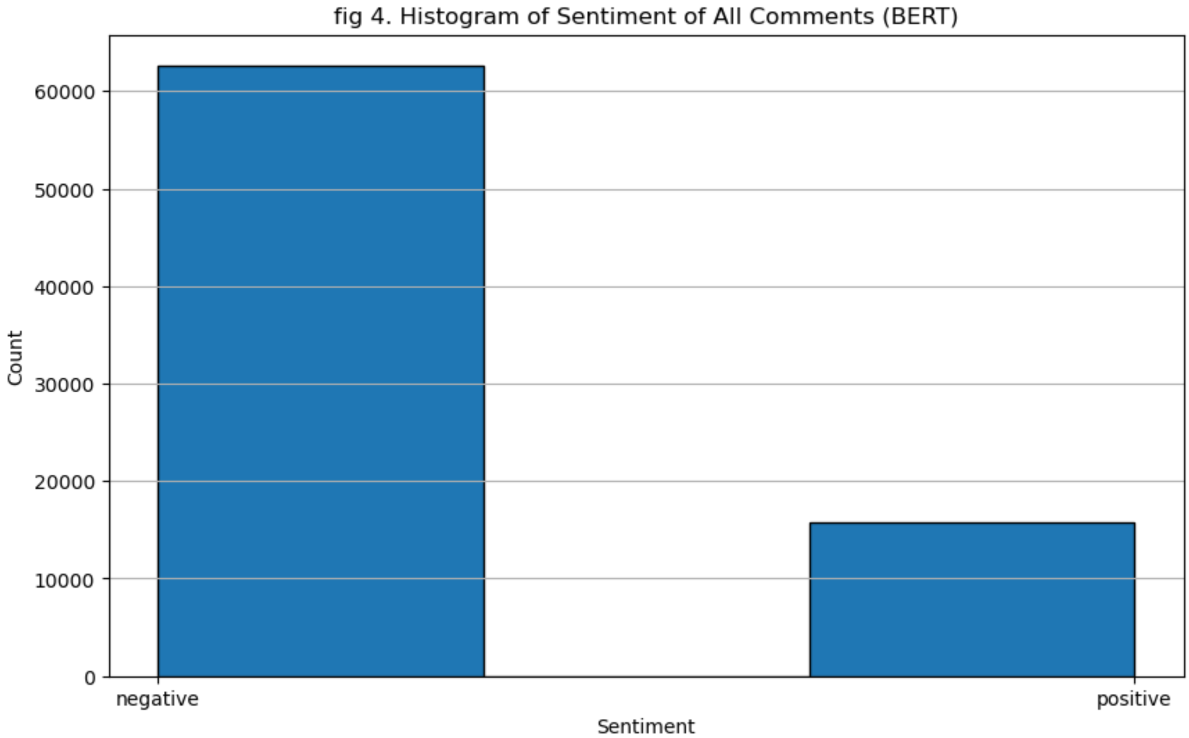

Figure 1: Sentiment Distribution-BERT

Moreover, the correlation of review score and Bert sentiment is 0.5179, which means there is a positive correlation between user's sentiment and it's review score.

Keyword extraction is the key to find out the reason that majority of the users are having negative sentiment.

Positive Keywords	**TF-IDF Score**	**Negative Keywords**	**TF-IDF Score**
music	0.119998	song	0.101859
app	0.090779	app	0.087978
song	0.084093	play	0.073879
love	0.077557	spotify	0.064396
spotify	0.075730	music	0.062172
good	0.071855	listen	0.053344
great	0.068227	playlist	0.051540
listen	0.066737	premium	0.051438
best	0.052691	use	0.050277
playlist	0.050568	ad	0.049915
like	0.048383	get	0.048828
premium	0.043506	like	0.048443
use	0.041542	cant	0.047533
play	0.040766	update	0.042985
really	0.037527	even	0.042940
ad	0.034648	time	0.038137
get	0.033452	want	0.037489
make	0.029722	dont	0.037455
add	0.029074	im	0.036504
find	0.028898	go	0.034573

Table 1: Top Positive and Negative Keywords with TF-IDF Scores

From table 1, there are some unique words such as *update* that only show up in negative keywords. However, most words are in both positive and negative keyword group. The negative keywords section will further process negative keywords and create topics for negative keywords to identify the true problem pointed out by users.

In our topic modeling, we found 5 topics for negative reviews, each topics contains 10 negative keywords.

Topic	**Top 10 Words**
Topic #0	year, connect, work, device, service, download, use, music, app, spotify
Topic #1	music, artist, want, listen, add, shuffle, like, play, playlist, song
Topic #2	pay, music, song, premium, listen, 30, minute, free, get, ad
Topic #3	issue, music, time, song, fix, phone, work, stop, play, app
Topic #4	skip, lyric, even, worst, spotify, listen, app, cant, premium, song

Table 2: Top 10 words for each topic

We can find the potential problems by top 10 words of each topic

- Topic 0
  The poor service and connection quality such as download and stop working on user's device.
- Topic 1
  The problem of features such as playlist, add, shuffle.
- Topic 2
  The unsatisfaction of free users experiencing ads.
- Topic 3
  The problem of stop working.
- Topic 4
  The problem of skip function and lyrics.

Sample of each topic:

Topic #0

Top 10 words: year, connect, work, device, service, download, use, music, app, spotify

> *Used to love this app, but I'm extremely frustrated with it now. I pay for premium, and no matter how good my Internet connection is, I get told I need to "go online to see the menu." I can't even access the songs I've downloaded - isn't the whole point of downloading them to be able to access them offline? This has been an issue for months now, but recently it happens pretty much any time I try to use the app. Think it's time to start looking for other services.*

Topic #1

Top 10 words: music, artist, want, listen, add, shuffle, like, play, playlist, song

> *Its great i found most of the music i wanted to hear, the only problem is im not able to listen to songs on their own anymore only playlists, and now any time i make a playlist spotify puts random songs in there that i never added and never even liked and it wont let me remove them, im also not allowed to skip more than 6 songs an hour which is annoying because then if i have another song i dont like and want to skip i just have to live through it, turn my volume on mute or go to youtube.*

Topic #2

Top 10 words: pay, music, song, premium, listen, 30, minute, free, get, ad

> *I've used this app for years and have had premium for a good portion of that well recently had to switch cards and didn't get to it before my benefits ended and wanted to try listening to music without it but god its horrible now. to get the song I'm even trying to play I have to wait possibly hours. The ads are outrageous now. And show before the songs. And I can't even see what songs are in my Playlists? Ridiculous and all behind a pay wall? Making YouTube music seem like a great option is sad.*

Topic #3

Top 10 words: issue, music, time, song, fix, phone, work, stop, play, app

> *This app is something I use every day, so I would be able to tell if something was off. And ever since the new update(I would believe) things just kept messing up. Like with commercials; it just keep skipping between three different commercials before deciding to just quit and continue with the music. How the spotify thing in the notif bar sometimes doesn't even appear while music is playing and how the music just stops abruptly without you even doing anything. Also, when they do show the songs on the notif bar it doesn't even show the correct song that's playing; because of this, I can't even click play or pause or even hear the song I'm listening to. And sometimes when on the app, after a commercial, most of the time the screen turns black, unless I go to home screen then enter the app again so the song playing will show up. I love this app, I truly do. But it is also very glitchy and not user friendly.*

Topic #4

Top 10 words: skip, lyric, even, worst, spotify, listen, app, cant, premium, song

> *I have been a long time user of spotify and while I still think it is one of the better options for music streaming out there, I am saddened by all of the features that are being taken away from free users. When I first started using spotify I was even able to choose each song I wanted to play without needing premium. Now I can't even add songs to my playlist from an artists discography. I understand that spotify needs a reliable way to make money but I see it better to add features instead.*

From the above topics and their samples, we found out that there are intersections between each topics. For example, topic 2 and topic 5 all contains premium, topic 0 and topic 3 pointed out the problem of not working on their phone/device, and free users are unsatisfied with ads and limitations. The topic modeling successfully identify potential problems hidden in negative reviews and all those problems are the key to improve the app in future updates.

5 Discussion

In our findings, the correlation of user's sentiment and review score is 0.5179, which is very close to the correlation tested in Guzman and Maalej's research [1] of 0.592. Therefore, our assumption that user's sentiment will not be revealed by review score is not fully supported, there is actually a positive correlation between sentiment and score.
The intersection of topics is inevitable for apps that do not contains lots of features, espicially for music app like spotify. Guzman and Maalej's research [1] also pointed out that they have the same problem of similar topics, one of the potential solution is to lower the number of the topics. But, the topic modeling in this research illustrated critical problems of spotify by filtering most of the insignificant words, each topic consists of several features and issues.
The reason that this research do not present the precision and recall score is because the data used in Alaparthi and Mishra's research [2] already classified each review's sentiment. Therefore, they can calculate the precision and recall score for BERT. In my case, the Spotify dataset contains over 84,000 rows of data before cleaned, and about 78,000 rows of data after cleaned, it is not possible for me to identify the sentiment of each review manually, and it is not realistic to use review score for calculating precision and recall score because in our original theory, the review score cannot reflect the user's sentiment correctly.

6 Conclusion

6.1 Key Findings

Sentiment Analysis

- The BERT sentiment analysis classified over 60,000 reviews as negative and approximately 15,000 reviews are positive. There is a positive correlation of 0.5179 between the user's sentiment and their review score, suggesting that higher ratings are generally associated with positive sentiments. The figure 2 in section 8 Appendix also support our finding.
  According to Figure 3 in section 8 Appendix, the overall trend of negative sentiment is increasing with a much higher rate compare to the number of positive reviews from year 2018 to year 2024. As a result, the problem of negative sentiment within Spotify users is becoming a critical problem for Spotify.

- Five topics were identified within negative reviews using the LDA algorithm suggests that the potential problem of Spotify is related to premium, connection, ads, stability, and functionality. The intersection of topics suggests that some of these problems are more critical than others such as premium and app not working.

7 References

References

[1] Guzman, E., & Maalej, W. (2014, August). How do users like this feature? a fine grained sentiment analysis of app reviews. In *2014 IEEE 22nd international requirements engineering conference (RE)* (pp. 153-162). IEEE.

[2] Alaparthi, S., & Mishra, M. (2021). BERT: A sentiment analysis odyssey. Journal of Marketing Analytics, 9(2), 118-126.

8 Appendix: Additional Figures

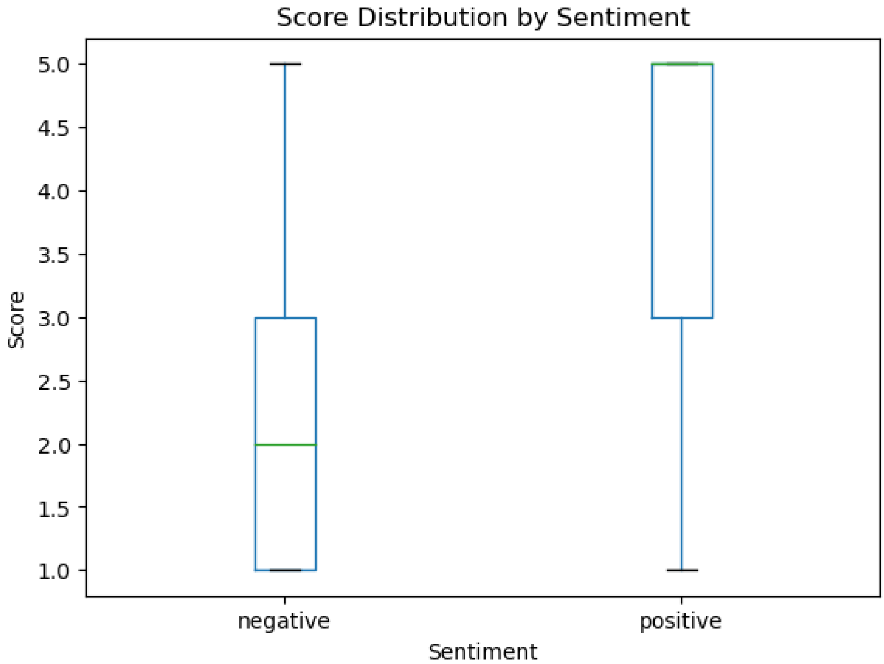

Figure 2: Score Distribution by Sentiment

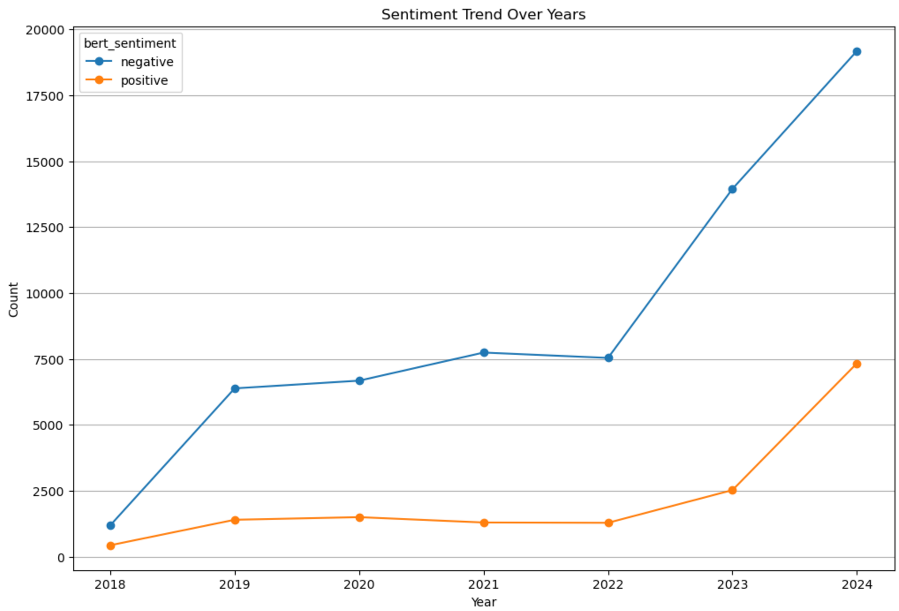

Figure 3: Sentiment Trend Over Time