Google Play Store Review Analysis for Bumble

Final Project Report

MSBX 5420 – Unstructured and Distributed Data Modeling and Analysis
Spring 2021

Team 10: *Santa's Unstructured Helpers*
Elena Heiner
Sarah Holmes
Krista Labovich
Erica Lee
Jack Santaniello

28 April 2022

Table of Contents

Background & Motivation

Meeting people can be hard…whether you want to find a romantic partner, a new friend, or even a new business partner, it can be hard to know where to go to meet the right person or people. Especially since the pandemic, Americans have experienced loneliness on a new level[1], and many have decided to look for social connection through social apps[2]. Searching for the right platform to meet people, may look different for different people. Some people may download a bunch of apps, and switch back and forth, trying to find the right one, while other people may do more careful research on whether an app seems like it may have the crowd and experience they are looking for. So how do apps communicate that they may be the right app for the user, even before the user downloads them? One way may be through previous reviews from other users. Reviews can be the make or break of a prospecting user debating whether an app is right for them, so it's important that a company approaches the reviews appropriately. This is especially important now, because coming out of the pandemic, social apps are noticing a change in users… users are starting to feel "much clearer" about what they are looking for, in their relationships[3].

By monitoring reviews, apps can tell what they are doing right and what they are doing wrong and can make changes to improve. Even more, by engaging with its reviews, the company can build trust with its users, improving their experiences. By improving the users' experiences, through engaging with reviews at least 25% of the time, the app can increase usage/revenue by an average of 35%[4].

---

[1] Alison Cashin, "The global pandemic has deeped the loneliness epidemic in America," *Harvard College*, https://mcc.gse.harvard.edu/reports/loneliness-in-america (accessed April 25, 2022).
[2] Tiyashi Datta, "Bumble going strong as love endures delta concerns," *Reuters*, https://www.reuters.com/technology/bumble-beats-quarterly-revenue-estimates-2021-08-11/ (accessed April 25, 2022).
[3] Jessica Bursztynsky "Say Goodbye to mindless swiping: Dating apps are turning to video and audio to link people up," *CNBC*, https://www.cnbc.com/2021/08/28/the-future-of-dating-apps-match-bumble-is-much-more-social.html (accessed April 25, 2022).
[4] GatherUp, "120+ Online Reviews Statistics You Need to Know in 2021," *GatherUp*, https://gatherup.com/100-online-review-statistics/ (accessed April 25, 2022).

The problem is, searching through each and every review to find out if they want/need to respond, can waste time and money for the platform. The following report will review the company Bumble, and attempt to help solve this problem for the social networking app. First, more about Bumble.

Abstract & Summary

Bumble is an application that has gained traction quickly since its inception in December of 2014. “Healthy relationships are central to living a positive, productive life.[5]” This quote is one of the first few one can find on Bumble’s about page on their website. Their mission is to help users *feel empowered* while *creating long lasting connection*s. It’s more than a dating app as it allows people to join to make friends or business connections.

People from different cities and countries come to the platform if they are looking to date, make friends, or grow their professional network. Founded in 2014 by Whitney Wolfe, the platform was developed to give women a voice in the dating world. It shifted the old-fashioned power dynamics to encourage equality on the platform; when members of the opposite sex match, women are required to make the first move. It has three subcategories: Bumble Date, Bumble BFF and Bumble Bizz. Wolfe was able to bring it to life with the help of Tinder departees Chris Gulczynski and Sarah Mick in the design process of the interface.

Learning all these great mission and vision ideas for the company, you would expect EVERYONE to love it. But just like any other platform, it can have issues that users catch. When users find anything faulty or great, the best way to leave feedback for the company is to leave a review on the respective user’s mobile app store. This helps companies improve their processes so similar mistakes can be avoided or to highlight the things they have done right. When it comes to Bumble’s reviews, the company seems to be selective in which

[5] https://bumble.com/

reviews they should respond to. Our team thought it would be interesting to create an algorithm that predicts whether or not they will.

Things that we wanted to keep in mind during this process were whether Bumble consistently responded to a certain type of review, what type of reviews they really wanted to be responding to and if their past responses reflected what they wanted to be responding to. For the sake of this project, our model assumes that Bumble is consistently responding to certain types of reviews and that their past responses do reflect what they would like to respond to in the future.

Dataset

The chosen dataset comes from Kaggle[6] and is primarily extracted from reviews on the Google Play store as well as the Bumble application itself. There are 10 variables in the dataset and 105,955 review instances written between November 2015 to April 2022.

Variable Name	**Variable Description**
reviewId	Unique identifier for users
userName	User names
userImage	URL for user's image
content	Description of the comment users made
score	Rating provided ranging from 1 to 5
thumbsUpCount	Number of thumbs up provided to a comment
reviewCreatedVersion	Review created at version numbe
at	Created at
replyContent	Replied by Bumble to the user
repliedAt	Replied at

[6] https://www.kaggle.com/datasets/shivkumarganesh/bumble-dating-app-google-play-store-review

The data included high level information about the user such as their name and profile picture. These were irrelevant to the analysis, as we did not focus on user information. The data also included a count of thumbs up per review, which we utilized in our regression. Lastly, there was timestamp information for each review and reply. We deemed the 'positive' reviews the ones that had a score of 4 or 5, and subsequently the 'negative' reviews ones with a score of 1, 2, or 3. Bumble did not respond to 42% of the reviews, or 44,397 of the reviews; of these, the average score was a 4.07 and the division of sentiment was 35,356 being positive and 9,040 being negative. On the other hand, they responded to 58% reviews, or the final 61,559; of these, the average score was 1.93 and the division of sentiment was 51,976 positive and 9,582 negative.

Models

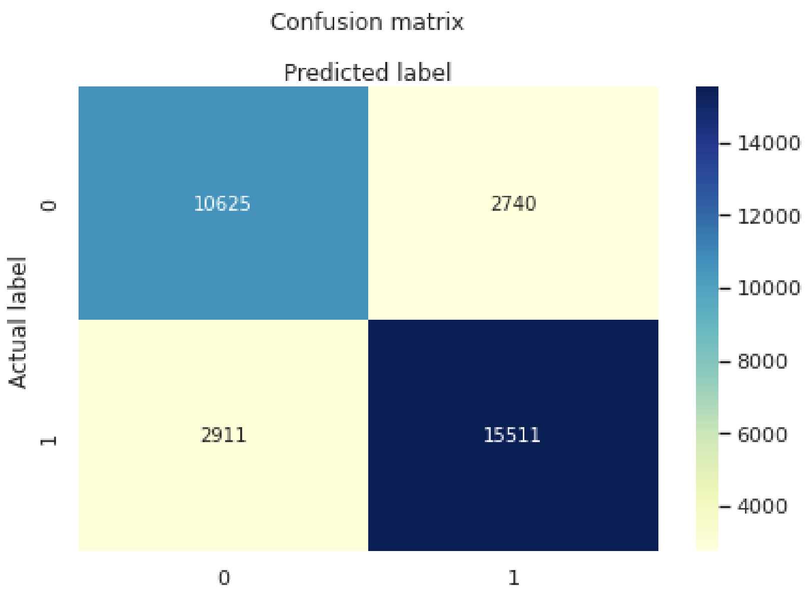
I. Regression

Multinomial logistic regression is an excellent tool for analyzing how different levels of a factor influence the probability of an outcome. Our team was excited to investigate how the sentiment of a score [in our case score - negative (1), neutral (2), positive (3)] and the number of likes a review received in the Google Play store would impact whether or not the review was responded to by application developers.

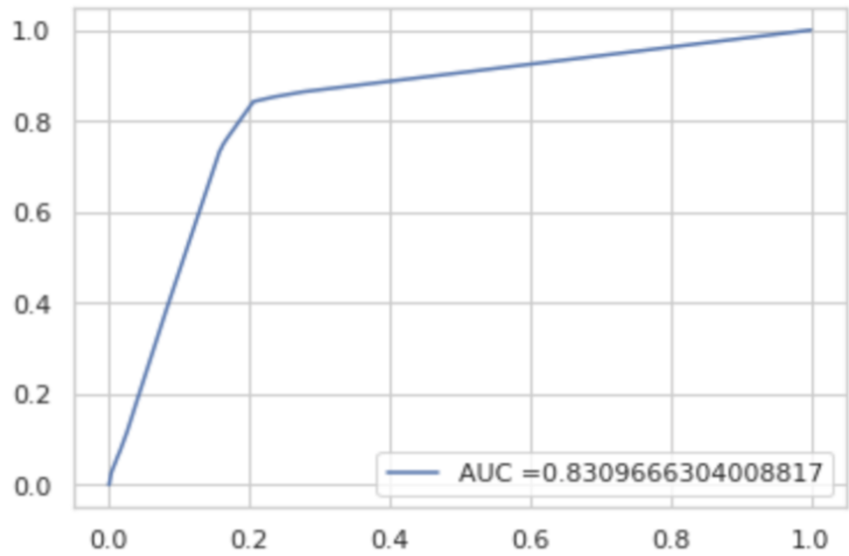
To begin, it was important to use feature engineering to convert the *repliedAt* variable into a binary feature. To do so, any observation that included a non-'NaN' value for the column was labeled a '1', indicating a developer response. Otherwise, it was labeled '0'. This feature engineering allows for the logistic regression model to produce a value between 0 and 1, which can be interpreted as a probability that event A (developer response) will occur.

In addition, the *score* feature needed to be transformed in order to be more useful. Users have an option to rate the application on a scale of 1-5. For the sake of interpretability, the team felt it important to separate numeric scores by sentiment. To accomplish this, scores of '1' or '2' were deemed to be 'negative' and were thus given a score of '1'. Scores of '3' indicate neutral sentiment, and are represented by a '2'. Finally, positive sentiment was considered to be the top two scores, '4' and '5', which were designated in the final model to be a '3'.

Ultimately, the model resulted in an accuracy score of 83%. This value can be interpreted as the proportion of correctly-identified data observations out of all observations. Furthermore, the model was approximately 85% precise. Precision is representative of the quality of the model, and indicates that out of all observations predicted to be responded to, 85% of them actually were. On the other hand, recall is a measure of the quantity of the data. It can be interpreted as the ability of the model to find all relevant instances in the data given. Our model produced a recall score of 84%. These values are reflected in the confusion matrix below.



To further evaluate the performance of the model we examined the Receiver Operating Characteristics curve, or ROC. The AUC score determines the separability of the data. A perfect model will have an AUC of 1, while a poor model will be below 0.5. As shown below, the AUC score of this model was 0.83, which shows good separation between classes.



Ultimately, a multinomial logistic regression model displayed acceptable results, however our team wanted to investigate a more complex issue which will be addressed in the following section: Which topics does Bumble determine to be the most critical to answer?

II. Topic Modeling

In order to further our understanding of what might cause Bumble to respond to a review we decided to perform Topic Modeling analysis to see if there is certain content that they prioritize. Latent Dirlchlet Allocation - LDA for short - is one of the techniques used for Topic Modelling. The process is quite simple, assigning it *K* topics, LDA loops through each document (in our case review) and randomly assigns each word into one of the topics. After this, it calculates a series of probabilities to determine a word likeness to each of the topics and their words, reassigning words to topics it corresponds to the most. This is repeated many times until the model is complete and you are left with *K* topics with words assigned to each topic.[7]

To begin, we feature engineered a couple additional columns to assist in our analysis. We created a new column named *sentiment* where all reviews with scores of 1, 2, or 3 were deemed *negative* and review scores of 4 or 5 were deemed *positive*. Then we utilized the *replyContent* column to split the dataset into two subsets: reviews that had been replied to by the app developer, and reviews that did not have replies. This was done to analyze potential differences in topics based on whether the developers acknowledge in the review or not by replying to it.

After implementing a topic modeling pipeline in PySpark, we originally decided to create 7 topics with 12 descriptive words per topic. We noticed at first that many topics had a strong overlap in descriptive words such as 'app', 'even', 'like', 'bumble', all of which are neutral and do not give us much context. Further optimization was completed to further refine the topic modeling pipeline and the final model excludes some of the common words that don't provide much information. We decided to finalize our model to include 10 topics with 10 descriptive words, and then chose three of the most unique topics from each subset to focus on for our final analysis.

Reviews with Responses

Pay-to-Match	**Poor Experience**	**Account/Support Issues**
match premium use pay facebook new first matches get limit	bad get people matches messages notifications message pay great fake	first account people one doesn match support message get time

[7] Ipshita. "Topic Modelling Using LDA." Medium. Analytics Vidhya, August 4, 2021. https://medium.com/analytics-vidhya/topic-modelling-using-lda-aa11ec9bec13.

For the reviews with responses, we identified three key themes between the ten topics: pay-to-match, account and support issues, and overall poor experience. We can see in pay to match that people talk about the premium subscription service, paying for the app, and the swipe limit if you are using a free account. For poor experience, users may have encountered fake accounts, did not get many matches, or had problems with notifications. Lastly, for account/support issues, users may have felt that Bumble support did not provide them with help for their accounts. Since Bumble replied to more negative reviews, it would make sense for them to look at the more negative reviews to identify ways they can improve their app to satisfy their users.

Reviews without Responses

Great Experience	**Better than Tinder**	**Easy to Use**
cool ok better tinder pretty people good excellent really decent	good nice far better works tinder people way concept much	people good dating great women use first get matches easy

For reviews without responses, three themes that seemed to be constant were an overall great experience, liking the app more than Tinder, and the ease of use. Great experience contains words of positive sentiment and hints at having attractive people on the app. The second topic suggests users like the concept of Bumble, where women message first, and that it works far better than the competing dating application Tinder. Lastly, users seem to find that Bumble is easy to use and is a good platform for women to use and is a great platform for dating. As mentioned above, since the distribution suggests Bumble focuses more on negative reviews, it would make sense for the themes of the reviews without replies to be more positive as the users are satisfied with their experience on the app.

Takeaways & Suggestions

At the moment, Bumble seems to be responding to the right reviews based on our topics and it doesn't seem like they are missing any major complaints. The topics are consistent with typical complaints but further research is needed to make a conclusion. Our Suggestion for Bumble is for them to continue to respond to messages that demonstrate unfavorable sentiment and have a low review score. By doing this, they can improve their business operations and ensure a favorable customer/user experience which will lead to more positive reviews in the future which ultimately improves the company's bottom line.

Further Research

After experimenting with both models, our team concluded a few things. One of the ways we could go about improving our topic modeling model is to consider n-gram analysis to get a better idea of groupings of words which results in a better assessment of user sentiment.We could look at reviews on other app services like the iOS app store or the Microsoft Store to get the responses of a bigger and more inclusive user population. Since we have information on timestamps for reviews, we could also conduct different kinds of analyses like a time series analysis.

In the end, just like any other business, Bumble wants to bring their customers the most positive experience possible which will be reflected in their responses in the customer reviews section. By using our model, Bumble will be able to save time and money by minimizing the amount of time spent on reading the reviews which therefore gives them more time to respond to what is truly important. Although these are just the first steps in this process, this is an important step for Bumble to be the best social networking app for all users alike.