Analysis on Reviews and Ratings of Dating Apps: Tinder, Bumble, and Hinge

Zixuan Wang, Ning Yang, Huiyu Jiang, Yuting Zhu

Columbia University

APAN5205: Applied Analytics Frameworks and Methods II

Vishal Lala, Davin Kaing, Eric Stewart

April 28, 2022

1 Statement of the research problem

Finding a suitable companion is one of the important life issues for most people. The process of finding the one has changed over time. Nowadays the dating app plays a significant role in the process along with the rise of smartphones and the changing of the social environment.

The dating application market is expanding at an incredible speed. The dating app revenues have increased every year since 2015, reaching $5.61 billion in 2021. Usage has also increased, with over 323 million people worldwide using dating apps. Tinder, Bumble and Hinge are the top three players in the market, which have a 69% worldwide market share in total. Tinder is the leader in the US dating app market, but Bumble has increased its market share every year since 2017. Hinge is also positioning itself as a potential leader in the near future. (Dating App Revenue and Usage Statistics 2022)

In the industrial era 5.0 (Society 5.0), product reviews are necessary for the sustainability of a company. Product reviews are a User Generated Content (UGC) feature which describes customer satisfaction. Meanwhile, customer satisfaction is considered to be important for a decision related to product purchase. Customer decisions often depend on the opinion and brand image of a product. The product reviews can be used to measure customer satisfaction based on its aspects.

The research is designed to investigate the feedback from the users of Tinder, Bumble and Hinge in the apple store. The research studies the rating and the features of these apps that are attractive to users and complained in the reviews. The research means to help the dating app companies to understand the users' demand and favor to improve the satisfaction by developing the apps.

The research aims to analyze comments and ratings of dating apps and investigate what features of the dating apps users care the most about. Which word in the comments has the most positive influence on rating? Which word in the comments has the most negative impact on rating?

2 Data Description

The content information, app ratings and reviews of the dating apps from the App Stores, are extracted by the 'appler' package in R. By applying the 'get_apple_reviews' function, we enter the id and country code of the three apps in the Apple Store.

The datasets based on a set of Tinder, Bumble, and Hinge application reviews are scraped from the Apple store website. The data include about 1000 reviews for each app, which are posted from January the 20th, 2022 to April the 14th , 2022. For Tinder and Bumble, the data include 1000 reviews. For Hinge, there are 942 reviews in the time period.

The data we are going to use includes the following 7 variables:

id: The id of the user who posted the review
review_time: The post date and time of each review from January the 20th, 2022 to April the 14th , 2022
author: The name of users who post each review
app_version: The version of the user's applications
title: The title of each review
review: Text of review posted on apple store
rating: Each review is rated by users with a five-star scale

For the research question, ‘review’ is an independent variable. We set words as tokens, and perform sentiment analysis to find the word in the reviews that have the most positive and negative impact on ‘rating’, as well as the words that show the features they care about in the comments.

3 Research Methods

In the research process, we utilized five different techniques to analyze the review data.

1. Sentiment Analysis: For the first research question, we create a list of words (called lexicon), associated with strongly positive or negative sentiments with “bing” dictionary from customer reviews. After counting the number of positive and negative words in the reviews, it can be found that the words that have the most positive and negative impact on ‘rating’, as well as the words that show the features they care about in the comments.

2. Word Cloud: A word cloud is a cluster of words depicted in different sizes. The bigger and bolder the word appears, the more often it’s mentioned within a given text and the more important it is.

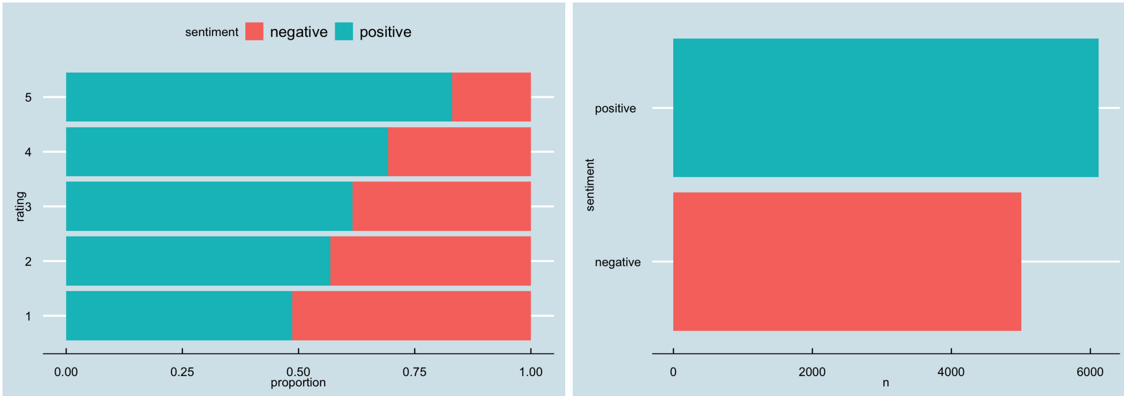
3. Feature Extraction(Bag of Words and TF-IDF): Bag of Words simply counts the frequency of words in reviews. Based on that , we use TF-IDF to determine how relevant those words are to their reviews.

4. Topic Modeling(LDA) : Beyond examining individual words, topic modeling is a useful way to discover themes about what customers care about from reviews. It is an extension of detecting customer attitudes based on sentiment analysis. To be specific, we assume the customer reviews documents are a mixture of topics, and topics are a mixture of tokens or words based on text preparation and tokenization processes we did. Latent Dirichlet Allocation(LDA) can automatically discover themes, using the probability distribution generating the words. And we can calculate and visualize the probabilities of each review being associated with each topic. Topic Modeling helps detect the features of the dating apps customers care about most and what we can improve to stay ahead of the dating app industry competition.

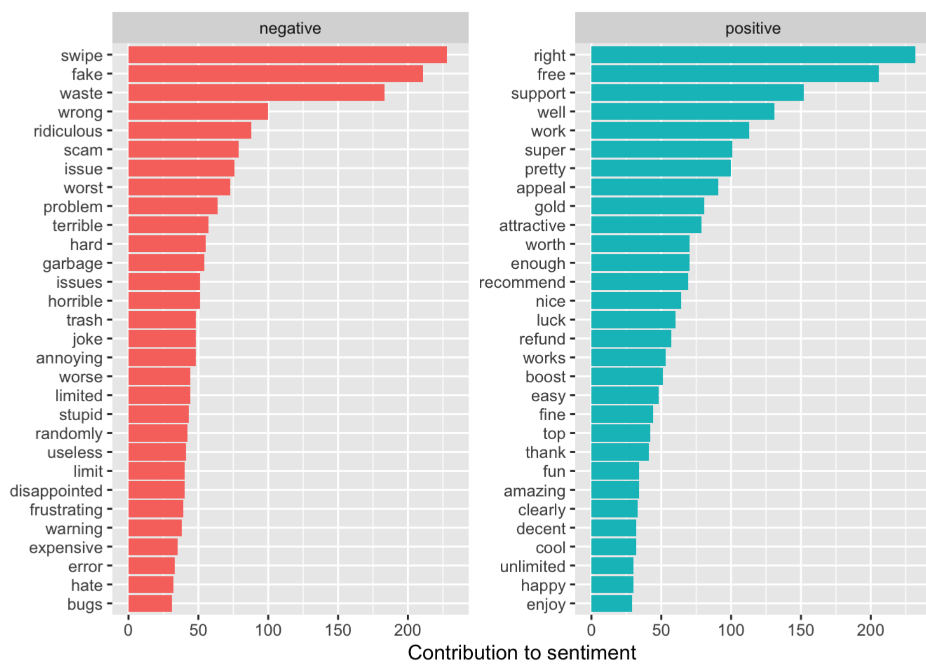
5. Latent Semantic Analysis: Latent semantic analysis (LSA) comes to the rescue because there are many features used in our sentiment analysis. LSA technique compresses the useful information into lower dimensions at the expense of some information loss noise.

4 Research Result

1. Sentiment Analysis: As expected, users who leave reviews with poor ratings are more likely to use negative words in their reviews, while users who leave positive reviews use more positive words. This supports the idea that a review’s overall sentiment can be approximated as the sum of the sentiment of each constituent word.



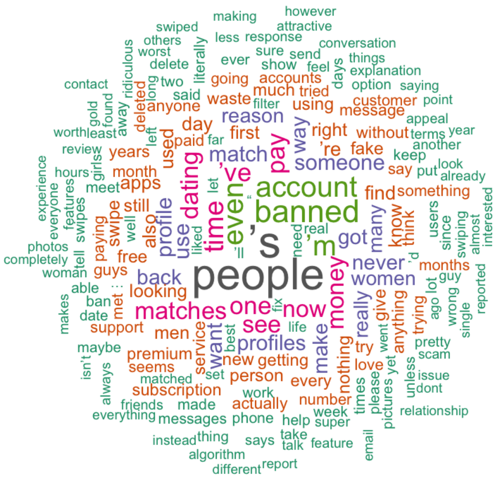
Focusing on positive reviews, we can quickly get a feeling of why users like dating apps. They love how easy it is to get a refund(customer service), the convenient and free use of some functions. As we can see, words such as “swipe”, “fake”, “scam”,“expensive”, and” error” show negative sentiments, so we can better understand which features(user profile accuracy audit, pricing, system operation) of the three dating apps should continue to be built out, and where we need to improve the most to stay with the competition of the industry.

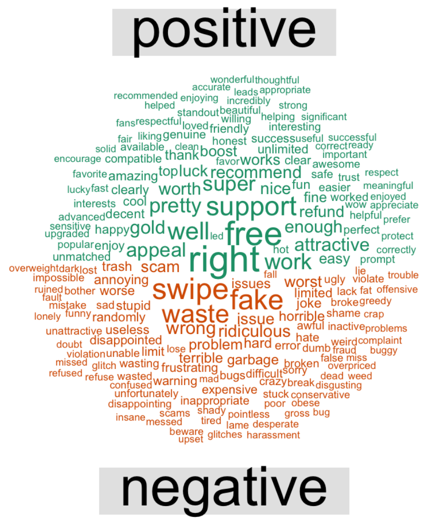


Although word frequencies give us an intuition in positive and negative reviews, we want to know what users love, what they find bad, and so on. Some words are not important in the text despite their high word frequency such as “like”, and “good” which can only show sentiments but not about the features, so we use the TF-IDF technique for the weight correction of features next.

2. Word Cloud

Word clouds are created to visualize the frequency of words in reviews.

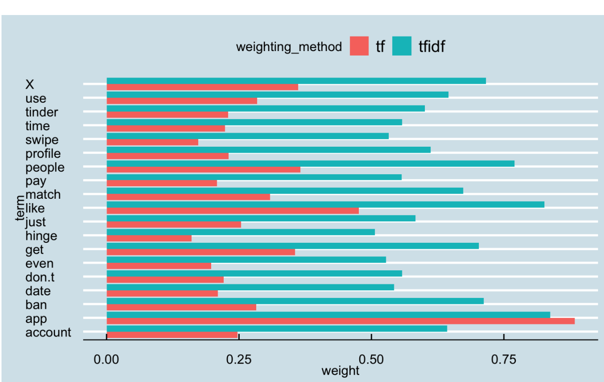




3. TF-IDF for feature extraction

As we can see, contrasting the weights of TF and IDF for the top 20 terms. For the first model, the term Frequency assigns" app" a heavyweight because "app" is the most frequently occurring term. But the word "app" appears in most of the reviews which means it has little diagnostic value. Accordingly, TF-IDF assigns "app" a much lower weight.

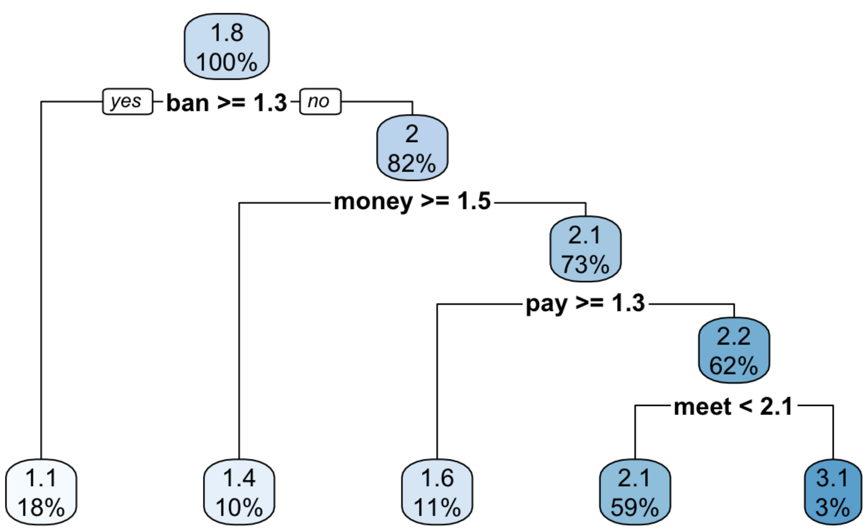
Being inspired, we found that some words are not meaningful to detect customer's attention to the specific features of the product, so we created own stopword list and remove 27 words('app','great','like','likes','good','better','bad','don.t','doesn.t','can.t','can','just','will','get','bumble','tinder','hinge','didn.t','that','that.s','won.t','you.re','i.ve','it.s','i.m','there.s','there.re'). The second TF-IDF model is created to increase the accuracy of feature extraction**.**



In the second TF-IDF model, it can be found that 16 words ("account", "ban", "meet", "show", "pay", "back", "date", "report", "women", "nothing", "now", "find", "fake", "wast", "money", "service", "anything") have p values below 0.05 in the linear regression model.

TF-IDF for prediction:
As we found in the analysis of positive and negative reviews, customer ratings act as a proxy that reflects customers' satisfaction with the service. Hence, we will look into predicting guests' behavior in the rating classification of reviews and see how the existence of certain attributes can increase the odds of a review in those apps being categorized as a high rating or low rating.

Next, take a look at a plot of the Decision Tree for prediction. The tree selected contains 4 variables with 4 splits. If we look at the plot and at the node descriptions, we will notice that splits have occurred on the variables 'ban', 'money', 'pay',and 'meet'. Nodes 2 is formed by splitting node 1, the root node, on the predictor variable 'ban'. The split point is the frequency of ban>=1.3; that is, node 2 consists of all rows with the frequency of ban< 1.3.It is predicted that the rating score will be 1.1 when the the frequency of 'ban' in reviews larger than 1.3, 18% of the total number observations(n=2059). If the the frequency of ban less than 1.3 but of 'money' larger than 1.5, the predicted rating score will be 1.4.In the end, When the the frequency of 'ban'<1.3, 'money'<1.5, 'pay'<1.3, and 'meet'>=2.1, the predicted score will be 3.1 which is higher. It is worth noting that at the last node, if people mention 'meet'<2.1, the average score will be 2.1, which accounts for 59% ,the largest proportion of all the observations.

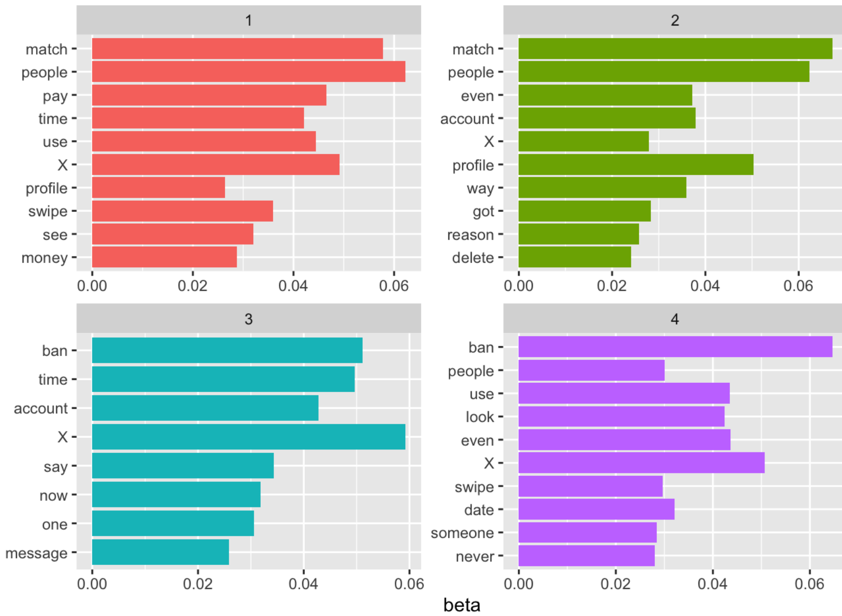


4. LDA model

First, we performed LDA analysis and categorized all words into 4 topics.

Topic 1:Payment system, swipe feature
Topic 2: User Accounts, Profiles
Topic 3: User account, message feature,Prohibited Features

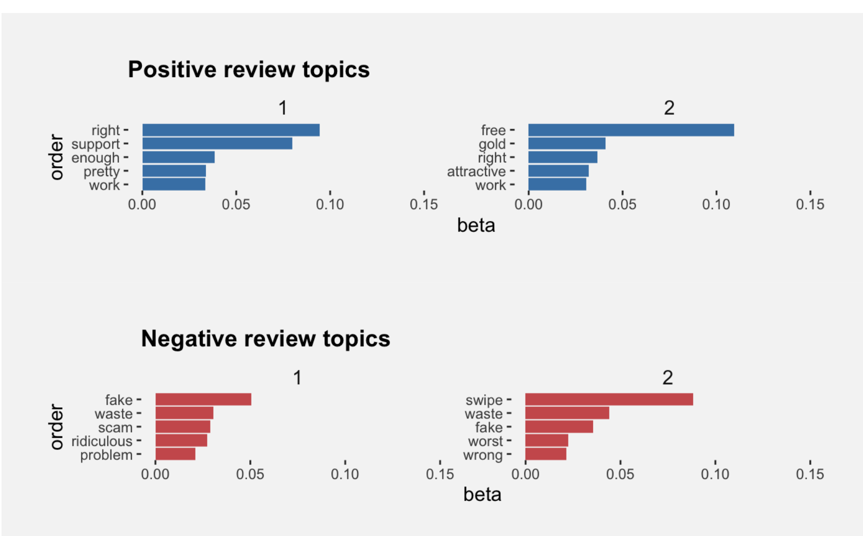
Topic 4 :Prohibited Features, swipe feature



Two-topic models of positive and negative reviews are created to find out which aspects of the customer experience come up in positive and negative reviews.

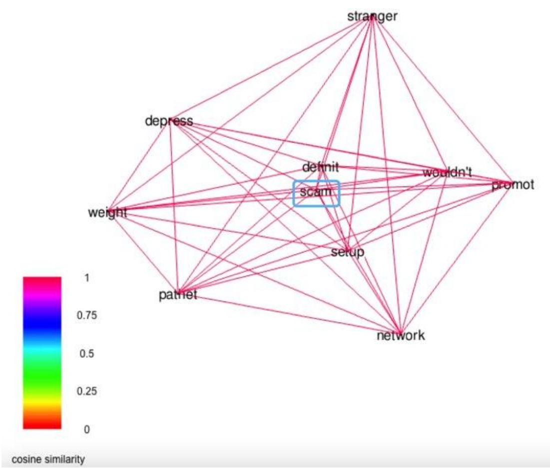
Positive Reviews:
Topic1: Customer support
Topic2: Free and premium features (Gold subscription)

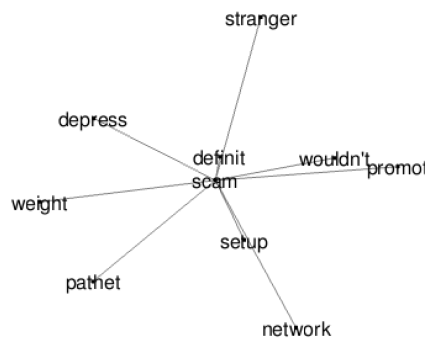
Negative Reviews:
Topic1: User Information authenticity
Topic2: Swipe design of the app



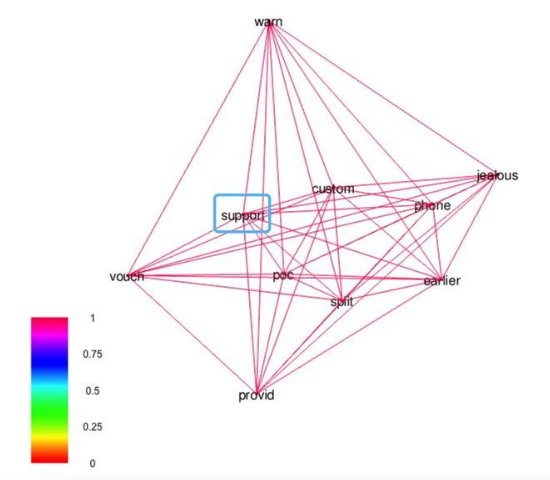
5. Latent Semantic Analysis(LSA)

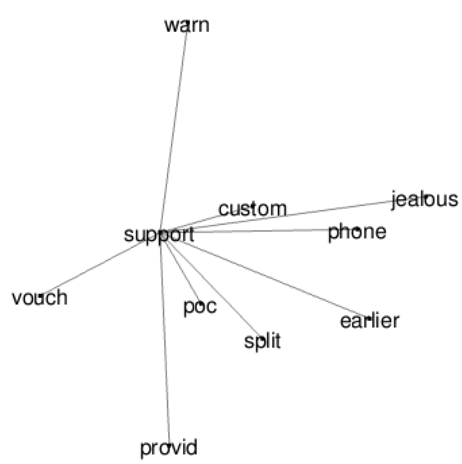
We created LSA model which assumes that words that are close in meaning will occur in similar pieces of text. According to our previous analysis of the LDA model, we find a positive topic is about customer support and service, and a negative topic is about user Information authenticity. So we use ‘LSAfun’ package to check the words’ neighbors in those two topics.





The term “scam” has relationships with “depress”, “stranger”, “definit”, etc. It means customers have customer experience of being scammed by strangers on the app, which reduced customer satisfaction.





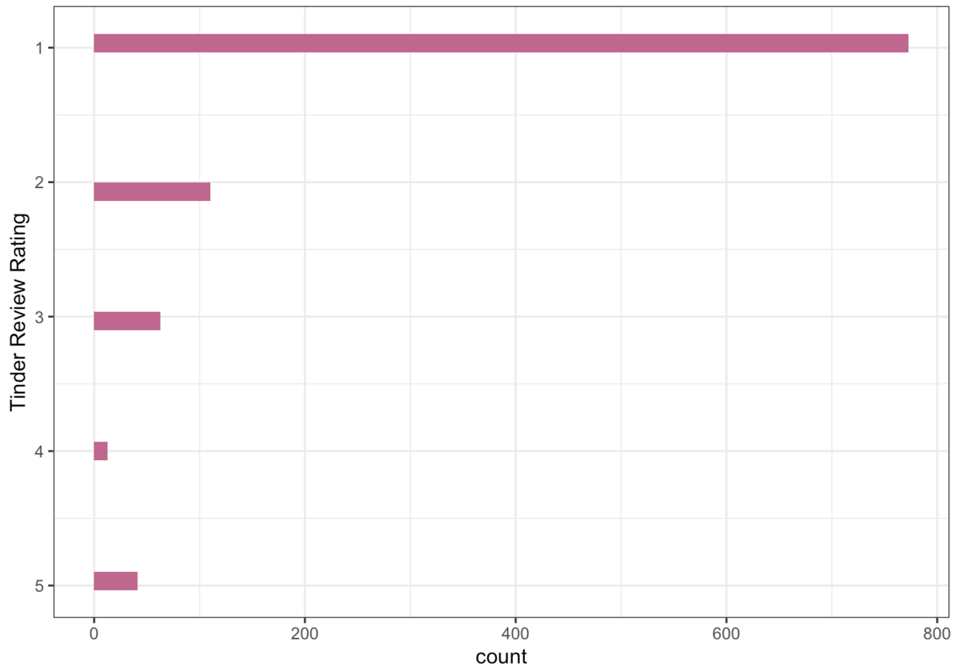
If we look at the positive topic, the term “support” has relationships with “custom”, “vouch”, “warn”, etc. Customers care about the coupons we provide, personalized customer service, and telephone contact to solve problems, which generate positive emotions and improve customer satisfaction.

5 Conclusion

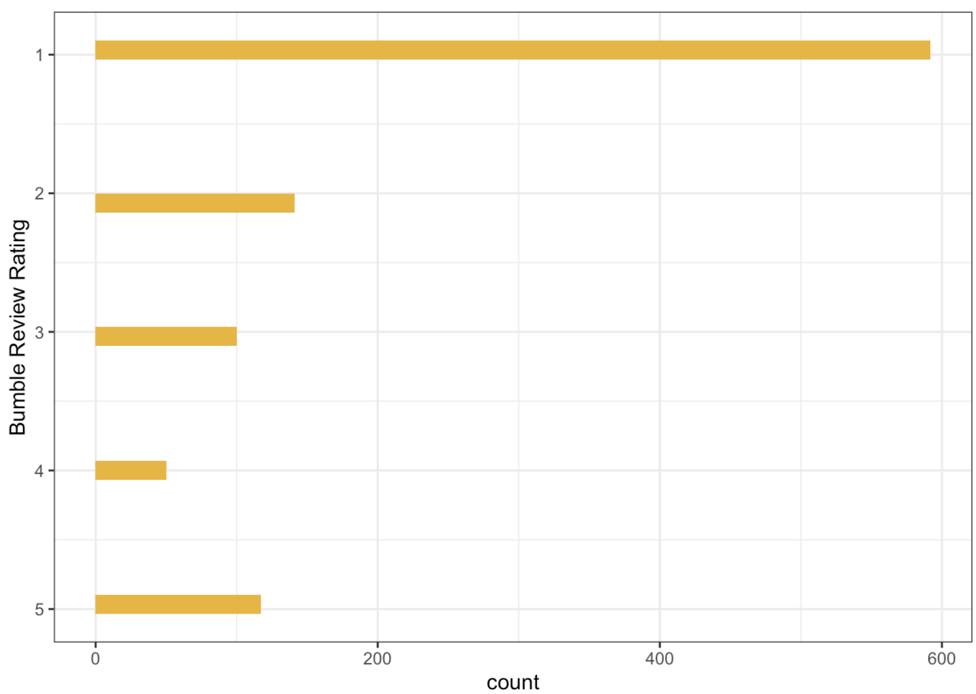
In conclusion, the positive emotions of users come from the commercial use of social software, and the functions of matching users and promoting social interaction with strangers make users satisfied. In addition, in customer service support, free download and use also have a good evaluation. The negative sentiment comes from other users' untrue information, receiving spam, and the design of the app's sliding function. Another category of dissatisfaction comes from paid-for-use features of apps, payment systems, and account-related activities.

Based on the above analysis, we recommend to decision-makers of dating app companies that they need to enhance users' identity reviews, such as the authenticity of photos and the authenticity of social identities, to prevent identity theft and fraud. In addition, the security optimization of the background needs to be improved. When sensitive words such as “transfer” and “bank card” appear in the chat between users, users should be prompted to prevent being deceived. In the operation design of matching between users, questionnaires can be sent to users, and the “swipe ”method favored by the public can be designed.

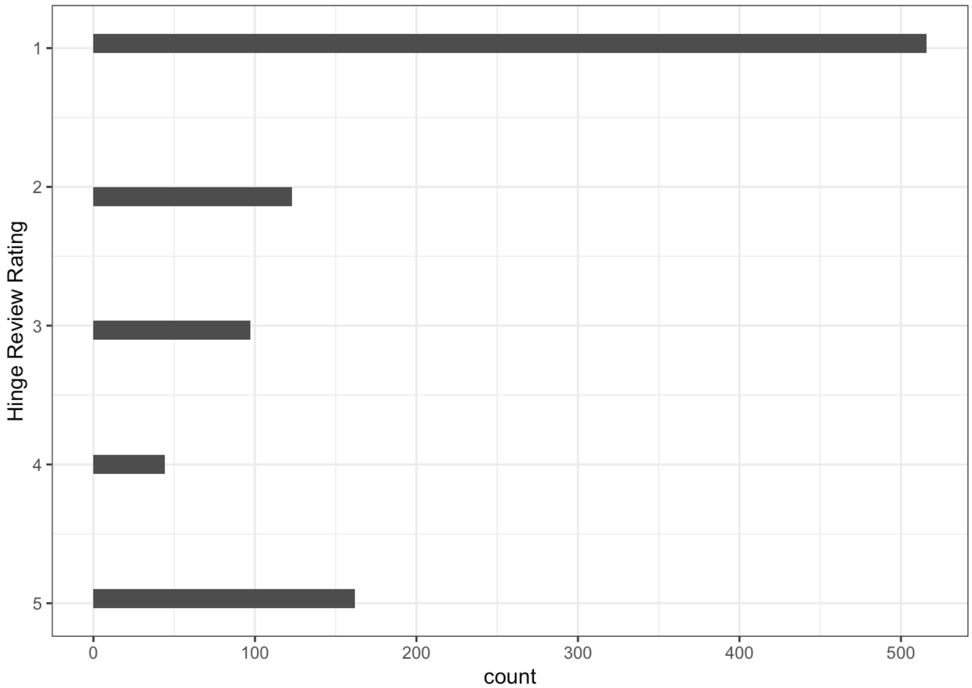
Appendix



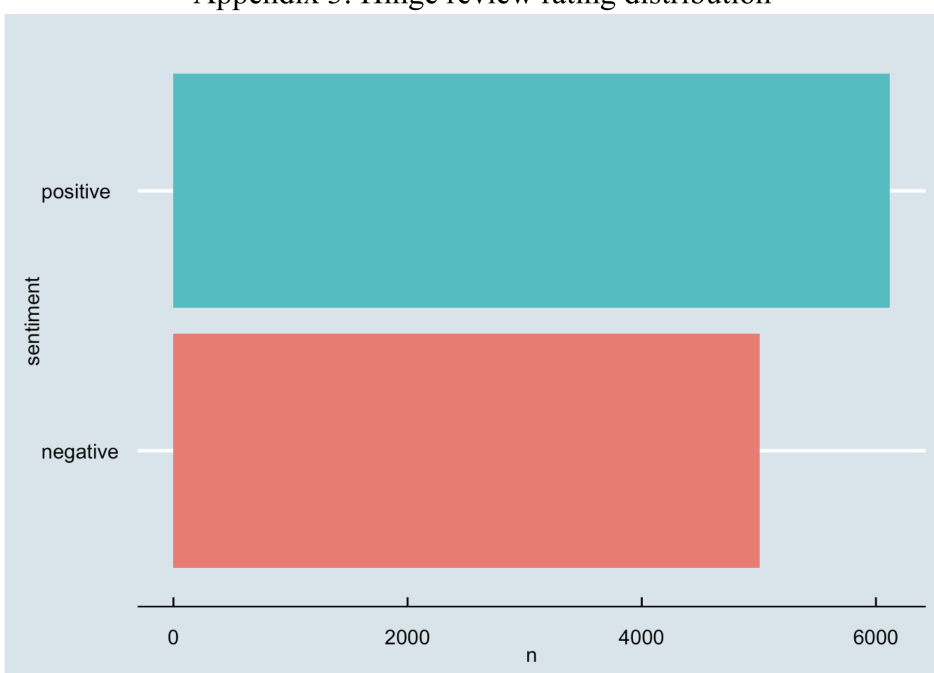
Appendix 1: Tinder review rating distribution



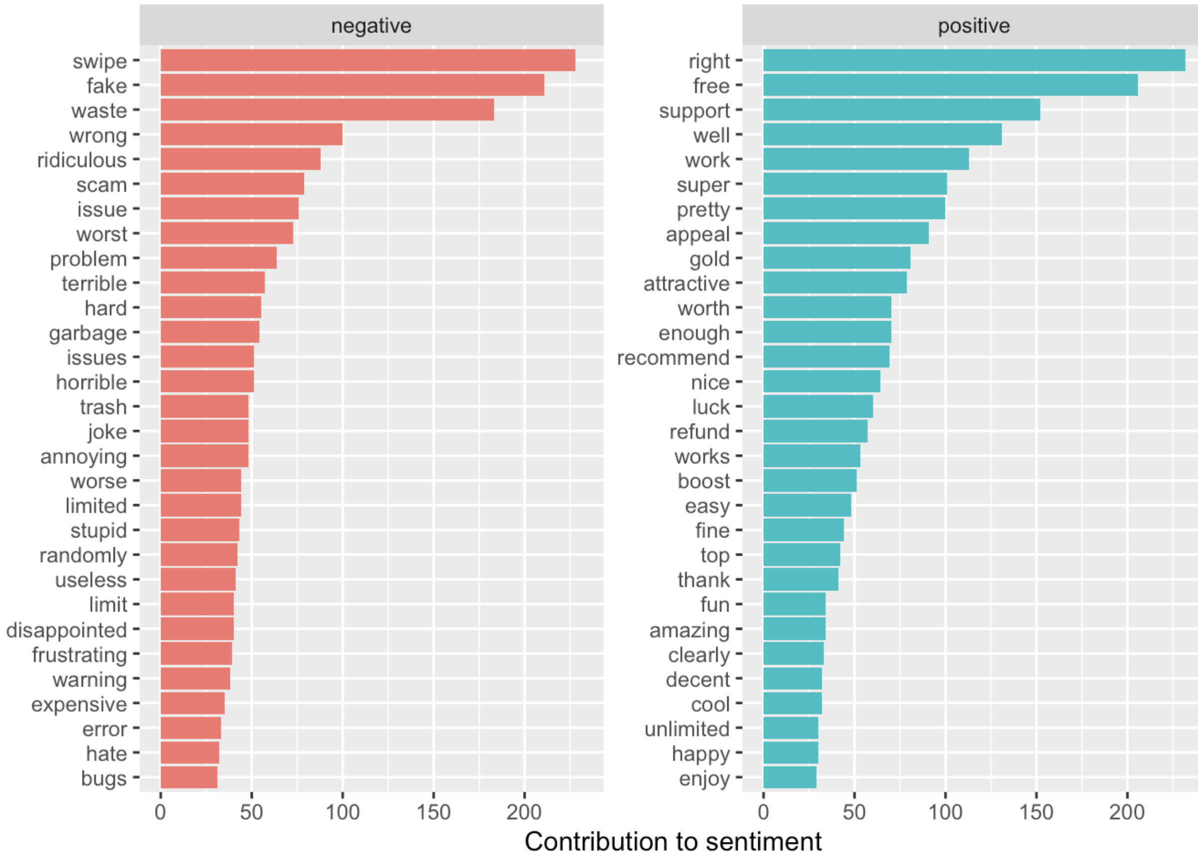
Appendix 2: Bumble review rating distribution



Appendix 3: Hinge review rating distribution



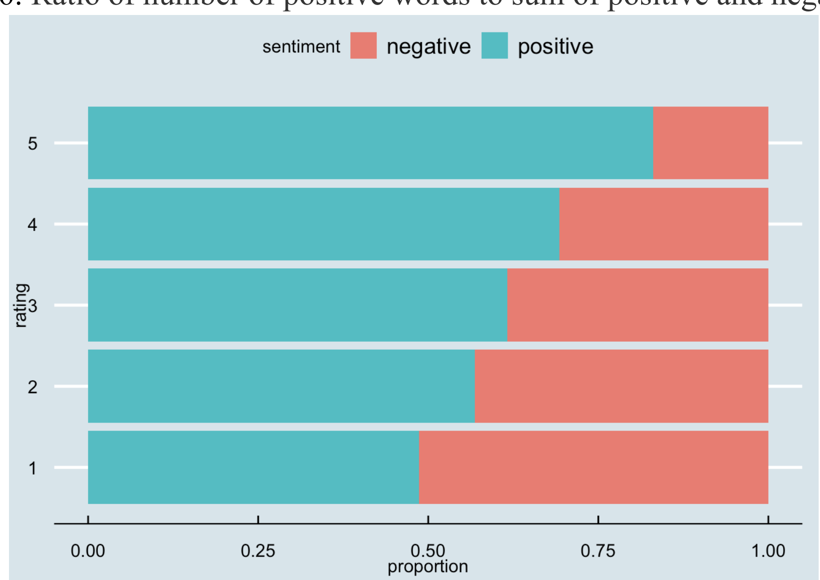
Appendix 4: Word count of negative and positive words in review



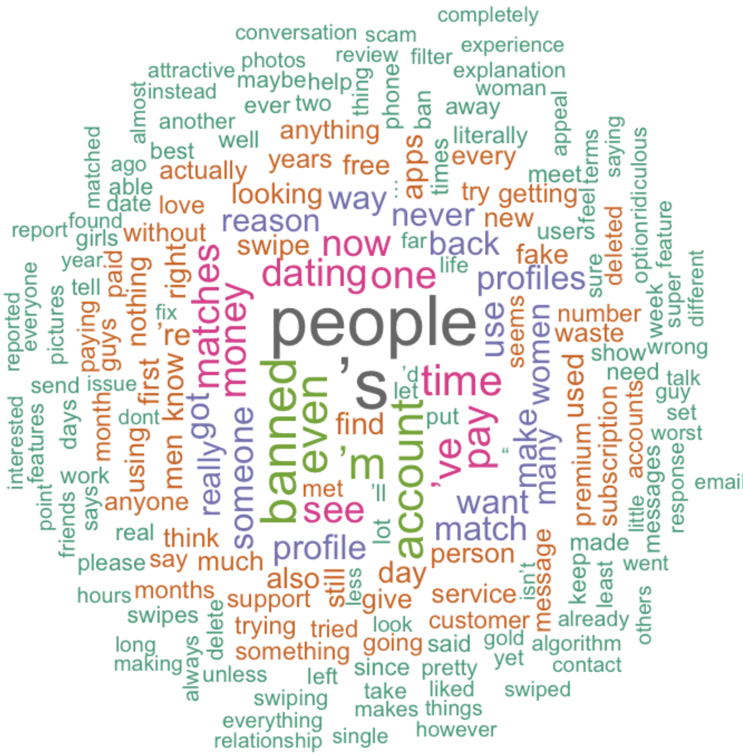
Appendix 5: Distribution of top 10 common negative words and top 10 common positive words

	sentiment	n	proportion
	<chr>	<int>	<dbl>
1	negative	5006	0.450
2	positive	6120	0.550

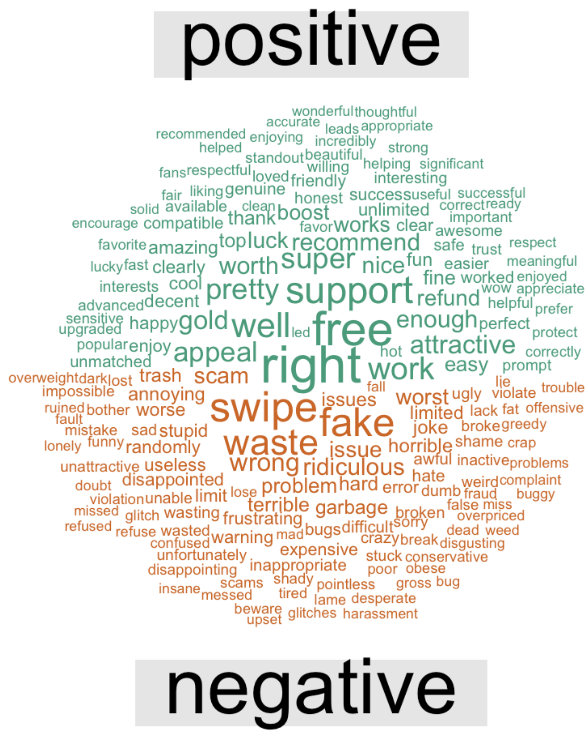
Appendix 6: Ratio of number of positive words to sum of positive and negative words



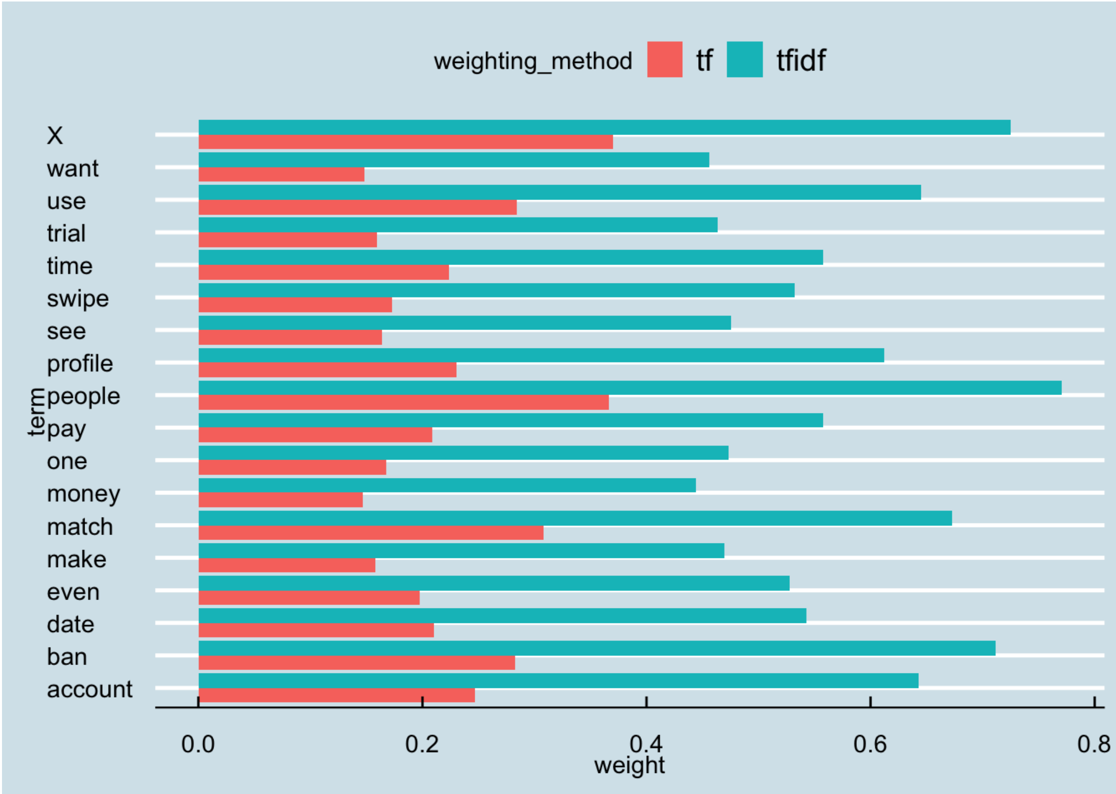
Appendix 7: Proportion of positive (or negative) reviews for each rating



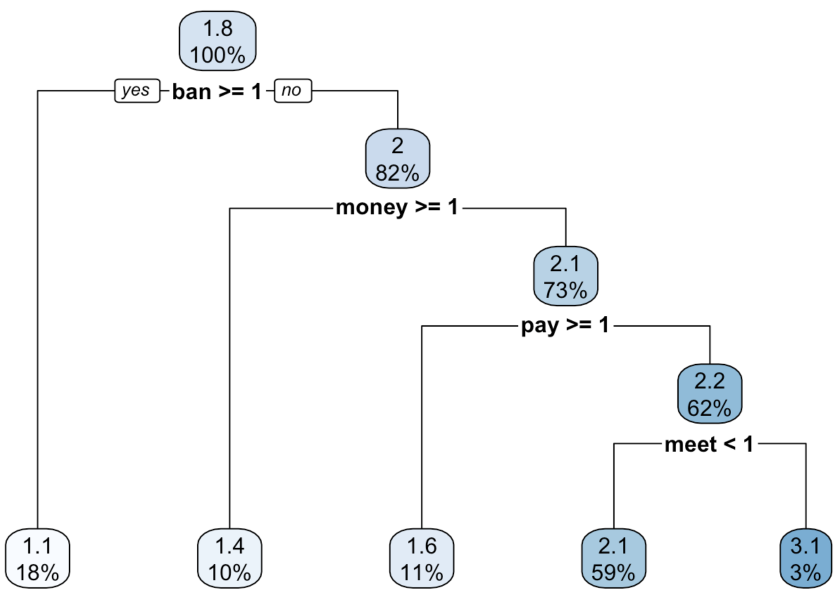
Appendix 8: Word cloud for all words in reviews



Appendix 9: Comparison word cloud for all words in reviews



Appendix 10: The analysis bar chart include both Term Frequency and Term Frequency Inverse Document Frequency



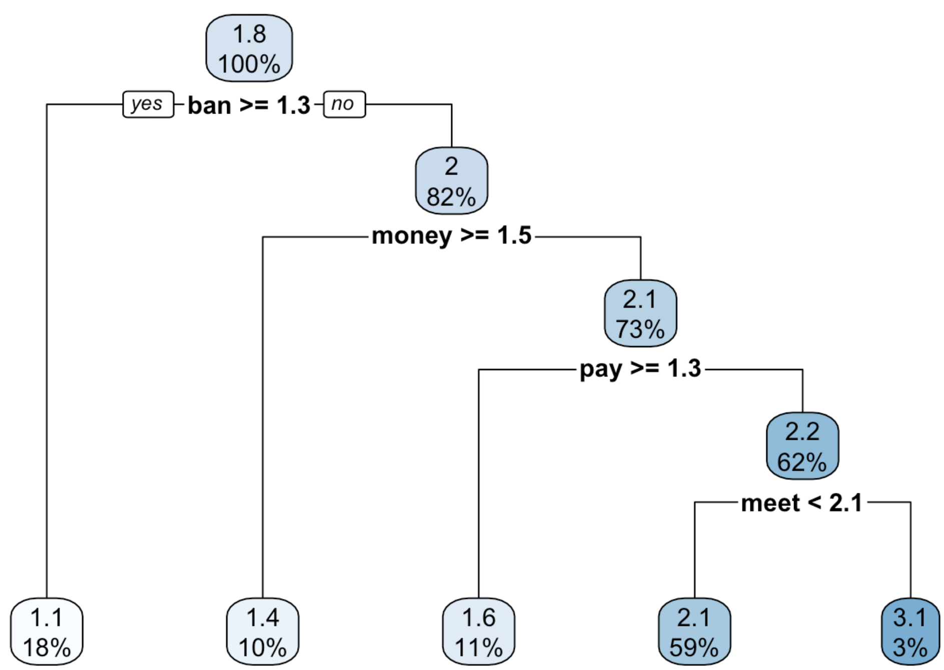
Appendix 11: The decision tree model using Term Frequency

```
Call:
lm(formula = rating ~ ., data = train)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4851 -0.9202 -0.4494  0.6044  3.9412

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.122404   0.041408  51.256  < 2e-16 ***
even         -0.148632   0.060707  -2.448  0.01444 *
say           0.014519   0.089021   0.163  0.87046
use           0.044793   0.052127   0.859  0.39028
account      -0.110795   0.055535  -1.995  0.04618 *
match        -0.016327   0.046890  -0.348  0.72772
ban          -0.235102   0.052100  -4.513 6.78e-06 ***
got          -0.103189   0.083888  -1.230  0.21881
first         0.091994   0.098267   0.936  0.34930
meet          0.470807   0.110964   4.243 2.31e-05 ***
new           0.063694   0.103314   0.617  0.53763
people       -0.014911   0.041036  -0.363  0.71637
show         -0.192811   0.083100  -2.320  0.02043 *
look          0.059181   0.071559   0.827  0.40832
make          0.025973   0.066436   0.391  0.69588
pay          -0.155134   0.054911  -2.825  0.00477 **
reason       -0.075833   0.091879  -0.825  0.40927
time         -0.048220   0.063444  -0.760  0.44732
want          0.046936   0.070561   0.665  0.50601
work          0.084354   0.099167   0.851  0.39508
back         -0.223285   0.082184  -2.717  0.00665 **
date          0.190923   0.060538   3.154  0.00164 **
day          -0.036949   0.074629  -0.495  0.62059
give         -0.033145   0.092523  -0.358  0.72020
keep         -0.154493   0.106780  -1.447  0.14810
men          -0.064101   0.073768  -0.869  0.38498
message      -0.059464   0.067038  -0.887  0.37518
never        -0.020906   0.082885  -0.252  0.80089
one           0.054649   0.068903   0.793  0.42779
report       -0.195329   0.091011  -2.146  0.03198 *
right         0.232765   0.097486   2.388  0.01705 *
thing         0.101324   0.118773   0.853  0.39371
## women        -0.189560   0.078645  -2.410  0.01603 *
## know          0.069881   0.090754   0.770  0.44139
## anyone       -0.231277   0.120543  -1.919  0.05518 .
## delete       -0.022123   0.084348  -0.262  0.79313
## nothing      -0.217482   0.110100  -1.975  0.04837 *
## now          -0.168668   0.075046  -2.248  0.02471 *
## swipe        -0.086548   0.058631  -1.476  0.14006
## trial        -0.129968   0.068374  -1.901  0.05747 .
## week          0.156964   0.110362   1.422  0.15511
## find          0.189665   0.077710   2.441  0.01475 *
## without       0.177008   0.118540   1.493  0.13553
## actual        0.019158   0.109659   0.175  0.86133
## fake         -0.280867   0.092549  -3.035  0.00244 **
## person        0.099762   0.083949   1.188  0.23483
## profile      -0.014791   0.050490  -0.293  0.76960
## seem         -0.085867   0.106724  -0.805  0.42116
## also          0.211921   0.096132   2.204  0.02760 *
## app           0.006107   0.091367   0.067  0.94671
## way           0.024126   0.085253   0.283  0.77721
## year          0.149096   0.092168   1.618  0.10589
## free          0.021740   0.095262   0.228  0.81951
## see          -0.055572   0.070356  -0.790  0.42970
## wast         -0.352592   0.115419  -3.055  0.00228 **
## manipulate    0.056310   0.084046   0.670  0.50294
## money        -0.147319   0.072417  -2.034  0.04205 *
## think         0.109288   0.105556   1.035  0.30063
## premium      -0.178658   0.108099  -1.653  0.09855 .
## much          0.188506   0.100175   1.882  0.06001 .
## service      -0.242025   0.101376  -2.387  0.01706 *
## get          -0.069087   0.096456  -0.716  0.47392
## subscription -0.111725   0.107848  -1.036  0.30035
## month        -0.098368   0.085025  -1.157  0.24744
## need          0.127903   0.104649   1.222  0.22177
## anything     -0.356659   0.118494  -3.010  0.00265 **
## guy           0.088981   0.076877   1.157  0.24723
## still        -0.008085   0.105280  -0.077  0.93879
## someone       0.036452   0.074406   0.490  0.62426
## take         -0.222532   0.114685  -1.940  0.05247 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.
##
## Residual standard error: 1.274 on 1989 degrees of freed
## Multiple R-squared:  0.1547, Adjusted R-squared:  0.125
## F-statistic: 5.276 on 69 and 1989 DF,  p-value: < 2.2e-
```

Appendix 12: The linear regression model using Term Frequency



Appendix 13: The decision tree model using Term Frequency Inverse Document Frequency

```
Call:
lm(formula = rating ~ ., data = train)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.4851 -0.9202 -0.4494  0.6044  3.9412 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.122404   0.041408  51.256  < 2e-16 ***
even        -0.055715   0.022756  -2.448  0.01444 *  
say          0.004279   0.026234   0.163  0.87046    
use          0.019754   0.022988   0.859  0.39028    
account     -0.042553   0.021329  -1.995  0.04618 *  
match       -0.007480   0.021482  -0.348  0.72772    
ban         -0.093303   0.020677  -4.513 6.78e-06 ***
got         -0.029854   0.024270  -1.230  0.21881    
first        0.023274   0.024861   0.936  0.34930    
meet         0.111597   0.026302   4.243 2.31e-05 ***
new          0.015389   0.024962   0.617  0.53763    
people      -0.007085   0.019497  -0.363  0.71637    
show        -0.048403   0.020861  -2.320  0.02043 *  
look         0.017255   0.020864   0.827  0.40832    
make         0.008740   0.022355   0.391  0.69588    
pay         -0.058085   0.020560  -2.825  0.00477 ** 
reason      -0.022585   0.027364  -0.825  0.40927    
time        -0.019329   0.025432  -0.760  0.44732    
want         0.015282   0.022974   0.665  0.50601    
work         0.020928   0.024603   0.851  0.39508    
back        -0.064196   0.023629  -2.717  0.00665 ** 
date         0.073915   0.023437   3.154  0.00164 ** 
day         -0.011577   0.023383  -0.495  0.62059    
give        -0.009258   0.025844  -0.358  0.72020    
keep        -0.036699   0.025365  -1.447  0.14810    
men         -0.015161   0.017448  -0.869  0.38498    
message     -0.016005   0.018044  -0.887  0.37518    
never       -0.006048   0.023979  -0.252  0.80089    
one          0.019364   0.024415   0.793  0.42779    
report      -0.045389   0.021149  -2.146  0.03198 *  
right        0.058774   0.024616   2.388  0.01705 *  
thing        0.024327   0.028517   0.853  0.39371    
women       -0.049685   0.020614  -2.410  0.01603 *  
know         0.019069   0.024765   0.770  0.44139    
anyone      -0.054344   0.028325  -1.919  0.05518 .  
delete      -0.006108   0.023289  -0.262  0.79313    
nothing     -0.055337   0.028014  -1.975  0.04837 *  
now         -0.054769   0.024368  -2.248  0.02471 *  
swipe       -0.028065   0.019013  -1.476  0.14006    
trial       -0.044695   0.023513  -1.901  0.05747 .  
week         0.037367   0.026273   1.422  0.15511    
find         0.053670   0.021990   2.441  0.01475 *  
without      0.041684   0.027915   1.493  0.13553    
actual       0.004744   0.027152   0.175  0.86133    
fake        -0.068566   0.022593  -3.035  0.00244 ** 
person       0.026572   0.022360   1.188  0.23483    
profile     -0.005570   0.019015  -0.293  0.76960    
seem        -0.021724   0.027000  -0.805  0.42116    
also         0.056347   0.025560   2.204  0.02760 *  
app          0.001598   0.023905   0.067  0.94671    
way          0.007034   0.024856   0.283  0.77721    
year         0.039502   0.024419   1.618  0.10589    
free         0.005253   0.023016   0.228  0.81951    
see         -0.019135   0.024226  -0.790  0.42970    
wast        -0.093749   0.030688  -3.055  0.00228 ** 
manipulate   0.015444   0.023051   0.670  0.50294    
money       -0.048737   0.023957  -2.034  0.04205 *  
think        0.027275   0.026344   1.035  0.30063    
premium     -0.042072   0.025456  -1.653  0.09855 .  
much         0.045830   0.024354   1.882  0.06001 .  
service     -0.060045   0.025151  -2.387  0.01706 *  
get         -0.017712   0.024729  -0.716  0.47392    
subscription -0.026597   0.025674  -1.036  0.30035    
month       -0.028192   0.024368  -1.157  0.24744    
need         0.031606   0.025859   1.222  0.22177    
anything    -0.085451   0.028389  -3.010  0.00265 ** 
guy          0.022338   0.019299   1.157  0.24723    
still       -0.002069   0.026941  -0.077  0.93879    
someone      0.010775   0.021994   0.490  0.62426    
take        -0.052862   0.027243  -1.940  0.05247 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.274 on 1989 degrees of freedom
Multiple R-squared:  0.1547,	Adjusted R-squared:  0.1254 
F-statistic: 5.276 on 69 and 1989 DF,  p-value: < 2.2e-16
```

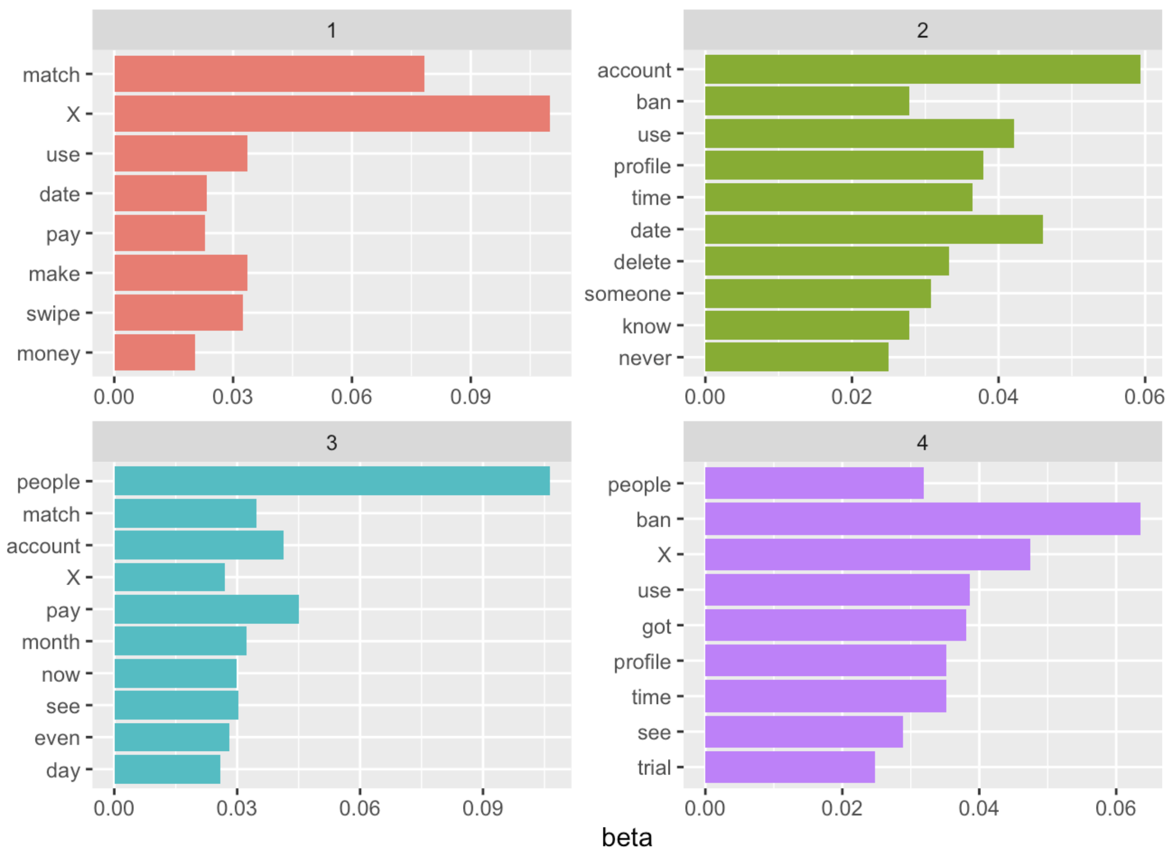
Appendix 14: The linear regression model using Term Frequency Inverse Document Frequency

	Topic 1	Topic 2	Topic 3	Topic 4
[1,]	"X"	"account"	"people"	"ban"
[2,]	"match"	"date"	"pay"	"X"
[3,]	"use"	"use"	"account"	"use"
[4,]	"make"	"profile"	"match"	"got"
[5,]	"swipe"	"time"	"month"	"profile"
[6,]	"X"	"delete"	"see"	"time"
[7,]	"X"	"someone"	"now"	"people"
[8,]	"date"	"know"	"even"	"see"
[9,]	"pay"	"ban"	"X"	"X"
[10,]	"money"	"never"	"day"	"trial"
[11,]	"now"	"one"	"money"	"make"
[12,]	"ban"	"trial"	"time"	"reason"
[13,]	"look"	"even"	"date"	"want"
[14,]	"one"	"want"	"X"	"X"
[15,]	"say"	"pay"	"one"	"even"
[16,]	"X"	"meet"	"profile"	"person"
[17,]	"think"	"first"	"give"	"much"
[18,]	"wast"	"look"	"look"	"say"
[19,]	"free"	"way"	"right"	"also"
[20,]	"someone"	"app"	"back"	"match"

Appendix 15: The topic model with 4 topics

	topic1	topic2	topic3	topic4
even	0.012789296	0.022758559	0.028177440	0.0239611898
say	0.015861972	0.007534588	0.006418359	0.0191658826
use	0.033682582	0.042075210	0.011785127	0.0385585506
X	0.031346127	0.004675532	0.001674353	0.0474928115
X.1	0.025307842	0.005712610	0.026923101	0.0127114746
account	0.001625226	0.059396268	0.041290787	0.0069948383
match	0.078180209	0.005676236	0.034664747	0.0181016215
ban	0.019682785	0.027854641	0.014149628	0.0635822364
got	0.001496806	0.006764712	0.002253930	0.0380590662
X.2	0.109898424	0.014361828	0.013355257	0.0265716143
first	0.002451625	0.018275998	0.010332001	0.0029684764
meet	0.003542549	0.018979328	0.004316045	0.0004058295
new	0.008231176	0.001574999	0.012828880	0.0068972757
people	0.013032282	0.011022573	0.106259988	0.0319562431
show	0.013090250	0.015197991	0.005433041	0.0013576190
look	0.016730567	0.017604074	0.016518335	0.0019969850
make	0.033551488	0.001207261	0.010837365	0.0244838468
pay	0.022863754	0.020525685	0.044970674	0.0041025816
X.3	0.001949756	0.008260241	0.008447797	0.0240107016
reason	0.014457974	0.002282837	0.007850159	0.0244065514

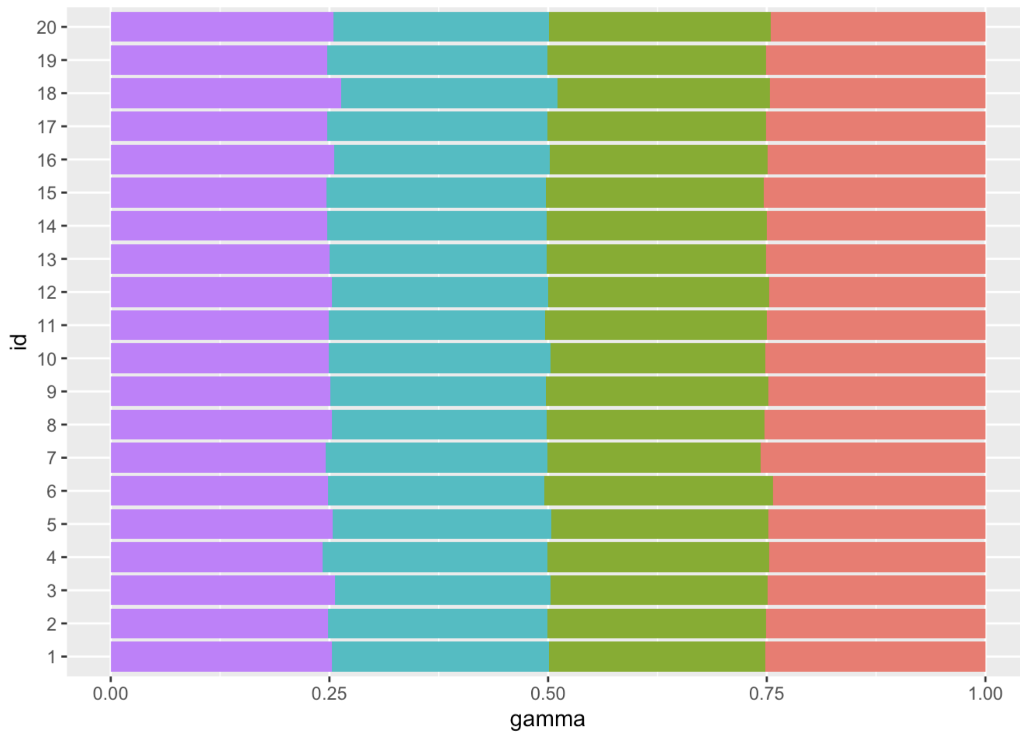
Appendix 16: Beta table of topic model with 4 topics



Appendix 17: Bar chart of words that differ across topics

	id	topic1	topic2	topic3	topic4
[1,]	1	0.2513469	0.2476550	0.2480615	0.2529366
[2,]	2	0.2505562	0.2505395	0.2508407	0.2480636
[3,]	3	0.2494009	0.2479262	0.2463421	0.2563309
[4,]	4	0.2470998	0.2539736	0.2565241	0.2424025
[5,]	5	0.2480772	0.2481591	0.2503509	0.2534128
[6,]	6	0.2427369	0.2612647	0.2476247	0.2483737
[7,]	7	0.2569442	0.2437719	0.2536284	0.2456555
[8,]	8	0.2521637	0.2494767	0.2456626	0.2526970
[9,]	9	0.2478237	0.2546206	0.2463482	0.2512075
[10,]	10	0.2516766	0.2452606	0.2542163	0.2488464

Appendix 18: Document-topic probabilities table



Appendix 18: Bar chart of document-topic probabilities for first 20 ids

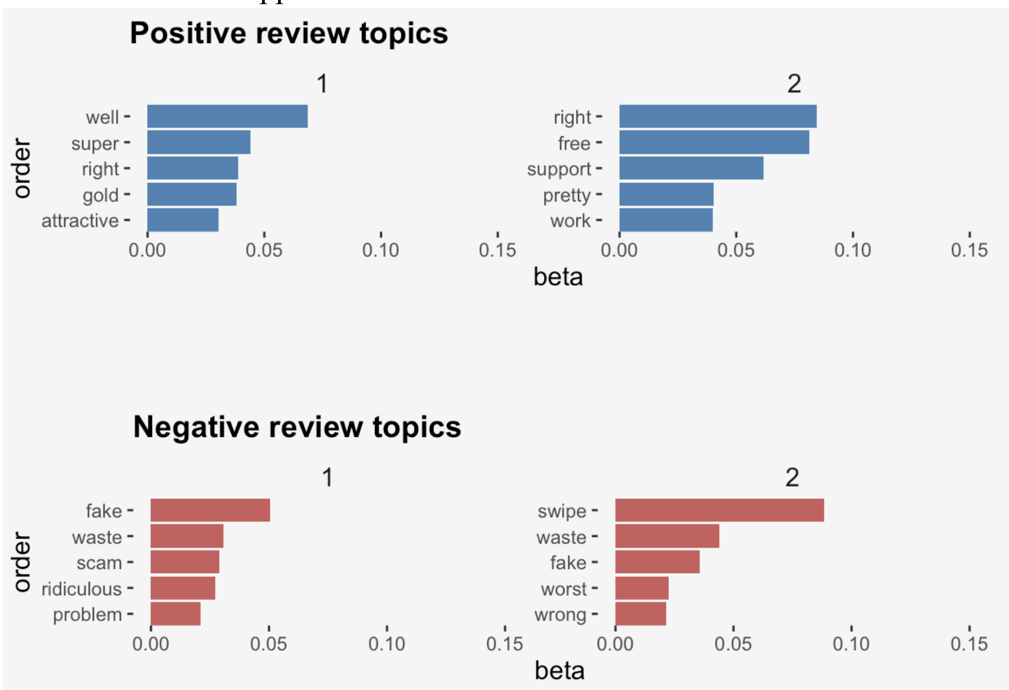
	id	topic1	topic2	topic3	topic4	rating
1	1	0.2513469	0.2476550	0.2480615	0.2529366	1
2	2	0.2505562	0.2505395	0.2508407	0.2480636	1
3	3	0.2494009	0.2479262	0.2463421	0.2563309	1
4	4	0.2470998	0.2539736	0.2565241	0.2424025	2
5	5	0.2480772	0.2481591	0.2503509	0.2534128	1
6	6	0.2427369	0.2612647	0.2476247	0.2483737	1

Appendix 19: Topic model based on on positive and negative(LDA)

topics(reviewLDA_pos)	n
1	767
2	898

topics(reviewLDA_neg)	n
1	1035
2	943

Appendix 20: Table of word count of LDA



Appendix 21: Topic model based on positive and negative(LDA)(shown top 5 words)

```
          dim1         dim2         dim3          dim4         dim5
1 -0.008113516 -0.007798542  0.006997838 -0.0004476023 -0.020610432
2 -0.004701204 -0.006704718  0.007705944 -0.0108580541  0.014272259
3 -0.009974874 -0.025788574 -0.009581060  0.0237722303  0.008510648
4 -0.020090662  0.002696763 -0.022463456 -0.0296277005 -0.001331326
5 -0.023008722 -0.024647354 -0.027709410  0.0149512983  0.005137237
6 -0.042014163 -0.052571582  0.029199983  0.0097422497  0.009618742
          dim6         dim7         dim8         dim9         dim10
1 -0.006686764 -0.012542926 -0.007702171  0.002852350  0.0034542761
2  0.014544818  0.017694182 -0.011484063  0.013764410 -0.0034239058
3  0.017624975 -0.020175614  0.030511636  0.003342545 -0.0002580005
4  0.010536434  0.008943708 -0.032026614  0.020176838 -0.0262914376
5  0.007986018  0.005230760  0.018766585 -0.014240520  0.0151792415
6  0.012192556  0.008187640 -0.022016506 -0.058916509 -0.0203541978
          dim11         dim12         dim13        dim14        dim15
1  0.0139719264 -0.0319026438 -0.0206337904  0.016585634 -0.003635301
2 -0.0010784019  0.0003989564  0.0024808616 -0.008136130 -0.004377858
3 -0.0030898696  0.0106232809 -0.0007696976  0.009756529 -0.013829997
4  0.0071810244 -0.0114314889 -0.0056641059  0.018580636  0.016100483
5 -0.0009657949  0.0315798495  0.0001031741 -0.024259371  0.016013459
6 -0.0192266697 -0.0150538426 -0.0327208822  0.040819065  0.018565323
         dim16        dim17        dim18        dim19        dim20        dim21
1  0.022466777  0.004497579  0.017160340 -0.006063337  0.009554553 -0.011394774
2 -0.002749153  0.005534290  0.007132104  0.002219278  0.006044337 -0.003452035
3  0.020073591 -0.002175822 -0.002105569 -0.004671921  0.007817605 -0.002939194
4 -0.008536996  0.027762216  0.003351894 -0.013470831 -0.010816700  0.003985046
5 -0.010972636  0.025682426 -0.003880079 -0.020033856  0.004874953  0.003140158
6 -0.078447367  0.005984852 -0.094774578 -0.068693631  0.012750014  0.052264024
         dim22        dim23         dim24         dim25         dim26
1 -0.023953845 -0.008836416  0.0023257476  0.0019960247 -0.0102529410
2 -0.005611422 -0.003020189  0.0002894168  0.0040983924 -0.0009718279
3 -0.002787906  0.013303967 -0.0084539225 -0.0003149309  0.0097939426
4 -0.006356760 -0.009789368 -0.0079588968  0.0008247979 -0.0052336805
5 -0.016747726  0.027240439  0.0059715276  0.0331493899  0.0087811458
6  0.042178836 -0.041962937  0.1002045316 -0.0508473856 -0.0009861481
```

Appendix 22: Latent Semantic Analysis (LSA)

References

CURRY, D. A. V. I. D. (2022, February 23). *Dating app revenue and Usage Statistics (2022)*. Business of Apps. Retrieved from https://www.businessofapps.com/data/dating-app-market/

Khotimah, D., & Sarno, R. (2019). Sentiment analysis of hotel aspect using probabilistic latent semantic analysis, word embedding and LSTM. *International Journal of Intelligent Engineering and Systems*, *12*(4), 275–290. https://doi.org/10.22266/ijies2019.0831.26

D. A. K. Khotimah and R. Sarno, "Sentiment Detection of Comment Titles in Booking.com Using Probabilistic Latent Semantic Analysis," 2018 6th International Conference on Information and Communication Technology (ICoICT), 2018, pp. 514-519, doi: 10.1109/ICoICT.2018.8528784. Retrieved from https://ieeexplore.ieee.org/abstract/document/8528784