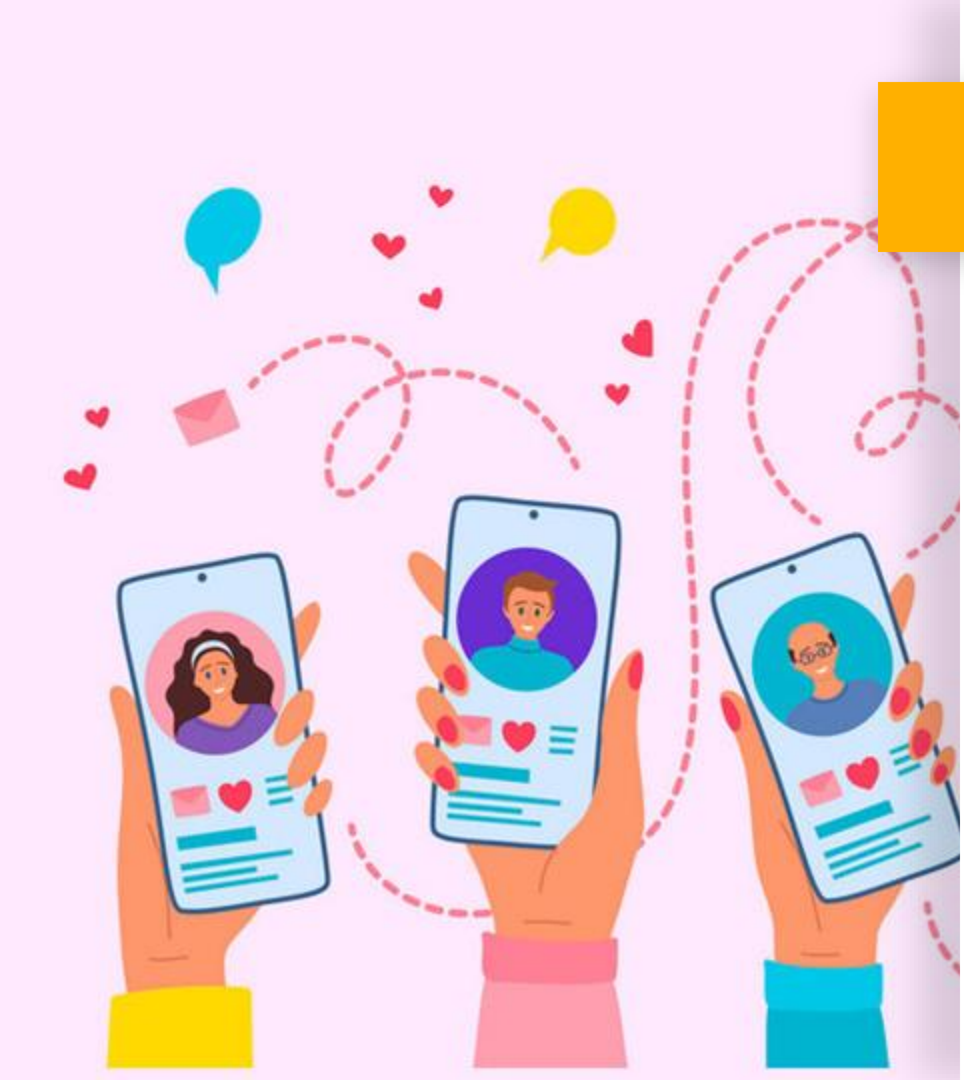
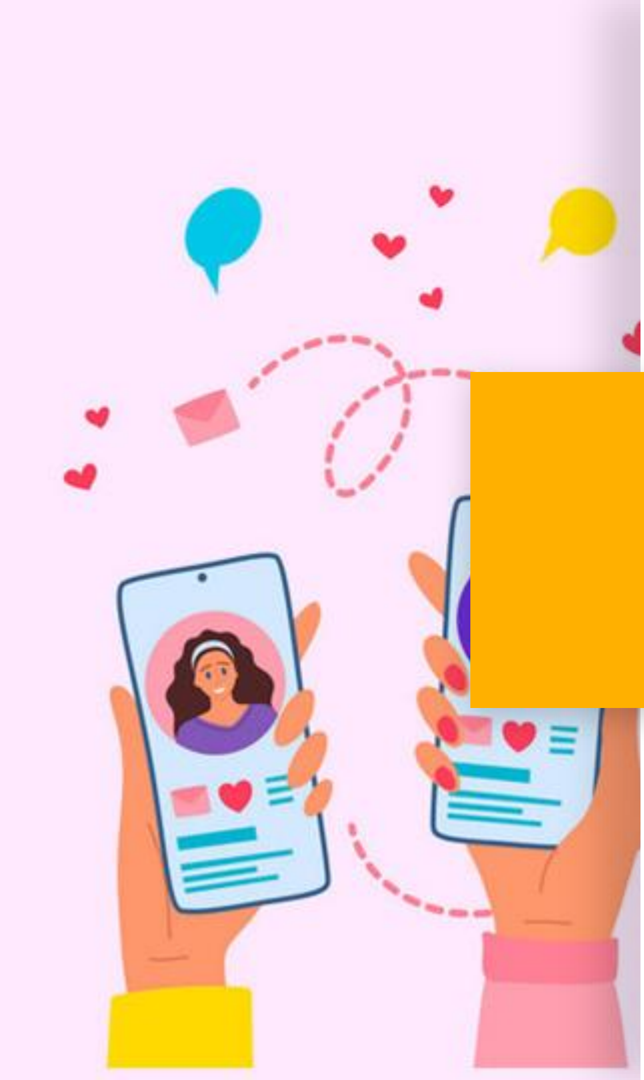
Applying Text Analytics in Dating App Reviews

Yufeng Xie, Zongqian Wu, Shuran Fu,
Xiaomeng (Laura) Yu, Xinyi (Pansy) Zhang



Contents



Overview

Top Dating Apps in the US

tinder

- Swipe-style matching
- Quick and easy
- More men than women
- Women get more attention
- Casual romance & hookups
- Charge more for age over 30

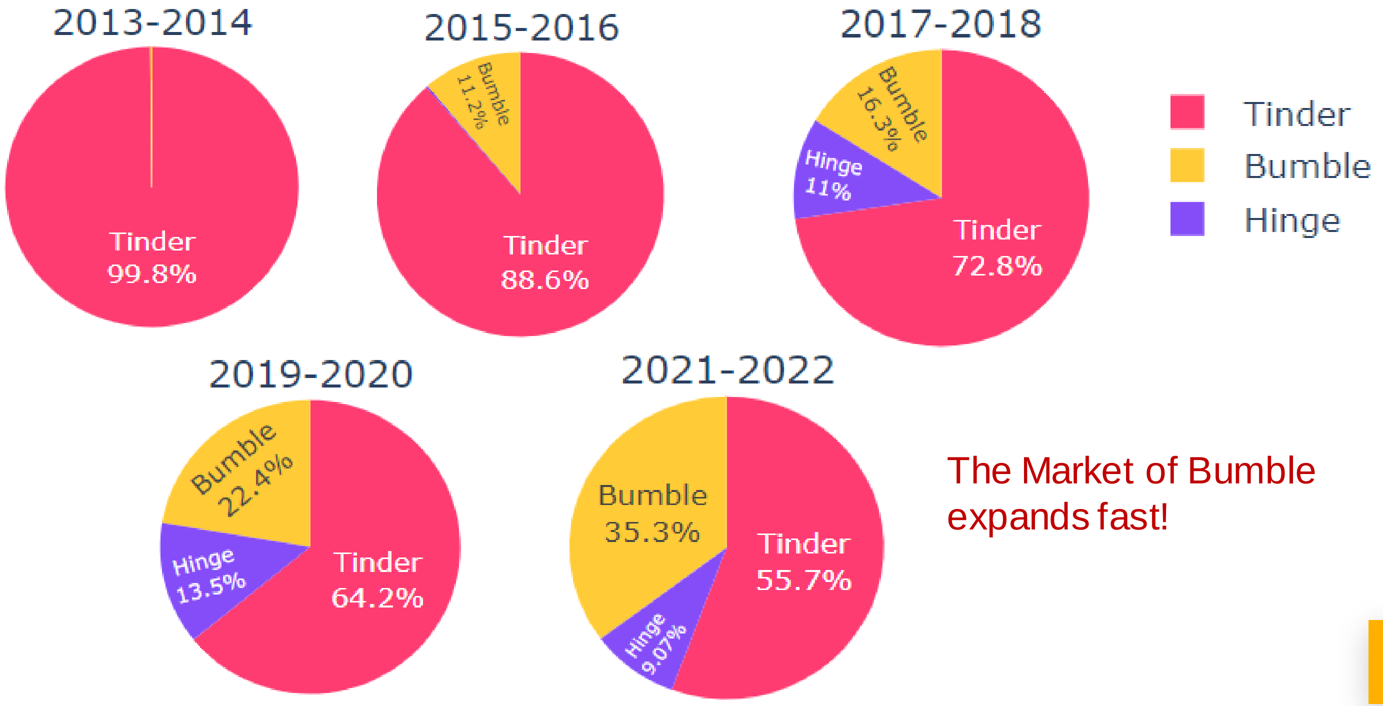
bumble

- Swipe-style matching
- Woman first
- Only woman can initiate messages
- Casual dating & friendship
- Real possibility for finding long-term partners
- Same charge for everyone

Hinge

- No swiping
- Individual match feeds
- Long sign up process
- Long profile
- Younger users
- Open to all sexual orientations

Market Share Over the Years



The Market of Bumble expands fast!

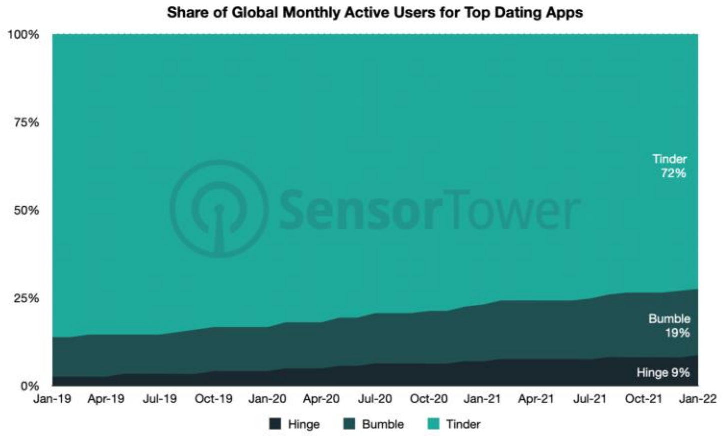
Note: Here the market share is roughly estimated by the number of reviews.

Bumble Has Huge Potential

- Bumble is the **second most popular** dating app in the US behind Tinder. Bumble's share of global monthly active users is about **19%** as of January 2022.

- Bumble's target customers are **single women** who have had bad experiences with unsolicited messages and are looking for more **initiative**. It also appeals to singles who want to find friends, casual flings, or **long-term romance** all in one place.

- We want to generate business insights for Bumble based on customers' reviews, so that Bumble can **leverage its feature differentiators to compete for a larger market share**.



Business Scope

- As the market share of Bumble becomes larger, the user experience becomes more important for decision making. From the Dating App reviews dataset, we have over 50K user reviews from their user experience. It includes tons of areas Bumble is good/bad at.

- With this dataset, we want to help Bumble identify its strengths and problems to attract more users. To make it happen frequently, we developed several NLP models to automate the topic finding process.

Business Solutions

Ratings Model

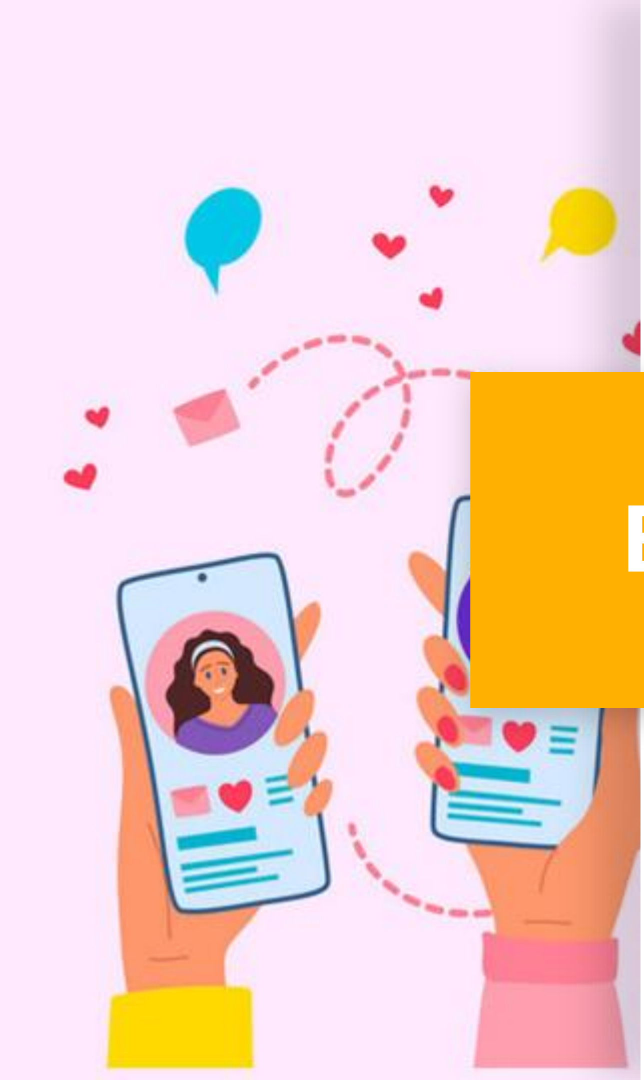
This model will generate adjusted scores. If the adjusted average score is lower than original average score, we will also put discrepancy reviews into topic modeling.

Thumbs Up model

At the end of the week, we can first use it to predict important reviews. Only important reviews will be used in topic modeling.

Topic Modeling

It helps generate insights and find out which parts Bumble should improve.



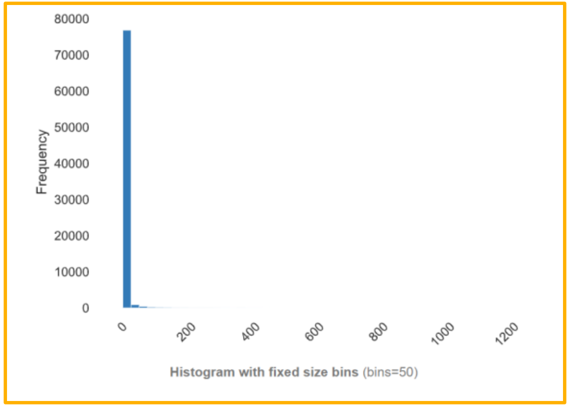
EDA & Data Processing

Filtering Data

- We did language detection for all the data and kept those reviews which are written in **English**
- Because our target App is **Bumble**, we filtered out the reviews for the other two Apps

EDA: # ThumbsUp & Rating

ThumbsUp



Rating

Value	Count	Frequency (%)
1	36164	45.8%
5	16133	20.4%
4	10019	12.7%
2	8976	11.4%
3	7601	9.6%

- 62% of the reviews did not get any thumb up
- The maximum number of thumbs up of a single review is 1276

- Most people scored 1 point for those dating app
- There is a polarization of scoring. Over 65% of users scored 1 or 5 point

Note: The EDA result is only from the data after filtering

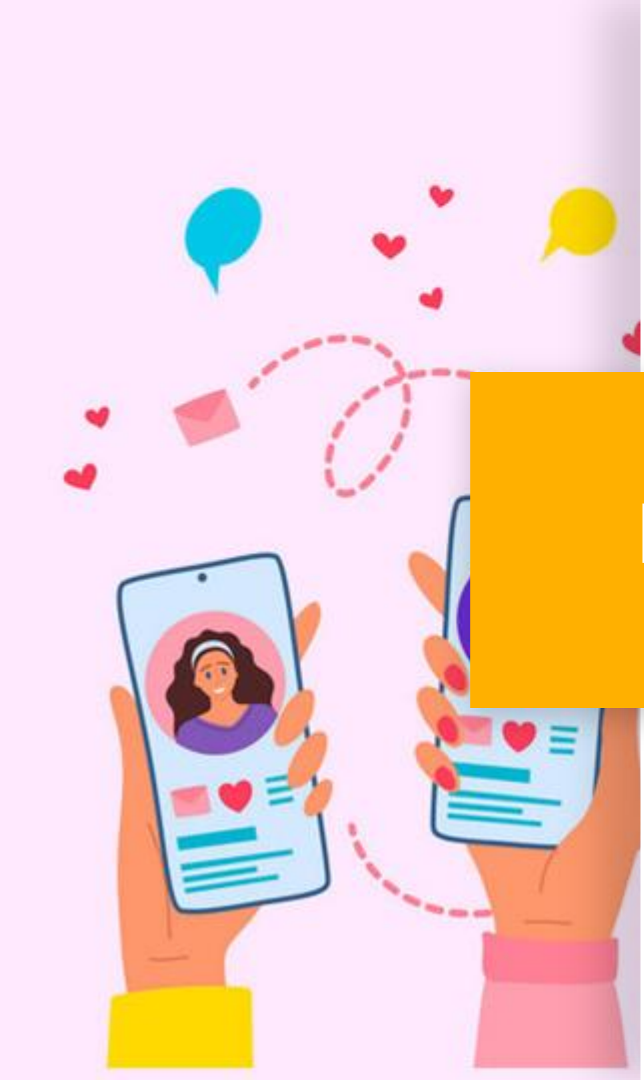
Text Preprocessing

1 Remove hashtags, emoji, URL, etc

2 Lowercase all the character

3 Lemmatization

4 Use Regex to group some words with similar meanings



Regression: Ratings Model

Regression: Ratings Models

Goal:

Build a regression model to predict the reviewers' rating based on reviews they provide. We assume that the description and words they use have a relationship with the rating scores, and most people have similar grading rubrics.

Business Use Case:

This model will help to adjust rating scores for apps and correct errors caused by malicious scoring and mismatch between reviews and rating. For example, some users who have negative feelings about the app would incorrectly rate highest scores and this model can be used to find it. This model may also add reviews data to subsequent topic modeling model.

Training Data:

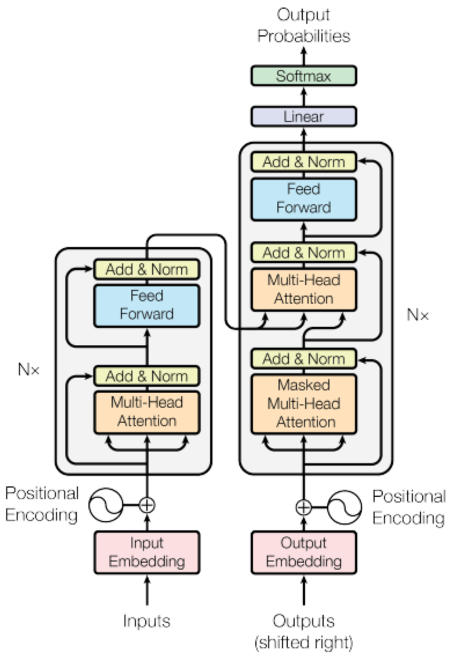
In the training dataset, we use rating scores as the dependent variables and reviews as the independent variables.

Regression: Model Performance

	MSE	R_Squared	Training time / epoch
TF-IDF+LightGBM	1.0433	0.5968	4 min
Word2vec+NN	0.9864	0.6122	14 min
DistilBERT	0.7161	0.7296	25 min
ALBERT	**0.6735**	**0.7457**	**43 min**

After training each model, we found that the fine-tuned **ALBERT** model will achieve lowest mean squared error and highest R-squared scores compared with other models, which are 0.6735 and 0.7457 respectively.

Regression: ALBERT Model Architecture



Model Structure:

- **Backbone:** encoder layers with GELU (Gaussian Error Linear Unit) activation function
- **Changes from BERT:** Factorization of the Embedding matrix; Cross-layer parameter sharing; Inter Sentence Coherence Prediction

Model Parameters:

- **Base model:** albert-base-v2
- **Learning rate:** 2e-5
- **Max length for input:** 256
- **Batch size:** 32
- **Training epoch:** 6

Regression: Adjustment Examples

Reviews	Rating	Prediction
fabulous experience could refer many more for such a lovely journey	1	4.77
I don't know why y'all keep blocking my account. Even when u have verified myself.. I didn't even know what I did..you keep blocking me . So annoying	5	1.12
have not use yet	1	3.26

Classification: Thumbs Up Model

Classification: Predicting Important Reviews

Goal:

Build a classification model to predict whether an App review will get more than 1 Thumbup. We assume those reviews with many Thumbs up are the "key opinions" from users and deserve more attention from the publisher or developer of the Dating App.

Business Use Case:

It may take up to a month for an "important" review to accumulate to 100 Thumbups. However, with this classification model, we can predict how many Thumbs up a review will get in the future immediately.

Training Data:

In the training dataset, we label all the reviews with 1 or more Thumbs up are "positive" and reviews with 0 or 1 Thumbups as "negative".

Classification: Down-sampling

Encoding:
- Positive(1): Thumbups > 1
- Negative(0): Thumbups = 0 or Thumbups = 1

Original Proportion:
- size of class 0 : size of class 1 = 4.17 : 1

Problem:
- the training model will spend most of its time on negative examples and not learn enough from positive ones
- We have tried to train on the true distribution, but the recall was extremely low.

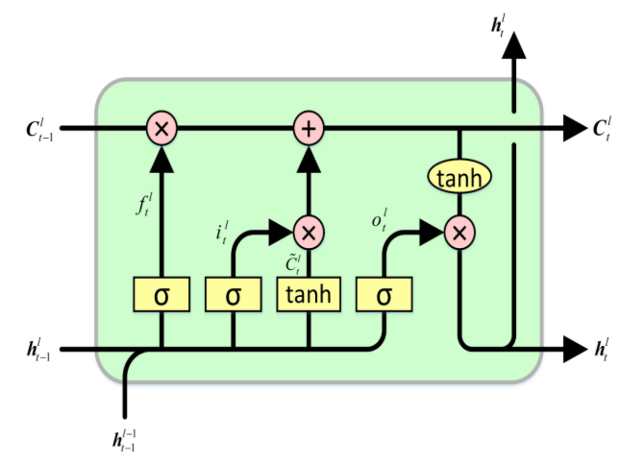
Down-Sampling:
- We used the 15258 rows of positive rows plus another 15258 negative rows sampled from 60k+ entire negative cases to make a balanced dataset.

Classification: Models Comparison

	Accuracy	Precision	Recall	ROC AUC	Runtime
GloVe+RNN	0.660	0.623	0.854	0.69	5 mins/epoch
GloVe+LSTM	0.732	0.734	0.723	0.8	7 mins/epoch
Word2Vec+Random Forest	0.692	0.682	0.718	0.75	2 mins/epoch

- Based on the calculation of overall accuracy, precision, and ROC-AUC, **LSTM** outperformed the other 2 models. The accuracy of LSTM on the test dataset reaches 71.3%.
- RNN gave the highest recall of 85%.
- Random forest, as a traditional machine learning model, cost less time for 1 epoch training

Classification: LSTM Model Architecture

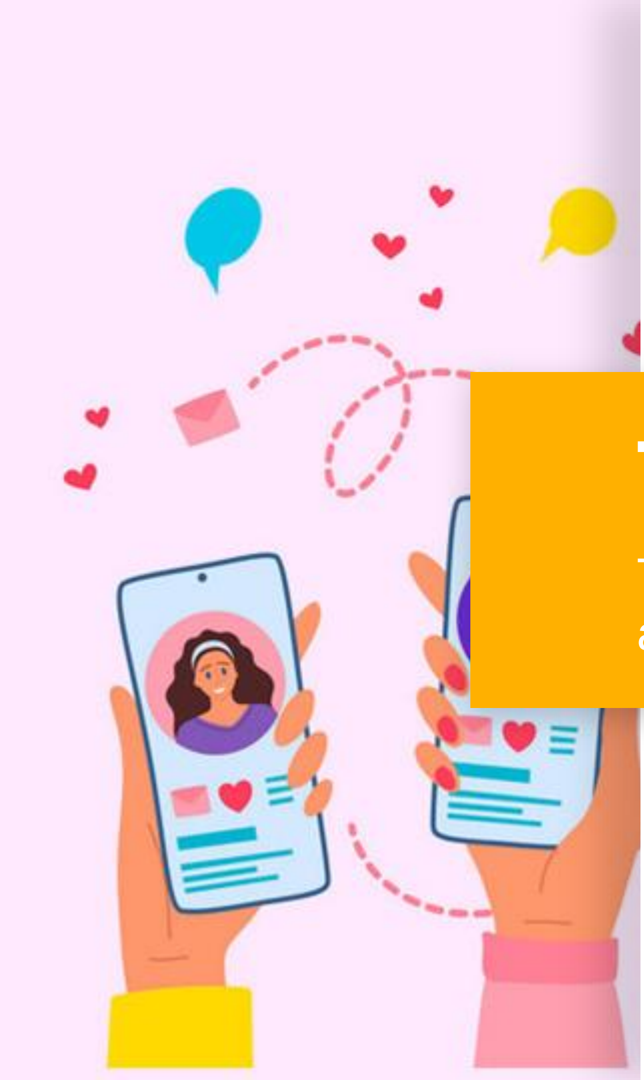


Model Structure:

- **Embeddings**: GloVe Embedding
- **LSTM layer**: the memory cell enables the model to contain long-term memory, and partially addresses the long-range dependencies problem with RNN
- **Output**: using sigmoid activation function to produce a single value of probability of belonging to the "positive" (>1) class

Model Parameters:

- **Vocabulary size**: **1.1** * #. of unique words in the corpus
- **Max length**: 855
- **Layers**: Embedding (size 100), Masking, LSTM (64 units), Dense (16 units), Dense (1 unit sigmoid)



Topic Modeling

Topics in positive reviews (rating 4-5)
and negative reviews (rating 1-3)

Topic Modeling

Goal:

Use a pre-train BERT model to accurately identify important topics based on negative and positive reviews.

Business Use Case:

There are millions of bumble app reviews online and a huge part of that information is useless. Topic modeling helps the company to filter the meaningless information and keep the important information, which enables the app analytics teams to work more efficiently to get insights.

Input Data:

In the input dataset, we label all the reviews with ratings from 1 to 3 are “negative" and reviews with ratings from 4 to 5 as “positive".

What users LIKE about Bumble

- **Woman first**
  Female users are protected from unsolicited and unwanted messages. Male users no longer have the pressure to initiate messages and come up with ice breakers while dating online.

```
[('female make', 0.05588102781453207 4),
 ('love female', 0.043583394271386 27),
 ('good female', 0.025604039620685203),
 ('fact female', 0.022937257196915757),
 ('female message', 0.021791697135693136),
 ('easy use', 0.021035061055218154),
 ('love fact', 0.018768025706761098),
 ('good app', 0.018163217939681 23),
 ('female good', 0.017751321586190853),
 ('female love', 0.017106473624986594)]
```

- **Easy to use**
  Profiles are short and concise, which makes signing up a breeze. Average sign-up time is about 3 minutes, which is lower than Tinder (~5 minutes) and Hinge (~10 minutes).

```
[('easy use', 0.43403594730901035),
 ('use easy', 0.1771208154336916),
 ('use good', 0.15796472551233734),
 ('concept easy', 0.08437370117124497),
 ('good concept', 0.05947340334626049),
 ('use fun', 0.05830470162097278),
 ('good easy', 0.04899777196161463),
 ('use quick', 0.04668780630354421),
 ('match easy', 0.04481974089801317),
 ('use love', 0.04344051684667035)]
```

What users DISLIKE about Bumble

- **24 hour response time**
  If you don't message a match within 24 hours, the match is lost. You have to purchase a match extension if you can't respond quickly.

```
[('female message', 0.011565095554194707),
 ('swipe right', 0.011493691655547865),
 ('dating apps', 0.009940158234739128),
 ('scam profiles', 0.009719840030359705),
 ('swiped right', 0.009388157288914789),
 ('24 hours', 0.009135427262358612),
 ('waste time', 0.008645490843183368),
 ('female love', 0.008334174718129198),
 ('look love', 0.008195467244529008),
 ('female match', 0.007993698887128312)]
```

- **High subscription costs**
  Paid membership plans starting $30.99 per month, which is higher than Tinder (starting $4.58 per month) and Hinge (starting $9.99 per month).

```
[('cancel subscription', 0.017747702672711125),
 ('credit card', 0.014494377550241723),
 ('cancelled subscription', 0.014111605165969177),
 ('customer service', 0.013472085681417721),
 ('month subscription', 0.010939801329070967),
 ('canceled subscription', 0.010397830974185427),
 ('day subscription', 0.009666198975182097),
 ('bank account', 0.009570009449711411),
 ('free trial', 0.009486985540956174),
 ('14 day', 0.009211746959266829)]
```

- **Customer support**
  Customer support can only be contacted via Facebook or Twitter, so it's very inconvenient for people who don't have these accounts.

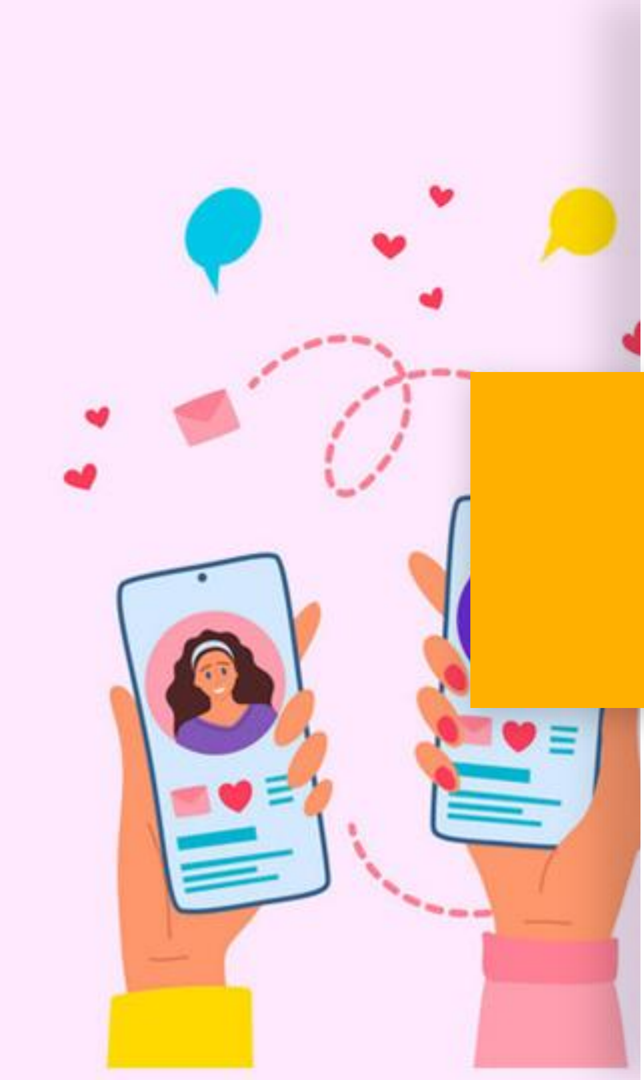
```
[('use facebook', 0.19588802314331638),
 ('facebook account', 0.13690036622235334),
 ('facebook use', 0.12452545928857754),
 ('facebook facebook', 0.11056832234716836),
 ('sign facebook', 0.08167369523066026),
 ('requires facebook', 0.07724481420097665),
 ('need facebook', 0.06849843165196277),
 ('facebook want', 0.057122446217753496),
 ('login facebook', 0.054877642206184324),
 ('account facebook', 0.05382019629797075)]
```

Business Implications

The dating app industry is becoming increasingly competitive as more and more apps enter the market and compete with their product differentiation strategies. As mentioned in the reviews, many users have tried several different apps and are ready to switch to another app if the current one doesn't work for them. Bumble needs to **promote its feature differentiators** and **fix the complaints** in order to increase its market share.

Business recommendations:

- Appeal more toward female users who like to initiate conversations and male users who want to avoid the pressure of reaching out first.
- Continue to optimize the sign-up process to make it fast, easy, and entertaining.
- Send reminders or nudge messages when the 24-hour time window is about to expire, so that users will not accidentally miss a match.
- Offer a one-month free trial with all the premium features before a formal subscription.
- Maintain a 24/7 customer support hotline to give users more accessibility.



ROI Analysis

ROI: Ratings Model

Earn (Yearly):

- Providing adjustment for topic modeling - $600k (Potential benefit for identifying the ignored topics in negative reviews with high ratings)
- Improving the rating in app stores - $600k (Potential benefit for improving the rating performance after solving identified problems)

Cost (Yearly):

- Technical implementation costs - $120k
- Mismatch check labor costs - $480k (Mismatching problems between review and ratings should be identified and analyzed by staff)

Return on Investment (Yearly):

$1.2M - $0.6M = $0.6M

ROI: Thumbs Up Model

Earn (Yearly):

Saving time for the Bumble publisher or developer to identify "keep opinions" from users so that they may act proactively to prevent user churn. The classification model will work much more efficiently than eyeballing thousands of online reviews – increasing the annual profit by 2% - $3M

Cost (Yearly):

Hiring Data Engineer or Data Scientist to clean the data, train models, and deploy the selected model into a real-world business application. - wages or other additional expenditure for the data team - $1.2M

Return on Investment (Yearly):

$3M – $1.2M = $1.8M

ROI: Topic Modeling

Earn (Yearly):

Improvement for the overall performance of the app - $3M (Potential benefit for solving the problems identified by topic modeling through negative reviews)

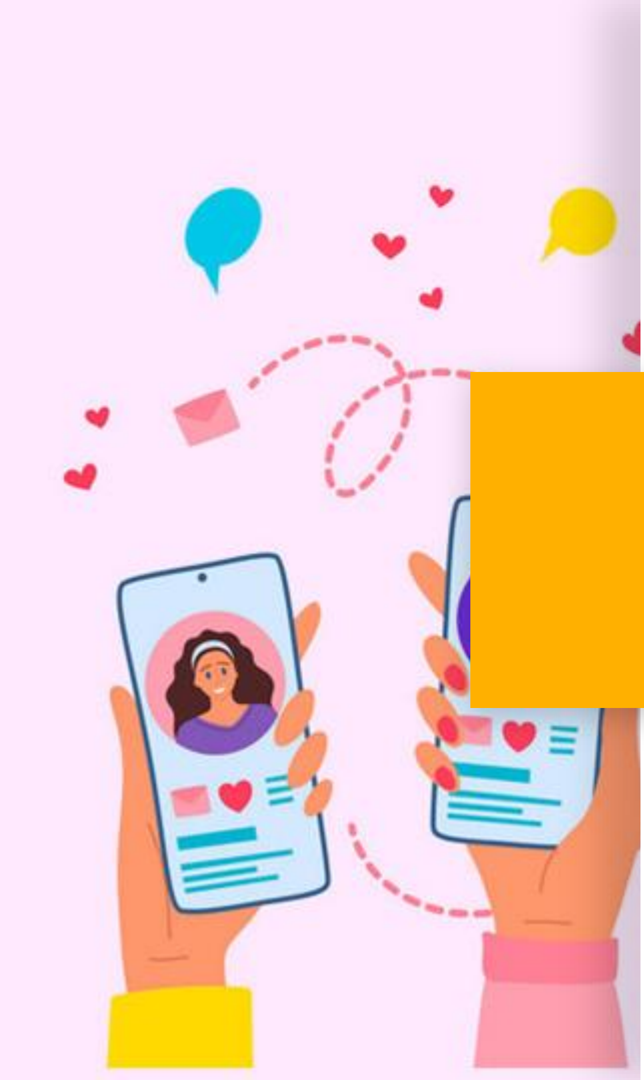
Cost (Yearly):

Data Science Team Labor costs - $600k

Technical implementation costs - $120k

Return on Investment (Yearly):

$3M - $0.72M = $2.28M



Conclusion

Conclusion

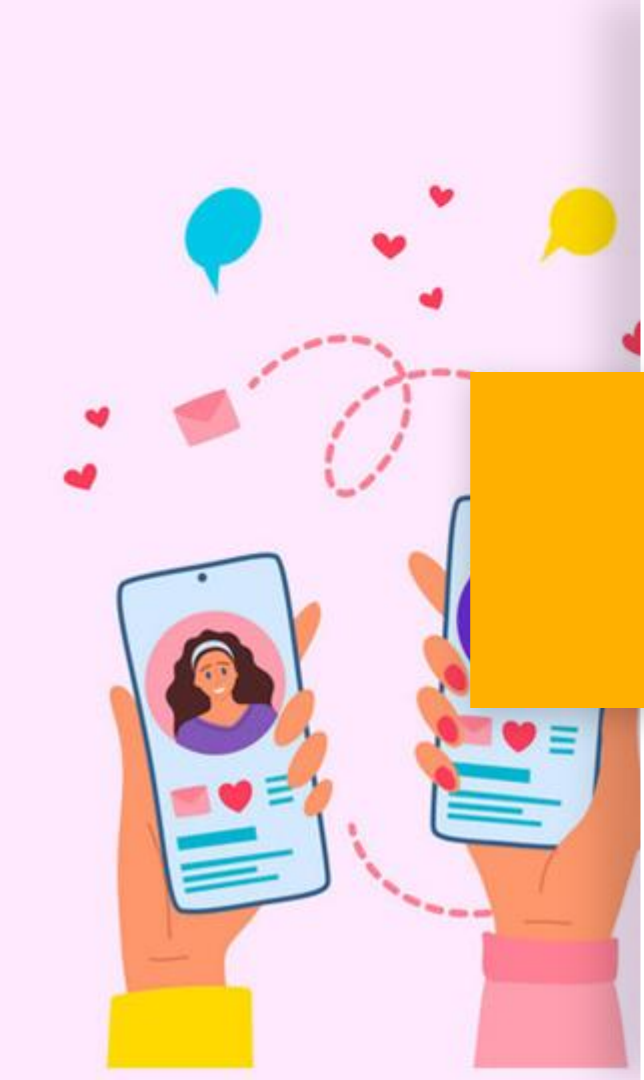
We developed three models: Ratings Regression, Thumbs Up Classification, and Topic Modeling. According to our estimate, the total Return on Investment will be **$4.68M**.

The first two models are used to detect potentially insightful reviews, which will be fed into the topic modeling model. The whole process can be performed on a weekly basis, and it will largely improve the operational efficiency than manually checking the reviews. The ways how our models help to improve operational efficiency are explained below.

- **Thumbs Up model**: Save waiting time for thumbs up; Select important reviews automatically.
- **Ratings Model**: Adjust original rating score; Select reviews for discrepancy analysis automatically.
- **Topic Modeling**: Identify potential problems; Automate report generating process.

References

- https://healthyframework.com/dating/review/bumble/
- https://sensortower.com/blog/dating-apps-2022/



Thank You!