

How to Be THE BEST Mexican restaurant in yelp*

GROUP 7 | HAN LIAO, QIAOCHU YU, YUJIE ZHANG, LINGFENG ZHU



CONTENTS

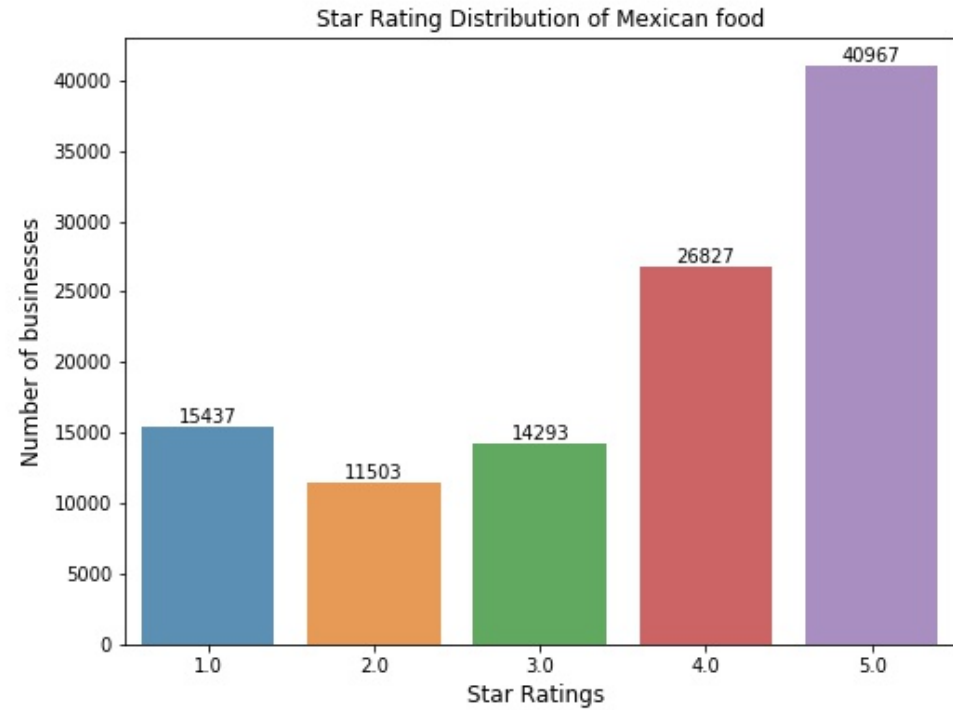
- Goal & workflow
- EDA
- Data Preprocessing
- Model

GOAL

- Offer specific suggestions to restaurants 'owners based on reviews on Yelp.

Why we choose Mexican restaurants?

- Large enough
(1738 restaurants,
109027 reviews).
- Relatively balanced.



WORKFLOW



EDA

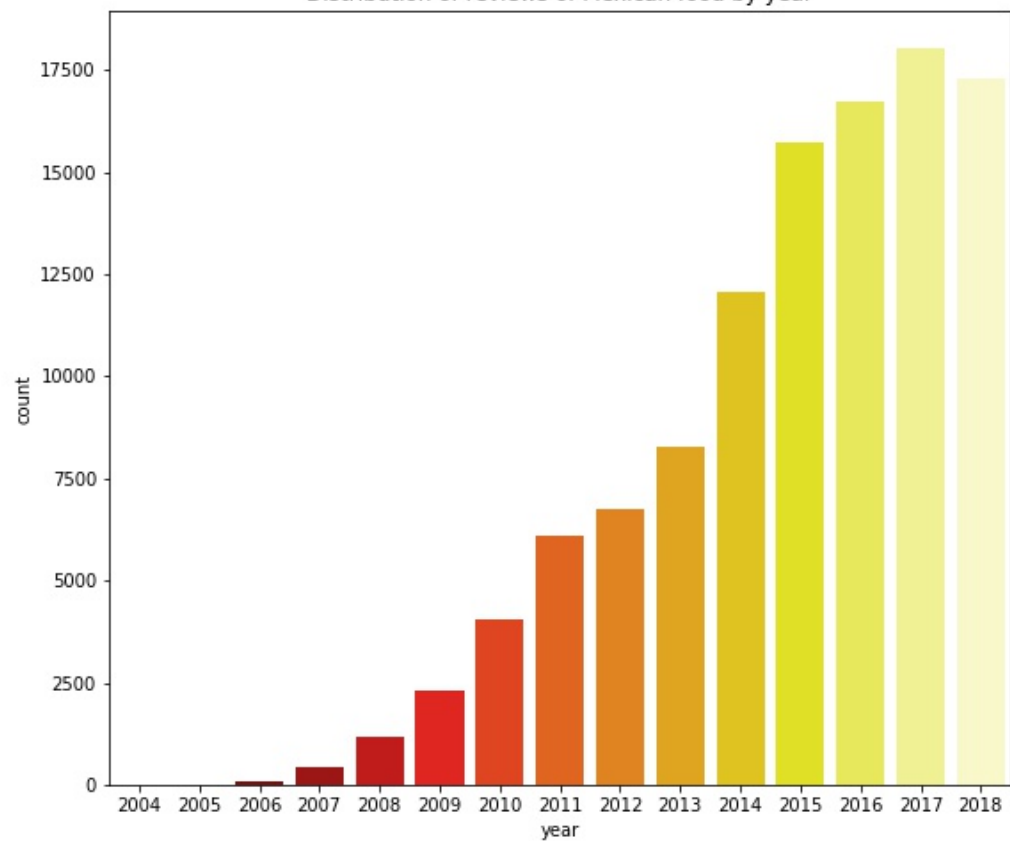
- Location.
- Time.
- wordcloud.
- words' frequency versus ratings.

Location

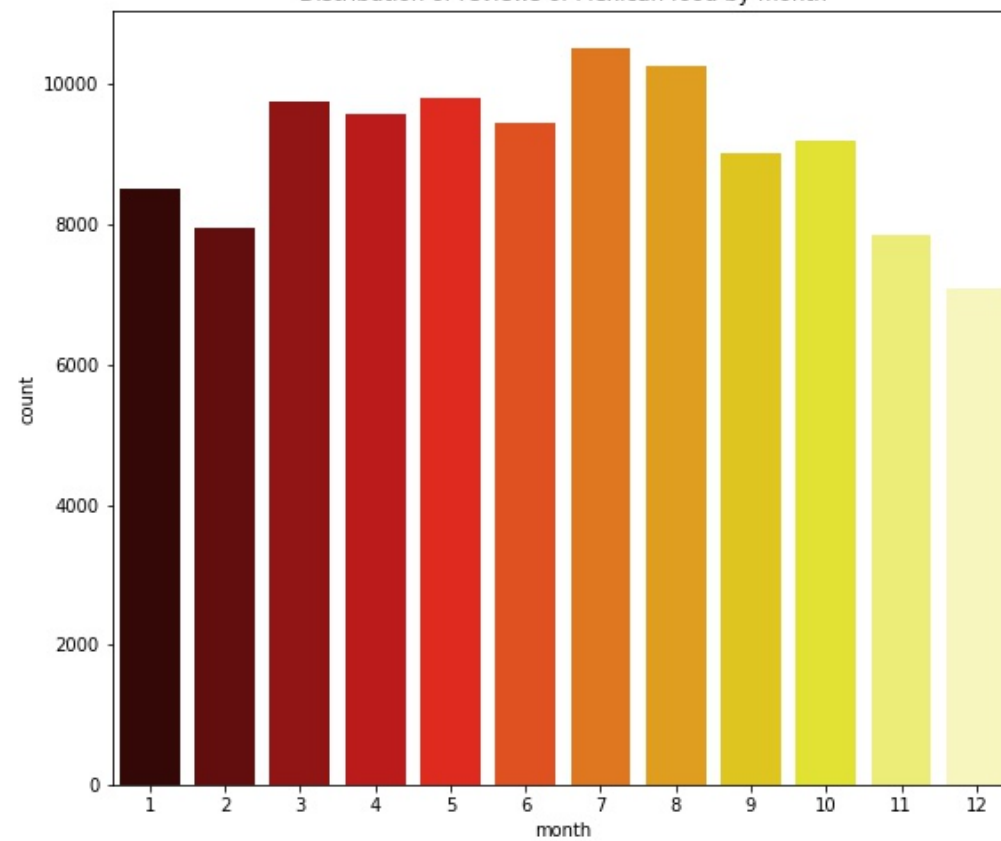


Time

Distribution of reviews of Mexican food by year



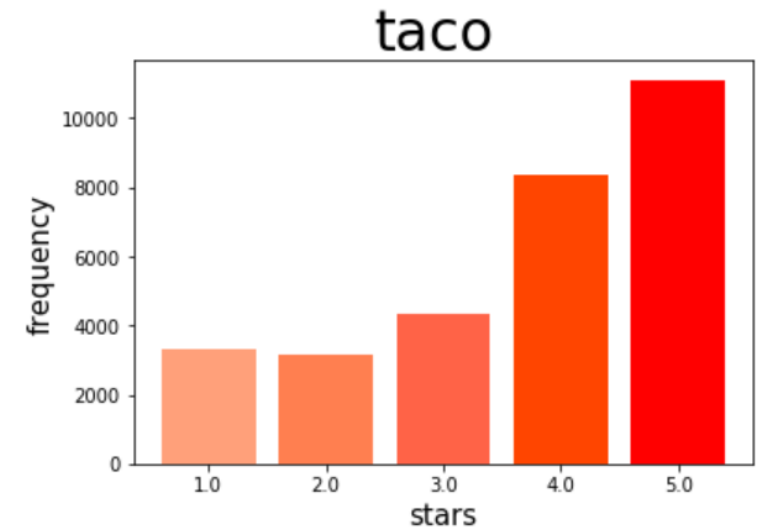
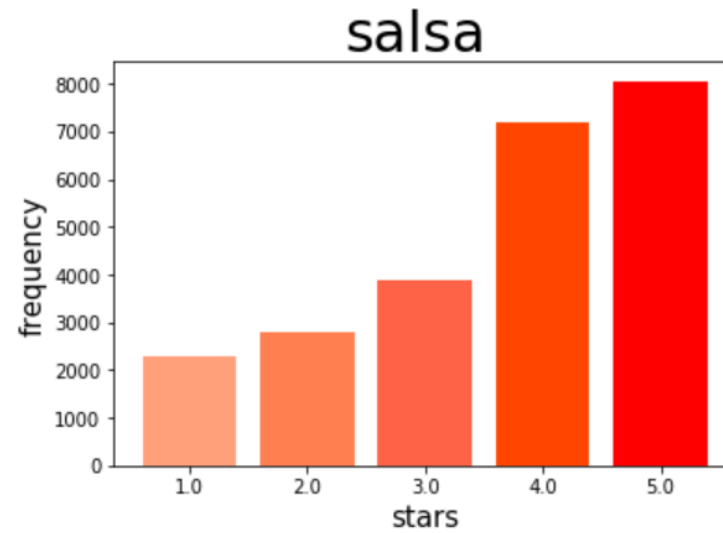
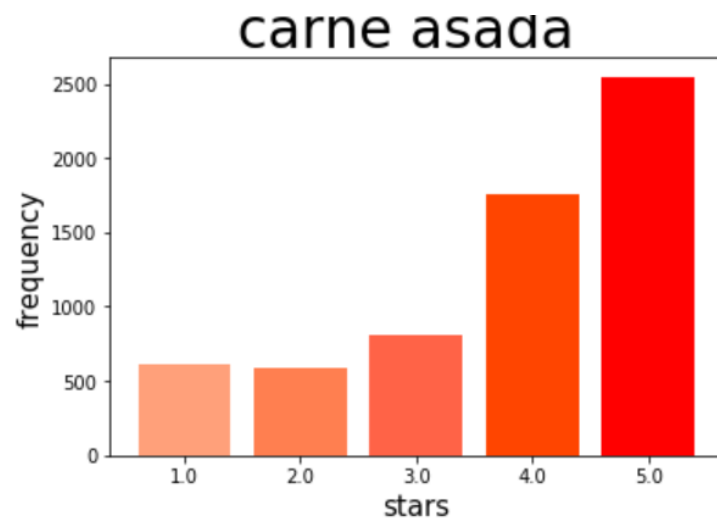
Distribution of reviews of Mexican food by month



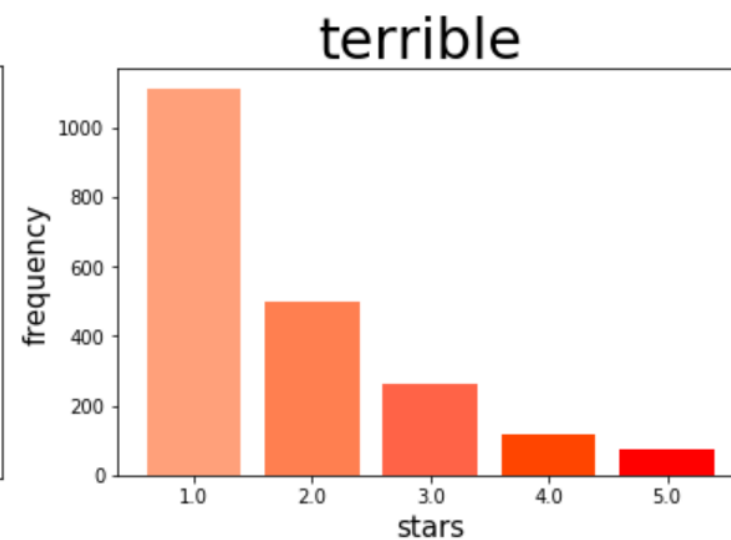
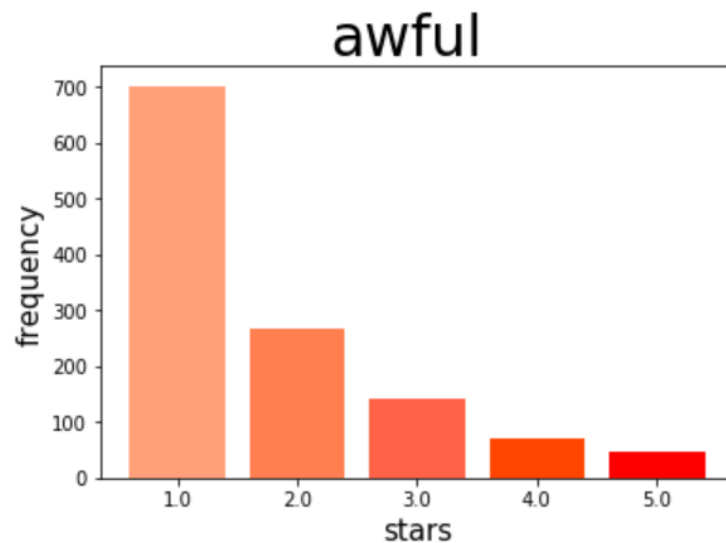
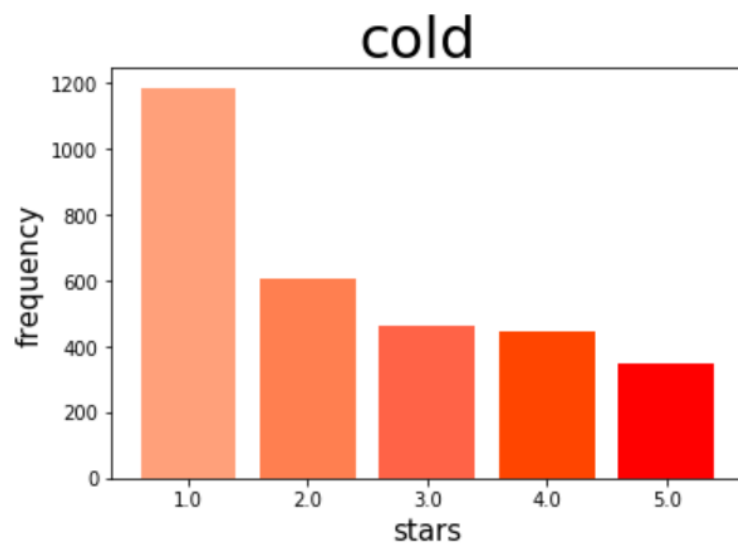
Wordcloud



Some Words' Frequency Versus Stars



Some Words' Frequency Versus Stars



Great food in a homely Mexican setting.
The tomato soup was amazingly delicious
and their chicken dishes are great.



great food in a homely mexican setting. the
tomato soup was amazingly delicious and
their chicken dishes are great.

DATA-PREPROCESSING

- Lowercase text
- Tokenization
- Lemmatization & Stemming
- Embedding

great food in a homely mexican setting. the
tomato soup was amazingly delicious and
their chicken dishes are great.



great/food/in/a/homely/mexican/setting/t
he/tomato/soup/was/amazingly/delicious/
and/their/chicken/dishes/are/great

DATA-PREPROCESSING

- Lowercase text
- Tokenization
- Lemmatization & Stemming
- Embedding

great/food/in/a/**homely**/mexican/setting/t
he/tomato/soup/**was**/**amazingly**/delicious/
and/**their**/chicken/**dishes**/**are**/great



great/food/in/a/**home**/mexico/**set**/the/tom
ato/soup/**be**/**amaze**/delicious/and/**they**/chi
cken/**dish**/**be**/great

DATA-PREPROCESSING

- Lowercase text
- Tokenization
- Lemmatization & Stemming
- Embedding

Embedding (GloVe)

Man (5391)	Woman (9853)	King (4914)	Queen (7157)	Apple (456)	Orange (6257)
$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$

One hot

*Jeffrey Pennington, Richard Socher, Christopher D. Manning. 2014. Glove: Global vectors for word representation.
<https://www.deeplearning.ai/>*

Embedding (GloVe)

	Man (5391)	Woman (9853)	King (4914)	Queen (7157)	Apple (456)	Orange (6257)
Gender	-1	1	-0.95	0,97	0	0.01
Royal	0.01	0.02	0.93	0.95	-0.01	0
Age	0.03	0.02	0.7	0.69	0.03	-0.02
Food	0.04	0.01	0.02	0.01	0.95	0.97
...

*Jeffrey Pennington, Richard Socher, Christopher D. Manning. 2014. Glove: Global vectors for word representation.
<https://www.deeplearning.ai/>*

SELECT LABELS FOR REVIEW

- Clustering (k-means, DBSCAN)

- Manual labeling

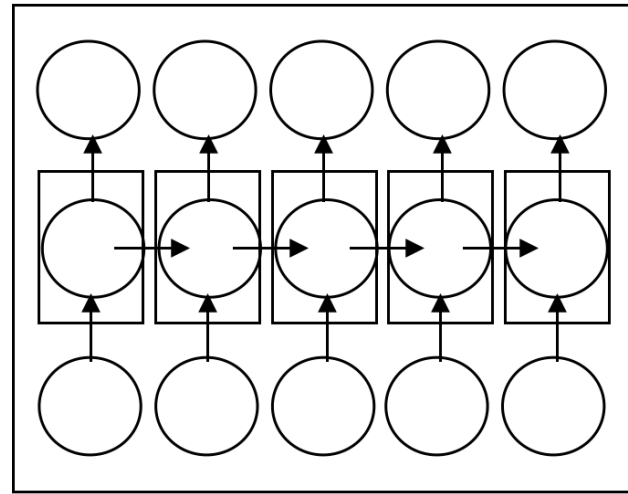
I.	Food quality	Negative/ positive
II.	Service quality	I. 1~3= Negative
III.	Environment	II. 4~5= Positive
IV.	Price	

MODEL (BIDIRECTIONAL LSTM RNN)

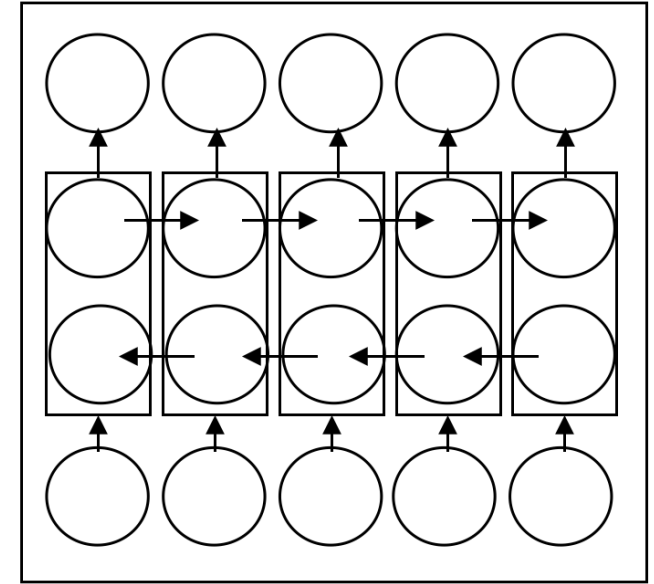
- Bidirectional recurrent neural network (BRNN)
- LSTM (Long short-term memory)
- Tree-based models (GBDT/Random Forest/ XGboost)

BRNN (Bidirectional Recurrent Neural Networks)

BRNN connect two hidden layers of opposite directions to the same output. With this form of generative deep learning, the output layer can get information from past (backwards) and future (forward) states simultaneously.



(a)



(b)

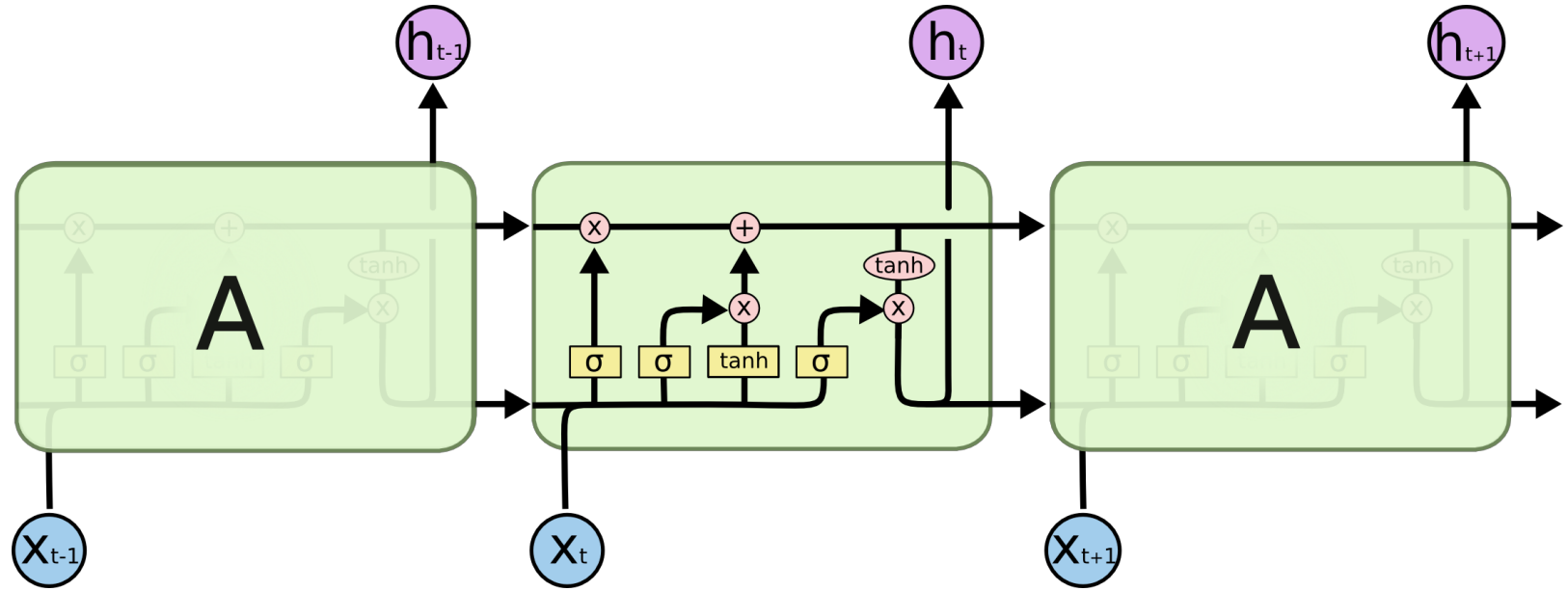
Structure overview

(a) unidirectional RNN

(b) bidirectional RNN

Long Short-Term Memory (lstm)

Long Short Term
Memory networks are
a special kind of RNN,
designed to avoid the
long-term
dependency problem.



Further Consideration

- Number of useful votes received by reviews (to distribute the weight among them)
- A review may have 2 labels at the same time.



Thank you!