How to Be THE BEST Mexican restaurant in yelp

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

CONTENTS

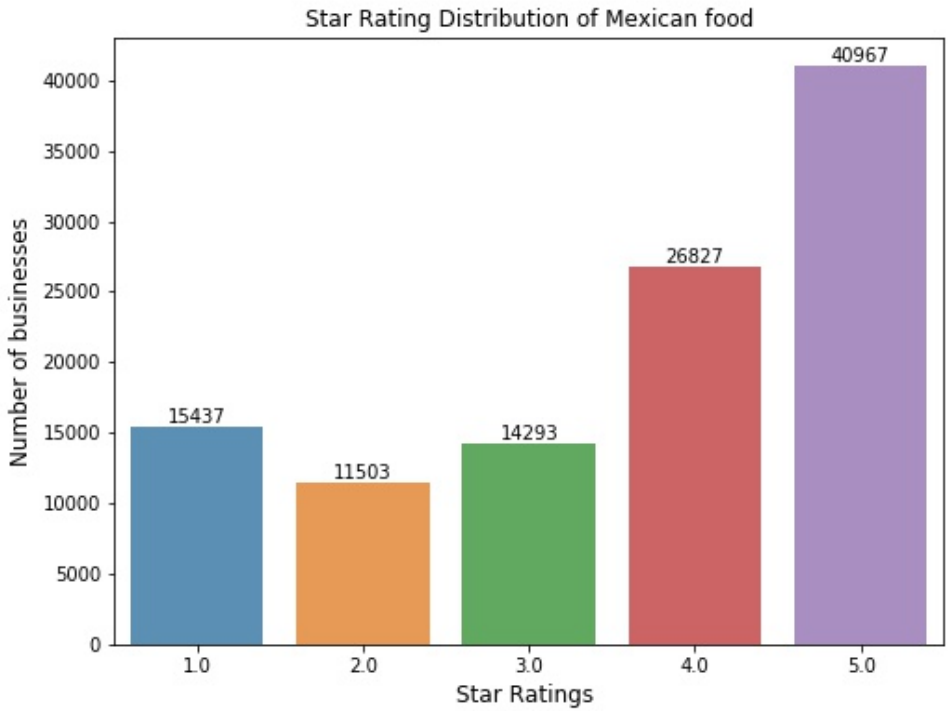
- Goal & workflow
- EDA
- Data Preprocessing
- Model

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


Data-Preprocessing
Lowercase
Tokenization
Lemmatization
Stemming
Embedding
Setting labels
Clustering (k-means/DBSCAN)
Manual labeling (Food quality / Service quality / Environment / Price)
Classification
BRNN with LSTM
Tree-based models(GBDT/RF/XGboost)

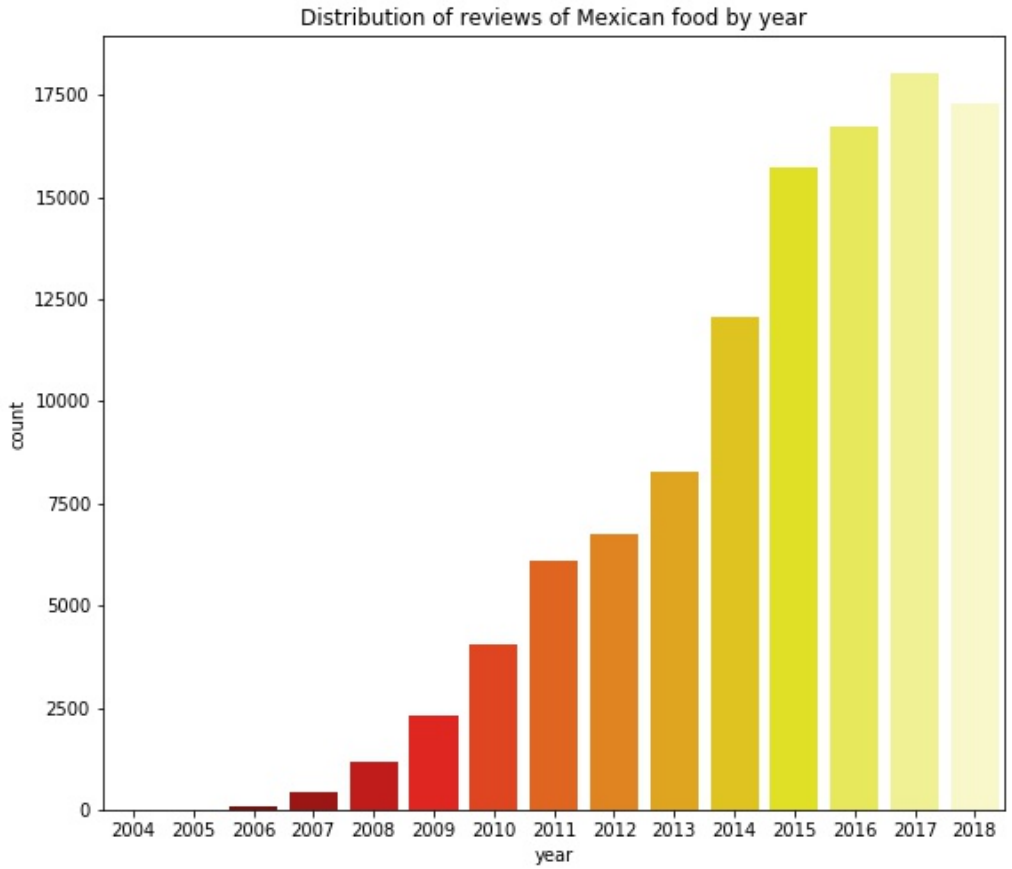
EDA

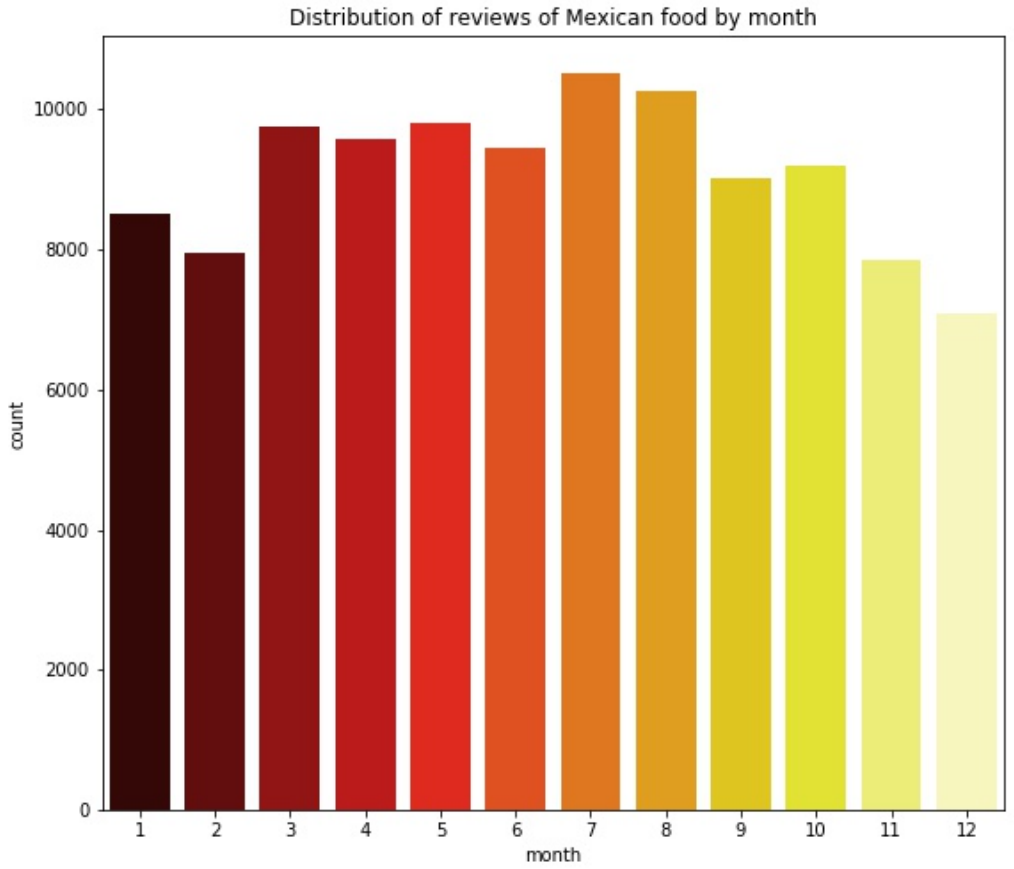
- Location.
- Time.
- wordcloud.
- words' frequency versus ratings.

Location


Montréal
Toronto
Pittsburgh
Las Vegas
Glendale
Phoenix
Charlotte

Time

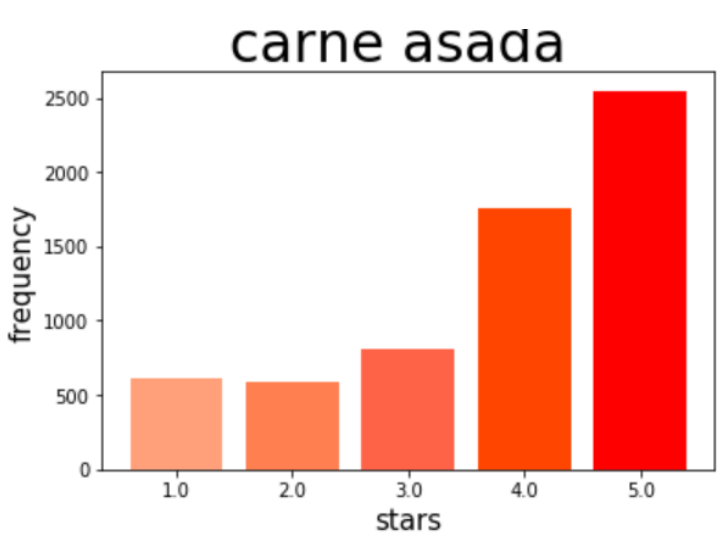

Distribution of reviews of Mexican food by year
count
year
0
2500
5000
7500
10000
12500
15000
17500
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018

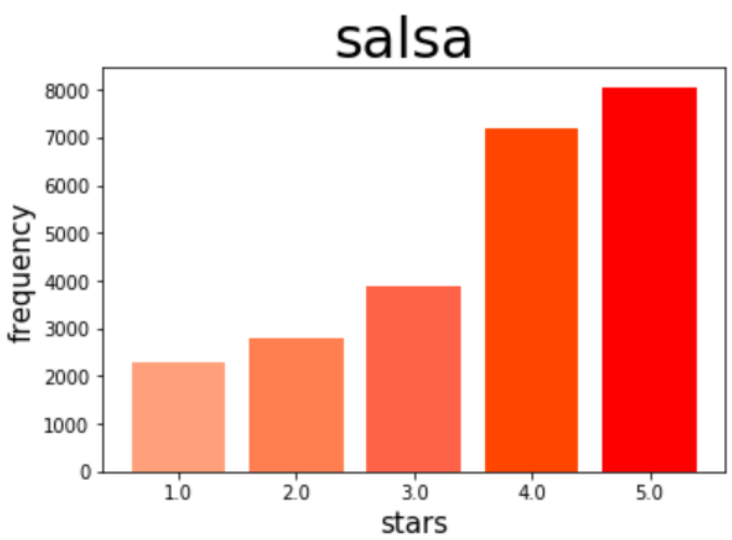

Distribution of reviews of Mexican food by month
count
month
0
2000
4000
6000
8000
10000
1
2
3
4
5
6
7
8
9
10
11
12

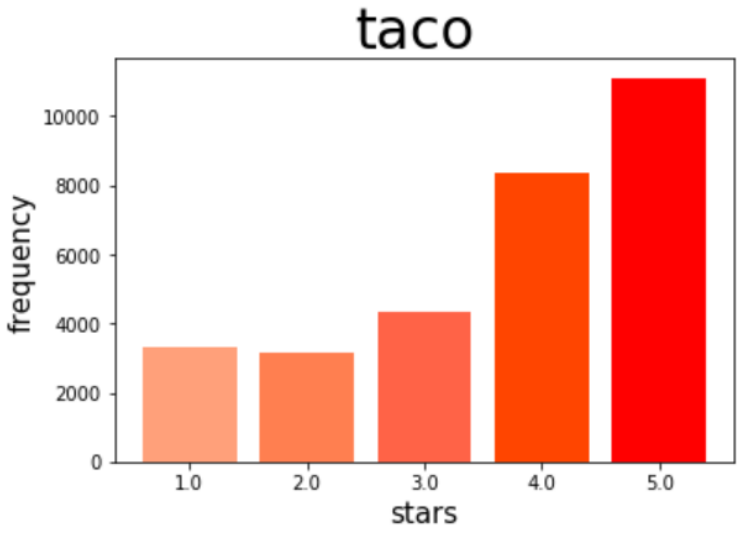
Wordcloud


Mexican food
carne asada
chips salsa
Mexican restaurant
one
food
good
now
rice bean
fish taco
really good
well
think
meal
know
authentic Mexican
great food
good food
first time
go back
even
got
eat
still
said
though
location
drink
say
customer service
best Mexican
sour cream
tried
margarita
restaurant
come back
thing
server
went
make
look
friend
order
food great
thought
taste
see
way
maybe
lot
dinner
pretty good
happy hour
flavor
great service
plate
usually
actually
breakfast burrito
will definitely
amazing
everything
table
awesome
salsa bar
every time
perfect
kind
next time
left
love
bad
inside
tasty
disappointed
staff friendly
work
special
choice
used
either
waiter
entree
eating
around
everyone
good Mexican
chicken taco
dish
waitress
come
spot
found
felt
cook
ok
refried bean
excellent
much
something
Oh
ye
sure
food place
big asada burrito
asked
bar
street taco
service great
shredded beef
feel
Mexican place
take
meat
find
plu
looking
owner
fantastic
dishe
will back
Thank
although
Taco Bell
guy
Overall
seemed
yet
want
almost
full
mean
okay
yummy
love place
delicious
especially
huge
great place
food service experience
least
believe
far
service good
put
highly recommend
check
small
offer
loved
nacho
guacamole
good service
going back
beef taco
told
asada taco
guess
came
day
may
right
people
fast food
anything
husband
instead
corn tortilla
option
ate
quesadilla
tortilla
today
serve
enjoy
hand
need
hot sauce
beans rice
probably
coming back
ask
tamale
top
served
even though
point
seem
going
area
menu
kid
al pastor
enough
tell
food delicious
use
wife

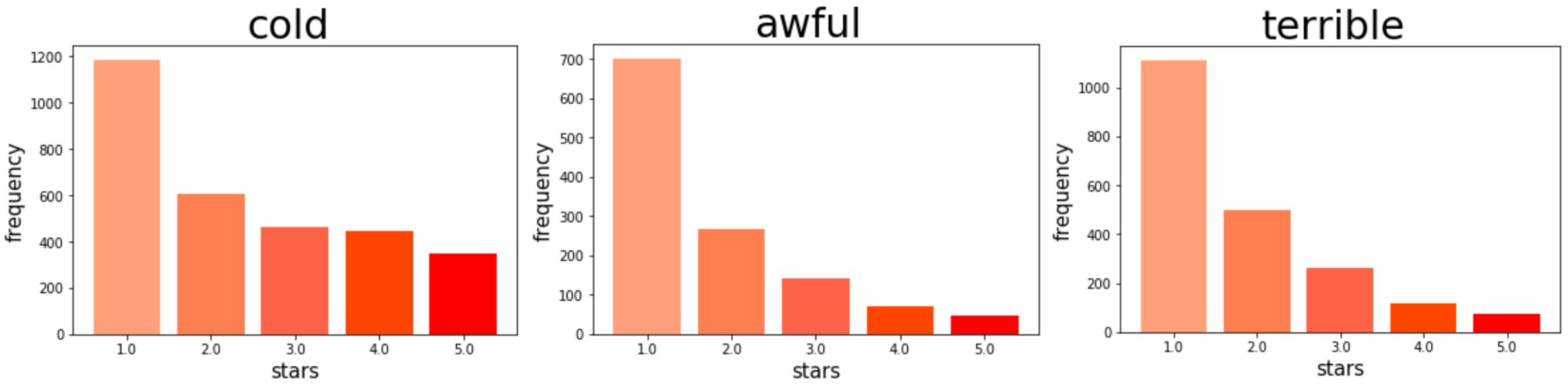
Some Words' Frequency Versus Stars







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/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/home/mexico/set/the/tomato/soup/be/amaze/delicious/and/they/chicken/dish/be/great

DATA-PREPROCESSING

- Lowercase text
- Tokenization
- **Lemmatization & Stemming**
- Embedding

Embedding (Word2vec/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 (Word2vec/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
II. Service quality
III. Environment
IV. Price

Negative/ positive

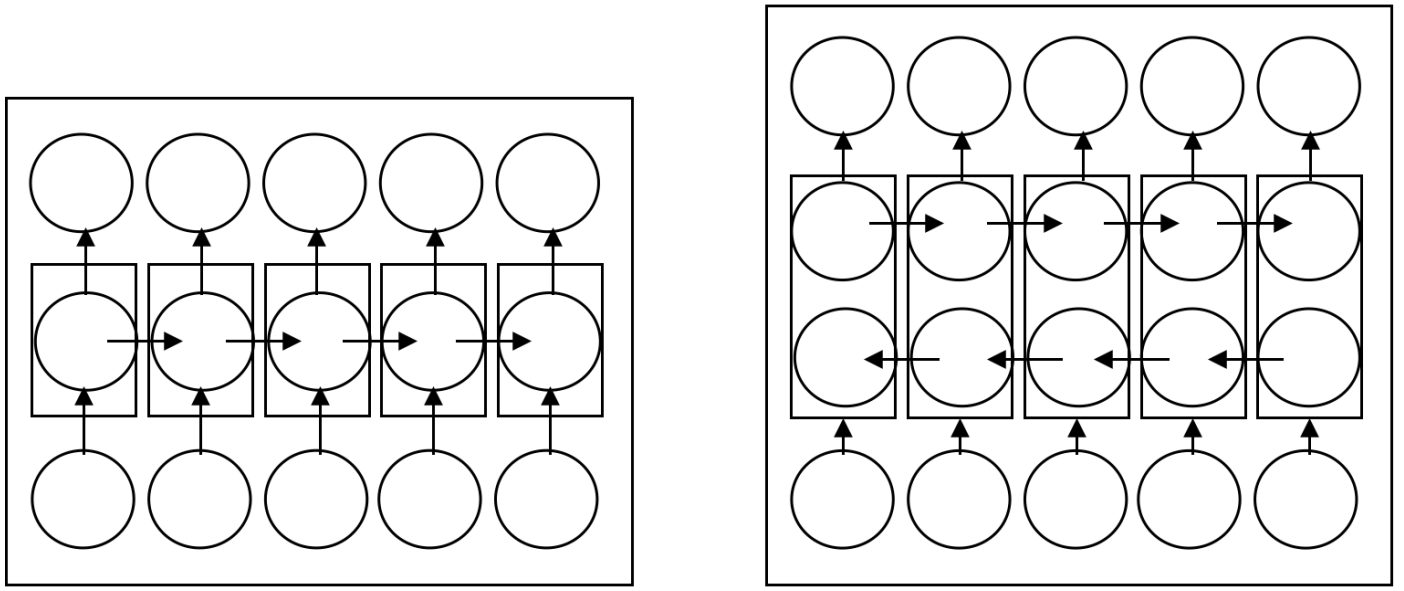
I. 1~3: Negative
II. 4~5: Positive

MODEL (BIDIRECTIONAL LSTM RNN)

- Bidirectional recurrent neural network (BRNN)
- LSTM (Long short-term memory)
- Tree-based models

BRNN (Bidirectional Recurrent Neural Networks)

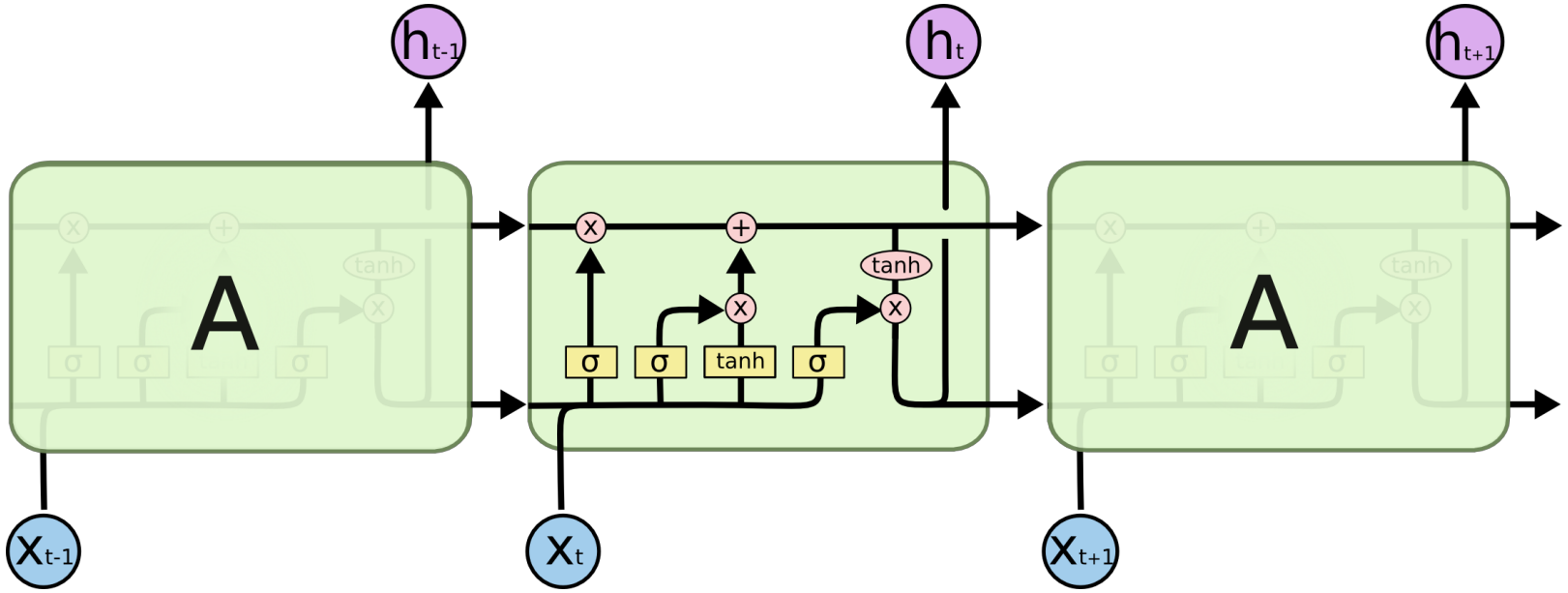
BRNN connect two hidden layers of opposite directions to the same output. With this form of generative deep learning, the output layer can get information from past (backwards) and future (forward) states simultaneously.



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.



Tree-Based Models

- GBDT
- Random Forest
- XGboost

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!