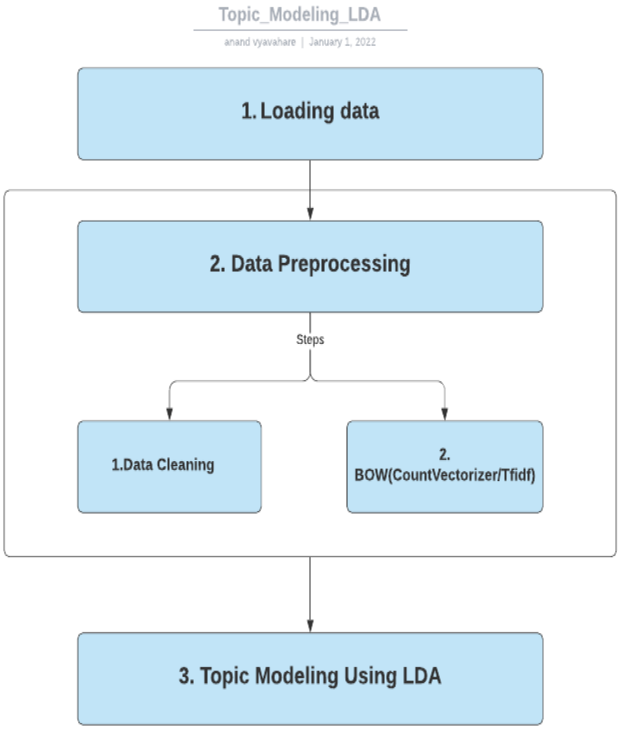
Topic Modeling Using LDA

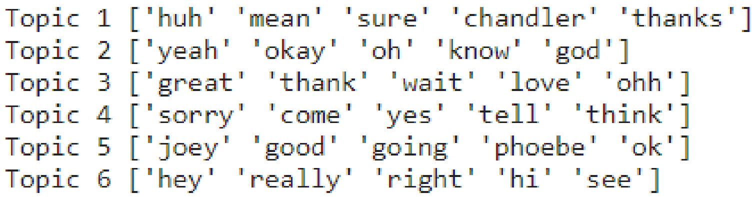
Topic modeling in NLP is the technique of identifying a set of topics that best describe the documents. It is unsupervised technique in which we identify and extract the topics from the underlying patterns like clustering algorithms.In LDA, latent indicates the hidden topics present in the data and Dirichlet is a form of distribution.

Following is the flowchart showing the steps we followed for topic modeling using LDA.


Topic_Modeling_LDA
anand vyavahare | January 1, 2022
1. Loading data
2. Data Preprocessing
Steps
1.Data Cleaning
2. BOW(CountVectorizer/Tfidf)
3. Topic Modeling Using LDA

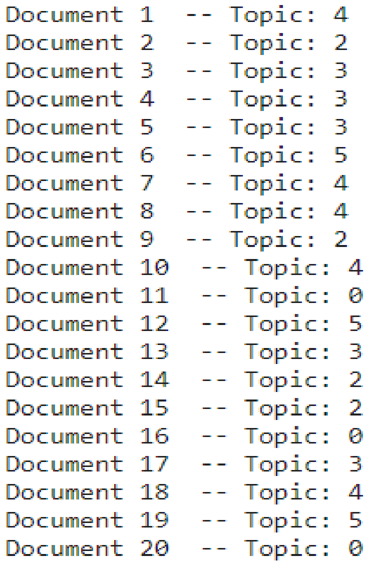
Results

We decided to go with 6 topics and below are the 6 topics and the words in each of them. We can see that since our corpus is not specific to any domain and is from a comedy sitcom, the topics and the words are expressions in the show, character names, etc. Same can be seen from the wordcloud below.



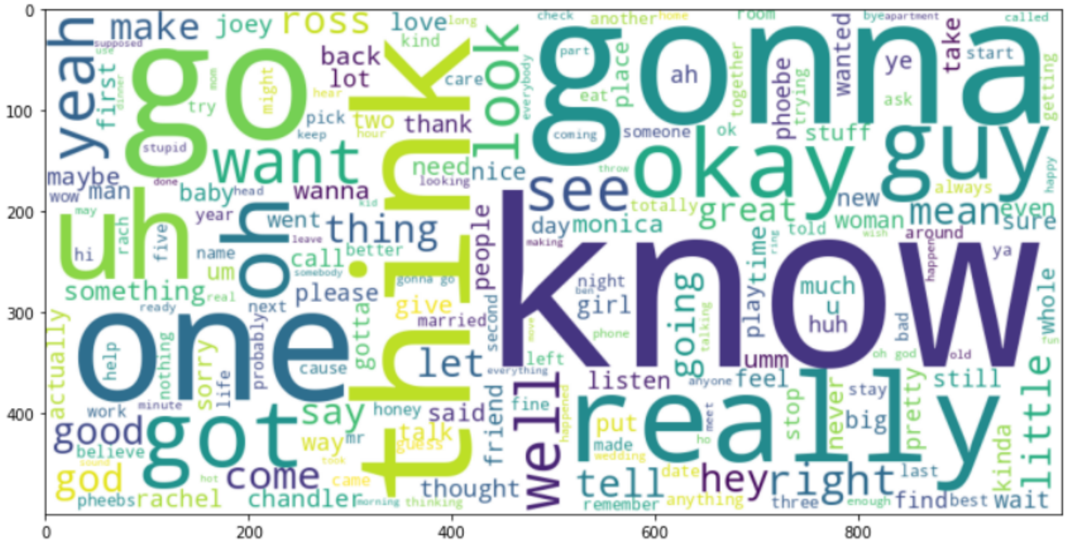
```
Topic 1 ['huh' 'mean' 'sure' 'chandler' 'thanks']
Topic 2 ['yeah' 'okay' 'oh' 'know' 'god']
Topic 3 ['great' 'thank' 'wait' 'love' 'ohh']
Topic 4 ['sorry' 'come' 'yes' 'tell' 'think']
Topic 5 ['joey' 'good' 'going' 'phoebe' 'ok']
Topic 6 ['hey' 'really' 'right' 'hi' 'see']
```

We can see which topic our documents(dialogues in our case) belong to



```
Document 1  -- Topic: 4
Document 2  -- Topic: 2
Document 3  -- Topic: 3
Document 4  -- Topic: 3
Document 5  -- Topic: 3
Document 6  -- Topic: 5
Document 7  -- Topic: 4
Document 8  -- Topic: 4
Document 9  -- Topic: 2
Document 10  -- Topic: 4
Document 11  -- Topic: 0
Document 12  -- Topic: 5
Document 13  -- Topic: 3
Document 14  -- Topic: 2
Document 15  -- Topic: 2
Document 16  -- Topic: 0
Document 17  -- Topic: 3
Document 18  -- Topic: 4
Document 19  -- Topic: 5
Document 20  -- Topic: 0
```

Wordcloud


know
gonna
go
think
one
really
okay
guy
uh
oh
got
yeah
want
see
look
well
make
joey
ross
love
right
tell
hey
great
mean
thing
going
good
god
come
chandler
rachel
monica
phoebe
people
say
let
little
much
huh
u
time
play
girl
night
something
please
call
back
lot
thank
two
need
nice
baby
wanna
maybe
man
first
actually
sorry
listen
feel
still
big
put
said
talk
way
thought
stop
never
pretty
kinda
whole
ah
ye
take
start
wanted
stuff
new
woman
even
sure
around
day
went
give
married
next
people
friend
fine
last
find
wait
three
pheebs
believe
help
nothing
life
cause
honey
remember
anything
date
ok
someone
place
eat
ask
always
totally
told
umm
bad
stay
left
second
gotta
probably
name
um
hi
five
rach
year
try
first
might
care
kind
another
room
called
getting
trying
together
ya
maybe
wow