Topic Modeling through LDA(Latent Dirichlet Allocation) of Student's Feedback

```
!pip install PyDrive
!pip install gensim
#!pip install pyldavis
!python -m spacy download en
```

```
    Requirement already satisfied: PyYAML>=3.0 in /usr/local/lib/python3.7/dist-packages
    Requirement already satisfied: oauth2client>=4.0.0 in /usr/local/lib/python3.7/dist-p
    Requirement already satisfied: google-api-core<2dev,>=1.21.0 in /usr/local/lib/python
    Requirement already satisfied: uritemplate<4dev,>=3.0.0 in /usr/local/lib/python3.7/d
    Requirement already satisfied: httplib2<1dev,>=0.15.0 in /usr/local/lib/python3.7/dis
    Requirement already satisfied: six<2dev,>=1.13.0 in /usr/local/lib/python3.7/dist-pac
    Requirement already satisfied: google-auth>=1.16.0 in /usr/local/lib/python3.7/dist-p
    Requirement already satisfied: google-auth-httplib2>=0.0.3 in /usr/local/lib/python3.
    Requirement already satisfied: setuptools>=40.3.0 in /usr/local/lib/python3.7/dist-pa
    Requirement already satisfied: requests<3.0.0dev,>=2.18.0 in /usr/local/lib/python3.7
    Requirement already satisfied: packaging>=14.3 in /usr/local/lib/python3.7/dist-packa
    Requirement already satisfied: pytz in /usr/local/lib/python3.7/dist-packages (from g
    Requirement already satisfied: googleapis-common-protos<2.0dev,>=1.6.0 in /usr/local/
    Requirement already satisfied: protobuf>=3.12.0 in /usr/local/lib/python3.7/dist-pack
    Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.7/dist-package
    Requirement already satisfied: pyasn1-modules>=0.2.1 in /usr/local/lib/python3.7/dist
    Requirement already satisfied: cachetools<5.0,>=2.0.0 in /usr/local/lib/python3.7/dis
    Requirement already satisfied: pyasn1>=0.1.7 in /usr/local/lib/python3.7/dist-package
    Requirement already satisfied: pyparsing>=2.0.2 in /usr/local/lib/python3.7/dist-pack
    Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-pac
    Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr/local/
    Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages
    Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-pa
    Requirement already satisfied: gensim in /usr/local/lib/python3.7/dist-packages (3.6.
    Requirement already satisfied: six>=1.5.0 in /usr/local/lib/python3.7/dist-packages (
    Requirement already satisfied: scipy>=0.18.1 in /usr/local/lib/python3.7/dist-package
    Requirement already satisfied: numpy>=1.11.3 in /usr/local/lib/python3.7/dist-package
    Requirement already satisfied: smart-open>=1.2.1 in /usr/local/lib/python3.7/dist-pac
    Collecting en_core_web_sm==2.2.5
      Downloading https://github.com/explosion/spacy-models/releases/download/en_core_web
        |████████████████████████████████| 12.0 MB 25.7 MB/s
    Requirement already satisfied: spacy>=2.2.2 in /usr/local/lib/python3.7/dist-packages
    Requirement already satisfied: setuptools in /usr/local/lib/python3.7/dist-packages (
    Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.7/dist-p
    Requirement already satisfied: thinc==7.4.0 in /usr/local/lib/python3.7/dist-packages
    Requirement already satisfied: wasabi<1.1.0,>=0.4.0 in /usr/local/lib/python3.7/dist-
    Requirement already satisfied: srsly<1.1.0,>=1.0.2 in /usr/local/lib/python3.7/dist-p
    Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.7/dist-p
    Requirement already satisfied: blis<0.5.0,>=0.4.0 in /usr/local/lib/python3.7/dist-pa
    Requirement already satisfied: numpy>=1.15.0 in /usr/local/lib/python3.7/dist-package
    Requirement already satisfied: plac<1.2.0,>=0.9.6 in /usr/local/lib/python3.7/dist-pa
    Requirement already satisfied: catalogue<1.1.0,>=0.0.7 in /usr/local/lib/python3.7/di
    Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.7/di
    Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.7/

    Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.7/dist
    Requirement already satisfied: importlib-metadata>=0.20 in /usr/local/lib/python3.7/d
```

```
Requirement already satisfied: zipp>=0.5 in /usr/local/lib/python3.7/dist-packages (f
Requirement already satisfied: typing-extensions>=3.6.4 in /usr/local/lib/python3.7/d
Requirement already satisfied: urllib3!=1.25.0,!=1.25.1,<1.26,>=1.21.1 in /usr/local/
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.7/dist-pa
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.7/dist-pac
✓ Download and installation successful
You can now load the model via spacy.load('en_core_web_sm')
✓ Linking successful
/usr/local/lib/python3.7/dist-packages/en_core_web_sm -->
/usr/local/lib/python3.7/dist-packages/spacy/data/en
You can now load the model via spacy.load('en')
```

```
import os
from pydrive.auth import GoogleAuth
from pydrive.drive import GoogleDrive
from google.colab import auth
from oauth2client.client import GoogleCredentials
import nltk
from nltk import FreqDist
nltk.download('stopwords')
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
True
```

```
!pip install pyLDAvis==3.2.2
```

```
Requirement already satisfied: pyLDAvis==3.2.2 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: joblib>=0.8.4 in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: numpy>=1.9.2 in /usr/local/lib/python3.7/dist-packages (f
Requirement already satisfied: numexpr in /usr/local/lib/python3.7/dist-packages (from p
Requirement already satisfied: pandas>=0.17.0 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: funcy in /usr/local/lib/python3.7/dist-packages (from pyL
Requirement already satisfied: jinja2>=2.7.2 in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: future in /usr/local/lib/python3.7/dist-packages (from py
Requirement already satisfied: wheel>=0.23.0 in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: scipy>=0.18.0 in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: MarkupSafe>=0.23 in /usr/local/lib/python3.7/dist-package
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.7/dist-p
Requirement already satisfied: pytz>=2017.2 in /usr/local/lib/python3.7/dist-packages (f
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (from
```

```
import pandas as pd
#pd.set_option("display.max_colwidth", 200)
import numpy as np
import json
import re
import gzip
import spacy

import gensim
import gensim
from gensim import corpora

import pyLDAvis
import pyLDAvis.gensim
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
/usr/local/lib/python3.7/dist-packages/past/types/oldstr.py:5: DeprecationWarning: Using
  from collections import Iterable
```

```
df=pd.read_csv('Responses.csv')
print(df.head(5))
```

```
   Resp ID  ...                                          Responses
0        1  ...  Who will responsible for students carrier beca...
1        2  ...  Taking ongoing online classes is quite tedious...
2        3  ...               Financial problem no network coverage
3        4  ...  In this pandemic no hope from anyone nobody ca...
4        5  ...  however my college is not helping/ demanding f...

[5 rows x 21 columns]
```

```
def parse(path):
  g = gzip.open(path, 'rb')
  for l in g:
    yield eval(l)

def getDF(path):
  i = 0
  df = {}
  for d in parse(path):
    df[i] = d
    i += 1
  return pd.DataFrame.from_dict(df, orient='index')
```

```


df.shape
```

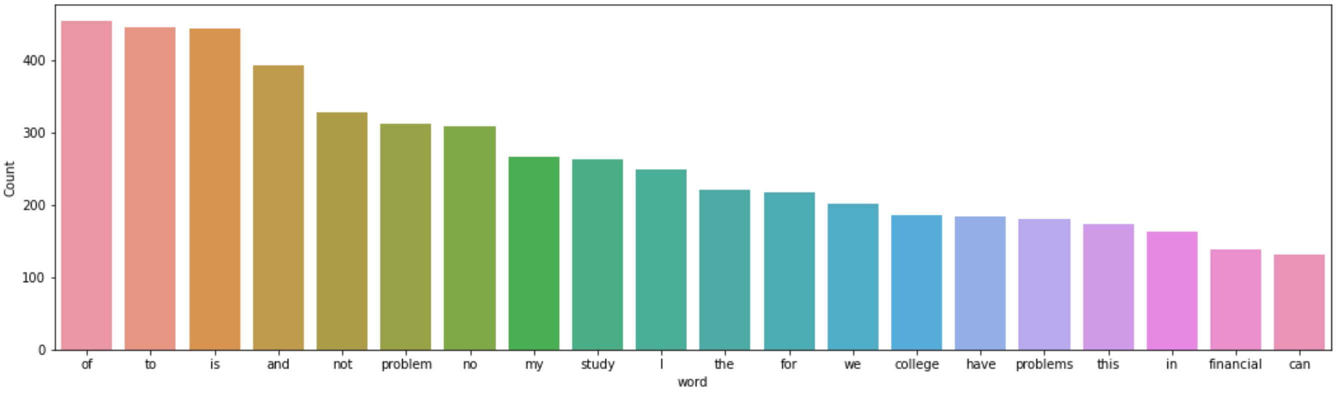
```
(1182, 21)
```

```
def freq_words(x, terms = 20):
  all_words = ' '.join([text for text in x])
  all_words = all_words.split()

  fdist = FreqDist(all_words)
  words_df = pd.DataFrame({'word':list(fdist.keys()), 'count':list(fdist.values())})

  # selecting top 20 most frequent words
  d = words_df.nlargest(columns="count", n = terms)
  plt.figure(figsize=(18,5))
  ax = sns.barplot(data=d, x= "word", y = "count")
  ax.set(ylabel = 'Count')
  plt.show()
```

```
freq_words(df['Responses'])
```



```
# replace "n't" with " not"
df['Responses'] = df['Responses'].str.replace("n\'t", " not")

# remove unwanted characters, numbers and symbols
df['Responses'] = df['Responses'].str.replace("[^a-zA-Z#]", " ")
```

```
from nltk.corpus import stopwords
stop_words = stopwords.words('english')
```

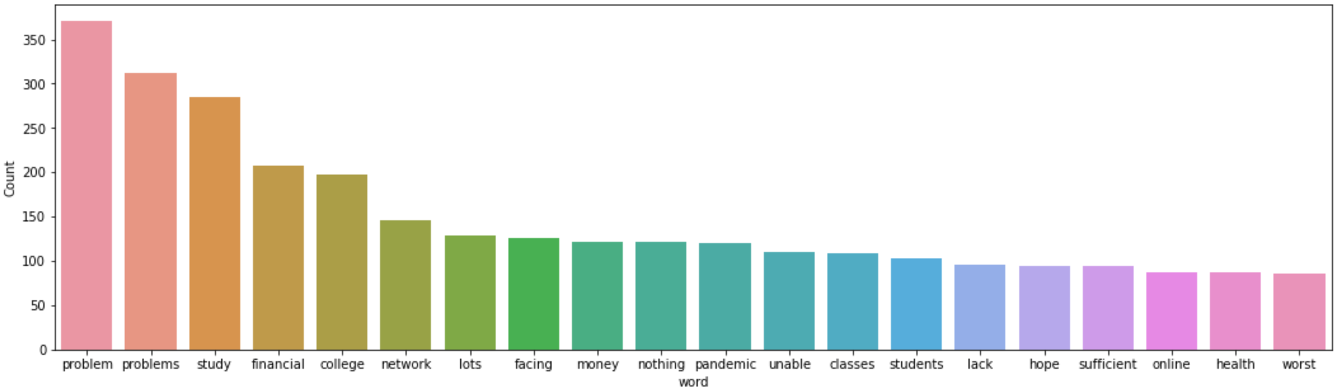
```
# function to remove stopwords
def remove_stopwords(rev):
  rev_new = " ".join([i for i in rev if i not in stop_words])
  return rev_new
```

```
# remove short words (length < 2)
df['Responses'] = df['Responses'].apply(lambda x: ' '.join([w for w in x.split() if len(w)>2]

# remove stopwords from the text
tweets = [remove_stopwords(r.split()) for r in df['Responses']]

# make entire text lowercase
tweets = [r.lower() for r in tweets]
```

```
freq_words(tweets, 20)
```



```
nlp = spacy.load('en', disable=['parser', 'ner'])
```

```
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectable(
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectable(
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:126: DeprecationWarning: Selectable(
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectable(
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
```

```
def lemmatization(texts, tags=['NOUN', 'ADJ']):
    output = []
    for sent in texts:
        doc = nlp(" ".join(sent))
        output.append([token.lemma_ for token in doc if token.pos_ in tags])
    return output
```

```
tokenized_tweets = pd.Series(tweets).apply(lambda x: x.split())
print(tokenized_tweets[1])
```

```
['taking', 'ongoing', 'online', 'classes', 'quite', 'tedious', 'another', 'issue', 'stay
```

```
print(tokenized_tweets[1])
len(tokenized_tweets[1])
```

```
['taking', 'ongoing', 'online', 'classes', 'quite', 'tedious', 'another', 'issue', 'stay
```

```
18
```

```
tweets_2 = lemmatization(tokenized_tweets)
print(tweets_2[1])
```

```
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
```

```
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
/usr/local/lib/python3.7/dist-packages/catalogue.py:138: DeprecationWarning: Selectab
  for entry_point in AVAILABLE_ENTRY_POINTS.get(self.entry_point_namespace, []):
```

```
print(tweets_2[3])
len(tweets_2[3])
```

```
set(tokenized_tweets[1]) - set(tweets_2[1])
```

```
tweets_3 = []
for i in range(len(tweets_2)):
    tweets_3.append(' '.join(tweets_2[i]))

df['tweets'] = tweets_3
```

```
freq_words(df['Responses'], 20)
```

```
# Create the term dictionary of our corpus, where every unique term is assigned an index
dictionary = corpora.Dictionary(tweets_2)
```

```
# Convert list of tweets (tweets_2) into a Document Term Matrix using the dictionary prepared
doc_term_matrix = [dictionary.doc2bow(rev) for rev in tweets_2]
```

```
# Creating the object for LDA model using gensim library
LDA = gensim.models.ldamodel.LdaModel
```

```
# Build LDA model
lda_model = LDA(corpus=doc_term_matrix,
                id2word=dictionary,
                num_topics=3,
                random_state=100,
                chunksize=50,
                passes=50)
```

```
lda_model.print_topics()
```

```
import pyLDAvis
import pyLDAvis.gensim
pyLDAvis.enable_notebook()
vis = pyLDAvis.gensim.prepare(lda_model, doc_term_matrix, dictionary)
```

```
vis
```

```
# Print the Keyword in the 10 topics
lda_model.print_topics()
```

```
#creating wordcloud
from wordcloud import WordCloud
WordCloud
from  nltk.corpus import stopwords #for stopwords
```

```
cloud  = WordCloud(max_words= 50, stopwords=stopwords.words("english") ).generate(str(tokeniz
plt.figure(figsize= (10 , 10))
plt.imshow(cloud)
```

! 36s completed at 5:40 PM ● ✕