

REG.	No. RA1911027010039
NAME	MAINAK CHAUDHURI
LAB	No. 9
TOPIC	Analysis using NLP

AIM : Classify complete and incomplete sentences.

1. Importing necessary packages for NLP

```
#import library
import numpy as np
import pandas as pd
import seaborn as sns
import nltk
nltk.download('stopwords')
nltk.download('wordnet')
from textblob import TextBlob
from wordcloud import WordCloud
import re
import string
string.punctuation
import matplotlib.pyplot as plt
plt.style.use('seaborn')
from sklearn.model_selection import train_test_split
from nltk.stem.porter import PorterStemmer
from nltk.stem import WordNetLemmatizer
from keras.preprocessing.text import Tokenizer
from keras.preprocessing.sequence import pad_sequences
from keras.models import Sequential
from keras.layers import Embedding, SimpleRNN, Dense, Dropout
from sklearn import metrics

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data] Unzipping corpora/wordnet.zip.
```

```
df = pd.read_json('https://raw.githubusercontent.com/Mainaknpositor/datasets/master/Finished_sentences.json')
```

Removal Punctuation

```
#defining to remove punctuation
def remove_punctuation(text):
    punctuationfree = "".join([i for i in text if i not in string.punctuation])
    return punctuationfree

text['sentence_clean'] = text['sentence'].apply(lambda text: remove_punctuation(text))
text.head()
```

	sentence	sentence_clean
0	Apple supplier AMS cuts forecast, indicating p...	Apple supplier AMS cuts forecast indicating po...
1	U.S. factory and services activity quicken in ...	US factory and services activity quicken in No...
2	Exclusive: Tesla expects global shortage of el...	Exclusive Tesla expects global shortage of ele...
3	World stocks climb on China trade relief, whil...	World stocks climb on China trade relief while...
4	Boeing, J&J, dismal China data drag Wall Stre...	Boeing JJ dismal China data drag Wall Street L...

Lower Casing

```
deflower_lower_case:
text['sentence_lower'] = text['sentence_clean'].str.lower()
text.head()
```

	sentence	sentence_clean	sentence_lower
0	Apple supplier AM5 cuts forecast, indicating p...	Apple supplier AM5 cuts forecast indicating po...	apple supplier am5 cuts forecast indicating po...
1	U.S. factory and services activity quicken in ...	US factory and services activity quicken in No...	us factory and services activity quicken in no...
2	Exclusive: Tesla expects global shortage of el...	Exclusive Tesla expects global shortage of ele...	exclusive tesla expects global shortage of ele...
3	World stocks climb on China trade relief, whil...	World stocks climb on China trade relief while...	world stocks climb on china trade relief while...
4	Boeing, J&J, dismal China data drag Wall Stre...	Boeing, J&J, dismal China data drag Wall Street L...	boeing j&j dismal china data drag wall street l...

Tokenization

```
defining function for tokenization
def tokenization(text):
    tokens = re.split('\s+', text)
    return tokens

Applying function to the column
text['sentence_tokenized'] = text['sentence_lower'].apply(lambda x: tokenization(x))
text.head()
```

	sentence	sentence_clean	sentence_lower	sentence_tokenized
0	Apple supplier AM5 cuts forecast, indicating p...	Apple supplier AM5 cuts forecast indicating po...	apple supplier am5 cuts forecast indicating po...	apple supplier am5 cuts forecast indicating po...
1	U.S. factory and services activity quicken in ...	US factory and services activity quicken in No...	us factory and services activity quicken in no...	us factory and services activity quicken in n...
2	Exclusive: Tesla expects global shortage of el...	Exclusive Tesla expects global shortage of ele...	exclusive tesla expects global shortage of ele...	exclusive tesla expects global shortage of el...
3	World stocks climb on China trade relief, whil...	World stocks climb on China trade relief while...	world stocks climb on china trade relief while...	world stocks climb on china trade relief whil...
4	Boeing, J&J, dismal China data drag Wall Stre...	Boeing, J&J, dismal China data drag Wall Street L...	boeing j&j dismal china data drag wall street l...	boeing j&j dismal china data drag wall street ...

Removal Stopwords

```
stop words present in the library
stopwords = nltk.corpus.stopwords.words('english')

defining the function to remove stopwords from tokenized text
def remove_stopwords(text):
    output = []
    for i in text:
        if i not in stopwords:
            return output

Applying the function
text['no_stopword'] = text['sentence_tokenized'].apply(lambda x:remove_stopwords(x))
text.head()
```

	sentence	sentence_clean	sentence_lower	sentence_tokenized	no_stopword
0	Apple supplier AM5 cuts forecast, indicating p...	Apple supplier AM5 cuts forecast indicating po...	apple supplier am5 cuts forecast indicating po...	apple supplier am5 cuts forecast indicating po...	apple supplier am5 cuts forecast indicating po...
1	U.S. factory and services activity quicken in ...	US factory and services activity quicken in No...	us factory and services activity quicken in no...	us factory and services activity quicken in n...	us factory and services activity quicken in n...
2	Exclusive: Tesla expects global shortage of el...	Exclusive Tesla expects global shortage of ele...	exclusive tesla expects global shortage of ele...	exclusive tesla expects global shortage of el...	exclusive tesla expects global shortage of el...
3	World stocks climb on China trade relief, whil...	World stocks climb on China trade relief while...	world stocks climb on china trade relief while...	world stocks climb on china trade relief whil...	world stocks climb on china trade relief whil...
4	Boeing, J&J, dismal China data drag Wall Stre...	Boeing, J&J, dismal China data drag Wall Street L...	boeing j&j dismal china data drag wall street l...	boeing j&j dismal china data drag wall street ...	boeing j&j dismal china data drag wall street ...

Stemming

```
defining the object for stemming
porter_stemmer = PorterStemmer()

defining a function for stemming
def stemming(text):
    stem_text = [porter_stemmer.stem(word) for word in text]
    return stem_text

Applying the function
text['sentence_stemmed'] = text['no_stopword'].apply(lambda x: stemming(x))
text.head()
```

	sentence	sentence_clean	sentence_lower	sentence_tokenized	no_stopword	sentence_stemmed
0	Apple supplier AM5 cuts forecast, indicating p...	Apple supplier AM5 cuts forecast indicating po...	apple supplier am5 cuts forecast indicating po...	apple supplier am5 cuts forecast indicating po...	apple supplier am5 cuts forecast indicating po...	apple supplier am5 cuts forecast indicating po...
1	U.S. factory and services activity quicken in ...	US factory and services activity quicken in No...	us factory and services activity quicken in no...	us factory and services activity quicken in n...	us factory and services activity quicken in n...	us factory and services activity quicken in n...
2	Exclusive: Tesla expects global shortage of el...	Exclusive Tesla expects global shortage of ele...	exclusive tesla expects global shortage of ele...	exclusive tesla expects global shortage of el...	exclusive tesla expects global shortage of el...	exclusive tesla expects global shortage of el...
3	World stocks climb on China trade relief, whil...	World stocks climb on China trade relief while...	world stocks climb on china trade relief while...	world stocks climb on china trade relief whil...	world stocks climb on china trade relief whil...	world stocks climb on china trade relief whil...
4	Boeing, J&J, dismal China data drag Wall Stre...	Boeing, J&J, dismal China data drag Wall Street L...	boeing j&j dismal china data drag wall street l...	boeing j&j dismal china data drag wall street ...	boeing j&j dismal china data drag wall street ...	boeing j&j dismal china data drag wall street ...

Lemmatizing

```
#defining the object for lemmatizing
lemmatizer = WordNetLemmatizer()

#defining a function for lemmatizing
def lemmatize_words(text):
    lemma_text = [lemmatizer.lemmatize(word) for word in text]
    return lemma_text

#applying the function
text['sentence_lemmatized'] = text['sentence_stemmed'].apply(lambda text: lemmatize_words(text))
text.head()
```

	sentence	sentence_clean	sentence_lower	sentence_tokenized	no_stopword	sentence_stemmed	sentence_lemmatized
0	Apple supplier AMS cuts forecast indicating po...	Apple supplier AMS cuts forecast indicating po...	apple supplier ams cuts forecast indicating po...	[apple, supplier, ams, cuts, forecast, indicating, po...	[apple, supplier, ams, cuts, forecast, indicating, po...	[apple, supplier, ams, cuts, forecast, indicating, po...	[apple, supplier, ams, cuts, forecast, indicating, po...
1	US factory and services activity quicken in No...	US factory and services activity quicken in No...	us factory and services activity quicken in no...	[us, factory, and, services, activity, quicken, in, n...	[us, factory, and, services, activity, quicken, in, n...	[us, factory, and, services, activity, quicken, in, n...	[us, factory, and, services, activity, quicken, in, n...
2	Exclusive Tesla expects global shortage of ele...	Exclusive Tesla expects global shortage of ele...	exclusive tesla expects global shortage of ele...	[exclusive, tesla, expects, global, shortage, of, ele...	[exclusive, tesla, expects, global, shortage, of, ele...	[exclusive, tesla, expects, global, shortage, of, ele...	[exclusive, tesla, expects, global, shortage, of, ele...
3	World stocks climb on China trade relief while...	World stocks climb on China trade relief while...	world stocks climb on china trade relief while...	[world, stocks, climb, on, china, trade, relief, whil...	[world, stocks, climb, on, china, trade, relief, whil...	[world, stocks, climb, on, china, trade, relief, whil...	[world, stocks, climb, on, china, trade, relief, whil...
4	Boeing JJ dismal China data drag Wall Street L...	Boeing JJ dismal China data drag Wall Street L...	boeing jj dismal china data drag wall street l...	[boeing, jj, dismal, china, data, drag, wall, street, l...	[boeing, jj, dismal, china, data, drag, wall, street, l...	[boeing, jj, dismal, china, data, drag, wall, street, l...	[boeing, jj, dismal, china, data, drag, wall, street, l...



Sentiment Analysis

```
df_new = text[['sentence_clean']]
df_new.head()
```

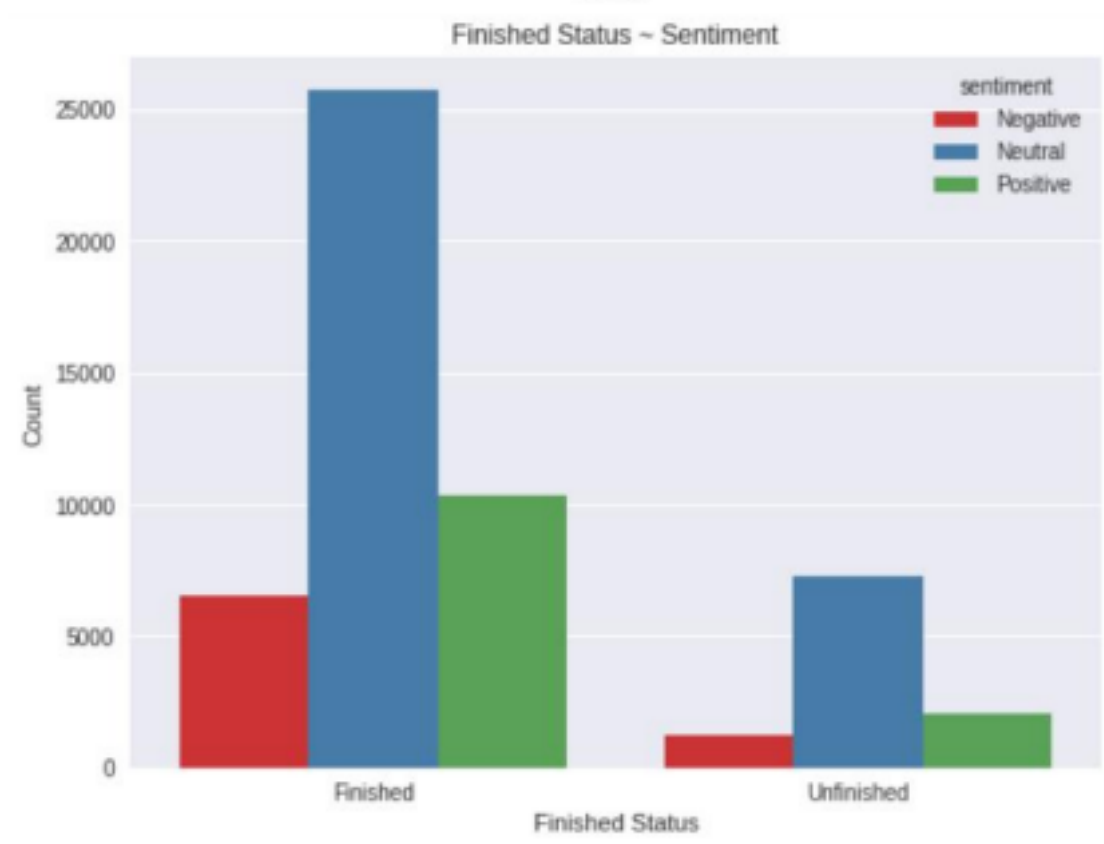
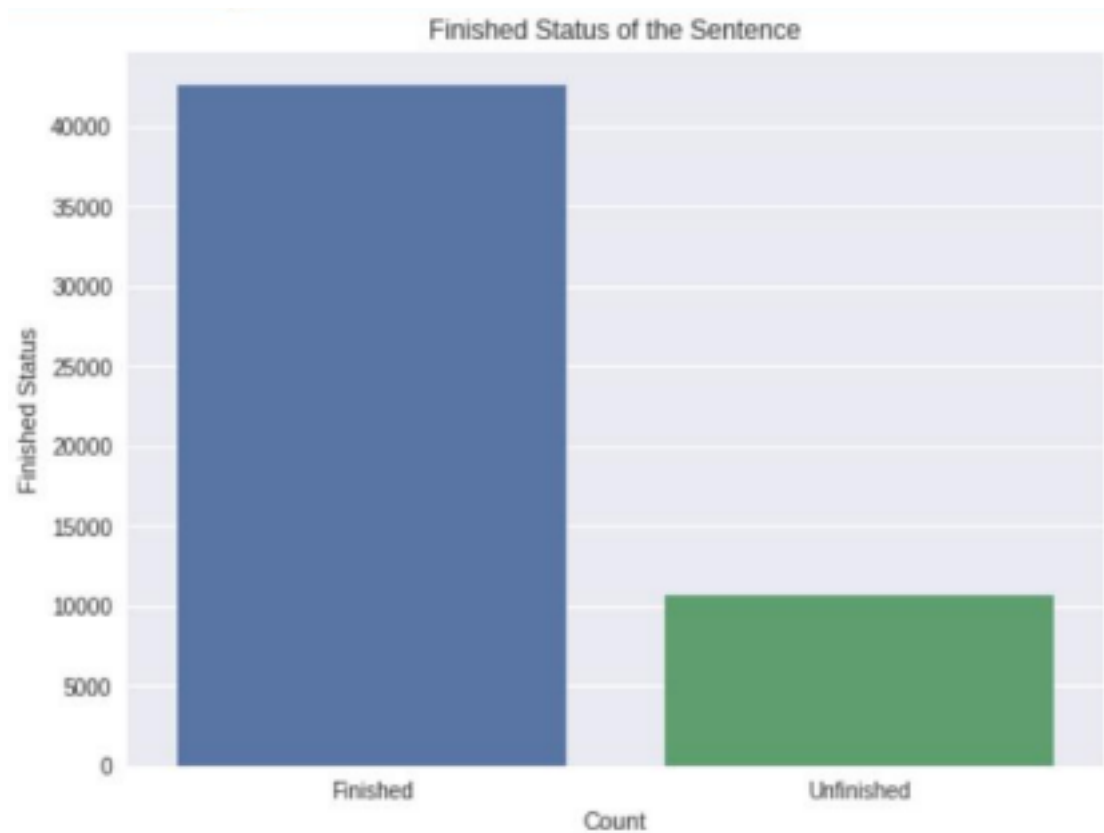
sentence_clean

```
0 Apple supplier AMS cuts forecast indicating po...
1 US factory and services activity quicken in No...
2 Exclusive Tesla expects global shortage of ele...
3 World stocks climb on China trade relief while...
4 Boeing JJ dismal China data drag Wall Street L...
```

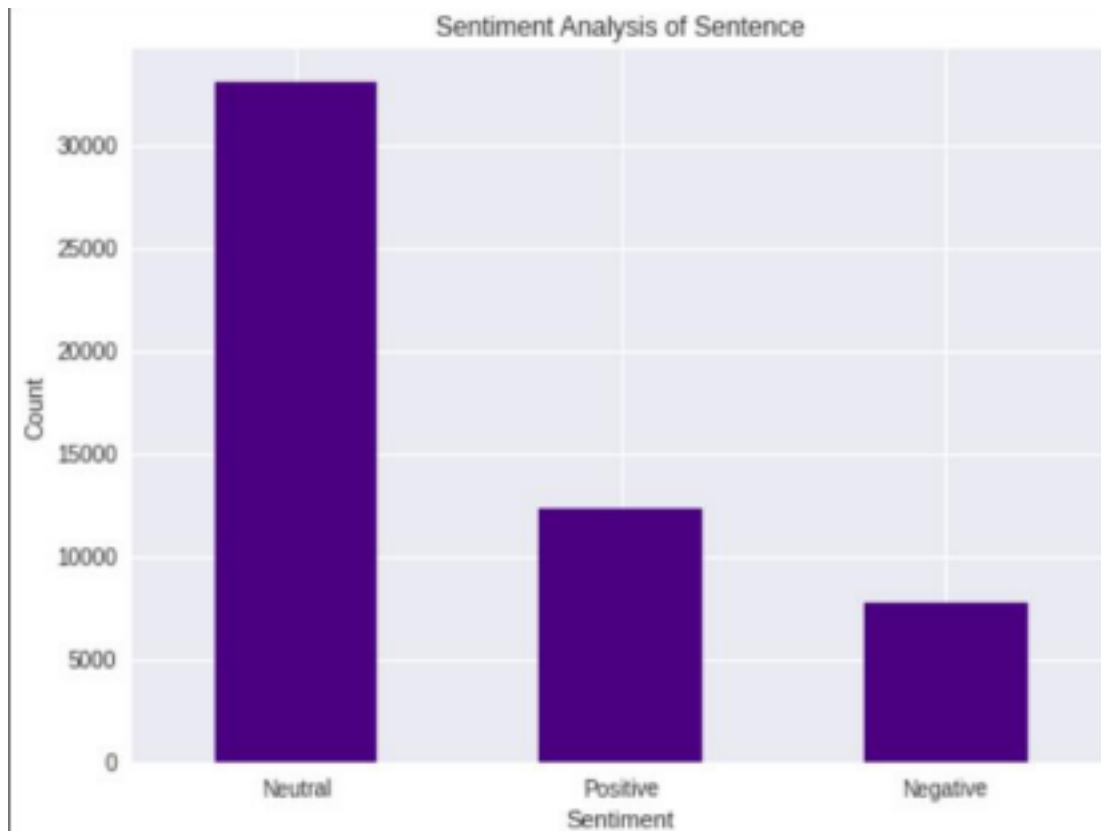
```
#create function to get subjectivity
def getSubjectivity(text):
    return TextBlob(text).sentiment.subjectivity

#create function to get polarity
def getPolarity(text):
    return TextBlob(text).sentiment.polarity

#apply function to data
df_new['subjectivity'] = df_new['sentence_clean'].apply(getSubjectivity)
df_new['polarity'] = df_new['sentence_clean'].apply(getPolarity)
df_new.head()
```



[illegible][illegible][illegible]



Reccurent Neural Network Model

```
#handling categorical data  
df['is_finished'] = df['is_finished'].astype('category').cat.codes  
df.head()
```

	sentence	is_finished
0	Apple supplier AMS cuts forecast, indicating p...	0
1	U.S. factory and services activity quicken in ...	0
2	Exclusive: Tesla expects global shortage of eL...	1
3	World stocks climb on China trade relief, whil...	0
4	Boeing, J&J, dismal China data drag Wall Stree...	0

Conclusion: Therefore the incomplete sentences are analysed with 91% accuracy and segmented properly.