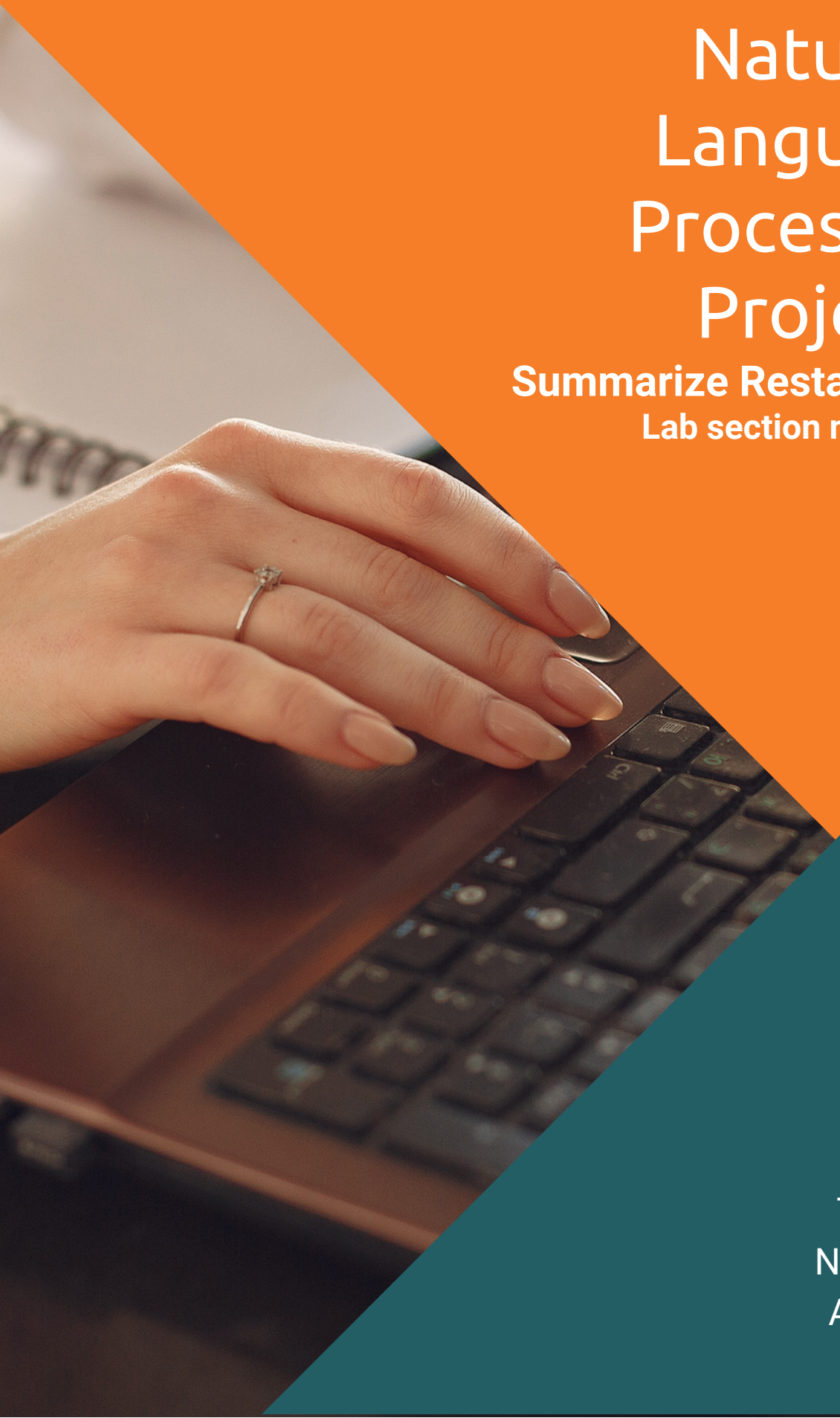
جامعة أم القرى
UMM AL-QURA UNIVERSITY



Natural Language Processing Project

Summarize Restaurant Reviews

Lab section number: 4

Prepared By:

Ohud Bukhari

Amjaad Saleh

Thekra Alhsani

Nosaibah Farhan

Afnan Farouqui

Team Members

Name	ID	Email	Task
Ohud Bukhari	441005142	s441005142@st.uqu.edu.sa	• Phase 1 (Chapter 1) • Bert Algorithm
Afnan Farouqui	441016120	s441016120@st.uqu.edu.as	• Phase 3 (chapter 3) • Filtering and reducing the dataset
Nosaibah Farhan	441016536	s441016536@st.uqu.edu.as	• Phase 2 (chapter 2)
Thekra Alhsani	437035939	s437035939@st.uqu.edu.sa	• Phase 3 (chapter 3)
Amjaad Saleh	441017415	s441017415@st.uqu.edu.sa	• Phase 3 (chapter 3) • Filtering and reducing the dataset

Chapter 1

01 Introduction

PROBLEM STATEMENT:

Millions of textual comments about goods and services are posted by customers and every day thousands are added, make it a big challenge to read and understand them to make them a useful structured data for customers and decision makers. Sentiment analysis or Opinion mining is a popular technique for summarizing and analyzing those opinions and reviews. We will use Natural Language Processing techniques to help us understand customer opinions and reviews (text comments) written in Arabic to understand customer trends.

MOTIVATION:

Since having a person read a large number of reviews is a time-consuming and tedious task. Building a sentiment analysis tool proved crucial to minimizing human participation in review analysis. A sentiment analysis tool makes the process totally automated, making it considerably faster and more accurate, providing real-time feedback that enables businesses to react quickly and enhance their services.

SUMMARY OF THE RESULTS:

To facilitate the process of understanding customer opinions and reviews, we will create a system that analyzes customer opinions and classifies them into positive or negative opinions.

PROJECT CODE:

Enter here

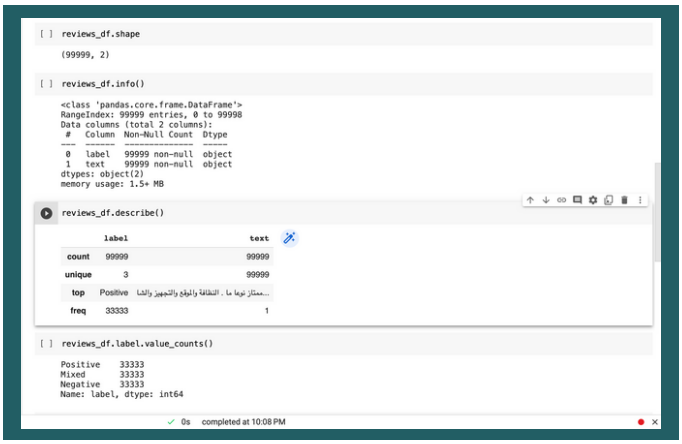
PROJECT DATASET:

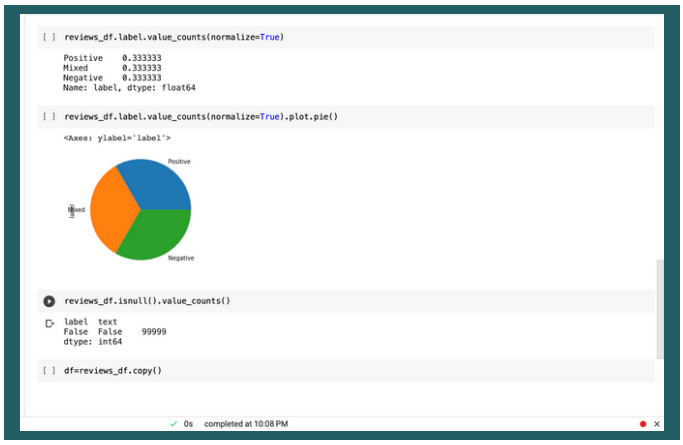
Enter here

Dataset details 02

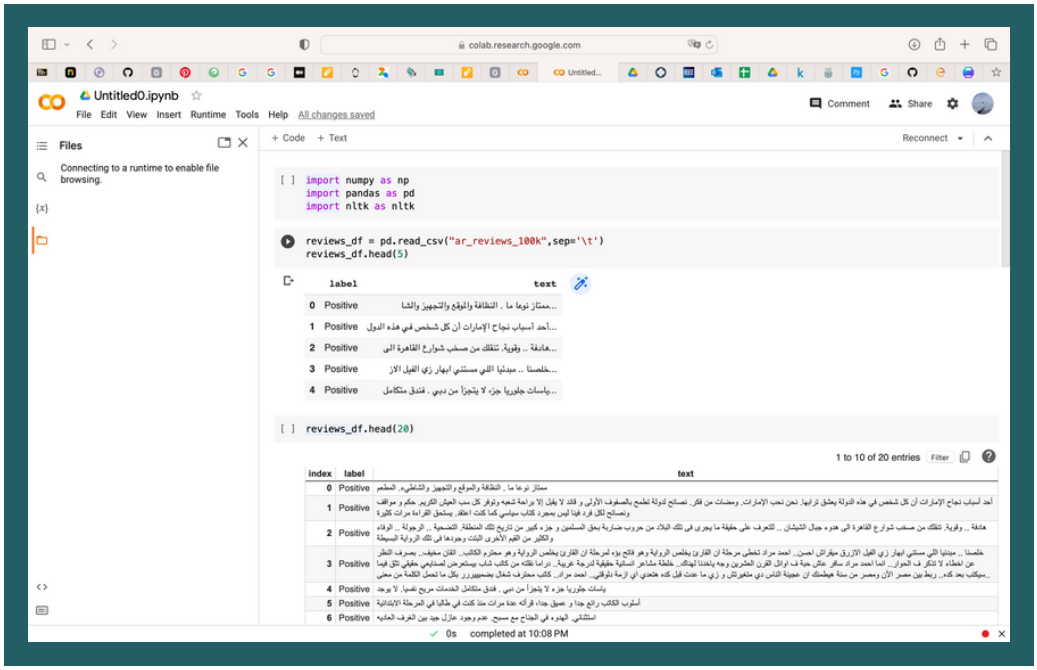
DESCRIPTION:

The dataset used in this project was taken from Kaggle website, it is Arabic datasets are available for classification comparison and other NLP tasks. This dataset is mainly a compilation of several available datasets and a sampling of 9999 rows of reviews, it contains two column, one to describe the reviews and the another for the reviews.





READING AND LOADING DATA SET IN PYTHON:



Filtering and Reducing

03

DESCRIPTION:

The dataset used in this project contained reviews about various things. this step ensures that the focus is only on book/ novel reviews.

Choose only Books/Novels entries/Reduce dataset

```
[6] ## Function for selecting books/novels reviews only

    from nltk.tree.tree import Nonterminal
    def books(text):
      for i in text.split(" "):
        if i in ['رواية','الرواية','كتاب','الكتاب','الكاتب','المؤلف','الحبكة']:
          return text
      return None
```

```
## remove all the None rows from reviews_df
reviews_df['text']=reviews_df['text'].apply(lambda x: books(x))
reviews_df = reviews_df.dropna(axis = 0)
reviews_df.reset_index()
reviews_df.head(10)
```

	label	text
0	Mixed	...هو بلا شك جدل لا ينتهي إلا ليبدأ، جدل يتصارع د
1	Negative	...لم اكن قد قرأت نصف الرواية حين اشمأزيت من كثره
2	Negative	...لو ده يعتبر كتاب ، فأنا علي كده كتبت كتب كتيره
3	Mixed	... في ذلك الكتاب للرائع أحمد بهجت رحمة الله عليه

SAMPLING THE DATASET

the books dataset had over 32000 records which delayed the processing and caused the session to crash multiple times. The following code takes 13% of the original data while keeping the label ratios even. then stores the sampled reviews in a new csv file.

```
## Reducing the size of the dataset since the Lemmatize function takes too long
# Create a new dataframe with the minimum number of rows for each label
reduced_df = reviews_df.sample(frac=0.13)
reviews_df = reduced_df
# Save the reduced dataframe
reduced_df.to_csv("reduced_dataset.csv", index=False)
```

Chapter 2

• REMOVING PUNCTUATION, STANDARDISATION

Second Phase

1.Removing punctuation,Standardisation

```
[111] import string
      string.punctuation

'!"#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~'

[112] def remove_punct(text):
          punc= string.punctuation+"،؟“”؛؟"
          text_no_punct="".join([char for char in text if char not in punc])
          return text_no_punct

[113] reviews_df['text_no_punct']=reviews_df['text'].apply(lambda x: remove_punct(x))
      reviews_df.head()
```

	label	text	text_no_punct
0	Positive	...ممتاز نوعا ما . النظافة والموقع والتجهيز والشا	...ممتاز نوعا ما النظافة والموقع والتجهيز والشاط
1	Positive	...أحد أسباب نجاح الإمارات أن كل شخص في هذه الدول	...أحد أسباب نجاح الإمارات أن كل شخص في هذه الدول
2	Positive	...هادفة .. وقوية. تنقلك من صخب شوارع القاهرة الى	...هادفة وقوية تنقلك من صخب شوارع القاهرة الى هد
3	Positive	...خلصنا .. مبدئيا اللي مستني ابهار زي الفيل الاز	...خلصنا مبدئيا اللي مستني ابهار زي الفيل الازرق
4	Positive	...ياسات جلوريا جزء لا يتجزأ من دبي . فندق متكامل	... ياسات جلوريا جزء لا يتجزأ من دبي فندق متكامل

To delete the punctuation marks, we imported the String library, then called the punctuation marks, and then created a function (remove_punct) in it for a loop to delete the punctuation marks if they are in the text. Then we added a new column to the dataset with the text applied to it (remove_punct).

• TOKENIZATION USING: 1- REGEX AND SPLIT 2- NLTK:

2.Tokenization using: 1- regex and split 2- NLTK

```
[171] #Tokeniziation (White-Spaceing)
      nltk.download('punkt')
      from nltk.tokenize import word_tokenize

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

reviews_df['text_tokenize']=reviews_df['text'].apply(lambda x: word_tokenize(x))
reviews_df.head()
```

	label	text	text_no_punct	text_tokenize
0	Positive	...ممتاز نوعا ما . النظافة والموقع والتجهيز والشا	...ممتاز نوعا ما النظافة والموقع والتجهيز والشاط	...ممتاز, نوعا, ما, ., النظافة, والموقع, والتجهي
1	Positive	...أحد أسباب نجاح الإمارات أن كل شخص في هذه الدول	...أحد أسباب نجاح الإمارات أن كل شخص في هذه الدول	... ,أحد, أسباب, نجاح, الإمارات, أن, كل, شخص, في
2	Positive	...هادفة .. وقوية. تنقلك من صخب شوارع القاهرة الى	...هادفة وقوية تنقلك من صخب شوارع القاهرة الى هد	...هادفة, وقوية, ., تنقلك, من, صخب, شوارع, ا
3	Positive	...خلصنا .. مبدئيا اللي مستني ابهار زي الفيل الاز	...خلصنا مبدئيا اللي مستني ابهار زي الفيل الازرق	...خلصنا, ..., مبدئيا, اللي, مستني, ابهار, زي, ال
4	Positive	...ياسات جلوريا جزء لا يتجزأ من دبي . فندق متكامل	... ياسات جلوريا جزء لا يتجزأ من دبي فندق متكامل	...ياسات, جلوريا, جزء, لا, يتجزأ, من, دبي, ., فن

import word_tokenize function used for (white-spaces) from tokenize library ,and adding the text_tokenize column, which is the text_no_punct column applied to it word_tokenize .

And then adding the text_tokenize_re column, which is the text_no_punct column applied to it split function from RE library.

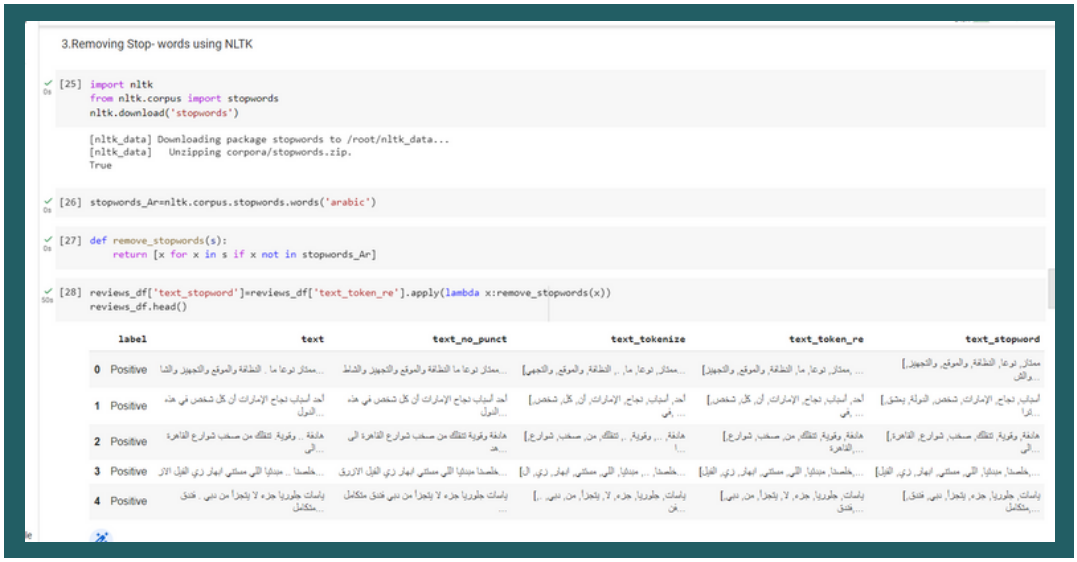
```
[23] import re

[24] reviews_df['text_token_re']=reviews_df['text_no_punct'].apply(lambda x:re.split('\W+', x))
     reviews_df.head()
```

	label	text	text_no_punct	text_tokenize	text_token_re
0	Positive	...ممتاز نوعا ما . النظافة والموقع والتجهيز والشا	...ممتاز نوعا ما النظافة والموقع والتجهيز والشاط	...ممتاز, نوعا, ما, ., النظافة, والموقع, والتجهي	...ممتاز, نوعا, ما, النظافة, والموقع, والتجهيز]
1	Positive	...أحد أسباب نجاح الإمارات أن كل شخص في هذه الدول	...أحد أسباب نجاح الإمارات أن كل شخص في هذه الدول	... ,أحد, أسباب, نجاح, الإمارات, أن, كل, شخص, في	... ,أحد, أسباب, نجاح, الإمارات, أن, كل, شخص, في]
2	Positive	...هادفة .. وقوية. تنقلك من صخب شوارع القاهرة الى	...هادفة وقوية تنقلك من صخب شوارع القاهرة الى هد	...هادفة, وقوية, ., تنقلك, من, صخب, شوارع, ا	...هادفة, وقوية, تنقلك, من, صخب, شوارع, القاهرة]
3	Positive	...خلصنا .. مبدئيا اللي مستني ابهار زي الفيل الاز	...خلصنا مبدئيا اللي مستني ابهار زي الفيل الازرق	...خلصنا, ..., مبدئيا, اللي, مستني, ابهار, زي, ال	...خلصنا, مبدئيا, اللي, مستني, ابهار, زي, الفيل]
4	Positive	...ياسات جلوريا جزء لا يتجزأ من دبي . فندق متكامل	... ياسات جلوريا جزء لا يتجزأ من دبي فندق متكامل	...ياسات, جلوريا, جزء, لا, يتجزأ, من, دبي, ., فن	...ياسات, جلوريا, جزء, لا, يتجزأ, من, دبي, فندق]

• REMOVING STOP- WORDS USING NLTK

import stopwords library from the script that under the nltk library, after that call the words function from stopwords and passed to it 'arabic' and set its values to stopwords_Ar. Then we created a function remove _stopwords that goes through each index and deletes stopwords if it exists. Add a text_stopword coulom and applay function remove _stopwords to it .



3.Removing Stop- words using NLTK

```
[25] import nltk
     from nltk.corpus import stopwords
     nltk.download('stopwords')

     [nltk_data] Downloading package stopwords to /root/nltk_data...
     [nltk_data]   Unzipping corpora/stopwords.zip.
     True

[26] stopwords_Ar=nltk.corpus.stopwords.words('arabic')

[27] def remove_stopwords(s):
         return [x for x in s if x not in stopwords_Ar]

[28] reviews_df['text_stopword']=reviews_df['text_token_re'].apply(lambda x:remove_stopwords(x))
     reviews_df.head()
```

	label	text	text_no_punct	text_tokenize	text_token_re	text_stopword
0	Positive	ممتاز نوعا ما . النظافة والموقع والتجهيز والش...	ممتاز نوعا ما النظافة والموقع والتجهيز والشاط...	[ممتاز, نوعا, ما, ., ., النظافة, والموقع, والتجهي...	[ممتاز, نوعا, ما, النظافة, والموقع, والتجهيز, ...	[ممتاز, نوعا, النظافة, والموقع, والتجهيز, والش...
1	Positive	أحد أسباب نجاح الإمارات أن كل شخص في هذه الدول...	أحد أسباب نجاح الإمارات أن كل شخص في هذه الدول...	[أحد, أسباب, نجاح, الإمارات, أن, كل, شخص, في, ...	[أحد, أسباب, نجاح, الإمارات, أن, كل, شخص, في, ...	[أسباب, نجاح, الإمارات, شخص, الدولة, يعشق, ترا...
2	Positive	هادفة .. وقوية تتفلك من صخب شوارع القاهرة الى ...	هادفة وقوية تتفلك من صخب شوارع القاهرة الى هد...	[هادفة, ..., وقوية, ., تتفلك, من, صخب, شوارع, ا...	[هادفة, وقوية, تتفلك, من, صخب, شوارع, القاهرة,...	[هادفة, وقوية, تتفلك, صخب, شوارع, القاهرة, الى...
3	Positive	خلصنا .. مبدئيا اللى مستنى ابهار زي الفيل الاز...	خلصنا مبدئيا اللى مستنى ابهار زي الفيل الازرق...	[خلصنا, ., ., مبدئيا, اللى, مستنى, ابهار, زي, ال...	[خلصنا, مبدئيا, اللى, مستنى, ابهار, زي, الفيل,...	[خلصنا, مبدئيا, اللى, مستنى, ابهار, زي, الفيل,...
4	Positive	ياسات جلوريا جزء لا يتجزأ من دبى . فندق متكامل...	ياسات جلوريا جزء لا يتجزأ من دبى فندق متكامل ...	[ياسات, جلوريا, جزء, لا, يتجزأ, من, دبى, ., ., ف...	[ياسات, جلوريا, جزء, لا, يتجزأ, من, دبى, ., فند...	[ياسات, جلوريا, جزء, لا, يتجزأ, من, دبى, فندق,...

• STEMMING

ISRIStemmer in isri is called under stem in nltk library. A variable p_stemmer of type stemmer is defined and an ISRIStemmer function is assigned to it. Create a stem function and let the variable p_stemmer do the stemming for the token. Create a text_stems column, which is a text_stopword column applay to it stem function.

4.Stemming

```
[29] import nltk
     from nltk.stem.isri import ISRIStemmer
     p_stemmer = ISRIStemmer()

[30] def stem(s):
       return [p_stemmer.stem(token) for token in s]

     reviews_df['text_stems'] = reviews_df['text_stopword'].apply(lambda x: stem(x))
     reviews_df.head(5)
```

	label	text	text_no_punct	text_tokenize	text_token_re	text_stopword	text_stems
0	Positive	ممتاز نوعا ما . النظافة والموقع والتجهيز والش...	ممتاز نوعا ما النظافة والموقع والتجهيز والشاط...	[ممتاز, نوعا, ما, ., ., النظافة, والموقع, والتجهي...	[ممتاز, نوعا, ما, النظافة, والموقع, والتجهيز, ...	[ممتاز, نوعا, النظافة, والموقع, والتجهيز, والش...	[ملز, نوع, نظف, وقع, جهز, شاطئ, [ظعم
1	Positive	أحد أسباب نجاح الإمارات أن كل شخص في هذه الدول...	أحد أسباب نجاح الإمارات أن كل شخص في هذه الدول...	[أحد, أسباب, نجاح, الإمارات, أن, كل, شخص, في, ...	[أحد, أسباب, نجاح, الإمارات, أن, كل, شخص, في, ...	[أسباب, نجاح, الإمارات, شخص, الدولة, يعشق, ترا...	[سبب, نجح, امر, شخص, دول, ... ,عشق, ترب, نحب, امر
2	Positive	هادفة .. وقوية تتفلك من صخب شوارع القاهرة الى ...	هادفة وقوية تتفلك من صخب شوارع القاهرة الى هد...	[هادفة, ..., وقوية, ., تتفلك, من, صخب, شوارع, ا...	[هادفة, وقوية, تتفلك, من, صخب, شوارع, القاهرة,...	[هادفة, وقوية, تتفلك, صخب, شوارع, القاهرة, الى...	[هدف, وقي, نقل, صخب, شرع, قهر, ... ,الى, هذ, جبل
3	Positive	خلصنا .. مبدئيا اللى مستنى ابهار زي الفيل الاز...	خلصنا مبدئيا اللى مستنى ابهار زي الفيل الازرق...	[خلصنا, ., ., مبدئيا, اللى, مستنى, ابهار, زي, ال...	[خلصنا, مبدئيا, اللى, مستنى, ابهار, زي, الفيل,...	[خلصنا, مبدئيا, اللى, مستنى, ابهار, زي, الفيل,...	[خلص, بدئ, الل, مست, بهر, زي, ... ,فيل, زرق, ميفراش
4	Positive	ياسات جلوريا جزء لا يتجزأ من دبى . فندق متكامل...	ياسات جلوريا جزء لا يتجزأ من دبى فندق متكامل ...	[ياسات, جلوريا, جزء, لا, يتجزأ, من, دبى, ., ., ف...	[ياسات, جلوريا, جزء, لا, يتجزأ, من, دبى, ., فند...	[ياسات, جلوريا, جزء, يتجزأ, دبى, ... ,فندق, متكامل	[ياس, جلر, جزء, جزأ, دبي, لدق, ... ,كامل, خدم, ريح

• LEMMATIZATION

IDownload qalsadi decisionry and summon lemmatizer from it. Define a lemmer variable of type lemmatizer that is assigned to the lemmatizer function located under the lemmatizer library. Create a lemtaization function and let the variable lemmer do the lemmatizer for the token. Create a text_lemmatize column, which is a text_stopword column applay to it lemtaization function.

*We have deleted the records from the dataset to complete the lemmatizer process, because it took a long time and was not executed .

5.Lemmatization

```
[208] !pip install qalsadi

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: qalsadi in /usr/local/lib/python3.9/dist-packages (0.4.6)
Requirement already satisfied: six>=1.10.0 in /usr/local/lib/python3.9/dist-packages (from qalsadi) (1.16.0)
Requirement already satisfied: pyarabic>=0.6.7 in /usr/local/lib/python3.9/dist-packages (from qalsadi) (0.6.15)
Requirement already satisfied: arramooz-pysqlite>=0.3 in /usr/local/lib/python3.9/dist-packages (from qalsadi) (0.4.1)
Requirement already satisfied: pickledb>=0.9.2 in /usr/local/lib/python3.9/dist-packages (from qalsadi) (0.9.2)
Requirement already satisfied: alyahmor>=0.1 in /usr/local/lib/python3.9/dist-packages (from qalsadi) (0.2)
Requirement already satisfied: tashaphyne>=0.3.4.1 in /usr/local/lib/python3.9/dist-packages (from qalsadi) (0.3.6)
Requirement already satisfied: libqutrub>=1.2.3 in /usr/local/lib/python3.9/dist-packages (from qalsadi) (1.2.4.1)
Requirement already satisfied: future>=0.16.0 in /usr/local/lib/python3.9/dist-packages (from qalsadi) (0.18.3)
Requirement already satisfied: naftawayh>=0.3 in /usr/local/lib/python3.9/dist-packages (from qalsadi) (0.4)
Requirement already satisfied: Arabic-Stopwords>=0.3 in /usr/local/lib/python3.9/dist-packages (from qalsadi) (0.4.3)

[209] import qalsadi.lemmatizer

[210] lemmer = qalsadi.lemmatizer.Lemmatizer()
      lemmer.lemmatize("")

'' Exception ignored in: <function WordFreqDictionary.__del__ at 0x7f67393a8ee0>
Traceback (most recent call last):
  File "/usr/local/lib/python3.9/dist-packages/arramooz/wordfreqdictionaryclass.py", line 130, in __del__
    self.db_connect.close()
sqlite3.ProgrammingError: SQLite objects created in a thread can only be used in that same thread. The object was created in thread id 140082713130816 and this is thread id 140081817966336.
```

```
[211] def lemtaization(s):
        return [lemmer.lemmatize(token) for token in s]

reviews_df['text_lemmatize']= reviews_df['text_stopword'].apply(lambda x:lemtaization(x))
reviews_df.head()
```

	label	text	text_no_punct	text_tokenize	text_token_re	text_stopword	text_stems	text_lemmatize
99499	Negative	يستاهل نجوم بس بالسالب	يستاهل نجوم بس بالسالب	[... ,يستاهل, نجوم, بس, بالسالب]	[يستاهل, نجوم, بس, بالسالب]	[يستاهل, نجوم, بالسالب]	[نجل, نجم, سلب]	[يستاهل, نجوم, سالب]
99500	Negative	لم تكن رواية عن هيباتيا بقدر ما هي مجموعة من ا...	لم تكن رواية عن هيباتيا بقدر ما هي مجموعة من ا...	[لم, تكن, رواية, عن, هيباتيا, بقدر, ما, ...هي, مج]	[لم, تكن, رواية, عن, هيباتيا, بقدر, ما, ...هي, مج]	[تكن, رواية, هيباتيا, بقدر, هي, ...مجموعة, المحاضن]	[تكن, روي, هيب, قدر, هي, جمع, حضر, [لسف ...,محاضر]	[تكن, رواية, هيباتيا, قدر, هي, مجموع, ...,محاضر]
99501	Negative	قعدت كتير أتهم نفسي بالغباء وقلة الاطلاع لأني ...	قعدت كتير أتهم نفسي بالغباء وقلة الاطلاع لأني ...	[قعدت, كتير, أتهم, نفسي, بالغباء, وقلة, ...الاطلا]	[قعدت, كتير, أتهم, نفسي, بالغباء, وقلة, ...الاطلا]	[قعدت, كتير, أتهم, نفسي, بالغباء, وقلة, ...الاطلا]	[قعد, كتر, تهم, نفس, غبء, وقل, طلع, ...لأن, مفهمت]	[قعد, كتير, تهم, نفس, غباء, قلة, اطلاع, ...لأن, م]
99502	Negative	الطباعي عن الرواية لم يكن جيدا. يمكن ما صوره لي	الطباعي عن الرواية لم يكن جيدا يمكن ما صوره لي	[الطباعي, عن, الرواية, لم, يكن, جيدا, ...يمكن, ما]	[الطباعي, عن, الرواية, لم, يكن, جيدا, ...يمكن, ما]	[الطباعي, الرواية, يكن, جيدا, يمكن, ...صوره, لي]	[طبع, روي, يكن, جيد, عكس, صور, لي, ...يمكن, روي, ح]	[الطباع, رواية, كتي, جيد, عكس, صور, ...لي, يمكن, ر]
99503	Negative	ضعيف جدا. لاشي ينذكر. كل شي	ضعيف جدا لاشي ينذكر كل شي	[ضعيف, جدا, ., لاشي, ينذكر, ., كل, شي]	[ضعيف, جدا, لاشي, ينذكر, كل, شي]	[ضعيف, جدا, لاشي, ينذكر, شي]	[ضعف, جدا, لشي, ذكر, شي]	[ضعيف, جدا, لاشي, ذكر, شي]

Chapter 3

• JOIN TOKENS

rejoin the tokens as a clean text to use it in countVector and TF_IDF to extract feature.

```
def join_text(set):
    text = ""
    for word in set:
        text += word + " "
    return text
reviews_df['text_cleaned'] = reviews_df['text_stems'].apply(lambda x: join_text(x))
```

• COUNT VECTOR

We use the count vector feature to count the number of times a word is repeated in each record. The fit_transform method is used to convert a collection of text documents into a matrix of token counts. And The pd.DataFrame(features_cv.toarray()) method is used to convert a sparse matrix into a dense matrix. The sparse matrix is a matrix that has mostly zeros. This is because most words in a document do not appear very often. The dense matrix is a matrix that has all of its values filled in.

```
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.naive_bayes import MultinomialNB
from sklearn.svm import SVC

# Count vector
from sklearn .feature_extraction.text import  CountVectorizer
count_vector = CountVectorizer()
features_cv = count_vector.fit_transform(reviews_df['text_cleaned'])
features_cv = pd.DataFrame(features_cv.toarray())
```

• TF-IDF VECTOR

We use the TF-IDF vector to calculate the weights of words by taking into account both the number of times a word appears in the dataset and the number of records in which it appears. In order to determine the important words for machine training.

```
## tf/idf vector
from sklearn.feature_extraction.text import TfidfVectorizer
tfidf = TfidfVectorizer()
features_tfidf = tfidf.fit_transform(reviews_df['text_cleaned'])
features_tfidf = pd.DataFrame(features_tfidf.toarray())
```

• MACHINE LEARNING USING RANDOM FOREST.

Before training the machine we remove all unnecessary columns and concatenate the two vectors countVector and TF-IDF vector together to increase the accuracy.

```
# remove all unnecessery columns
target = reviews_df['label']
reviews_df = df.drop(['label','text'], inplace=True, axis=1)

## concatenate the two vectors
reviews_df = pd.concat([reviews_df,pd.DataFrame(features_cv)], axis = 1)
reviews_df = pd.concat([reviews_df,pd.DataFrame(features_tfidf)], axis = 1)
```

After that we split the data for training size 75% and for test size 25% .

```
##split data for training test size 25%
X_train, X_test, y_train, y_test = train_test_split(reviews_df, target, test_size=0.25)
```

After training the machine using the random forest model, the accuracy was 54.38%, which is less than the perfect accuracy, which can be improved by increasing the amount of data, but we dispensed with that because it takes a lot of space and time.

```
# Train the model
model = RandomForestClassifier(n_jobs=-1)
model.fit(X_train, y_train)

# Evaluate the model on the test set
accuracy = model.score(X_test, y_test)
print('Accuracy: {0:.2f}%'.format(accuracy*100

Accuracy: 54.38%
```