

- [tubeComments.rar](#) Arabic sentiment for YouTube comments

Prediction accuracy 83%

very large file	tubeComments.rar	499,746,673	WinRAR archive
	tubeComments		File folder
	nltk_packages		File folder
	google_api_python_cl...		File folder
	arabic_natural_langua...		File folder

- 1 First, install MongoDB from MongoDB.com

```
https://www.youtube.com/watch?v=ckA1JWBTB4  
https://www.youtube.com/watch?v=p-EdBcGuKuk  
https://www.youtube.com/watch?v=s0Joygr_ILc  
https://www.youtube.com/watch?v=ZkLI1_mN_Uc  
https://www.youtube.com/watch?v=9S50dXEv41k  
https://www.youtube.com/watch?v=2ygKU0GyvFU
```

2

input VideoID.txt

```
ckA1JWBTB4  
p-EdBcGuKuk  
s0Joygr_ILc  
ZkLI1_mN_Uc  
9S50dXEv41k  
2ygKU0GyvFU
```

3

run

RUN.analyze.bat

4

output

```
[nltk_data] Downloading package punkt to nltk_packages...  
[nltk_data] Package punkt is already up-to-date!  
[nltk_data] Downloading package stopwords to nltk_packages...  
[nltk_data] Package stopwords is already up-to-date!  
- Collecting comments of [redacted] video...  
10643 comments found and added to db  
- Finding sentiment...  
10643 comments were analyzed.  
Positive comments: 8548 = 80.32%  
Negative comments: 2095 = 19.68%
```

83% accuracy


```

Python 3.8.6 Shell
File Edit Shell Debug Options Window Help
الجامعة' اليك' الخطاب' كم' ونفاد' سهل' احسن' جديدة' ر'
لاعتقد' احتيا' اريد' ابداء' ينبغي' والنطق' تضحك' امن'
غنام' النوار' اُحِبُّكَ' آي' المدرب' بالرياض' ماعليك' ا' اشارك' ال
عروض' ا' ا' ا' ا' ا' ا' ا' ا' ا' ا' ا' ا' ا' ا' ا' ا' ا' ا' ا' ا' ا' a'
[اخذ' ويوم' كون' اجماعة' الديرسي' اطلع' بيننا' ا' ا' ا' a'
-----
generating features for documents ...
splitting documents into train and test ...
data set size 21083
data set: Counter({'neg': 14335, 'pos': 6748})
train: Counter({'neg': 11530, 'pos': 5336})
test: Counter({'neg': 2805, 'pos': 1412})
training NaiveBayes classifier ...
training is done
accuracy: 0.8937633388664927
positive f-score: 0.848135593220339
negative f-score: 0.9183078045222465
Most Informative Features
contains(☺) = True          pos : neg = 516.4 : 1.0
contains(☹) = True          pos : neg = 332.0 : 1.0
contains(:) = True          pos : neg = 155.0 : 1.0
contains(☺) = True          pos : neg = 144.8 : 1.0
contains(نشر) = True        neg : pos = 128.8 : 1.0
contains(☺) = True          pos : neg = 109.0 : 1.0
contains(☹) = True          neg : pos = 105.7 : 1.0
contains(☺) = True          pos : neg = 92.7 : 1.0
contains(☹) = True          neg : pos = 89.5 : 1.0
contains(☹) = True          neg : pos = 84.9 : 1.0
contains(☺) = True          neg : pos = 78.4 : 1.0
contains(+غذاء) = True      neg : pos = 74.5 : 1.0
contains(المنتجات) = True  pos : neg = 69.3 : 1.0
contains(تصل) = True        pos : neg = 68.3 : 1.0
contains(حسابك) = True      neg : pos = 55.6 : 1.0
[REDACTED] = True           neg : pos = 55.5 : 1.0
contains(☺) = True          neg : pos = 55.5 : 1.0
contains(وشكرا) = True      neg : pos = 49.0 : 1.0
contains(☺) = True          neg : pos = 47.7 : 1.0
contains(يقوم) = True       neg : pos = 45.3 : 1.0
>>>
Ln: 67 Col: 4

```

```

Python 3.8.6 Shell
File Edit Shell Debug Options Window Help
training is done
Most Informative Features
has(موتق) = True           pos : neg = 238.5 : 1.0
has(☺) = True              neg : pos = 202.0 : 1.0
has(☺) = True              neg : pos = 171.3 : 1.0
[REDACTED] = True          pos : neg = 170.1 : 1.0
has(وصلوا) = True         pos : neg = 166.9 : 1.0
has(☹) = True              neg : pos = 164.2 : 1.0
has(الشروط) = True        pos : neg = 151.4 : 1.0
has(وتابع) = True          pos : neg = 143.9 : 1.0
has(بيكي) = True           neg : pos = 143.6 : 1.0
has(☹) = True              neg : pos = 132.4 : 1.0
has(السحب) = True         pos : neg = 118.4 : 1.0
has(☹) = True              pos : neg = 116.5 : 1.0
[REDACTED] = True          neg : pos = 113.3 : 1.0
has(يشجعون) = True         pos : neg = 100.5 : 1.0
has(ركبت) = True           neg : pos = 100.2 : 1.0
has(بضل) = True           neg : pos = 98.8 : 1.0
has(نشبة) = True          neg : pos = 97.4 : 1.0
has(☺) = True              neg : pos = 97.3 : 1.0
has(☹) = True              neg : pos = 97.2 : 1.0
has(وكانهم) = True        neg : pos = 94.7 : 1.0
has(☺) = True              neg : pos = 91.9 : 1.0
has(☺) = True              neg : pos = 91.3 : 1.0
has(:) = True              neg : pos = 87.5 : 1.0
has(☹) = True              neg : pos = 85.0 : 1.0
has(شديدة) = True         neg : pos = 82.4 : 1.0
has(الزرقاء) = True       pos : neg = 82.3 : 1.0
has(توجيه) = True         neg : pos = 79.5 : 1.0
[REDACTED] = True          neg : pos = 78.8 : 1.0
has(فص) = True             neg : pos = 78.1 : 1.0
has(بمناسبة) = True       pos : neg = 74.9 : 1.0
has(☹) = True              neg : pos = 71.3 : 1.0
has(انظر) = True          pos : neg = 70.7 : 1.0
has(☹) = True              neg : pos = 67.5 : 1.0
has(قيمنه) = True         pos : neg = 67.5 : 1.0
has(☺) = True              neg : pos = 66.7 : 1.0
has(☺) = True              neg : pos = 66.0 : 1.0
has(الايفون) = True      pos : neg = 62.9 : 1.0
has(مبلغ) = True          pos : neg = 62.1 : 1.0
has(المسابقة) = True     pos : neg = 61.6 : 1.0
has(♥) = True              pos : neg = 61.0 : 1.0
accuracy: 0.8913283975831844
accuracy: 0.8913283975831844
Ln: 509795 Col: 0

```

✓ has (نشبة) = True

neg : pos = 97.4 : 1.0

```
Python 3.8.6 Shell
training is done
Most Informative Features
has(موتق) = True          pos : neg = 238.5 : 1.0
has(Ⓜ) = True            neg : pos = 202.0 : 1.0
has(Ⓜ) = True            neg : pos = 171.3 : 1.0
has(Ⓜ) = True            pos : neg = 170.1 : 1.0
has(وسلوا) = True        pos : neg = 166.9 : 1.0
has(Ⓜ) = True            neg : pos = 164.2 : 1.0
has(الشروط) = True       pos : neg = 151.4 : 1.0
has(وتابع) = True        pos : neg = 143.9 : 1.0
has(يسكي) = True         neg : pos = 143.6 : 1.0
has(Ⓜ) = True            neg : pos = 132.4 : 1.0
has(النسحب) = True       pos : neg = 118.4 : 1.0
has(Ⓜ) = True            pos : neg = 116.5 : 1.0
has(العروبر) = True       neg : pos = 113.3 : 1.0
has(Ⓜ) = True            pos : neg = 100.5 : 1.0
has(ركبتا) = True        neg : pos = 100.2 : 1.0
has(يخلل) = True         neg : pos = 98.8 : 1.0
has(نشبة) = True         neg : pos = 97.4 : 1.0
has(Ⓜ) = True            neg : pos = 97.3 : 1.0
has(Ⓜ) = True            neg : pos = 97.2 : 1.0
has(وكانهم) = True       neg : pos = 94.7 : 1.0
has(Ⓜ) = True            neg : pos = 91.9 : 1.0
has(Ⓜ) = True            neg : pos = 91.3 : 1.0
has(:) = True            neg : pos = 87.5 : 1.0
has(Ⓜ) = True            neg : pos = 85.0 : 1.0
has(ثديدا) = True        neg : pos = 82.4 : 1.0
has(بمناسبة) = True     pos : neg = 74.9 : 1.0
has(Ⓜ) = True            neg : pos = 71.3 : 1.0
has(انظر) = True        pos : neg = 70.7 : 1.0
has(Ⓜ) = True            neg : pos = 67.5 : 1.0
has(قيمتنا) = True       pos : neg = 67.5 : 1.0
has(Ⓜ) = True            neg : pos = 66.7 : 1.0
has(Ⓜ) = True            neg : pos = 66.0 : 1.0
has(الايون) = True       pos : neg = 62.9 : 1.0
has(مبلغ) = True         pos : neg = 62.1 : 1.0
has(المصايغة) = True      pos : neg = 61.6 : 1.0
has(Ⓜ) = True            pos : neg = 61.0 : 1.0
Ln: 509795 Col: 0
```

✗ has (وكانهم) = True

neg : pos = 94.7 : 1.0