Twitter Sentiment Analysis

File Edit View Insert Runtime Tools Help Last saved at 1:55 AM

Comment Share

+ Code + Text

RAM Disk Colab AI

```
#Installing the Kaggle Library
! pip install kaggle
```

```
Requirement already satisfied: kaggle in /usr/local/lib/python3.10/dist-packages (1.5.16)
Requirement already satisfied: six>=1.10 in /usr/local/lib/python3.10/dist-packages (from kaggle) (1.16.0)
Requirement already satisfied: certifi in /usr/local/lib/python3.10/dist-packages (from kaggle) (2023.11.17)
Requirement already satisfied: python-dateutil in /usr/local/lib/python3.10/dist-packages (from kaggle) (2.8.2)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from kaggle) (2.31.0)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from kaggle) (4.66.1)
Requirement already satisfied: python-slugify in /usr/local/lib/python3.10/dist-packages (from kaggle) (8.0.1)
Requirement already satisfied: urllib3 in /usr/local/lib/python3.10/dist-packages (from kaggle) (2.0.7)
Requirement already satisfied: bleach in /usr/local/lib/python3.10/dist-packages (from kaggle) (6.1.0)
Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-packages (from bleach->kaggle) (0.5.1)
Requirement already satisfied: text-unidecode>=1.3 in /usr/local/lib/python3.10/dist-packages (from python-slugify->kaggle) (1.3)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests->kaggle) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->kaggle) (3.6)
```

Upload your Kaggle.json file

```
[103] # configuring the path of Kaggle.json file
      !mkdir -p ~/.kaggle
      !cp kaggle.json ~/.kaggle/
      !chmod 600 ~/.kaggle/kaggle.json
```

```
[104] # API to fetch the dataset from kaggle
      !kaggle datasets download -d kazanova/sentiment140
```

```
sentiment140.zip: Skipping, found more recently modified local copy (use --force to force download)
```

Importing Twitter Sentiment Dataset

```
[105] #extracting the compressed dataset

      from zipfile import ZipFile
      dataset = '/content/sentiment140.zip'

      with ZipFile(dataset,'r') as zip:
        zip.extractall()
      print ('The dataset is extracted')
```

```
The dataset is extracted
```

```
[106] ! pip install kaggle
```

```
Requirement already satisfied: kaggle in /usr/local/lib/python3.10/dist-packages (1.5.16)
Requirement already satisfied: six>=1.10 in /usr/local/lib/python3.10/dist-packages (from kaggle) (1.16.0)
Requirement already satisfied: certifi in /usr/local/lib/python3.10/dist-packages (from kaggle) (2023.11.17)
Requirement already satisfied: python-dateutil in /usr/local/lib/python3.10/dist-packages (from kaggle) (2.8.2)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from kaggle) (2.31.0)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages (from kaggle) (4.66.1)
Requirement already satisfied: python-slugify in /usr/local/lib/python3.10/dist-packages (from kaggle) (8.0.1)
Requirement already satisfied: urllib3 in /usr/local/lib/python3.10/dist-packages (from kaggle) (2.0.7)
Requirement already satisfied: bleach in /usr/local/lib/python3.10/dist-packages (from kaggle) (6.1.0)
Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-packages (from bleach->kaggle) (0.5.1)
Requirement already satisfied: text-unidecode>=1.3 in /usr/local/lib/python3.10/dist-packages (from python-slugify->kaggle) (1.3)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests->kaggle) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->kaggle) (3.6)
```

Uploading your Kaggle.json file

Twitter Sentiment Analysis

File Edit View Insert Runtime Tools Help Last saved at 1:55 AM

Comment Share

+ Code + Text

RAM Disk Colab AI

```
[107] # configuring the path of Kaggle.json file
      !mkdir -p ~/.kaggle
      !cp kaggle.json ~/.kaggle/
      !chmod 600 ~/.kaggle/kaggle.json
```

Importing Twitter Sentiment Dataset

```
[108] # API to fetch the dataset from google
      !kaggle datasets download -d kazanova/sentiment140
```

```
sentiment140.zip: Skipping, found more recently modified local copy (use --force to force download)
```

```
[109] #extracting the compressed dataset

      from zipfile import ZipFile
      dataset = '/content/sentiment140.zip'

      with ZipFile(dataset,'r') as zip:
        zip.extractall()
      print ('The dataset is extracted')
```

```
The dataset is extracted
```

```
[110] import numpy as np
      import pandas as pd
      import re
      from nltk.corpus import stopwords
      from nltk.stem.porter import PorterStemmer
      from sklearn.feature_extraction.text import TfidfVectorizer
      from sklearn.model_selection import train_test_split
      from sklearn.linear_model import LogisticRegression
```

Connected to Python 3 Google Compute Engine backend

Twitter Sentiment Analysis

File Edit View Insert Runtime Tools Help Last saved at 1:55 AM

Comment Share

+ Code + Text

RAM Disk Colab AI

```
[110] from sklearn.model_selection import train_test_split
      from sklearn.linear_model import LogisticRegression
      from sklearn.metrics import accuracy_score
```

```
[111] import nltk
      nltk.download('stopwords')
```

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

```
[112] #printing the stopwords in english
      print (stopwords.words('english'))
```

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'you', "you're", "you've", "you'll", "you'd", 'your', 'yours', 'yourself', 'yourselves', 'he', 'him', 'his', 'himself', 'she', "she's", 'her', '
```

```
[113] #loading the data from csv file to pands dataframe
      twitter_data = pd.read_csv("/content/training.1600000.processed.noemoticon.csv",encoding = 'ISO-8859-1')
```

```
[114] # checking the number of rows and colums
      twitter_data.shape
```

```
(1599999, 6)
```

```
[115] #printing the first 5 rows of the dataframe
      twitter_data.head()
```

Twitter Sentiment Analysis

File Edit View Insert Runtime Tools Help Last saved at 1:55 AM

Comment Share

+ Code + Text

RAM Disk Colab AI

[115]

	0	1467810369	Mon Apr 06 22:19:45 PDT 2009	NO_QUERY	_TheSpecialOne_	@switchfoot http://twitpic.com/2y1zl - Awww, that's a bummer. You shoulda got David Carr of Third Day to do it. ;D
0	0	1467810672	Mon Apr 06 22:19:49 PDT 2009	NO_QUERY	scotthamilton	is upset that he can't update his Facebook by ...
1	0	1467810917	Mon Apr 06 22:19:53 PDT 2009	NO_QUERY	mattycus	@Kenichan I dived many times for the ball. Man...
2	0	1467811184	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	ElleCTF	my whole body feels itchy and like its on fire
3	0	1467811193	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	Karoli	@nationwideclass no, it's not behaving at all....
4	0	1467811372	Mon Apr 06 22:20:00 PDT 2009	NO_QUERY	joy_wolf	@Kwesidei not the whole crew

[116]
```
# naming the columns and reading the dataset again

column_names = ['target', 'id', 'date', 'flag','user','text']
twitter_data = pd.read_csv ('/content/training.1600000.processed.noemoticon.csv', names=column_names, encoding  = 'ISO-8859-1')
```

```
# checking the number of rows and colums
twitter_data.shape
```

```
(1600000, 6)
```

[118]
```
#printing the first 5 rows of the dataframe
twitter_data.head()
```

	target	id	date	flag	user	text
0	0	1467810369	Mon Apr 06 22:19:45 PDT 2009	NO_QUERY	_TheSpecialOne_	@switchfoot http://twitpic.com/2y1zl - Awww, t...
1	0	1467810672	Mon Apr 06 22:19:49 PDT 2009	NO_QUERY	scotthamilton	is upset that he can't update his Facebook by ...
2	0	1467810917	Mon Apr 06 22:19:53 PDT 2009	NO_QUERY	mattycus	@Kenichan I dived many times for the ball. Man...
3	0	1467811184	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	ElleCTF	my whole body feels itchy and like its on fire

Connected to Python 3 Google Compute Engine backend

Twitter Sentiment Analysis

File Edit View Insert Runtime Tools Help Last saved at 1:55AM

Comment Share

+ Code + Text

RAM Disk Colab AI

```
[119] #counting the number of missing values in the dataset
      twitter_data.isnull().sum()
```

```
target    0
id        0
date      0
flag      0
user      0
text      0
dtype: int64
```

```
[120] # checking the distribution of target cloumn
      twitter_data['target'].value_counts()
```

```
0    800000
4    800000
Name: target, dtype: int64
```

```
[121] twitter_data.replace({'target':{4:1}}, inplace=True)
```

```
      # checking the distribution of target cloumn
      twitter_data['target'].value_counts()
```

```
0    800000
1    800000
Name: target, dtype: int64
```

```
[123] port_stem = PorterStemmer()
```

```
[124] def stemming (content):
```

Connected to Python 3 Google Compute Engine backend

Twitter Sentiment Analysis

File Edit View Insert Runtime Tools Help Last saved at 1:55 AM

Comment Share

+ Code + Text

RAM Disk Colab AI

```
[124]  stemmed_content = re.sub('[^a-zA-Z]',' ',content)
       stemmed_content = stemmed_content.lower()
       stemmed_content = stemmed_content.split()
       stemmed_content = [port_stem.stem(word)for word in stemmed_content if not word in stopwords.words('english')]
       stemmed_content = ' '.join(stemmed_content)

       return stemmed_content
```

```
[125] import nltk
      from nltk.stem import PorterStemmer

      stemmer = PorterStemmer()

      def stemming(text):
          return " ".join([stemmer.stem(word) for word in text.split()])
```

```
[126] twitter_data['stemmed_content'] = twitter_data['text'].apply(stemming)
```

```
[127] twitter_data.head()
```

	target	id	date	flag	user	text	stemmed_content
0	0	1467810369	Mon Apr 06 22:19:45 PDT 2009	NO_QUERY	_TheSpecialOne_	@switchfoot http://twitpic.com/2y1zl - Awww, t...	@switchfoot http://twitpic.com/2y1zl - awww, t...
1	0	1467810672	Mon Apr 06 22:19:49 PDT 2009	NO_QUERY	scotthamilton	is upset that he can't update his Facebook by ...	is upset that he can't updat hi facebook by te...
2	0	1467810917	Mon Apr 06 22:19:53 PDT 2009	NO_QUERY	mattycus	@Kenichan I dived many times for the ball. Man...	@kenichan i dive mani time for the ball. manag...
3	0	1467811184	Mon Apr 06 22:19:57 PDT 2009	NO_QUERY	ElleCTF	my whole body feels itchy and like its on fire	my whole bodi feel itchi and like it on fire

Connected to Python 3 Google Compute Engine backend

Twitter Sentiment Analysis

File Edit View Insert Runtime Tools Help Last saved at 1:55 AM

Comment Share

+ Code + Text

RAM Disk Colab AI

```
[128] print(twitter_data['stemmed_content'])
```

```
0          @switchfoot http://twitpic.com/2y1zl - awww, t...
1          is upset that he can't updat hi facebook by te...
2          @kenichan i dive mani time for the ball. manag...
3               my whole bodi feel itchi and like it on fire
4          @nationwideclass no, it' not behav at all. i'm...
                                ...
1599995    just woke up. have no school is the best feel ...
1599996    thewdb.com - veri cool to hear old walt interv...
1599997    are you readi for your mojo makeover? ask me f...
1599998    happi 38th birthday to my boo of alll time!!! ...
1599999    happi #charitytuesday @thenspcc @sparkschar @s...
Name: stemmed_content, Length: 1600000, dtype: object
```

```
[129] print(twitter_data['stemmed_content'])
```

```
0          @switchfoot http://twitpic.com/2y1zl - awww, t...
1          is upset that he can't updat hi facebook by te...
2          @kenichan i dive mani time for the ball. manag...
3               my whole bodi feel itchi and like it on fire
4          @nationwideclass no, it' not behav at all. i'm...
                                ...
1599995    just woke up. have no school is the best feel ...
1599996    thewdb.com - veri cool to hear old walt interv...
1599997    are you readi for your mojo makeover? ask me f...
1599998    happi 38th birthday to my boo of alll time!!! ...
1599999    happi #charitytuesday @thenspcc @sparkschar @s...
Name: stemmed_content, Length: 1600000, dtype: object
```

```
[130] print(twitter_data['target'])
```

```
0          0
1          0
```

Connected to Python 3 Google Compute Engine backend

Twitter Sentiment Analysis

File Edit View Insert Runtime Tools Help Last saved at 1:55 AM

Comment Share

+ Code + Text

RAM Disk Colab AI

[130]
```
print(twitter_data['target'])
```

```
0          0
1          0
2          0
3          0
4          0
          ..
1599995    1
1599996    1
1599997    1
1599998    1
1599999    1
Name: target, Length: 1600000, dtype: int64
```

[131]
```
#separating the data and label
X = twitter_data['stemmed_content'].values
Y = twitter_data['target'].values
```

[132]
```
print(X)
```

```
["@switchfoot http://twitpic.com/2y1zl - awww, that' a bummer. you shoulda got david carr of third day to do it. ;d"
 "is upset that he can't updat hi facebook by text it... and might cri as a result school today also. blah!"
 '@kenichan i dive mani time for the ball. manag to save 50% the rest go out of bound'
 ... 'are you readi for your mojo makeover? ask me for detail'
 'happi 38th birthday to my boo of alll time!!! tupac amaru shakur'
 'happi #charitytuesday @thenspcc @sparkschar @speakinguph4h']
```

[133]
```
print(Y)
```

```
[0 0 0 ... 1 1 1]
```

Connected to Python 3 Google Compute Engine backend

splitting the data to training data and test data

```
[134] X_train, X_test, Y_train, Y_test = train_test_split(X,Y, test_size=0.2, stratify=Y,  random_state=2)
```

```
[135] print(X.shape, X_train.shape, X_test.shape)

      (1600000,) (1280000,) (320000,)
```

```
[136] print(X_train)

      ['about to watch saw iv and drink a lil wine' "@hatermagazin i'm in!"
       'even though it my favourit drink i think it the vodka and coke that wipe my mind all the time think im gonna have to find a new drink'
       ... 'is eager for monday afternoon'
       "hope everyon and their mother had a great day! can't wait to hear what the guy have in store tomorrow!"
       'i love wake up to folgers. too bad my voic wa deeper than his.']
```

```
[137] print(X_test)


      ["@mmangen - m do fine! i haven't had much time to chat on twitter. hubbi is back for the summer &amp;amp; tend to domin my free time."
       'at ah may show w/ ruth kim &amp;amp; geoffrey sanhueza'
       '@ishatara mayb it wa onli a bay area thang dammit!' ...
       '@destini41 nevertheless hooray! for 4700 member and have a wonder and safe trip!'
       'not feel too well' '@supersandro !!!! thank you!']
```

Feature extraction vectorizer

```
[138] # converting textual data to numarical data

      vectorizer = TfidfVectorizer()
```

Twitter Sentiment Analysis

File Edit View Insert Runtime Tools Help Last saved at 1:55AM

Comment Share

+ Code + Text

RAM Disk Colab AI

```
X_train = vectorizer.fit_transform(X_train)
X_test = vectorizer.transform(X_test)
```

```
[139] print(X_train)
```

```
  (0, 555970)   0.42532987400715944
  (0, 311959)   0.39943217862098934
  (0, 166830)   0.3586993165413134
  (0, 56947)    0.1541598714398116
  (0, 254267)   0.4975163889508718
  (0, 453119)   0.3393001361168507
  (0, 548737)   0.2582573574441168
  (0, 519169)   0.1251682928604393
  (0, 40279)    0.24921248655385633
  (1, 247278)   0.21546651956326146
  (1, 226262)   0.9765112282750745
  (2, 375366)   0.15418764469918453
  (2, 193769)   0.1867791654247997
  (2, 226497)   0.11093606860080028
  (2, 214788)   0.17300813303941862
  (2, 245846)   0.14905964709328745
  (2, 516781)   0.14121557901439757
  (2, 51092)    0.13232152007961004
  (2, 349419)   0.22287868000539435
  (2, 556379)   0.31105619176256455
  (2, 507481)   0.10868332108979104
  (2, 129252)   0.28852240010234204
  (2, 545131)   0.30348832488514077
  (2, 507873)   0.15389117177429465
  (2, 513658)   0.2962743072556722
  :     :
  (1279998, 184747)     0.27940108825167487
  (1279998, 552329)     0.20454757140378815
  (1279998, 236205)     0.2214524643637042
  (1279998, 222374)     0.2167702817289201
```

Connected to Python 3 Google Compute Engine backend

Twitter Sentiment Analysis

File Edit View Insert Runtime Tools Help Last saved at 1:55 AM

Comment Share

+ Code + Text

RAM Disk Colab AI

```
[140] print(X_test)
```

```
  (0, 530652)   0.16406640641232267
  (0, 519169)   0.15246547985671288
  (0, 516781)   0.28905256268144397
  (0, 507873)   0.07874952233647492
  (0, 505261)   0.3173064852404718
  (0, 492694)   0.20033417376282853
  (0, 390585)   0.11211826947842242
  (0, 366485)   0.09288398012183556
  (0, 364066)   0.16204963433757727
  (0, 354510)   0.41225351040602914
  (0, 251929)   0.10105303919622356
  (0, 238369)   0.2631863878189709
  (0, 226531)   0.2118117681451351
  (0, 222374)   0.15352997153281683
  (0, 200572)   0.21642485492921723
  (0, 198156)   0.10402819171551929
  (0, 193829)   0.23231905667160468
  (0, 163021)   0.34277901517152526
  (0, 161994)   0.1364625632059021
  (0, 117476)   0.2448178674664643
  (0, 74691)    0.14771930001474182
  (0, 55622)    0.15630416888403628
  (1, 466638)   0.24069634981937746
  (1, 445346)   0.47513615904510464
  (1, 337825)   0.2801236227174083
  :     :
  (319995, 220218)      0.2337302958701797
  (319995, 166915)      0.2564683351198874
  (319995, 165114)      0.2831014966255927
  (319996, 571349)      0.17456322686703027
  (319996, 551020)      0.40253287464031445
  (319996, 503610)      0.8466100675416142
  (319996, 44865)       0.301241362028454
  (319997, 557810)      0.2168598849701581
  (319997, 525589)      0.23609255174637112
```

Connected to Python 3 Google Compute Engine backend

Training the Machine Learning Model

```
[141] model = LogisticRegression(max_iter=1000)
```

```
[142] model.fit(X_train, Y_train)
```

Model Evaluation

Accuracy score

```
[143] #accuracy score on the training data
      X_train_prediction = model.predict(X_train)
      training_data_accuracy = accuracy_score(Y_train, X_train_prediction)
```

```
[144] print('Accuracy score on the training data :', training_data_accuracy)
```

```
[145]  # accuracy score on the test data
       X_test_predection = model.predict(X_test)
       test_data_accuracy = accuracy_score(Y_test, X_test_predection)
```

```
[146] print('Accuracy score on the training data :', test_data_accuracy)
```

Model Accuracy is 79.8 %

Saving the Trained Model

Twitter Sentiment Analysis

File Edit View Insert Runtime Tools Help Last saved at 1:55 AM

Comment Share

+ Code + Text

RAM Disk Colab AI

```
[147] import pickle
```

```
[148] filename = 'trained_model.sav'
      pickle.dump(model,open(filename,'wb'))
```

Using the saved model for future predections

```
[149] #loading the saved model
      Loaded_model = pickle.load(open('/content/trained_model.sav','rb'))
```

```
X_new = X_test[200]
print(Y_test[200])

prediction = model.predict(X_new)
print(prediction)

if (prediction[0] ==0 ):
   print('Negative Tweet')

else:
  print ('Positive Tweet')
```

```
[151] X_new = X_test[3]
      print(Y_test[3])

      prediction = model.predict(X_new)
      print(prediction)
```

Connected to Python 3 Google Compute Engine backend

Twitter Sentiment Analysis

File Edit View Insert Runtime Tools Help Last saved at 1:55 AM

Comment Share

+ Code + Text

RAM Disk Colab AI

Using the saved model for future predections

```
[149] #loading the saved model
      Loaded_model = pickle.load(open('/content/trained_model.sav','rb'))
```

```
[150] X_new = X_test[200]
      print(Y_test[200])

      prediction = model.predict(X_new)
      print(prediction)

      if (prediction[0] ==0 ):
         print('Negative Tweet')

      else:
        print ('Positive Tweet')
```

```
[151] X_new = X_test[3]
      print(Y_test[3])

      prediction = model.predict(X_new)
      print(prediction)

      if (prediction[0] ==0 ):
         print('Negative Tweet')

      else:
        print ('Positive Tweet')
```

Connected to Python 3 Google Compute Engine backend